

Det här verket har digitaliserats vid Göteborgs universitetsbibliotek.  
Alla tryckta texter är OCR-tolkade till maskinläsbar text. Det betyder att du kan söka och kopiera texten från dokumentet. Vissa äldre dokument med dåligt tryck kan vara svåra att OCR-tolka korrekt vilket medför att den OCR-tolkade texten kan innehålla fel och därför bör man visuellt jämföra med verkets bilder för att avgöra vad som är riktigt.

This work has been digitised at Gothenburg University Library.  
All printed texts have been OCR-processed and converted to machine readable text.  
This means that you can search and copy text from the document. Some early printed books are hard to OCR-process correctly and the text may contain errors, so one should always visually compare it with the images to determine what is correct.



*Sture Allén*

# Modersmålet i fäderneslandet



# MEIJERBERGS ARKIV

FÖR

SVENSK ORDFORSKNING

UTGIVET

AV

STYRELSEN FÖR MEIJERBERGS INSTITUT  
VID GÖTEBORGS UNIVERSITET

25

GÖTEBORG 1999



# MEIJERBERGS ARKIV

FÖR  
SVENSK ORDFORSKNING

UTGIVET  
AV  
STYRELSEN FÖR MEIJERBERGS INSTITUT  
VID GÖTEBORGS UNIVERSITET  
GENOM  
BO RALPH

25

GÖTEBORG 1999



# MODERSMÅLET I FÄDERNESLANDET

Ett urval uppsatser under fyrtio år

AV

STURE ALLÉN

Ingår som del 25 i Meijerbergs arkiv för svensk ordforskning

© Sture Allén och  
Meijerbergs institut för svensk etymologisk forskning  
Göteborgs universitet  
Box 200  
SE-405 30 Göteborg

ISSN 0348-7741  
ISBN 91-630-8022-2

Novum Grafiska AB, Göteborg 1999



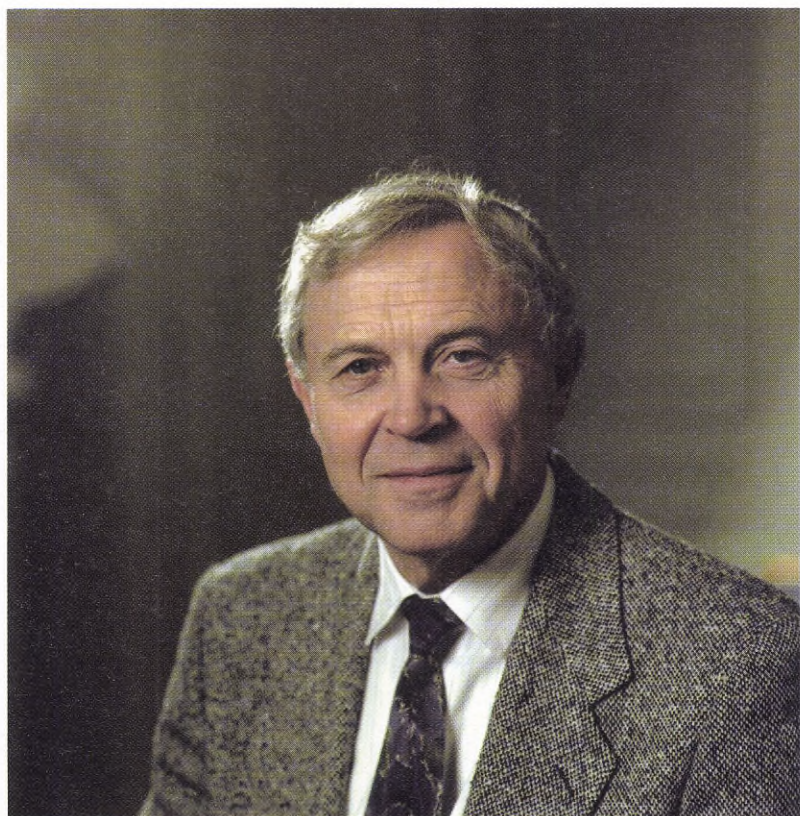
## Utgivarens förord

Sture Alléns insatser för svensk språkvetenskap och svenskt kulturliv är välkända på ett övergripande plan. Däremot fördelar sig hans skriftliga produktion över en rätt oöverskådlig mängd publikationer, varav vissa dessutom numera är svåråtkomliga.

Här presenteras ett fylligt urval av Sture Alléns artiklar, huvudsakligen med svenska språket som ämne, i en sammanhållen volym, som fått titeln *Modersmålet i fäderneslandet*. Vid urvalet har författaren själv haft ett avgörande inflytande. Den ursprungliga texten har endast lätt justerats där det nya sammanhanget så krävt; tabeller och figurer har i görligaste mån harmoniserats till formen. Bidragen härrör från fem av seklets decennier.

Det är naturligt att detta skrifturval, som i så hög grad behandlar språkets lexikaliska sida, ges ut i serien Meijerbergs arkiv för svensk ordforskning. Som ung nordist hade Sture Allén under många år ett excerperingsuppdrag för institutet, och senare var han länge ledamot av dess styrelse. Institutet tackar för ett troget engagemang.

Att boken tillkommit just nu skall emellertid främst ses som en hyllning till Sture Allén med anledning av hans 70-årsdag den 31 december 1998. Flera gamla och nya medarbetare och vänner har funnit glädje i att få bidra med arbetsinsatser av olika slag och omfattning vid bokens sammanställning, och ännu fler har sällat sig till gratulanterna.



Shure Allen

## Tabula gratulatoria

LARS AHRENBERG, LINKÖPING  
SIGVARD AKSÉN, UPPSALA  
LARS ALFVEGREN, GÖTEBORG  
BENGT ALTENBERG, LUND  
ÅKE ALTÉUS, STOCKHOLM  
ANDERS-BÖRJE ANDERSSON, GÖTEBORG  
ERIK ANDERSSON, ÅBO  
LARS-GUNNAR OCH NINNI ANDERSSON, ASKIM  
SVEN-ERIK ANDERSSON, GÖTEBORG  
ANNA HELGA HANNESDÓTTIR, GÖTEBORG  
BRITT-MARIE ARKHAMMAR, SKÖNDAL  
LENNART ARONSON, GÖTEBORG  
STINA AULIN, STENUNGSUND  
BALDUR JÓNSSON OCH GUÐRÚN STEFÁNSDÓTTIR, REYKJAVÍK  
FREDRIK BELFRAGE, GÖTEBORG  
ALICE OCH SVEN BENSON, GÖTEBORG  
STURE BERG, GÖTEBORG  
BILLSTRÖMSKA FOLKHÖGSKOLAN, KÅLLEKÄRR  
CURT BLADH, SUNDSVALL  
CLAS BLIX, DJURSHOLM  
BIRGITTA BLOM, BROMMA  
BOK & BIBLIOTEK I NORDEN AB  
BOKFÖRLAGET ATLANTIS  
ALBERT BONNIERS FÖRLAG  
JAN-ERIK BRANTÅS, DJURSHOLM  
TORE BROWALDH, STOCKHOLM  
EVA BRYLLA, UPPSALA  
CARINA OCH LARS BURMAN, UPPSALA  
DANIEL BÖRTZ, KUNGSÅNGEN  
JÖRGEN CEDERBERG, STOCKHOLM  
YVONNE CEDERHOLM, GÖTEBORG  
INGRID DAHLBERG, STOCKHOLM  
KERSTIN OCH LARS DAHLBÄCK, STOCKHOLM  
MARGARETHA OCH GUNNAR DAHLSTRÖM, GÖTEBORG  
BERTIL DANIELSSON, STOCKHOLM  
PERNILLA DANIELSSON, GÖTEBORG  
DANSK SPROGNÆVN

DEN NORSKE NOBELKOMITÉ  
DET NORSKE NOBELINSTITUTT  
DIALEKT-, ORTNAMNS- OCH FOLKMINNESARKIVET I GÖTEBORG  
DIALEKT- OCH ORTNAMNSARKIVET I LUND  
ELA DURA, GÖTEBORG  
LARS-ERIK EDLUND, HOLMSUND  
JAN EINARSSON, LUND  
ALVAR ELLEGÅRD, VÄSTRA FRÖLUNDA  
LENNART ELMÉVIK, UPPSALA  
ELISABET ENGDAHL, MÖLNDAL  
CHRISTER ENGSTRÖM, NYHAMNSLÄGE  
GUNNEL OCH LARS ENGWALL, STOCKHOLM  
BIRGITTA ERNBY, SÄRÖ  
ANNA OCH KJELL ESPMARK, STOCKHOLM  
BERTIL FALCK, GÖTEBORG  
BIRGIT FALCK-KJÄLLQUIST, SVEDALA  
MAGNUS FLORIN, STOCKHOLM  
FORSKNINGSCENTRALEN FÖR DE INHEMSKA SPRÅKEN  
FØROYAMÁLSDEILDIN FRÓÐSKAPARSETUR FØROYA  
QARIN FRANKER, GÖTEBORG  
INGEGERD OCH SIGURD FRIES, UMEÅ  
MARTIN FRITZ, GÖTEBORG  
ERIK FRYKMAN, GÖTEBORG  
KARL ERLAND GADELI, PARIS  
ROLF GAVARE, GÖTEBORG  
MARTIN GELLERSTAM, GÖTEBORG  
MAUD GISTEDT, HISINGS KÄRRA  
LARS GRAHN, STOCKHOLM  
OLLE GRANATH, STOCKHOLM  
BO GRANDIEN, STOCKHOLM  
MICKEL GRÖNROOS, GÖTEBORG  
DAG GUNDENSEN, OSLO  
ANNA GUNNARSDOTTER GRÖNBERG, GÖTEBORG  
LARS OCH YVONNE GURMUND, GÖTEBORG  
RUNE GUSTAFSSON, ALINGSÅS  
SVERKER GÖRANSSON, GÖTEBORG  
GÖTEBORGS UNIVERSITET  
ULF ERIK HAGBERG, BROMMA  
STIG HAGSTRÖM, STOCKHOLM  
ELISABETH OCH PAUL HALLBERG, GÖTEBORG  
ELSA OCH GUNNAR HAMBREÆUS, SOLLENTUNA  
K. G. HAMMAR, UPPSALA  
OLLE HAMMERMO, KARLSTAD  
BENGT HANE, GÖTEBORG  
STINA HANSSON, GÖTEBORG  
GUNNAR HARDING, BROMMA  
SVEN-OLOF HEDENGREN, KUNGSÄNGEN  
KOLBJØRN HEGGSTAD, BERGEN

STAFFAN HELLBERG, SOLNA  
PER HOLMBERG, GÖTEBORG  
GUNNAR HOPPE, SOLNA  
BOJEN OCH LARS HULDÉN, HELSINGFORS  
TOR G. HULTMAN, YSTAD  
HUMANISTISKA INSTITUTIONEN, HÖGSKOLAN I ÖREBRO  
KJELL HÄRNQVIST, GÖTEBORG  
INSTITUT FOR NORDISK FILOLOGI, KØBENHAVNS UNIVERSITET  
INSTITUTIONEN FÖR KULTUR OCH KOMMUNIKATION/SVENSKA SPRÅKET,  
HÖGSKOLAN I KARLSTAD  
INSTITUTIONEN FÖR NORDISK FILOLOGI, ÅBO UNIVERSITET  
INSTITUTIONEN FÖR NORDISKA SPRÅK, JYVÄSKYLÄ UNIVERSITET  
INSTITUTIONEN FÖR NORDISKA SPRÅK, LUNDS UNIVERSITET  
INSTITUTIONEN FÖR NORDISKA SPRÅK, STOCKHOLMS UNIVERSITET  
INSTITUTIONEN FÖR NORDISKA SPRÅK, UMEÅ UNIVERSITET  
INSTITUTIONEN FÖR NORDISKA SPRÅK, UPPSALA UNIVERSITET  
INSTITUTIONEN FÖR NORDISKA SPRÅK, VASA UNIVERSITET  
INSTITUTIONEN FÖR SVENSKA SPRÅKET, GÖTEBORGS UNIVERSITET  
INSTITUTIONEN FÖR TYSKA, FRANSKA OCH NORDISKA SPRÅK,  
ULEÅBORGS UNIVERSITET  
INSTITUTT FOR NORDISK SPRÅK OG LITTERATUR, UNIVERSITETET I OSLO  
ÍSLENSK MÁLNEFND  
ÍSLENSK MÁLSTÖÐ  
ANN-MARIE IVARS, HELSINGFORS  
GUNILLA OCH CARL-OLOF JACOBSON, UPPSALA  
HILLEVI OCH GUNNAR JACOBSSON, GÖTEBORG  
BENGT JANGFELDT, STOCKHOLM  
INGER JANSSON, MÖLNDAL  
PIA JARVAD, KØBENHAVN  
LORE JOHANNISSON, GÖTEBORG  
GUNNAR L. JOHANSSON, GÖTEBORG  
KARL G. JOHANSSON, GÖTEBORG  
MONICA JOHANSSON, KUNGÄLV  
SOFIE JOHANSSON KOKKINAKIS OCH DIMITRIOS KOKKINAKIS, PARTILLE  
ING-MARIE JOHANSSON-ZERE, GÖTEBORG  
INGE JONSSON OCH KERSTIN FREDGA, SALTSJÖ-DUVNÄS  
FOLKE JOSEPHSON, GÖTEBORG  
JERKER JÄRBORG, GÖTEBORG  
SIGRID OCH JOHN KAHLE, UPPSALA  
ULLA KAMPMAN, ULEÅBORG  
BRITT OCH HUGO KARLSSON, SÄTILA  
ARTO KIRRI, OULU  
GÖRAN OCH LISBETH KJELLMER, GÖTEBORG  
EINAR KORPUS, GÖTEBORG  
KRISTINN JÓHANNESON, GÖTEBORG  
KUNGL. MUSIKALISKA AKADEMIEN  
KUNGLIGA INGENJÖRSVETENSKAPSAKADEMIEN, IVA  
REINERT KVILLERUD, KARLSTAD

INGRID OCH ROGER KÄLLSTRÖM, VÄSTRA FRÖLUNDA  
HANS LANDQVIST, VÄSTRA FRÖLUNDA  
BO LARSSON, STOCKHOLM  
SVEIN LIE, OSLO  
GERT LINDBERG, DEN GYLDENE FREDEN  
ELSA OCH ÖRJAN LINDBERGER, BROMMA  
BIRGITTA LINDOHF, STOCKHOLM  
PER LINELL, LINKÖPING  
JAN LING, GÖTEBORG  
ANNE-MARIE LONDEN, HELSINGFORS  
SIGNE LUND-ASPENSTRÖM, STOCKHOLM  
AINA LUNDQVIST, GÖTEBORG  
BJÖRN LUNDSTRÖM, KASTELLHOLMENS FASTIGHETSUTVECKLING AB  
BENJAMIN LYNGFELT, GÖTEBORG  
LARS LÖFGREN, NORDISKA MUSEET  
SVEN LÖVFORS, AXVALL  
BENTE MAEGAARD, VALBY  
SVEN-GÖRAN MALMGREN, GÖTEBORG  
GÖRAN MALMQUIST, STOCKHOLM  
MÄRTA MALMSTRÖM, ULEÅBORG  
SÖREN MANNHEIMER, GÖTEBORG  
KJELL A. MATTSSON, SKEE  
BARBRÖ MEDVECKZY, LANDVETTER  
LENNART MOBERG, UPPSALA  
HENRY MONTGOMERY, STOCKHOLM  
KATARINA MÜHLENBOCK, GÖTEBORG  
TORA OCH LEIF MÆHLE, OSLO  
PETER MÖLLER, BROMMA  
BENGT NAGEL, TÄBY  
GUNVOR NILSSON, ÖREBRO  
JAN S. NILSSON, GÖTEBORG  
NOBELSTIFTELSEN  
BENGT NORDBERG, UPPSALA  
KERSTIN NORDENSTAM OCH SUNE ÖRNBERG, GÖTEBORG  
NORDISK SPRÅKRÅD  
NORDISKA SPRÅK, TAMMERFORS UNIVERSITET  
CHARLOTTE OCH JON ERIK NORDSTRAND, KULLAVIK  
MARI OCH CLAES NORDSTRÖM, STOCKHOLM  
KERSTIN NORÉN, GÖTEBORG  
ERLING NORRBY, STOCKHOLM  
NORSTEDTS FÖRLAG  
NORSTEDTS ORDBOK AB  
PIRKKO NUOLIJÄRVI, HELSINGFORS  
INGEGERD NYSTRÖM, HELSINGFORS  
BRITT-MARI NÄSSTRÖM, LERUM  
BERNT OLSSON, HELSINGBORG  
TORE ÖTTERUP, GÖTEBORG  
PER A. PETTERSSON, VIKINGSTAD

PHILIP PARTNERE ADVOKATFIRMA, KØBENHAVN  
MAGNUS VON PLATEN, STOCKHOLM  
JÓHAN HENDRIK W. POULSEN, KIRKJUBØUR  
BO RALPH, GÖTEBORG  
YLVA RALPH, OSLO  
GUNNEL OCH MÅRTEN RASCH, STOCKHOLM  
MAJ OCH VIDAR REINHAMMAR, UPPSALA  
SIV OCH MIKAEL REUTER, HELSINGFORS  
PER S. RIDDERSTAD, LIDINGÖ  
RIKSARKIVET  
NILS RINGERTZ, STOCKHOLM  
CHRISTINA ROGESTAM OCH LENNART WIDING, GÖTEBORG  
LENA ROGSTRÖM, GÖTEBORG  
JAN-CARL VON ROSEN, STOCKHOLM  
ALLAN ROSTVIK, UPPSALA  
EVA RUDMAN, STOCKHOLM  
HANNE RUUS, KØBENHAVN  
KERSTIN OCH ANDERS RYBERG, BROMMA  
BENGT RYDÉN, STOCKHOLM  
ANNELI RÄIKKÄLÄ, VANDA  
MIRJA SAARI, HELSINGFORS  
BENGT OCH KARIN SAMUELSSON, DJURSHOLM  
BO SAMUELSSON, GÖTEBORG  
LILLEMOR SANTESSON, HÖRBY  
PAULI SAUKKONEN, HELSINGFORS  
SEKSJON FOR LEKSIKOGRAFI OG MÅLFØREGRANSKING, UNIVERSITETET I OSLO  
INGVAR OCH EVA SELIN LINDGREN, GÖTEBORG  
SEMINARIET FÖR NORDISK NAMNFORSKNING, UPPSALA UNIVERSITET  
ULLA OCH BENGT SIGURD, LUND  
JOHN SINCLAIR, PESCIA  
ANDERS SJÖBERG, ROMELANDA  
CHRISTIAN SJÖGREEN, GÖTEBORG  
SKANDINAVISKA ENSKILDA BANKEN STIFTELSE OCH FÖRETAG  
PER SKÖLD, STOCKHOLM  
MICHAEL SOHLMAN, STOCKHOLM  
SPRÅK- OCH FOLKMINNESINSTITUTET  
STOCKHOLMS UNIVERSITET  
RIGMOR OCH JAN-OLOF STRANDBERG, STOCKHOLM  
SVANTE STRANDBERG, UPPSALA  
STIG OCH GUNILLA STRÖMHOLM, UPPSALA  
ANNA GRETA STÄHLE, LIDINGÖ  
BARBRO STÄHLE SJÖNELL, LIDINGÖ  
TORHEL STÄLMARCK, LIDINGÖ  
STAFFAN SUNDBERG, STOCKHOLM  
BO SUNDQVIST, UPPSALA  
JAN OCH GUNILLA SVARTVIK, LUND  
JOHAN SVEDJEDAL, UPPSALA  
BO SVENSÉN, STOCKHOLM

SVENSKA AKADEMIEN  
SVENSKA AKADEMIENS NOBELBIBLIOTEK  
SVENSKA AKADEMIENS ORDBOKSREDAKTION  
SVENSKA SPRÅKNÄMNDEN  
LARS SVENSSON, MALMÖ  
HENRIK SVENUNGSSON, STOCKHOLM  
SVERIGES FÖRFATTARFÖRBUND  
MARIE-CHRISTINE OCH KRISTIAN VON SYDOW, GÖTEBORG  
BO SYLVAN, NORRKÖPING  
MARGARETA SÖRENSON, STOCKHOLM  
MARIKA TANDEFELT, HELSINGFORS  
TEKNISKA NOMENKLATURCENTRALEN  
ULF TELEMAN, LUND  
BENGT E. TELLANDER, SOLLENTUNA  
SEIJA TIISALA, HELSINGFORS  
UNIVERSITETET I OSLO  
HÅKAN WAHLBERG, HUDDINGE  
BO VAHLNE, SOLNA  
PETER WALLENBERG, STOCKHOLM  
BARBRO WALLGREN-HEMLIN, GÖTEBORG  
LARS-G. WARNE, STOCKHOLM  
KJELL WEIMER, HELSINGFORS  
DAG WESTERSTÅHL, GÖTEBORG  
MARGARETA WESTMAN, STOCKHOLM  
LENA WETTERLING, KUNGÄLV  
GUN WIDMARK, UPPSALA  
ANNE MARIE WIESELGREN, VÄXJÖ  
MARIA WINE, SOLNA  
KAROLINA WIRDENÄS, GÖTEBORG  
JOHAN WREDE, HELSINGFORS  
ÅBO AKADEMI  
ÅBO AKADEMIS BIBLIOTEK  
KENNETH ÅSTRÖM, KRISTIANSTAD  
VERONICA OCH ANDERS R. ÖHMAN, STOCKHOLM



## Innehåll

Svenskans beskrivning. Öppningsanförande vid tjugotredje sammankomsten för svenskans beskrivning 1998 (Utkommer i: Svenskans beskrivning 23) .....	1
Uarda-Akademiens ordbok (Sydsvenska Dagbladet 22 maj 1998) .....	6
Överträffa Shakespeare? Några aspekter på Nobelpriset i litteratur (Artes 1997: 3) .....	8
Den goda smakens återupprättande. Om akademiernas roll på tjugohundralet (Svenska Dagbladet 9 oktober 1997) .....	20
LEXIS. Opening Speech at the Seventh Euralex International Congress, Göteborg, August 13-18, 1996 (International Journal of Lexicography 10: 4. 1997) .....	26
Tidender och tidningar (Den moderna dagspressen 350 år. 1996) .....	36
Klassiker och klassiker (Språket lever! Festskrift till Margareta Westman den 27 mars 1996) .....	42
Har Europa närmat sig varandra? (Magasin Lyran 1996: 3) .....	47
The Relation between Language and Mind. Opening Address at a Nobel Symposium 1994 (Of Thoughts and Words. The Relation between Language and Mind. Proceedings of Nobel Symposium 92. 1995) .....	49

Om Akademiälmanackan (Kulturrådet 1994: 6) .....	54
Runmärkt (Anförande vid invigningen av runutställningen på Stockholms medeltidsmuseum 1994) .....	59
Ifall ett språk du äger. Föredrag vid Svenska litteratursällskapets i Finland årshögtid den 5 februari 1992 (Historiska och litteraturhistoriska studier 68. 1993) .....	63
Tal inför överlämnandet av Svenska Akademiens nordiska pris till Thor Vilhjálmsen (Svenska Akademiens Handlingar 1992) .....	72
Corpus Linguistics. Opening Address at a Nobel Symposium 1991 (Directions in Corpus Linguistics. Proceedings from Nobel Symposium 82. 1992) .....	75
Något om NEO och somt om SAOL (Finlandssvenskans framtid. Rapport från en språkvårdskonferens i Vasa den 4 och 5 maj 1992) .....	78
Vad skall husen heta? (Göteborgs-Posten 12 juli 1992) .....	89
Ture Johannisson – nordist (Humanistdag-boken nr 9. Göteborgs universitet 1996; även snarlikt i: Kungl. Vitterhets Historie och Antikvitets Akademiens Årsbok 1991) ....	91
Vilka är de intellektuella? (Svenska Dagbladet 24 februari 1991) .....	97
<i>Textil, slöjd och textilslöjd</i> (Jubileumsskriften Hjärna – Händer. Göteborgs universitet 1991) .....	99
Bibeln och Akademien (Ingår i: Svenska Akademien och Bibeln – en triptyk. I: Den svenska bibeln. Ett 450-årsjubileum. 1991) .....	102
Från grafem till lexem (Forskning i ett föränderligt samhälle. Stiftelsen Riksbankens Jubileumsfond 1965–1990. 1990) .....	109

Liseberg i våra hjärtan (Göteborgs-Posten 15 juli 1990) .....	129
Possible Worlds in Humanities, Arts and Sciences. Opening Address at a Nobel Symposium 1986 (Possible Worlds in Humanities, Arts and Sciences. Proceedings of Nobel Symposium 65. 1989) .....	130
Språkfrågor. Om ordbruket i det här landet (Kungl. Vitterhets Historie och Antikvitets Akademiens Årsbok 1989) ..	136
Om Svensk ordbok (Svenskläraren 1988: 2) .....	143
Akademiens språkliga engagemang (Tal och texter 4–5 april 1986. Svenska Akademien 200 år. 1986) .....	146
På ren svenska (Språk. Svenskläraryrskriftens årsskrift 1986) .....	150
<i>Information</i> (Information om <i>information</i> . 1985) .....	153
Teknik och språk i informationssamhället (Språk og samfunn i Norden etter 1945. 1985) .....	157
Skandinavisk datalingvistik (The Nordic Languages and Modern Linguistics 5. 1984) .....	168
Översättningsdatoren – finns den? (Mänskligt och omänskligt i informationsteknologin. 1984) .....	200
Inte bara idiom (De nordiska datalingvistikdagarna 1983. 1984) .....	207
Översätarnas språk — finns det? (Mänskligt och omänskligt i informationsteknologin. 1984) .....	214
En forskningsstrategi för språkvetenskaplig databehandling. Perspektivskiss (Språkdata 1983) .....	222
Andas (Struktur och variation. Festskrift till Bengt Loman 7.8.1983) .....	234

Text Processing as a Theme. The Capriccio and Ostinato of Texts. Opening Address at a Nobel Symposium 1980 (Text Processing. Text Analysis and Generation, Text Typology and Attribution. Proceedings of Nobel Symposium 51. 1982) .....	241
Om förändringar i svenskan sedan 1950 (Svenskans beskrivning 13. 1982) .....	249
The Lemma-Lexeme Model of the Swedish Lexical Data Base (Empirical Semantics II. A Collection of New Approaches in the Field. 1981) .....	268
Tjugohundratalet och några andra tidsord (Språkvård 1981: 2; även i: Festskrift till Carl-Eric Thors 8.6.80) .....	279
Herr talman tolv tusen gånger (Dagens Nyheter 30 september 1981) .....	287
Hur djupt går SAOB? (Nysvenska studier 59–60. 1979–1980. 1980) .....	290
The Language Bank Concept (Data Bases in the Humanities and Social Sciences. 1980) .....	302
Carl Ivar Ståhle. Inträdestal vid Svenska Akademiens högtidssammankomst den 20 december 1980 (Svenska Akademiens handlingar 88/1980. 1981) .....	311
Språkvetenskaplig databehandling (Presentationer av ämnesområden utgivna i samband med Göteborgs universitets professorsinstallation lördagen den 20 oktober 1979) .....	324
Skriva som vi talar? (Språkform och språknorm. Studier tillägnade Bertil Molde. 1979) .....	335
Om språkvårdens principer (Språk i Norden 1979) .....	342
Dator och datamaskin, fackspråk och allmänspråk (Språkvård 1978: 4; även snarlikt i: Tankar om språket. 1982) .....	354
Terminologi i nordiskt samarbete (Språk i Norden 1978) .....	360

Text-Based Lexicography and Algorithmic Text Analysis (Association for Literary and Linguistic Computing Bulletin 5: 2. 1977) .....	370
Morf-o-log-i och poly-sem-i. Om NFO4-gruppens forskningsarbete (Nysvenska studier 57. 1977) .....	380
Några könsskillnader i journalistspråk (Könsroller i språk. 1977) .....	393
On Phraseology in Lexicology (Cahiers de lexicologie 29. 1976) .....	396
Alla svenskars efternamn (Dagens Nyheter 25 januari 1976) .....	407
Alla svenskars förnamn (Dagens Nyheter 27 december 1975) .....	413
Den svenska språkvårdens organisation (Språkvård 1975: 4) .....	421
I ordens ljus tonar samhället fram (Dagens Nyheter 6 februari 1975) .....	431
Tionde upplagan av Svenska Akademiens ordlista (Språkvård 1973: 4) .....	437
Stavning (Tre uppsatser om stavning. Ord och stil 5. 1973) .....	449
Denominala verbalfraser (Svenskans beskrivning 5. 1972) .....	460
Ordströmmen som dataflöde. En aspekt på språklig databehandling (Humanistisk forskning 1972: 2) .....	480
Om textattribution. Kring en avhandling av Marina Mundt (Arkiv för nordisk filologi 86. 1971) .....	487
Vocabulary Data Processing (The Nordic Languages and Modern Linguistics. 1970) .....	523

Ordklasser (Studier över nusvenskans vokabulärsystem. Språkdata 1969) .....	549
Semantische Strukturen. Bemerkungen zur Arbeit von Inger Rosengren (Studia Neophilologica 40: 2/1968. 1969) .....	570
Engelskt och svenskt. En översikt av det språkliga inflytandet (Språkdata 1968) .....	583
Om nominalkomponerade verb i nordiska språk (Språkdata 1968) .....	601
Verbal allomorfi i modern svenska (Språkdata 1967) .....	637
Förhållandet mellan skrift och tal (Språk, språkvård och kommunikation. 1967) .....	661
Ordleksepisoden i Gutasagan och textdokumentens stemma (Språkdata 1967) .....	679
Ordforskaren och datamaskinen (Göteborgs Handels- och Sjöfarts-Tidning 31 mars 1964) .....	700
Massmedium, etermedium, mediotek (Svenska Dagbladet 6 oktober 1964) .....	704
Indelningen av nusvenskans pronomen (Arkiv för nordisk filologi 73. 1958) .....	709

## Svenskans beskrivning

Öppningsanförande vid tjugotredje sammankomsten för svenskans beskrivning 1998.

Kära sammankommande,

Den 19 augusti 1772 gjorde Gustaf III sin statskupp med en vit bindel om vänster överarm. Dagen efter talade han på Stortorget till stadens invånare. Han gjorde succé, rapporteras det, ty han talade svenska. Det hade ingen kung gjort sedan Karl den XII. Gustaf III främjade svenskt språk och svensk kultur på mångahanda sätt.

Det finns andra hållningar. En går under namnet toalettprincipen. Innebörden är att det skall vara minst två utlänningar i en bolagsstyrelse, så att man inte börjar prata svenska, när den ende utlänningen går på toaletten. Principen har attribuerats till Percy Barnevik.

Trycket på vårt språk är verkligen stort. En av de värsta hållningarna formuleras så: barnbarnen skall undervisas på engelska, för svenskan kommer ju ändå att vara död, när de har vuxit upp. Som referens får jag ange "personal communication".

En sammankomst för svenskans beskrivning är i detta perspektiv ett viktigt, för att inte säga livsviktigt, möte. Det representerar en av de tre grundläggande språkaktiviteterna: att använda språket, att beskriva det och att vårda det. Lyckligtvis rör det sig egentligen om två av dem, eftersom vi också använder språket.

Det finns mer att säga om språksituationen. Svenskundervisningen i skolan, eller snarare direktiven för svenskundervisningen, har inte ett fläckfritt förflutet under vår tid. Högre makters ryckighet har kraftigt försvårat lärarnas arbete. Förhoppningsvis är det vi nu ser en omsvängning till det bättre.

Svenskan har ställning som riksspråk och är erkänt som ett av EU:s språk. Men kan vi lita på att våra europeiska representanter talar med kraft och insikt om vikten av mångspråkighet? Låt mig nämna här, att vi inom Svenska språknämndens styrelse har ut-

arbetat ett förslag till handlingsprogram, där vi bland annat föreslår lagstiftning om svenskan som landets huvudspråk.

Många oberoende iakttagelser av de företag vi kallar massmedier går ut på att ordförrådet på en del håll förråas. Silningen och vinklingen av informationsflödet gör att vi har hamnat i vad forskare betecknar som det redigerade samhället. Förhoppningsvis, återigen, skall mediernas självsanering växa sig starkare.

Internet har inneburit, att en ny kanal för språklig aktivitet har öppnats. Hittills har dock inte tjatrandet nått någon högre språknivå. Jargongen i allmänhet kring utrustningen är inte heller värd något beröm: jag hämtar ner fajlen från webben och mejlar över den till er.

Dessutom blåser isiga vindar över de humanistiska fakulteterna. Problemen hopar sig i forskarutbildningen, i projekten och på institutionerna med indragning eller vakantsättning av tjänster. Men sammankomsterna för svenskans beskrivning fortsätter, gudskelov.

Efterhand som tiden har gått, har våra sammankomster utvecklats till ett forum, där man får tillfälle att pröva idéer, även halvbakta, att redovisa resultat, att känna av klimatet och mycket annat. Här möts äldre och yngre forskare, mindre i en salig blandning än i en mångstämmig kör. I början var det inte så.

Den första sammankomst jag deltog i hölls 1964 i Lund. Jag talade då om grafem och fonografem. Det mötet hade föregåtts av ett i Stockholm året innan. Bakgrunden var att det på några håll i landet hade börjat sticka upp ett intresse för att sätta sig in i och använda nya metoder för strukturell språkbeskrivning. Bengt Loman har skissat det dåvarande läget i ett anförande vid femtonde sammankomsten i denna sal 1985. Han säger (s. 45): ”Rörelsen kämpade i motvind. Ord som strukturell och fonem betraktades i traditionella kretsar som onödiga påhitt. För att förstå klimatet bör man också veta att de dåvarande högskolekurserna i nordiska språk (som ämnet då hette) var kraftigt koncentrerade på fornspråken, dvs. på färdighet att läsa och översätta Snorre, Eddan och Västgötalagen samt på memorering av paradigm och böjningsmönster. Kurser eller föreläsningar i grammatik förekom inte, och i läskursen representerades nusvenskan av



Beckmans svenska språklära, som ju utkommit i sin första upplaga redan 1904 och egentligen var avsedd för 'den högre elementarundervisningen'." Det Bengt Loman här säger var något som flera av oss upplevde mycket starkt. Jag vill gärna tillägga, att vi behöll vårt historiska intresse. Det var i grunden behovet av internationell orientering som vår sammetsrevolution gällde.

Här i Göteborg ordnades den tredje sammankomsten 1965. Vi var inte fler än att vi satt i ett rum i den villa på Viktor Rydbergsgatan, där institutionen för nordiska språk då höll till. Sammankomsterna i Göteborg avspeglar på ett åskådligt sätt utvecklingen. Vid tredje sammankomsten 1965 hade vi självfallet bara en session, vid den nionde 1975 hade vi två sessioner och vid den femtonde 1985 tre. Nu, vid den tjugotredje 1998, blir det som vi vet fyra. Serien av sammankomster fick stabilitet genom tillkomsten av Forskningsnämnden för modern svenska, som i början också hade projektsamordnande uppgifter. Av stor vikt har det varit, att sammankomsterna har främjat kontakt och utbyte mellan de finlandssvenska och de rikssvenska forskarna.

Hur har då fokuseringen skiftat genom de 35 åren? De ursprungliga frågorna kvarstår, men fältet har vidgats i många riktningar. Här har presenterats bidrag inom fonologi, grafonomi, morfologi, syntax, semantik, lexikologi, pragmatik, textlingvistik, stilistik, språkinläring m.m. Generativa trender har passerat revy, språksociologin har gjort sitt inträde, korpuslingvistiken var på plats innan termen hade uppfunnits. I år tornar speciellt en företeelse upp sig: vart fjärde föredrag handlar om samtal.

Jag tar också tillfället i akt att säga något om Svenska Akademiens arbete för språket. I slutet av året kommer tolfte upplagan av SAOL, som bildar epok i ordlistans historia genom att presentera uppslagsorden i bokstavsordning. Redaktör är Martin Gellerstam, som har ett nära samarbete med Sture Berg. De rapporterar till Akademiens språkkommitté, som beslutar i principiella frågor, avgör en del kvistiga ärenden och i enstaka fall hänskjuter spörsmål till Akademien. Ordlistan har också stramats upp i lexikologiskt avseende, vilket ju som en av sina förutsättningar har en långt driven datamaskinell metodik. Dessutom ger den bland annat ordklassbeteckning för varje uppslagsord och

markering av ordledsgränserna. Det senare har ett belärande och ett disambiguerande syfte men blir därtill en hjälp för dem som använder morfologisk avstavning.

När jag 1985 föreslog Akademien att fullfölja stadgarnas uppdrag rörande en svensk grammatik, var förslaget och blev beslutet ett femårsprojekt. Grammatiken är nu i provsättningskedet, och det har alltså snarast blivit ett femtonårsprojekt. Jag beklagar ingalunda, att aptiten har vuxit och att vi får ett verk på åtminstone 2 500 sidor. Två av huvudförfattarna finns på sammankomstens deltagarlista: Staffan Hellberg och Erik Andersson. Den tredje är Ulf Teleman, gruppens ledare. Vi får en innehållsrik grammatik på klassisk grund och i modern tappning. Jag kan intyga detta, eftersom jag har läst och kommenterat det hela. Det gäller också de båda övriga i Akademiens språkkommitté, Göran Malmqvist och Gunnel Vallquist, liksom Bo Svensén, adjungerad sekreterare i kommittén. Grammatiken har ytterligare den förtjänsten, att den under gång nämner var vi saknar tillräckliga kunskaper. Här finns många uppslag till inlägg vid sammankomsterna framöver.

Mot bakgrunden av de alltmer arbetskrävande uppgifterna och det stadigt växande antalet språkfrågor från olika håll har det blivit nödvändigt att förstärka Akademiens sekretariat med en språkvetenskapligt välmeriterad person. Beslut har fattats om att så skall ske. Låt mig tillägga att Akademien i år har tilldelat Svenska språknämnden både ett årligen återkommande anslag och ett extra anslag.

Arbetet på den stora ordboken, SAOB, har för ett bra tag sedan sprängt sekelvallen. Man slutför nu bokstaven S och har börjat på redigeringen av T. Akademien har beslutat, att ordboken skall vara klar år 2017, det är 25 år efter 1992 då beslutet fattades. Som bekant finns ordboken sedan en tid också i form av en databas, kallad OSA. Det dagliga arbetet på den leddes länge av Rudolf Rydstedt och leds numera av Yvonne Cederholm. I dagarna har den gjorts allmänt tillgänglig via Internet.

Efter den inviten vill jag till sist gärna instämma i Lars-Gunnar Anderssons välkomsthälsning och därtill foga en välgångsönskan. Det som återstår är den akt som nu följer. Tjugotredje sam-

mankomsten för (att dryfta frågor rörande) svenskans beskrivning är öppnad.

## Uarda-Akademiens ordbok

Sydsvenska Dagbladet 22 maj 1998.

Det finns många sätt att roa sig med ord. Ett är vitsen, ofta med en galghumoristisk underton, som när en tätt trafikerad led längs älven kallas *gasaremsan*. Ett annat är delar av det som går under beteckningen politiskt korrekt språk, där exempelvis *corpulent* föreskrivs bli ersatt av *kroppsligt begåvad*. Till repertoaren hör också Oxfords bidrag, spoonerismerna: (you have deliberately) *tasted two worms* för det som skulle vara *wasted two terms*. Bland formelskämten har vi inte minst det som bygger på *jag trodde mina lakan var vita tills jag såg* etc. med varianter som *jag trodde jag var ateist tills jag upptäckte jag var Gud*.

Man kan förstås också som Uarda-Akademiens ordbok vränga ord i allmänhet och vid sidan av det vanliga *lexikon* och Bra Böckers *hexikon* presentera *exikon* 'uppslagsverk över fördettingar' och *mexicon* 'aztekisk ordlista'. Det kan räcka med att smussa med ett tecken. Vi får *blundin* 'sovande skönhet', *ängelska* 'himmelskt språk', *nedröstningskonferens* 'överläggning vid vilken varje förslag förkastas med förkrossande majoritet' och *sfärord* 'runt eller eljest anstötligt ord'. Ibland kan man tillåta sig att lägga till ett tecken: *deklantera* 'tappa upp vin slarvigt', *skrönika* 'osannfärdig historieskrivning'. Några gånger kan man gå mera våldsamt tillväga: *fåtöljetång* 'förströelselektyr', *papput* 'hanne av utdöd elefantart'.

Man kan också läsa orden som de står, även om de var tänkta på ett annat sätt. Så avslöjas bland många andra *cd-rom* 'västindisk diplomatdryck', *drivhus* 'snöhydda', *flottbas* 'motsatsen till smörtenor' och *stämjärn* 'dryck som klarar strupen'. Och det går att bygga vidare: *Giss dur* 'tonart man bara kan ana sig till', *laga förfall* 'reparera'.

Det är sådana här knep som med stor framgång används i den nya upplagan av Uarda-Akademiens ordbok. Den profilerar sig därigenom tydligt mot Svenska Akademiens ordbok, som ju tillkommer i environgerna. Men boken har också en vidare syftning.

Den skall enligt utgivarna innebära ett verkligt kunskapslyft. Dock har innebörden i det ordet redan klarlagts av månadstidningen Grönköpings Veckoblad: man flyttar helt enkelt Nationalencyklopedin från den nedre hyllan i bokskåpet till den övre.

Redaktörerna berikar vårt ordförråd med åtskilliga hundra ord. Låt oss hoppas att traditionen förs vidare. Antalet möjliga ord med de bokstäver, ljud och regler vi har är emellertid så stort, att det förtjänar att betecknas som oändligt: sundsvall sundstall sundssall sundsrall sundspall sundsnall sundsmall sundslall sundskall sundsjall sundshall sundsgall sundsfall sundsdall sundsball – sandsvall sandstall sandssall etc. – sondsvall sondstall sondssall etc. etc. Att ordboken inte är uttömmande är mot den bakgrunden rent beröm.

Herrar Landin, Meurling, Meurling och Strömberg lyckönskas till verket. Det framstår på de lexikala vattnen som en veritabel ordbåk.

## Överträffa Shakespeare? Några aspekter på Nobelpriset i litteratur

Artes 1997: 3. Uppsatsen bygger på ett anförande vid forumet "Creativity toward the 21st Century", Tokyo och Sapporo 1996.

En gång om året ankommer det på Svenska Akademien att fånga det väsentliga i en författares verk på två rader. Detta går stundom under beteckningen kollektiv diktning. Ett exempel som ligger nära till hands är motiveringen för Kenzaburo Oe som mottagare av 1994 års Nobelpris: "som med poetisk kraft skapar en imaginär värld, där liv och myt förtätas till en skakande bild av människans belägenhet i nutiden".

Motiveringar har formulerats ända sedan priset utdelades för första gången, fastän det inte står något därom i Alfred Nobels testamente. Medan detta är alldeles klart, finns det å andra sidan en passus i själva testamentet som har ansetts mindre klar.

En av de många saker som August Strindberg ondgjorde sig över i "Tal till svenska nationen" 1910 var valet 1901 av den förste Nobelpristagaren i litteratur, Sully Prudhomme: "Vidare skulle pristagaren ha skrivit idealiskt (senare tummat till idealistiskt som är något annat), men han var materialist och hade översatt Lucretius." Strindberg fortsätter med att notera, att priset tilldelades Prudhomme "mot statuter och testamente".

### Alfred Nobel och Nobelpriset i litteratur

Vad Strindberg syftar på är att Alfred Nobel i sitt testamente 1895 hade föreskrivit, att Nobelpriset i litteratur skulle tilldelas den som hade producerat "det utmärktaste i idealisk riktning". Som samtida med Nobel har Strindberg säkert rätt när han påpekar, att *idealistisk* är något annat än *idealisk* och oförenligt med testamentet. Men han säger inte vari skillnaden består. Frågan om vad ordet *idealisk* betyder hos Nobel – uppfattat som en egendomlighet eller rentav en anomali – har diskuterats i nästan

hundra år, och många tolkningar har förts fram. Som Kjell Espmark formulerar det i sin bok "Det litterära Nobelpriset. Principer och värderingar bakom besluten" 1986: "I själva verket framträder litteraturprisets historia till en väsentlig del som en räkna försök till tolkning av ett oklart testament."

Den danske litteraturkritikern Georg Brandes uppgav i ett brev, att han hade frågat Nobels vän Gustaf Mittag-Leffler vad Nobel kunde ha menat. Svaret blev, att Nobel var anarkist och att han med *idealistisk* menade 'som intar en polemisk eller kritisk hållning till religionen, kungadömet och äktenskapet, till samhällsordningen i dess helhet'. Om detta är att säga, att det är en kommentar till *idealistisk* men att det ord det gäller är *idealistisk*. Inte desto mindre ger den en antydning om Nobels självständiga hållning.

Anders Österling ville betona den roll Shelleys utopiska idealism spelade för vad han kallade Nobels praktiska idealism. Enligt honom tänkte sig Nobel "ett verk av positiv och humanistisk tendens". Han tillade, att endast den vidast möjliga tolkning kunde göras fruktbar.

Enligt Gunnar Brandell talar inte Nobels bibliotek för att han skulle vara särskilt influerad av Shelley. Brandell tvivlar på att det någonsin blir möjligt att utröna vad Nobel menade. För egen del vågar han ändå en gissning, nämligen att Nobel genom att foga ett litteraturpris till de andra priserna ville bidra till internationell förståelse med litteraturens hjälp.

Artur Lundkvist gav en gång en intervjuare följande svar: "Priset ska ha en idealistisk inriktning, företräda en humanism. Man kan inte ge det till dem som förespråkar våld." Att notera är, att Lundkvist inte begränsar sig till rent litterär produktion utan syftar på författarens åsikter över huvud taget.

Man får avstå från att hämta någon närmare ledning av det diskuterade uttrycket, menar Lars Gyllensten. Enligt hans uppfattning får man nöja sig med att konstatera, att seriös och kvalitativt högtstående litteratur på det hela taget främjar vår kunskap om människan och hennes villkor och strävar efter att berika hennes liv och förbättra hennes villkor.

Den senaste kommentaren finner man i tredje delen av Torgny Segerstedts stora verk "Svenska Akademien i sin samtid" 1992. Segerstedt påpekar, att priserna för naturvetenskap och fred tyder på att Nobel ville främja framsteg som kunde befordra folkförsoning och mellanfolkligt samarbete. Litteraturpriset skulle då inspirera författare att göra de ideal levande och attraktiva som vi, om vi bara ville, skulle kunna uppnå tack vare naturvetenskapernas framsteg. Alfred Nobel skulle kunna betecknas som radikal etisk idealist.

Åter till Strindberg! Eftersom han så tydligt markerar en semantisk skillnad mellan de båda adjektiven *idealisk* och *idealistisk*, ligger det nära till hands att undersöka, hur han själv använder dem. Nationalupplagan av Strindbergs verk är på god väg, och texterna finns i Språkbanken vid Göteborgs universitet. En genomgång av de körningar som har gjorts ger en mycket intressant bild av Strindbergs repertoar. Jag citerar några av fallen.

"De voro förträffliga ynglingar, idealiskt anlagda man kallade, med vackra uppsåt och svärmande för okända dunkla ideal. --- Där fanns ... en känslösam yngling som svärmade ännu, för alla de gamla idealen men mest Heine ..." (Tjänstekvinnans son, 1886). Här lägger man märke till att *idealisk* hänför sig till kontextens ideal, av vilka ett utpekats som Heine.

"Efter allt vad jag kan förstå bjuder Nora ut sig – mot kontant betalning. Det är ju idealiskt och täckt. Allt, naturligtvis av kärlek till mannen! För att rädda honom!" (Giftas, 1884.) Detta ställe ger liksom det följande exempel på Strindbergs användning i ironisk kontext av samma betydelse som i det första fallet. "Samhället hade nu nått höjdpunkten eller idealet av bakvändhet. Det nyttiga hade kommit i förakt och det onyttiga var ärat. --- Mest idealiskt utvecklade sig dock livet i städerna. --- Härtill bidrog även i hög grad det idealiska byggnadssättet. Man hade nämligen i stället för att lägga husen i terrasser mot solsidan, byggt dem i långa strängar, mellan vilka lupu gångstigar, och dessa så smala att endast ett fåtal av innevävarne hade ljust i sina rum." (Svenska öden och äventyr, 1883.)

"1a september i år lämnar jag fem nya original. Ett sorgespel en komedi ett lustspel två proverb. Kunna även lämnas tidigare.



Personalen endast åtta personer: en käring en fru två flickor (blond och brun) en gubbe du en älskare (idealisk) en d:o (realisk, ful)" (Brev till August Lindberg, 1887). Det ideal som frammanas här är motsatsen till *realisk* i sammanställning med *ful*.

"Christopher Jacob Boström var en kammarfilosof, en 'spekulativ filosof från Piteå' som infångat sig själv i ett filosofiskt system, daterande sig från Platos idealistiska världsåskådning. I korthet: vår sinnevärld saknar den högsta realiteten såsom varande blott en grumlig spegelbild av idévärlden vilken existerar på andra sidan graven." (Tal till svenska nationen, 1910.) Detta är ett exempel på användningen av *idealistisk* i den vanliga filosofiska betydelsen.

"Denna idealistiska uppfattning av konungadömet med Guds nåde, vilket förtolkats såsom om monarken stode över alla mänskliga lagar och endast sorterade under Guds högsta domstol, inplantades hos mina samtida, och fostrade ett trälsinnat, servilt släkte av biplaneter, som hämtade en bit makt uppifrån, och därigenom glömde att de voro tjänare." (Tal till svenska nationen, 1910.) I detta fall står *idealistisk* för en konservativ syn på livet och samhället, om än med en filosofisk bakgrund. De exempel jag har givit gör det tydligt, att båda adjektiven tillhörde Strindbergs aktiva ordförråd och att de hade klart olika betydelser.

Det finns likartade exempel hos andra författare, bland dem Carl Adolph Agardh, som 1857 skrev följande: "Om man icke betraktar den christna pligtläran såsom idealisk, d.v.s. såsom framställande ett oändligt mål, hvartill vi skola närma oss, så kunna vi icke rätt fatta densammas mening." (Samlade skrifter, 1.) Adjektivet får här en explicit angiven betydelse, ungefär 'som avser ett mål'.

Ett annat fall som jag anser hör hit finns hos filosofen Kristian Claëson i en text från 1888: "Vi påstå ej därför, att vår tid är idealisk. — — — Men en tid är ej en avslutad storhet; och hon får därför ej mätas blott efter vad hon gjort och är, utan även efter vad hon vill och kan bliva. Vår tid har satt sig höga mål före; hon går — åtminstone älska vi tro det — ännu framåt."

En intressant parallell erbjuder Strindberg själv: ”Men reformationen kom, hänsynslös, praktisk, materialisk, framgångsrik; avgudabilderna (som de kallades) blevo störtade och templet rensat, i synnerhet på ädla metaller, ty trätavlorna och träbilderna fingo mångenstans sitta kvar in på våra dagar, och sitta där än.” (Det nya riket, 1882.) Adjektivet *materialisk* används här i betydelsen ’som avser material’ och är alltså analogt med *idealisk* och *realisk*.

Det är uppenbart att *idealisk* i betydelsen ’som avser ett ideal, idealorienterad’ på Nobels tid användes av åtskilliga personer som ett naturligt inslag i deras språk. Lägg också märke till Strindbergs ord om de unga männen i ett av exemplen ovan: ”idealiskt anlagda *man kallade*”. Detta metalingvistiska uttalande gäller tidens språkbruk och får därmed ett särskilt intresse. Det är därför inte så överraskande som det har kunnat synas att finna det diskuterade ordet i testamentet.

Nu är det så, att just ordet *idealisk* är resultatet av en ändring som Nobel gjorde i sitt handskrivna testamente. De båda slutbokstäverna *sk* har skrivits direkt på några andra tecken. Detta har väckt min nyfikenhet – kanske kunde den ursprungliga lydelsen ge en ny utgångspunkt.

Jag gjorde åtskilliga försök att med mer eller mindre obehägnat öga komma åt det ursprungligen skrivna men beslöt snart vända mig till kriminalteknisk expertis. Jan-Erik Karlsson hade vänligheten att undersöka saken. Genom att utnyttja olika optiska metoder kunde han först konstatera, att allt bläck i det granskade området var av samma slag. Med hjälp av mikroskopi, kontrastförstärkning och teckenjämförelser kom han så fram till att det från början hade stått *idealirad*. Ett av de fotografier han tog återges här intill. Han avslutade sitt utlåtande med observationen, att vissa personer kan utelämna bokstäver, stavelser eller hela ord när de skriver, kanske genom att tanken löper före skrivandet.

Något *idealirad* finns emellertid inte i svenskan, och det är högst begripligt att Nobel ändrade. Det ord som föresvävade honom i skrivögonblicket var säkerligen *idealiserad*. Varför ändrade han då inte till det? Han kunde ha skrivit *se* ovanför raden

och markerat var det skulle in med hjälp av ett invisningstecken. Så gjorde han t.ex. med *af*, som skulle föras in mellan *freds-förfäktare* och *ett utskott* några rader längre fram i testamentet. Min tanke är, att han inte var nöjd med *idealiserad* – det kanske var hans funderingar på det som vållade felskrivningen – utan ville ha ett ord som markerade lyftning utan att så tydligt associera till försköning. Med några raska drag skrev han så *sk* på slutbokstäverna *rad* och hade därmed åstadkommit *idealisk*.



BILD 1. Rättelse i Alfred Nobels testamente.

Inom språkvetenskapen brukar man skilja mellan klassificerande adjektiv och karakteriserande adjektiv. Exempel på de klassificerande är *ortografisk* 'som avser ortografi(n)' i *ortografisk reform* och *fysikalisk* 'som avser fysik(en)' i *fysikaliskt experiment*. Karakteriserande å andra sidan är exempelvis *degig* 'som är som deg' och *lymmelaktig* 'som är som en lymmel'. Jag tänker mig att *idealisk* hos Nobel, liksom hos Strindberg, Agardh och Claëson, är ett klassificerande adjektiv, som alltså har betydelsen 'som avser ett ideal, idealorienterad'.

Lösningen tycks alltså vara, att Nobels uttryck "i idealisk riktning" betyder 'i riktning mot ett ideal, i idealorienterad riktning'. Idealets räjong utpekas i sin tur av Nobelprisernas grundkriterium mänsklighetens nytta. Detta betyder exempelvis, att skrifter som förespråkar låt oss säga folkmord inte är i linje med testamentet, hur briljanta de än må vara. Jag kommer alltså från mitt filologiska håll fram till en helhetssyn som resultatet av Torgny Segerstedts idéhistoriska resonemang väl låter sig förenas med.

Gunnar Brandell frågade sig i den artikel jag var inne på tidigare, hur man skall kunna avgöra om ett slags litteratur är mer

”idealisk” än ett annat slags litteratur. Det blir man emellertid befriad från. Klassificerande adjektiv har absolut betydelse och kan alltså inte kompareras utom under mycket speciella omständigheter, och de är inte aktuella här. Komparationen är i stället relevant i det uttryck som föregår ”i idealisk riktning”, alltså ”det utmärktaste”.

När man väl har skaffat sig en uppfattning om hur saken kan ha tett sig för Nobel, blir det varje generations sak att bedöma, hur tidens litteratur åstadkommer detta. Det är kanske inte givet, att Nobel själv skulle ha sett Becketts verk som vi ser det, för att ta ett exempel. Och från vår horisont hade möjligen den förste Nobelpristagarens namn sett annorlunda ut.

### Svenska Akademien och Nobelpriset i litteratur

Akademiens diskussion de första Nobelåren med ständige sekreteraren Carl David af Wirsén som tongivande ledde trots balanserande inlägg av Esaias Tegnér d.y. och andra till en idealistisk tolkning av testamentet. Det skedde i den anda som karakteriserade den ovan nämnde filosofen Christopher Jacob Boström, den tyske estetikern och idealrealisten Friedrich Theodor Vischer och med dem alltså Wirsén. Det är detta Strindberg har iakttagit och vänt sig mot.

Naturligt nog reflekteras Akademiens hållning i prismotive-ringarnas historia. Här följer de formuleringar som innehåller *ideal(-)* med engelska motsvarigheter.

	Original	Auktoriserad översättning
1901 Prudhomme	hög idealitet	lofty idealism
1908 Eucken	ideal världsåskådning	idealistic philosophy of life
1909 Lagerlöf	ädel idealitet	lofty idealism
1910 Heyse	ideal uppfattning	idealism

1911 Maeterlinck	poetisk idealitet	poetic fancy
1916 Rolland	upphöjd idealism	lofty idealism
1926 Shaw	idealitet	idealism
1927 Deledda	hög idealitet	idealistically inspired

---

I nästan var tredje motivering under de tre första decennierna användes dessa ord på ett eller annat sätt.

Andra aspekter blev emellertid synliga efterhand. Med utnyttjande av tillgängliga källor såsom utlåtanden och brev har Kjell Espmark i den bok jag nämnde undersökt kriterierna bakom prisbesluten. Hans upplysande studie har visat på en serie begrepp som har dominerat under de olika perioderna, såsom litterär neutralitet (under första världskriget, något som gynnade nordiska författare), den stora stilen (i Goethes efterföljd), allmänmänskligt intresse (i motsats till nationellt eller på annat sätt begränsat), banbrytare (varigenom innovativt och exklusivt författarskap uppmärksammades) och pragmatisk besinning (som lyfte fram betydelsefulla men okända författarskap, då ingen kan anses vara bäst när det inte finns någon gemensam måttstock). Detta är sålunda olika slags interna argument som har förts fram vid prisdiskussionerna genom åren.

Å andra sidan utgör de externa prismotiveringarna en officiell källa, som ger oss Akademiens explicita uttalanden om skälen. Ett studium av dem visar helt naturligt reflexer av de underliggande kriterier som just nämndes. Men återspeglingsarna är inte ymniga, och bilden är något komplex. Likväl kan man iaktta ett antal återkommande argument.

Det utan jämförelse vanligaste inslaget i prismotiveringarna är att de refererar till författarens hemland eller ibland världsdel. Det gäller mer än en av tre motiveringar, från Grazia Deleddas plastiskt åskådliga skildring av livet på hennes fäderneö Sardinien, Anatole Frances franska lynne och Yasunari Kawabatas fina känsla för japanskt väsen i dess egenart till Toni Morrisons visionära kraft och poetiska pregnans som levandegör en väsentlig

sida av amerikansk verklighet och Pablo Nerudas naturkraft som levandegör en världsdels öden och drömmar.

Särskilt under den första tiden, och i varje fall inte senare än Sjolochovs "Donska epos" i motiveringen från 1965, angav Akademien ofta ett enskilt verk som huvudskäl till priset. Ytterligare åtta böcker nämns på det sättet, nämligen Mommsens "Römische Geschichte", Spittelers "Olympischer Frühling", Hamsuns "Markens grøde", Reymonts "Bönderna", Thomas Manns "Buddenbrooks", Galsworthys "The Forsyte Saga", Martin du Gard's "Les Thibault" och Hemingways "The Old Man and the Sea".

Bara en gång betecknas en pristagare i motiveringen som banbrytare, och det gäller T.S. Eliot. Men Pirandello prisas för sin djärva och sinnrika nyskapelse av dramats och scenens konst och Faulkner för sin kraftfulla och självständigt konsnärliga insats i Amerikas nya romanlitteratur. Dessutom gottskrivs Hemingway och Sartre det inflytande de har utövat.

Omvänt belönades Benavente för det lyckliga sätt varpå han hade fullföljt den spanska dramatikers ärofulla traditioner. I samma anda nämndes i motiveringarna för prisen till Bunin, Pasternak och Solsjenitsyn deras insatser inom ramen för den ryska litteraturens klassiska/stora/omistliga traditioner.

Den synpunkt som oftast har förts fram under det senaste halvsekle – ett dussin gånger sedan Gide 1947 – är skildringen av mänskliga villkor. Man finner den till exempel i motiveringen för Beckett, vars författarskap hämtar sin konstnärliga resning ur nutidsmänniskans blottställdhet, Singer, som levandegör universella mänskliga villkor, Golding, vars romaner belyser mänskliga villkor i dagens värld, och Oe, vars verk alltså ger en skakande bild av människans belägenhet i nutiden.

Idag kan vi notera en dragning *tillbaka till testamentet* i ljuset av ansträngningarna i det förflutna. Betrakta som exempel motiveringen för Nadine Gordimer: "som genom storartad episk diktning har – med Alfred Nobels ord – gjort mänskligheten den största nytta".

Nobel försåg oss i sitt testamente med fem kriterier. Tre av dem är av allmänt slag och gäller för samtliga fem priser. Två är speciellt utformade för det litterära priset.

(1) "till dem som ... hava gjort mänskligheten den största nytta". Detta är det grundläggande kriteriet, presenterat redan i det aktuella styckets första mening.

(2) "under det förlupne året". Av uppenbara skäl tolkas detta så, att författaren skall vara i livet vid nomineringen – inget utrymme för Shakespeare – och att författarskapet skall vara av fortsatt intresse.

(3) utan att "avseende fästes vid någon slags nationalitetstillhörighet". Detta var ett framsynt kriterium som ingalunda var förenligt med det slutande 1800-talets nationalromantiska strömningar. Under det senaste tiotalet år har det kunnat få vidsträckt tillämpning.

(4) "till den som har åstadkommit ... det utmärktaste". Det säger sig självt, att detta betyder litterär excellens.

(5) "i idealisk riktning". Detta innebär enligt tolkningen ovan, att riktningen är idealorienterad inom ramen för det första kriteriet.

Tack vare Nobelstiftelsens effektiva förvaltning av Nobels förmögenhet har vår Akademi resurser att engagera ett nätverk av experter för utlåtanden rörande vilket kulturområde som helst. De som är berättigade att nominera kandidater och utnyttjar denna förmån bildar ett annat internationellt nätverk. Men grunden för utvärderingen är Akademi-ledamöternas läsning. Prioriteringar är naturligtvis subjektiva, men de uppkommer ur gruppens vida erfarenhet och prövas i grundliga diskussioner vid sammankomsterna. Ledamöterna är väl medvetna om att Nobelpriset rykte vilar på listan över pristagare.

## **Framtiden och Nobelpriset i litteratur**

I en essä i *Scientific American* i oktober 1994 skriver Marvin Minsky: "As a species, we seem to have reached a plateau in our intellectual development. There is no sign that we are getting

smarter. Was Albert Einstein a better scientist than Isaac Newton or Archimedes? Has any playwright in recent years topped William Shakespeare or Euripides?" Hans svar är nej, och det tillfredsställer honom inte. Vi borde finna vägar att överskrida våra intellektuella gränser. Han menar, att detta kan göras genom konstruktion av utbyteskroppar och utbyteshjärnor med hjälp av nanoteknik. En option blir då odödlighet.

I ett första skede kommer forskarna att öka vår hjärnkapacitet genom att ansluta elektroniska apparater till våra hjärnor med hjälp av stora mängder mikroskopiska elektroder. Efterhand kommer de emellertid att kunna ersätta varje del av hjärnan. Därigenom gör vi oss till maskiner, konstaterar Minsky. Dessa nya, intelligenta maskiner föredrar han att kalla våra tankebarn efter Hans Moravec i boken "Mind Children: The Future of Robot and Human Intelligence", 1988. Eftersom chips är miljontals gånger snabbare än hjärnceller, kan tankebarn formges så, att de tänker oerhört mycket snabbare än vi gör. För ett sådant väsen blir enligt Minskys uppskattning en halv minut lika lång som ett av våra år och varje timme lika lång som en människas hela liv.

Som jag ser det kan en maskin eller robot möjligen programmeras att i viss utsträckning simulera känslor som medlidande eller tacksamhet eller sorg, men den är troligen oförmögen att erfara dem. Skulle den någonsin kunna känna sin puls slå av upphetsning? Som Douglas Hofstadter påpekar i sin bok "Fluid Concepts & Creative Analogies. Computer Models of the Fundamental Mechanisms of Thought", 1995, är det något helt annat att få en datamaskin att känna tacksamhet än att få den att säga tack. Simulering av exempelvis känsla, avsikt, identitet är enligt min mening inte nödvändigtvis känsla, avsikt, identitet i egentlig mening.

Härvidlag instämmer jag helhjärtat med Roger Penrose i hans bok "The Emperor's New Mind. Concerning Computers, Minds, and The Laws of Physics", 1991, med dess anspelning på H.C. Andersen. Även om mycket av det som faktiskt inbegrips i vår mentala aktivitet kan vara algoritmiskt, hävdar han, måste det finnas ett väsentligen icke-algoritmiskt inslag i medvetandets funktion. Det är detta som David J. Chalmers i en artikel i Scientific



American i december 1995 kallar medvetandets svåra problem, frågan om hur fysiska processer i hjärnan kan ge upphov till subjektiv erfarenhet.

Bland annat uppskattar jag också Roger Penroses påpekande om vårt moraliska ansvar i det fall, att det fanns en maskin som påstods vara ett tänkande, förnimmande, känsligt, förstående, medvetet väsen: "Simply to operate the computer to satisfy our needs without regard to its own sensibilities would be reprehensible."

Dessutom förefaller mig en annan fråga inställa sig. Förväntas dessa antropomorfa maskiner kunna hysa kärlek och älska? Med tanke på det överflöd av erfarenheter, nyanser, intryck, tolkningar osv. som inbegrips i detta mycket mänskliga fenomen anser jag att svaret är nej. Lägg också märke till att ett utbyte av hjärnan påverkar mottagarens identitet, en fundamental faktor i en relation som kärlek. För mig antar detta skepnaden av ett avgörande argument.

Överträffa Shakespeare? Nåja, även om det inte är en förutsättning, skulle det vara mycket välkommet. Vad robotar beträffar är knappast någon verklig konkurrens att förvänta. Skulle för övrigt en maskin uppskatta ett pris? Låt oss fröjdas åt Wisława Szymborskas visdom i "Notes from a nonexistent Himalayan expedition":

We've inherited hope –  
the gift of forgetting.  
You'll see how we give  
birth among the ruins.

Yeti, we've got Shakespeare there.  
Yeti, we play solitaire  
and violin. At nightfall,  
we turn lights on, Yeti.

## Den goda smakens återupprättande. Om akademiernas roll på tjugohundratalet

Svenska Dagbladet 9 oktober 1997. Artikeln grundar sig på ett högtidstal hållet vid den katalanska akademiens 90-årsjubileum i Barcelona 1997.

För några år sedan gjorde Svenska Akademien ett besök i Prag. Vid det tillfället förenade sig Vaclav Havel med oss i en skål för den goda smakens återupprättande. Det förefaller mig som om denna spontana gest är värd att tänka på en gång till. I själva verket är ju "Snille och smak" det motto Svenska Akademien fick 1786 av sin grundare Gustaf III. Uppenbarligen finns det ett nära samband mellan begreppen smak och akademi.

Vad är då smak? Enligt Nationalencyklopedins ordbok betyder ordet 'uppfattning om vad som är vackert, passande eller bra' som i *ha dålig smak, var och en väljer efter sin smak*. Det används också om en uppfattning som leder till speciell uppskattning av en viss företeelse som i *en internationell publik med smak för det eleganta*. Det finns alltså en typisk grundbetydelse som har vida tillämpningar. God smak framstår i många stycken som en hållning.

När den norske filosofen och ekologen Arne Næss fick Svenska Akademiens nordiska pris förra året, höll han ett tal där han pekade på vad han betraktade som mänsklighetens tre största uppgifter under nästa århundrade. De var att förhindra krig, att utrota fattigdom och att lösa den ekologiska krisen. Det är inte min tanke att fokusera dessa tre värn, men jag är övertygad om att mitt tema har avsevärd betydelse i sammanhanget.

Det har försports att ett antal rymdsonder är på väg ut ur vårt solsystem. De för med sig diverse budskap från jorden – ifall. På Voyagersonderna finns det videor med över hundra foton, mer än en timmes musik från olika håll, hälsningar på sextio språk och låtet av en val. Denna rymdens flaskpost kan behöva våra välgångsönskningar.

Det är inte överraskande att ett av de sextio språken ombord är engelska. Men detta språk har exempelvis inget ord för just flask-

post, inget för allemansrätt och, naturligtvis, inget för uppehållsväder. Nog var det motiverat att skicka med femtionio språk till.

I detta sammanhang kan man komma att tänka på Drakes ekvation. Idén är att försöka uppskatta antalet intelligenta civilisationer i vintergatan, vår galax. Ekvationen innehåller en handfull parametrar, vilkas värden är någorlunda kända i vissa fall, ren gissning i andra, t.ex. sannolikheten för att liv uppkommer och utvecklas till en högteknologisk civilisation. Beroende på vilka värden som väljs kan resultatet bli allt mellan endast en civilisation och tio miljarder civilisationer. Vi kan med andra ord vara unika. Låt oss betrakta denna möjlighet som ett gynnsamt tillfälle!

Faktiskt har vi åtminstone femtusen språk på jorden. En förskräckande stor del av dem håller nu på att passera gränsen mellan levande och utslocknat. Detta är en stor kulturell fråga med tanke på att varje språk representerar ett särskilt sätt att uppfatta och förhålla sig till verkligheten. Genom sin märkvärdiga komplexitet är språken till stor hjälp i undersökningar av hjärnan och dess organisation liksom av medvetandet. Förlusten av språk reducerar våra chanser att vinna dessa cerebrala insikter.

Oberoende resultat från historisk språkvetenskap, genetik och arkeologi bidrar till att ge människosläktets historia en ny kontur. En av de mest slående iakttagelserna är kongruensen mellan språk och etniska grupper. Detta innebär att språklig mångfald är en viktig källa till kunskapen om utvecklingen av *homo sapiens sapiens*. Förlusten av språk underminerar också detta historiska studium. Vi måste se till att säkra språkens överlevnad.

På åttiotalet skrev John Naisbitt sin bok *Megatrends*, där han diskuterade tio nya tendenser som kunde förväntas omvandla våra liv. En av dem var globaliseringen av ekonomierna, vilket han menade skulle komma att åtföljas av en renässans i fråga om språklig och kulturell medvetenhet. Svenskarna skulle bli mer svenska, kineserna mer kinesiska och fransmännen, Gud hjälpe oss, mer franska. Vid sidan av detta skulle engelskan fortsätta att växa som affärsspråk. För att bli verkligt framgångsrik, tillade han, borde man kunna tala flytande med datorn.

God smak innefattar behärskning av ens eget språk, och eftersom det bara finns ett språk som var och en verkligen behärskar, är varje språk en nådegåva. Den oundvikliga slutsatsen är återigen att vidta åtgärder för att stödja vart och ett. De som talar x-iska, ett språk med begränsad spridning, skall aldrig mer tvingas se eller höra sådant som "Förbjudet att spotta och tala x-iska".

Bör Europeiska unionen vara en organisation, där man kan missförstå varandra på elva – eller fjorton eller arton – språk? Mitt svar är ja, utan tvekan. Alternativet är oacceptabelt: ett eller två språks överhöghet och som en följd därav en eller två kulturella varieteters. Mångfald är rikedom. Den pluralistiska linje som EU har antagit är värd varje peseta.

Låt mig särskilt understryka, att det inte finns något sådant som ett litet språk. Det uttrycket är stötande för dem som talar ett språk som betecknas så och alltså har det som en integrerande del av sin personlighet. Det är en historisk tillfällighet att vissa språk talas av fler människor än andra. Och idén att reducera antalet språk genom att välja t.ex. ett av de nordiska språken är lika absurd som att välja exempelvis det ena av portugisiska och spanska eller nederländska och tyska.

Akademier och universitet bör anta denna utmaning i den goda smakens namn. Det behövs mer av översättar- och tolkutbildning, projekt inom datorstödd översättning, kurser i skriftlig framställning, mångkulturella studier osv. Som alltid förutsätts historiska insikter, minne.

En fotnot: för dem som talar ett språk med begränsad spridning finns det däremot inget behov av att stå där och knacka på dörren och be att få bli insläppt. Varje kulturfär har intressanta ting att bidra med. Öppenhet är i allas intresse.

I stället för att be är det tid att handla. Akademiska institutioner måste satsa på språkets utforskning och odling. Att upprätta och konsolidera språkbanker, att utveckla redskap för databehandling av språk, att utarbeta detaljerade lexikaliska databaser är hörnstenar i sådana program. Det är också EU:s initiativ att skapa ett nätverk som Parole mellan centra som arbetar efter dessa linjer.

Språksamfundet måste uppmuntras att använda sitt eget språk. Detta innebär, att jag särskilt påminner om den fjärde och sista

av de huvudprinciper som har förts fram i den allmänna diskussionen: identitet, kreativitet, mångfald och medverkan. Författare spelar en betydelsefull roll härvidlag som professionella språkanvändare. Det är också mycket viktigt att sätta press på dem som är ansvariga för valet av språk i gränssnittet i olika slags dataprogram.

Yttrandefrihet är grundläggande i vårt överordnade perspektiv och förvisso i akademiernas liv. Den bör underbyggas inte genom politiska uttalanden eller utspel utan genom synen på denna frihet som en självklar del av den plattform man arbetar från. För att ta ett exempel är detta skälet till att Nobelpriset i litteratur kan utdelas som ett litterärt pris utan sidohänsyn.

Enligt Mark Twain är alla bibliotek som inte har Jane Austens böcker bra bibliotek, även om de inte har några andra böcker. En dynamisk tolkning av detta är, att alla rum är potentiella bibliotek. En annan, mindre dynamisk men observerad, är att inga böcker behövs i datorernas tidevarv. Men systemen för behovstryck – för att inte nämna alla de andra argumenten – visar att så inte är fallet.

Detta för mig tillbaka till datasystem och språk. Några av språken, t.ex. katalanska, tyska och tjeckiska, har ett antal specialtecken i sina alfabet, ibland kallade nationalkaraktärer. I själva verket tycks de flesta skriftspråk ha det. Inte desto mindre låtsas man inom system för datakommunikation, att dessa tecken inte existerar. I svenskan och åtskilliga andra språk är ju bokstäverna ä och ö vanliga. När vi får e-post, får vi stå ut med ersättare som tecknen större än eller lika med ( $\geq$ ) respektive promille (‰). En sådan försumlig och kortsynt attityd måste motarbetas tills den korrigeras. Det är inte längre fråga om tekniska begränsningar. Dessvärre ser det ut som de obotfärdigas förhinder.

I jämförelse med detta är vad som är inne bland dem som pratar på Internet ett mindre problem. Lustifikationer som *cul8r* för *see you later*, *LOL* för *laughing out loud* och *IMHO* för *in my humble opinion* kommer och går som sådana ting gör. Det som verkligen är nedslående är den vårdslösa syntax och stavning som man ofta ser. Smaklöst.

Det finns ändå några ljuspunkter i datorernas värld. Jag tänker särskilt på visitkort i skorna. Vanliga kort behövs inte längre. Så snart man skakar hand, överförs den relevanta informationen genom kroppen till en dator i klacken. Vår fysiska kapacitet är omkring 100 000 bitar i sekunden, vilket är fullt tillräckligt för informationen på ett kort. Datorn drivs av den energi som frigörs när man går. Väl hemma eller på kontoret kan man skriva ut sina klackdata och se vem man har skakat hand med.

Nyligen rapporterade Ronald E. Buckman från Seattle om en talande erfarenhet. Som svar på önskemål från ingenjörshåll hade han utarbetat en kurs i ingenjörspraxis, som bland annat omfattade ingående diskussioner om vardagens etik. En utgångspunkt var, att studenterna genom sin uppfostran och utbildning förstod skillnaden mellan rätt och fel.

Lärarna märkte snart, att detta antagande var ogrundat. Några av de skäl som kom fram var familjens försvagning, den minskande fokuseringen på religiösa värden, bristande ansvarskänsla, alltför liberaliserad utbildning och kraftigare betoning på det egna jaget. Detta hade lett till en situation, där förvrängningar i nyhetsmedierna, dubiös reklam, oetisk affärspraktik och åsidosättande av hävdvunnen moral stillatigande accepterades av samhället. Man fick helt enkelt utarbeta en ny kurs i tillämpad etik.

Jag tror det är dags att vi påminner oss om Nobelpristagaren Joseph Brodskys maxim: estetiken är etikens moder. För samhället gäller det, skulle jag vilja säga, att åstadkomma en kaskad av sinnesintryck i skolan genom ökade möjligheter till läsning, skrivning, teckning, musikutövning, slöjd, gymnastik. Sensibilitet förtjänar att senteras.

Ett sådant program kräver en förändring av dagens politik. Kulturen måste återfå en central plats. Detta betyder, att jag helhjärtat stöder tankarna i Förenta Nationernas rapport "Our Creative Diversity" och Europarådets rapport "In From the Margins". Den senare har ett dubbelt syfte: att föra miljoner av ogynnsamt ställda européer in från samhällets marginaler och att föra kulturpolitiken in från maktens marginaler. För min del betraktar jag språken – informationsbärare och konstnärliga medier – som grundläggande instrument i detta företag.

Medan opinioner och attityder tenderar att växla, ofta mellan extrema punkter, kan självständiga akademier arbeta på en solid bas av erfarenhet och rentav sunt förnuft. Dyrbarheter som guld och silver är resultat av katastrofer som stjärnexplosioner. För dyrbarheter som dem jag har diskuterat här borde inga explosioner vara nödvändiga.

## **LEXIS: Opening Speech at the Seventh Euralex International Congress, Göteborg, August 13–18, 1996**

International Journal of Lexicography 10: 4. 1997.

Ladies and gentlemen, dear colleagues and friends,

It is a pleasure to join in welcoming you to Göteborg University. This is a city university with five faculties and a group of artistic departments. At its start last century there were 7 professors and 21 students. This wonderful teacher/student ratio has changed drastically, and nowadays our Alma Mater is one of the four big universities of Sweden with more than 200 full professors and 34,000 students.

Overlooking this magnificent Euralex audience, I can feel my verbal heart beat with joy. The two volumes of congress papers bear witness to the width and depth of the field, with topics like dictionary, lexicon, lexical database, thesaurus, dictionary for learners, children's dictionary, corpus, frequency, lemmatization, tagging, collocation, idiom, metaphor, figurative meaning, polysemy, disambiguation, lexical function, synonym, etymology, meaning, definition, cultural boundness, obscurity, creativity, theoretical basis, bilingual dictionary, parallel corpora, network, translation, processing, tool, neologism, dialect, slang, insulting items, agnonyms, terminology, morphology, compound, noun, verb, adjective, interjection, connective, grammar, passive, active. I think we have got a lot to do the next few days.

Allow me to strike a modest note. I think that non-trivial natural-language processing – in particular corpus work, lexical work, and tool development – has contributed, significantly so, to saving the science of linguistics from losing its foothold. We have seen a certain amount of venturesome speculation that hardly stands the test of real language. That is why a steady cry for *authenticity* has been emitted from here.



On the other hand, it goes without saying that corpus work and lexical work as scientific domains involve theoretical considerations. This means that I put in a word for cross-fertilization between empirical and theoretical efforts. Using another metaphor, we can think of it as meandering *à la grecque*.

A theoretical issue of considerable interest is Michael Halliday's proposition in a recent paper that grammar and vocabulary are the same thing seen by different observers and that grammar is to be treated as inherently probabilistic. As a corollary, vocabulary, being the same thing as grammar, is evidently probabilistic, too. I agree up to the point that lexicon and grammar are two descriptive aspects of the same entity, but I wonder whether each lexical unit of language really carries some probability in the true sense of that term. Actually, Halliday talks, in very general terms, of the system's probability profile and of registers as variations in probabilities. Still, it seems to me that the occurrences of linguistic entities in language manifestations are probably not governed by quantitative attributes in the language system but by a very complex set of factors at work in infinitely changing situations, with their endlessly varying conditions. When the frequency of a word like Swedish *bogvisir* 'bow-visor' suddenly soars in many types of text – and remains soaring – it is not due to a change in the system but to an accident at sea. Admittedly, the tendencies discussed by Halliday are interesting enough, but that is another story. I put in as an aside here that I have talked this over with him.

What, then, is a corpus? In my view, it is a body of text for linguistic investigation or inspection. I think we should be careful in this connection, reserving the term for specific purposes. There is a clear relation between the concepts of corpus and frequency – and thus probability – where frequency is the number of occurrences in a corpus. Consequently, a corpus is finite in size, and frequency figures are exact and interpretable through their dependence on the corpus.

From this point of view, it seems we are invited to reconsider the use of a term like *corpus* in a phrase like *monitoring corpus*, specified as a huge stream of language in motion. Of no finite

size, such a current flows across a set of filters extracting linguistic evidence, and then the text is discarded. This means that the strict concept of corpus tends to dissolve and its precise relation to the concept of frequency as well.

Of course I am all for the idea of a monitoring setting as such. In the late sixties and early seventies Kolbjørn Heggstad and I talked about the possibility of being connected to a news agency and making a computer record new words and forms. In 1972 I wrote a paper called *The Stream of Words as a Flow of Data*, where I argued for the tapping and study of the wordstream. In the language-bank framework this is mainly realized as off-line tapping.

This brings me to another theoretical issue of great importance, the question of representativity. Randolph Quirk has recently argued that if we evade the challenge of total accountability, our use of a corpus will be no advance on Jespersen's slips or Murray's file boxes. With respect, I must say I disagree with this somewhat pointed statement. As another aside: he knows I do. Now, total accountability would mean responsibility for covering all aspects of the language studied, in minute detail. For a long time I have been convinced that this is impossible. No one has ever described the population to be sampled in the way needed. What proportion of, e.g., lyric verses by Philip Larkin and TV reports of Olympic beachvolley should be included in the sample? I notice that Sue Atkins and colleagues have made clear, in a recent paper, that, in their opinion, too, the collection of a representative sample of total language production is not feasible.

My interpretation of Sir Randolph's brave claim is that he wants us to aim high. Notice that he talks about the *challenge* of total accountability. When he goes on to tell us that the corpus he is building will be underpinned by a detailed design specification, so that it will be as representative as possible of modern British English, he has in fact transformed the demand for total accountability into very good representativity. And now we all agree.

On this occasion, I think it is fair to say a few words from my horizon of what is going on here in the field. Let me do so by presenting some basic ideas underlying the Swedish Dictionary of the National Encyclopaedia, the third and final volume of which was published last month in co-operation with Bra Böcker Publishing Group. One of the starting-points was the understanding that an encyclopaedia, describing the world, and a dictionary, describing language, together form a powerful source of knowledge.

Since the mid-seventies we have been busy making up a lexical database resting on theoretical, methodological and technical notions developed for the purpose. This is of course quite different from the more common procedure of transmuting a printed dictionary. The first book printed from the database was the one-volume Swedish Dictionary of 1986 in co-operation with Norstedt Publishing House.

Fundamentally, our dictionaries are organized in accordance with the lemma-lexeme model. The headwords are lemmas, based on criteria of spelling, pronunciation, inflexion, and part of speech. Each lemma has one or more lexemes, which are the main units representing discrete basic meanings. Each lexeme can have zero, one, or more nuances in terms of extended, specialized, transferred, or figurative meanings. The dictionary has a cyclic structure, in that each lexeme gets the same types of information in the same order.

The new dictionary is a defining dictionary but it contains a characteristic set of information types not usually found in such works. Our ambition has been to present as full a picture as possible of the entries for the benefit of the users. Thus, the specification of pronunciation includes the tone accent. There is a morphological analysis of each lemma (for use also in word-division), and it includes a signal of the main morphological boundary in compounds. For each verb the corresponding verbal nouns are specified. A selection of semantic relations between lexemes is presented, involving co-hyponyms, hyponyms, hyperonyms, and synonyms. There are features of usage like style, subject domain, and notes on Finland-Swedish differences.

Morphological and syntactic examples are authentic as far as possible. Idioms in a narrow sense of the term are recorded to a large extent. The description of verbs, nouns, and adjectives includes valencies. There is also a historical part accounting for the year of the first occurrence in the language and the etymology of each lexeme. Finally, the margin is used for quotations of aphorisms, repartees, verses, etc., giving further aspects of the sense potential and cultural environment of the entries. There are no pictures – so typical of encyclopaedias – in the dictionary. We rely on words as our meta-language, taking care to avoid narrow circles.

A prerequisite for the lexical database and the dictionary based on it was the work done in other projects from the sixties on. In fact, the first concordance was produced in 1961–62 as part of my dissertation. Two milestones were the project presented in print as the *Frequency Dictionary of Present-Day Swedish 1–4* (1970–80) and the establishment of the national Swedish Language Bank at the department in 1975. On the basis of a corpus created in 1965 by means of paper tapes from newspapers, the four volumes of the *Frequency Dictionary* account for (1) graphic words and homograph components, (2) lemmas, (3) collocations, (4) morphemes and meanings. The introductions to the four volumes are all written in English as well as in Swedish. From this experience, we recognize many of the problems dealt with in papers submitted to this congress. There are eternal challenges.

In the course of these and other activities, the research group known as *Språkdata* was set up in 1966. Our thirty years' celebration took place a couple of months ago. The group is now one of the partners in the Department of Swedish. In 1972 a chair in natural-language processing was established. At present it is vacant and announced in natural-language processing with special reference to lexicology. As a matter of course, lexicology, corpus work, and tools are included in the post graduate programme.

Two long-term projects are run here in co-operation with the Swedish Academy, the linguistic and literary Academy whose

main objective is to work for the clarity, vigour, and standing of the Swedish language, although it is also in charge of the Nobel Prize in Literature. The purpose of one of the projects is to convert the historical dictionary of the Academy – 31 volumes from A to S so far – into a database by means of optical reading and systematic structuring. A complete concordance is available on the Internet as well as a number of lists covering, e.g., compounds and derivatives. There is a brand-new report on this undertaking available at the congress and there is a paper by Yvonne Cederholm. The other project is the elaboration of successive editions of the one-volume glossary of the Academy, functioning as the norm for the spelling, pronunciation, and inflexion of present-day Swedish. This task is accomplished by Martin Gellerstam and colleagues in close co-operation with the Language Committee of the Academy.

There is also a faculty programme, involving several departments, which is devoted to research in lexicology. This programme is co-ordinated by Martin Gellerstam. One of the participants is Göran Kjellmer, the man behind the thorough investigation of the Brown Corpus published as *A Dictionary of English Collocations*.

As far as publications are concerned, Sven-Göran Malmgren of the Organizing Committee is one of the two editors of *LexicoNordica*, the journal of the Nordic Association for Lexicography. In fact, all the members of the Committee are of course active in the field. Furthermore, with Ladislav Zgusta and Ulrich Heid among us, we are three speakers at this congress who are editors of the series *Lexicographica*. Let me add that I recently edited a book called *Of Thoughts and Words*, which contains the proceedings of a Nobel Symposium on the Relation Between Language and Mind.

It seems to me that collocations in a wide sense make one of the most intriguing wonders of language. There are several reasons for this. One is the enormous number of them in running text. I shall never forget the moment when I realized that there were 660,000 instances of recurrent word combinations in the one-million corpus of the frequency dictionary I mentioned. I

like to think of collocations as the *ostinato* of language as opposed to the *capriccio*. Another reason is the fact that some of them are vastly more common than expected, i.e. as compared with the theoretical frequency calculated on the assumption that the material were ordered in a completely random fashion. By way of example, the Swedish combination *på ett helt annat sätt* meaning 'in a quite different way' was found to be 180 million times more common than expected. A third reason is that there are so many types of them: from idioms in the strictest sense of the word, like *to carry coals to Newcastle*, to everyday word-strings like *in any case*, which just happens to be the wording required for idiomatic phrasing. Incidentally, the Swedish idiom corresponding to *to carry coals* etc., means 'to offer bread to bakers' children'.

Furthermore, collocations form an important link between lexicon and grammar. The concept of lexicalization indicates a kind of sub-field in its own right. It also seems to me that the wealth and diversity of collocations ought to influence theoretical models more profoundly, at least those which aim at psychological relevance. In algorithmic text analysis collocations still await their full-scale exploitation, surprisingly enough.

There are several other aspects that could be brought up. I shall touch upon a few. In a way, a text is a kind of dictionary of itself. The main drawback is that it is impractical. A concordance is of course a great help. But a concordance is not enough either; witness hapax legomena in the text. A dictionary based on wider vistas is necessary. But a dictionary is limited in size. The point is that a dictionary connected to a corpus is a very strong couple. And this applies to bilingual as well as monolingual relations. The tendency in this direction is one of promise.

It is clear that there is a very great general interest in language. The lexical ability of readers should not be underestimated. But this ability should not be overestimated either. The use of dictionaries in a competent way – we can call it lexical maturity – is not as widespread as desirable, not even always in circles where it could be expected. And it is wellknown that introductions to dictionaries are not very much frequented. In other words, we

have got an immense pedagogical problem. And this is not alleviated, I am sorry to say, by the fact that there are questions we have difficulty clarifying, like the following: To what extent does a descriptive dictionary tend to be prescriptive?

George Orwell's creation of *duckspeak*, *doubleplusgood*, etc., has already brought up the problem of what to record in a general-purpose dictionary. In the days of PC, in the secondary sense of 'politically correct', we are now confronted with the question of recording things like *herstory* as an alternative to *history*.

It is evident that our field has got a new dimension in the society of today. Time and again the European Union has stressed the importance of protecting and promoting the rich linguistic diversity of Europe with about fifty national and regional tongues. Languages are recognized by the Union as a key component in the evolving information society. The Commission supports the development of multilingual tools and services like translation, as well as centres of language resources and a network connecting them. Our department is a node in the network as an active partner in projects like Parole and Aventinus. We have also been entrusted with the evaluation of projects, in the first place Eurotra. The Union's attitude is of special importance to such languages as are spoken by relatively limited numbers of people. Notice that I do not speak of "small" languages. It is not by historical coincidence, like the number of speakers, that a language is great. I trust that you back me up in this. Anyway, we are strongly encouraged to carry on in our lexical art, documenting our languages and relating them to others.

Not long ago, a group of six researchers and industrialists sent a common letter to Mr. Santer, President of the Commission of the European Union, against the background of the rapid development of international networks – more specifically the Internet – and the linguistic and cultural consequences of this type of communication. The letter was signed by Danzin and Quemada of France, Oakley of Great Britain, Zampolli of Italy, Martín-Municio of Spain, and myself. We drew attention to the fact that non-specialists can take advantage of the new possibili-

ties, which means a profound change in the state of native languages and cultural influence. It is necessary, we said, to take action now in order to pay respect to the richness of our continent, not least in the light of what has taken place in North America and the Far East. The newly established European Language Resources Association, ELRA, of which our group is a member, can serve, I hope, as a counterbalance in this context. The issue I have touched upon here is, in my view, the crucial general point today. Euralex is on the right track.

Drawing near to the end of my address, having given some topical glimpses of problems and activities in our vast field, I can feel the tangible presence of a phrase from Seamus Heaney, Nobel Laureate, alias Famous Seamus. What he tells me is: "Whatever You Say Say Nothing".

In conclusion, however, let me quote the first permanent secretary of the Swedish Academy, Nils von Rosenstein. At the meeting of the Academy on the 12th of February, 1801, he commented on one of the tasks given to the Academy by its founder King Gustavus III in 1786, namely the creation of a dictionary of Swedish. In addition to knowledge, he said, this heavy work seems to require youthful merriness and health. Ladies and gentlemen, I wish you all the best in both respects.

## References

- Allén, S. (1972). Ordströmmen som dataflöde. En aspekt på språklig databehandling [The Stream of Words as a Flow of Data. An Aspect of Natural-Language Processing], *Humanistisk forskning*, 1972. 2: 19–22.
- Allén, S. (ed.) (1995). *Of Thoughts and Words. Proceedings of Nobel Symposium 92: The Relation Between Language and Mind*. Imperial College Press, Singapore, River Edge, NJ, and London.
- Allén, S. et al. (1970–80). *Nusvensk frekvensordbok baserad på tidningstext* [*Frequency Dictionary of Present-Day Swedish*]



- Based on Newspaper Material*] 1–4. Almqvist & Wiksell, Stockholm.
- Allén, S. et al. (1986). *Svensk ordbok* [A Dictionary of Swedish]. Esselte Studium, Stockholm.
- Allén, S. et al. (1995–96). *Nationalencyklopedins ordbok* [The Swedish Dictionary of the National Encyclopaedia] 1–3. Bra Böcker, Höganäs.
- Allén, S. et al. (1996). Om svar anhålles. Rapport från projektet OSA [Please Reply. Report from the Project OSA [RSVP]]. Research Reports from the Department of Swedish, Göteborg University, GU-ISS-96-4.
- Atkins, S., Clear, J., and Ostler, N. (1992). Corpus Design Criteria, *Literary & Linguistic Computing*, 7. 1: 1–16.
- Cederholm, Y. (1996). A Historical Lexical Database of Swedish: The O.S.A. Project. In M. Gellerstam et al. (eds.), *Eura-lex '96. Proceedings I*. Göteborg University, Department of Swedish, 65–72.
- Halliday, M.A.K. (1991). Corpus Studies and Probabilistic Grammar. In K. Aijmer and B. Altenberg (eds.), *English Corpus Linguistics. Studies in Honour of Jan Svartvik*. Longman, London and New York, 30–43.
- Kjellmer, G. (1994). *A Dictionary of English Collocations. Based on the Brown Corpus* 1–4. Clarendon Press, Oxford.
- Quirk, R. (1992). On Corpus Principles and Design. In J. Svartvik (ed.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82*. Mouton de Gruyter, Berlin and New York, 457–469.
- Svenska Akademien (ed.) (1986). *Ordlista över svenska språket* [A Glossary of Swedish]. 11th ed. Norstedts, Stockholm, in distribution.
- Svenska Akademien (ed.) (1898–). *Ordbok över svenska språket* [A Dictionary of Swedish] 1–. Gleerups, Lund, in distribution.

## Tidender och tidningar

Inledningsanförande vid konferensen Den moderna dagspressen 350 år. 1996. Publicerat i konferensrapport med samma titel.

Nyheter är färskvara, och så har det alltid varit. Låt oss gå tillbaka ett tusental år i tiden till den poetiska Eddan. Den kanske elegantaste dikten där är Kvädet om jätten Trym eller Hammarhämtningen. Tors vrede var våldsamt och uppståndelsen i Asgård stor, när det upptäcktes att hammaren var borta. Loke fick låna Frejas fjäderhamn eller flygdräkt, och med den på sig lämnade han Asgård och flög in i Jotunheimen, jättarnas land. Där fick han veta, att Trym hade gömt hammaren åtta mil ner i jorden och att ingen skulle få tag i den som inte förde honom Freja till brud. När Loke återvände till Asgård, stod Tor och väntade. Han lät inte stackars Loke ens få ta mark, innan han ville ha besked: "Säg i luften / långväga tidender!" Det var bråttom. Episoden är på sätt och vis också ett uttryck för den tidens dröm om den globala byn. Budbäraren från Marathon till Aten på 400-talet f. Kr. sprang, och Stenbocks kurir från Helsingborg till Stockholm på tidigt 1700-tal red. De var båda markbundna, men i våra förfäders myter kunde drömmen om överjordiskt snabba förbindelser gestaltas.

Både *tidende* och *tidning* fanns i fornsvenskan. Betydelsen var ursprungligen 'det som har hänt, händelse' och sedan 'under rättelse om det som har hänt, nyhet'. Båda orden bygger mycket riktigt på *tid*. I ett uttryck som Ordinari Post-Tijdender, som egentligen betyder 'nyheter komna med posten', är det lätt att uppfatta *tidender* som syftande på ett nyhetsblad. På motsvarande sätt är det med Post- och Inrikes Tidningar. Förd fram till vår tid får vi alltså en betydelseutveckling som går från 'händelse' till 'nyhet' till 'nyhetsblad' till 'företag som ger ut nyhetsblad'. Ordet *tidning* tog efterhand befälet på bekostnad av *tidende*. Det blev också svenskans standarduttryck för 'nyhetsblad', men ännu på 1800-talet hade det konkurrens av *avisa* och *gasett*.

Som ansvarig utgivare för Post- och Inrikes Tidningar gläder jag mig åt att det är den som är utgångspunkten för det här symposiet. Den har ju onekligen de grå tidningarnas charm. Det var ur nyhetsbehovet under 30-åriga krigets tid som tidningen uppkom, och den började ges ut regelbundet 1645 under namnet Ordinari Post-Tijdender. Bara några år tidigare, 1636, hade postverket inrättats på initiativ av Axel Oxenstierna. Tidningen bytte namn flera gånger under århundradena och fick 1821 sitt nuvarande namn, Post- och Inrikes Tidningar. Den innehöll tidigare nyheter och bidrag från olika samhällsområden. Där fanns också en litterär avdelning, där inte minst Carl David af Wirsén på sin tid flitigt medverkade. När Anders Österling blev redaktör för tidningen 1936, fick han i kontraktet med justitiedepartementet förbinda sig att använda ett vårdat språk. Det torde han också ha gjort i den enda artikel han bidrog med. Nu för tiden är tidningen nästan enbart ett organ för legala annonser – en gång om året publiceras en redogörelse för Svenska Akademiens högtidssammankomst den 20 december.

Just den 20 december, året var 1791, kungjorde Gustav III, att Svenska Akademien för evärdeliga tider skulle förunnas rättigheten att ensam förlägga och utge tidningen och därigenom få en viss inkomst. Det är tack vare denna utomordentliga idé som Akademien har kunnat realisera stadgarnas åläggande att utarbeta en svensk ordbok. Svenska Akademiens ordbok är en beskrivning av ordförrådet och dess utveckling från Gustav Vasas tid till skrivande stund. Den första volymen kom ut på 1890-talet, och hittills är 31 band utgivna. Den alfabetiska genomgången av ordförrådet har därmed nått till ordet *stå*. Vi får räkna med ytterligare två decennier, innan hela verket är avslutat i den här formen. Jag vill gärna tillägga, att ordboken genom samarbete mellan Akademien och Språkdata här vid universitetet nu också finns som databas och kan nås via Internet i den form som kallas konkordans.

Ett aktuellt ord när det gäller tidningsverksamhet är *journalism*. Det är emellertid inte nytt i språket. Belägg finns redan på 1830-talet, då betydelsen är 'tidningsväsende'. Hundra år senare får det en ny användning, 'verksamhet såsom tidningsman', och

används i den meningen av Sigfrid Siwertz bland andra. Det betecknade då också det som vi nu vanligen kallar *journalistik*.

I den analys man finner hos Jan Ekecrantz och Tom Olsson innebär *journalism* en strävan att få mottagarna att tillägna sig journalisternas världsbild. Härigenom uppstår vad de med ett talande uttryck kallar "det redigerade samhället". Dagens journalist är enligt deras mening mindre referent än medaktör med beskrivningsmakt. Journalisten är också mer fokuserad på nuet, mindre historieintresserad än tidigare.

Mot bakgrund av liknande tankar, framför allt Olof Petersons beskrivning av journalismen som en ideologi vars mål är att få kontroll över informationsströmmarna, betecknar Ami Lönnroth i sin installationsföreläsning från i våras denna journalism som typiskt manlig. Mot den ställer hon feminismen, kvinnliga journalisters strävan att vidga journalistikens intresseområden och perspektiv. Också härigenom kan man skaffa sig makt, påpekar hon. I varje fall är det i den riktningen hon ser det paradigmskifte som hon finner nödvändigt.

Ordet *makt* återkommer gång på gång i analyserna. Man kan höra en chefredaktör säga, inte utan tillfredsställelse, att det var hans tidning som avsatte en minister. I sådana fall må ens demokratiska sinnelag tillåtas reagera. Problemet är, att tidningar är affärsdrivande företag, vilkas ledare har valts på andra grunder än demokratiska. I vidrigt fall kan detta leda till det som på engelska kallas *infotainment* 'nyhetsunderhållning'. Men lika lite som vi skulle vilja lägga ner mänskligheten för att rädda jorden, vill vi lägga ner pressen för att rädda demokratin. När man talar med seriösa journalister om denna ödesfråga, hänvisar de till de tecken på självsanering som de tycker sig urskilja. Jag hoppas också på en sådan.

Ordet *journalism* finns även på engelska. Det syftar där i första hand på journalistyrket. Det kan därutöver avse det skrivande som karakteriseras av brådska och ytlighet. Stundom är man frestad att citera G.K. Chesterton: "Journalism consists largely in saying 'Lord Jones Dead' to people who never knew that Lord Jones was alive." Engelskan håller sig också med ordet *news-*

*paperism*, som är beteckningen för skandalsidan av journalistiken.

Ibland har jag lekt med tanken på en alternativ tidning, en publikation som kunde fungera som komplement och korrektiv. Den skulle i så fall arbeta från en plattform som möjliggjorde, att man satte information och analys främst, utan sidoblickar, utan vinklingar, utan ägarinflytande, utan vinstintresse, utan maktanspråk och utan signaturer och porträtt. I medvetande om att ingen är fullkomlig inser jag, att en sådan alldeles oberoende gasett vore en nåd att stilla bedja om. Kunde emellertid staten i parlamentarisk enighet avsätta ett engångsbelopp på ett par miljarder kronor, skulle avkastningen kunna användas i den här riktningen. Förr eller senare måste problemet lösas på ett eller annat sätt.

En fundamental faktor i vår världsbild är ordets frihet. I mina mycket unga år tog jag djupa intryck av Torgny Segerstedts *Handelstidning*. Jag minns ännu de svarta sjok som markerade censurens ingrepp. När det gäller ordets frihet befinner sig läsarna alltid i underläge. De måste ha möjlighet att komma till tals i tidningen, som inte får bli ett slags informationshinder. I en uppsats nyligen har Torsten Steinby presenterat det program som Hufvudstadsbladet antog vid sin omorganisation för ett antal år sedan. Där sägs att man fattar ordets frihet inte endast som en garanti mot ingripanden i yttrandefriheten från statens sida. Som en central synpunkt betonar man också läsarens rätt att bli mångsidigt och korrekt informerad, att sakligt uttala sina egna uppfattningar och att försvara sig mot angrepp. ”Den ärliga informationen, öppna debatten och rätten till fri, saklig kritik bildar en väsentlig del av ordets frihet.” Under allra senaste tid har det öppnats en ny möjlighet att komma till tals, när ett medium inte släpper fram tillrättalägganden från angripna. Man kan helt enkelt lägga ut sin text på Internet, där var och en som så önskar kan läsa den. Vad detta kommer att betyda framöver är sannerligen värt att tänka på.

Också från ett helt annat håll kan man få en bild av det stora ansvar som vilar på tidningarna. Vid ett tillfälle gjorde vi vid vår institution en undersökning av hur mycket text på svenska som trycks under ett år. Vi fann att det rör sig om mellan tre och fyra

miljarder löpande ord. Det anmärkningsvärda är, att dagstidningarna svarar för tre fjärdedelar av den textmassan. Man frågar sig, vad det betyder för språket, för kulturen i vidare mening, för samhället i dess helhet att mycket få personer skriver i tidningarna, medan väldigt många läser vad de skriver. Språkliga fenomen kan få kraftiga genomslag. Förändringar kan inträffa snabbare än förr. Opinioner kan kantra. Man har stora möjligheter att agera på de mest skilda sätt.

Det finns förvisso goda texter i dagens tidningar, och min förhoppning är att de har en förebildlig verkan. Men språket har ju både ett uttryck och ett innehåll, och innehåll av typen tidningsankor avstår vi gärna från. Det finns också ett par uttryck som beskriver den faktiskt förekommande vulgariteten. Jag tänker på "drevet går" och "i allmänhetens intresse". Jag tänker också på den kravfria enkelspårigheten i frågor till enskilda personer av typen: "Har *du* hört talas om Oe? – Har *du* hört talas om Oe? – Har *du* hört talas om Oe? ..." Jag kan väl också nämna den pseudointimitet som användningen av förnamn nu tycks locka till: "Har du frågat Mona?"

Naturligtvis finns det också de oavsiktligt komiska formuleringar som kommer till i misshugg. Spalten Pressgrannar i Grönköpings Veckoblad ger ofta exempel på den saken. Vid ett tillfälle citerades en tidningsartikel, som handlade om hur Galileo Galilei satt i husarrest i en liten by utanför Florens under de sista åtta åren av sitt liv. "Här hade han utsikt över sitt älskade Florens som han aldrig fick återse."

Ingenjörsvetenskapsakademiens avdelning XII, vars inriktning är informationsteknik, tar i dagarna upp temat "Dagspressen inför informationsåldern". Bakgrunden är att tidningarna får räkna med att distributionen på papper i framtiden kommer att vara ett av flera olika sätt att nå kontakt med läsarna. Några tidningar har ju redan börjat anpassa sig till den nya situationen. Fysiskt är vi här rätt långt ifrån Frejas fjäderhamn, men innehållsligt ack så nära.

Att pressen spelar en viktig roll i samhällslivet behöver ju inte understrykas igen. Det betyder å andra sidan, att den presshistoriska forskningen liksom pressforskningen över huvud taget får

stor vikt. Det symposium som förestår har i det perspektivet en mycket intressant uppläggning: presshistoriska utblickar, utvecklingsmönster, tryckfrihet och etik, lokala tidningsmarknader, tidningsmakeriets villkor. Jag önskar lycka till med detta och skall nu till sist utföra, vad vi språkvetare i samförstånd med filosoferna kallar en performativ akt. Det är en språkhandling med innebörden, att när orden är sagda, är också själva saken utförd. Alltså: härmed förklarar jag symposiet "Den moderna dagspressen 350 år" invigt.

## Klassiker och klassiker

Språket lever! Festskrift till Margareta Westman. 1996.

Vad ord betyder, hur de används, vilka valörer de har, hur väl-etablerade de är i språket och åtskilligt annat lockar till ständigt nya fjät i riktning mot den fulla insikt som aldrig kan uppnås men alltid hägrar. Detta är i och för sig skäl nog att ta ett ord som *klassiker* i betraktande. Det har emellertid också slagit mig, att ordet har börjat utnyttjas alltmer under senare år och att det har hänt något med dess innebörd. Dessutom har de första böckerna i serien ”Svenska klassiker utgivna av Svenska Akademien” kommit ut.

En rimlig fråga är alltså: Vad är en svensk klassiker? Jag har botaniserat i Språkbankens olika textmaterial, där det finns mängder av exempel på ordets användning och där mångahanda svar kan utletas. I vissa fall har jag naturligtvis också utnyttjat andra källor.

Ett av svaren – om än inte det viktigaste – är för ovanlighetens skull klart och koncist. En svensk klassiker omfattar långlopp på skidor (Vasaloppet eller Engelbrektsloppet), långlopp på cykel (Vätternrundan), långdistanslöpning (Lidingöloppet) och långdistanssimning (Vansbrosimningen) – allt under ett och samma år. Och den som har klarat dessa bedrifter får diplom och blir själv en svensk klassiker. Bland exemplen finns:

en svensk *klassiker*, namnet klingar respektingivande  
 en svensk *klassiker* startades upp 1971  
 att jag bara genomfört en fjärdedel av *klassikern*.

Bakgrunden till benämningen är att de tävlingar som ingår gärna betecknas som klassiska. Man kan kanske tillägga, att Vasaloppet skall åkas i vad som kallas klassisk stil, dvs. med skidorna i spåret på hävdvunnet sätt och inte vinklade på skridskovis. En skidlöpare som är specialist på den klassiska stilen kan rentav också kallas klassiker:

Christer, yngst i landslaget, har räknats som en utpräglad *klassiker*.



Över huvud taget är det inte ovanligt att anrika evenemang får den aktuella beteckningen:

den fotbollscup som sedermera blev en *klassiker*  
 Wimbledonturneringen – tennisens *klassiker*  
 tremålsförlust i returen av *klassikern* mot ärkerivalen Danmark  
 den *klassiker* som ändå ett skid-SM utgör  
 applåder för en *klassiker* – ”De Gamlas Resa” är världens äldsta  
 dagstidningsarrangemang.

På liknande sätt är det med personer som svarat för ovanliga prestationer:

addera med gamle *klassikern* Peter Stastny, som spelat 13(!) år  
 i NHL.

Inte så få fall knyter sig till mat(varor) och matställen:

en svensk *klassiker* – löjromstoast  
 lax är en *klassiker* till valborg och pingst  
*klassikern* stekt skogssvamp på rågbröd har gått upp  
 scampi fritti är en *klassiker* som står och faller med goda råvaror  
 på bekostnad av *klassikern* – den röda standardmjölken  
 Gamle Port och Weise finns med bland de bästa *klassikerna*  
 i kvarteret intill ligger *klassikern* Flygarns Haga.

En diverselåda innefattar saker och ting som tilldrar sig (återkommande) uppmärksamhet snarare än uppvisar speciellt anmärkningsvärda egenskaper:

den äldsta är en *klassiker* av -47 års modell  
 bandyportföljen har ju under alla år varit en *klassiker*  
 den gamla *klassikern* sextimmarsdagen  
 ’Iran-contrasaffären’ jämförs i alla fall med andra *klassiker*.

Skrifter med andra syften än konstnärliga kan få klassikerstatus:

marxistiska och filosofiska *klassiker* köpes  
 filosofiska och socialistiska *klassiker*.

En motsvarighet till engelskans *minor classic* är också belagd:

”Meta-sociologisk essä” blev en mindre *klassiker*.

Det är på de konstnärliga områdena som ordet traditionellt har sin hemortsrätt. Musiken:

en av operans *klassiker*  
 de s.k. *wienklassikerna*  
 en av de *klassiker* lyssnarna får höra är Louis Armstrongs West End  
 Blues  
 vi spelar *klassiker* från 40-talet och framåt, evergreens och bebop.

### Måleri och jämförbara aktiviteter:

man finner *klassiker* som Rubens, Bruegel, Monet och Renoir har lärt grundligt av många stora *målarklassiker* där pågår just nu en utställning under titeln Kameran *klassiker* Bruno Mathsson visar sina *klassiker* på Illums Bolighus "Fiskartröjor och andra *tröjklassiker*".

### Program i etermedierna:

väl förspant i radion med flera fasta programtider förutom *klassikern* "Smoke Rings" den är en svensk *TV-klassiker*, oändligt älskad.

### Film och annan underhållning:

*klassiker* ur pilsnerfilmsrepertoaren  
*klassikern* Snövit  
*science fiction-klassikern* Star Trek  
*kultklassikern* Beyond the Valley of the Dolls  
*snyftklassikern* Dimmornas bro  
Glaset i örat är en svensk *klassiker*.

### Skönlitteratur och dans:

den svenska *klassikern* Nils Holgersson  
Samuel Becketts *klassiker* I väntan på Godot  
Birgit Cullbergs moderna *klassiker* Fröken Julie  
läste metervis av *klassiker* och gick på gym.

### Ursprungligen gällde det författare från antikens litterära blomstringstid:

en humanism där våra *klassiker* från antiken var våra dagliga följeslagare  
revolutionärer som på allvar började spela de grekiska *klassikerna*.

### Senare kunde det också gälla konstnärer som tagit djupa intryck av de gamle:

*klassiker* och romantiker på en gång.

### Den som framstår som specialist på den antika kulturen tillerkänns samma beteckning:

en genombildad *klassiker*.

### Här gäller också att det skall böjas i tid:

som *halvklassiker* betecknas en svensk gymnasieelev som läser latin;  
en *helklassiker* läser dessutom grekiska.

### Tidsaspekten är viktig och gäller även förhoppningsfulla prognosmakare:

föreläsningen kommer att bli en *klassiker*  
 Axel och Bengts Taube-CD borde bli en *klassiker*  
 en blivande '*klassiker*' utlovar fliktexten  
 "Dynamite Dance", tre CD-skivor fyllda med de bästa låtar som har  
 gjorts under 90-talet – "Dynamite Dance" är redan en *klassiker*.

Kulturrådets administrativa stipulation att ett verk måste vara  
 minst tjugo år gammalt för att få

klassikerstöd  
 är en sak för sig.

Det semantiska fältet är sålunda mycket brokigt numera. Om vi  
 försöker samla det i lexikaliska termer, kunde det se ut ungefär så  
 här, i ett slags omvänd ordning:

- (a) företeelse som tilldrar sig återkommande uppmärksamhet  
 en *klassiker* som sextimmarsdagen
- (b) företeelse som har en väletablerad position på sitt område  
 en *klassiker* som löjromstoast
- (c) företeelse av traditionsrik karaktär  
 en *klassiker* som Vasaloppet
- (d) person som har utfört en minnesvärd bedrift  
 en *klassiker* som Peter Stastny
- (e) skidåkare som är specialist på s.k. klassisk stil  
 en *klassiker* som Christer Majbäck
- (f) gymnasieelev som läser (grekiska och) latin  
 en *klassiker* som Pernilla
- (g) kännare av antikens kultur  
 en *klassiker* som Axel Boëthius
- (h) författare från antiken  
 en *klassiker* som Vergilius
- (i) författare eller annan konstnär som tagit intryck av antika  
 eller antikiserande mönster  
 en *klassiker* som Schiller

(j) författare eller annan konstnär som har skapat verk av bestående värde

en *klassiker* som fru Lenngren

(k) verk av bestående konstnärligt värde

en *klassiker* som "Röda rummet".

Betydelsekärnan har vi i de senare momenten (g–k): högklassiga författare eller andra konstnärer och deras verk med tillägg av antika författare och konstnärer jämte kännare av deras kultur. Ännu 1936, då artikeln *klassiker* i Svenska Akademiens ordbok trycktes, var det endast dessa moment som noterades. Härifrån har fältet som vi sett vidgats i olika riktningar från elementet kvalitet mot vad som snarast är traditioner och bedrifter av alla möjliga slag. Ibland är det knappast ens fråga om det, bara någontings mer eller mindre uppmärksammade förekomst. Stundtals har det med förlov sagt karaktär av exploatering av ordet. Vad vi skådar är under alla omständigheter ett av dessa epidemiska språkliga moden.

Det första belägget i svenskan har formen *classicus* och är från 1819, då Lorenzo Hammarsköld i senare delen av "Svenska vitterheten" använde det om en författare:

och Lehnbergs ära såsom *Classicus* var förlorad.

Formen *klassiker* har vi från kyrkan. Det bör tilläggas att Rosenstein använde ordet *klassisk* redan 1787 och då om en författares verk:

en bok, som i Frankrike då ansågs för *klassisk*.

Ytterst går uttrycken tillbaka på latinets (*auctor* eller *scriptor*) *classicus*, där adjektivet avser förnämsta rang. Bakom det ligger i sin tur substantivet *classis*, som anger den främsta kategorin av samhällsmedlemmar.

I grunden är det således egenskapen kvalitet det handlar om. Karakteristisk och berömvärd är en nutida distinktion som denna:

Boken blir möjligen en bästsäljare, aldrig en *klassiker*.

## Har Europa närmat sig varandra?

Magasin Lyran 1996: 3.

*Ömsesidig* är ett vackert ord för en vacker tanke – i varje fall när man gör sitt bästa på ömse håll. *Reciprok* är kanske inte lika vackert, men tanken är den samma. Vi har ordet från språkläran, där *varandra* och det litet vardagligare *varann* kallas *reciproka pronomen*.

Ömsesidigheten fanns med från början. I fornsvenskan hette det *hvar ... annan* 'var och en ... den andre'. Hos Birgitta kan vi läsa "änglane skodha hwar annars fäggrind" – i modernare form "änglarna skådar varandras fäggring" – som egentligen betyder 'änglarna skådar var och en den andres fäggring'. Det är lätt att se, att samma grundtanke finns i tyskans *einander*, engelskans *one another* eller *each other* och franskans *l'un ... l'autre*.

För att något skall kunna bli ömsesidigt krävs naturligtvis minst två parter. Fullgoda exempel är "till sommaren skall vi äntligen få vila ut och riktigt rå om varandra", "deras vägar korsade varandra flera gånger", "grannarna vattnar varandras blommor" och "lyckliga svävar de ut på dansgolvet och släpper inte varann, förrän det är dags för orkestern att ta paus".

Ett av problemen uppkommer när man tar för sig ord som till formen är singulara men genom sitt kollektiva innehåll leder tankarna till flera. Det ger betänkliga fall som "laget ställer upp för varandra", "då lärde bybefolkningen varandra att skriva" och "det såta paret har återfunnit varandra". Genom kollisionen mellan form och innehåll alstras en kraftig störning, som man med fördel kan undvika genom att till exempel säga "spelarna ställer upp för varandra", "då lärde byborna varandra att skriva" och "de båda kumpanerna har återfunnit varandra".

Ord sådana som *människa* kan användas på liknande vis: "sätt människan i centrum" betyder ungefär det samma som "sätt människorna i centrum". Men det blir mycket skralt när *människan* kombineras med *varandra* som här: "människans beskrivningar av varandras utseende är rasismens rötter". Och här

saknar *varandra* helt förankring i texten: ”respekten för varandras åsikter är större nu”.

Några av våra pronomen vållar också bekymmer i förhållande till *varandra*. Till dem hör *man*, som ju numera rätt ofta ses och hörs med plurala adjektiv: ”man var ganska nöjda”. Det är ingen föredömlig konstruktion, och det är inte heller samma andas ”man tog varandras händer”. Det är rimligt att ta hänsyn till alla dem som reagerar negativt inför sådana uttryck. Motsvarande krumelur med *en och annan* är ännu mindre föredömlig: ”en och annan kände inte igen varandra”. Alldeles urbota är ”ingen brydde sig om varandra”.

Namn har entydig syftning. *Europa* är namnet på en världsdel. Lika litet som man kan tala om att ”världsdelens har närmat sig varandra” kan man utan att väcka löje säga att ”Europa har närmat sig varandra”. Byt ut ”Europa” mot ”Europas länder” och risken är ur världen!

Som så ofta i språkliga sammanhang kan vi alltså iaktta en skala från goda och av alla accepterade användningar av orden till de användningar som ingen accepterar. En sak är säker: frågan ”Har Europa närmat sig varandra?” är fel ställd – i varje fall språkligt sett.

## The Relation between Language and Mind. Opening Address at a Nobel Symposium 1994

Of Thoughts and Words. 1995.

Romeo.  
Peace, peace, Mercutio, peace!  
Thou talk'st of nothing.

Mercutio.  
True. I talk of dreams;  
Which are the children of an idle brain,  
Begot of nothing but vain fantasy.

*Shakespeare*

Dear Colleagues and Friends,

A participant in a holiday course for learners of Swedish was out one evening and came back fairly late. He then found he had forgotten his key and couldn't enter his room. He went to the teacher-in-charge and said, in his rudimentary Swedish: "Jag ute, nyckel inne" ("I out, key in"). This fine effort of a beginner's mind was immediately understood and there was a happy end. I intend to come back to this utterance in a few minutes.

This Nobel Symposium is about a relation, that is what one thing, Language, has to do with another, Mind. What kind of something are they? In what way does the one relate to the other? What kind of connection or correspondence prevails between them? And, excitingly enough, what does one tell us or at least indicate about the other? How is it possible to suppress less relevant information and survey what is more relevant? Furthermore, to what answers can electronic computing contribute?

It is a great privilege and pleasure for me to greet this distinguished group of experts in many fields on this occasion. I do this on behalf of the Swedish Academy, the responsible institu-

tion, and with thanks to the Nobel Foundation, our sponsor, as well as its Nobel Symposium Committee.

Nobel Symposia are arranged in accordance with the intentions of Alfred Nobel, the founder of the Nobel Prizes. In recent years, a series of symposia has been organized in the fields indicated by the Prizes. This is one of a number of symposia supported by the Swedish Academy, whose domain is language and literature. Our Academy was founded in 1786 by King Gustavus III in order to promote the Swedish language and Swedish literature. The Academy has eighteen members. Some hundred years ago she accepted the wider task of selecting the winners of the Nobel Prize for Literature, entrusted to the Academy by Alfred Nobel in his will.

The concept of mind, as well as the concept of language, was characterized as elusive in the invitation to this Symposium. It would be a considerable achievement of our interdisciplinary gathering if we could contribute to some reduction of this elusiveness, which is what our preprints promise. Among the many facets touched upon there are the questions of creativity, inspiration, linguistic ability, psychic space, soul, consciousness, subconscious processing, system-and-process, modularity, spirit, self-confidence, understanding, intelligence, skills, memory, thinking, dreaming, lying, mind-reading, cognition, inner environment, neural Darwinism, interaction between neurons, practiced and novel conditions, timing and free will, information and exformation, semantic transparency, ambiguity, deformation, gestalts, benign instability, etc. In any case, a remaining paramount riddle is the so-called binding problem: how does a unified perception come about? A lexical approximation of *mind* as 'seat of consciousness, thought, volition, and feeling' is too simplistic and carries the dangerous association of *seat* meaning 'site or location'.

The I-out-key-in-episode is another illustration of what I consider the central component in language ability: the lexicon, words. In fact, not only separate words but all sorts of more or less lexicalized strings of words, collocations, or blocks as I call them. In their capacity of lexical units, blocks are of course part



of the paradigmatic system, but since they contain sequences of words they have a syntagmatic quality at the same time. One out of two or three breathless experiences in my life as a linguist was when I saw the immense impact of these connections: set and half-set phrases, continuous and discontinuous strings, complete and open-ended strings, nominal and verbal and prepositional phrases, full-fledged clauses, and entire sentences – and all this in hundreds of thousands of items. I marvelled at the capacity of the cerebrum and the versatility of mind. The following is a microscopic sample from English, centered around *time*: to stand the test of time, to gain time, the good old times, time immemorial, in time, time and again, there is a time for everything, she is near her time, now is the time to press your point, it is time I was going. No wonder I declared myself a collocationist.

This happened to me about thirty years ago in the early years of a comprehensive project in corpus linguistics, featuring the ineluctable conjunction between authentic text, linguistic theory, method, and technics. That is also where my real interest in the theme of this meeting was born.

In this vein, the inquiry about the key could perhaps be worded in the following way: "Excuse me for disturbing you. I was out this evening and my preparations left a lot to be desired. I must confess I forgot to bring my key. Could you please let me in?" This is of course nothing but a design of blocks. It seems to me that, as a matter of principle, they have a bearing on general linguistics: an explicit theory has to cope with this incredibly varied category as opposed to the discrete parts of language, often heavily emphasized; also notice that time is a critical factor in cerebral language processing making the large number of choices presupposed in abstract models less realistic than the handling of blocks. On psycholinguistics: a psychologically interesting model of perception and generation should incorporate the large amount of blocks; it is plausible to assume that they are treated as elements in their entirety. On neurology: how are they stored and, not least, how are they accessed; it seems there is a function overriding the ordinary storing function. On computational linguistics and artificial intelligence: if you are unable to

take care of them efficiently, even the most sophisticated morphological or syntactic analysis or synthesis is of limited significance, relatively speaking. And of course they have a bearing on literary enterprises, one of whose endeavours is to combat them.

The mass of blocks also plays an important part with respect to such weighty categories in language use as style and fluency. This is not to say that quantities or priorities, although carrying great consequence, are part and parcel of the language system as such. I am convinced they are not. But if you don't master the blocks, you will never approach anything like fluency in a given language.

In addition, the key-episode gives us a neat picture of one of the fundamental aspects of mind: the possibility of having an intention and acting in accordance with it. This is, by the way, where the computer fails most obviously. Actually, intention and attention combined will take us a long way.

The use of language for artistic purposes is another characteristic feature of the human mind that I think ought to be stressed. The extremely interesting question of whether a piece of literary art is wholly or partly the outcome of pure chance is a special problem here.

In this connection, why not hold up wit as an essential aspect of mind? By way of example, let me quote a compatriot of mine who made the following observation (rendered in English):

"There are probably intelligent creatures on other planets. Otherwise, we should have had them here by now."

Or listen to a Dane trying to draw attention more widely to Danish literature:

"Something is written in the state of Denmark."

Let me add that, to my mind, one of the finest qualities encountered in this world is the deep insight inherent in humour, real humour.

In the true spirit of Nobel Symposia we shall try to come to grips with the questions posed from a variety of viewpoints. I am happy to be able to say that we have got a wealth of stimulating thought&debatestuff. The headings of the sessions are of course not sacrosanct. You can think of them as a smorgasboard with at least five dishes, alternatively designated Litterae, Lingua, Psyche, Cerebrum, and Instrumentum Computatorium.

In sum: welcome and let us get started – if you don't mind!

## Om Akademialmanackan

Kulturrådet 1994: 6. Uppsatsen bygger på ett anförande inför Observatoriekullens Vänner 16 november 1994 under rubriken "Idag är det Edmund".

Den namnlängd som kom att gälla från 1901 innebar i själva verket en anpassning av almanackan till reformationen, generöst avrundat 350 år i efterhand. En hel del stoff från det gamla helgonkalendariet ersattes sålunda av namn från vår egen kulturkrets. Vissa ändringar hade gjorts under århundradenas lopp men inte i en omfattning som är tillnärmelsevis jämförlig. När man skred till verket, gjorde man det med gott humör: Primus byttes mot Börje, Konkordia mot Frida och Zephyrinus mot Östen; Jesper anslöts till Sjusovardagen.

Fram till nu har endast tre ändringar gjorts i 1901 års namnlängd: Nore till Sverker (unionsupplösningen), Estrid till Astrid (prinsessan) och Kasper till Sibylla (prinsessan). Sedan Vetenskapsakademien 1972 framtogs almanacksprivilegiet, har emellertid kaos utbrutit. Vem som vill kan lansera en namnlängd.

Den 8 november 1993 sände jag via TT ut ett meddelande om att Svenska Akademien hade beslutat ge ut en almanacka med den traditionella namnlängden. Initiativet hade fått stöd av Vetenskapsakademien, och det fick även Vitterhetsakademiens stöd.

Åtminstone tre slags namnlängder har förekommit på olika håll efter 1972: förutom den etablerade ettnamns­längden (1901) också en tvånamns­längd (1993) och en trenamns­längd (1986). Svenska Akademiens syfte är att vårda, vitalisera och vidareutveckla den ettnamns­längd som har funnits under hela 1900-talet och som i väsentliga delar har långt äldre anor. Om man inför den uppgiften överväger principerna för almanackans namnlängd, är det flera önskemål som anmäler sig.

Namns vanlighet växlar från år till år, och det är inte förutsägbart hur utvecklingen skall bli. Det framgår med all tydlighet av den undersökning Staffan Wåhlin och jag gjorde av samtliga namn som bars av alla svenska medborgare den 1 januari 1973 (se Förnamnsboken, 2 uppl. 1979). Det rörde sig sammanlagt om

inemot 18 miljoner namnförekomster. Typiska exempel är de nu aktuella namnen Amanda och Joel. Amanda var 1890 och 1910 tillräckligt populärt för att ha kunnat pocka på en plats i en frekvensbaserad namnlängd, men så var inte fallet 1930, 1950 och 1970. Joel var inte gouterat 1890 men blev det 1910 för att sedan åter falla tillbaka och förbli lågfrekvent ända tills nu. Att följa namnmodet är alltså att förbinda sig till återkommande revisioner, en obeständighet som är långt ifrån önskvärd.

Namnen i namnlängden är sedan lång tid – hundra år eller ofta flera hundra år – knutna till bestämda dagar. De har därigenom kommit att ingå i en kulturtradition som det finns all anledning att vårda. Omflyttningar motverkar vården av kulturarvet och vållar oro och missnöje. I den tvånamnslängd som har lanserats är ett fyrtiotal namn – Gudmund, Rakel, Rebecka, Viktor osv. – flyttade från sina traditionella dagar. Namnen i namnlängden bör ligga kvar på sina traditionellt givna dagar.

Man skall inte underskatta den roll namnet spelar i en människas liv. Det ligger fortfarande något av magi i att uppkalla barn efter förfäder eller personer som på något sätt har gjort intryck. Namnets ursprungliga betydelse – Katarina 'den rena' osv. – spelar ofta också in. Hos den uppväxande blir namnet en integrerad del av personligheten. Stavningen ingår som ett viktigt moment, och namnsdagen blir på många håll en fast punkt.

Det är viktigt att namnlängden utvecklas så, att den fångar in fler namn. Det bör inte ske genom att språkligt obesläktade namn – t.ex. Bror och Bruno, Lydia och Carola – ställs sida vid sida samma dag. I stället bör varje dags namn betraktas som ett samlingsnamn för alla namn som är språkligt besläktade med det. Ett namn som Anna har varianter som Anita, Anja, Ann(e), Annette, Annie och Annika. Alla dem skall man tänka sig tillsammans i almanackan på Annadagen den 9 december. Egentligen är det alltså fel att beteckna en sådan almanacka som en ettnamnslängd. Att föra språkligt besläktade namn till samma dag är en riktlinje som har den extra fördelen att den aktiverar språkintresset. Därigenom främjas namnvården och alltså språkvården, en uppgift för Svenska Akademien sedan två sekler.

En given dags uppsättning av språkligt besläktade namn kan i själva verket variera från något enstaka namn till flera tiotal. Av praktiska skäl bör de inte införas i själva namnlängden, där endast samlingsnamnet bör stå, utan upptas i en namnlista knuten till almanackan. Namnlistan är helt enkelt en praktisk lösning på utrymmesproblemet.

Sammansatta namn måste kunna anslutas enkelt till modellen. De bör hänföras till de samlingsnamn som ingår i dem. Ofta finns båda leden i namnlängden: Ann(e)-Marie kan firas på Anna och/eller Maria och Britt-Marie på Birgitta och/eller Maria, medan Åse-Marie kan ta sikte på Maria. Principen om språkligt släktskap tar i stor utsträckning också hand om utifrån inkomna eller inkommande namn. För exempelvis Jacqueline och Jaakko som namn på svenska medborgare svarar på ett naturligt sätt Jakob.

Namn är inte sällan besläktade över könsgränsen. Det vållar inga svårigheter. Lika väl som Amandus må firas på Amanda, går Ylva till Ulf och som sagt Jacqueline till Jakob. Det är naturligtvis också önskvärt att antalet namn i namnlängden och namnlistan reflekterar könsfördelningen. I själva verket slår detta förvånansvärt väl ut, när de språkligt besläktade namnen systematiskt tas tillvara. Det blir en svag övervikt för kvinnonamnen, det är allt. Förklaringen är att en hel del kvinnonamn – sådana som Elisabet, Maria och Margareta – är starkt produktiva i namnbildningen, varigenom det större antalet mansnamn i själva namnlängden gott och väl uppvägs.

Svenska folket har tiotusentals olika förnamn. De är ungefär 80 000 om man räknar varje stavning för sig. Hur almanackans namnbestånd än utformas, blir därför de allra flesta namnen över. För Akademi-almanackans del blir det de som inte är språkligt besläktade med något samlingsnamn. För dem kan emellertid en möjlighet öppnas till en gemensam namnsdag, kallad allmän namnsdag. Den förnamnslösa nyårsdagen ligger nära till hands. Enligt Alf Henriksons Alla årets dagar har den dagen inget förnamn i någon almanacka någonstans. Det gamla kyrkliga temat Jesu omskärelse har inte gjort dagen till någon särskild Jesusdag

i det allmänna medvetandet. Lea, Åsa, Stig och jag kan alldeles lugnt börja året med att fira namnsdag.

När man betraktar den oerhörda mängden förnamn, inser man, att almanackan aldrig kan representera annat än en liten bråkdel av namnskicket. Dess huvudfunktion är som har framgått en annan. Namnskicket beskrivs i namnböcker av olika slag, och de utgör med almanackan som tillägg de hjälpmedel föräldrar kan begagna, när de söker namn till sina barn.

Akademialmanackan bygger på de tankar som har förts fram här, och den uppfyller de principiella krav som har ställts. Den innehåller alltså den välkända etnamnslängden utan flyttningar av namn, kompletterad med en namnlista som upp tar ett stort antal språkligt besläktade namn. Eftersom namnlängden inte bygger på (alltid tidsbundna) frekvenser, är den inte gjord med tanke på att revideras. Tillsammans med namnlistan redovisar den en stor mängd namn, ungefär lika många av vardera könet. Den tar hand om sammansatta namn och namn som kommer utifrån. Genom markering av nyårsdagen som allmän namnsdag ger den alla som så önskar möjlighet att fira namnsdag.

Namnlistan i Akademialmanackan för 1995 innefattar drygt tusen namn. Principen om språklig släktskap gör den emellertid till en öppen lista, som kan kompletteras av var och en som finner sitt namn höra dit. Låt oss se på den faktiskt förekommande namnvariationen i en av de mest omtalade veckorna i almanackan, fruntimmersveckan den 18–24 juli.

- |    |           |   |
|----|-----------|---|
| 18 | Fredrik   | Fred, Freddy, Friedrich   |
| 19 | Sara      | Zarah   |
| 20 | Margareta | Margit, Margita, Margot, Marguerite, Marga, Maggie, Maisie, Marit, Maret, Mait, Meit, Meta, Greta, Gret, Gretel, Gretchen |
| 21 | Johanna   | Jane, Jen, Janna, Janet, Jeanna, Jeanne, Jeanette, Jonna, Jonnetta, Jonni, Hanny, Hanni, Hannele, Yvonne                  |
| 22 | Magdalena | Madeleine, Magda, Malena, Marlene, Malla  |
| 23 | Emma      | Emilia, Emerentia   |

- 24 Kristina Kristin, Kristine, Kristinia, Kersti,  
Kerstin, Kirsten, Kirsti, Kickan, Kicki,  
Chris, Christa, Christel, Christiane

Det rör sig i det här urvalet om ett sextiotal namn. Under de sju dagarna kan alltså personer som bär något eller några av dessa namn fira namnsdag. Ändå har jag inte nämnt Greta-Kajsa, Greta-Lisa, Anna-Greta, Ulla-Greta osv. i en lång rad. För Yvones del, må det tilläggas, är den förmedlande länken Ivan, den ryska formen av Johannes.

Den tvånamnslängd som har förts fram som ett alternativ till den traditionella almanackan har flera otillfredsställande drag. Det beror inte på oskicklighet utan på de utgångspunkter som har valts. Till olägenheterna hör att den till stor del är grundad på namnfrekvenser; att den i anslutning härtill ges ut med den uttalade avsikten att fortlöpande revideras; att ett fyrtiotal namn har flyttats från sina traditionella dagar, vilket går ut över kulturarvet; och att språkligt obesläktade namn har ställts sida vid sida. Grundläggande svagheter som dessa har starkt bidragit till att göra Akademialmanackan nödvändig.

Svenska Akademiens lansering av Akademialmanackans idé har fått ett mycket uppmuntrande mottagande. Entusiastiska personer har hört av sig. Till akademiernas stöd kommer en rad företags intresse för saken. Filofax, som är känt för sina planeringskalendrar, har gått över till ettnamnslängden, likaså Bra Böcker i Årets Dagar. Filofax ger i samarbete med Akademien ut Akademialmanackan som fickkalender. Också Verbum har gett ut en fickkalender för 1995 med ettnamnslängden. Dock har dess redaktör Carl Henrik Martling ansett sig böra göra ett tjugotal avvikelse, huvudsakligen genom återinsättande av ett antal helgon, martyrer och biskopar. Göteborgs-Posten har i sin dagliga namnspalt gått över till ettnamnslängden, och Aftonbladet har hälsat den med kraftigt bifall. Det finns all anledning att gå vidare.



## Runmärkt

Anförande vid invigningen av runutställningen på Stockholms medeltidsmuseum 1994.

När jag gick i gymnasiet – som det hette då och som jag tycker att det fortfarande borde heta – undervisade vår svensklärare, minns jag, om språkets uppkomst. Det fanns två teorier: hejhejteorin och ojojteorin. Den ena tog fasta på att man behövde säga åhej eller så, om man skulle göra något gemensamt. Den andra innebar förstås, att man sade ifrån när det gjorde ont. Vi lyssnade och drog i varje fall slutsatsen, att talet var äldre än skriften. Senare har jag konstaterat, att det finns forskare som hävdar att skriften är äldst. Jag tillhör inte dem, men jag tillmäter skriften en betydande självständighet i förhållande till talet.

Vad är då skrift? Om någon skickar en kolbit till någon annan för att överbringa budskapet ”Mitt hjärta brinner av kärlek”, är det ett exempel på skrift? Nej, det är vad vi kallar föremålskommunikation. Skrift är ett konventionellt system för kommunikation genom på föremål gjorda tecken. Men om avsändaren då ristade ett hjärta på kolbiten? Nej, det räcker inte. Det skulle alternativt kunna utläsas ”Jag kan inte leva utan dig” etc. Skrift i egentlig mening måste också uppfylla kravet på en i princip entydig konvertering till tal.

Med denna innebörd har skriften en dokumenterad historia på ungefär 6 000 år. I runt tal 3 000 år f. Kr. fanns det dels en sumerisk, dels en egyptisk ordskrift. För båda gällde alltså regeln ett ord–ett tecken. Den egyptiska ordskriften eller logografin var vad vi brukar kalla hieroglyferna. Frammot 1500 f. Kr. hade det ur dem utvecklats semitisk stavelseskrift, syllabografi, enligt regeln en stavelse–ett tecken. Det var sedan grekerna som omkring 900 f. Kr. tog det stora språnget. De skapade alfabetisk skrift genom att med utgångspunkt i den feniciska varianten av semitisk skrift systematiskt införa tecken för vokalerna.

När dyker då runorna upp? I stort sett vid tiden för Kristi födelse. De återgår på det grekiska alfabetet och i viss mån på

därur utvecklade system som den latinska skriften. Med andra ord: den kortaste sammanfattningen av runornas förhistoria är att de ytterst härstammar från hieroglyferna.

För skriften har man vid olika tider och på olika håll använt olika material: lava, sten, trä, vax och horn liksom förstas papyrus, pergament och papper. Låt oss slå fast att runorna är en alfabetisk skrift för användning på trä. Vi kan lägga märke till att ordet *bokstav* just betyder 'runa ristad på ett stycke bokträ'. Från början var det 24 bokstäver i runalfabetet eller futharken, men senare reducerades de till 16.

Runorna uppfattades i ett tidigt skede som gudomligt givna. De fick epitetet *reginkunnr* 'som härstammar från gudarna'. Oden, vismannen, fick äran att vara upphovsman. Om det kan man läsa i Havamal. "Råde den som runvis är", står det med en lätt magisk formulering på Ågerstastenen. Vi kan också påminna oss, att ordformen *runa* är sen, belagd först på 1400-talet. Den är en utvidgning av ett äldre *run*, som ytterst betyder 'hemlighet'.

Bruket av trollformler och lönnrunor hör hit. Om man bar en brakteat med runor på, kunde man vara skyddad mot olyckor. Om man ville ha framgång i strid, skulle man rista runor på svärdet, segerrunor som de kallas i Eddan. Runor kunde ge framgång i kärlek eller användas i förbannelser över fienden eller den som skändade ett minnesmärke. Man försökte också binda de döda vid graven med runor, till värn för de levande.

För språkhistorien är runmaterialet självklart ovärderligt. Det är främst genom runinskrifter som vi känner *urnordiskan* från århundradena efter Kristi födelse fram till c. 800. Detta ger intryck av att vara det ålderdomligaste av alla germanska språk. Man kan få rysningar av språkhistoriskt välbehag. Det närmast jämförliga är Wulfilas gotiska bibelöversättning, som jag håller mycket högt.

Nästa period, 800–1225, har helt enkelt fått sitt namn av runorna: *runsvenskan*. Det är runornas glansperiod. Sedan tar den nykomna latinska skriften successivt över. Efter mitten av 1300-talet, digerdödens tid, blir runskriften perifer – men den lever vidare trots allt genom århundradena, i olika skepnader. Exempelvis gav Johannes Bureus 1611 ut en liten skrift, Runa-abc–

boken, där han ger alfabetet i både latinsk skrift och runskrift och där man bland annat kan läsa fadervår, trosbekännelsen och budorden i båda skriftformerna.

Runologin är en viktig och resultatrik gren av språkforskningen. Vi får acceptera att den ibland balanserar på avgrundens rand, som när den har att diskutera möjligheten av stavfel hos den mödosamt huggande runmästaren eller när sprickbildningar på en häll kan tolkas som en uppgift om att en viss person inte äger sjön, varför man nu skall kungöra det. I ett längre perspektiv kan sådant vara uppfriskande.

Minnesmärkena var ofta ståtliga, bildprydda, målade i rött och andra färger. Historikerna har stor glädje av deras texter som källor. Många ristningar berättar om vikingafärder och andra ting. Man kan få uppgift om den inhemska organisationen, handelsförhållandena, religionen. Flera ristningar slutar med de tänkvärda orden "Gud hjälpe hans själ bättre än han förtjänade". Också latinsk text finns skriven med runor: "Ave Maria gratia plena" osv.

Ägaruppgifter är rätt vanliga. På en måttsticka från Lödöse står det helt sonika: "Helga äger mig". Detta är något helt annat än när "Iarlabanki" fann sig föranlåten att hugga in på en sten, att han ensam ägde hela Täby.

Det finns exempel på ristningar som är genealogiska, helt enkelt räknar upp ett antal generationer. Och kalendariska användningar är väl kända. Runstavarna var ett slags evighetskalendrar, och de användes långt fram i tiden.

Inte sällan får vi veta vem som har svarat för en viss ristning. "Skamhals högg dessa runor", står det t.ex. på Västerljungstenen. Vi känner också Öpir och andra.

Det medeltida vardagslivet, inklusive kroglivet, får vi en tidigare aldrig ens anad inblick i genom fynd av träristningar i Bergen, Lödöse och på andra håll. Det är speciellt detta som utställningen lyfter fram. Särskilt har jag sedan länge fäst mig vid det meddelande på en träbit från Bergen som kort och gott lyder: "Gyda säger att du skall gå hem". Det finns de ristningar som är värre. De finns bakom galler på utställningen. Låt mig ändå tillägga, att Gydas adressat försökte rista ett svar på samma träbit,

men han var tydligen inte i det tillståndet att han kunde åstadkomma något läsligt.

Inte minst Sven B.F. Jansson har tryckt på runtexternas roll som litterära vittnesbörd med Rökstenen som paradnummer. Här finner vi en genomarbetad rytmisk konstprosa, ofta allittererad och mättad med allusioner till ett nu oåtkomligt litterärt stoff som tydligen var välbekant på 800-talet. Här finns också den starka strofen på fornyrdislag om "Tjodrik den djärve" med alla de associationer den ger. Vi har alltså på Rökstenen en text som är flera hundra år äldre än de äldsta bevarade uppteckningarna av eddadiktningen.

Men också i det senare medeltida materialet från Bryggen i Bergen finns reflexer av den gamla poesin:

Jag ristar botrunor,  
jag ristar skyddsrunor,  
en gång mot alferna,  
två gånger mot trollen,  
tre gånger mot tursarna.

Detta är en del av en poetisk-magisk ristning på en träkavle, en ristning som skulle försäkra diktaren om en kvinnas kärlek.

En sladderristning från samma ställe är skriven på fornyrdislag:

De är båda  
i boden samman,  
Klaufa-Kari  
och Viljalms kvinna.

Sist och slutligen, som finlandssvenskarna säger, måste vi understryka, att runristningar är samtida texter, original, som alltså kan upplysa oss om samhällslivet, om kulturella och sociala förhållanden och mycket annat. Det ger en viktig aspekt på den nu aktuella utställningen.

Jag lyckönskar till initiativet och tillönskar framgång. Runor och runristningar har en stor lockelse. Det finns ett sug från vikingatid och medeltid. Vi är nog alla i en eller annan mening runmärkta.

## Ifall ett språk du äger. Föredrag vid Svenska litteratursällskapets i Finland årshögtid den 5 februari 1992

Historiska och litteraturhistoriska studier 68. 1993.

Väster om Östersjön har vi en akademi med uppgiften att främja svenska språket och litteraturen. För en tid sedan fick jag ett förslag till hur den saken skulle kunna ordnas: "Detta är Svenska Akademiens ständiga telefonsvarare. Börja dikta nu!"

Jag vidarebefordrar uppslaget för eventuellt beaktande inom Litteratursällskapet, som ju har i hög grad befryndade uppgifter. Det gäller både att vårda klassikerna och att stödja och uppmuntra dem som har satt segel.

Klassiker inom konst och vetenskap är de som har skapat verk av bestående värde. Vem kunde passa bättre i kostymen än Johan Ludvig Runeberg? Så firar också Litteratursällskapet ännu i sitt andra sekel årshögtiden på hans dag.

För mig är Runeberg av ett särskilt skäl förknippad med Staffan Björck. Jag hade förmånen att läsa litteraturhistoria med poetik för honom under hans Göteborgstid. När det var dags för det maktpåliggande slutsamtalet mellan fyra ögon, gav han mig med ett blitt leende temat "allen i svensk poesi".

Vägen med den dubbla trädraden lades åt sidan efter en reverens för Harriet Löwenhjelms: "Vi gingo i en lindallé/ på gula, våta blad,/ och tårar sköljde på min kind/ och jag var ändå glad." Månadsrimmet "februari tjuugoåtta allen/ alla de övriga trettioen" gav anledning till några ord om ramsor, men sedan fick det bli *allena*. Med Wallins version av den gamla psalmen "Allena Gud i himmelrik" och med Lina Sandell-Bergs "och hans mål är blott det ena:/ barnets sanna väl allena" öppnades ett resonemang om den religiösa diktningen. Det ledde så småningom fram till Pär Lagerkvists "Den väg du går allena/ för bort ifrån dig själv" och därmed till den Lagerkvistska tankevärlden.

”Allena var jag,/ han kom allena” förde oss in på Runeberg och hans nyskapande idyll- och epigramdiktning. Vi talade utförligt om hans versepos, om de identitetskapande Sägnera med den inledningsdikt som just i år har en särskild innebörd och om åtskilligt annat. Efterhand tog vi oss fram till senare tiders Edith Södergran, Elmer Diktonius, Gunnar Björling osv. i 1900-talets rika finlandssvenska poesi.

Runeberg hade självfallet varit en tillgång för Svenska Akademien. Det finns emellertid ett kungligt brev från 1811 – utfärdat med anledning av Armfelts manövrer den gången – som slår fast, att man måste vara svensk medborgare för att kunna vara En av De Aderton. När Akademien vid årshögtiden 1839 ville utmärka Runeberg, som det var Akademien ”förment att tillägna sig”, tilldelades han Stora priset, och man betecknade honom i samband därmed som Akademiens osynlige ledamot. Det var samma hedersnamn som en gång – av andra skäl – hade använts om fru Lenngren.

I Runebergs *Dikter* från 1833 står ett poem med titeln ”Den ångrande”. Däri finner vi halvstrofen ”Säg honom då, ifall ett språk du äger,/ De tysta ord jag nu med bäfvan säger”. Den som tilltalas av den ångerköpta flickan är poppeln, i vars skugga det kyska mötet hade ägt rum.

Naturen har många röster i den svenska dikten. ”Säv, säv, susa/ våg, våg, slå,/ I sägen mig var Ingalill den unga månde gå?” Så Fröding, tonsatt av Sibelius. ”Vad suckar häcken?/ Vad nordans storm, som i tallen gnyr?/ Vad viskar bäcken, där genom dalen han sakta flyr?” Så Stagnelius som världssjälens tolk. ”Äppelträd och körsbärsträd, de skrattar tyst åt Salome/ de blommar i min tunnel.” Så Tranströmer i dikten ”Sena maj”.

Vad är i själva verket ett språk? Låt oss notera, att ingen annan varelse än människan har något språk i ordets egentliga mening. Vad vi kräver är då det som har kallats dubbel artikulation, detta att med hjälp av ett litet antal element, t.ex. ljud, kunna bilda praktiskt taget oändligt många betydelsebärande enheter, t.ex. ord. Inte desto mindre är naturligtvis studiet av andra kommunikationssystem – schimpansens, biets, delfinens och andras – av stort intresse.

Det lilla ordet *äger* i Runebergs dikt har ungefär betydelsen 'har tillgång till'. Som bekant finns det många andra schatteringar. Låt mig kontrastera Runebergstället med några fall där man gör anspråk på eller faktiskt har laglig rätt att helt förfoga över språkliga ting.

Vi känner alla till veteraner som *dynamit*, *masonit* och *diesel*, ursprungliga varumärken respektive personnamn som har fått en vidare användning och tagits upp i allmänspråket. I och med detta har de med det hemska ord som används i sammanhanget "degenererat". Vad som i själva verket har skett är motsatsen: benämningarna har visat sig livskraftiga och vunnit en plats i det allmänna språket.

Ett aktuellt fall är *galleria*, som betecknar en täckt gata med butiker. Från början fanns det en galleria i Stockholm; nu finns det gallerior på många håll i den staden och i andra städer, och ordet används som vilket appellativ som helst. I Italien har det funnits gallerior sedan länge. Frågan är nu om den som först importerade ordet och företeelsen skall få ensamrätt till ordet i svenskan och därigenom principiellt ha möjlighet att lagsöka dem som använder det om andra anläggningar liksom att tvinga ordboks författare att markera det som varumärke.

Naturligtvis behövs en skyddsmekanism för varumärken. Det är inte det saken gäller. Vad det däremot gäller är lyhördhet för allmänspråkets behov. Ord som redan är drabbade av äganderätt är exempelvis *keso* och *tetra*. Det görs också ansträngningar att få rena ordled till exklusiv egendom, exempelvis *radi-*. Om något sådant skedde, skulle konsekvenserna som lätt inses bli oöverskådliga. Man skänker en tacksamhetens tanke åt tandkrämen Tandkräm.

Ibland blir det i och för sig trångt på språksamfundets allmänning. Det gamla *eko* (av grek. *ekhó*) som står för 'genljud, återskall' lever och frodas. Det har också utvidgat sin räjong genom *Dagens eko* och *ekonyheter* i radion och genom *eko-frågor* i grammatiken, sådana som ("Vem är den gamle mannen?" –) "Vilken gamle man?". För några år sedan började man så tala om *ekobrott* för *ekonomiska brott* (till grek. *oi'kos* 'hus'). På sistone har ordet också använts som kortform av

*ekologiska brott*. Och därmed blir den ena villan värre än den andra. Ett ord som *ekofråga* blir åtminstone femtydigt: en fråga om genljud, en fråga för ett visst nyhetsprogram, en fråga av viss grammatisk typ, en fråga om hushållning och en fråga om miljön. Det är lätt att tänka sig en mera ändamålsenlig egenomsförvaltning.

"Vem äger språket?" frågar Sandro Key-Åberg i en dikt, och han svarar: "Ensam äger ingen språket." När man arbetar med en ordbok, en encyklopedi eller ett program i TV eller radio förekommer det, att personer eller grupper tar kontakt och med olika grader av emfas gör anspråk på att bestämma över ett ords form, innehåll eller användning. Man kan vilja diktera ordets betydelse, förtiga de konnotationer eller bibetydelser ordet har eller rentav utplåna ordet eller en av dess betydelser.

I frikyrkliga kretsar är man ofta avvisande mot ordet *frireligiös*. Delvis beror detta på ett missförstånd: en hopblandning av *frireligiös* med det principiellt annorlunda *religionsfri*. Ordet *frireligiös* innebär att man står fri i förhållande till den officiella trosriktningen, ordet *frikyrklig* i stället att man står fri i förhållande till den officiella kyrkan.

Till stor del täcker betydelserna varandra, men det är också uppenbart att båda orden behövs. De står följdriktigt i Svenska Akademiens ordlista. Man kan inte tillmötesgå kravet att stryka *frireligiös*, men vi bör naturligtvis ta hänsyn till de frikyrkligas önskemål i vårt språkbruk.

Från kyrkligt håll vill man gärna bli av med den användning av *döpa* som gäller namngivning: "De döpte statyn till Sjöjungfrun", "De döpte den nya mässbyggnaden till Högmässan" osv. I de mest utpräglade fallen uppfattas detta som hädelse. Återigen finns det anledning att tänka på vad man säger, men att utrota uttrycket *döpa någon eller något till något* är både ogörligt och orimligt. Upphovet är ju dessutom kyrkans egen månghundraåriga förknippning av dopceremonin med föräldrarnas namngivning.

Också oberoende av vad jag nu har sagt kan vi konstatera, att det dessvärre finns fördomar om språket. Den mest svårartade gäller det egna tyckandets supremati. Det heter *kex* [k-] och



*kilo* [k-], säger man, ovillig att erkänna de vitt spridda försvenskade uttalen *kex* [ç-] och *kilo* [ç-]. I Luciasången heter det "Kring jord som soln förgät/ skuggorna ruva", för så har vi alltid sjungit hos oss, säger man, trots det grammatiskt omöjliga utbytet av textens *förlät* 'övergav' mot *förgät* 'glöm!'. Fördomar som dessa bottnar säkert ofta i bristande kunskap och stundtals också i ovist nit. Det ter sig onekligen som försök till privat ägande av språket men skulle nog kunna tolkas som ett utslag av språkets fundamentala roll i den enskildes liv.

Fördomar yttrar sig också på andra sätt, inte minst i ett okritiskt övertagande av svepande generaliseringar som fått bestämmda språkliga uttryck, exempelvis *negrer är lata*. Det är detta som Tage Danielsson har fångat i en dräpande grimas i boken *Grallimatik*: "Det var mig en flitig neger. Han måste vara vit."

Ibland säger man, att engelsmännen inte längre äger sitt språk. Det man då syftar på är de språkliga konsekvenserna av att engelskan numera hanteras av användare på alla möjliga håll och på alla möjliga nivåer. Inget språk förblir självfallet opåverkat av en sådan situation.

Svenskan blir möjligen inte ett världsspråk på samma sätt som kusinen från England. Inte desto mindre används det dagligen av hundratusentals invandrade talare med rikt växlande bakgrund. Såvitt vi hittills har kunnat se tycks detta likväl inte få någon genomgripande betydelse för språkutvecklingen. Osvuret är bäst, men enstaka nedslag i ordförrådet – just nu bland annat *wokade* grodlår – framstår som de mest märkbara spåren. Allmänt sett kan det väl också sägas vara rimligt, att de som med allvarligt uppsåt söker sig till ett nytt land strävar efter att tillägna sig landets språk.

Hur är det nu, äger finlandssvenskarna ett eget språk, skilt från rikssvenskan? Om detta har det ju förts en diskussion här i landet under senare år. Det man främst har pekat på är vissa drag i uttalet, särskilt kortstavigheten, vissa egenheter i ordförrådet och influenserna från finskan.

Jag måste bekänna att jag är föga imponerad av de argument för ett separat språk som har förts fram. Rikssvenskan har

visserligen inga korta rotstavelser, men de finns i dialekterna på flera håll. Och om en finlandssvensk säger *väri* har vi inga som helst svårigheter med att uppfatta det som *varit*. Om en finlandssvensk använder ordet *råddig*, smäller det inte högre än om en västsvensk säger *tyken* för 'uppnosig' eller en östsvensk säger *vart* för 'blev'. Variation är ett utmärkande drag för alla språk, även för riksspråkliga former. Den stör i allmänhet inte kommunikationen, är ofta rentav uppfriskande. Vad finskans inflytande beträffar är det å ena sidan så, att frånvaro av påverkan skulle vara uppseendeväckande. Å andra sidan hävdar sig finlandssvenskan uppenbarligen väl. Från min utsiktspunkt noterar jag alltså inte något separat finlandssvenskt språk. Vi talar alla svenska.

I Svenska Akademiens ordlista noteras följaktligen finlandssvenska specialiteter såsom *aktionär* 'aktieägare', *gulnäbb* 'nyinskriven student', *halare* 'överdragskläder', *lokal* 'bostadslägenhet' och *ren* 'redan' likaväl som sydsvenska egenheter som *mölla* 'kvarn', *rälig* 'otäck' och *ifjor* 'i fjol'. Allt som allt redovisar den senaste upplagan av SAOL – den elfte 1986 – 138 ord som finlandssvenska eller provinsiella och finlandssvenska. Många av dem är av denna senare typ, alltså representerade också i rikssvenska regioner. Hit hör exempelvis *blida* 'blidväder', *skrinna* 'åka skridsko', *långledas* 'ha långtråkigt' och *stickelbär* 'krusbär'. Detta är varken mer eller mindre än språklig rikedom.

Urvalet av finlandssvenska ord och uttryck i SAOL är kanske inte helt idealiskt. När det nu börjar dra ihop sig till en ny upplaga – den tolfte 1996 om det går som planerat – hoppas vi på assistans från härvarande expertis.

Svenska språket har några viktiga år framför sig, egentligen avgörande år. Det hänger förstås samman med Sveriges ansökan om medlemskap i EG. Ibland formuleras det som att vi söker inträde i Europa. Det är en alldeles ohållbar formulering med tanke på att Sverige liksom Finland ovedersägligen är ett europeiskt land. Hur som helst kom det i somras signaler om att Sveriges representanter i Bryssel inte räknade med att satsa

på svenskan som ett av EG-språken, i det fall att Sveriges ansökan om medlemskap skulle godtas. Det stod faktiskt så!

Min mening är att svenska språket som kulturspråk står och faller med detta ärende. Svenskan är värd ett bättre öde än att bli betraktad som ett slags polarpladder. Jag har påtalat hotet i TV och radio. Nordiska språkmötet i Reykjavik i höstas sände ett uttalande till svenska regeringen om nödvändigheten av att svenskan får samma ställning inom EG som andra nationella språk. Jag tackar Finlands representanter för hjälpen. Regeringen har gjort vissa utfästelser, bland annat med utgångspunkt i att EG:s regler blir en del av den svenska rättsordningen. Låt mig inskjuta att 15 000 sidor text ur EG:s regelverk nu har översatts till svenska. Under alla förhållanden kan vi inte slå oss till ro med utfästelser. Skarp utkik föröver anbefalls.

Vi skall verkligen inte förhäva oss, men vi behöver heller inte stå där som supplikanter med hatten i handen. Det svenska språket och den svenskspråkiga kulturen har något att ge. Redan nu är det fråga om ett ömsesidigt tagande och givande. Det är alltså fullt möjligt att sträcka på ryggen inför EG-förhandlingarna om språket.

Till detta vill jag också lägga, att ambitionen inom EG är att vidmakthålla medlemsstaternas språkliga mångfald. Det understryks dessutom att språklig information och kommunikation intar en framskjuten, man säger rentav primär, ställning som bas för ekonomisk tillväxt.

Denna hållning stämmer väl med det intryck jag har fått i mina kontakter med EG, främst i samband med uppdraget att i ett par omgångar utvärdera maskinöversättningsprojektet Eurotra. Redan storleken hos den satsningen – i runt tal 150 forskare i 10 år – talar sitt tydliga språk.

Samma inställning har Europarådet, som genom manifestet i Tours för några år sedan riktade en uppmaning till regeringar och personer i ledande befattningar att slå vakt om Europas språkliga mångfald, betecknad som en av världsdelens rikedomar. Slaget är i varje fall inte förlorat på förhand.

Från en synpunkt sett kan man fråga sig, om det kanske snarare är språket som äger oss än tvärtom. Varje språk är ett uttryck för ett särskilt sätt att förhålla sig till tillvaron, en särskild världsbild. Harry Martinson har talat om "språkets rörelse över tingen". Med språkets hjälp kategoriserar vi vår omvärld – den materiella och den andliga – som alltså ter sig olika för olika språksamfund. Internationella förhandlingar lider ofta svårt av detta.

Vad betyder det inte för verklighetsuppfattningen att vi har olika ord för 'värd' och 'gäst' till skillnad från språk som har samma ord, att vi har samma ord *låna* för både det att låta någon annan utnyttja något man har och det att själv få utnyttja vad en annan äger, att vi har olika ord för 'se', 'höra' och 'känna', medan det finns språk som har samma ord för detta? Hur ser verkligheten ut för den som har att ange om en handling är fullbordad eller ej i stället för som hos oss hur den ligger i tiden? Våra släktskapsord, våra färgord och mycket annat bildar system som skiljer sig från andra språks. Gunnar Fredriksson har noterat en typisk bestämmelse i mellanfolkligt samarbete: "Med jultomte avses även julhare."

Så ojämförliga ting som *allemansrätt* och *lutfisk* kan också få tjäna som exempel. Och tänk på de associationer som knyter sig till poetiska formuleringar som "vår andes stämma i världen" och "jag trivs bäst i öppna landskap" eller politiskt färgade uttryck som "vi nådde inte ända fram" och "det hade jag ingen aning om". Att ta till sig ett nytt språk är att öppna ett nytt fönster mot verkligheten. Sammantagna är de förhållningssätt de olika språken erbjuder en oerhörd tillgång för mänskligheten. För filosofer, språkvetare, författare, översättare, jurister och många andra rör det sig om en första rangens utmaning.

Allt tal om ägande blir i sista hand ändå fåfängt. Kvar står det individuella förhållandet till språket; språket som personlighetens kärna och medvetandets bas; språket som tillvarons universalredskap; språket som vår mötesplats med de gamla och med de ofödda. Om språket hotas av undergång, drabbar det generationers liv och möda. Som Ulf Linde har formulerat

det inför de uttjanta tingen på en loppmarknad i Paris: "Här dör de döda en gång till." Eller med Kjell Espmarks ord om landet där ett språk har dött: "Visst hör vi fåglarna i trädet/ men var har sången blivit av?"

## Tal inför överlämnandet av Svenska Akademiens nordiska pris till Thor Vilhjálmsen

Svenska Akademiens Handlingar 1992.

Enligt gudasagan skapades en man som fick namnet Kvaser av gudarnas spott. Hans visdom blev stor. Men han dräptes av dvärgar, och de blandade hans blod med honung i kitteln Odrörers. Den som drack av detta mjöd blev visman och diktare. Därför kallades diktarkonsten Kvasers blod och Odrörers vät-ska.

Sedan kom jätten Suttung över mjödet. Då listade sig Oden dit, tömde det i tre klunkar och flög därifrån i örnhamn. Suttung upptäckte det och gav sig efter. När asarna såg Oden komma, satte de ut kar på gården, och Oden spydde upp mjödet i dem. Men Suttung hade kommit så nära, att Oden var tvungen att släppa en del mjöd den bakre vägen. Det kunde vem som helst ta för sig av; det blev de usla skaldernas lott.

Så berättas det i Snorres Edda. Med sin humor och sitt allvar är detta en underbar myt om diktarkonstens upprinnelse. Den stakar också ut en hållning, som ännu är starkt levande på Island.

Det är en märkvärdig nation. Där är den poetiska Eddan, en samling dikter utan motstycke; Egill Skalla-Grímsson, som tre-årig diktade sina första strofer, sju år gammal utförde sitt första dråp och, när den dagen långt senare kom, skrev sin djupt personliga sorgedikt Sonatorrek 'Förlusten av söner'; detta var i mitten på 900-talet, nästan tre sekler före Snorre; där är Hallgrímur Pétursson med passionspsalmerna på 1600-talet; det senaste sekelskiftets Einar Benediktsson, jurist, lyriker, visionär, verklighetens motsvarighet till den centrala gestalten i Thor Vilhjálmsens intensivt levande roman Gråmossan glöder; där är Sigurður Nordal, som förändrade synen på sagalitteraturen; Gunnar Gunnarsson och Halldór Laxness, litterära

fyrbåkar. Och i Vigdís Finnbogadóttir har man nu en president med starkt kulturellt engagemang, en inspiratör i kampen för språk och identitet.

Gestalter och platser ur sagorna finns med i det dagliga livet. Och vättarna ger sig tillkänna. Vägen får kröka, när stenen är bebodd. Islänningarna måste befolka landet med fantasi, har Thor sagt, eftersom det är så glest mellan människorna. Om man som många av Akademiens ledamöter har suttit i gräset på Lidarände, lyssnat till stycket om Gunnar ur Njåls saga och sett ut över Markarfljót, där Gunnar vände – ”Jag vill rida hem igen och fara ingenstans” – då har man drabbats av förtrollningen.

Välkommen hem, säger flygvärdinnan när planet har landat i Keflavík. Islänningar längtar ut och far ut, men de far alltid utan, dvs. tillbaka igen. Så också Thor. Men han är resenär inte bara i rummet utan också i tiden och texterna. Hans färder i världen och litteraturen präglar hans diktning. Han är rapportör och introduktör men också frontkämpe med traditionen i blodet. Prosan glimmar av naturlyrik, och han är naturligtvis först som sist en storartad berättare med humor och ironi. Hans penna ger klassiska isländska begrepp som ära och öde nya dagrar.

Thor är frågeställare utan väjningsplikt inför makter och myter:

Vad tänker du på? säger hon när han hade slagit sig ner.

Han teg.

Skål, sa hon.

Han sa: Tänka. Vet du att det om en miljon år kommer att gå så att solen utvidgas. Och havet på jorden kommer i kokning och går upp i ånga. Stenen smälter och blir till luft. Allt liv, forntid, nutid – alla människans minnen, alla tankar, konst blommor människor, och fjärilar, allt. Till och med döden dör. Allt.

Om en miljon år, säger kvinnan förvånad.

En miljon år, svarar han, en miljon år.

Kvinnan släckte sin cigaretten i askfatet som var stort med en bild av en bredbent man i svarta knähöga stövlar, gråa ridbyxor, röd jacka, han hälsade med cylinderhatten högt i vädret: Hej hej. Cigaretten var halvrukt.

Hon gjorde det långsamt och målade askfatets kanter svarta när glöden var död, kramade sedan sönder cigaretten och strödde

tobaken kring fatet, spred den omsorgsfullt med en utbränd tändsticka och fortsatte att rota i tobaken, dessa gula strimlor:

Men Gud, sa hon med försiktigt allvar: tror du att Gud finns?

Om Gud är lågt räknat på tjugofem miljoner ljusårs avstånd från jorden, så tar det ett halvt århundrade för en bön att nå fram till honom och få svar. Bönhörelse.

Ja, säger hon: femti år. Ett halvt århundrade, och de blå ögonen var så ljusa och blåa: men tror du att han finns?

Vem? sa han: va?

Gud förstås.

Om han är på tiotusen kilometers avstånd från jorden ... började han.

Det är hemskt att tänka så, sa hon och lutade sig intill honom och strök med handen över hans rygg.

Vetenskapen är handlingskraftens moder, sa han.

Vi dansar i stället, sa kvinnan.

Detta är hämtat ur *Fort fort*, sa fågeln, en Eliotinspirerad bok om tid, ensamhet och gemenskap, rik och mångtydig i texten på Thors mästerliga vis.

År 1992 går till hävderna som ett märkesår för den isländska litteraturen. Olika bedömargrupper har funnit skäl att saluera på olika sätt. Hrafnhildur Hagalín Guðmundsdóttir fick det nordiska dramatikerpriset, Fríða Sigurðardóttir Nordiska rådets litteraturpris, Guðrún Helgadóttir det nordiska barnbokspriset och så Thor Vilhjálmsson Svenska Akademiens nordiska pris.

Priset kommer ur den Gierowska donationsfonden, och det skall enligt bestämmelserna gå till någon i de nordiska länderna som har gjort betydelsefulla insatser inom något av Akademiens verksamhets- eller intresseområden. Det gäller mottagarens hela produktion eller samlade insats, inte ett enskilt verk. De som hittills har fått priset är Villy Sørensen, Danmark, William Heinesen, Färöarna, Rolf Jacobsen, Norge, Nils Erik Enkvist, Finland, Henrik Nordbrandt, Danmark och Tomas Tranströmer från vårt eget land.

Belöningen är på 250 000 kronor, det är 2,5 miljoner isländska kronor. Efter Nobelpriset är detta det största av de i runt tal femtio priser som Akademien delar ut. Akademiens direktör i höst, herr Segerstedt, kommer nu att överlämna priset till Thor Vilhjálmsson, denne isländske författare som helt visst har skålat i Odens mjöd.



## Corpus Linguistics. Opening Address at a Nobel Symposium 1991

Directions in Corpus Linguistics. 1992.

Ladies and gentlemen, dear colleagues and friends,

There was a time when members of the linguistic corps regarded a corpus as a corpse. If people are less than interested in corpora, they are welcome to it. My comment is that it shouldn't be necessary to push a position to extremes.

As time goes by, things will change. Nowadays corpus work is clearly gaining ground. A few weeks ago, I participated in the Machine Translation Summit in Washington, D. C. There was a trend, which I supported, towards using corpora for the purpose of evaluating machine translation systems. There were even attempts at basing the whole system on corpora. I mention this as an example from a field not commonly associated with our topic.

It seems to me that a key notion in the field is that of authenticity. It is certainly reasonable to take a look at real manifestations of language when discussing linguistic problems.

Now, when you delve into authentic text you will often be surprised at what you actually find. At an earlier Nobel Symposium, on text processing, I called this the *capriccio* of language as opposed to the equally characteristic *ostinato*. As a linguist, you are never safe.

Of course the development of theories and models is fundamental to linguistic research including corpus linguistics. What our observations imply, however, is that the idealization of language and speaker comes out as having a somewhat precarious existence. It remains to be demonstrated that language as a whole is a well-formed system.

Two more things. Corpus linguistics proper is not the heaping of data for its own sake but rather the investigation of data for scientific purposes, witness the impressive preprints of this Symposium. Secondly, we should remember that corpora are not the

only way of obtaining information on language. Tests and introspection are among procedures that readily come to mind.

I am happy to say that problems of current and future interest are abundant: what is a corpus in theory and practice? indeed, what is language in a corpus perspective? to what extent is formalization possible? how to design a corpus or a language bank? how to ensure representativeness of samples? how to cope with the dynamics required?, etc.

Nobel Symposia are arranged in the spirit of Alfred Nobel, who in his will of 1895 set up the Nobel Prizes, awarded annually since 1901 to those who "have conferred the greatest benefit on mankind" in the fields of physics, chemistry, physiology or medicine, literature, and peace. Recently, a prize in economic sciences "in memory of Alfred Nobel" has also been instituted. The Symposia are financed by a fund established by the Nobel Foundation, the Bank of Sweden Tercentenary Foundation (this bank, incidentally, is the national bank of our country), and the Knut and Alice Wallenberg Foundation. Applications are dealt with by the Nobel Symposium Committee, referred to the prize-awarders for their opinion, and finally decided upon by the board of the Nobel Foundation.

The Swedish Academy – not to be confused with the Royal Swedish Academy of Sciences where the public sessions will be held – is dedicated to the promotion of language and literature and is in charge of the Nobel Prize for Literature. Accordingly, she supports the Nobel Symposia in her province. Earlier themes belonging here have been Modern Chinese literature and its social context, Theory and practice of translation, Text processing, The Slavic literatures and modernism, Possible worlds in arts and sciences, Strindberg, O'Neill and the modern theatre, and Conceptions of national history. On two of these occasions I have been delighted to cooperate with this friendly Education Center.

With respect to our present field, a particularly wide perspective is opened up by Lars Huldén, the Finno-Swedish philologist and poet, in one of his poems (*Långdansen*, 1976, p. 59). I have tried to render it in English, a totally unauthorized undertaking.

You know, my dear Teophilus, that  
in Heaven there is a concordance  
of all the words that people ever  
said and/or thought as well as  
context enough for assessing  
their inward sense.

But what perhaps you don't know  
– nor did I before –  
is that Hell has been given a  
microfilm of it all for  
the sake of the devils' studies.

In any case, I feel confident of the successful outcome of this earthly meeting of distinguished scholars in a field of great scientific importance and great relevance to society. It is my pleasure and privilege to welcome you on behalf of the Swedish Academy and to do so with the compliments of the Nobel Symposium Committee and the board of the Nobel Foundation.

## Något om NEO och somt om SAOL

Finlandssvenskans framtid. 1992.

Jag kastar mig rakt in i materialen och börjar med en glimt av Nationalencyklopedins ordbok – NEO – i en utformning som inte är ett provtryck utan en preliminär utskrift från ett datasystem (figur 1). Jag återkommer litet mer utförligt till detta. Låt mig säga från början, att jag kommer att tala rätt mycket om standardspråk, men att jag också går in på de regionala aspekter som avtecknar sig mot standardspråket.

Vi kan ta en titt på en av de kortare artiklarna, t.ex. *saltarell'o*, som står i högra spalten nedanför mittens. Man ser där att uttalet anges genom att trycket markeras och genom att det är fråga om accent 1 – till skillnad från *sal'tbad* där det finns en markering av accent 2. Man får ordklassen, böjningen och sedan, inom en särskild parentes, en uppgift om var man lämpligen kan bryta när man avstavar. Som ni ser är det en morfologisk avstavningsmodell som tillämpas. Så kommer vid pricken en definition, 'en tarantellaliknande dans i 6/8-delstakt'. Därefter följer en uppgift om första belägg, 1818, och så etymologin: av italienska *saltarello* osv.

*Saltbad* därefter har som sagt accent 2. I lexem 1 följs definitionen av en stiluppgift. Andra lexemets definition kompletteras av ett definitionstillägg och ett morfologiskt exempel. Den som har sett Svensk ordbok (SOB) tidigare lägger märke till att vi här har tagit ytterligare ett steg i riktning mot att betona lexemens stora betydelse, alltså betydelseenheterens ledande roll. Jag kommer in något på lemma–lexemmodellen senare, men märk alltså att ettorna och tvåorna, som markerar lexemen, är utdragna i vänsterkanten.

Det viktigaste när det gäller de teoretiska utgångspunkterna för arbetet är, att vi betraktar lexikonet som en central komponent både i språkförmågan och i beskrivningen av språkssystemet. Det är i det perspektivet man skall se den här ordboken. Det är också en av motiveringarna till att man kallar området lexikologi. Det

saltpatient subst. ~en ~er (saltpati-ent-en)

- sjukhuspatient som ligger på allmän sal och inte på enskilt rum [1945]

<sup>1</sup> salt [sal't] adj., neutr. ~ (salt-are)

- som smakar som havsvatten, tårar eller svett en av de fyra grundläggande smakerna jämte sött, surt och beskt: saltkött; Skagerrak är ~are än Kattegatt; ~ sill; ~a mandlar; ~a jordnötter o även a) något utvidgat: en ~ vind ifrån havet b) överfört om företeelse eller person som gör ett starkt eller fränt intryck: ett ~ fruntimmer; en ~ sjöman; ~ kritik [före 1520; fornsv. salter; till <sup>2</sup> salt]

<sup>2</sup> salt [sal't] subst. ~et ~er (salt-et)

1 (knappast plur.; ibland förtydligt till koksalt)

- ett vitt kristalliskt ämne som smakar som havsvatten, tårar eller svett och huvudsakligen används för att förhöja smaken på mat eller för att konservera livsmedel {SYN. natriumklorid (mest fackspråkligt); -<sup>2</sup> salt 2}: saltbemängd; → salthalt; bordssalt; havssalt; kryddsalt; medelhavssalt; vittlöksalt; örtsalt; peppar och ~; utvrinna ~ ur havsvatten o även överfört om något som utmärker sig genom styrka, frånhet e. d.: sarkasmer med ~ och sting; kärleken är livets ~ \* strö ~ i såren på ngu fortsätta att angripa någon som redan lider; ta ngt med en uypa ~ inte tro okritiskt på något [före 1520; fornsv. salt; gemensamt germanskt ord, besläktat med lat. sal med samma betydelse]

2 • ämne som i vattenlösning bildar positiva och negativa joner {o hydroxid}: ferrosalt; lödsalt; → närtsalt; → vägsalt; ~er bildar kristaller i fast form men upplöses i vätskor; svavelsyrans ~er kallas sulfater o även utvidgat om vissa andra (vattenlösliga) ämnen: → badsalt; → fruktsalt; → luktsalt [1592; se 1]

salt|a verb -ade -at; -ning, -ande (salt-ar)

- 1 • tillsätta (kok)salt till livsmedel, för att förhöja smaken {Δ smaksätta}: ~ och peppra köttet efter behag o även a) utvidgat, i konserveringsyfte: genomsaltad; låtsaltad; -ad skinka; -at smör b) ytterligare utvidgat: här i staden -ar man vägarna vintertid i stället för att sanda dem c) överfört, särskilt i uttryck för från kritik, ironi e. d.: ~ sitt tal med sarkasmer |~ (ngt) (med ngt), ~ (på ngt) (med ngt), ~ in el. ner el. ned (ngt)/(med ngt)| [före 1520; fornsv. salta; till <sup>1</sup> salt, <sup>2</sup> salt]

2 • lägga oskäligen kostnader på räkning e. d.: en -ad hotellräkning; ~ notan |~ ngt| [1700; se 1]

salt|a in' verb -ade -at; insaltning, insaltande (salt-ar)

- salta för konservering: ~ sill |~ (ngt)/(med ngt)| [1650]

saltarell'o subst. ~n ~r (saltar-ellon)

- en tarantellaliknande dans i 6/8-takt [1818; av ital. saltarello med samma betydelse; till saltare 'hoppa'; av lat. saltare 'dansa'; jfr saltomortal]

saltbad subst. ~et, plur. ~ (salt-bad-et)

- 1 • bad i saltvatten (något älderdomligt) [1824]
- 2 • vätska med lösta salter i vilken metallföremål doppas ned för att hårdas: saltbadsugr [1860]

saltbassäng subst. ~en ~er (salt-bassäng-en)

- bassäng med havsvatten ur vilket man utvinner salt då vattnet avdunstar {o salin, saltbassäng; Δ saltverk} [1761]

saltbehov subst. ~et, plur. ~ (salt-behov-et)

- behov av tillförsel av koksalt till kroppen hos människan och många högre däggdjur: tillfredsställa ~et [1781]

finns till och med de som anser att lingvistik över huvud taget är just detta. Så långt vill jag inte gå, men det finns anledning att betona lexikonets mycket framträdande roll.

En annan utgångspunkt är att de lexikaliserade språkliga enheterna utgörs av solidariteter mellan diskreta begrepp och diskreta former. Man kan komma in på resonemang om det språkliga tecknet, men det är inte precis samma sak. Dessa solidariteter mellan begrepp och former har en bestående karaktär i den meningen, att de i viss mån är permanenta över tiden.

Det tredje är, att betydelseenheter kan beskrivas i ord. Det är inte alldeles självklart att man skall se det på det sättet, men vi har kommit fram till att det är det enda helhetsgrepp man kan ta på ett språkssystem.

Det finns många olika sätt att definiera ord. Ostensiva definitioner, utpekande av saker och ting, kan givetvis inte vara aktuella i en ordbok. Avbildningar, illustrationer, kunde delvis komma i fråga, men hur skall man avbilda t.ex. en lukt? Nätverk har man använt som en modell, i varje fall för delar av det semantiska systemet, men att göra ett nätverk som klarar hela språkssystemet har inte varit möjligt. Komponentanalys har också varit på tal. En sådan kräver en uppsättning särdrag, men ännu har ingen tillnärmelsevis presenterat en särdragsmodell för ett helt språk. Prototyper har diskuterats, i synnerhet inom språkpsykologin. I det fallet är man tillbaka vid särdragsanalysen. Prototyper är i grunden ett slags särdragsknippen. Vissa språkpsykologer har en processmodell, där schema och begrepp, extension och intension, relateras till varandra. Det är alltså fråga om utpekandet i världen, denotationen, å ena sidan och det inomspråkliga begreppssystemet å den andra. Detta leder till ett slags beslutstabeller, som har en förvillande likhet med särdragsnotationer, och så är man tillbaka igen i problemet att särdragen in corpore inte finns explicit angivna. Vi har som sagt stannat för att arbeta uteslutande med ord i våra definitioner.

Metodologiskt sett utgår vi från lemma–lexemmodellen, där lemmat definierar den fundamentala uppslagsformen, den man använder som grund för alfabetiseringen. Lemmana är helt baserade på formella egenskaper: ordklass, böjning, stavning och

uttal. Lexemen är de diskreta begreppen, och vi har vissa regler för att urskilja dem. Det som är problemet om man går till lexikologi som inte arbetar med någon liknande modell utan har ett slags etymologisk bakgrund är, att man inte kan få någon enhetlighet i beskrivningen. För det första känner man inte etymologin på alla håll i ordförrådet, och för det andra finns det ingen som konsekvent har försökt genomföra en etymologisk modell i ett helt lexikaliskt system. Tittar man i ordböcker, kan man finna att *kratta* i betydelsen 'redskap' och *kratta* i betydelsen 'ynkrygg' i den ena ordboken är två olika uppslagsord, i den andra två varianter av samma uppslagsord. *Kruka* på samma sätt. I vissa ordböcker behandlar man redskapet *kratta* och ynkriggen *kratta* som två enheter men förvaringskärl *kruka* och ynkriggen *kruka* som en enhet. Lemma–lexemmodellen är ett försök att bringa ordning i det hela.

Våra definitioner är inriktade på kärnbetydelser eller grundbetydelser, som alltså är diskreta enheter, från vilka man kan avleda betydelsenyanser, t.ex. genom specialisering eller utvidgning. Alla definitioner skall vara utbytbara mot det definierade ordet. Lexem 1 av prepositionen *om* definieras som 'med position i (mer eller mindre slutet) krets runt'. Ett krus i en svängrem *om* halsen är alltså ett krus i en svängrem *med position i (mer eller mindre slutet) krets runt* halsen. Problemet med att definiera i ord är att det finns risk för utbredd cirkularitet – man kan få se *kolossal* förklarad som 'ofantlig' och *ofantlig* som 'kolossal'. Detta har vi försökt komma ifrån genom att begränsa cirkulariteten till ett litet antal axiomatiska element, omkring 4 000, inom vilka cirkularitet är tillåten men utom vilka den inte är det. Vi har inte lyckats hundraprocentigt ännu med detta men arbetar med det som mål.

Olika definitionsformat har utvecklats för olika ordtyper: ordklasser eller subklasser. Större vikt läggs vid centrala ord än vid mindre centrala, semantiskt sett. Det är alltså inte fråga om frekvens utan om semantisk centralitet. Låt oss som ett enkelt exempel ta *äpple* i relation till *papaja*. Vi har *matäpple*, *äpplepaj*, *knycka äpplen*, *adamsäpple*, *äpplet faller inte långt från trädet* osv. Det är en mycket användbar enhet. Men det är ingen

som talar om *matpapaja*, *papajapaj*, *knycka papaja* osv. Det har inte samma centralitet i det semantiska systemet.

Autentiska språkprov ser vi som viktiga för belysningen av de lexikaliska enheterna. Av ordbokens användare krävs härvidlag en viss mognad. Det finns en del som tror eller låtsas tro, att vi hyser de uppfattningar som kommer till uttryck i exemplen. Jag för min del har blivit anklagad för att vara en bottenlös kvinnoföraktare, en våldsam kommunist, en rabiat Amerikafiende osv. bara därför att det finns exempel i SOB som belyser hur orden används. Det är en stor uppgift för pedagogiken och renhållningen att komma till rätta med detta.

Regionala specialiteter, som denna konferens delvis skall handla om, tar vi också upp. Vi har i stort sett samma hållning som den SAOL uppvisar. Därför tror jag att jag uppskjuter min kommentar, tills jag går in på nästa upplaga av SAOL. Standard-språket finns ju i båda våra länder, och regionala ting noteras. De må förekomma i Sverige eller i Finland.

Kanske skulle jag också antyda hur det ser ut, om man arbetar med den utrustning som nu för tiden finns. Jag skall säga till dem som är intresserade, att vi kör på HP-datorer med det stora databssystemet INGRES. Man kan med det utveckla tillämpningar som ger fantastiska arbetsmöjligheter. Christian Sjögren är i full färd med detta. Vad man får är ett antal fönster, och det syns på bilden av skärmen (figur 2), hur man kan plocka fram ett lexem, t.ex. all, med tillhörande uppgifter: definition, definitionstillägg, fri kommentar, morfologiska exempel, syntaktiska exempel, hänvisningar etc. När man arbetar med sin artikel, kan man gå in var som helst i de olika delarna och göra ändringar eller tillägg. Är det fråga om hänvisningar, får de omedelbart genomslag i hela systemet. Detta underlättar förstås arbetet i hög grad. Så småningom skall hela ordboken komma ut på en CD-skiva.

Forskningsbakgrunden är i korthet den, att efter Nusvensk frekvensordbok i de fyra delarna och efter vår medverkan i det hektiska slutskedet av 11:e upplagan av SAOL detta har varit huvuduppgiften inom den lexikologiska delen av institutionens verksamhet. Jag skrev 1976 planen för projektet Lexikalisk databas. Som ett första resultat utkom 1986 Svensk ordbok, och den



Program: **Lexemeter Neo**

### Lexemets kärnbetydelse.

Lexem	Ordens	SI
all 1/1	pn	0
Fermsif kommentar (7)		
mest i böjda former		
tyckstform (7c)		
1		
Definition (8)		
som omfattar hela uttalsformen av		
Definitionstilläg (7)		
den givna el. underförstådda mängden		
Hersifbommentar (8f)		
Dold hersifbommentar ((8>>))		
Frihomsbommentar (15)		
SAOI: kallar det pron. adj. (det tycker jag är bra); (diom: vad/var/hur i B		

Morfologiska exempel (10)		Factsområden (8)		Stil. brud. m.m. (8)	
Exempel	Fri kommentar	SI	SI	SI	SI
allting	0				

Syntaktiska exempel (9)		
Exempel	Parasit (8g)	Fri kommentar
hon ägnar åt sin fristid åt jogging		0
Häll ut åt vattnet		0

Måttenheter (8e)						
Ord	nr	Typ	Ex. ersättningsst	Måttenh. till	Kommentar (8d)	SI
a	1	Jämför		hel 1/1	hel 1/1	0
b	1	Jämför		varje 1	varje	0
c	1	Jämför		samtliga 1	samtliga	0

FIGUR 2. NEO: bild på dataskärm.

har nu fått en förändrad skepnad, därför att Norstedts förenhetligar alla sina ordböckers utseende, och den heter från och med i år Norstedts stora svenska ordbok. Det finns nämligen också en förkortad version, som ser litet annorlunda ut och som kom 1991. Dess titel är Norstedts svenska ordbok. Den innehåller i stort sett samma ord men är något begränsad, särskilt i fråga om exempel.

Under 1990 skrev vi kontrakt med Bra Böcker, som ger ut Nationalencyklopedin, om att utarbeta Nationalencyklopedins ordbok. Den är på intet sätt identisk med Svensk ordbok men utgår från den lexikaliska databasen i det skick den har fått genom arbetet under åren. Ordboken skall omfatta tre band i Nationalencyklopedins format. Den skall enligt överenskommelse utkomma 1995, innan ännu hela Nationalencyklopedin är utgiven.

Jag vill i anslutning till detta säga, att det är av stor vikt att se lexikaliskt arbete som ett oavslutbart arbete. Det var innebörden i den första mening jag skrev i förordet till SOB. Hela tiden hänger det saker och ting med språket, och man upptäcker ständigt nya egenskaper hos orden och nya sammanhang mellan dem. SOB är på det sättet en rapport från ett forskningsfält, en forskningsrapport. Vi inbjöd till kommentarer och fick en mängd sådana av olika slag. Dem har vi utnyttjat i det fortsatta arbetet.

Fem av de åtta som medverkar i arbetet på NEO har erfarenhet från tidigare projekt. Därtill har vi som uttalskonsult Per Lindblad, docent i fonetik i Göteborg. Vi har många ideellt arbetande fackexperter, och vi kan utnyttja Nationalencyklopedins experter i den mån vi behöver det för fackordens granskning.

För att ge en fylligare bild av den kommande ordbokens karaktär tar jag nu för mig en litet längre artikel. Om vi betraktar lemma 2 av *salt* (figur 1), som står i vänsterspalten strax ovanför mitten, finner vi efter stickordet en uttalsuppgift inom parentes. Den är förkortad; det skall komma till en uppgift om vilken accent ordet *salt* har i plural. Jag tror inte någon ordbok har givit det tidigare, men vi avser att ge den musikaliska accenten för plural av enstaviga ord. Det heter ju å ena sidan *sal`ter* med accent 2 och å andra sidan *gre`ker* med accent 1. Det finns åtskilliga i Sverige som behöver den uppgiften.

Efter ordklassen och böjningen kommer som synes en avstavningsuppgift, och i den har vi valt bestämd form. Böjningsformerna är minst lika intressanta från avstavningssynpunkt som de oböjda formerna, och när det gäller den oböjda formen i det här fallet är den ju ointressant i sin korthet. Definitionen och definitionstillägget följs av en synonym och en hänvisning till lexem 2. Framför ett av de morfologiska exemplen, *salthalt*, står det en pil, vilket innebär att det också finns som eget uppslagsord. Så följer de syntaktiska exemplen. Efter den lilla ringen noteras en överförd betydelse med exempel, och därefter kommer stjärnan som markering för idiomen: *strö salt i såren på någon* etc. I den etymologiska parenteser anger "före 1520", att lexemet är belagt i fornsvenskan. Efter detta följer härledningen till latinets *sal* med samma betydelse.

Under lexem 2 finns det exempel på en annan viktig nyhet, nämligen uppgifter om semantiska fält. Den högerriktade triangeln anger, att *hydroxid* är en kohyponym till lexemet. I grunden är synonymer – som *natriumklorid* under lexem 1 – också kohyponymer, men de är självklart ett specialfall. Lexem 1 av verbet *salta* ger oss exempel på en uppåtriktad triangel. Den står framför *smaksätta*, som är det överordnade begreppet, hyperonymen. Från *smaksätta* kan man sedan komma till *peppra*, *söta* osv. På det sättet behandlar vi de semantiska fälten.

Före de etymologiska uppgifterna i det aktuella lexemet kan man urskilja en parentes, som här är markerad med ett lodrätt streck. Där finner man valensuppgifter, alltså uppgifter om konstruktionsförhållandena, ett betydelsefullt moment som fanns med i den ursprungliga projektbeskrivningen. Vad som anges är att man kan *salta*, *salta något*, *salta något med något*, *salta på något* osv.

Givetvis tillkommer en mängd nyord i NEO: *läktarvåld*, *hålltider*, *växthuseffekt*, myntenheten *ECU* osv. Vi relaterar dessa liksom andra ord till SAOL på det sättet, att vi vill undvika onödiga avvikelser. Det kan exempelvis gälla ordningsföljden mellan jämnöda böjningsvarianter.

Ett annat tillägg avser verbalabstrakter. Det kan man också se under *salta*, där man finner *saltning* och *saltande*. Detta är egent-

ligen ett mycket stort åtagande, därför att det inte är generellt givet, vilka ändelser olika verb kan ha. *Yngling* går ju inte till *ynгла*, vilket däremot *ynglande* gör. När det gäller *mygla* kan man ha *myglande* och *mygling* men även *mygel*. Till *springa* hör *springande* men inte *springning*, däremot *spring* osv.

I fråga om första belägg utnyttjar vi naturligtvis Svenska Akademiens ordsamlingar i Lund, men de är inte tillräckliga, inte bara därför att verket inte är tryckt längre än till S utan också därför att excerperingen inte har förts fram jämnt i tiden över hela alfabetet. Ett ord som *allsvenska*, en idrottsterm, är inte belagt förrän 1962 i samlingarna. Vi vet att det är mycket äldre. *Aja baja* är belagt första gången 1989.

En aspekt på NEO som jag vill nämna är, att vi bland annat genom de semantiska fälten försöker göra den till inte bara en perceptuell utan också en aktiv ordbok. Om man inte kommer ihåg var det är lejonet håller till eller räven eller kanske björnen, kan man slå upp något av de ord man minns. Väljer man *gryt*, får man hänvisning till det överordnade ordet *bo*, och från *bo* kan man sedan komma ner genom hänvisning till *kula*, *ide* etc. Den typen av aktivitet vill vi gärna fånga i ordboken.

När det gäller SAOL:s 12:e upplaga är det några saker som är aktuella. Arbetet har påbörjats, och det är Martin Gellerstam som leder det. Med all säkerhet kommer ordförrådet att utökas och den lexikologiska modellen att stramas upp. Det blir fler böjningsuppgifter; de flesta ord i SAOL har ju inga sådana. Man har utgått från att de kan hämtas i de osammansatta orden, men det är många problem som uppkommer vid sammansättnings pluralbildning. En diskussion om huruvida man skall gå över till bokstavsordning kan förväntas. Hur det blir vet jag inte än.

I fråga om regionala uttryck får jag kanske anknyta till vad jag sade i Helsingfors på Runebergsdagen i år. Jag tog då upp de regionala aspekterna från allmän synpunkt och framhöll, att om en finlandssvensk använder ordet *råddig*, det inte är märkligare än om en västsvensk använder ordet *tyken* i betydelsen 'uppenosig' eller om en östsvensk säger *vart* i stället för *blev*. Det är sålunda rimligt att man i SAOL och alltså också i NEO förtecknar sådana specialiteter. I SAOL finns *aktionär* för *aktieägare*,

*lokal* för *bostadslägenhet* och många fler. På samma sätt noteras sydsvenska *mölla* i betydelsen 'kvarn' och *i fjor* i stället för *i fjol*, som är svenskt standardspråk.

Vi gjorde en liten körning på SAOL och konstaterade, att det är 138 ord i SAOL:s 11:e upplaga som betecknas som finlandssvenska eller provinsiella och finlandssvenska. Många av dem finns i svenska varianter, exempelvis *långledas* 'ha långtråkigt' och *skrinna* 'åka skridskor'. För mig är detta inget problem. Jag ser variationen som språklig rikedom helt enkelt. Men urvalet är verkligen inte lätt att göra, och jag hoppas på ett gott samarbete med dem som kan finlandssvenskan. De finns i det här rummet.

Ett litet ord också om projektet OSA (Om svar anhålles eller Ordbok över svenska språket utgiven av Svenska Akademien). Det är ju så att SAOB, som nu har kommit ut i 30 band, innehåller en oerhörd mängd information. Den är dock till stor del svåråtkomlig, eftersom den enda ingången är bokstavsordningen. Detta var utgångspunkten för att läsa in alla banden med optisk teknik och lagra dem som en databas. Allt som är tryckt är nu inläst. Varje band innehåller 1700 spalter, vilket ger 50 000 spalter totalt. Det blir annorlunda uttryckt 200 miljoner svärtade tecken.

Redan nu kan man använda databasversionen av SAOB för vissa typer av excerperingar. Man kan köra igenom hela materialet efter bestämda teckensträngar och då få ut en konkordans över alla belägg på en given sträng. Vi har fått en förfrågan från Strindbergsutgivningen om att få alla Strindbergsbelägg. Dem kan man nå genom att söka på bokstavsföljden *Strindberg* och skriva ut de rader där den förekommer, vilket också ger de aktuella beläggen. Från en annat land har man bett om hjälp med att få fram uppgifter om det tyska inflytandet på det svenska ordförrådet. I det fallet söker man i den etymologiska parenteser efter förkortningarna för olika versioner av tyska och får ut de uppslagsord där de förekommer.

Målet är att databasen skall vara strukturerad så, att man kan gå in i olika typer av information genom en rad olika ingångar gällande morfologi, stil, definitioner, språkprov etc. Det arbetet pågår, och vi siktar på att det skall vara klart innan året är slut.

Det jag här har talat om ingår i ett strategiskt mönster. Den historiska aspekten täcks ju av SAOB i kombination med databasversionen OSA. Den nuspråkliga aspekten täcks av SAOL när det gäller den normativa delen och av NEO när det gäller den beskrivande delen. Det finns många viktiga och behövliga specialordböcker utöver dessa. Det har emellertid varit naturligt för mig att presentera det perspektiv som jag arbetar i.

Med detta har jag verkställt hotelsen att säga något om NEO och somt om SAOL och inte aktat för rovt att dessutom nämna OSA.

Håll svenskan igång!

## Vad skall husen heta?

Göteborgs-Posten 12 juli 1992.

Den språkliga uppfinningsrikedomen är stor i Göteborg. Den mattas inte, och väl är det. Det finns ett hemligt samförstånd bakom Högmässan och Hedendomen och alla de andra benämningar som vi gläder varandra med och testar våra besökare på.

Men smeknamn är en sak och officiella namn en annan. De officiella, och bara de, har en sådan karaktär att de skall kunna sättas upp på husen. Om ett smeknamn hamnar på en skylt, blir förtrollningen bruten. Det är därför man står med blandade känslor framför skylten Feskekörka.

Universitetet växer in i staden och tar nya byggnader i bruk. De skall också ha namn. Humanistiska fakultetens byggnad vid Näckrosdammen har efter en pristävlan fått det officiella namnet Humanisten. Det kan ses som en kortform av Humanistiska fakulteten i analogi med traditionella benämningar som Anatomen, Patologen o.s.v. I Mölndal har vi på samma sätt fått Pedagoggen. Snart invigs Artisten, där konstnärliga institutioner skall flytta in. Namnet har rekommenderats av universitetets namngrupp. Beslutet har fattats av rektorsämbetet.

Det fanns många fantasirika förslag: Grannlåten, Klanglådan, Notstället, Spelhuset o.s.v. Några föreslog Lyran, intressant men riskfyllt: Var är han? – Han är på Lyran. Det mycket vitsiga Scenfärdigheten fördes också fram. Alla de här är tänkbara som smeknamn, men som officiella namn – icke.

Så kom turen till samhällsvetarnas nya viste i Haga, ett ärende som G-P har ägnat sig åt vid ett par tillfällen. Bland förslagen dominerade smeknamnstypen: Sam-anhanget, Samvetet, Samvetet, Sam-förståndet o.s.v. Namngruppen drog ut linjen från de tidigare officiella namnen och föreslog Samhällsvetaren, som blev fastställt. I dagligt tal används gärna Samvetet, och inget ont i det.

Senast har institutionerna på Brogatan varit aktuella. Där ryms grupper som driver studier av människans villkor i ett globalt

perspektiv. Som samlande namn har man tänkt på Världshuset (G-P 29 april) med anspelning på både värld och världshus.

Lustifikationen har ingenting med verksamheten att göra utan ger förslaget den smeknamnskaraktär som vi har sett så många exempel på. Namngruppens tankar gick till Kosmopoliten, som ju betecknar en som inte känner sig knuten till något särskilt land eller samhälle utan känner sig hemma var som helst i världen. Någon belastning i form av antisemitiska övertoner, som gjorts gällande, finns inte noterad i ordböckerna.

Det är en grannliga uppgift att lämna förslag i namnändringen som dessa, särskilt när man själv gläder sig åt den språkliga lekfullheten. Distinktionen mellan officiella namn och smeknamn är emellertid viktig. I grunden är den en stilfråga.



## Ture Johannisson – nordist

Humanistdag-boken nr 9. Göteborgs universitet 1996. En något utökad version av texten i Kungl. Vitterhets Historie och Antikvitets Akademiens Årsbok 1991.

Det konsistorium vid Lunds universitet som kallades det mindre akademiska utfärdade 1945 ett betyg för Ture Gustaf Johannisson. På tidens ciselerade ämbetsmannaspråk vitsordade man med synnerligt nöje den berömliga skicklighet och det osparda nit varmed han fullgjort sina åligganden som docent sedan 1939. ”Då härtill kommer”, anförde man i det avslutande stycket, ”att docenten Johannisson städse iakttagit ett värdigt och hedrande uppförande, räknar Konsistoriet för en angenäm skyldighet att honom sålunda vitsordad, hos alla och en var till det bästa anmäla.” Det har funnits goda skäl för olika instanser vid skilda tillfällen och i växlande sammanhang att uttala sig väl om Ture Johannisson. Det var en språkforskare med ovanlig resning som gick ur tiden den 16 december 1990.

Johannissons bakgrund var sydlig och nordlig. Fadern, verk-mästare Johan Johannisson, föddes i Algutsboda i Småland. Sitt småländska ursprung var Ture Johannisson starkt medveten om, och familjen har fortfarande anknytning till landskapet. Fadern ägnade sig åt byggenskap med verksamheten efterhand förlagd till orter som Sundsvall, Sollefteå och Östersund. Han karakteriseras som kunnig och varmhjärtad, någon gång möjligen i överkant ömsint i förhållande till branschens kutym. Modern Julia, född Boman, var från Gnarp i norra Hälsingland, några mil söder om Sundsvall. Hennes duglighet är ofta omvittnad. Makarna hade fyra barn, och sonen Ture föddes i Sundsvall 1903.

Skolpojken begåvning blev tidigt uppenbar. Familjens omständigheter var emellertid knappa. Det krävdes uppoffringar, extraarbete och hjälp av en prästman för att uppehållet skulle klaras fram till studentexamen i Sundsvall 1923.

Ett år senare skrev Johannisson in sig vid universitetet i Lund. Efter två år var han fil. kand. och efter ytterligare två fil. mag. Då var det 1928, och det året trycktes hans första skrift, som hand-

lade om stark och svag adjektivböjning i den äldre Eddan. Fram till licentiatexamen 1934 försörjde han sig först som adjunkt i Sundsvall och sedan som svensklektor i Jena, Marburg och Giessen. Lektoratet i Marburg uppehöll han från 1930 till 1938. Året därpå spikade han sin doktorsavhandling i Lund. Samma år gifte han sig med Lore Schmidt, som hade drivit medicinska studier i Marburg.

Hans akademiske lärare var till en början Emil Olson, mest bekant för sin studie över de appellativiska substantivens bildning i fornsvenskan. Aktuella forskningsfält i denna miljö var ljudhistoria, formlära, etymologi och textfilologi. Till tentamen hos professor Olson hade kandidat Johannisson att inställa sig i frack. Längre fram blev hans handledare Erik Noreen, en forskare som bland annat studerat författarfrågor i Lejonkulans dramatik. Det var för honom han disputerade.

Johannisson framhöll ofta värdet av kamratkritik, och i Lund tillhörde han en krets som betydde mycket i både allvar och gamman. Till dessa språkvetare och ordlekare hörde bland andra Birger Bjerre, Gösta Mattsson och John Svensson. Härutöver gav honom vistelsen i Tyskland möjlighet att pröva sina idéer i en omgivning, där kontakten med de junggrammatiska tankarna var starkt förnimbar.



Avhandlingen "Verbal och postverbal partikelkomposition i de germanska språken" var en banbrytande undersökning på ordbildningens och etymologins områden. Följdriktigt tilldelades författaren 1940 den vart fjärde år utdelade belöningen ur Axel och Ebba Kocks fond. I sin avhandling visade han, hur en rad nordiska substantiv och adjektiv med betonad förstavelse kunde förklaras som bildningar till gamla sammansatta verb med obetonad och senare bortfallen förstavelse. Härigenom fick en lång rad problem en sammanhållen lösning. Undersökningen förutsatte en utomordentligt omfattande excerpering av språkligt material. I särskilda uppsatser fortsatte han redovisningen och kunde på analogt sätt exempelvis klara ut uppkomsten av ord som *ärende*.

och *idrott*. Substantivet *ärende* hör således ihop med det fornsvenska verbet *inna*, svarande mot ett urnordiskt verb *\*urinþan* med betydelsen 'utföra'.

Hans huvudarbete på det syntaktiska området var "Hava och vara som tempusbildande hjälpverb i de nordiska språken" (1945). Här klarlade han i vilka uttryckstyper de båda verben uppträder och hur utvecklingen har gått. En förutsättning för studien var en analys av de fenomen som innefattas i begreppen aspekt och aktionssätt. I historiskt avseende visade det sig bland annat, att hava-konstruktionen vidgar sin domän från slutet av 1600-talet och att det från mitten av 1700-talet finns en kraftig tendens att utelämna hjälp verbet i bisats. Vi får en konstruktion med rent supinum som i *Han bekräftade, att brevet kommit fram*.

En 1600-talstext som "Nils Matson Kiöpings Resa" intresserade honom av flera skäl. Det viktigaste var den nya språkform som Lars Salvius gav texten vid tryckningen av tredje upplagan 1743 och som på ett mycket åskådligt sätt belyser den yngre nysvenskans första skede. Ändringarna gällde såväl ordförråd som satsbildning och stil. Johannissons kommenterade utgåva av utdrag ur andra och tredje upplagan publicerades 1961. I oklanderlig överensstämmelse med den kraftfulla kulturvården här i landet har man verkställt nermalning av restupplagan.

I boken "Ett språkligt signalement" (1973) redovisade han sina omfattande undersökningar av författarskapet till de anonyma breven i samband med biskopsvalet i Strängnäs 1952. Johannisson arbetade med dels särskiljande kriterier, dvs. ett batteri av alternativa uttryck sådana som *båda – bägge*, dels individualismer, dvs. avvikelser från normen som kontaminationen *ebba ut i sanden*. Härtill kom också utomspråkliga kriterier såsom förfarandet med skrivmaskiner och kännedomen om vissa sakuppgifter i breven. Samtidigt som Johannissons resultat med rätta kom att tillmätas en mycket stor betydelse för utgången av ärendet, gav studierna fördjupade insikter i den moderna svenskan.

Tankegången var enkel och klar men visade sig ändå ligga över horisonten på några håll. En kritiker ansåg sig sålunda ha

utdelat ett dråpslag när han påpekade, att Johannissons egen utredning överensstämde med de anonyma breven i fråga om bruket av negationer och relativter. Han ställde frågan, om det alltså var Helander som hade skrivit utredningen, och framhöll att frågan inte var skämtsamt menad utan avsedd att vara en tankeställare. ”På detta kunde jag bara svara”, konstaterade Ture Johannisson senare, ”att jag naturligtvis fick en tankeställare, men väl inte den avsedda.”

Nusvenskan blev i själva verket ett allt tydligare intresseområde för honom. I en hyllningsskrift från 1975 med titeln ”Ordbildning och ordbetydelse” samlades en rad uppsatser av hans hand. Här kan man läsa om inte bara *ärende* och *idrott* utan också *plast*, *akademiker* och *övermaga*. Dessutom återfinner man bland mycket annat hans kända uppsats ”Tendenser i nutida svensk ordbildning”. Det var också Johannisson som i samverkan med K.G. Ljunggren gav ut ”Svensk handordbok. Konstruktioner och fraseologi”.



Om hans uppväxt och utbildning hade nord-sydlig orientering, Norrland-Sydsverige-Tyskland, följde hans fortsatta gärning en väst-östlig linje: Göteborg-Stockholm. Johannisson kallades genom beslut av ett enhälligt lärarråd till lärostolen i nordiska språk i Göteborg 1945 och var verksam som professor där till sin pensionering 1970. I denna egenskap var han också undersökningsledare för Institutet för ortnamns- och dialektforskning i Göteborg fram till dess en särskild befattning som arkivchef tillskapades 1957 med Verner Ekenvall som förste innehavare. Åren närmast före Göteborgs högskolas omvandling till universitet 1954 var han sitt lärosätes prorektor. Under hela den aktiva professorstiden deltog han i arbetet inom Göteborgs gatunamnsberedning, och han var inspektor för Västra Sveriges folkbildningsförbund.

Han engagerade sig i arbetet inom Statens humanistiska forskningsråd, Svenska institutet och Längmanska kulturfonden. Under åtskilliga år var han vice ordförande i Svenska språk-

nämndens styrelse och under en tid också ordförande i styrelsen för Tekniska nomenklaturcentralen. I ett decennium var han redaktör för huvudtidskriften på sitt forskningsområde, Arkiv för nordisk filologi.

I Vitterhets Historie och Antikvitets Akademien valdes han in 1951. Han tog sitt inträde med ett föredrag över ämnet "En aktuell språkvårdsfråga och dess historiska bakgrund". För tio år sedan höll han här parentation över Carl Ivar Ståhle, en annan av nordistikens märkesmän.

Vid sin bortgång var han Svenska Akademiens äldste ledamot, invald 1955. I den kretsen gjorde han under lång tid en mycket betydelsefull insats i det språkliga arbetet, i synnerhet vad gällde ordboken och ordlistan. Han var också ledamot av Vetenskaps-societeten i Lund, Vetenskaps- och vitterhetssamhället i Göteborg och Gustav Adolfs Akademien. Dessutom tillhörde han lärda samfund i Finland, Norge och Island.



Ture Johannisson var en människa som man minns med glädje. Genom sin varma personlighet, sin gedigenhet och sina lågmälda men träffande synpunkter var han mycket avhållen bland kolleger och studenter. Och han var med Eddans ord sina vänners vän. Han hade också en stark känsla för de utsatta i tillvaron.

Hans kunnskap kombinerades med en utpräglad formuleringsförmåga. Det kunde gälla omdömen i vetenskapliga sammanhang, formella anföranden eller repliker i vardagslivet. Vi är åtskilliga som bär med oss minnesvärda exempel på denna konst.

När hans sista dag som aktiv professor randades, kom också den siste tentanden. Johannisson hade då med sin vanliga grannlagenhet gått igenom hyllor och skåp och rensat ut sådant som kunde undvaras av efterträdaren. Det utgallrade låg i ett par välfyllda säckar vid skrivbordet, och i dem skyntade åtskilliga papper med brevhuvuden som lät ana avsändare på högre ort. Sobra sopor, förklarade han med en gest och ett leende.

Ture Johannisson satte sig samvetsgrant in i de arbeten som hans studenter på olika stadier lade fram. Vid bedömningen var

han noga med att visa på det som var positivt i samband med att han gav kritik. Kraftiga påminnelser var naturligtvis stundom av nöden, och då kunde han efterlysa inte mera kött på benen utan mera ben i köttet.

Han var öppen för nya riktningar inom språkforskningen utan att falla i farstun för dem. Med pennan i hand kunde han säga ifrån ganska bestämt, när han fann en sak viktig. Den skepsis han i sådana sammanhang gav uttryck åt gällde i första hand, vad han uppfattade som överdrivna anspråk från olika skolbildningars sida.

Hans lärdom var förenad med klarhet och stringens. Detta genomsyrade hans vetenskapliga arbeten, som präglas av en synnerligen njutbar methodsäkerhet. Självklart bidrog det också till att ge den karakteristiska lyftningen åt hans undervisning. Vad han alltid framhöll för sina elever var vikten av saklighet enligt maximen starkt i sak men mildt i form – fortiter in re, suaviter in modo.

Deltagarna i högre seminariet var välkomna till regelbundna postseminarier i Ture och Lore Johannissons hem. Vi som hade förmånen att vara med minns dem som högtidsstunder, då den vetenskapliga diskussionen övergick i samtal. På samma gång som det blev en skola för livet, grundlades en livslång vänskap.



En av Ture Johannissons passioner var att hugga ved. Det är en sysselsättning som kräver styrka och uthållighet men också förmåga att lösa problem. Dessa egenskaper hade Ture Johannisson i rikt mått, och han utnyttjade dem på ett enastående sätt i sin gärning. Hans livsverk består.

## Vilka är de intellektuella?

Svenska Dagbladet 24 februari 1991.

Enskilt och offentligt samtal om språket tillhör de centrala aktiviteterna i informations- och kunskapssamhället. Intresset är också stort i vida kretsar, mycket större än man kunde ana, om man bara hade tidningarnas kultursidor att gå efter. Där är ju språkliga ärenden och språkligt orienterade insatser rejält underrepresenterade.

En av de frågor som ofta ställs i levande livet gäller, vilka de intellektuella egentligen är. Saken har övervägts på åtskilliga håll: av Jefferson Morley i *The New Republic* härförleden, av Paul Johnson i hans bok *Intellectuals*, av Kjartan Fløgstad i *BLM* nyligen osv. Det är omväxlande och delvis stimulerande läsning, dock mera kommenterande än klargörande.

Mot den bakgrunden finns det anledning att göra ett litet språkligt experiment. Låt oss titta efter hur uttrycket används i massmedierna. Det är något som vi kan göra med hjälp av material från Språkbanken i Göteborg.

Om någon exempelvis skriver om vilka önskemål de intellektuella och popdyrkarna har i ett visst fall, är det i princip två olika grupper som nämns. Belägg av den typen ger oss alltså möjlighet att skilja ut popdyrkarna från de intellektuella i det språkbruk vi observerar. Det blir ungefär som att konstatera, att peppar och salt är olika saker och att Andersson, Pettersson och Lundström är olika personer. Vi kan gå vidare på den vägen genom att ta oss igenom språkprov efter språkprov och se vad vi efterhand får kvar i gruppen de intellektuella. Sättet att gå tillväga har likheter med proceduren när man äter kronärtskocka.

Till en början kan vi avskilja en del specialkategorier på basis av belägg som "de omgav sej med celebriteter och intellektuella" och "Paul Hogan möter rånare, hallickar, transvestiter och intellektuella och sätter dem alla på plats med enkelt bondförnuft och en och annan rättfram smäll på käften".

Vi kan avföra politikerna med hjälp av t.ex. ”generationer av politiker och intellektuella i Västeuropa”. Denna och ytterligare en kategori försvinner med fall som ”kritiken från politiker, intellektuella och handelsmän”.

En stor men minskande grupp faller ifrån genom ”en befrielse-rörelse med basen bland bönder och intellektuella”. Andra stora grupper går samma väg: ”de bittraste minnena har de intellektuella, de okvalificerade arbetarna och tjänstemännen”.

Inte heller teologer blir kvar: ”med all respekt för intellektuella och vänsterpräster”. Men de konstnärligt aktiva och vetenskapsmännen? Ingalunda: ”alla konstnärer, författare och intellektuella”, ”samspelet mellan makthavare, intellektuella, konstnärer, forskare och medier”.

Det finns naturligtvis samhällsgrupper som inte har blivit nämnda i exemplen, men det finns också många fler språkprov. Politiker, handelsmän, bönder, arbetare, tjänstemän, konstnärer, författare, forskare osv., alla hamnar de till synes utanför gruppen de intellektuella. Vad dessa texter pekar mot är en mängd utan element, den tomma mängden. Man kan tolka detta som att uttrycket de intellektuella har blivit en joker, en benämning som används för det syfte som passar för tillfället eller som används släpphänt och oöverlagt i största allmänhet av pennföraarna.

Givetvis möter vi en rad exempel, där uttrycket förekommer utan att någon annan kategori nämns i sammanhanget. Problemet med de fallen blir, att den oskärpa vi har noterat i de citerade texterna ger oklarhet också åt de övriga.

Om inte annat kan det vi har iakttagit vara en nyttig påminnelse om att ord blir vad man gör dem till. Vad skulle kräva större varsamhet i användningen än ett uttryck som detta?



## *Textil, slöjd och textilslöjd*

Jubileumsskriften Hjärna – Händer. Göteborgs universitet 1991.

Om man ser till den språkliga härledningen, är *textilslöjd* en hybrid. Första ledet går tillbaka på den latinska verbstammen *tex-*, som betyder 'fläta, väva'. Andra ledet däremot är av urgammalt nordiskt stoff. Det bygger på adjektivet *slög* 'fyndig, skicklig'. Sammantaget gäller alltså för *textilslöjd*, att det rör sig om skicklighet i att knyta samman trådarna. En liten extra poäng är att *-t-* i *subtil* också går tillbaka på *tex-*. Ordet betyder egentligen 'fint vävd'.

Låt oss emellertid börja hemmavid, med *slöjd*. Den gamla betydelsen 'färdighet' levde länge och vidgades efterhand till det mera allmänna 'syssla, verksamhet'.

Mitt bästa är att tyst vid håll'n  
Få minnas flydda dar,  
Då brasan brinner ned, och qväll'n  
Ej flere slöjder har.

Så skrev Runeberg, och året var 1831.

Just under 1800-talet uppstod så småningom allt starkare strömningar med inriktning på att främja skicklighet i handarbete och bevara kulturarvet. Vi fick den nuvarande betydelsen hos *slöjd*, den som med det speciella språket i Svenska Akademiens ordbok betecknas som 'icke yrkesmässigt bedrivande av manuell tillverkning av föremål (t.ex. genom bearbetning av trä eller metall eller genom sömnad eller vävning)'.

Ordet *hemslöjd*, som är känt sedan 1807, blev ett honnörsord längre fram. Det fick i svenskan aldrig någon allvarlig konkurrens från *husflit* – i och för sig belagt sedan 1830-talet – som är norskans och danskans ord (*husflid*).

Utvecklingen ledde också till pedagogisk slöjd eller *skolslöjd*. Själva ordet är belagt sedan 1883. Genombrottet på den här flanken var tillkomsten av Nääs slöjdseminarium 1874 under ledning av Otto Salomon. På den textila sidan fick Maria Nor-

denfelts högre handarbetsseminarium från 1890 stor betydelse. Här användes *handarbete* som term för den textila slöjden. Ordet har ju en mycket allmän innebörd, men med denna speciella inriktning (naturligtvis dock inte i skolsammanhang) är det känt sedan mitten av 1700-talet. Föreningen Handarbetets vänner stiftades för övrigt 1874. Yrkesbenämningarna *handarbetslära-  
rinna* och *slöjdlärlarinna* förekom sida vid sida på 1880-talet.

Ordet *textil* har en intressant historia. När det först dyker upp i svenskan, är det faktiskt som adjektiv. I det äldsta kända belägget, från 1873, talas det om ”den textila ... konsten”. Jag vill gärna inskjuta här, att jag när det gäller textilord har fått kopior på språkprov ur samlingarna på Svenska Akademiens ordboksredaktion (den tryckta ordboken har ännu inte kommit till T).

Sammansättningar skapas redan tidigt. På 1880-talet möter vi sådana som *textilämne*, *textilindustri*, *textilindustriell* och *textil-  
konst*. Man skall försöka känna dessa bildningar så som man upplever *blåsippa*, *fullmåne* och andra ord med adjektivisk förled.

Substantivet presenterar sig inte förrän efter sekelskiftet och då i pluralformen *textilier*, känd från 1907. Det dröjer ända till in på 1930-talet, innan singularen *textil* börjar synas (1934):

Före sparbankernas och det rationella jordbrukets tid bestod ett allmogehems lösa förmögenhet till icke ringa del just av textil.

Troligen har ordet till väsentlig del lösts ut ur sammansättningar av typen *textilindustri* – jämför *möbelindustri*, *skoindustri* osv. Processen kan också ha fått stöd av tillbakabildning från den alternativa pluralformen *textiler*, som uppträder på 1920-talet. På parallellt sätt uppkom singularen *förälder* ur den gamla pluralformen *föräldrar*.

När substantivet *textil* väl hade etablerats, flödade sammansättningarna ännu rikligare än tidigare, och som alltid speglar orden samhällsutvecklingen: *textilkonstnärinna* 1937, *textilingenjör* 1945, *textilforskningsinstitut* 1945, *textilkunskap* 1946, *textilkris* 1957, *textilformgivare* 1965, *textillinje* 1966, *textilbadare* 1970 (naturisternas benämning på den som badar med badklä-

der) och *textilburen* (plast) 1981. *Textillärare* nämns första gången 1957:

Utbildningen av textillärare kommer i praktiken att innebära en sammanslagning av de nuvarande kategorierna handarbetslärare och vävlärlärare.

Intressant nog är *textilslöjd* mycket äldre, belagt med jämna mellanrum ända sedan 1897, utställningsåret.

Textilslöjd beskrivs ibland som ett praktisk-estetiskt ämne. Det innebär språkligt sett, att det är fråga om ett estetiskt ämne med dragning åt det praktiska. Det sköna, det konstnärliga är då den dominerande aspekten.

Om man å andra sidan vill betona handaskickligheten, bör man beteckna ämnet som estetisk-praktiskt. Då framstår det som ett färdighetsämne med skönheten i blickpunkten.

Utfallet kan kanske många gånger bli likartat, men vid utformning av studieplaner och vid uppläggning av undervisningen kan det vara nog så betydelsefullt att bestämma vilken fot man skall stå på. Under alla omständigheter är kombinationen av det praktiska och det estetiska en stor tillgång som det är av vikt att slå vakt om.

## Bibeln och Akademien

Den svenska bibeln. Ett 450-årsjubileum. 1991.

Under Svenska Akademiens två sekler har en enda bibelöversättning blivit stadfäst, den från 1917. Den dominerande gestalten vid dess tillkomst var Esaias Tegnér d.y., ledamot av Akademien sedan 1880-talet.

Det var också från en ledamot, rättare sagt en blivande ledamot, som den skarpaste kritiken mot översättningen kom. H.S. Nybergs attack vid 1934 års kyrkomöte lämnade knappast sten på sten: "Den bibelsvenska vi nu läsa påminner åtskilligt om salig Chronschougs sätt att skriva." Nyberg talar om platthet i uttrycken och brist på udd, skärpa och glans. Särskilt trycker han på en pratsamhet, som han menar stegras till det odrägliga i Nya testamentet.

Vad var det som väckte en sådan ilska? Bibelkommissionen med ärkebiskop Anton Niklas Sundberg och Tegnér bland medlemmarna hade som mål att åstadkomma "en texttrogen översättning på nutidssvenska, med så mycken hänsyn som möjligt till vår gamla kyrkobibel". Detta ställde stora krav på filologen Tegnér, som strävade efter att uttrycka författarens tanke så som författaren själv skulle ha uttryckt den, om han hade skrivit på svenska. Det ledde bland annat till användning av en mängd småord, som skulle bidra till begripligheten och stilens ledighet men som sålunda också gav upphov till kritik för mångordighet.

Jag tar ett exempel ur det evangelium som jag själv särskilt har fäst mig vid, det som också anses vara det äldsta. Markus 1: 2-5:

Så är skrivet hos profeten Esaias: "Se, jag sänder ut min ängel framför dig, och han skall bereda vägen för dig. Hör rösten av en som ropar i öknen: 'Bereden vägen för Herren, gören stigarna jämna för honom.'" I enlighet härmed uppträdde Johannes döparen i öknen och predikade bättringens döpelse till syndernas förlåtelse. Och hela judiska landet och alla Jerusalems invånare gingo ut till honom och läto döpa sig av honom i floden Jordan, och bekände därvid sina synder.

Det är inte utan att stilkänslan reagerar inför en smått kameral vändning som *i enlighet härmed* och ett inskott som *därvid*.

Polemikern Nyberg fann efterhand, att han inte hade gjort Tegnér rättvisa. Det är en betydligt dämpad pennförare man möter i minnesteckningen över Tegnér bortåt trettio år senare. Nyberg ställer sig fortfarande kritisk, men han erkänner att han ensidigt hade skjutit in sig på svagheterna och förbisett Tegnér's stora förtjänster om bibeln.

En parallell till det nu noterade finner vi drygt hundra år tidigare. Mot Tegnér svarar Johan Adam Tingstadius, invald i Akademien på 1790-talet. Han arbetade på egen hand som bibelöversättare och publicerade exempelvis översättningar av Ordspårboken och Psaltaren. År 1786 blev han medlem av den då ännu unga bibelkommissionen, som i honom fick en drivande kraft. Till 200-årsminnet 1793 av Uppsala möte lade kommissionen fram en reviderad bibelöversättning, som dock inte betraktades som slutgiltig. Arbetet fördes vidare, och 1815 var man i slutskedet av en provöversättning av Nya testamentet.

Då slog Johan Olof Wallin till. I ett memorial till prästeståndet försvarade han det ärvda bibelspråket i starka vändningar. Det vore otillåtligt att, "förebärande smakens fordringar, putsa, omflytta eller bortbyta de heliga Urkundernas bekanta ord och ordställningar". Och han fortsatte: "Huru vill man der krossa den yttre formen, utan att det inneboende Heliga förlorar något af sin kraft för *vårt* sätt att känna och fatta?" I ett slags sammanfattning av sin nyromantiskt influerade hållning skriver han: "*Det fordras, säger man, ett vårdadt och städadt språk. Men är då fädernas manliga, välljudande, böjliga och för oss ännu tydliga språk ovårdadt, derföre att det ej talas, ej bör och kan talas af Svenska Akademien?*"

Tingstadius, som var bibelkommissionens sekreterare, gick naturligtvis i svaromål. Lika naturligt var att han återopade instruktionen för kommissionen. I den föreskrevs att man skulle "följa Textens innehåll och rätta ordaförstånd och det genom Svenska ordalag af lika betydelse, uti en ren och otvungen mening uttrycka". Tingstadius, som liksom Tegnér var filolog (och hade orientaliska språk som sin specialitet), var i och för sig väl

skickad att arbeta med de bibliska texterna, i synnerhet inom det slutande 1700-talets rationalistiska ram.

Han hade också vunnit erkännande. Den översättning av Psaltaren som han publicerade 1791 belönades med Lundbladiska priset av Svenska Akademien. Kellgren berömde i Stockholms-Posten hans översättningar av Psaltaren och Höga Visan. I sitt svar på Tingstadius' inträdestal i Akademien framhöll kanslern, Abraham Niclas Edelcrantz, särskilt "renheten i uttrycken, klarheten i tankarne, färgen i målningarne, högheten förenad med enfald i stilen".

Å andra sidan är det inte bara Wallin som har framfört kritik. Från senare tider hör vi åter H.S. Nybergs röst, nu i minnesteckningen över Tingstadius: "Med lätt överdrift skulle man kunna kalla Tingstadius' språk gustaviansk herrgårds- och salongssvenska." Denna salva motiverar han med att översättningarna stilistiskt präglas av "tidens modeord, dess urbana och sirliga men en smula bleka och abstrakta uttryckssätt".

Kommissionens provöversättning av Nya testamentet trycktes 1816. Så här utföll den valda passusen:

Såsom skrifwet är af Propheterna: Si! jag sänder mitt budskap, framför Dig, hwilket bereda skall din wäg för Dig; En röst ropar: Jöknen bereder wäg för HERran: Baner stigarne för Honom; framträdde Johannes i öknen att döpa, och predikade bättring med dop, till syndernas förlåtelse. Till honom utgick hela Judeens land, samt med Jerusalems invånare, och alle döptes af honom i floden Jordan, sedan de bekänt sina synder.

Även om en ordställning som *bereda skall* och en imperativ som *bereder* återgår på Gustaf Vasas bibel från 1541, visar översättningen en betydligt modernare språkform, samtidigt som den gör ett självständigt intryck.

Vid samma tid bildades emellertid bibelsällskap på åtskilliga håll, och de gav spridning åt den officiellt erkända översättningen, Karl XII:s bibel av 1703, som stod Gustav Vasas bibel mycket nära. Det gustavianska bibelverket hade förlorat slaget om 1800-talet.

Ännu vid det senaste sekelskiftet kunde konfirmanderna således få sin bibeltext i en form, för vilken fraktur var ett adekvat typsnitt. Jag citerar ur ett tryck från 1898:

Såsom skrifwet är i propheterna: Si, jag sänder min ängel framför ditt ansigte, hwilken bereda skall din wäg för dig; Ens ropandes röst är i öknene: Bereder HErrans wäg, görer hans stigar rätta! Johannes war i öknene, döpte och predikade bättringenes döpelse till syndernas förlåtelse. Och till honom gingo ut hela Judiska landet och de utaf Jerusalem och läto sig alle döpa af honom i Jordans flod och bekände sina synder.

Bortsett från ett par småting (*Ens* för *Een* och *Judiska* för *Judiska*) är det bara stavningen som skiljer texten från den i Gustaf Vasas bibel 350 år tidigare. Under alla omständigheter ger detta ytterligare relief åt Tegnér's insats.

I Svenska Akademiens ändamålsparagraf har stiftaren ålagt ledamöterna "att arbeta uppå Svenska Språkets renhet, styrka och höghet, så uti Vetenskaper, som serdeles i anseende till Skaldekonsten och Vältaligheten uti alla thes tillhörande delar, jemväl uti then, som tjenar att tolka de Himmelska Sanningar". Här finns alltså en direkt anknytning till det religiösa språket, men som synes inte till bibeln själv. Akademien som sådan har följdenligt varit mycket litet engagerad i översättningsprojekt genom åren. Det har varit en uppgift för den långlivade bibelkommissionen, den som inrättades några år innan Akademien stiftades och som kom att verka jämsides med Akademien i 131 år, fram till 1917.

Som framgått har däremot enskilda ledamöter gjort betydande insatser i flera olika skeden. Ofta har de varit medlemmar av bibelkommissionen, men stundom har de lockats att ge sig i kast med texterna ändå, kanske stimulerade också av den litterära utmaningen. Utöver dem som redan nämnts kan man för 1800-tallets vidkommande framhålla Johan Henrik Thomander och Hans Magnus Melin, som båda fick mycket goda vitsord för sina insatser. Thomander utgav en översättning av Nya testamentet 1835 och Melin en bibelöversättning 1858–65.

Bibelkommissionens arbete gick vidare, nu med Melin som alltmer tongivande. En annan ledamot av Akademien, Viktor

Rydberg, publicerade 1873 en uppsats, vari han kritiskt granskade den samma år utgivna versionen av Nya testamentet. I uppsatsens andra del med den kända rubriken "Tysk eller nordisk svenska?" vidgade han perspektivet till att gälla en allmän diskussion av svenskans förhållande till främmande språk.

Även om Viktor Rydberg sköt över målet, fick hans kritik betydelse för kommissionens arbete. En reviderad version av Nya testamentet kom 1877, och 1883 var det dags för den s.k. normalupplagan. Den finns självfallet också i bakgrunden för Tegnér i hans arbete.

[Begynnelsen af Jesu Kristi, Guds Sons, evangelium] såsom det är skrifvet hos profeterna: "Se, jag sänder din ängel före ditt ansigte, hvilken skall bereda din väg framför dig; en ropandes röst i öknen: bereden Herrens väg, gören hans stigar räta!" Johannes uppträdde döpande i öknen och predikande bättringens dop till syndernas förlåtelse. Och till honom utgingo hela Judiska landet och Jerusalemens invånare, och de läto alla döpa sig af honom i floden Jordan, bekännande sina synder.

Bibeln 1917 upphörde inte att liva andarna. H.S. Nybergs inlägg är redan nämnda. Tidigt på 1960-talet skrev Viveka Heyman en artikel i Dagens Nyheter om vad hon kallade den censurerade Höga Visan. Hon vände sig däri mot okänslighet och missförstånd i 1917 års översättning. I intervjuer apropå artikeln uttalade sig Eyvind Johnson, Gunnar Ekelöf, Erik Lindegren och Östen Sjöstrand, alla kritiskt. Flera sade sig använda äldre översättningar, bland dem Eyvind Johnson som anvisade Karl XII:s bibel. Ekelöf uppgav, att han inte hade läst 1917 års översättning. Sjöstrand betecknade den som rent dålig. Den ende som sade något uppskattande om den var Lindegren, som påpekade att man hade bemödat sig om en mera svensk ordföljd än tidigare. Han ville emellertid gärna se en ny översättning och beklagade samtidigt, att inte "gamle Stiernhielm" hade gjort ett försök.

Några år senare tillkallades Karl Ragnar Gierow och den blivande Akademiledamoten Carl Ivar Ståhle som sakkunniga i 1963 års bibelkommittés utredning om behovet av en ny översättning av Nya testamentet. Gierow skrev en personlig bilaga till betänkandet under rubriken "Bibelns språk och tidens". Ståhle



bidrog bland annat med en längre uppsats med titeln "Några drag i det svenska bibelspråkets historia 1526–1917". Huvudsynpunkten hos Ståhle är, att vi vet att språket hela tiden utvecklas men att vi inte kan förutsäga hur. Någon hållbar språkdräkt för bibeln kan därför aldrig garanteras. Jag delar helt hans uppfattning.

I sitt remissvar anslöt sig Akademien till tanken på en ny översättning av Nya testamentet. Man accepterade också målet för en sådan: en idiomatisk översättning som samtidigt låg nära originalet. Däremot avvisade man förslaget om två auktoriserade översättningar, en kyrkobibel för gudstjänstbruk och en mera lättillgänglig folkbibel.

Till det som gör arbetet i Svenska Akademien så stimulerande hör de livliga diskussionerna med goda argument från olika håll. Det är en gammal tradition inom Akademien att inte försitta en möjlighet till en väl turnerad replik i åsiktsutbytet. Karl Ragnar Gierows dekonstruktion av det betänkande som lades fram av 1971 års bibelsakkunniga återges i del 2 av Akademiens bidrag.

Akademien yttrade sig senare om ett förslag till ny psalmbok. Vi tog bland annat upp de ofta hårdhänta bearbetningarna av äldre psalmer och framhöll, att det ingalunda är givet att en större pietèt skulle motverka psalmbokens syfte. Särskilt borde sådana texter som har en stark ställning i traditionen kvarstå oförändrade. En naturlig och effektiv lösning vore diskreta ordförklaringar nertill på sidorna. På så sätt skulle större hänsyn till texterna och deras upphovsmän förenas med gemenskap över tiden och andlig stimulans. Synpunkterna fick ett visst genomslag men blev inte tillräckligt beaktade i 1986 års psalmbok.

På liknande sätt betraktar jag serien av revisioner av bibeln. Allt i en religiös text behöver säkerligen inte omedelbart förstås. Om det hade varit fråga om ett rent intellektuellt förhållande till texten, hade denna synpunkt fallit. Men det rör sig ju också om en emotionellt, traditionellt och litterärt betingad upplevelse. Det är tydligt att den motverkas, när texten trivialiseras och banden bakåt klipps av. Mot den bakgrunden finns det fog för dämpad entusiasm inför 1981 års översättning av Nya testamentet.

Som det står skrivet hos profeten Jesaja: *Se, jag sänder min budbärare före dig, han skall bereda vägen för dig. En röst ropar i öknen: Bana väg för Herren, gör hans stigar raka* – så uppträdde Johannes döparen ute i öknen och förkunnade syndernas förlåtelse genom omvändelse och dop. Och hela Judeen och alla i Jerusalem kom dit ut till honom, och de bekände sina synder och döptes av honom i Jordan.

Varför skall man störa kontakten med Franzéns psalm genom att skriva *Bana väg för Herren*?

I sitt direktörstal 1981 tryckte Erik Lönnroth just på översättningens traditionsbrytande karaktär. Den sakrala verkan var som bortblåst, konstaterade han, medan texten däremot hade fått en möjlighet att nå den stora massan. Lönnroth tangerar här den ständigt aktuella frågan om kyrkobibel kontra folkbibel. I del 3 av vårt bidrag tar Gunnel Vallquist upp detta spörsmål mot bakgrund av en rad iakttagelser i 1981 års översättning.

Till sist: den jag för min del sätter högst av översättningar till germanska språk är den som ombesörjdes av Wulfila, den lilla vargen. Hans gotiska bibel från 300-talet har en kraftfull enkelhet i stilen som är efterföljansvärd. Låt oss inte väja för språkformen (läs *ai* som *e*, *au* som *o*, *ei* som *i* och *þ* som läspljud):

swe gameliþ ist in Esaïin praufetau: saf, ik insandja aggilu meinana faura þus, saei gamanweiþ wig þeinana faura þus. Stibna wopjandins in auþidai: manweiþ wig frauþins, raifhtos waúrkeiþ staigos gudis unsaris.

En tänkbar invändning vore, att man inför en text som denna kan hamna i den anakronistiska fällan. Jag tar risken.

## Från grafem till lexem

Forskning i ett föränderligt samhälle. Stiftelsen Riksbankens Jubileumsfond 1965–1990. 1990.

När Stiftelsen Riksbankens Jubileumsfond firar sitt 25-årsjubileum, kan Institutionen för språkvetenskaplig databehandling vid Göteborgs Universitet (Språkdata) räkna in 24 år. Dess ursprung, Forskningsgruppen för modern svenska, bildades nämligen 1966 – tack vare ett projektanslag från den nybildade Jubileumsfonden. Projektet hette Datamaskinell undersökning av tidningsprosa.

Åren kring 1960 var vi några nordister på olika håll i landet, som började orientera oss mot nutidens språk och mot teoretiska och metodologiska frågor i ett mera internationellt perspektiv. Vi hade fått vår utbildning i en kraftfull filologisk tradition och arbetat inom den med stor behållning. Själv var jag sysselsatt med stimulerande textfilologiska problem för min avhandling om Johan Ekeblads brev från mitten av 1600-talet. Nya signaler kändes emellertid också lockande. Dessutom hade ett arbetsredskap av principiellt nytt slag, datamaskinen, dykt upp. Det hade visat sig vara en stor tillgång i avhandlingsarbetet och borde prövas närmare.

Det fanns många områden att undersöka. Av dessa tilldrog sig fonologin och morfologin stor uppmärksamhet internationellt sett. Syntaxen behandlades ännu sparsamt. Grafonomin förekom bara sporadiskt. Semantiken och lexikologin (inte lexikografen) låg nästan i träda.

Medan jag ännu höll på med min avhandling, skisserade jag ett projekt, i vilket jag kombinerade det (filologiska) ordintresset, de grafonomiska studierna, intresset för det levande språket och nyfikenheten på datamaskinens möjligheter som hjälpmedel för språkvetenskaplig forskning. Det var det som blev projektet Datamaskinell undersökning av tidningsprosa.

De frågor som ställdes var följande. Vilka är de i vid mening lexikaliska enheterna i svenskan, representerad av en stor kor-

pus? Hur vanliga är de, och hur fördelar de sig över olika slags texter? Resultaten skulle i första hand läggas fram i en frekvensordbok.

Materialet hämtades ur fem morgontidningar, vilkas språk kunde karakteriseras som standardspråkligt. För den typ av svenska det gällde använde jag beteckningen bruksprosa. Tidningarna representerade tre regioner, nämligen Stockholm, Göteborg och Malmö. Texterna täckte hela år 1965 enligt ett särskilt utarbetat schema. Totalt omfattade de en miljon löpande ord.

Från Statens humanistiska forskningsråd hade jag tidigare fått ett anslag för att samla in materialet. Texterna levererades i form av hålremsor från tidningarnas sätter. Därigenom kunde man komma ifrån den kostsamma och tidsödande inkodningen via hålremsstansar. Nu för tiden kan man förstås komma undan ännu bekvämare. Det fanns också en del maskinprogram från de närmast föregående åren, som kunde utnyttjas i projektets första skede.

## Nusvensk frekvensordbok 1–2: ord och åter ord

De kvalitativa frågorna, framför allt utformningen av kriterier för de språkliga undersökningsenheterna, kom att ta en väsentlig del av tid och krafter i anspråk. Den strategi som utvecklades för projektet återges i förenklad form i figur 1.

Beskrivningsmodellen tog sin utgångspunkt i den teckenföljd som bildade texten. Denna segmenterades i sådana strängar som stod mellan två konsekutiva blanka positioner. Kriteriet kan betecknas som syntagmatisk kontrast. De erhållna textorden befriades från interpunktionstecken m.m. (*lever!* och *lever*) normaliserades till *LEVER*), distinktionen mellan versaler och gemena neutraliserades (*lever* och *Lever* blev också *LEVER*) och avstavade ord sattes samman (*le- ver* och *lev- er* till *LEVER*). Vi betecknade enheterna som grafiska ord eller graford. Som resultat av segmenteringen (och den nämnda normaliseringen) utgjorde de den första beskrivningsnivån.

Huvudmoment	Beskrivningsnivåer, huvudkriterier, ordboksdelar		
	1 Ord	2 Ordförbindelser	3 Ordled
I Segmentering	<i>Graford</i> Syntagmatisk kontrast /NFO 1/	<i>Kombinationer</i> Rekurrens /NFO 3/	<i>Grafiska morfer</i> Kommutation /NFO 4/
II Separering	<i>Homograf-</i> <i>komponenter</i> Grammatik /NFO 1/	<i>Konstruktioner</i>  Grammatik /NFO 3/	<i>Homomorf-</i> <i>komponenter</i> Grammatik /NFO 4/
III Klassificering	<i>Lemman</i> Grammatik /NFO 2/	–	<i>Formella morfem</i> Grammatik /NFO 4/
IV Differentiering	–	<i>Idiom</i>  Semantik /NFO 3/	<i>Polysem-</i> <i>komponenter</i> Semantik /NFO 4/

FIGUR 1. *Projektstrategi.*

Den andra ordnivån uppnåddes genom analys av den grammatiska mångtydigheten, homografin. Härigenom kunde till exempel singularen *koppar* skiljas från det *koppar* som är plural av *kopp* och det *koppar* som är presens av *koppa*. På motsvarande sätt separerades singularen *basar* (med betoningen på andra stavelsen) från pluralen av *bas* och presens av *basa* liksom adverbet *åt* från prepositionen *åt* och verbformen *åt*. I fallet *lever* kom substantivet *lever* att stå mot presensformen *lever*.

Enheterna på denna nivå betecknades som homografkomponenter. Analysen utfördes som en genomgång belägg för belägg av en maskinellt framtagen konkordans till hela materialet. Dock undersöktes ett sampel när det gällde beläggen på var och en av de 24 vanligaste homograferna, vilket i sin tur ledde till skattning av frekvenser och beräkning av konfidensintervall.

De båda nu antydda beskrivningsnivåerna presenterades tillsammans med en mängd kvantitativa uppgifter i första delen av Nusvensk frekvensordbok baserad på tidningstext (NFO 1, 1970). Som en detalj kan också nämnas, att grafemens frekven-

ser gavs i ett av appendixen. Det visade sig, att böjningsmorfologens element – e a n t r s – dominerade bilden stort. De svarade tillsammans för 51 % av alla grafemförekomster i texten.

Homografin är allmänt sett en av språkets mest karakteristiska egenskaper. Företeelsen är så utbredd, att (med vår analys) två ord av tre i en löpande text är homografa. Här möter allt från de uppenbara fallen som *lever* till de mera sökta men principiellt existerande som *leverans*. Detta exempel hänger inte bara ihop med *leverera* utan också (annorlunda uttalat) med *lever* och *ans* i den hittills säkerligen aldrig skådade innebörden 'vård av lever'.

De kvantitativa uppgifterna är framför allt av två slag: frekvens och spridning. Grundläggande är den observerade frekvensen  $F$ , dvs. antalet förekomster i det studerade materialet. Eftersom ett och samma frekvenstal kan återspegla helt olika fördelningar på deltexter, är det viktigt att också ge mått på enheternas spridning.

Vi valde två slags spridningsmått, betecknade som dispersion och kontribution. Dispersionsvärdet  $D$ , som bygger på antalet deltexter, medelvärdet och standardavvikelsen, kan variera mellan 1 för perfekt fördelning över deltexterna och 0 för belägg i en enda deltext. Som deltexter betraktades dels de fem tidningar som hade bidragit till korpusen, dels de sex ämnessfärer som materialet hade indelats i. Med kontribution betecknade vi ett grövre spridningsmått, som angav vilka och därmed hur många deltexter som hade bidragit till en viss enhets frekvens.

En sammanvägning av frekvens och spridning företogs på så sätt, att den observerade frekvensen multiplicerades med dispersionen över tidningarna. Det resulterande värdet benämndes modifierad frekvens  $F_{\text{mod}}$ . De enheter som hade ett  $F_{\text{mod}}$ -värde på minst 4,125 benämndes basvokabulär och presenterades särskilt ingående. Det angivna minimivärdet var det som erhöles, om en enhet hade ett belägg i varje tidning.

Genom homografsepareringen ökar antalet observerade enheter. Det är emellertid självfallet av stort intresse att sammanföra böjningsformer och variantformer under gemensamma

grundformer, lemman. De båda *lever* går till var sitt lemma: substantivet *lever* respektive verbet *leva*. Under det senare samlas *lev, leva, levande, levat, levde, leve, lever, levt* m.fl. Flest belagda former, 24 stycken, visade sig *ge* ha: *gav, gavs, ger, ges* (infinitiv och presens), *gett* (supinum), *giv, giva, givande, give, given, gives, givet, givit, givna* (perfekt particip neut-rum, utrum och pluralis), *gåve* m.fl. Denna procedur genomfördes till sin huvuddel automatiskt på grundval av en genomgång av svenskans morfologi. Programmet förde hela 97 % av ord-formerna till rätt lemma. Detta var naturligtvis till mycket stor hjälp, men resultatet fick självfallet ändå granskas i sin helhet, eftersom det gällde att lokalisera felen (som när *flygmajor* hade avlästs som pluralis och förts till ett nyskapat *flygmaja*). De en miljon löpande orden hade gett oss 103 000 olika graford, som hade uppdelats i 112 000 olika homograf-komponenter, som sammanfördes i 71 000 olika lemman. Resultaten på lemmanivån presenterades i NFO 2 (1971).

I figur 2 ges ett kort utdrag ur listan Alfabetiskt ordnad basvokabulär. Av förkortningarna betyder Onr ordningsnummer (rangnummer baserat på Fmod), DT och DÄ dispersion över tidningar respektive ämnessfärer, Tidn och Ämn kontribution från tidningar respektive ämnessfärer och Sub antalet belagda sublemmaformer (böjnings- och variantformer).

Onr	Enhet	Fmod	F	DT	DÄ	Tidn	Ämn	Sub
4289	kärnvapen <i>nn -et</i>	12,48	18	,693	,174	1-5	123	3
8178	kärra' <i>nn -n</i>	4,338	6	,723	,171	1345	56	2
4335	kärv <i>av -t</i>	12,31	17	,724	,627	1235	1-6	6
3076	kö <i>nn -n</i>	20,79	25	,832	,487	1-5	1345	5
7372	köa <i>vb -ad</i>	5,149	10	,515	,281	235	35	6
2071	kök <i>nn -et</i>	36,26	79	,459	,306	1-5	2356	6
6648	kölvatten <i>nn -et</i>	6,098	9	,678	,534	1235	235	2
3749	kön <i>nn -et</i>	15,51	25	,620	,803	1-5	1-6	5
7840	könsroll <i>nn -en</i>	4,661	10	,466	,539	1234	1236	4
2868	köp <i>nn -et</i>	23,00	27	,852	,732	1-5	1-6	2
357	köpa <i>vb -te</i>	234,6	259	,906	,666	1-5	1-6	12
3083	köpare <i>nn -n</i>	20,73	28	,740	,298	1-5	12345	5
1479	Köpenhamn <i>pm -s</i>	56,40	72	,783	,780	1-5	1-6	2

FIGUR 2. Ur Alfabetiskt ordnad basvokabulär (NFO 2).

Ett urval ur materialet i NFO 1–2 publicerades i boken Tio-tusen i topp (1972). Den innehåller också vissa specialbearbetningar.

### Nusvensk frekvensordbok 3: ordförbindelser

Ord fungerar sällan ensamma i kommunikationen. Många ord är mycket hårt bundna till andra. Ett typiskt sådant är *del*: *en del av förklaringen till, få del av, ta del av, en betydande del av, en hel del annat, för en del år sedan, för min del, i denna del av världen* osv. I nästa skede av projektet riktade vi in oss just på ordförbindelserna (eller kollokationerna).

Den metodologiska grundidén var att vaska fram de ordförbindelser, som återkom gång efter annan i korpusen. Bland dem borde en stor mängd lexikaliskt intressanta enheter finnas. Detta att uppträda med minst två identiska belägg i materialet fick benämningen rekurrens. Sekvenser som uppfyllde kravet benämndes kombinationer. Det visade sig att rekurrenskriteriet bar frukt på ett överväldigande sätt. Vi fick 660 000 belägg på rekurrenta kombinationer, från fraseologiskt relevanta som de ovan nämnda till fraseologiskt irrelevanta som *det inte att*.

Det gällde nu att skilja agnarna från vetet. De fraseologiskt relevanta ordförbindelserna är på en gång lexikaliskt selekterade och grammatiskt styrda (välformade). De grammatiska urvalskriterier som vi uppställde kan grupperas i fyra kategorier: nominalfraser, verbalfraaser, konnektiver och satser. Sjutton typer utkristalliserades: *den stora frågan, i linje med, ta form, sköta om, kommer att fortsätta, för att undersöka, är på väg, mycket ung, även om, om också, för att, in i, men låt oss inte gå händelserna i förväg, det är givet att, kort sagt, som människa, jo då*.

Den fortsatta bearbetningen av konstruktionerna innefattade ganska komplicerade moment, bland annat beroende på enheternas överlappning. I den återkommande förbindelsen *umgås med sina vänner* ingår de båda var för sig återkommande förbindelserna *umgås med* och *med sina vänner*. Detta måste



redas ut och ges en genomskinlig presentationsform. En annan av de faktorer som krävde speciell uppmärksamhet var de diskontinuerliga förbindelserna, sådana som *ta (x) ställning* och *på (x) sätt* i exempelvis *på ett eller annat sätt*. Vi beslöt att redovisa de rekurrenta fallen av dessa förbindelser i deras fullständiga form.

Analysen på denna beskrivningsnivå gav 50 000 olika konstruktioner med 180 000 belägg som s.k. maximala enheter. Ur denna stora mängd letade vi fram idiomerna, de konstruktioner vilkas betydelser inte kan förutsägas med utgångspunkt i de ingående ordens betydelser. Ett typexempel är *i elfte timmen*. Vi tillämpade kriteriet restriktivt och fick fram 300 enheter med 1700 belägg. Ordförbindelserna av typerna kombinationer, konstruktioner och idiom lades fram i NFO 3 (1975).

Figur 3 visar ett exempel från listan Lexikon per i enheterna ingående ord. Här står F för total observerad frekvens, Fx för frekvens som maximal enhet och För för antal författare som använt konstruktionen i fråga (ett slags kontributionsuppgift). Det sammanlagda antalet författare var 569. Uppgiften 258/563 innebär att 258 av de sammanlagt 563 beläggen på *ligger* i texten är knutna till de här förtecknade konstruktionerna. Kvoten 46 % benämner vi konstruktionsbenägenhet.

Enhet	F	Fx	För	Enhet	F	Fx	För
<b>LIGGER 258/563</b>				<i>ligger</i> inte i	2	2	2
<i>ligger</i> bakom* ab	5	3	5	<i>ligger</i> kanske	2	2	2
<i>ligger</i> bakom* pp	14	14	13	<i>ligger</i> kring	2	2	2
<i>ligger</i> bra till	3	3	3	<i>ligger</i> mellan* pp	7	7	7
<i>ligger</i> dock	5	5	5	<i>ligger</i> nu	2	2	2
<i>ligger</i> där	3	3	3	<i>ligger</i> nära	8	2	8
<i>ligger</i> däremot	3	3	2	<i>ligger</i> nära till hands	6	2	6
<i>ligger</i> emellertid	3	1	3	<i>ligger</i> närmare	4	4	4
<i>ligger</i> emellertid inte	2	2	2	<i>ligger</i> närmast	3	3	3
<i>ligger</i> för* pp	3	3	3	<i>ligger</i> också	11	11	11
<i>ligger</i> för närvarande	3	1	3	<i>ligger</i> oftast	2	2	2
<i>ligger</i> för närvarande hos	2	2	2	<i>ligger</i> omgivet av	2	2	2
<i>ligger</i> före* pp	2	2	2	<i>ligger</i> på* ab	1	1	1
<i>ligger</i> givetvis	2	2	2	<i>ligger</i> på* pp	39	35	38
<i>ligger</i> hela tiden	2	2	2	<i>ligger</i> på ett annat plan	2	2	2
<i>ligger</i> helt enkelt	2	2	1	<i>ligger</i> på gränsen till	2	2	2
<i>ligger</i> helt i linje m:d	3	3	3	<i>ligger</i> som	3	3	2
<i>ligger</i> hos	4	4	4	<i>ligg</i> . sill grund för	4	4	4
<i>ligger</i> i* pp	69	56	65	<i>ligger</i> under* pp	7	7	7
<i>ligger</i> i att	7	7	7	<i>ligger</i> utanför* pp	3	3	3
<i>ligger</i> i dag	2	2	2	<i>ligger</i> vid* pp	7	7	7
<i>ligger</i> i linje med	2	2	2	<i>ligger</i> väl framme	2	2	2
<i>ligger</i> i sakens natu.	4	4	4	<i>ligger</i> väl till för	2	2	2
<i>ligger</i> inom	5	5	4	<i>ligger</i> öster om	2	2	2
<i>ligger</i> inte	9	5	9	<i>ligger</i> över* pp	3	3	3
<i>ligger</i> inte heller	2	2	2	det <i>ligger</i> nära till hands att	4	4	4
				som <i>ligger</i> bakom* ab	2	2	2

FIGUR 3. Ur Lexikon per i enheterna ingående ord (NFO 3).

Som framgår av figuren utfördes ingen lemmatisering av konstruktionerna. Detta hänger samman med att enheterna inte längre är enskilda ord utan ordförbindelser. Det skulle vara opraktiskt att exempelvis reducera *tillfälligheternas spel* till *tillfällighet spel* och missvisande att sammanföra *i tiden* och *i tid*.

Ett bestämt intryck av undersökningen av kollokationerna var, att den oerhörda mängden rekurrenta konstruktioner representerar en fundamental del av den mänskliga språkförmågan. Det blir nödvändigt att tänka sig att en mycket stor mängd färdiga ordförbindelser ingår i det mentala lexikon, som språkbrukarna förfogar över. Jag har i annat sammanhang betecknat detta fenomen som språkets ostinato i motsats till språkets capriccio, den kreativa aspekten. En psykologiskt trovärdig modell för språklig produktion och perception måste i hög grad beakta detta. Det har också visat sig bli en alltmer intressant faktor vid utvecklingen av exempelvis system för maskinöversättning.

#### **Nusvensk frekvensordbok 4: ordled**

I NFO-projektets sista skede riktades intresset mot ordens inre struktur. Osammansatta ords uppträdande som självständiga enheter och deras samspel med andra i ordförbindelserna borde kompletteras med deras roll som ordled. Detta skulle ge en betydligt fylligare bild av enheternas faktiska användning i språket.

Som material för analysen valdes lemmanas uppslagsformer i NFO 2 med vissa uteslutningar sådana som utländska enheter, proprier (egennamn), förkortningar och siffertal. Antalet analyserade lemman blev härigenom 55 000. Deras frekvens i texten var så stor som 930 000, dvs. de täckte 93 % av korpusen.

Lemmanas uppslagsformer konstitueras av ett litet antal olika slag av ordled: prefix, grundmorf, ordbildningssuffix, fog och i enstaka fall ordböjningssuffix (nämligen i ord av typen *flerveckorsskägg*). För att finna gränserna mellan dem utnytt-

jade vi med vissa inskränkningar s.k. kommutation, d.v.s. serier som *chefsdomare* – *lärdomsaptit* – *ungdomlig* – *folksjukdom* – *domstol* – *dom*. Segmenteringen skedde interaktivt med ett adaptivt program, ett program som tog fasta på och utnyttjade beslut som fattats under analysens gång. Det var grundat på en modell för tillåtna teckenkombinationer (grafotax) och hade förmågan att approximera stavelser. Analysen gav 7800 grafiska morfer med en textuell frekvens av 1 652 000.

Separeringen i homomorfkomponenter, som i praktiken genomfördes i nära anslutning till segmenteringen, gällde de olika ordledstyperna grundmorfer, suffix osv. Av *dom* fick vi en grundmorf och ett substantivbildande suffix som i *sjukdom*. Både *lever* och *lev* blev givetvis grundmorfer. Totalt framkom 7900 olika grundmorfer och 560 olika affix. Av dessa senare var 460 avledningssuffix, 240 av dem substantiviska. Växlingen mellan prefix, grundmorfer, fogar och suffix kan belysas med det ganska lättmanövrerade exemplet *o lyck s fall s för säkr ing s av del ning*.

Materialet för klassificeringen i formella morfem utgjordes av grundmorferna. Som kriterier utnyttjades olika slag av variation. Exempel på några typer är *elv(a)/elf(te)*, *fabel/fabul(ös)*, *ess/äss*, *simm(a)/sim(ning)*, *(upp)bär(a)/(bil)bur(en)*. Bearbetningen gav 5300 olika enheter med 6100 varianter. Liksom de båda föregående nivåerna redovisades denna i NFO 4 (1980).

Den sista beskrivningsnivån, den mest intressanta när det gällde ordled, uppnåddes genom att vi utförde en semantisk differentiering av de formella morfem, som framstod som polysema. För att en uppdelning skulle komma till stånd skulle den urskilda huvudbetydelsen (kärnbetydelsen) vara möjlig att parafrasera och att särskilja från andra huvudbetydelser grundade på samma formella morfem. I ett typfall som *tipp* fick vi på detta sätt tre polysemkomponenter: *tipp* 1 'spets' som i *svanstipp*, *tipp* 2 'gissa' som i *topptippad*, *tipp* 3 'stjälpa ut' som i *slamtippning*. I huvudlistan i NFO 4 förtecknades alla de 8300 framkomna enheterna med alla sina belägg i lemmanas uppslagsformer. Listan utgör i själva verket en avancerad

semantisk ordledskonkordans. Apropå antalet enheter kan nämnas, att ett ordinärt kinesiskt lexikon innehåller omkring 10 000 (ord)tecken. Likheten i siffertalen är nog inte bara en tillfällighet.

Några rader ur den nämnda listan återges i figur 4. Ftext avser observerad frekvens i texten och Flex antalet olika lemmar som enheten förekommer i.

I den starkt schematiska översikt av de olika momenten, kriterierna och beskrivningsnivåerna, som jag nu har lämnat, har de tusentals enskildheterna i språkets myllrande liv, de ständiga överraskningarna, de återkommande modelljusteringarna och den stadigt växande mängden av tillämpningskonventioner knappast ens kunnat antydast. De var vårt dagliga bröd.

Enhet och kontext	Ftext	Flex	Enhet och kontext	Ftext	Flex
lev	1386	141			
lev 'åga organisk existens'	651	55	lev nad s reg el nn -er	1	
lev a vb -de	454		lev nad s standard nn -en	16	
lev a nde ab	1		lev nad s sätt nn -et	2	
lev a nde nn -t	1		lev nad s teckn ing nn -en	4	
lev a nde gör a vb -de	6		lev nad s trött av -t	1	
lev e bröd nn -et	10		lev nad s van a nn -n	6	
lev e rop nn -et	1		lev nad s vill kor nn -et	3	
lev erne nn -t	5		lev nad s vis dom nn -en	2	
lev nad nn -en	3		lev nad s ålder nn -n	1	
lev nad s ban a nn -n	3		lev nad s öd e nn -t	2	
lev nad s be skriv ning nn			efter lev a vb -de	3	
-en	1		efter lev nad nn -en	5	
lev nad s bild nn -en	1		fort lev a vb -de	4	
lev nad s för håll a nde nn			fram lev a vb -de	2	
-t	5		in lev else nn -n	22	
lev nad s glad av -t	2		in lev else för måg a nn -n	1	
lev nad s konst nn -en	5		kvar lev a nn -n	9	
lev nad s konst när nn -en	4		kvar lev a vb -de	1	
lev nad s kost nad nn -en	1		ren lev nad s man nn -en	1	
lev nad s kost nad s in			sam lev nad nn -en	7	
dex nn -et	1		sam lev nad s norm nn -en	1	
lev nad s lopp nn -et	1		träd lev a nde av -ø	2	
lev nad s lyck a nn -n	1		väl lev nad s man nn -en	1	
lev nad s minn e nn -t	1		öken lev a nde av -ø	1	
lev nad s mönster nn -et	1		över lev a vb -de	35	
lev nad s niv å nn -n	1		gen om lev a vb -de	2	
lev nad s norm nn -en	1		halv ut lev d av -t	1	
lev nad s om kost nad nn			liv s lev a nde av -ø	1	
-en	1		samm an lev nad nn -en	1	
lev nad s om stånd ig			drif t s ut lev else nn -n	1	
het nn -en	1				

FIGUR 4. Ur *Lexikon med kontexter* (NFO 4).

De fyra volymerna av NFO redovisar på sammanlagt 4400 sidor drygt 150 olika bearbetningar på olika beskrivningsnivåer med allt efter omständigheterna varierande typer av kvantita-

tiva uppgifter. Vi känner inte till någon motsvarighet på något annat språk.

Forskningsresultaten publicerades i den serie, *Data linguistica* (Almqvist & Wiksell International), som grundades i och med utgivningen av NFO 1 1970. Serien är också ett allmänt forum för arbeten på området och omfattar dels undersökningar av språkligt material (svenska och utländska), dels teoretiskt och metodologiskt inriktade studier. Hittills har arton volymer utkommit.

### **Institution**

Verksamheten fick litet fastare form redan 1970 genom att forskningsrådet då tillskapade en forskartjänst i språklig data-teknik. Två år senare inrättades i stället en professur i språklig databehandling vid rådet. I samband därmed bildades Avdelningen för språklig databehandling vid Institutionen för nordiska språk vid Göteborgs universitet, och forskningsgruppen kom nu att arbeta inom dess ram. Gruppen upplöstes formellt 1976, då anslagen från Jubileumsfonden upphörde efter att under hela tio år ha varit en fast grund. Arbetslaget arbetade emellertid vidare. Andra projekt tillkom parallellt, bland annat Algoritmisk textanalys, som framför allt var inriktat på automatisk morfologisk analys; det av Jubileumsfonden stödda Syntaktisk taggning (SynTag), som gällde syntaktisk analys och märkning av en del av NFO-korpusen; Svenska namn, som gav underlag till Förnamnsboken och andra bearbetningar; O.S.A., som avser inläsning, lagring och strukturering av Svenska Akademiens ordbok; Lexikon för invandrare (Lexin); Lexikalisk databas/Svensk ordbok; En lexikonorienterad parser för svenska; Translator's Aid (Traid); de båda senare i samarbete med Anna Sågvall Hein.

Efterfrågan från olika håll på språkliga data växte hela tiden. Framställningar till myndigheterna om behovet av ett serviceorgan ledde till att språkbanken Logoteket med två fasta tjänster kunde upprättas 1975. Den har genom åren försett många

mottagare inom forskning, språkvård och näringsliv med uppgifter. Inte minst har det gällt den grafiska industrin. Det finns material och bearbetningar att ta av: text till ett omfång av över 30 miljoner löpande ord ur tidningar, romaner, riksdagsprotokoll, författningssamlingar m.m. och stora lexikaliska bearbetningar av dessa material. Dessutom lagras resultaten av genomförda projekt. Här kan också nämnas, att föreståndaren Martin Gellerstam är redaktör för Svenska Akademiens ordlista och att regelbundna körningar görs för den pågående utgivningen av Strindbergs samlade verk.

En studieplan för forskarutbildning i ämnet som ett alternativ inom allmän språkvetenskap fastställdes 1973, och konsolideringen av området fortsatte både nationellt och internationellt. År 1977 omvandlades avdelningen till en självständig institution. I samband därmed fastställdes en ny studieplan för forskarutbildning i ämnet som ett självständigt område. Två år efter institutionsbildningen överfördes professuren till Göteborgs universitet. I samband därmed preciserades ämnet till språkvetenskaplig databehandling, och den benämningen kom också att ingå i institutionens namn. Sedan lång tid hade kortnamnet Språkdata använts, och det passade också för det nya institutionsnamnet. Idag är ett tjugotal personer engagerade vid institutionen. I samband med att jag valdes till Svenska Akademiens ständige sekreterare 1986 fick jag tjänstledigt med 75 % från professuren. Min vikarie fram till utgången av 1989 var Anna Sågvall Hein. Våren 1990 vikarierade Martin Gellerstam.

Ämnesbeteckningen *språkvetenskaplig databehandling* tillkom, innan beteckningen *datalogvistik* hade etablerats i svenskan. Båda svarar mot den engelska beteckningen *computational linguistics*. Engelskans *natural-language processing* har en vidare innebörd och ligger närmare *språklig databehandling*. Såsom variant av ämnesbeteckningen *datalogvistik* förekommer *datorlingvistik*, som ibland används med särskild tanke på den gren inom datalingvistik, som gäller simulering av språkliga processer. Datalogvistik i stort omfattar bland annat också datorstödd upparbetning av den lingvistiska kom-

petens – inte minst lexikalisk – och den empiriska kunskap, som krävs för framgångsrik utveckling av system för analys, syntes, dialog och översättning. De olika aktiviteterna övergår ofta i varandra. Genom beteckningen *språkvetenskaplig databehandling* markeras, att det är fråga om ett språkvetenskapligt och därmed humanistiskt ämne.

Av dem som har fått sin vetenskapliga träning genom projektarbete i kretsen är det lika många som har lagt fram sina avhandlingar i annat ämne som i institutionens ämne. Denna konsekvens av den tvärvetenskapliga inriktningen har jag aldrig vänt mig mot, eftersom jag menar att resultaten under alla omständigheter har kommit den humanistiska fakulteten till godo.

Grundutbildningen i ämnet hade länge karaktären av orienteringskurser, senare s.k. enstaka kurser. År 1984 inrättades emellertid datalingvistiklinjen, en fyraårig utbildning som ger tillträde till forskarutbildningen. I undervisningen samverkar institutionen med institutionerna för lingvistik, filosofi och informationsbehandling.

## **Svensk ordbok**

Efter NFO, där särskilt i fjärde delen betydelseförhållandena spelade en viktig roll och främst då skillnaderna mellan ordleds olika betydelser, var det naturligt att ställa den fråga, som hela tiden hade svävat över vattnen: Vad betyder orden närmare besett? Tiden föreföll vara mogen för en semantisk framstöt mot de ord, som vi hade behandlat på så många kvalitativa och kvantitativa sätt. Det blev projektet Lexikalisk databas/Svensk ordbok.

Från en synpunkt sett svarar detta nya projekt mot en tom plats i strategifiguren, nämligen differentiering på ordnivå. Emellertid rörde det sig inte om någon korpusstudie, utan materialet hämtades från många olika håll. Några kvantitetsuppgifter kom därför inte i fråga. Inte heller var det uteslutande

ordbetydelser som behandlades. Det skulle också ges upplysningar om konstruktioner, idiom och morfem.

Jag skrev projektplanen 1976. Året därpå var vi färdiga med en förstudierapport. Esselte Studium beslöt finansiera huvuddelen av projektet, och arbetet kunde börja 1978. Det pågick till 1986, då Svensk ordbok (SOB) kom ut. I mera begränsad skala har arbetet emellertid fortsatt hela tiden. En ordbok blir aldrig färdig.

SOB har en tyngdpunkt i definitionerna av ordens betydelser. Som uppslagsord valdes lemmarna enligt vår tidigare specifikation. Men lemmat är grundat på ordens formella egenskaper och måste därför kompletteras med en semantisk enhet. Som beteckning för en kärnbetydelse hos ett lemma valdes lexem. Ett lemma kan således innehålla ett eller flera lexem.

För olika typer av ord utvecklades olika definitionsformat. En viktig riktlinje var vidare, att det definierade ordet skulle vara utbytbar mot definitionen i sats sammanhangen. Vi strävade också efter att begränsa och kontrollera (den principiellt ofrånkomliga) cirkulariteten i definitionerna.

Inom lexemen kunde en eller flera betydelseyanser anges. Dessa betraktades som härledda ur kärnbetydelserna genom t.ex. utvidgning eller överföring. De presenterades på samma sätt som kärnbetydelserna genom en cyklisk uppbyggnad av ordboksartiklarna.

Karakteristiskt för allmänspråket är att dess ord har ganska vagt avgränsade betydelser, annorlunda uttryckt att orden är mycket flexibla i språkanvändningen. Detta gäller alltså också sådana facktermer, som har funnit sin väg in i allmänspråket. Man skall inte vänta sig att språkbrukarna använder t.ex. *predikat* eller *export* eller *elektromekanisk* på ett sådant sätt, att det exakt svarar mot användningen på respektive fackområde. Ordboken skall återspegla allmänspråket men kan fästa uppmärksamheten vid särskilt noterbara skillnader gentemot fackspråket genom upplysningar som "ej i vetenskapl. sammanhang" (så vid *stjärna*), "noggrannare def. i mat. sammanhang" (så vid *derivata*) o.d.



SOB presenteras i inledningen som en rapport från ett forskningsfält. Språkets karaktär gör ju, att man ständigt ser nya sammanhang och att man kan finputsas utan ände. Dessutom är språket hela tiden i rörelse. Vi kunde emellertid konstatera, att våra övergripande teoretiska och metodologiska idéer höll. På åtta år hade vi genomfört urvalet av uppslagsord, skrivit 66 000 definitioner av kärnbetydelser, specificerat 25 000 nyanser (utvidgningar m.m.) av dessa, behandlat 4000 idiom, anfört 25 000 indirekt definierade morfologiska exempel (ordbildningsexempel), valt ut 100 000 satsexempel och givit upplysningar om stilvärde och bruklighet, i förekommande fall också fackområde och långivande språk. Resultatet hade redovisats i en volym på 1500 sidor. Denna synnerligen arbetskrävande genomarbetning av ordförrådet på så begränsad tid hade inte varit möjlig utan den myckna erfarenheten från NFO-projektet.

Figur 5 visar artikeln *leva* jämte några partikelverb. Den lilla kvadraten □ markerar betydelsenyans, stjärnan \* idiom och slingan ~ upprepning av hela uppslagsordet respektive den stundom flerordiga uppslagsformen.

Det var tydligt att det fanns ett behov av en ordbok som SOB. Hittills har den sålts i inemot 70 000 exemplar. Vår inbjudan i inledningen till användarna att komma med synpunkter och förslag har fått gensvar. I de successiva tryckningarna har vi kunnat dra nytta av åtskilliga iakttagelser. Dialogen fortsätter.

## Utblick

På vår dåvarande avdelnings initiativ arrangerades nordiska sommarskolor i språklig databehandling under en rad år under 1970- och 1980-talet. Den första kursen höll vi i Marstrand 1972. Sommaren 1977 var kursen förlagd till Palo Alto i Kalifornien, där vi har viktiga kontakter. Genom dessa och andra aktiviteter byggdes ett nordiskt samarbete upp, som har betydtt åtskilligt för ämnesrådets utveckling.

**lev/a** verb *-de, -t el. -at* **1** vara vid liv (jfr liv **1**): överleva; *-er dina föräldrar ännu?*; *hon -de ytterligare två år*; *man -er bara en gång* □ ofta med tonvikt på livsinnehållet: *de -er ett stillsamt liv*; *man ska ~ som man lär*; *de -er under ytterst knappa villkor*; ~ på existensminimum; ~ livets glada dagar □ spec. njuta av livet: *hon hade arbetat hela sitt liv och aldrig haft tid att ~*; *det kändes som om han skulle börja ~ igen* □ spec. äv. ha sin tillvaro förlagd (till) viss plats: *de -er numera på landet*; *att ~ i storstaden* \* få veta att man *-er* bli hårt ansatt; *hur -er världen med dig?* hur har du det? (vard.); ~ för ngn/ngt helt ägna sitt liv åt ngn/ngt; ~ livet utnyttja livets glädjeämnen, roa sig; ~ rullan/loppan festa, svira (vard.) **2** (göra intryck av att) vara fylld av liv och kraft om platser m. m.: *stadens centrum -er igen*; *badorten -er upp på sommaren* □ spec. om konstnärligt verk e. d. ha förmåga att fångsla: *han är död, men hans böcker -er*; *sångerna skall alltid ~ inom oss*; *tyvärr -er inte hans romanfigurer* □ äv. (i tävlingssammanhang) om boll e. d. vara i spel □ äv. om tävling e. d. (alltjämt) vara spännande: *när NN reducerade till 2-1 -de matchen plötsligt igen* **3** med prep. livnära sig på visst sätt: *det är svårt för två att ~ på en lön*; ~ på grönsaker; ~ på räntor; *de hade inget att ~ av* □ äv. överfört, spec. stimuleras (av): ~ på gamla minnen □ spec. äv. ha såsom förutsättning: *romanen -er helt på personskildringarna* \* ~ på luft livnära sig på (nästan) ingenting **4** med

prep. el. partikel ha såsom (nära) umgängespartner: *flera familjer måste ~ tillsammans i det lilla huset* □ särsk. ha såsom sexuell partner: *hon -er ihop med en äldre man*; *de -de tillsammans i många år innan de gifte sig*; *makarna hade -t samman i 30 år* □ äv. med avs. på abstrakta företeelser ständigt sysselsätta sig (med): *jämlikhetsfrågorna har hon -t med i många år* □ spec. med avs. på negativa företeelser uthärda: *mänskligheten måste ~ med kärnvapenhotet* **5** vara livlig (och föra oväsen) (syn. stoja) (vard.): *det var värst vad barnen -er i kväll* □ äv. överfört om naturföreteelse: *vinden -er i trädkronorna* □ spec. om segel fladdra (utan att fyllas av vinden)

**lev/a me'd** verb *-de, -t el. -at* \* ofta med prep. med stor inlevelse höra eller se: *barnen -de helt med i pjäsen*

**lev/a om'** verb *-de, -t el. -at* särsk. föra ett utsvävande liv (syn. svira): *när han fick arvet åkte han till Paris och -de om ordentligt*

**lev/a sig in'** verb *-de, -t el. -at* \* med prep. med stor inlevelse föreställa sig: *han kunde inte ~ i hennes situation*; *en skådespelare som helt -er sig in i sina roller*

**lev/a upp'** verb *-de, -t el. -at* **1** få nya krafter el. nytt liv: *efter framgången -de hon upp en smula*; *den gamla sången -er upp på nytt* **2** göra slut på tillgångar e. d. (syn. förlösa): *han -de upp hela arvet* **3** med prep. till motsvara: *han kunde inte ~ till förväntningarna*

**lev/a u't** verb *-de, -t el. -at* \* ge fritt utlopp för känslor, aggressioner e. d.: ~ *sina drifter*

## FIGUR 5. *Ur Svensk ordbok.*

Bland annat ledde denna nordiska samverkan till att konferensserien Nordiska datalingvistdagar startades. Den första sammankomsten hölls i Göteborg 1977, den senaste i Reykjavík 1989.

En stor upplevelse var det att få arrangera ett Nobelsymposium, till vilket det var möjligt att bjuda in forskare på högsta internationella nivå. Det tema jag hade valt var Text Processing. Symposiet hölls på IBM:s utbildningscentrum på Lidingö, och året var 1980. Sex år senare var det dags för

ytterligare ett Nobelsymposium och det hölls på samma ställe. Den gången handlade det om Possible Worlds in Arts and Sciences.

Association for Literary and Linguistic Computing är en internationell organisation på forskningsområdet. ALLC ger bland annat ut en tidskrift och anordnar konferenser. XIVth International ALLC Conference 1987 förlades till Göteborg. Den konferensserie som går under namnet Coling – International Conference on Computational Linguistics – ger också rika möjligheter till internationella förbindelser.

Ytterligare en kontaktpunkt erbjuder serien Lexicographica (Max Niemeyer Verlag, Tübingen). Jag är medredaktör för den sedan starten 1984. Fram till nu har 29 volymer utkommit.

Under senare år har Europarådet börjat engagera sig på det område, som man kallar *language industries*. Ett viktigt skäl har varit nödvändigheten att slå vakt om Europas språkliga mångfald i en språkligt sett engelskdominerad informationsindustri. Vi tillhör en samarbetsgrupp, som driver ett flerspråkigt datalingvistiskt projekt inom rådets ram.

Ett av de allra största internationella forsknings- och utvecklingsprogrammen på vårt område är EG:s maskinöversättningsprojekt Eurotra med 150 forskare i tolv länder engagerade under lång tid. Vi har fått förtroendet att i två olika skeden utvärdera projektets (data)lingvistiska resultat.

### **Blick framåt**

Tryckoriginalen till SOB genererades programvägen från den lexikaliska databasen. Denna är i sig själv ett synnerligen intressant forskningsobjekt, som öppnar möjligheter för ingående studier av ordförrådets semantiska struktur. De strikta formaten för definitionerna av de olika slagen av ord gör det möjligt att efter analys av själva definitionerna i olika avseenden (disambiguering, lemmatisering, syntaktisk märkning) undersöka exempelvis funktionsklasser och hierarkiska klasser inom vokabulären. Några av medlemmarna i en instrumentell funktions-

klass är för att ta ett exempel *spetta* 'bryta med spett', *skyffla* 'flytta med skyffel' och *kratta* 'jämna med kratta'. Som exempel på de många hierarkiska klasserna kan jag nämna *kasack – blus – plagg – föremål – företeelse – ngt* i stigande abstraktionsordning. Överst tronar alltid i sådana här sammanhang ngt dvs. något. En begreppsordbok eller tesaur, byggd på språkets semantik (och inte på en tentativ världsbeskrivning), hägrar. Den skulle innebära ett fullföljande av en länge driven forskningslinje. Tillämpningarna i olika slags kunskapsbaserade system, vid semantisk parsning (satsanalys), textgenerering och maskinöversättning liksom i andra sammanhang är uppenbara och får efterhand ökad betydelse i institutionens forskningsarbete.

Här ligger forskningsfronten för närvarande. Det har börjat byggas upp likartade projekt på andra håll i världen, dock med något undantag baserade på redan tryckta ordböcker och inte på en egen, dynamiskt modifierbar, lexikalisk databas. En avgörande faktor är, att lexikonet och den lexikaliska semantiken allt tydligare framstår som centrala delar i de allt mer omfattande datalingvistiska system som byggs upp.

Låt mig till sist notera, att det med åren blivit en rätt lång rad likasinnade som jag har haft glädjen att dela forskningsäventyret med. De är i hög grad närvarande i min skiss.

## Referenser

*I. Skrifter i serien Data linguistica. University of Göteborg: Department of Computational Linguistics*

1. Allén, S. (1970). *Nusvensk frekvensordbok baserad på tidningstext. 1. Graford, homografkomponenter.*
2. Allén, S., & Hellberg, S. (1971). *Introduktion i grafonomi, det lingvistiska skriftstudiet.*
3. Thavenius, J. (1971). *Konkordans till Hjalmar Gullbergs lyrik.*

4. Allén, S. (1971). *Nusvensk frekvensordbok baserad på tidningstext. 2. Lemman.*
5. Gavare, R. (1971). *Graph description of linguistic structures. A linguistic approach to the theory of graphs.*
6. Allén, S. (1971). *Tiotusen i topp. Ordfrekvenser i tidningstext.*
7. Ralph, B. (1972). *Introduktion i historisk språkvetenskap.*
8. Sågwall, A-L. (1973). *A system for automatic inflectional analysis implemented for Russian.*
9. Ljung, M. (1974). *A frequency dictionary of English morphemes.*
10. Allén, S. et al. (1975). *Nusvensk frekvensordbok baserad på tidningstext. 3. Ordförbindelser.*
11. Gellerstam, G., & Gellerstam, M. (1977). *Ord och fras i psalmboken.*
12. Berg, S. (1978). *Olika lika ord. Svenskt homograflexikon.*
13. Hellberg, S. (1978). *The morphology of present-day Swedish. Word-inflection, word-formation, basic dictionary.*
14. Allén, S., Berg, S., Järborg, J., Löfström, J., Ralph, B., & Sjögreen, C. (1980). *Nusvensk frekvensordbok, baserad på tidningstext. 4. Ordled, betydelser.*
15. Lindvall, L. (1982). *Jean Renart et Galeran de Bretagne. Étude sur un problème d'attribution de textes.*
16. Allén, S. (Ed.). (1982). *Text processing. Text analysis and generation. Text typology and attribution.* (Proceedings of Nobel Symposium 51.)
17. Engwall, G. (1984). *Vocabulaire du roman français (1962–1968). Dictionnaire des fréquences.*
18. *Studies in computer-aided lexicology* (1988).

## II. Övrig litteratur

- Allén, S. (1965). *Grafematisk analys som grundval för text-  
edering med särskild hänsyn till Johan Ekeblads brev till  
brodern Claes Ekeblad 1639–1655.* (Nordistica Gotho-  
burgensia 1.) Göteborg: Acta Universitatis Gothoburgensis.

- Allén, S. (1965). *Johan Ekeblads brev till brodern Claes Ekeblad 1639–1655. Utg. med inledning, kommentar och register.* (Nordistica Gothoburgensia 2.) Göteborg: Acta Universitatis Gothoburgensis.
- Allén, S. (1976). On phraseology in lexicology. *Cahiers de Lexicologie* 29, 83–90.
- Allén, S. (1977). Text-based lexicography and algorithmic text analysis. *Association for Literary and Linguistic Computing Bulletin* 5 (2), 126–131.
- Allén, S. (1979). Lexical morphology. A model and an application. In H. Bergenholtz & B. Schaefer (Eds.), *Empirische Textwissenschaft*. Königstein/TS: Scriptor.
- Allén, S. (1980). The language bank concept. In J. Raben & G. Marks (Eds.), *Data bases in the humanities and social sciences*. Amsterdam, New York, Oxford: North-Holland.
- Allén, S. (1981). The lemma-lexeme model of the Swedish lexical data base. In B. Rieger (Ed.), *Empirical semantics 1–2*. Bochum: Studienverlag Brockmeyer.
- Allén, S. (1985). Computational linguistics in a future-requirement framework. In S. Bäckman & G. Kjellmer (Eds.), *Papers on language and literature presented to Alvar Ellegård and Erik Frykman*. (Gothenburg Studies in English 60.) Göteborg: Acta Universitatis Gothoburgensis.
- Allén, S. (Ed.) (1988). *Possible worlds in humanities, arts and sciences. Proceedings of Nobel Symposium 65*. Berlin, New York: Walter de Gruyter.
- Allén, S., Hamesse, J., & Zampolli, A. (Eds.) (1988). *Literary & linguistic computing 3 (2)*. Oxford: Oxford University Press.
- Allén, S. & Petöfi, J.S. (Eds.) (1979). *Aspects of automatized text processing*. Hamburg: Helmut Buske.
- Svensk ordbok* (1988). Stockholm: Esselte Studium. (2 uppl. Utarbetad vid Språkdata.)

## Liseberg i våra hjärtan

Göteborgs-Posten 15 juli 1990.

- Har du varit på Liseberg nyligen, frågade en som ville veta.
- Var lugn för svenska folket! En så vacker park med så fina traditioner! Och mitt i alltihopa blinkar Fyr Elise.
- Då vet du kanske också vad de vill förströ oss med i år?
- Attraktionerna? Jo då. Om du tänker på spel, kan du välja på till exempel Skee-Ball, Waterball, Basketball, Boom Ball och Woody. Det låter väl trevligt?
- Det är säkert trevligt. Hur det låter är en annan sak.
- Vad då? Vill du ha något att pigga upp dig med? En åktur kanske? Prova Rainbow eller Break-Dance – eller Flume-Ride om du vill åka båt.
- Nej, jag är nog mera för rymden.
- I så fall skall du till Spaceport. Där kan du ta dig en Simula Tour. Och efteråt kan du gå på Star Dust och få dig ett Space-meal.
- Det ljuder mäktigt. Jag likar att ha nånting enkelt.
- Där är ju Popcorn, men annars frestar de med Big Bang Burger och Soft Frozen Yogurt.
- Väl, för en start önskar jag bara att titta runt.
- All right. Du vill aldrig förgäta Lajseberg.

## **Possible Worlds in Humanities, Arts and Sciences. Opening Address at a Nobel Symposium 1986**

Possible Worlds in Humanities, Arts and Sciences. 1989.

Dear Colleagues,

It is my pleasure and privilege to welcome you on behalf of the Swedish Academy and its Organizing Committee. I trust that you are familiar with the theme of this Nobel Symposium, "Possible Worlds in Arts and Sciences". We note with great satisfaction that such a distinguished group of scholars have come together to discuss such a challenging subject in the year of the 200th anniversary of the Academy.

With your permission, I also bring a greeting from a possible world (Figure 1). As you can see, it is a rather superreflective one. The picture by Guy Billout is part of a series called "Imaginary Gardens".

The theme of the Symposium is truly interdisciplinary, encompassing several academic disciplines, even representing what I hate to call "the two cultures". Thus, we are happy to see participants from natural science as well as philosophy, linguistics, literature, and the visual arts.

Nobel Symposia are sponsored by the Nobel Foundation through its Nobel Symposium Fund. They are arranged in accordance with the intentions of Alfred Nobel, the founder of the Nobel Prizes. Over the last twenty years, a series of symposia have been organized, attracting a great deal of attention in their respective fields. This is one of several Nobel Symposia supported by the Swedish Academy, whose domain is language and literature.

Some of you, a very small minority, were right here at the IBM Nordic Education Center exactly the same days six years ago, when we held the Nobel Symposium on "Text Processing", dis-





FIGURE 1.

Discussing aspects of text representation, text analysis and generation, and text typology and attribution. As veterans you will recall that, in our deliberations on text-understanding problems, the notion of possible worlds was touched upon. That was where the idea of the present Symposium was born.

Even if, accordingly, the impetus of the Symposium was linguistic, it is obvious that a concept of possible worlds is used as a tool in many other fields, particularly those I mentioned.

This comes natural if we think of the general setting put forward in the invitation.

In the Actual World, there are various Media of Communication, which have developed gradually. Such media of great importance are natural language in speech and writing, formal languages including mathematical and logical notation as well as programming languages, and also visual art and music.

It is possible to use these media for a wide range of purposes, such as practical, scientific, and artistic. The result in each case is a Manifestation. This can be a text of a matter-of-fact or belletristic type, a piece of art, or a scientific theory or model.

The design and the interpretation of such a manifestation can be regarded as the establishment of a Possible World. Both processes involve a number of interesting problems.

The main point concerns the relationship between man, the medium, the actual world, and the possible world established. This includes, of course, the problem of representation or mimesis versus self-reflexivity or autonomy, as well as the fundamental questions of whether a possible world created by art can be an instrument for understanding the actual world and, conversely, whether scientific theories and models can also be regarded as fictional in some way.

In the contributions presented, this ground is covered admirably, a host of important issues being brought up for discussion. Among these there are those of differences between language communities, truth-valuation and modalities, extension and intension in semantics, universes of discourse, fictionality and performative speech acts, ontological landscapes, polyphony and revelation, syntactic fiction and illusion, quantum mechanics and physical ambiguity as well as the role of a conscious observer and creation "ex nihilo".

The normal sense of *possible* seems to be verbal: 'that can exist, be done, or happen' (The Concise Oxford Dictionary). The underlying Latin word is derived from *posse* 'be able'. This is also reflected in ordinary, general-purpose language.

Consider one of the pictures by the artist Oscar Reuterswärd (Figure 2). It has been presented in the following way: "He has

designed something which doesn't exist, which is not even possible." What does *possible* mean here?

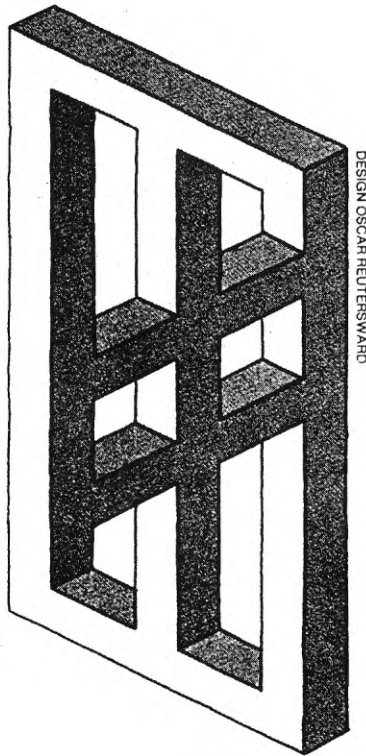


FIGURE 2.

Since the picture is there, the painting of it is evidently feasible or realizable in the actual world. However, the quotation is not concerned with the picture but with the object, and such an object is certainly not feasible in that sense. Thus, in the current negated context, *possible* can be taken to mean 'feasible'. For obvious reasons, feasibility is a main concept involved in the use of *possible* in everyday language. Sometimes there is the additional flavour of suitability, a sort of functional modality, as in one reading of "the only possible answer".

In a wider perspective, it is equally evident that the shape created, although not feasible, is imaginable. In terms of imaginability, *possible* includes, consequently, both feasibility and non-feasibility. The latter type isn't as common as the former in everyday language.

Now consider the following sentence: "It is possible that he knows." This example is different in an important way. It means 'there is a chance, or maybe a risk, that he knows'. In other words, this is a quantitative utterance, although vague. It seems that the probability involved is usually rather low. In this way we have discerned three aspects: imaginability, feasibility, and probability.

These days, when foreign submarines are said to traverse Swedish waters and are in fact sometimes embodied, a new quantitative expression has been coined. Certain observations are interpreted as indicating "a possible submarine". This phrase can probably be taken to be a nominalization of an adverbial construction like "it is, possibly, a submarine".

In everyday usage there are, in addition, a number of set phrases: *as fast/much/soon as possible*, where *possible* has an intensifying function, *if possible*, *make possible*, etc. There is also a substantival version: *possibles* meaning 'reasonable candidates' is a weaker forecast than *probables* meaning 'likely candidates'.

For language users in general, it is obvious that the actual world has priority among the worlds. From a theoretical point of view, it need not be so. Yet, in a human perspective, one might think of all possible worlds, in one mode of thought, as ultimately entered from the actual world.

Special-purpose language is chiefly general-purpose language. Its main characteristic is a set of special terms and maybe some particular phrasing. This also applies to the papers of this Symposium. The terminology used is inventive and revealing, as shown by the following examples: *possibility*, *possibly true statements*, *lexically possible worlds*, *non-possible-worlds theories*, *possible psychological reality*, *impossibility*, *compossibility*,

*physically possible worlds, non-actual possibles, fictionally authentic impossible worlds, possibilist, possibilism.*

One of the fascinating aspects of our theme, observed in several contributions, is that of dimensions. The superstrings proposed elsewhere as the vehicles of microcosmos are reported to require tens or hundreds of dimensions for their description. The theory of relativity offers four, three spatial and one time-like. If we restrict ourselves to the dimensions normally perceived, we find that, basically, sculpture works in three dimensions, painting in two, and language in one dimension. There are evidently things which you cannot sculpture but which you can paint. There are likewise things which you cannot paint but which you can express in speech and writing, as, e.g., a paper on possible worlds.

If we interpret this, in an admittedly straightforward way, as a case of inversely proportional entities – the decrease of dimensions being accompanied by an increase of expressiveness – we obtain zero, the empty space or the empty page, as the all-time-high of expressive force, the black hole, as it were.

This, in turn, evokes the second stanza of the famous "Ode on a Grecian Urn" by John Keats:

Heard melodies are sweet, but those unheard  
Are sweeter, therefore, ye soft pipes, play on;  
Not to the sensual ear, but, more endear'd  
Pipe to the spirit ditties of no tone.

Dear colleagues, fellow-participants!

Quoting Keats, I certainly don't invite you to keep silent. After all, our joint venture has begun.

## Språkfrågor. Om ordbruket i det här landet

Kungl. Vitterhets Historie och Antikvitets Akademiens Årsbok 1989.  
Inträdeshörelse i Kungl. Vitterhets Historie och Antikvitets Akademien  
6.9 1988.

Det är inte utan att det vattnas i munnen på en språkvetare, när man läser i skrifterna, att hela världen är en text. Detta kan väl uppfattas så, att det snarast är fråga om en metafor eller ett förhållningssätt, men ändå. Vad det innebär är bland annat, att språket fokuseras på ett markant sätt. Detta sker också från ett helt annat håll. Bakgrunden är i det fallet språkets nyckelställning i vad som kallas informations- eller numera ofta kunskapssamhället. Språkvetenskapens metoder och strävanden har helt enkelt vidgats och transponerats till flera andra fält.

Det fanns en värld före det mänskliga språket och därmed texterna men naturligtvis ingen världsbild i vår mening. De världsbilder som nu är gängse inom de olika språkgemenskaperna är i väsentlig mån språkens verk. Det paradoxala är att språket på samma gång används som ett redskap för att utforska världen och världsbilderna. Inte minst inom språkvetenskapen är denna paradox en välkänd följeslagare: språket beskrivs med hjälp av språket. Och inte nog med det. Språket refererar ofta till sig självt i texterna via uttryck av typen *som det heter, så att säga*. Man får sålunda lägen, i vilka man metaspråkligt beskriver metaspråk i språket. Det glimtar till av metametaspråk i vårt arbete.

Närmare än att betrakta världen som en text ligger det kanske att betrakta texten som en värld. Att tränga in i en text blir enligt ett viktigt synsätt att försöka bygga upp en möjlig värld på textens grund. Denna möjliga värld kan mycket väl vara omöjlig från den faktiska världens synpunkt. Begreppet eller snarare begreppsklassen "möjlig värld" har en lång historia och lever fortfarande ett friskt liv i både humanistiska och naturvetenskapliga kretsar. Nyligen ägnades ett Nobelsymposium däråt.

Språket har många andra dimensioner. Det varierar geografiskt och socialt, efter kön och ålder och mellan olika stilistiska nivåer. Orden har stavning, uttal, böjning, ordbildningsegenskaper, be-

tydelse, fraseologi och syntaktiska egenskaper. Samtidigt som språket rymmer stelnad metaforik, ligger det öppet för nyskapande. Det används i såväl dialog som prosa och poesi. Till det allmänna språket kommer de många fackspråken. Allt detta och mer än så flätas samman till en intrikat och intressant väv.

En språkbeskrivning – en grammatik eller en ordbok – är att betrakta som en möjlig lösning av ett problem utan facit. Man bottnar helt enkelt aldrig riktigt i det mänskliga språkets förunderliga sjö. Och där finns inga givna enheter, som det bara gäller att fiska upp. Det är tydligt att det saknas utrymme för eventuella språkliga fundamentalister.

Detta är verkligen inte något som leder till uppgivenhet utan verkar tvärtom stimulerande. I en ständig växelverkan mellan teori och empiri kan man steg för steg utveckla sina definitioner och skärpa sina iakttagelser. Hela tiden erfar man nya tendenser, nya överlappningar, nya flertydigheter. Varje seriös språkbeskrivning blir en rapport från ett forskningsfält.

Dessa vetenskapliga och metodologiska spörsmål får tjäna som bakgrund till några aktuella språkfrågor. Vi kan först lägga märke till att substantivet *fråga* är flertydigt. I en svensk ordbok från senaste tid definieras de två huvudbetydelseerna som "begäran om upplysningar /vanl. betr. ngt frågeställaren inte känner till/" respektive "(visst) aktuellt ärende /som bör diskuteras igenom/". Jag har tänkt mig att komma in på båda aspekterna. Det kan tilläggas att det finns en tredje betydelse hos *fråga* i vissa uttryck, nämligen "tvivel" i fall som *ifrågasätta* och *utom all fråga*.

Jag styrks dagligen i uppfattningen, att det finns ett stort och växande intresse för vad som är gott svenskt språk, "hur det skall heta" som det gärna formuleras. Här gäller det ofta de formella aspekterna: stavning, uttal och böjning. Skall det vara accent över *e* i *jubileum* och *museum*? Nej. Heter det *kakafoni* eller *kakofoni*? Det senare. Vad heter *video* i pluralis? *Videor*. Det är detta behov av en formell norm som Svenska Akademiens ordlista skall tillgodose.

Dagligen får jag också nya bevis på ett allt tydligare intresse för ords betydelser, deras nyanser, deras laddningar, deras an-

vändning i olika kontexter. Skulle inte *egennyttan* kunna innebära ett positivt intresse för den egna personen, frågade någon häromdagen. Svaret är, att ordet skulle kunna göra det, men att språket har lexikaliserat den egoistiska innebörden. Inte sällan är dessa frågor kopplade till en klar insikt om den kunskapsevolution som samhället nu genomgår.

Behovet av ett verk med tyngdpunkten i betydelsedefinitioner och kontextexempel är vad Svensk ordbok är avsedd att tillgodose. Här får beskrivningen en mer framträdande roll än i ordlistan. Det är naturligt att ordboken blir en riktpunkt för gott ordbruk, men att direkt föreskriva vad ord skall betyda är en vanskelig väg, en väg som ytterst kan leda till Newspeak. Låt mig markera, att jag inte talar om fackspråk. Låt mig också understryka, att det är en artskillnad mellan en ordlista och en ordbok.

Naturligtvis behöver vi också en dokumentation av nysvenskans grammatiska struktur med tyngdpunkt i morfologin och syntaxen. En sådan är under arbete sedan början av 1986. Jag syftar på projektet Svenska Akademiens grammatik, som nu har hunnit halvvägs. Tre professorer är tjänstlediga under fem år för att genomföra arbetet med stöd av assistenter och kommentatorer, med hjälp av datateknik och i nära kontakt med Akademiens språkkommitté. I denna beskrivning inarbetas också språkvårdsaspekter.

Språket kryllar av liv. Man kan se det i Svensk Bokhandel, när höstens böcker presenteras. Man kan höra det vid kaffebordet. Man kan avläsa det i det klåfingriga klottrandet. Man märker det i sentenser och maximer. "Cherchez la femme" är exempelvis ett tema, som aldrig upphör att engagera och som nu kan ses citerat i formen "Bakom varje framgångsrik man står en förvånad kvinna."

Journalister tar gärna ut svängarna. Språket torde tåla det. Hur långt läsarna orkar med är en annan fråga. När "de män av makt och mod som styr världens öden" kallas för "adrenalinstinna hannar", vad är då formuleringskonst och vad är fördom?

Uttrycken kan varieras. I rubriken över ett reportage om ett av de ärenden vi har undfågnats med på sistone stod det om X: han *sa inte hela sanningen*. I själva artikeln hette det först: han *har*



varit oärlig, när han lämnat uppgifter angående sitt agerande. Längre fram i texten blev det: han *har brustit i uppriktighet*, när han redovisat sitt engagemang. Till sist triumferade den raka journalistiken: *X ljög* vid förhören.

En sak jag finner betänklig är tendensen att salta ordströmmen i överkant. Språket och språkklimatet tar skada av överdriven saltning, menar jag, lika väl som vägar och broar gör det. Det finns åtskilligt i våra texter av *fähundar*, *skitstövlar* och *kulturskymning*. Skulle det inte gå an att föra ett samtal i stället för att leverera verbala råsopar?

”Jag trivs bra på det här plejset”, kan man höra anställda anförtro sina interlokutörer. ”När det gäller att vårda svenskan saknar vi know-how”, har det också sagts. Svengelskan finns ständigt omkring oss och kräver vaksamhet. Trots allt är vi tämligen överens i språkvetenskapliga kretsar om att det inte rör sig om en dödlig fara. Men om vi sysslar med det som på engelska heter *desk top publishing*, skall vi då verkligen *desktoppa*?

Ett ord som *koncept* i betydelsen ”begrepp, idé” har kommit i svang under senare tid. Betydelsen ”utkast, kladd” kvarstår. De båda betydelserna levde sida vid sida i flera hundra år i svenskan, innan den förra försvann i slutet av 1700-talet. Återupplivningen kan otvivelaktigt skrivas på engelskans konto.

När det gäller *å*, *ä* och *ö* har företagen *Skanska*, dvs. Skånska Cement, *Samhall*, dvs. Samhällsföretag, och *Gotabanken*, dvs. Götabanken, den ovanskliga äran av att ha gått i spetsen för utvecklingen. Vad föranleder denna lindrigt talat osolidariska hållning till svenskan, kan man undra. Ett av skälen är att man inte vill låtsas om utomlands, att man är svensk. Den belastningen slipper man, om man uttrycker sig på skanska. Tyngst faller ansvaret på Samhall, som är ett statligt företag.

Vår tid har också sina sanningsvittnen. När det gäller språket har de på senare år då och då lyckliggjort kultursidorna, också dem som i övrigt hör till landets mest respekterade. Vad sägs till exempel om följande? I uttryck av typen *ett kvarts varv*, *ett tjugofjärdedels dygn* uppges *ett* vara bestämning till *kvart* respektive *tjugofjärdedel*. Substantiven *kvart* och *fjärdedel* är därför neutrum i svenskan. Detta är en grammatisk analys med säll-

samma utvecklingsmöjligheter. Men riktigt så bottenlös är inte språkets sjö.

Det ligger en stor fara i att kraven på professionalitet i just språkliga ting nonchaleras på ledande publicisthåll. Språket är inte en kuriositet i stil med högvuxna solrosor och hopvuxna morötter. Man kan hoppas på en viss självläkning, men det är mycket viktigt att det klagörs från intellektuellt håll, att dilettantism är lika förkastlig i fråga om språk som i fråga om litteratur eller fysik. Man ser inga artiklar som hävdar, att ljudets hastighet är högre än ljusets eller att Gustaf Fröding var en surrealistisk dramatiker från Småland.

En gång i tiden *gick man hädan*. Sedan *lämnade man in*. Nu *loggar man ut*. Metaforen är hämtad från databehandlingens språk. Det är helt normalt att en ny teknik liksom andra nya företeelser på detta sätt avspeglas i språket. Jag ser inte det som något stort problem. Det verkligt viktiga är, att vi ser till att vi kan använda svenskan i kommunikationen med datasystemen. Och det skall vara full svenska, inte någon vingklippt version av den typ som går under benämningen kontrollerat språk. Jag ser tecken på att det börjar finnas en viss medvetenhet om detta. Den stärks av Europarådets betydelsefulla initiativ med syfte att verka för Europas språkliga mångfald, en av världsdelens rikedomar.

Det enda som är konstant i livet är förändringen, heter det. Språket står aldrig stilla. Likväl fryser vi det efter bästa förmåga ett ögonblick, när vi gör en lexikalisk eller grammatisk beskrivning av nusvenskan eller något annat skede. Det är viktigt för både språkforskaren och läsaren att tillägna sig detta grepp och rätta sin presentation respektive sin läsning därefter. Exempelvis blir en betydelsedefinition i detta fall just en uppgift om innebörden, inte en etymologisk upplysning. Den kan ges på annat sätt.

Efter en nedgång i det historiska intresset, både inom och utom språkvetenskapen, kan man nu skönja en viss återhämtning. Fortfarande är dock historielösheten en faktor att räkna med. I en diskussion om visor och vistexter ställdes frågan om varför tidningarna bevakar området så dåligt. Från ansvarigt håll svarades då, att det är svårt att skriva om gamla visor. ”Det är inte fräscht.

Det är ointressant.” Tilläggas bör att de gamla visor som det gällde – och som fortfarande står på repertoaren – var från 1960-talet.

Vad är det då språkbrukarna själva frågar om? Jag kan ge några synpunkter på detta efter att ha bearbetat ett material, som har kommit in till ett återkommande inslag i det medium som förr kunde heta bildradion. Jag har fått drygt 3 000 frågor den vägen från alla delar av landet. Frågorna fördelar sig jämnt mellan kvinnor och män.

En glädjande iakttagelse är att var tredje fråga gäller innehållsliga problem. Till dem räknar jag spörsmål om betydelse och ordval. Vad betyder *postmodernistisk*, undras det till exempel. Det måste ju anses vara en bra fråga. Är salt en krydda, frågas det från ett annat håll. Jo, salt är både en krydda och ett konserveringsmedel, det är lexikograferna ense om.

Som väntat samlar substantiven det ojämförligt största intresset. Var fjärde fråga tar sikte på dem. Också verben, adjektiven och adverbena ligger väl framme. Alla ordklasserna är för övrigt representerade. Bland de allmänna aspekter som läggs på språket dominerar starkt kommentarer till språkbruket i radio och TV, något som knappast är oväntat.

Det heter ju *jag delar din mening*, skriver en av de många skarpsynta tittarna, hur kan då *delade meningar* som det verkar betyda ”olika meningar”? I språkets värld är just ambiguiteten något av det mest fångslande. Verbet *dela* drar åt både söndring och gemenskap. Å ena sidan har vi *frågan delar partiet i två läger*, alltså *delade meningar*, och å andra sidan *de delar ljuvt och lett* liksom *jag delar din mening*.

Prepositionerna ger anledning till många frågor. I ett brev noteras med rätta, att *på* håller på att tränga undan flera kolleger: *förslag på något* i stället för *till*, *förteckning på något* i stället för *över*, *kunskaper på något* i stället för *om* osv. Här finns rum för upplysning och vägledning.

För ett drygt år sedan talade herr Bergh i denna krets om de utrotningshotade justeringsmännen, dessa förtroendevalda som tenderar att bli justeringspersoner. Relationerna genus kontra sexus och människa kontra man tilldrar sig också frågeställarnas

intresse. Som ersättning för *riksdagsman* föreslår en av dem, att vi som alternativ till *riksdagsledamot* skall använda *riksdagsdaman*. Som han matematiskt korrekt påpekar består ju senare delen *daman* till 60 % av *dam* och till 60 % av *man*.

En av språkets många dimensioner är som sagt den stilistiska. Min lilla grimas i dagens underrubrik hör hit: *i det här landet*. Det är ett ganska nytt modeuttryck, som florerar särskilt i politikerspråk. Åtskilliga tittare tar upp det i sina brev. De tycker det är slapt och nedvärderande, låter avståndstagande och främmande, anonymt och rentav förolämpande. Det väcker kort sagt irritation. Uttrycket passar, om man är i ett annat land och talar om det landet, påpekar man, men inte annars. En av brevskrivarna ger som exempel en notis om en landskamp på Råsunda: ”Det här landet mot England”.

I riksdagen talar man av naturliga skäl mycket om landet. Ett favorituttryck är det associationsrika *i vårt land*, som yttras 700 gånger på ett år i kammaren att döma av ett undersökt material. Det stilistiskt obelastade *här i landet* ligger omkring 250. Nästan lika vanlig är faktiskt frasen *i det här landet*. Dess uppgång innebär naturligtvis en nertoning. Som så ofta kan man för övrigt ta ett steg till: *i det här avlånga landet*.

Som väl är finns det goda krafter på språkets sida. Det finns utmärkta skribenter och klipska läsare, välartikulerade talare och uppmärksamma lyssnare, hängivna svensklärare och välutbildade ungdomar, uthålliga språkvårdare och lysande författare.

Till syvende och sist – för det sjunde och sista med andra ord – är det oss alla här i landet, tillsammans med en viktig grupp i vårt östra grannland, som svenska språket har att förlita sig på. Här gäller regeln ”Tiga är silver men tala är guld”. Ett ord betyder så mycket.

## Om Svensk ordbok

Svenskläraren 1988: 2.

Svensk ordbok ger 66 000 kärnbetydelser, 25 000 indirekt definierade ordbildningsexempel, 25 000 specificerade betydelsenyanser och 4 000 idiom (typen *vara på tapeten*). Det blir sammanlagt 120 000 innehållsligt behandlade språkliga enheter. Allt detta belyses med 100 000 satsexempel, till stor del autentiska, hämtade ur Språkbankens stora material. Härigenom tydliggörs också ordens konstruktioner. Tyngdpunkten ligger alltså i ordens betydelser och användning. Svensk ordbok är i första hand en definitionsordbok.

Sitt omfång till trots är ordboken liksom alla ordböcker ett urval. Stor vikt har lagts vid att göra urvalet så täckande som möjligt från användarnas synpunkt.

Man bör också tänka på vad Svensk ordbok inte är. Den är ingen encyklopedi 'uppslagsverk om mänskligt vetande [på ett, flera el. alla kunskapsområden; alfabetiskt el. systematiskt ordnat]' (s. 263 i ordboken). Den är inte någon fackordbok 'ordbok med förklaringar av termerna inom ngt område' (s. 282). Den är inte heller någon ordlista 'ordförteckning utan ingående ordförklaringar' (s. 866).

I stället är Svensk ordbok sålunda ett verk, som syftar till att ge ganska ingående definitioner av allmänspråkets ord. Detta är en spännande uppgift, fast svår av flera skäl. Mest noterbart härvidlag är att ordens betydelser i allmänspråket inte är så väl specificerade som i fackspråken. Ordet *tvilling* definieras exempelvis 'ettdera av två samtidigt födda syskon [vanl. i förh. till det andra]' (s. 1318). Naturligtvis kan tvillingar inte födas samtidigt i egentlig mening, men den angivna betydelsen torde ändå fånga innebörden i allmänspråkets ord.

Den vaghet och tänjbarhet som betydelserna i allmänspråket uppvisar är i själva verket en förutsättning för deras användbarhet i den normala kommunikationen. Det är förvisso inte heller så, att betydelserna bildar något slags konsekvent

system, ett faktum som man inser måste få genomslag i beskrivningen. Inte desto mindre ankommer det naturligtvis på ordboken att karakterisera betydelseerna så väl som omständigheterna medger.

Det har också gällt att skapa ett teoretiskt underlag för beskrivning av ords betydelser liksom mönster för definitioner av olika slags ord. Någon egentlig föregångare på svenskt språkområde finns inte, utan de närmaste motsvarigheterna är utländska. Vid genomförandet av projektet måste man dessutom ta hänsyn till det utrymme som står till förfogande, i vårt fall ett band.

Svensk ordbok inleds med påpekandet att det att utarbeta en ordbok på sätt och vis är ett oavslutat arbete. Man gör ständigt nya upptäckter i fråga om orden och sambanden mellan dem, och man ser möjligheter till förbättringar på olika punkter. Oaktat ett omfattande kontrollarbete på Språkdata (där ordboken tillkommit som ett led i forskningsprocessen), manuskriptgenomgång på Svenska språknämnden och korrekturläsning på förlagets ordboksredaktion kan dessutom vissa onöjsaktigheter ha slunkit igenom. Av båda dessa skäl är en levande dialog med användarna mycket värdefull. En sådan har också kommit till stånd och har, bortsett från viss desinformation i pressen, givit uppslag till både modifikationer och tillägg i de successiva tryckningarna (den fjärde pågår nu).

De stavnings-, uttals- och böjningsformer som redovisas tillhör det vedertagna svenska riksspråket, om inte annat anges. De kan därför användas så som de står. Skillnaderna i förhållande till Svenska Akademiens ordlista är marginella och där de förekommer vanligen betingade av att ordlistan med sitt normativa syfte ger förslag på sådana punkter, där inget etablerat språkbruk föreligger. I vissa fall upptas också provinsiella egenheter. De är att betrakta som exempel och gör självfallet inte några andra anspråk.

Ordbokens ord står i strikt alfabetisk ordning för att underlätta användarnas sökande. Det görs hänvisningar från variantformer och oregelbundna böjningsformer.

Svensk ordbok är utformad med tanke på att kunna användas både när man tar del av andras språkproduktion och när man själv är den som producerar. Det har visat sig, att ordboken är till hjälp för såväl dem som har svenska till modersmål som dem som arbetar med svenska från andra utgångspunkter. Det har också visat sig, att både skolelever på olika stadier och vuxna har nytta av den. Prova gärna!

## Akademiens språkliga engagemang

Tal och texter 4–5 april 1986. Svenska Akademien 200 år. 1986.

I språket är allting möjligt. Det kan låta så här:

- Kommer du inte ihåg vad som hände för tio år sedan?
- Jo, jag minns det – men jag har glömt det.

Eller så här, med ett annat läte:

O Romeo ...  
 'Tis but thy name that is my enemy.  
 Det är blott ditt namn som är min fiende.

Eller rentav så här:

Det finns flera tecken som tyder på att Jesus föddes omkring år 5 före Kristus.

Språket är med andra ord också det omöjligas värld.

Svenska Akademien är inte en litterär akademi. Den är en språklig och litterär akademi. Det är den i kraft av sina stadgar och sin arbetspraxis. När Gustaf III i sitt stiftelsebrev säger, att vi har saknat ”en Inrättning, ägnad endast til Språkets förbättring”, så säger han också, att det gäller ett språk som har saknat ”den namnkunnighet, hvilken store Scribenter endast förskaffa”.

De båda uppgifterna är intimt förbundna med varandra. Vi kan formulera det så, att det är fråga om det språkliga systemet och om dess användning. Arbetet med språkssystemet kräver insatser av språkvetenskaplig natur. Den mest avancerade användningen återigen har litterära förtecken.

Vi får bland annat tänka oss Akademiens tillkomst mot bakgrunden av en ostadig språkmiljö. Stavningen fluktuerade, ordböjningen likaså. Satsbyggnaden kunde vara motsträvig, och ordförrådet gav inte möjlighet att uttrycka allt man önskade. Det var upplagt för satsning på dokumentation och rekommendationer.



Enligt sitt uppdrag har Akademien givit ut en grammatik. Året var 1836. Publiceringen skedde under viss vanda, och kritiken inom och utom Akademien var inte alldeles oberättigad. Den språksyn som kom till uttryck ter sig också annorlunda än vår: "Modersmålet [har] omsider antagit en form, hvilken – man dristar tryggt påstå det – aldrig kan föråldras."

Nej, språket är dynamiskt, utvecklas ständigt. Det rymmer inre spänningar, och det reflekterar förändringar i det samhälle som det tillhör.

Låt mig ta några exempel med anknytning till Akademien. Om jag hade haft uppgiften att tala vid 100-årsjubileet, är det mycket möjligt att jag inte hade sagt *Svenska Akademi'en* utan *Svenska Akade'mien*. Det är ett uttal med klassiska anor. Vi kan jämföra med att världens äldsta ännu verksamma akademi, som nyligen firade sitt 400-årsjubileum i hemstaden Florens, heter *Accade'mia della Crusca*.

Den gamla traditionen att inleda ett inträdestal med "Mine herrar" blev också föråldrad. Den förste som sade "Mina herrar" var Esaias Tegnér d.y. 1883. Nu inleder man ju för övrigt sitt inträdestal utan något anrop alls.

Till dessa exempel från ljudläran och formläran kan vi lägga ett från syntaxen. Presentationen av den första upplagan av Svenska Akademiens ordlista 1874 öppnades i inledningen med en komplicerad mening på jämnt 100 ord: "När åt den läsande verden öfverlemnas en sammanträngd bild af modersmålets grammatiska och ortografiska former, hvarvid fördelen kunnat beredas att inom en ringa synvidd sammanföra många af egenheterna och mycket af rikdomen i språkets yttre dänning, har tillfället jemväl iakttagits att i afseende på ett antal enskilda ord och uttryck vidröra språkrigtigheten, der af en eller annan orsak ett missförstånd uppkommit, en missbildning för någon tid eller inom en mindre omkrets vunnit insteg, men kan med ett penn- drag afhjelpas på sitt eget lilla och ej ofta beträdda område, utan att störande ingripa i det hela eller verka en allmännare rubbning." Här är vi ett stycke från dagens uttrycksvanor.

Nu har Akademien startat ett nytt projekt kallat "Svenska Akademiens grammatik". Tre högt kvalificerade forskare har en-

gagerats att genomföra arbetet på fem år i samråd med Akademiens språkkommitté. En beskrivning som den det är fråga om här bör kunna öka intresset för det egna språket, främja frimodigheten och på så sätt stimulera språkanvändningen.

Åt harmonisering av stavningen, som är så betydelsefull för kontakten inom språksamfundet och mellan generationerna, har Akademien ägnat åtskilligt arbete under de 200 åren, och det pågår i enskilda detaljer ännu. Ett par av milstolparna i arbetet var Leopolds "Afhandling om Svenska Stafsättet" 1801 och den nyss nämnda första upplagan av Akademiens ordlista 1874. Ordlistan gick den gången från *abborre* till *östra*, inte till *övrig* som nu för tiden. Det beror förstås på att stavningen då var *öfrig*, vilket placerade ordet före *öga* och *öra* och många andra ord. Att det tog tid att nå den enhetlighet i stavningen som vi nu drar nytta av framgår inte minst av att ännu åttonde upplagan av ordlistan angav, att exempelvis *brett* av *bred* alternativt kunde stavas *bredt*. Den upplagan var gällande fram till 1950.

Vi räknar med att senare under jubileumsåret kunna ge ut den elfte upplagan av ordlistan. Liksom tidigare kommer denna att innehålla ett rikt och levande ordförråd med rekommendationer om i första hand stavning, uttal och böjning inom det allmänna kommunikationsspråket. Man kan slå upp både *härdsmlta* och *hällristning*, både *likviditet* och *livskvalitet*. Med tanke på allt som händer i språket skulle ordlistan nog fylla sina uppgifter bäst, om den kom ut i ny upplaga vart tionde år.

På sin ordbok började Akademien arbeta redan innan ett år hade gått. Man lottade ut bokstäverna bland ledamöterna och skrev provartiklar. Arbetet rann emellertid ut i sanden – av naturliga skäl, kan vi nu säga. Att göra en ordbok kräver en modell, ett stort material och lexikografisk kompetens. Dessa förutsättningar var endast uppfyllda i liten mån.

Det var inte förrän hundra år senare som ordboksarbetet verkligen kom igång. Det första häftet, som omfattar *a-afbild*, publicerades 1893. Redaktionen hade förlagts till Lund av personskäl. Hittills har 29 volymer kommit ut av denna ingående beskrivning av det svenska ordförrådet och dess utveckling alltifrån

1520-talet. Efter hundra års möda har man kommit till *spånta*. Även om man tar modern teknik till hjälp, ligger det inemot ett halvsekel till framför de tjugo i redaktionen.

Mycket kan emellertid göras under tiden. Utöver grammatiken, ordlistan och ordboken driver Akademien ett fjärde projekt. Det heter "Om svar anhålles" och går ut på att läsa in hela ordboken och lagra den som en databas. Detta kommer att ge helt nya möjligheter till sökning i den väldiga ordboken liksom till bearbetningar av olika slag. Det inbjuder också till komplettering där den är föråldrad eller ofullständig.

När man väl har ordboken i digital form, kan man dessutom utnyttja den nya lasertekniken. Den gör det möjligt att lagra hela verket med sina 200 miljoner tecken på en enda kompaktskiva. Tillgängligheten ökar språngartat.

Svenska Akademien arbetar ju inte längre ensam i språkets och litteraturens tjänst. Universiteten och olika språkvårdsorgan spelar mycket viktiga roller. Skolan och massmedierna har ett stort ansvar. Dessutom har i dessa dagar frågan om de nationella språken lyfts upp på det europeiska planet.

På många håll i Europa har man liksom hos oss känt trycket utifrån, framför allt från amerikansk engelska. Härvidlag är det av stor och växande betydelse på vilket språk datasystemen kommunicerar med användarna. Vid ett kollokvium i Tours i år i Europarådets regi antogs ett manifest i detta angelägna ärende. Det mynnar ut i en uppmaning till vetenskapliga institutioner, företag och regeringar att stödja den nationella utvecklingen av sådana system för att därigenom bevara det språkliga arvet. Detta må tagas på allvar.

Ytterst är språket på en gång allas vårt redskap, författarnas lera, rikets kitt och nationens minne. Det är denna mångskiftande och oändligt fascinerande materia som Svenska Akademien ägnar sig åt. Den bästa sammanfattningen erbjuder kanske ordet mysterium.

## På ren svenska

Språk. Svenskläraryrskriftens årskrift 1986. Tal vid Positiva Sveriges frukost på nationaldagen den 6 juni 1986.

Varför skall vi använda så många främmande ord, när vi har en adekvat och relevant svensk vokabulär? Detta yttrande, som tillskrivs en av Gamla Stans nuvarande invånare, har en djupare innebörd än man kanske anar. *Adekvat, relevant, vokabulär* – javisst, lånord med latinets i bakgrunden. Men *använda* och *främmande* är också lånord – från tyskan.

När vi idag firar just *nationaldagen*, sker det med latinets hjälp. När vi gör det här i Berns *Salonger*, tar vi till ett franskt lån, som i sin tur visserligen går tillbaka på vårt germanska ord *sal*. Till och med ordet *språk* har kommit utifrån, nämligen från tyskan.

Nå, alla dessa lånord och många andra har anpassat sig väl till svenskans struktur och betydelsesystem. Detta är avgörande. Språket har behållit sin identitet.

Svenska Akademien skall ju arbeta på svenska språkets renhet, styrka och höghet. Vi kan notera, att *renhet* här snarast syftar på klarhet, knappast på purism, att *styrka* avser uttrycksfullhet och att *höghet* tar sikte på anseende. Det gällde ju att göra svenskan till vad den nu är, ett kulturspråk på höjd med andra sådana.

However, there is a conspicuous amount of English influence on Present-Day Swedish in areas like, e. g., pop music and technology. Resultatet kallas ju ofta *Swenglish* eller *svengelska*, men jag har på senare tid understundom kallat det för *skanska*.

Det engelska inflytandet tar sig varierande uttryck. Man har exempelvis funnit det lämpligt att motverka säkerheten på våra vägar genom att ständigt påminna om dryckeskärlet *stop*. Inte sällan maskerar sig engelskan som en ren teckensträng: *CAD* 'computer-aided design', *CAI* 'computer-aided instruction' osv. Nu senast har man nått fulländningen i denna serie i och med *CAC* 'computer-aided computing'.

Denna överdrivna svansviftning för det utrikiska i språk och kultur är både onödig och farlig. Vi behärskar nämligen inte

engelska på långt när så bra som vi tror. Handen på hjärtat, skulle ni utan vidare klara en instruktion, en bruksanvisning, en manual på engelska under tidspress i en nödsituation? Det jag vill slå ett slag för kan vi kalla *språklig ergonomi*. Den skall vara baserad på svenska, menar jag.

Om vi skall kunna visa en god sida utåt, får vi kanske putsa litet på några andra ställen också. I dessa dagar kan vi eventuellt tänka efter vad *fattig* betyder i Sverige och utomlands. Vi kanske inte nödvändigtvis behöver utdela *råsopar* i samhällsdebatten och kalla varandra för *utstuderade fähundar*, som tidningarna rapporterar. Och varför måste man just ha *bra på fötterna*, när man *lägger förslag på* hur man skall *managera* den här *biten* med nationaldan?

Allt handlar ju om människor. Vad är då det? Människan är den enda varelse som kan rodna – och som har anledning att göra det, är det sagt. Jag håller delvis med Mark Twain, men bara delvis. Man kan alternativt tala om människan som den förnuftiga varelsen, *homo sapiens*. Också det är jag i någon mån beredd att lämna öppet. Men det finns en karakteristik som går utanpå alla de andra: människan har ett rikt nyanserat språk som också används för konstnärliga syften. Hon är en språkvarelse. För mig är hon med andra ord *homo loquens*.

Ibland, om än sällan, är människan svensk. Det rör sig om ett par promille av jordens befolkning. Vad är en svensk? En som med jämnmod kånkar på sitt personnummer? Nåja. En som älskar björkarnas sus? Jodå. Men trots allt: svensk är i grunden den som kan säga *sjuksköterska*. Det är *homo loquens suecicus*. Jag får nationellt sett reservera mig för en folkgrupp som har min högaktning. Det är de grannar som säger *sjuksköterska* (på finlandssvenskt sätt).

Språket är inte minst ett instrument för att utforska världen omkring en. Det ger i sig självt konturerna till en världsbild, i vissa stycken olika från språk till språk. Att lära sig ett nytt språk blir att tillägna sig en delvis ny, vidgad syn på världen. Ju mer förtrogen man är med sitt eget språk och dess förutsättningar, desto lättare har man att tränga in i andras. Men – riktigt hemma är man bara i sitt eget språk, tillvarons bas.

Därför kära landsmän, svenska män och kvinnor, låt oss språka med varandra! Och låt oss göra det på ren svenska och med gott hopp! En sak till bara: hoppet är en god frukost men en dålig kvällsvard.

## Information

Information om *information*. 1985.

Språkligt sett hör ordet *information* ihop med ordet *form*. I bakgrunden finns latinets verb *informare*, som har utvecklat flera olika betydelser: 'ge form åt, skapa, gestalta', 'göra sig en föreställning om, föreställa sig', 'göra en bild av, teckna, skildra' och 'ge utbildning åt, skola'. Den sista av dessa betydelser reflekteras i vårt nu föråldrade ord *informatör*, 'hemlärare'. Jämför också att *informationsmästare* på sin tid var titel för (musikalisk) lärare. Det som har levt vidare är främst betydelseelementet 'föreställning, idé'.

Betydelsekärnan i nusvenskans *information* kan anges som 'en viss mängd fakta eller föreställningar'. Formuleringen innefattar både sådant som syftar till att avspegla den faktiska verkligheten och sådant som avser en önskad eller frammanad värld i t.ex. reklamässigt eller propagandistiskt sammanhang. Det är genomgående alltså fråga om en mental storhet. Exempel på denna användning är *få information*, *lämna information*, *publicera information*, *samla information*, *sitta på information*, *undanhålla information*; *fel information*, *korrekt information*, *omkombinerbar information*. Några belysande sammansättningar är *informationsbärare*, *informationsexplosion*, *informationsflöde*, *informationsförlust*, *informationsförmedling*, *informationsförsörjning*, *informationsjakt*, *informationsklyfta*, *informationskälla*, *informationslager*, *informationsläcka*, *informationsskrift*, *informationssökning*, *informationstavla*; *andrahandsinformation*, *bakgrundsinformation*, *bildinformation*, *sakinformation*. Det som flödar, det man jagar, det man har skrivit om osv. är en viss mängd fakta eller föreställningar.

Rimligt nog tenderar mängden fakta eller föreställningar att vara i rörelse mellan sändare och mottagare. Vi ser den i kommunikationssituationens perspektiv. Det som utgår från sändaren är då någon sorts representation av mängden fakta eller föreställningar. Denna representation betecknar vi som *data*.

För mottagaren gäller det att tolka dessa data för att utvinna den mängd fakta eller föreställningar som sändaren avsett (data kan betraktas som potentiell information). Detta kan naturligtvis lyckas i större eller mindre grad beroende på en mängd faktorer såsom motivation, kunskapsbakgrund och uppfattbarhet. För hela det skissade förloppet används också ordet *information*. Vi får härigenom en andra huvudbetydelse, nämligen 'överföring av en viss mängd fakta eller föreställningar'. Processen som helhet åsyftas i exempelvis *grundlig information på modersmålet* och i sammansättningar som *informationsdag*, *informationsmöte*, *informationsteknik* och *informationsverksamhet*. Det är ju här fråga om dag, möte, teknik och verksamhet som gäller överföring av en viss mängd fakta eller föreställningar.

I olika sammanhang kan olika delar av processen träda i förgrunden. Avsändaren framtonar i exempel som *informationsavdelning*, *informationschef*, *informationsplikt*, *informationssyfte*, *säljinformation* och *muntlig information*. Här är det fråga om t.ex. plikten att överföra en viss mängd fakta eller föreställningar. I ett fall som "Presskonferensen var av informationskaraktär" är det enkelriktade flödet från avsändaren mycket starkt markerat.

På motsvarande sätt fokuseras mottagaren i andra fall. Till dem kan man räkna uttrycket *för information* och förbindelsen *rätten till information*. Det finns anledning att påpeka att detta också gäller Regeringsformens *informationsfrihet*, som bestäms som frihet att inhämta och mottaga upplysningar samt att i övrigt ta del av andras yttranden.

Man kan också finna *information* kännetecknat på ett sådant sätt, att just mottagaren spelar en avgörande roll. Begreppet karakteriseras då som något som uppfattas av en mottagare och framkallar något slags reaktion hos denne (t.ex. Linell 1982). Denna aspekt kan ha sitt berättigande i vissa teoretiska sammanhang. Som framgått gäller den emellertid inte generellt för språkbruket. Budskapet kan mycket väl gå fram utan att någon reaktion utlöses och likväl betecknas som (t.ex. *oanvändbar information*).



I vissa tillämpningar, inte minst på området informationsbehandling, fäster man stor vikt vid att den tolkning som mottagaren gör av inkommande data tillför honom något nytt i fråga om fakta eller föreställningar (t.ex. Spang-Hanssen 1979). Detta är ett specialfall. Uppfattningen är motiverad i de sammanhang där det gäller att förse mottagaren med vad han söker utan att belasta honom med ointressanta eller redan bekanta ting. Den rubbar emellertid inte det allmänna fallet: om två mottagare förses med samma data och tolkar dem på samma sätt, har båda fått tillgång till samma mängd fakta eller föreställningar, dvs. information, oberoende av om den har varit känd tidigare eller ej. Jämför uttryck som *redundant information* och *överflödigt information*. Ett fall som *relevant information* förutsätter att det finns information som är irrelevant. Ett likartat motsatspar utgör textlingvistikens *ny* respektive *gammal information*. Lägg också märke till förbindelser av typen *förstöra information*.

Ibland är det mer eller mindre oklart om det är mängden fakta eller föreställningar eller om det är överföringen därav som i första hand föresvävar språkbrukaren. Naturligtvis kan också båda aspekterna finnas med samtidigt. Några fall att tänka på är *bristande information*, *informationskampanj*, *informationsmaterial*, *informationsnerskräpning* och *produktinformation*.

Man kan lägga märke till att ordet *informationsbehandling* exemplifierar ett annat specialfall. Det som behandlas på detta område är i grunden data, representationen av information, varför *databehandling* vore den egentliga beteckningen. Den mentala storheten information förutsätter ju mottagarens tolkning.

En helt annan innebörd har *information* i termen *informationsteori*. Här är det inte fråga om fakta eller föreställningar utan om signalers sannolikhet. Som alternativ på detta område har *signaltransmissionsteori* föreslagits (Langefors 1968). Om två koder har samma antal signaler, har den kod högst informationsvärde vars signaler är lika sannolika. På detta område studeras företeelser som medelinformation, koders effektivitet, redundans m.m. men alltså inte semantiskt innehåll.

En besläktad icke-mental innebörd finner vi i uttryck som *informationen i ett skivspår, genetisk information och informations-RNA* (dvs. *ribonukleinsyra*). Här är det fråga om processstyrande koder eller signaler av olika slag, teknikbaserade respektive biologiskt grundade. Naturligtvis kan människan skaffa sig kännedom om dessa och lagra denna kännedom mentalt. Det är då fråga om information om information, dvs. *metainformation*.

En särskild kommentar förtjänar *informationssamhället*. Egentligen vore detta ord att hänföra till dem där förleden aktualiserar både mängden fakta eller föreställningar och dess överföring (*informationsnerskräpning* m.fl.). Detsamma kan sägas om uttrycket *informationsåldern*.

Emellertid är *informationssamhället* också knutet till ord som *informationsindustrin* och *informationssektorn*. I den vidaste meningen räknar man här in inte bara forskning, utveckling, utbildning och förvaltning (till stor del) utan också såväl massmediernas, förlagens, reklambyråernas osv. verksamhet som produktionen av datamaskiner, elektroniska komponenter, teveapparater, skrivmaskiner, tryckpressar m.m. (se t.ex. Masuda 1984). Det är härigenom man kommer fram till att omkring hälften av arbetskraften i vissa länder på ett eller annat sätt är knuten till informationsområdet. Denna vida, vaga och snart sagt allomfattande innebörd i ordet *information* finns det anledning att inta en reserverad hållning till.

## Litteratur

- Langefors, Börje*: Introduktion till informationsbehandling, Stockholm 1968.
- Linell, Per*: Människans språk: en orientering om språk, tänkande och kommunikation. 2 uppl. Lund 1982.
- Masuda, Yoneji*: Informationssamhället. Stockholm 1984.
- Spang-Hanssen, Henning*: Terminologisk dokumentation i forbindelse med terminologiske specialer. (i: Seminar über terminologische Diplomarbeiten. Handelshochschule Kopenhagen 1979, s. 57–60).

## Teknik och språk i informationssamhället

Språk og samfunn i Norden etter 1945. 1985.

”Långa köer vindlade utanför Grand Palais i Paris under några dagar i slutet av mars. Orsaken till den stora folksamlingen var boken, denna seglivade anakronism i en elektronisk och digital värld.” Om boken som anakronism skrev Carl Rudbeck i en artikel i Svenska Dagbladet (1984).

”Vem har lagt detta *papper* på mitt skrivbord?” frågar sig en person på en teckning i Nordiska-Datanytt (8:15). Skrivbordet består helt och hållet av textskärmar och tangenter. I samma riktning går titeln på Anthony Smiths bok från 1980: *Goodbye Gutenberg*.

Tanken att vi är på väg in i ett papperslöst samhälle återkommer ofta. Skulle den förverkligas, vore det naturligtvis en faktor av stor betydelse för språket. Det är frågor av detta mera övergripande slag som skall stå i förgrunden här.

Att språket inte förblir opåverkat av teknisk utveckling över huvud taget är uppenbart. Varje förändring i samhället reflekteras i själva verket i språket. Låt oss som exempel tänka på hantverkskulturen under medeltiden, industrialiseringen på 1800-talet och filmen på 1900-talet. Det intressanta blir alltså inte *om* utvecklingen påverkar språket utan *hur*.

För att vi skall känna vittringen av skeendet ger jag här några exempel hämtade ur Svenska språknämndens samling av nyord från efterkrigstiden (jag tackar Claes Garlén för hjälpen med att ta fram underlaget):

*analogdator, astronaut, atomubåt, bankomat, bildskärm, borrhög, bromsservo, bärraket, charter, chips, container, diabild, dialys, digitaldator, elektronikkomponent, elgitarr, ergonomi, fjärranalys, flygsimulator, fotosättning, frysbox, flerkanalsstereo, färdskrivare, gaffeltruck, genbank, glasfiber, gyrostabilisator, halogenljus, halvledarteknik osv. fram till varmluftsridå, videokamera, vindsurfa, vätskefjädring, xerografi, ytaktiv, zooma, ångstrykjärn och överljudsplan.*

## Samhälle

Det samhälle som hyser språket kallas numera gärna informationssamhället. Det är detta som Alvin Toffler har betecknat som den tredje vågen. Föregångarna var förstas industrisamhället och jordbrukssamhället, i omvänd kronologisk ordning.

Vi är idag i ett läge där vi kan värdera ett antal framtidsstudier rörande detta samhälle, eftersom vi sitter med facit i handen. Det visar sig att man i studier från 1960-talet har varit alltför optimistisk beträffande utvecklingen (se t.ex. Vedin 1981). Man förutsade att vi skulle ha helt automatiserade bibliotek i mitten av 1970-talet och automatisk språköversättning med felfri grammatik i slutet av 1970-talet. Då skulle vi också kunna konversera med datamaskiner på talad engelska, och boken skulle ha ersatts av magnetband.

Det skall inte förnekas att det finns en vetenskapligt grundad futurologi. Man får emellertid tolka resultaten av den nämnda sortens studier med stor försiktighet. Det är mycket ovisst om vissa av de prestanda som förespeglats någonsin kommer att bli verklighet. Farten in i informationssamhället är hög som den är.

I en kommentar till vad som förestår skriver John Naisbitt i sin bok *Megatrends*: "To be really successful, you will have to be trilingual: fluent in English, Spanish, and computer." Ståndpunkten är inte orimlig, och den har uppenbar relevans för dagens ämne.

Till dem som har insett detta hör våra japanska kolleger. Deras stora satsning på projektet Fifth Generation Computer Systems innefattar utveckling av maskiner och program som lämpar sig för kunskapshantering, slutsatsdragning m.m. Till detta kommer arbete på utformningen av kontaktytan (interfacet) mellan användarna och systemet. Här tänker man sig i stor utsträckning nyttja naturligt språk. Det är också intressant från vår synpunkt att de tre viktigaste tillämpningarna anges vara frågebesvarande system, översättning och taligenkänning. Till detta kan läggas, att satsningar i samma storleksordning

också görs i USA och att man likaså har startat stora projekt inom den europeiska gemenskapen.

## Information

På databehandlingens område var man i början mest inriktad på maskinvaran. Efterhand gled intresset över mot programvaran. Nu kan man säga att informationen som sådan står i centrum. Detta avspeglas också i användningen av ordet självt, som på sitt sätt exemplifierar teknikens inflytande på språket.

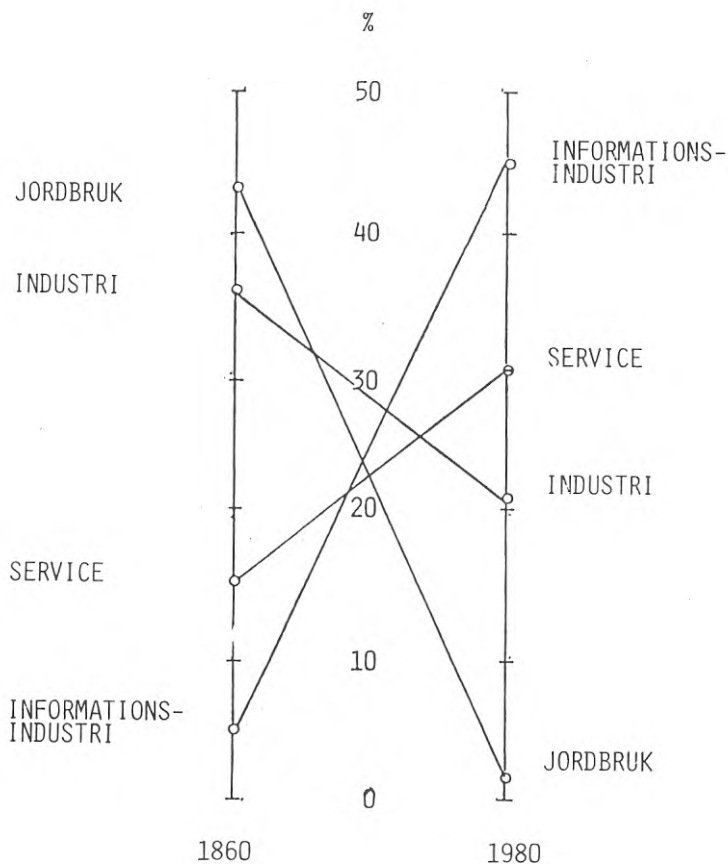
Efter att ha levat ett någorlunda stillsamt liv fram till andra världskriget har *information* under efterkrigstiden svingat sig upp till sådana höjder, att det nu får ge namn inte bara åt samhället utan också åt tidsåldern: *informationsamhället*, *informationsåldern*.

I botten ligger en betydelse som kan anges som 'en viss mängd fakta eller föreställningar' såsom i *informationsökning*. Denna mängd tenderar att sätta sig i rörelse, och man finner mycket riktigt att *information* ofta används för att beteckna en process: 'överföring av en viss mängd fakta eller föreställningar' (t.ex. i *informationsverksamhet*). Till dessa mentala aspekter kommer de icke-mentala, som gäller signaler och koder såsom i *informationsteori* och *genetisk information*. Redan detta ger alltså en rätt brokig bild.

Än mer vidgas perspektivet när vi betraktar ord som *informationsindustri*. Under denna beteckning räknar man in producenter av datamaskiner, skrivmaskiner, tryckpressar, TV-apparater osv., producenter av förlagsartiklar, annonser, radioprogram osv., hela utbildningssystemet, all forskning och utveckling samt en stor del av den offentliga förvaltningen.

Med denna (alltför?) vida syn på informationen får man en ny bild av arbetslivet. I figur 1 visas USA:s arbetskrafts fördelning på fyra kategorier år 1860 respektive år 1980 (efter Marc Uri Porat). Under loppet av de 120 åren har sålunda jordbruk och informationsindustri bytt plats som procentuellt största respektive minsta kategori. Informationsindustrin tar nu hand

om ungefär hälften av arbetskraften. Detta innebär under alla omständigheter att det naturliga språket alltmer kommer i centrum.



FIGUR 1. USA:s arbetskraft 1860–1980 (efter Marc Uri Porat).

## Medier

Nya medier knutna till denna utveckling sköljer över oss. Medan *text-TV* utgör en form av envägskommunikation (med

vissa valmöjligheter inom en meny), erbjuder *teledata* tvåvägskommunikation med bland annat sökning i databaser. *Teledata* är den svenska term som svarar mot den internationella *videotex*. *Datavision* är televerkets teledatasystem.

Till det spridda kommunikationssystemet *telex* har nu lagts det elektroniska *teletex*. Detta är mycket snabbare och innehåller dessutom möjligheter till ordbehandling. Med hjälp av systemet *telex* kan man också överföra bilder. Låt oss i detta sammanhang notera att bildens roll i flera av våra medier är mycket framträdande. Detta saknar självfallet inte heller betydelse för språket och dess bruk. Fenomenet skulle förtjäna sin egen behandling.

Datanät byggs upp i rask takt. *Datex* är televerkets allmänna elektroniska datanät. *SUNET* är Sveriges universitets datanät. Genom nät som dessa kan man komma vidare ut på den internationella arenan. Det finns också lokala nät, som knyter samman exempelvis installationer vid ett universitet.

Elektronisk post och datamaskinella konferenssystem hör till det som möjliggörs. Här uppkommer samtidigt problem beträffande kommunikation utan den personliga närkontaktens hela register. Vad av speciellt de elektroniska "brev" som skall bevaras för dokumentation och framtida forskning är också en påträngande fråga.

Ett ord som nu dyker upp som beteckning för sammanfattningen av teleteknik, datateknik och mikrominiaturisering är *telematik*. Det används t.ex. av televerket. Ordet möter också utomlands med anpassade stavningsvarianter.

## **Människa och maskin**

En viktig punkt gäller kommunikationen mellan användare och datasystem. Från dataindustrins sida vill man gärna göra gällande, att man har utvecklat system som kommunicerar med användarna i naturligt språk. Som jag tidigare nämnt kan detta gälla frågebesvarande system av olika slag. Det kan t.ex. också gälla den interaktiva fiktionslitteratur som nu växer fram

("databoken"), där användaren är en av personerna i handlingen (se t.ex. Fichtelius 1984). Också bland mera seriösa författare finns planer i denna riktning.

Att kommunikationen sker i naturligt språk är emellertid endast delvis sant, och *delvis* är här av fundamental betydelse. Det är nämligen så, att alla aktuella system arbetar med en delmängd av naturligt språk, en delmängd med stora restriktioner i fråga om ord, uttryck och konstruktioner. Från språkvårdens, och i vidare sammanhang samhällets, synpunkt måste moteld anläggas. Bland annat är det viktigt att utbilda ungdomar som förenar språkvetenskaplig och datavetenskaplig kompetens.

Programspråken som sådana ser jag inte som någon större riskfaktor. Vi har under lång tid haft matematikens, logikens, kemins och fysikens termer och formler inpå oss utan vådliga effekter. Detsamma bör gälla programspråken.

Däremot är det nödvändigt att vara på sin vakt mot dåliga system för lagring och hantering av data om personer. Av system med sådana uppgifter måste krävas att de bygger på psykologisk, social och språklig insikt. De måste bland annat ha utrymme för fria kommentarer, där nyanser, kritik, värderingar, osäkerhet osv. kan noteras. Detta är främst en fråga om vilja och kompetens hos de systemansvariga.

I anslutning till diskussionen om system av den typ jag nu har nämnt görs det ibland gällande, att det är datorns begrepp som används för att gestalta verkligheten (se t.ex. Helmersson 1983). Detta bottnar i en felsyn. Datorn har inga begrepp. Det är människan som förser den med möjlighet att vidarebefordra sådana. Kritiken riktar sig därför egentligen mot systemkonstruktörerna, som jag redan tidigare har apostroferat. Att arbetssituationen för dem som utnyttjar datasystem förändras ligger i sakens natur. Deras yrkeskunskap kommer i viss mån att förskjutas. Med väl utformade system behöver den emellertid inte obligatoriskt försämrats. Och såvitt bekant finns det hittills inga belägg för att deras eget privata språk skulle ha förflockats. Diskussionen bör emellertid föras vidare.



## Språk och maskin

Ordbehandlingssystem underlättar kontorsarbete och författande i allmänhet. Man kan arbeta med sin text på ett mycket effektivt sätt och snabbt få den utskrivnen. I vissa tillämpningar kan man också utnyttja textmallar, som hämtas ur apparatens minne. Detta kan vara på gott och ont. Samtidigt som det gör det möjligt att höja standarden hos vissa typer av dokument, kan det eventuellt verka i monotoniserande riktning.

Stundom pekar man på de hittillsvarande systemens svagheter i fråga om orddelning eller avstavning. "En vanlig skolklocka kan höras skolk-locka", skriver Kajenn i en dagsvers med titeln Ordfog (Svenska Dagbladet 25 mars 1984). I och för sig fäster man enligt min mening alltför stor vikt vid denna fråga, i synnerhet i jämförelse med de strödda kommentarer som ägnas åt texternas faktiska innehåll. Uppenbart är emellertid att rutinerna kan förbättras, även i betraktande av att de sista tiondels procenten ställer mycket stora krav på språklig analys.

Intresset för maskinell översättning har ökat igen efter att ha varit mycket svagt i tjugo år. Det har emellertid skett med andra förtecken än tidigare. Visserligen arbetas det på vissa håll med automatisk översättning av bestämda texttyper, men huvudvikten ligger nu på att utveckla arbetsstationer för mänskliga översättare. De blir ett slags översättningslaboratorier som tillhandahåller textbehandlingssystem, ordbanker, dokumentsökningssystem m.m. Effektiviteten och den terminologiska konsekvensen har visat sig må väl av detta.

I fråga om taligenkänning är det lång väg kvar. Vad man kan göra idag är att klara korta ordsträngar, företrädesvis med paus mellan orden, inom ramen för en mycket begränsad vokabulär och med kalibrering på vissa bestämda personer. Det finns tillämpningar även av sådana rudimentära system, t.ex. i branscher där räkneorden spelar en viktig roll. Ett genombrott på området bedömer jag inte som närliggande. Det är inte bara fråga om att utveckla teknik. Den avgörande svårigheten är språkets ytterligt komplexa natur, speglande människans egen komplexitet.

## Ordböcker på chips

Av växande betydelse från språklig synpunkt är det faktum att man kan ha lexikaliska verk i maskinläsbar form. Redan nu är vissa enklare ordlistor och parlörer tillgängliga i apparater liknande fickräknare. På ganska kort sikt kan vi räkna med att ordinära ordböcker kommer att finnas på chips för användning under lite annorlunda förutsättningar än hittills.

Man kan också ha en terminal eller persondator hemma. Den kan vara knuten till en språkbank, i vars stora lexikaliska, textuella och grammatiska material man kan söka svar på sina frågor. Terminalen ger också tillgång till ordbehandlings-system, den kan användas för pedagogiska språkspel av olika slag osv. Det är härutöver möjligt att dra in talet i bilden. När ett ord eller uttryck anropas, kan man få höra det uttalas med hjälp av analog eller digital teknik. Detta har naturligtvis också pedagogiska tillämpningar.

Hjälpmiddel av detta slag, som kompletterar bokvärlden i vanlig mening, ger också vissa andra möjligheter av betydelse från språklig synpunkt. Den motsägelse som ligger i att man måste veta hur ett ord stavas för att kunna slå upp det i syfte att ta reda på hur det stavas kan lösas upp. Om programsystemet inte finner det önskade (felstavade) ordet, kan det t.ex. erbjuda uttalsmässigt närliggande alternativ (*kassett* för ett försöksvis inkodat *kasett* osv.). Man kan också räkna med snabbt resultat, eftersom den tid det tar att bläddra sig fram genom en bok försvinner.

När vi nu är inne på ordförrådet och dess hantering, kan vi också anteckna den roll som tekniken och inte minst data-tekniken spelar som förmedlare av lånord, främst från engelskan. Det är inte bara själva ordbeståndet som berörs utan också ordbildningen och böjningen. För språkvården, den allmänspråkliga och den fackspråkliga, gäller det att båda upp svenskans rika resurser inklusive dess internationella stammar och suffix för att konsolidera språket.

## Exemplets makt

Den tekniska utvecklingen har som vi noterat försett oss med en rikt varierad uppsättning medier. Detta har givit vissa samhällsgrupper tidigare osedda möjligheter att påverka språk och attityder. Låt mig i detta sammanhang peka på tre ting som framstår som oroande.

Här är några prov på vad man kan finna på en sida i en rikspridd morgontidning (inklusive citat från andra tidningar):

*avskryvärt, frihetsförkävande, hänsynslöst, egoistiskt, vildsint, förödande, insinuationer, vulgariteter, starka invektiv, frustande anklagelser, människor utan samvete och moral, utstuderade fä-hundar.*

Det är alltså i detta tonläge som debatten förs. En sak vore om ordväxlingen skedde mer eller mindre i enrum, men här sker den med full exponering. Och mediernas genomslagskraft är som bekant kolossal.

En andra punkt gäller begreppsvärlden. Man kan till nöds acceptera att *krympning* får heta *dimensionsändring* och kanske rentav smått road notera, att medan *skattesänkning* kallas *skattesänkning*, *skattehöjning* gärna kallas *inkomstförstärkning*. Men när man inom sociallagstiftningens ram omöjliggör användningen av ett ord som *fosterbarn*, då ligger man mycket nära gränsen till Newspeak.

Alltför ofta iakttar man att personer som brukar fokuseras i medierna använder språket som informationshinder. Man glider undan, ger oklara svar, talar om någonting annat. I politiska kretsar anses detta vara legitimt. Från språkanvändningens synpunkt är det olyckligt i den mån det leder till efterföljd. Det strider mot själva grundförutsättningen för språklig kommunikation.

Låt mig som en motvikt mot detta peka på att de gravt språkhandikappade, åtskilliga av dem skadade vid trafikolyckor, får allt bättre hjälpmedel genom datatekniken. Hit hör system med ideografisk skrift (Bliss), alfabetisk skrift eller punktskrift. De kan t.ex. vara utrustade med textskärmar och talsyntesapparater

och ha olika slag av språkliga rutiner inbyggda. Härigenom skapas kommunikationskanaler som ofta öppnar nya världar för dem som berörs.

## **Papper och elektronik**

Under gång har jag gjort en rad kommentarer om språket och dess villkor i informationssamhället. Låt mig runda av genom att först knyta an till boken igen. En ökad användning av elektronik tror jag inte kommer att leda till ett papperslöst samhälle, så långt jag nu kan se. Boken är beprövad som medium för långsiktig lagring av information. Den är lätt att ha med sig och praktisk att göra markeringar och noteringar i. Den är oberoende av elnät och batteri. Och den kan inte minst vara estetiskt tilltalande. Vi kommer att få användning för både papper och elektronik.

Vad språket och språkbrukarna beträffar finns det som sagt vissa riskmoment vid sidan om de positiva dragen. Under förutsättning av fortsatt vakthållning tror jag emellertid inte att undergångsstämningar är motiverade. Enligt min syn på saken låter sig inte människan heller hanteras hur som helst.

God vakthållning återigen kräver att språkvårdens företrädare själva arbetar med de nya medierna. Det är enda sättet att skaffa sig den inblick som behövs för att kunna göra de strategiskt riktiga insatserna.

## **Litteratur**

Allén, Sture, Översätarnas språk – finns det? (Mänskligt och omänskligt i informationsteknologin. Red. Tor Kihlman. Chalmers tekniska högskola. 1984, s. 88–96.)

Allén, Sture, Översättningsdatorn – finns den? (Mänskligt och omänskligt i informationsteknologin. Red. Tor Kihlman. Chalmers tekniska högskola. 1984, s. 37–43.)

- Fichtelius, Erik, Detektiven, det är jag. (Upp & ner, 1984:5, s. 26–33.)
- Helmersson, Dick, Låt inte datorerna förstöra språket! [Intervju med Ingela Josefsson.] (Ny teknik. 1983:13, s. 1, 30.)
- Naisbitt, John, Megatrends. Ten new directions transforming our lives. New York 1982.
- Porat, Marc Uri, The information economy. US Department of commerce. 1977.
- Rudbeck, Carl, Noterat i Paris. (Svenska Dagbladet 5 april 1984.)
- Smith, Anthony, Godbye Gutenberg. Oxford 1980.
- Toffler, Alvin, The third wave. London 1980.
- Vedin, Bengt-Arne, Datorer med ett mänskligt ansikte? Stockholm 1981.

## Skandinavisk datalingvistik

The Nordic Languages and Modern Linguistics 5. 1984.

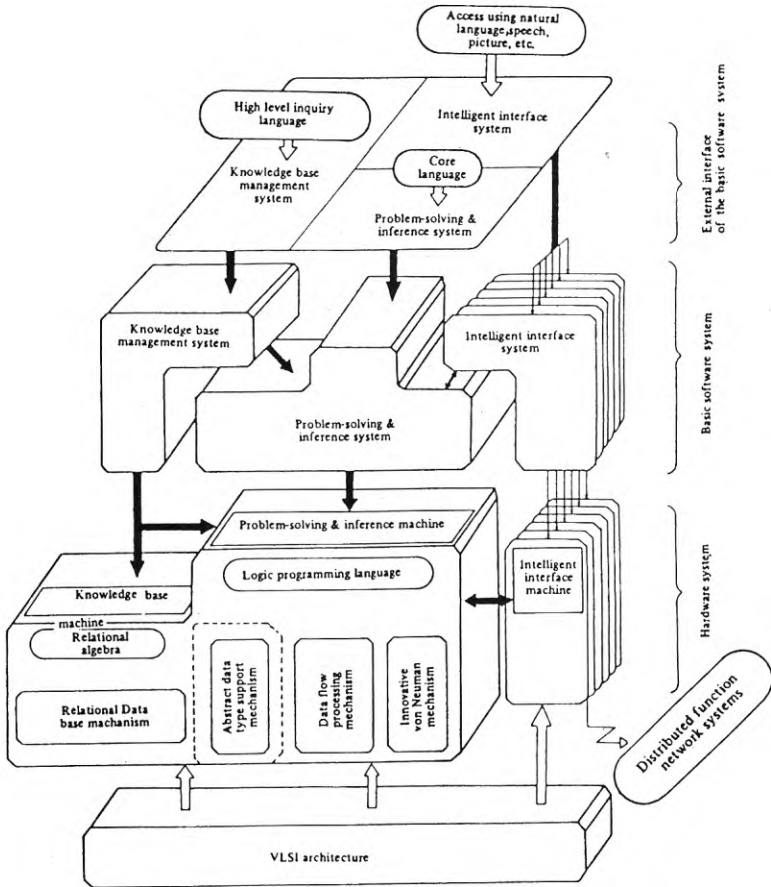
Jag skall börja med en bild för att visa en struktur. Figur 1 är en översiktlig presentation av ett av de största projekten på det datamaskinella området för närvarande, det japanska Fifth Generation Computer Systems. Satsningen är enorm. Man sätter in ungefär fyra miljarder kronor i projektet under en tioårsperiod. Från datalingvistisk synpunkt är det mycket intressant att se vad man tänker sig att göra. Figuren skall läsas så, att man överst har ett så kallat interface, en kontaktyta mellan det datamaskinella systemet och omvärlden, därunder mjukvaran eller programvaran och underst hårdvaran eller maskinvaran.

Man kan konstatera att interfacet planeras ge access eller tillgång till systemet genom naturligt språk (tal, skrift), bilder etc. I interfacet ingår också komponenter som har avseende på problemlösning, slutsatsdragning och kunskapsdatabaser. Interfacets tredelning har direkta motsvarigheter i fråga om både programvara och maskinvara. Inte minst är det värt att observera, att man skall bygga särskilda maskiner som är databasmaskiner, slutsatsdragningsmaskiner osv., en utveckling mot en kraftigt förbättrad situation.

Projektet är på sitt sätt ett uttryck för en ökande insikt om tyngdpunktsförskjutningen på det datamaskinella området. Under de första två decennierna låg tyngdpunkten mycket starkt på maskinsystemet. Sedan ökade intresset för programvaran, och nu kan det med en viss förenkling sägas att informationen själv står i centrum. Man har till och med börjat använda beteckningen informationsvara analogt med programvara och maskinvara.

Det är också intressant att titta på vilka tillämpningar som japanerna i första hand satsar på. De är huvudsakligen tre: språköversättning, frågebesvarande system och talanalys. Detta betyder att det naturliga språket på ett mycket markant sätt gör sig gällande i projektet, både med hänsyn till dess uppbyggnad

och med hänsyn till dess applikationer. Det blir viktigt att vi som humanister, lingvister, datalingvister försöker påverka den här typen av satsningar, så att de resulterande systemen så långt de kan utvecklas fungerar på människans villkor.



FIGUR 1. Översikt av Fifth Generation Computer Systems enligt presentation av projektet.

Av presentationerna framgår att optimismen är stor i projektet. Jag har i olika uttalanden och uppsatser sagt, att jag tycker målspecifikationerna är väl avancerade och att jag inte tror att man kommer att uppnå målen helt. Däremot kommer det med säkerhet att utvinnas mycket av intresse ur projektet. Vi skall också notera att det inte bara är japanerna som satsar. IBM har stora forskningscentra, likaså Bell (inte minst när det gäller talanalysen), Xerox osv.

Datalingvistik är en humanistisk vetenskap och som sådan en språkvetenskap. Det är min grundsyn. Inom språkvetenskapen har vi särspråkliga ämnen och allmänspråkliga ämnen. Bland de allmänspråkliga har vi det som fackmässigt heter allmän språkvetenskap, vi har datalingvistik, fonetiken, sociolingvistik och flera andra prefixlingvistiker, som man ibland kallar dem. Datalingvistik heter i mitt fall språkvetenskaplig databehandling. Båda dessa beteckningar – liksom de i Danmark och Norge förekommande datamatisk lingvistik och datamaskinell språkbehandling – svarar mot den engelsk-amerikanska termen *computational linguistics*.

Diskussioner om avgränsning mellan vetenskaper är egentligen inte helt centrala. I den mån det gäller utarbetande av doktorsavhandlingar eller inrättande av tjänster eller yttranden över ansökningar om forskningsanslag osv. får man ändå ha en ram att arbeta inom. Närmast skissar jag en förenklad modell som jag har haft i bakhuvudet i sådana sammanhang.

Vad jag tänker mig är av naturliga skäl att man ser till syftet. Detta är i korthet att vinna insikt i naturligt språk med datamaskinella metoder. En aning utförligare kan ämnet beskrivas som forskning och utbildning som syftar till ökad insikt i naturligt språk genom fördjupad teoretisk och praktisk kunskap om datamaskinella metoder för språklig analys, syntes och kommunikation. Man kan tänka sig ett slags språkligt programsystem som i görligaste mån kan analysera text, generera text respektive befinna sig i dialog med en mänsklig partner. Om man vill kan man se det som ett slags (ofullkomlig) simulering av en språkbrukare. För att ett sådant system skall kunna arbeta behöver det



resurser av olika slag. Jag kallar dem med heltäckande benämningar för lingvistisk kapacitet och extralingvistisk kapacitet. I båda fallen är det frågan om dels processer, dels data.

Naturligtvis står den lingvistiska kapaciteten i centrum inom datalingvistiken. Det är fråga om den lexikaliska komponenten, om morfologi och syntax, om textlingvistiska aspekter, om pragmatiska aspekter, om allt som över huvud taget behöver komma in. Det gäller de språkliga enheterna själva men också deras kvantitativa egenskaper. Den extralingvistiska kapaciteten har att göra med strukturering av den information som kommer in, den slutsatsdragning som jag talade om nyss och liknande ting. Det är också fråga om kunskap om världen och om vad som har hänt i världen.

Denna allmänna ram är ju synnerligen vid, så inriktningen inom den är viktig. Man skall bland annat också lägga märke till att jag menar, att det inte bara är en legitim uppgift utan en väsentlig del av datalingvistiken att utveckla de resurser som systemet behöver. Det sker oftast i samarbete mellan människan och datamaskinen.

Viktigt för diskussionen är det faktum att datamaskinen är en symbolbehandlande apparat, den första i mänsklighetens historia. Det gör den på ett särskilt sätt ägnad som instrument i utforskningen av språket, symbolsystemet par préférence. Vid utvecklingen av systemet/beskrivningen ställs rigorösa krav. Absolut explicitet i stort och smått erfordras, något som existerande lexikon och grammatiker inte kan erbjuda. Resultterande system skall vidare utgöra en fungerande helhet, ett grundkrav som inte ställs inom andra språkvetenskapliga discipliner. Som nämnts måste också en mängd nya uppgifter om det faktiska språkbruket inhämtas. Allt som allt öppnas stora forskningsfält för språkvetenskapen.

Naturligtvis finns det överlappningar mellan datalingvistiken och andra ämnen. Området artificiell intelligens skall ju enligt sitt mål skapa system som visar ett slags intelligent beteende. Det gäller olika delar av mänsklig verksamhet, i viss mån också språket. Uppenbarligen finns det vidare överlappningar med filosofi, cognitive science, textlingvistik, sociolingvistik, matematisk

lingvistik och annat. Jag lämnar frågan med denna allmänna glimt.

När det gäller forskningsstrategier på området är det viktigt, att man på något sätt kan visa att man har kommit åt innehållet i behandlade texter. Det kan man göra genom att låta systemet skriva sammanfattningar eller parafraaser, ge förslag till översättningar osv. Jag arbetar med ett slags allmän modell som kan kallas för språkbanksmodellen. Den innebär att man vid analys av texter skall kunna placera identifierade element och kategorier av alla slag på rätt plats i en språkbank av teoretiskt avancerad karaktär. Detta är en utgångspunkt för de synpunkter som jag anlägger fortsättningsvis.

Min tanke är att till stor del diskutera vad som sker i förhållandet mellan text och lexikon, vilket leder in på olika typer av forskningsuppgifter, efter hand också på grammatiska och andra ting. Det är en glidande skala hela vägen här.

Det första jag då vill säga är, att en text på sätt och vis är ett lexikon. En text är nämligen ett slags lexikon över sig själv. Underlaget till figur 2 är helt enkelt ett stycke ur en artikel av Peter Bøgh Andersen som handlar om programspråket Simula.

Derfor har vi set en successiv tilnærmelse af programmerings-sprogene til naturligt sprog. Maskinsprogene lå langt fra de naturlige sprog; for at lette programmøren arbejdet konstruerede man de højere programmeringsprog, og arbejdet med at konstruere datasprog, der ligner naturlige sprog, indenfor f.eks. QA, er næste skridt. Datasprogene vil dog rimeligvis ikke komme til at falde sammen med de naturlige sprog. Ser vi på sprogudviklingen i gottonsk opdager vi, at de højere programmeringsprog ligger i smuk forlængelse af denne udvikling.

FIGUR 2. *Belägg på strängen sprog i en artikel av Bøgh Andersen.*

Jag har kursiverat de fall där strängen *sprog* förekommer. Man får *sprog*, *programmeringsprog*, *maskinsprog*, *datasprog*, *sprogudvikling* osv. Faktum är att man kan ha sin text maskinläsbar, sitta vid en terminal och när man vill ta fram belägg på de

ord man intresserar sig för. Texten blir därigenom som sagt ett slags lexikalisk tillgång, och allting finns ju i själva texten. Det är bara det att den inte är analyserad på något sätt. Den utgör ett rent råmaterial, och man kan säga att organisationen är opraktisk i många avseenden.

Därför har man arbetat med en mängd olika modeller för att göra materialet mera tillgängligt. Vanligast är att göra en konkordans. Vid denna konferens har en dansk runkonkordans presenterats. Exemplet i figur 3 är hämtat från Strindbergsprojektet. Det ges ut en ny stor nationalupplaga, en textkritisk sådan, av Strindbergs samlade verk. Man använder arbetskonkordanser som underlag när man slutgiltigt etablerar texten och när man kontrollerar den. Utförandet är en KWIC-konkordans med uppgift om beläggställe. Det visade avsnittet är hämtat ur Det nya riket. Efter hand görs också definitiva konkordanser baserade på de tryckta volymerna. De är tillgängliga på mikrokort.

Interaktiva konkordanser kan gestaltas på många olika sätt. En modell är den som visas i figur 4 (använd av Jonas Löfström). Antag att man vill studera ordet *förstå*. Man anger hur stor kontext man vill ha och får ut det material som då är relevant. Det är fråga om en talspråkskorpus med olika repliker. Talspråksmaterialet som sådant erbjuder som exemplet visar en lång rad intressanta problem.

.13304	jag kan	<i>förstå</i> det faktiskt	189 03240:03 A #20
18903180B	inte det		
18903190A	nää		
18903200B	däremot upplever jag mej som kuratorn där xx		
18903210A	det gör du ja		
18903220B	jaa definitivt		
18903230A	ja		
18903240A	jag kan	<i>förstå</i> det faktiskt	
18903250B	å det		
18903260A	på nåt sätt		
18903270B	ja å det är väldigt jobbigt		
18903280A	jaa		
18903290B	när du har med labila personer		

FIGUR 4. Exempel på interaktiv talspråkskonkordans efter Löfström.

göra gamla ämbetsmän galna, han titulerar kunglig horisonten) och samma år till president i kunglig utnämmandet till överste skattnästare vid kunglig kejsarens hov, äro det oaktat försedda med icke fick fläcka. Resultatet blev en kallelse till erare från konungens egna livgarder, kaptenar och ver festtalet och den offentliga lögnen!! Claris - Men straffet då? - Det kommer! - Claris , egentligen var det ingenting. Hans poesi var en övrigt. En hyckalisch imeslutande Chefens es värdelösa tomen låtit uppföra ett palats vars å okunniga i rikets grundlagar att de tro jag har

2605 majestäts kanslistar för notarier, kongliga biblio  
4296 majestäts (katarren är här; han viskar med sloöbna  
4294 majestäts orden 1816, till ondenskansler 1818, till  
2433 majollika-kakelugnar, införskrivna från den kunglig  
1319 majoren, där ett förberedande förhör anställdes.  
370 majorer i riktiga uniformer med guld och silver. J  
3024 Majorum exemplis eller Ärftilgheten utan Moral.  
3462 Majorum exemplis? Genom förfäders lysande föredöm  
2115 majstång, bestående av en naken stång påklädd med  
555 maka, barn och pigor anslände till 18gerplatsen i F  
2413 make man aldrig förr sett och som tävlar i skönbet  
151 makt att skänka något; ni hurrar åt kungen, därför

FIGUR 3. *Ur konkordans till Strindbergs Det nya riket.*

Herr talman! I går kväll hann jag om en utredning. // Utskottsminoriteten, som r än avgiftsbilgömg. På den frågan har jag fall alla vara överens om. Samhället värderar s. Indonestien erövrade sin självständighet men eten på dessa utbildningsområden, men det har till stånd en rekonstruktion av företaget har vandrare. Men realistiskt sett vet vi att det rvägsnätets utformning och omfattning har det hjälpa de värst drabbade. På kort sikt kan vi dmedsagsprediknan, men verkligheten låter sig n. förövändning sedan kommer och säger att den

tyvärr inte fullfölja mitt anförande inom den utmätt  
tyvärr inte funnit anledning att gå in i debatten hä  
tyvärr inte fått något svar från kommunikationsminis  
tyvärr inte föräldramas insatser utifrån den utgång  
tyvärr inte gav frihet åt den lilla kolonin på Finor  
tyvärr inte gett några utslag från riksdagens sida.  
tyvärr inte givit något resultat. Konkursförvaltaren  
tyvärr inte går att dra i gång en sådan verksamhet ö  
tyvärr inte gått att nå enighet i utskottet. Centern  
tyvärr inte göra mer än att visa mänsklig solidarite  
tyvärr inte göras mera presentabel. // Man kunde väl  
tyvärr inte hann eller att lagrådet tog så lång tid

023077-28  
010008-33  
034087-02  
154032-13  
049043-21  
127072-06  
042040-41  
116035-05  
164003-20  
102021-07  
124008-13  
068023-09

FIGUR 5. *Ur kollokationskonkordans till riksdagsspråk.*

Man behöver ju inte stanna vid orden utan kan gå till ordkombinationerna, kollokationerna som jag gärna kallar dem. I figur 5 har vi en konkordans som tar fasta på dem. Den gäller svenskt riksdagsspråk. Det framgår också att man kan bygga ut beläggsuppgiften så, att man anger vem som står för det aktuella belägget, vad vederbörande har för politisk hemvist, vilket kön och vilken ålder talaren har osv. Det möjliggör givetvis detaljerade studier av många olika slag. *Tyvärre inte* är en av de viktigaste kollokationerna, i varje fall en av de vanligaste, i Sveriges riksdag.

låt	Ftext Flex	
	976	68
låt/låt 1 'ljuda'	135	21
låt nn -en	12	
låt a vb -ø	83	
låt skriv a nde nn -t	1	
låt e nn -t	3	
beat låt nn -en	1	
folk låt s mo tiv nn -et	1	
hit låt nn -en	1	
pop låt nn -en	2	
standard låt nn -en	1	
tyst låt en av -t	3	
tyst låt en het nn -en	1	
ut låt a vb -ø	1	
ut låt a nde nn -t	16	
beat - låt nn -en	1	
sirap s låt nn -en	1	
Waters - låt nn -en	1	
de charge ut låt a nde nn -t	1	
låk ar ut låt a nde nn -t	1	
50 - tal s låt nn -en	1	
sak kunn ig ut låt a nde nn -t	1	
tjän st e ut låt a nde nn -t	2	
låt 2 'lämna'	31	10
kvar låt en skap s skatt nn -en	1	
upp låt a vb -ø	7	
upp låt a nde nn -t	1	
över låt a vb -ø	15	
över låt else nn -n	1	
över låt else av tal nn -et	1	
över låt else dat um nn -et	1	
ur akt låt a vb -ø	2	
ur akt låt en het 'nn -en	1	
ur akt låt en het s synd nn -en	1	
låt 3 'utgjuta'	4	2
åder låt a vb	1	
åder låt ning nn -en	3	
låt 4 'främja oförhindrat förlopp'	736	16
låt - gå - system nn -et	1	
låt a vb -ø	546	
låt a nde nn -t	1	

FIGUR 6. *Ur morfosemantisk konkordans i Nusvensk frekvensordbok 4.*

Det som presenteras i figur 6 är ett resultat av en mer ingående analys. Denna typ av konkordans har antagligen inte många varit i kontakt med ännu. Den är publicerad i fjärde delen av Nusvensk frekvensordbok. Det är en morfosemantisk konkordans. I botten ligger en morfematisk analys genomförd på formella grunder. De morfem som framkommer analyseras sedan semantiskt. Vi kallar de resulterande enheterna polysemkomponenter. De blir fyra stycken när det gäller elementet *låt*. De parafraseras såsom 'ljuda', 'lämna', 'utgjuta' och 'främja oförhindrat förlopp'. Den morfosemantiska konkordansen har den karaktären att alla lemman som innehåller den parafraserade semantiska enheten är representerade med sin analys under denna. Man kan alltså se precis vad som är gjort och använda det, om man tycker det är relevant, annars ändra analysen.

<i>låt</i> <i>nn -en</i>	12
<i>låt</i>	6
<i>låtar</i>	5
<i>låten</i>	1
<i>låt-gå-system</i> <i>nn -et</i>	1
<i>låt-gå-systemet</i>	1
<i>låta</i> <i>vb -ø</i>	629
<i>låt</i>	126
<i>låta</i>	163
<i>låten</i> <i>imp</i>	1
<i>låter</i>	175
<i>låtit</i>	56
<i>lät</i>	108
<i>låtande</i> <i>nn -t</i>	1
<i>låtanden</i>	1
<i>låtsa</i> <i>vb -ad</i>	20
<i>låtsade</i> <i>ptp plu</i>	1
<i>låtsades</i> <i>prt</i>	1
<i>låtsas</i> <i>inf</i>	6
<i>låtsas</i> <i>prs</i>	11
<i>låtsats</i> <i>sum</i>	1
<i>låtsasdjur</i> <i>nn -et</i>	1
<i>låtsasdjuret</i>	1
<i>låtsaslek</i> <i>nn -en</i>	1
<i>låtsassand</i> <i>nn -en</i>	1
<i>låtsassanden</i>	1
<i>låtskrivande</i> <i>nn -t</i>	1
<i>låtskrivandet</i>	1

FIGUR 7. Ur lemmatiserad ordlista i Nusvensk frekvensordbok 2.

'hårfrisyr  
 "beatlesfrisyr  
 "hövolmsfrisyr  
 'blesstyr  
 "småblesstyr  
 dressstyr  
 'hunddressyr  
 'kammardressyr  
 'fältadressyr  
 'hästdressyr  
 miniatyr  
 karikatyr  
 advokatyr  
 'satyr  
 'appretyr  
 polityr

'schellackpolityr  
 'shellackpolityr  
 'möbelpolityr  
 'nagelpolityr  
 'snickarpolityr  
 garnityr  
 'tandgarnityr  
 'möbelgarnityr  
 "wienerschnit-  
 'zelgarnityr  
 'pärlgarnityr  
 'skinn-garnityr  
 'ledargarnityr  
 'skrivbordsga-  
 rnityr

FIGUR 8. Ur Svensk baklängesordbok.

MUSIC:1/B =D  
 VOICE:1/ =D  
 SEL:1/A , WHICH MAKES:2/B A RINGING:32/W  
 STRIKE:2/2 .P  
 RING:2/WA =D CAUSE:2/ A BELL:1/ TO  
 RING:2/WB :R  
 BLEW:2/VB =E BLOW:2/VB A HORN:1/WA =P  
 RING:2/WC =D  
 SOUND:1/W PRODUCED:32/1 WITH:7/H ART:1/2 2633  
 SOUND:1/W PRODUCED:32/3 IN:7/5 SPEAKING:12/A AND:6 4348  
 SOUND:1/W WHEN:4/B STRUCK:32/1 0355  
 SOUND:2/WA 3934  
 SOUND:2/WA 3336  
 SOUND:2/WA / ( OF:7/D A BELL:1/ 3337  
 SOUND:2/WA =D CAUSE:2/ TO MAKE:2/ A NOISE:1/ 3773  
 SOUND:2/WA BY:7/A MEANS:1/W2 OF THE BREATH:1/ 0415  
 SOUND:2/WA OUT:4/C CLEARLY:43/1A LIKE:4/V A BELL:1 3338  
 SOUND:2/WB =D EXAMINE:2/A A THING:1/L EG MAN:S:1 3774  
 SOUND:2/Y =D LONG:3/V NARROW:3/\*A PIECE:1/ OF WATE 3776  
 SOUND:2/Z =D MEASURE:2/ THE DEPTH:1/A OF WATER:1/ 3769  
 SOUND:3/VA :N UN+32BROKEN:9/2 3770  
 SOUND:3/VB :P HEALTHY:3/\*A 3771  
 SOUND:3/VC :P STRONG:3/AK 3771

FIGUR 9. Ur en lexikalisk konkordans efter Petöfi.

Denna konkordanstyp förutsätter att man har gjort en lemmatisk analys i ett tidigare skede. Man sammanför variantformer under någon form av stickord som figur 7 visar.

Ett i huvudsak automatiskt system som skall lemmatisera en text kan man utforma med eller utan lexikon. Gör man det utan lexikon, som vi gjorde i Göteborg för 15 år sedan, får man i vissa fall sådana fenomen som att ordet *major* av systemet uppfattas som pluralis, varvid grundformen *maja* ansätts: en *maja*, flera *major*.

Har man sitt ordmaterial i maskinläsbar form, kan man bland annat sortera det ordslutsvis (figur 8), vilket är nyttigt för många uppgifter. Den raka högerkanten hör till det som markerar att det rör sig om en baklängesordbok, finalalfabetisk ordbok eller, som det heter i det danska fallet där Henrik Holmboe står bakom, retrograd ordbok.

Nu är det omvänt så, att man kan se lexikonet som en text. Man kan sålunda göra konkordanser på lexikonet, vilket också det ofta är mycket användbart. Figur 9 är hämtad ur en konkordans som baserar sig på ett arbete av János Petöfi rörande definitionsvokabulären i Michael Wests engelska lexikon. West uppger att han definierar alla sina ord med hjälp av 1200 enheter. Det visar sig emellertid vara enheter i mycket speciell mening. Petöfi har undersökt de olika betydelseinnehållen i alla beläggen på de ord som ingår i definitionerna. När det exempelvis i lexikonet står *blow a horn* med parafrasen 'sound by means of the breath' har Petöfi lagt in en analys av vilken betydelse hos *sound* det är fråga om. Det framgår att de olika beläggen ger en mängd olika analyser. Vi noterar substantivet, verbet och adjektivet med varierande innehåll. Antalet enheter i definitionsvokabulären får multipliceras med 4 eller 5 för att ge en semantiskt acceptabel bild.

På samma sätt är naturligtvis själva definitionerna i ett lexikon att betrakta som text, och de kan göras till föremål för grammatisk analys, parsing. I figur 10 ses ett resultat av en körning i en parser av en sådan definition. Det som definieras som 'formellt krav på beslut om viss åtgärd' är det svenska ordet *yrkande*. Parsar man det, får man ut strukturen, som utgörs av en nominal-



fras som består av ett attribut, ett substantiv och en prepositionsfras som i sin tur har en viss byggnad osv. En poäng med detta är att man i strukturen kan se vilket som är huvudordet, och huvudordet är ju genus proximum i lexikologisk terminologi. När man gör liknande analyser av definitionerna i ett lexikon, får man möjlighet att etablera pekare i sin databas mellan huvudorden, som ju bär upp centrala delar av de lexikaliska enheternas betydelser. Man kan då få kedjor eller slingor i lexikonet, som sammanhåller begreppsmässigt relaterade ord, kort sagt ett slags tesaurus.

FORMELLT KRAV PÅ BESLUT OM VISS ÅTGÄRD

```

(N00567 D1)
  (N00551 NP1)
    (N00422 ATTR1)
      (T00421 AV FORMELLT)
      (T00425 NN KRAV)
      (N00534 PP1)
        (T00435 PREP PA)
        (N00527 NP1)
          (T00436 NN BESLUT)
          (N00522 PP1)
            (T00460 PREP OM)
            (N00518 NP1)
              (N00463 ATTR1)
                (T00461 AV VISS)
                (T00515 NN ÅTGÄRD)

(N00563 D1)
  (N00547 NP1)
    (N00422 ATTR1)
      (T00421 AV FORMELLT)
      (T00425 NN KRAV)
      (N00532 PP1)
        (T00435 PREP PA)
        (N00526 NP1)
          (T00437 NN BESLUT)
          (N00522 PP1)
            (T00460 PREP OM)
            (N00518 NP1)
              (N00463 ATTR1)
                (T00461 AV VISS)
                (T00515 NN ÅTGÄRD)

```

FIGUR 10. *Resultat av parsning av en definition.*

En term som har kommit i svang inom datalingvistikens under de senaste åren är *taggning*. Den svarar mot engelskans *tagging*. Vi har tagit tjuren vid hornen och anpassat benämningen till svenska. Vi *taggar* texter, och vi sätter *taggar* på enheterna i texterna.

Egentligen har man hållit på att tagga i alla datalingvistiska år. Det är bara det att man inte har kallat det så tidigare. Vad det är fråga om är helt enkelt att man mer eller mindre automatiskt lägger på resultatet av någon form av analys på de olika elementen i en text eller ett lexikon.

Exemplet i figur 11 återgår på Magnus Ljungs projekt SEMTAG i Stockholm. Det går ut på att lägga information från Longmans Dictionary of Contemporary English på orden i ett par engelska textkorporusar. Markeringarna innebär att det första ordet är ett egennamn, det andra och fjärde är entydiga i lexikonet och det tredje uppvisar den första betydelsen hos *primary* 1. Andra sätt att göra taggningen kan vara att stoppa in taggarna som ett slags bihang till de ord de hänför sig till eller att göra en uppställning i kolumnform. Alla de här modellerna och fler till förekommer i de olika projekten.

Atlanta's	recent	primary	election
npr	x	lsl	x

Atlanta's/npr	recent/x
primary/lsl	election/x

Atlanta's	npr	---
recent	x	---
primary	lsl	---
election	x	---

FIGUR 11. *Taggning av text efter Ljung.*

I Norge har man arbetat mycket med taggning. Stig Johansson i Oslo och hans medarbetare har i samarbete med bergensarna taggat den s.k. LOB-korpusen – Lancaster-Oslo-Bergen Corpus – som är en korpus över brittisk engelska. I Göteborg har Alvar Ellegård arbetat med att tagga Brown Corpus, som är motsvarande amerikanska material. I Lund har Jan Svartvik och hans

medarbetare taggat London-Lund-korpusen, som är en tal-språkskorpus baserad på det material som Randolph Quirk har samlat in. Där har också tillämpats en intressant probabilistisk modell för taggning som Mats Eeg-Olofsson har utarbetat. I Göteborg skall vi börja ett projekt kallat SynTag, vars syfte är en analys av huvudord och argument i satserna i texter ur den korpus som Nusvensk frekvensordbok bygger på (Press 65).

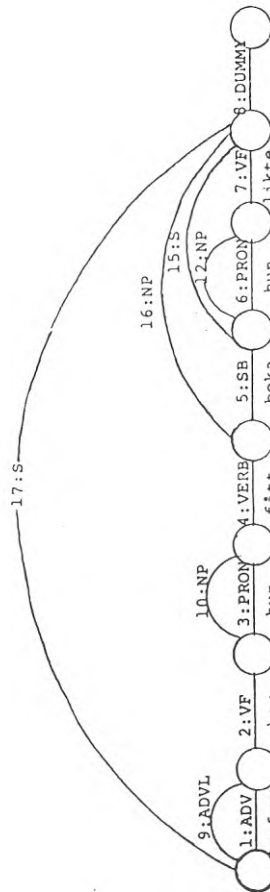
Som jag nämnde bildar de olika forskningsuppgifterna ett continuum. Man kan inte klart skilja studiet av texter från studiet av lexikon, taggning från parsning, databaser från filsystem osv. Snarast är det i stor utsträckning fråga om olika tyngdpunkter. Jag skall nu göra ytterligare en förskjutning. Från den angivna forskningsstrategins synpunkt har jag anledning att också gå in lite närmare på parsning.

Jag skall kanske först säga att vi sedan början av 1970-talet har haft en serie sommarskolor i språklig databehandling. Det började i Göteborg (Marstrand) och har fortsatt i Bergen, Köpenhamn och Reykjavík. Vi har dessutom haft en skandinavisk workshop i Palo Alto nära San Francisco. En förhoppning är att vi snart skall komma till Finland i denna serie. En central lärare vid sommarskolorna har varit Martin Kay. Hans huvudintresse ligger på det morfologiska och syntaktiska analysplanet. Han är framför allt intresserad av att göra experimentsystem med vars hjälp man kan utveckla sådan metodik.

Den här verksamheten har satt spår också på det sättet, att det hålls symposier i flera av länderna, regelbundet i Danmark och Sverige, då tillfälle ges även på andra håll. Vi har vidare en serie som heter Nordiska datalingsvistikdagarna. Den närmaste sammankomsten hålls hösten 1983 i Uppsala hos Anna Sågvall Hein.

En av dem som fångade upp idéerna från sommarskolorna var Svein Lie i Oslo i samverkan med bl.a. Knut Hofland i Bergen. Svein Lie har skrivit en rapport om automatisk syntaktisk analys, som jag kan rekommendera som en pedagogisk introduktion. Hanne Ruus har som framgår av hennes bidrag liksom Bente Maegaard också arbetat med den här metodiken. Den data-

struktur som används kallas för *chart*, och efter den benämns detta slags parsning *chart analysis*. Strukturen innehåller som figur 12 visar noder förenade med bågar på vilka man kan lägga lexikalisk information, som man sedan utnyttjar i sin grammatik för att bygga upp satslösningar, ofta alternativa lösningar.



FIGUR 12. *Chart efter Lie.*

I Sverige har Anna Sågvall Hein arbetat med en vidareutveckling av denna formalism. Det finns en parser i Uppsala med grammatik, lexikaliska inslag och fungerande program. Den är testad på meningar i en invandrarlärobok i svenska. Det finns också idéer om hur den skall kunna byggas ut. Den presenteras som och är en experimentell parser. I vilken utsträckning denna typ av syntaktisk textanalys kan användas på större material är fortfarande en öppen fråga.

En som anser att man skall gå helt andra vägar är Benny Brodda i Stockholm. Han talar i stället om heuristisk parsning. Tillvägagångssättet blir att ta fasta på signalerna i den syntaktiska ytstrukturen och basera parsern på dem. Han har utvecklat ett system kallat Beta, som mycket kort kan beskrivas som ett slags substitutionsgrammatik baserad på en s.k. finite state automaton. Det är alltså fråga om regler som innebär att någonting omskrivs som någonting annat i en viss kontext.

Med utnyttjande huvudsakligen av formorden och ändelserna i texten kan man komma fram till en analys som den i figur 13. Den presenterades på det svenska kollokviet i språklig databehandling 1983. Den analys som exemplet är taggat med har erhållits efter åtta regelomgångar, dvs. sedan systemet körts åtta gånger med olika innehåll varje gång. Det enda ord som inte har fått någon analys är *undersökt*. Vissa morfologiska gränser har kunnat identifieras och en prepositionsfras har kunnat etableras (markerad med ett plus).

Vad	ska	ni
conj	aux	pron
gör'a	med+den	
v-inf	prep-p.	
fråg=ade	småbruk=arn	
v-fin	nom	
som	under>sökt	
conj	o	
om	den	dög
conj	nom	v-fin

FIGUR 13. Resultat av heuristisk parsning efter Brodda & Källgren.

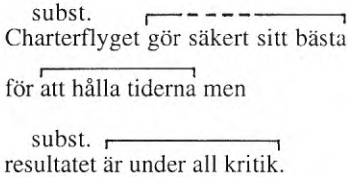
Brodda menar att hans system är betydligt snabbare än chart-analys. Det blir intressant att se hur långt man kan komma med de enkla medel som det är fråga om här. Det finns inte någon som helst semantisk information i systemet. Det finns heller inget lexikon i det, och det är oklart hur ett sådant skulle kunna inlemmas. När det gäller körningar på finsk morfologi rapporterar emellertid Brodda i samarbete med Fred Karlsson mellan 80 och 90 % träffsäkerhet. Det man inte får veta vid en körning är vilka enskilda fall som har analyserats rätt respektive fel. En generell osäkerhet kvarstår alltså, men vid vissa tillämpningar är man ändå hjälpt av resultatet.

En annan typ av grammatisk textanalys som jag vill nämna är den som har redovisats av Bente Maegaard och Ebbe Spang-Hanssen. Den gäller segmentering av skriven fransk text, nämligen segmentering i satser. Maegaard och Spang-Hanssen arbetar också med ytsignaler men dessutom med ett verbstamslexikon på ungefär 5000 enheter. Det är identifieringen av det finita verbet som står i centrum. Man rapporterar att man har testat systemet på olika texttyper och kommit upp i en säkerhet av 95 %, förutsatt att man inte arbetar med de allra svåraste texterna, särskilt inte med texter skrivna av de Gaulle.

Det finns andra parsningprojekt. I Umeå har man också börjat arbeta med en finite state-modell, fast i ett annorlunda utformat system. Det är Eva Ejerhed i samarbete med Kenneth Church som står för detta. Från Helsingfors förväntas en hithörande avhandling utarbetad av Kimmo Koskenniemi. Även Koskenniemi arbetar med en finite state automaton. Vad han gör är att utveckla ett generellt system för morfologisk analys av naturliga språk, som han själv tillämpar på finska.

I Göteborg har vi också arbetat en del med parsningens problem. Vi har tagit fasta på vad man inte brukar göra i textgrammatisk analys, nämligen lexikonets roll, ett stort lexikons möjligheter. Vad som antyds i figur 14 är, att om man utgår från de heterografa orden och de återkommande ordförbindelserna, täcker man mycket av texten. Man får en ganska tät serie öar, som naturligtvis är försedda med grammatisk information att bygga vidare på. Det framstår som rimligt att göra en lexikalisk

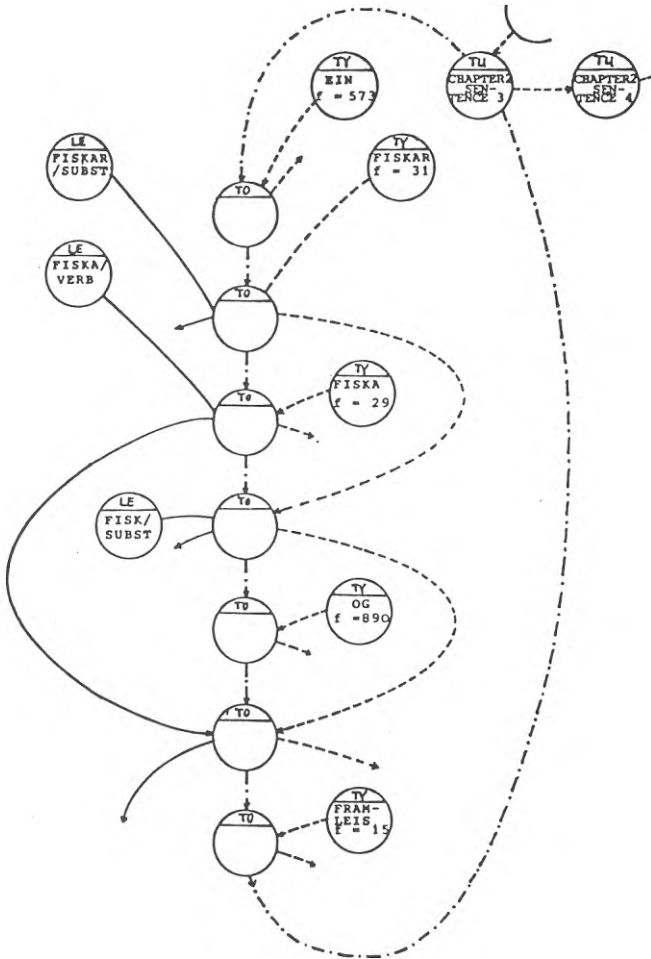
parsning baserad på ett stort lexikon med bland annat detta innehåll.



FIGUR 14. *Heterografer och rekurrenta kollokationer.*

Men det är inom datalingvistikens inte bara fråga om att arbeta med meningar, satser, kollokationer, ord och morfem, utan man kan ju sträcka sig längre ut i texterna. Betrakta figur 15 som jag har hämtat ur en uppsats av Michael Gillow, som under åtskilliga år samarbetade med Kolbjørn Heggstad i Bergen, när de banade väg för datalingvistikens i Norge. Den visar en modell för hur man i ett databassystem kan lagra en text med information om inte bara textens linjära struktur utan också dess semantiska och lexikaliska innehåll. En nod i figuren talar som synes om att man befinner sig i en viss textenhet, i det här fallet kapitel 2, mening 3. Därpå förs man till de på varandra följande orden i meningen: *ein fiskar fiska fisk* osv. Man kan fortsätta att följa slingan genom texten, men man kan alltså också uppsöka andra typer av information, t.ex. lemmatisk eller kvantitativ, som är fogad till de textuella ordens noder.

En annan form av nätverk kan illustreras av figur 16. Det är en tämligen informell representation av början på *Candide*, gjord av Bengt Sigurd. Den visar de storheter som är verksamma i texten, deras relationer till varandra, deras attribut osv. Här finns alltså ett slott som är beläget i Westfalen, en baron som har en ägarrelation till detta och som har namnet Thunder-Ten-Tronckh, en lady som har egenskapen att vara ung osv. Man lägger in allt detta i nätverket och försöker fylla ut det på bästa sätt genom slutsatsdragning. Textprogressionen får man fram genom att följa den markerade slingan.



FIGUR 15. Textuell och lexikalisk databas efter Gillow.





Skärningsområdet mellan datalingvistik och textlingvistik aktualiserades vid det Nobelsymposium jag anordnade i Stockholm 1980 med rubriken *Text Processing*. De föredrag och diskussionsinlägg som förekom där finns utgivna i en volym med samma namn. Frågor om textrepresentation diskuteras av János Petöfi, Roger Schank, Yorick Wilks, Umberto Eco, Sven Öhman, Bengt Sigurd, Östen Dahl och andra. Inlägg om textanalys och textgenerering görs av Michael Halliday, Charles Fillmore, Maurice Gross, Martin Kay, Ulf Teleman, Jan Svartvik, Nils-Erik Enkvist och andra. Texttypologi och textattribution behandlas av Joseph Grimes, Teun van Dijk, Robert Longacre, Wolfgang Dressler, Alvar Ellegård, Hans Karlgren och andra.

Hans Karlgren, som jag just nämnde, har i stor utsträckning intresserat sig för lingvistiskt orienterad informationsökning. Han har bland annat varit inne på att man inte i semantiskt avseende exakt skall behöva specificera det man söker utan kunna göra en approximation av det innehåll man är intresserad av. Han kallar detta fenomen för *homeosemi*. Det behandlas i en uppsats som står i en bok han har utgivit tillsammans med Donald Walker och Martin Kay och som just heter *Natural Language in Information Science*. För närvarande arbetar han med problem kring lagring av och sökning i lagtexter. Tillsammans med Benny Brodda har Karlgren också utvecklat en teknik för beskrivning av likhet mellan varunamn.

Hittills har jag mest uppehållit mig vid textanalys av många olika slag. Jag skall ge en liten glimt av textgenerering också. Nyligen hade jag anledning att göra en demonstration och valde då att utforma ett system som genererar haikudikter. En haikudikt är ju så väl strukturerad formellt sett att den består av tre rader med fem, sju respektive fem stavelser. Vad som behövdes var att skriva en liten speciellt avpassad svensk grammatik och att göra ett begränsat lexikon som huvudsakligen innehåller ord som suggererar centrallyrisk stämning, litegrann av det man förväntar sig av en haikudikt. Sedan skrev Christian Sjögreen ett program som använder den lilla grammatiken och lexikonets ungefär 150 ord plus en slumptalsgenerator. Denna väljer dels de regler som skapar de grammatiska strukturerna, dels de lexi-

kaliska enheter som skall sättas in i de terminala noderna. Varje gång man trycker på en viss tangent genereras en ny dikt på terminalens textskärm, t. ex.

Ord dröjer stundom.  
Gräs skall lysa under bron.  
Sången klagar vagt.

En bidragande orsak till att jag tar upp den här genereringen är att jag vill understryka, att det inte är datamaskinen som har skrivit dikten utan jag. Maskinen är enbart mitt medium. Jag skulle ha kunnat sitta vid skrivbordet helt utan maskin och åstadkomma precis samma dikt på basis av grammatiken och lexikonet. Synpunkten har som lätt inses generell betydelse.

Ett projekt där man arbetar med textgenerering genomförs i Lund av Bengt Sigurd och Jan Fornell. Det kallas Commentator. Man har ett lexikon på omkring trettio ord och en grammatik som medger att man kan uttala sig om lägen i ett koordinat-system. Två markeringar kallade Adam och Eva rör sig i förhållande till varandra och till ett par fasta punkter. Systemet kan säga: Eva är nära porten, Adam närmar sig trädet osv. Det är än så länge ganska enkelt, men projektet har intressanta utvecklingsmöjligheter på både det semantiska och det syntaktiska planet.

Så är det dags att gå lite närmare in på de semantisk-lexikaliska frågorna och säga något om en lexikalisk databas som är under arbete. Det gäller i det här projektet att utforma betydelsebeskrivningar av ungefär 75 000 centrala ord och ange deras fraseologi, morfologi, stavning och uttal. I arbetet används en semantiskt grundad definitionsvokabulär av ungefär 5000 enheters storlek. Den har tagits fram på olika vägar under projektets gång. Med dessa enheters hjälp definieras omkring 25 000 enheter och parafrazeras alltså omkring 50 000.

Ett sådant arbete kräver förutom att man utvecklar en definitionsvokabulär dels att man skapar definitionsformat, alltså bestämda sätt att definiera olika typer av substantiv, adjektiv, verb osv., dels att man utarbetar ett system med vars hjälp man

kan urskilja de olika lexemen i det underliggande ordmaterialet. Vi arbetar med begreppet kärnbetydelse och olika regler för hur en kärnbetydelse kan utvidgas och specialiseras. Mycket av informationen hämtas naturligtvis in på automatisk väg, men definitionerna och en del annat läggs in i ett fast inkodningsformat genom interaktiv bearbetning. Olika personer kan i ett sådant system göra olika saker. En blir specialist på lemmafrågor, alltså den formella delen av enheternas karakterisering, andra arbetar med definitionerna, någon ägnar sig åt uttalet osv. De olika typerna av information lagras i olika filer: en lemmafil, en definitionsfil, en frasfil m.m. och en sammanhållande identitetsfil som pekar på de olika element i de olika filerna som har med samma enhet att göra. Avsikten är att senare omforma filsystemet till en databas i teknisk mening.

Projektet har dubbelt syfte. Den lexikaliska beskrivningen av ordmängden kan föras vidare i en formalisering, som efterhand skall kunna hjälpa oss att göra textanalys av det slag som jag talade om tidigare. Dessutom har projektet emellertid det praktiska målet att avkasta en svensk ordbok. För det ändamålet har vi utformat en algoritm som ur de aktuella filerna plockar ut dels de enheter som skall ingå i det tryckta lexikonet, dels de uppgifter om dessa enheter som skall tas med. När utskriftsalgoritmen har fungerat, får man på papper en sådan utskrift som i figur 17. Den ser naturligtvis inte ut så som den tryckta versionen kommer att göra, men den är bra som arbetsunderlag. Man ser stickordet (*ljud*), ordklassangivelsen, morfologiska uppgifter, definitionen med definitionstillägg, morfologiska exempel (*ljudeffekt* etc.), specialiseringar inom kärnbetydelsens ram, syntaktiska exempel och efter ett särskilt märke ett idiom, nämligen *nu blev det annat ljud i skällan* med parafrasen 'nu ändrade man attityd'. Därefter kommer första verblemmat *ljuda* och andra verblemmat *ljuda* med sina olika uppgifter osv.

En lexikalisk databas av detta slag har många andra användningar. Den är i sig själv ett forskningsobjekt. Man kan undersöka vokabulärens komplexitet på olika plan, det intressanta förhållandet mellan morfologi och semantik, olika typer av semantiska relationer mellan enheterna, tesauristiska klasser osv. Detta

kan i sin tur komma inte bara det fortsatta datalingvistiska arbetet utan också annan forskning tillgodo.

ljud subst. -et = \* fysikaliskt fenomen som ger upphov till hörsselförnimmelser genom att sätta öronens trumhinnor i rörelse; såväl om fenomenet i sig som om enskilda förekomst därav: *ljudeffekt; ljudstyrka; missjud; strupljud; välljud; inte ett - hördes; -et av krossat porslin; -ets hastighet är 330 m/sekund; bilen gav ifrån sig ett märkligt - \* spec. om ljud med språklig funktion vokal eller konsonant: *ljudlära; a-ljudet; omljud; utljud \* ibl. med spec. syftning på innehållet ord: han var så förvånad att han inte fick ur sig ett - \* sv. friare livstecken: vi har inte hört ett - från honom på flera månader # nu blev det annat - i skållan nu ändrade man attityd**

<sup>1</sup>ljud\A verb *ljöd -it \* avge (kraftigare) ljud <SYN. höras, låta>: genljuda; de hörde basunerna -; hennes ord ljöd ännu i hans öron*

<sup>2</sup>ljud\A verb *-ade -at \* uttala (ord) ljud för ljud vid läseövning e. d.*

FIGUR 17. *Ur automatiskt genererat manus till Svensk ordbok.*

Den abstrakta forskningsstrategi som jag sammanfattade under rubriken språkbank kan givetvis också konkretiseras. Att driva en faktisk språkbank kräver rätt mycket övervägande. Bland annat är urvalsproblematiken stor. Numera är det inga problem att få så mycket maskinläsbart textuellt och lexikaliskt material man önskar. Problemet är att göra ett relevant urval. Man bör täcka olika genrer i lagom stora mängder. Olika typer av lagrat material bör kunna kombineras för olika ändamål. Man måste vidare tänka på tillgängligheten. Förutom texter och lexikon bör språkbanken också innehålla bearbetningar av råmaterialen, utförda inom dess egen ram. Till detta kommer att den bör omhänderta maskinläsbara resultat av relevanta forskningsprojekt.

Det finns en språkbank i Bergen, Norsk textarkiv, som är ett samarbetsprojekt mellan PDS (Projekt för datamaskinell språkbehandling) och NAVFs EDB-senter. Det finns material i Köpenhamn hos Hanne Ruus och Bente Maegaard i projektet Danord eller DANWORD. Det finns material i Århus, i Trondheim, i Stockholm, i Lund och på andra ställen, men det finns inga aktiva organiserade språkbanker på fler ställen än i Bergen

och Göteborg. Som en underavdelning av språkbanksbegreppet kan man se termbankerna. Vi har Danterm i Köpenhamn, Termdok i Stockholm och Norsk termbank i Bergen. Termer är för mig ett specialfall av ord, och det motiverar att jag nämner termbankerna i detta sammanhang.

Låt mig också göra en kommentar om korpus och introspektion, en av de frågor som Erik Andersson är inne på i sitt bidrag. Jag skulle vilja säga att en empirisk undersökning inte skall börja med datainsamling. Man bör ha teoretiska föreställningar, idéer av hypoteskaraktär innan man börjar samla in data i ett forskningsprojekt. Vidare vill jag hellre tala om introspektion än intuition som pendang till korpus. Jag är helt enig med Erik Andersson om att korpus och introspektion är två aspekter som är förenliga. Korpus är resultatet av många informanternas språkproduktion, introspektion resultatet av en informants språkproduktion, nämligen vederbörande forskares egen. Sett från data-lingvistisk synpunkt tillkommer en tredje aspekt: experiment.

Självfallet kan man också göra diakroniska iakttagelser i materialen. Jag skall ge två exempel, det första av mikrotyp. Som jag har påpekat i annat sammanhang befinner sig det svenska verbet *betala* i morfologisk utveckling. Det är böjningen i preteritum och framför allt i supinum det är fråga om. I tidnings-språkmaterial från 1965 och 1976 heter det *betalade* i båda fallen men huvudsakligen *betalt* i det förra respektive *betalat* i det senare. Verbet tycks tendera i riktning mot ren första konjugation. Det skall bli intressant att se hur det ser ut i mitten på 1980-talet.

Man kan också arbeta på ett makroplan, och det för mig in på Svenska Akademiens ordbok. Den håller på att läsas in med hjälp av en apparat som heter Kurzweil Data Entry Machine. Det är en avancerad optisk klartextläsare. Ordboken (A-S) omfattar 28 volymer som innehåller drygt 200 miljoner tecken. Texten är tryckt i åtta fonter. Avsikten är att läsa in hela materialet och lagra det i en välstrukturerad form, så att man kan få svar på frågor av olika typ. Att jag nämner det i det här sammanhanget beror på att en uppenbar användning är att göra historiska studier, bland annat studier av betydelseutveckling. Natur-

ligtvis kan man göra lånordsstudier, morfologiska studier etc. också, men jag pekar särskilt på detta. Man kan tillfoga att 1520-talet, som är den bakre gränsen för SAOB, inte är någon vattendelare när det gäller ordens betydelseutveckling. Det är befogat att också överväga inmatning av fornspråkliga lexikon och texter för att på så sätt få ett sammanhängande helt bort till runstenarna.

Låt mig till sist ge en ansats till en blick in i hjärnan. Jag kommer tillbaka till de rekurrenta ordförbindelserna. De har fånglat mig under lång tid, och jag har skrivit om dem i olika sammanhang, inte minst i inledningen till tredje delen av frekvensordboken. Jag vill framhålla att jag inte bara syftar på idiom i trängre mening, utan att det är fråga om kollokationer av alla möjliga slag. De har egenskapen att tillsammans vara så vanliga att de bygger upp stommen i alla texter. En del är idiom, dvs. semantiskt oförutsägbara i förhållande till de ingående orden, men huvuddelen är av mera rättfram natur, helt enkelt ordinära kombinationer som uttrycker vissa ting enligt vanliga grammatiska regler. Det är bara det att det är just så det brukar uttryckas i språket. Kollokationerna utgör en avgörande komponent i förmågan att tala flytande. Till kretsen kring *låta* hör: *låt oss anta att, låt oss börja, låt oss säga, men låt oss inte gå händelserna i förväg, låta påskina, låta uppföra, låta ana, låta sig påverkas, inte låta vänta på sig, låta meddela att* osv.

Man finner en sådan överväldigande mängd av dessa förbindelser i alla texter man studerar, att jag i början på 1970-talet drog slutsatsen att det språkliga medvetandet måste ha dem aktuella. Det har viktiga implikationer, menar jag, vad gäller språk teori och språkbeskrivning. I inledningen till tredje delen av frekvensordboken kallar jag constellationerna för block. Det visar sig att man får direkta metaspråkliga referenser till dem i texter: *Han är chefens förlängda arm, som det heter.* Det påpekas alltså att det är så det brukar uttryckas. *Detta är s.k. analys genom syntes* innehåller en annan typ av referens. Ibland anger man att man citerar: *Strängt taget handlar alla böcker om läsaren, för att citera Olle Hedberg.* Jag har inte klart för mig hur lexikonet i denna vida mening är organiserat i hjärnan men

skulle gärna vilja veta det. För närvarande ter det sig som ett slags strukturerat kaos.

Jag deltog i den första konferensen i den här serien, den som hölls 1969 i Reykjavík. När Hreinn Benediktsson tog kontakt med mig den gången, enades vi om att jag skulle tala om Vocabulary Data Processing. Det handlade då om vissa tidiga resultat på ett delområde. När Kristian Ringgaard tog kontakt med mig inför denna femte konferens 1983, gällde det att ge en statusrapport om skandinavisk datalingvistik. De båda ämnenas formuleringar tycker jag antyder att det har runnit en del vatten under broarna under de år som har gått.

Man kan notera att datalingvistiken har etablerat sig som ett ämnesområde med akademiska tjänster. Man arbetar på undervisning i datamatisk lingvistik här i Århus. Det finns sådan i Köpenhamn. Bergen har planer i samma riktning osv. Datalingvistiken har också ställt redskap och resultat till förfogande för nordiska språk, som är den ena komponenten i den här konferensserien. Jag tycker mig se att datalingvistiken har utövat ett visst inflytande även på den allmänna lingvistiken. Denna var avogt inställd till datalingvistiken i början. Men när man nu ser en utveckling från en mera axiomatisk modell till ett mera språknära tänkande i den allmänna lingvistiken, kan det i varje fall hänga samman med datalingvistiken.

Det finns också en tendens till en mera processuell inriktning inom lingvistiken i motsats till den tidigare mera statiskt strukturella inriktningen. Kanske är det, som t.ex. Nils Erik Enkvist nämner, ett paradigmskifte som börjar skymta. Nu är det rentav så att en del allmänna lingvister vill klä sig i datalingvistisk dräkt. Det är naturligtvis fullt i sin ordning, förutsatt att man sätter sig in i vad som krävs på området.

Man kan fundera över vad som skulle finnas att rapportera på en konferens i denna serie om ytterligare något decennium, om det skulle slumpa sig så, att det blev aktuellt med ett inlägg då också. Det står var och en fritt att göra sina spekulationer med utgångspunkt från nuläget. Utan att gå närmare in på saken vill jag för min del säga, att det finns anledning till vissa förhopp-



ningar. Men – för att sluta där jag började – det kommer nog inte att handla om något fullfjädrat femte generationens system.

## Summary

As indicated, e.g., by the research and development plans of the huge Japanese project called Fifth Generation Computer Systems, natural language has attracted more and more attention in the computational field in recent years. In consequence, the significance of computational linguistics is becoming increasingly obvious.

This linguistic field involves research and education aimed at improving our insight into natural language through deeper theoretical and practical knowledge of computational methods of linguistic analysis, synthesis, and communication. A resulting system can be regarded as a kind of (imperfect) simulation of a language user. Such a system needs linguistic and extralinguistic resources. It has to be rigorously explicit, form a functioning whole, and be based on solid facts about language use. An overall research strategy can be founded on the abstract notion of a sophisticated language bank as a framework for text decoding. Vast new linguistic domains are opened up.

Textual and lexical investigations of various kinds form a gliding scale. A text is, in a way, a dictionary of itself and can be accessed as such. A variety of concordances have been devised, however, in order to refine the organization of the raw material. They can focus on graphic words, lemmas, collocations, morpho-semantic units, etc. They can be lineprinted or presented on microfiche, or made available in an interactive mode. As linguistic tools, concordances are very helpful.

Advanced types of concordances are often based on the lemmatization of graphic words. Inflectional forms and variants are then brought together under lexical entries by means of more or less automatic procedures. Quantitative investigations of this type of linguistic units as well as other specified types of units are also undertaken.

Inversely, a dictionary can be regarded as a text. Concordances of dictionaries can reveal interesting relations between entries. Being texts, definitions can be parsed, which means, among other things, that head-words can be identified. Thus, pointers can be established and thesauruslike classes set up.

There are several methods of tagging texts. This term means that the result of some kind of analysis is attached to the linguistic units of a text (or a dictionary). Often grammatical information is inserted but semantic or other information can also be treated in this way.

Parsing or sentence analysis is an adjacent branch. A number of formalisms and algorithms have been employed. One of these is chart analysis designed to obtain all the possible solutions in each case. Another is a heuristic approach taking advantage of morphological signals in the surface structure and emphasizing speed and approximation. A third gives priority to the identification of the finite verb. The potential of lexically based parsing is also being explored.

There are certain models for making use of database techniques in text processing. Textual, lexical, semantic, and quantitative information can be combined by such means. Information retrieval along linguistic lines should also be mentioned in this connection.

Some work on text generation has been done. Speech synthesis has been included in the set-up.

In the lexicosemantic domain, there is work in progress on a lexical database. This requires, among other things, the development of a system for the identification of lexemes, a defining system, and a defining vocabulary. It should provide a useful basis for further formalization aiming at a lexical component for use in text processing. In fact, it is also a semantic research object in its own right.

Language banks have been organized in some places. They contain and supply machine-readable texts and lexical material as well as other kinds of processed material. It is pointed out that evidence from corpus work, introspection, and experiments all contribute to linguistic progress.

Diachronic observations can of course be made in the large bodies of material at hand. In addition, printed sources like historical dictionaries can be fed into the computer by means of optical reading. This makes it possible to study semantic change, etc., in a new manner.

One of the lasting impressions of studies in the field is the fundamental role of recurrent collocations of all kinds in language ability and, thus, in text processing. They are also repeatedly referred to by speakers and writers in metalinguistic comments.

Computational linguistics has been established as an academic field. There are institutes, posts, dissertations, etc. It seems that this field has had an influence on general linguistics, where trends towards more authentic evidence on one hand and procedural models on the other can be perceived.

## Litteratur

- Allén, Sture (under tryckning): Inte bara idiom. I Anna Sågwall Hein (ed.): *Nordiska datalingvistikdagarna 1983*. Centrum för datorlingvistik, Uppsala.
- Allén, Sture (under tryckning): Språkdata Lexibase System. An integrated view of a lexical project. *Proceedings of European Science Foundation Workshop 1981*. Pisa.
- Allén, Sture (ed.) 1982: *Text processing*. Data linguistica 16. Almqvist & Wiksell International, Stockholm.
- Allén, Sture 1980: The language bank concept. I Joseph Raben & Gregory Wilks (ed.): *Data bases in the humanities and social sciences*. North-Holland, Amsterdam, s. 171–176.
- Allén, Sture et al. 1970–80: *Nusvensk frekvensordbok 1–4*. Data linguistica 1, 4, 10, 14. Almqvist & Wiksell International, Stockholm.
- Brodda, Benny (under tryckning): Heuristisk satsanalys. *4:e svenska kollokviet i språklig databehandling*. Institutionen för lingvistik, Lund.

- Brodda, Benny & Karlsson, Fred 1981: *An experiment with automatic morphological analysis of Finnish*. Department of General Linguistics, Helsingfors.
- Eeg-Olofsson, Mats 1982: En språkstatistisk modell för ordklassmärkning i löpande text. I Gunnel Källgren (ed.): *Taggning. Föredrag från 3:e svenska kollokviet i språklig databehandling*. Institute of Linguistics, Stockholm, s. 27–30.
- Ejerhed, Eva & Church, Kenneth (under tryckning): Finite state parsing. I Fred Karlsson (ed.): *Papers from the Seventh Scandinavian Conference of Linguistics 9*. Helsingfors.
- Fornell, Jan (under tryckning): Commentator – ett mikrodatorbaserat forskningsredskap för lingvister. *4:e svenska kollokviet i språklig databehandling*. Institutionen för lingvistik, Lund.
- Gavare, Rolf 1979: Konkordanser. *Föredrag från 2:a svenska kollokviet i språklig databehandling*. Survey of Spoken English, Lund, s. 7–16, 51 ff.
- Gillow, Michael 1979: Linguistic data-bases. I Sture Allén & János Petöfi (ed.): *Aspects of automatized text processing*. Helmut Buske, Hamburg, s. 19–35.
- Järborg, Jerker 1982: SynTag. I Gunnel Källgren (ed.): *Taggning. Föredrag från 3:e svenska kollokviet i språklig databehandling*. Institute of Linguistics, Stockholm, s. 31–39.
- Koskenniemi, Kimmo (under tryckning): *Two-level morphology: A general computational model for word-form production and recognition*. Helsingfors.
- Källgren, Gunnel: Satsanalys utifrån morfologiska kriterier: Att skutta mellan nivåer. *4:e svenska kollokviet i språklig databehandling*. Institutionen för lingvistik, Lund.
- Lie, Svein 1980: *Automatisk syntaktisk analyse. 1. Grammatikken*. NAVFs EDB-senter, Bergen.
- Ljung, Magnus 1982: SEMTAG-projektet. I Gunnel Källgren (ed.): *Taggning. Föredrag från 3:e svenska kollokviet i språklig databehandling*. Institute of Linguistics, Stockholm, s. 49–57.
- Maegaard, Bente & Ruus, Hanne 1978: DANWORD. Hyppighedsundersøgelser i moderne dansk: Baggrund og materiale. *Danske Studier* 1978 (februar), s. 42–70.

- Maegaard, Bente & Spang-Hanssen, Ebbe 1978: *Segmentation automatique du français écrit*. Editions Jean-Favard, Paris.
- Outline of research and development plans for Fifth Generation Computer Systems*. Institute for New Generation Computer Technology, Tokyo. 1982.
- Sigurd, Bengt 1977: Om textens dynamik. *Papers from the Institute of Linguistics, University of Stockholm* 34. Stockholm.
- Sågvall Hein, Anna 1983: *A parser for Swedish*. Centrum för datorlingvistik, Uppsala.
- Walker, Donald E., Karlgren, Hans & Kay, Martin (ed.) 1977: *Natural language in information science. Perspectives and directions for research*. Skriptor, Stockholm.

## Översättningsdatorn – finns den?

Mänskligt och omänskligt i informationsteknologin. 1984.

I den svenska texten stod det:

Hur sköter ni er hud?

Förmedlat till engelska av en (mänsklig) översättare med gott hjärta blev det:

How do you look after your dog?

Människan läser ju gärna helheter, gestalter, vilket kan leda fel, även om det sällan resulterar i fadäser av det här slaget. Datamaskinen däremot arbetar sig fram tecken för tecken och gör därigenom knappast sådana fel. Frågan är emellertid vad den i övrigt kan trimmas till, när det gäller översättning av naturligt, dvs. mänskligt språk.

Konstigt nog var översättning en av de första tillämpningar man tänkte på när datamaskinen hade konstruerats på 1940-talet. Det satsades efterhand rätt mycket på olika projekt, och en del system utvecklades. De var i huvudsak ord-för-ord-system. När ett ord hade flera motsvarigheter i målspråket, drog man gärna till med en hel serie och lämnade valet åt läsaren:

In experiments on/in/at/to mammals the special/particular/peculiar attention/consideration was given to the research/analysis/investigation of the state/condition/position of the system of the blood/hemogenesis formation, the determination/definition/decision of intermediate products ...

Hade man problem med ordmotsvarigheter, kunde man låta källspråkets ord stå:

The short-time orbital flights below of the radiational belts of earth in the absence that was raised by the SOLNECNOI one of activity in a radiational attitude are BEZOPASNYMI.

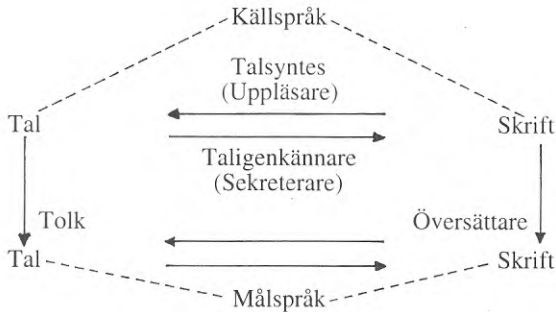
Det finns över huvud taget inget par av språk som uppvisar något ett-till-ett-förhållande i fråga om ord. Inte ens när det gäller en så relativt välbestämd ordklass som räkneorden är det så. Här är en liten glimt av de två närbesläktade språken svenska och engelska:

två böcker	two books
halv två	half past one
båda två	both of them
två gånger	twice
på tvåhundralet	in the third century
tvåspråkig	bilingual
tvåbokstavsgrupp	digraph
två sina händer	wash one's hands (of it)

Försöken att komma tillrätta med svårigheterna genom diverse ad-hoc-lösningar hade ringa framgång. I mitten på 1960-talet stod det klart, att man inte var rustad att hantera de språkvetenskapliga problemen, och projektanslagen drogs i stort sett in. I gengäld blommade den lingvistiska forskningen inom syntax och semantik upp.

Om man ser översättning som överföring av information mellan olika språkformer, kan man konstatera att mönstret är mer sammansatt än man kanske i förstone tänker på. Både källspråk och målspråk har ju både en talad och en skriven form. Se figur.

Översättning i vanlig mening innebär att man går från ett språks skriftform till motsvarande form i det andra språket. En tolk brukar däremot omsätta källspråkets tal till tal i målspråket. Från datamaskinell synpunkt medför denna typ av informationsöverföring särskilt stora problem. Trots betydande ansträngningar på åtskilliga håll är möjligheterna till taligenkänning mycket starkt begränsade. Svårigheterna vållas bland annat av att närstående ljud delvis artikuleras samtidigt, att vissa ljud eller ljudföljder undertryckes helt, att ord och ordsträngar mycket ofta är mångtydiga och att olika individer har högst varierande röstläge. I fråga om talsyntes, generering av tal, har man kommit längre, även om mycket återstår att göra där också.



Låt oss titta på några andra av de besvärligheter som översättare och tolkar – och därmed programsystem – ställs inför. Jag nämnde nyss språkets mångtydighet, och det är den fundamentala svårigheten.<sup>6</sup> Ambiguiteten finns på alla plan. Här är ett lexikaliskt exempel:

*bar* 'ställe där starka drycker serveras'  
 'ställe med självservering av mat'  
 'enhet för tryck'  
 'naken'  
 'lyfte och flyttade på'

Det finns naturligtvis många fler nyanser än så: *på bar gärning* osv. Ofta(st) hjälper den omgivande texten till för att klara ut betydelsen, men inte sällan är närkontexten otillräcklig: *hon slank in på en bar*.

Också längre ordsträngar är ofta svåra att reda ut. Hur många olika kategorier av personer syftar man på i vart och ett av de här fallen:

långa och stiliga män  
 långa och stiliga män och kvinnor  
 långa och korta män  
 långa och korta män och kvinnor

Det som krävs är en specificering av vad som gäller i varje enskilt belägg i en text.

Inte minst här i trakten utnyttjas ju mångtydigheten i språket till vitsar.



- Hur kan ni sälja kött så billigt?  
 –Vi har skurit ner personalen.

På engelska skulle *skära ner* i betydelsen 'reducera' ungefär svara mot *cut down*, medan 'stycka' snarast skulle svara mot *cut up*. Det systematiska utnyttjandet av språkets mångtydighet är ett av skälen till den stora svårigheten att översätta mästervitsaren Shakespeare.

Ofta dyker nya ord upp, inte minst sammansättningar, och det gäller att ha beredskap för att klara dem. Förleder och efterleder kan förhålla sig på många sätt till varandra. Några otvetydiga regler har vi inte tillgång till.

sandlåda	innehåller sand
trälåda	gjord av trä
	innehåller trä
krabblåda	har krabba som huvudingrediens
	innehåller krabbor
draglåda	kan dras ut
	används för avdunstning
koklåda	används för upphettning
brevlåda	används för ankommande brev
	används för avgående brev
resonanslåda	ger genklang
boklåda	är en bokhandel
	innehåller böcker
språklåda	avser talförmåga
	är en typ av radioprogram

Var och en kan själv åstadkomma många fler *lådor* med mer eller mindre mångtydigt innehåll.

Många gånger finns det ingen motsvarighet alls i målspråket. Det kan bero på skillnader i natur, kultur, samhällsskick osv. Prova att översätta *lussekatt* och *lutfisk*! Och tänk på hur man skulle bära sig åt för att förmedla hela den mängd av bibetydelser och associationer som knyter sig till uttryck som *dra i långbänk* och *det hade jag ingen aning om!*

Även om det finns en motsvarighet kan det lätt gå fel. Tag som exempel *pfalziska tronföljdskriget* (1688–1697). På engelska heter *pfalzisk Palatine* och *tronföljdskrig war of succession*. Men det finns inget krig som kallas *the Palatine War of Succession*. Engelsmännen har tagit fasta på en annan aspekt

och kallar det *the War of the Great Alliances*. Vad det i själva verket gäller är att ha tillgång till en stor fond av kunskaper om världen i gången tid och nutid, allt dessutom sett ur olika folkgruppers olika perspektiv.

Ett speciellt kruz är all den implicita, ytligt sett utsagda, informationen i texter. *Det drar* kan i själva verket betyda *var snäll och stäng fönstret, ge mig papperet* kan betyda *jag är beredd att skriva under, är han nykter idag?* kan innebära *han brukar vara onykter*, vilket i sin tur kan ha stor betydelse för analysen av texten i övrigt osv. För att komma tillrätta med detta behöver man bygga upp en analysapparat, som kan dra slutsatser ur texten och på den grunden skapa en informationsmängd som förutsättning för det fortsatta arbetet med texten. Det finns bara mikroskopiska ansatser till sådana system.

En summarisk sammanställning av några av de moment som måste behandlas vid en översättning från tal till tal kan se ut så här:

taligenkänning resp. talsyntes  
lexikonsökning  
ordböjning  
ordbildning  
satsbildning  
meningsbildning  
underförstådd information  
kunskapsgrund  
betydelseinnehåll

Dessa faktorer och flera andra leder till ståndpunkten att en helt automatisk översättning av hög kvalitet och oberoende av texttyp inte kan väntas. Det är inte bara en fråga om att satsa pengar. Språket är ett mentalt fenomen och som sådant svår-fångat.

Under de senaste åren har vi likväl sett ett ökande intresse för maskinöversättning. Det sker t.o.m. en viss marknadsföring. Man kan fråga sig vad denna nya trend beror på.

Orsakerna är flera. Den lingvistiska semantiken har utvecklats ganska markant. Datalingvistikens metodik har vidgats, fördjupats och effektiviserats. Inom datalogin har området

artificiell intelligens kommit med intressanta bidrag. Datorernas kapacitet har fortsatt att förbättras i raskt tempo. Och behovet av översättning i stor skala har accentuerats, t.ex. inom den europeiska gemenskapen, i det tvåspråkiga Canada, i det marknadsinriktade Japan och supermakterna emellan.

Jag kan nämna två mycket uppmärksammade projekt. Det utan jämförelse största av dem är det japanska Fifth Generation Computer Systems. På det satsar man flera miljarder kronor under en tioårsperiod. Kontaktytan mot omvärlden skall enligt planen vara ett interface som ger tillgång till systemet genom naturligt språk (tal, skrift, bilder). De tillämpningar som man framför allt siktar på är språköversättning, frågebesvarande system och taligenkänning.

Det andra projekt jag har i tankarna är den europeiska gemenskapens satsning på Eurotra. Arbetet är upplagt så, att översättning skall ske från vart och ett av de berörda språken till vart och ett av de övriga språken inom gemenskapen. Hitills har den huvudsakliga delen av insatsen gjorts i fråga om utformningen av en formalism, som skall kunna användas genomgående i projektet.

Enligt min erfarenhet råder det en ganska långtgående enighet om att översättning kan dra stor nytta av datalingvistiska metoder, även om en helautomatisk, högkvalitativ översättning av all slags text inte ligger inom räckhåll.

Interaktion mellan människa och maskin är grundläggande för alternativa utformningar. Med en sådan samverkan som utgångspunkt kan man skissa flera olika modeller. Jag skall antyda ett par.

En modell skulle man kunna kalla Translab. Idén är att med hjälp av datalingvistiska metoder åstadkomma en så god arbetsmiljö som möjligt för en mänsklig översättare. Detta innebär bland annat att översättaren från sin terminal snabbt skall kunna slå i ordbanker (inklusive termbanker), utnyttja textredigeringsprogram, databehandla källtexten för att exempelvis lokalisera likartade textpartier och samhörig vokabulär, genom dokumentsökning få fram likartade texter osv. Detta ger inte bara större effektivitet utan också högre säkerhet. Det har t.ex.

visat sig vara mycket svårt för översättare att hålla en konsekvent terminologi.

En annan modell är den s.k. synergistiska. Den gäller det fallet att källtexterna skall översättas till flera språk. Här sätter man in de datamaskinella och mänskliga resurserna på att åstadkomma en representation av texten som är tillräckligt bra för att kunna tjäna som gemensam utgångspunkt för maskinell överföring till olika målspråk. De resulterande översättningarna kontrolleras av experter på respektive språk (posteditering).

Sammanfattningsvis kan sägas, att översättningsdatorn inte finns och att något stort genombrott i fråga om helt automatisk översättning av hög kvalitet och gällande all slags text knappast är att vänta. Till att börja med måste vi utesluta de litterära texterna. Sedan är det nödvändigt att inrikta sig på begränsade ämnesområden. Inom sådana ramar kan dock intressanta resultat nås, i synnerhet om människan på ett eller annat sätt inlemmas i systemet. Men det krävs mycket mer forskning.

Betrakta inte detta som en pessimistisk syn. För begränsade praktiska ändamål kan också mindre framsteg bli nyttiga. Och vi lär oss mycket om språket.

Att fullvärdig översättning i princip är en omöjlig uppgift – för människor och därmed för programsystem – är något vi får leva med.

## Inte bara idiom

De nordiska datalingvistikdagarna 1983. 1984.

Detta skall handla om kollokationer eller ordförbindelser. Vissa grundfrågor inom språkvetenskapen upphör inte att fångla en, och kollokationerna har för mig varit en sådan grundfråga. Bland mycket annat har de ett intressant förhållande till parsning.

Jag minns med stor intensitet mina intryck, när jag såg de första resultaten av konkordanskorningar för ganska många år sedan. De öppnade nya språkliga vyer. Man såg i ett huj vilken betydelse kollokationerna måste ha i den språkliga aktiviteten. Detta har varit en viktig utgångspunkt för ständigt uppdaterade tankar kring kollokationer.

En av de första jag stötte på som över huvud taget hade tänkt i de här banorna med datamaskinen i perspektivet var John Sinclair (1970). Han menade att man skulle ta fasta på vad som är statistiskt signifikant för att få fram kollokationerna. De ord som uppträdde tillsammans oftare än man hade anledning att förvänta sig, sett mot en slumpmässig bakgrund, skulle ha kollokationell karaktär. Kriteriet gav både relevanta och irrelevanta (*if take* osv.) förbindelser. Resultatet av arbetet var således inte övertygande i detta avseende, inte heller för Sinclair själv. Också andra experiment har sedan visat att det naturligtvis inte är så enkelt, även om det är en del av sanningen.

De första tankarna med utgångspunkt i en allmän uppfattning om språket och om vad konkordanserna visade ledde till uppläggningsen av arbetet på tredje delen av Nusvensk frekvensordbok (1975) och presentationen av resultaten. Jag talade under 1970-talets första hälft bland annat i Åbo, Pisa och London om kollokationerna och deras roll i språket (föredragen trycktes 1973, 1976 respektive 1977).

Det som var utgångspunkten för tredje delen av frekvensordboken var iakttagelsen att kollokationerna ständigt återkommer, är rekurrenta i den mening som anges i inledningen till

frekvensordboken. Villkoret i denna undersökning var att det skulle finnas minst två identiska belägg på en miljon löpande ord. Med det villkoret fick vi ut 660 000 belägg på ordkombinationer. Det var kanske den första överväldigande siffran för oss. Den vanligaste av alla förbindelserna var *det är*, sedan kom *och den*, *sig i*, *än att*, *på ett helt annat sätt* osv., alltså välformade och icke välformade ordkombinationer om vartannat.

Ur de 660 000 beläggen analyserade vi fram vad vi kallade för konstruktioner. På dem lade vi villkoren att de skulle vara grammatiskt styrda och lexikaliskt selekterade. De utgör sålunda ett urval ur kombinationerna. Vi fick 50 000 olika konstruktioner, som vi klassade i 17 olika huvudtyper: *den stora frågan*, *i linje med*, *ta form*, *sköta om*, *kommer att fortsätta*, *för att undersöka*, *är på väg*, *mycket ung*, *även om*, *om också*, *för att*, *in i*, *men låt oss inte gå händelserna i förväg*, *det är givet att*, *kort sagt*, *som människa*, *jo då*. Dessa exempel markerar de olika typerna.

Ur konstruktionerna gjorde vi i sin tur ett urval. Det gällde idiomerna i ordets snäva bemärkelse, dvs. de kollokationer som har oförutsägbar betydelse sett från de ingående ordens betydelsers synpunkt. Vi tillämpade kriterierna ganska strängt och fick fram 300 olika idiom: *ge sig i kast med*, *i elfte timmen*, *lägga sista handen vid*, *det är inte utan att*, *göra avkall på*, *slå dövörat till* osv. Idiomerna utgör alltså en mycket liten del av den stora mängden kollokationer.

I den nämnda inledningen påpekar jag, att det är ganska tydligt att resultaten bör få konsekvenser för språkteorin. Lexikonet ser antagligen inte ut på det sätt som man har tänkt sig tidigare, t. ex. inom den generativa inriktningen. Vad man kan kalla lexikaliska block av många olika slag bör finnas i språkmedvetandet. Strax efter utgivningen av frekvensordbokens tredje del publicerade Jan Anward och Per Linell (1976) en uppsats om lexikaliserade fraser i svenskan. De tog fonologi, grammatiska egenskaper m.m. i betraktande och sammanfattade sitt resonemang så här: varje enskild konstituent i en lexikalisk enhet (dvs. i en kollokation) kan inte exploatera sin betydelse fullt ut och kan inte syntaktiskt varieras i någon större

utsträckning; kan inte ha egen referens (säger man *sparka boll*, syftar man inte på en speciell boll); kan inte böjas fritt; kan inte modifieras fritt; kan inte fritt befrågas, negeras eller affirmeras; kan inte varieras med avseende på ordföljd och prosodi i någon större utsträckning; kan å andra sidan utmärkas av morfologiska och andra oregelbundenheter.

Om man tittar lite närmare på detta, ser man att flera av villkoren är av sådan karaktär att de karakteriserar vad jag vill kalla idiom snarare än kollokationer i allmänhet. Men slutsatsen är i alla fall densamma som den jag hade dragit från andra utgångspunkter, nämligen att lexikonet spelar en större roll än man kanske trott och att grammatiken spelar en något mindre roll.

Charles Ruhl och Adam Makkai förde en diskussion om idiom i en volym från 1976. Utgångspunkten var att Makkai hade skrivit en bok om *Idiom structure in English*. Ruhl menade att ett fel med den boken var att Makkai inte hade tagit hänsyn till den omedvetna delen av språkmanifestationen. Han tog som exempel *take off* som ju på engelska betyder 'become airborne' men som i sin nominaliserade form också kan betyda 'parody'. Makkai menade att man här har två olika lexikaliska enheter medan Ruhl efter en genomgång av ett hundratal belägg på *take off* noterar hur man från den ena betydelsen gradvis kommer över i den andra.

På detta svarar Makkai genom att ta fram alla ord som i engelskan slutar på *-ic(k)* (*bic, brick, chick, dick* osv.) och tillsammans med sina studenter resonera sig fram till hur vart och ett kan anknytas till dels 'liv' eller 'död', dels 'skatter'. Med andra ord: från vad som helst kan man resonera sig fram till vad som helst. Det är ett sätt att ironiskt avfärda den tanke som Ruhl hade. Man måste försöka identifiera de olika lexikaliska enheterna, menar Makkai. Han använder rentav benämningen *institutions* för vissa av de här kollokationerna – de har så att säga en stadfast roll i språket.

Ungefär samtidigt hade vi i Göteborg ett mindre projekt som hette Algoritmisk textanalys. Jag förbigår här det som gjordes på programsidan och i någon mån på syntaxsidan och nämner

de lexikaliska delresultaten. Vi utarbetade ett paradigmmärkt baslexikon på ungefär 8000 enheter. Det är publicerat av Staffan Hellberg i en bok från 1978. Baslexikonet upptar stam, uppslagsform och paradigmnummer och täcker därmed i princip hela morfologin. Vidare upprättades ett speciallexikon över heterografer (ickehomografer) på ungefär 900 enheter. Ett tredje resultat var ett speciallexikon över disambiguerande kollokationer på omkring 1600 enheter. Det omfattar sådana kollokationer som innehåller homografer som blir disambiguerade genom att ingå i kollokationerna. Ett exempel är *komma hem*, där *komma* kan vara verb eller substantiv och *hem* kan vara substantiv eller adverb, medan *komma hem* är entydigt. Till detta kom ett ändelselexikon för sannolikhetsklassificering av ord som saknas i baslexikonet. Ett av de ständigt återkommande problemen är ju att man stöter på nya ord.

Bland det mest intressanta var att de två speciallexikonerna på 1600 respektive 900 enheter vid körningar visade sig täcka 50 procent av en okänd text av normal typ. Det är tankeväckande från parsningens synpunkt. Jag kan tillägga att de 8000 enheterna i baslexikonet täcker i runda tal 90 % av en text. Då skall vi emellertid komma ihåg att de innefattar en mängd homografer. Det fina med de två speciallexikonerna är att de ger stycken av fast mark som analysen kan bygga vidare på.

Låt mig nämna ytterligare några som på olika sätt har arbetat med kollokationer. En av dem är Harald Burger som har publicerat en intressant bok om idiom (1973), där han framför allt är inne på de teoretiska problem som knyter sig till begreppet. Maurice Gross (1982) har undersökt vad han kallar "frozen sentences", vilket också är ett slags kollokationsbegrepp. I hans fall har det gällt franska. Göran Kjellmer vid engelska institutionen i Göteborg är sysselsatt med en genomgång av hela Brownkorpusen för att ta fram kollokationsmaterialet ur den. Syftet är främst att utarbeta en frasordbok.

Så är vi framme vid Lexikalisk databas. Det är det största projektet vid Språkdata för närvarande. I det definierar vi omkring 75 000 lemman ur det moderna svenska språket och ger



uppgifter av många olika slag, bl.a. just beträffande fraseologi och idiomatik.

Här är ett utdrag ur kollokationsuppgifterna rörande lemmat *land*, som representerar tre olika lexem (lexikaliska enheter baserade på kärnbetydelser). Jag ger inte deras definitioner, utan vi kan se på exemplen vad de avser. Till det första lexemet hör *inom landets gränser*, *flytta till ett annat land*, *de afrikanska länderna* och idiomerna *det heliga landet* eller *det förlovade landet* och *skuggornas land*. Det andra omfattar *en sjöman går i land*, *land i sikte*, *på torra land*, *färdas till lands* och idiomerna *förstå hur landet ligger*, *gå i land med något* och *nu går skam på torra land*. Det tredje har *hon var från landet*, *resa till landet under veckoslutet* och idiomerna *ingen mans land*. En av tankarna med projektet Lexikalisk databas är just att man med utgångspunkt från i första hand kollokationerna och definitionerna skall arbeta vidare i riktning mot ett lexikon för parsning. I existerande system är de såvitt bekant, liksom lexikaliska uppgifter över huvud, svagt företrädda. Låt mig ändå i detta sammanhang nämna Kaplan & Bresnan (1980), Sager (1981) och Limmermann, Kroupa & Keil (1983).

På senare tid har jag kommit att uppmärksamma en annan typ av indikationer som jag anser vara viktig från såväl teoretisk som praktisk och psykologisk synpunkt. Det gäller vad jag kallar för de metaspråkliga kommentarerna i texter. Det är alltså så att språkbrukarna själva i viss utsträckning talar om hur deras lexikon ser ut, något som man kanske inte har uppmärksammat tidigare. De metaspråkliga kommentarerna gäller rätt ofta enskilda ord men inte sällan just kollokationer. Man markerar dem med uttryck av typen *som det heter*, *som det så vackert heter* osv. Ett litet urval exempel följer: *ett rörligt intellekt, som det heter*; *karavanen rör sig trots att hundarna skäller, som det heter i ett gammalt arabiskt ordstäv*; *administrativ databehandling, som det heter*; *ett förslag till, som det så vackert heter*, *en förenklad deklarationsblankett*; *bidrag för att, som det heter*, *förbättra konkurrensmöjligheterna på den internationella marknaden*; *hon är väl död vid det här laget, om*

*man så säger; vi nådde alltså, för att lätt travestera ett slitet uttryck, ända fram.*

Själva de metaspråkliga uttrycken är rikt varierade och baserar sig på en rad olika verb: *för att använda en kliché, för att använda ett gammalt ordspråk, för att använda ett slitet uttryck, för att använda herr NNs egen formulering; för att citera NN; som det heter, som det numera heter, som det brukar heta, som det så vackert heter; som man brukar kalla det, som det kallas, så kallad; som frasen lyder, som uttrycket lyder; som man säger, så att säga, om jag så säger, som det brukar sägas; för att travestera ett gammalt uttryck; om man så vill; om uttrycket tillåts, om man så får uttrycka saken, som NN uttrycker det.* Alla de metaspråkliga kommentarerna syftar inte alltid på kollokationer i den mening jag avser här, men mycket ofta visar det sig vara fallet. Man kan dela in dem i några huvudtyper. *Som det heter* syftar ofta på en kurant kollokation. *Som X lyder* syftar gärna på ett ordspråk eller ordstäv. *För att citera* anger direkt källan: en författare, en lagtext, en förordning eller någonting sådant. *För att travestera* kan i princip syfta på alla de olika typerna. Som travesti åberopar den indirekt en kollokation av något slag.

De metaspråkliga kommentarer som man kan plocka fram på det här sättet ger ytterligare en inblick i människans lexikon. De fogar sig också till de tidigare typerna av kollokationella kriterier som har varit rekurrensen, de grammatiska konstruktionskriterierna, idiomkriteriet och de tillkommande lingvistiska kriterier som Anward och Linell har pekat på. De bidrar alltså till att ge en ny bild av lexikonet och följaktligen också en ny bild av hur vi fungerar i språkproduktionen och i perceptionen. Därmed är de av fundamentalt intresse vid utvecklingen av parsningsystem.

## Litteratur

Allén, Sture, Om fraser i svenskan. (Svenskans beskrivning 7. Ed. Christer Hummelstedt. Åbo 1973, s. 24–31.)

- Allén, Sture, On phraseology in lexicology. (*Cahiers de lexicologie* 29 (1976), s. 83–90.)
- Allén, Sture, Text-based lexicography and algorithmic text analysis. (*ALLC Bulletin* 5: 2 (1977), s. 126–131.)
- Allén, Sture, et al., Nusvensk frekvensordbok baserad på tidningstext. 3. Ordförbindelser. 1975.
- Anward, Jan, & Linell, Per, Om lexikaliserade fraser i svenskan. (*Nysvenska Studier* 55–56 (1975–76), s. 77–119.)
- Burger, Harald, *Idiomatik des Deutschen*. Tübingen 1973. (*Germanistische Arbeitshefte*. Ed. Otmar Werner & Franz Hundsnurscher. 16.)
- Gross, Maurice, Simple sentences. (*Text processing. Proceedings of Nobel Symposium* 51. Ed. Sture Allén. 1982, s. 297–315.)
- Hellberg, Staffan, *The morphology of Present-Day Swedish. Word-inflection, word-formation, basic dictionary*. 1978.
- Kaplan, R.M., & Bresnan, J.W., *Lexical-functional grammar: a formal system for grammatical representation*. Occasional Paper 13. MIT Center for Cognitive Science. Cambridge, Mass. 1980.
- Makkai, Adam, Idioms, psychology, and the lexemic principle. (*The Third Lacus Forum*. Ed. Robert J. Di Pietro & Edward L. Blansitt, Jr. Columbia, South Carol. 1976, s. 467–478.)
- Ruhl, Charles, Idioms and data. (*The Third Lacus Forum*. Ed. Robert J. Di Pietro & Edward L. Blansitt, Jr. Columbia, South Carol. 1976, s. 456–466.)
- Sager, Naomi, *Natural language information processing: a computer grammar of English and its applications*. Reading, MA 1981.
- Sinclair, J. McH., Jones, S., & Daley, R., *English lexical studies*. Department of English. Birmingham 1970.
- Zimmermann, Harald H., Kroupa, Edith, & Keil, Gerald, *CTX. Ein Verfahren zur computergestützten Texterschliessung*. Saarbrücken 1983.

## Översåtarnas språk – finns det?

Mänskligt och omänskligt i informationsteknologin. 1984.

Den röda gubben vid övergångsstället, vem är han? En avlägsen släkting till Storebror? Du står kvar fast gatan ligger tom. Eller – tar du ditt öde i egna händer och korsar den?

Trafikljus fanns naturligtvis på Orwells tid. Men faktum är att television och datamaskiner också fanns, liksom för den delen atombomber. Det är kanske inte främst teknologiskt som Orwells roman imponerar utan psykologiskt, som skildring av ett inre tillstånd präglad av andlig otrygghet och ofrihet. Också härvidlag fanns det åskådningmaterial i tiden, närmast Stalins totalitära Sovjet och den nyss störtade Hitlers tyska diktatur. I den litterära bakgrunden fanns en mästare som Kafka.

Förhållandet mellan översåtar och undersåtar kan ta sig uttryck på många olika sätt. Här skall vi först och främst se det från språklig synpunkt. Men låt oss då komma ihåg, att intet mänskligt är språket främmande.

*Översåte* är nytt som ord men gammalt som begrepp. Ordet har kommit i bruk under vårt århundrade som en motsvarighet till *undersåte*, på liknande sätt som *underdrift* har bildats efter *överdrift*.

Vilka är då översåtarna? Kort sagt är det de som har makt, men mera de som har makt i det stora än i det lilla. De finns bland företrädarna för regering, riksdag, kommuner och lands-ting, bland partipolitikerna, bland administratörerna, bland intresseorganisationernas ledare och bland massmediernas företrädare.

Journalisterna sticker t.ex. inte under stol med att de har makt – ”tredje statsmakten”. Den framställs då gärna som riktad mot överheten. I den mån så är fallet är det fråga om en nytthet. Emellertid kan man märka ett tilltagande partipolitiskt engagemang. Nyheter och kommentarer flyter in i varandra. Den gamla journalistiska regeln att endast publicera vad som är ”sant och intressant” får konkurrens av sin karikatyr:

”Kontrollera aldrig en intressant uppgift – den kan visa sig vara osann.” Därmed minskas trovärdigheten. Det är en stor förlust vi gör om och när pressen själv blir en del av överheten.

Debattklimatet och därmed språkklimatet omkring oss är hårt och oförsonligt. I det allt häftigare ordflödet tror man sig hela tiden behöva öka trycket, höja tonen och utplåna nyanserna. Öppna en dagstidning vilken som helst och läs: *eländigt, botten, urholkad, gigantisk, självgodhet, häpnadsväckande, social nedrustning, hyckleri, passar inte i tryck, fanatism, vårdslöst, oansvarigt, grova övertramp, stötande, katastrofal, uppseendeväckande, flagrant, magstarkt, överkörd, vulgaritetens retorik, patologiskt, förhastat, oöverlagt, beklämmande* etc.

En av de bärande linjerna i Orwells Newspeak var ju att tillhandahålla ett språk, som var ett uttryck för överhetens världsbild och som gjorde alla andra tänkesätt omöjliga. Texter inom en annan ideologisk ram skulle vara oöversättbara. De kunde endast fångas upp i det samlande ordet *crimethink*. Det yttersta målet var att talet skulle komma från struphuvudet ensamt, utan inkoppling av hjärnan. Beteckningen för detta idealtillstånd var *duckspeak* ’snattra som en anka’. En verkligt god talare var en *doubleplusgood duckspeaker*.

Det som aktualiseras här är grundläggande mentalhygienska frågor. Kan språket styras? Hur kan det i så fall utnyttjas?

Det har länge funnits en tendens i språket att ord med särskilt känsliga bitoner efterhand förbrukas. Ett av de oftast citerade exemplen är *dårhus* som ersattes av *sinnessjukhus* som har ersatts av *mentalsjukhus*. Till denna tendens har man t.ex. knutit an inom kriminalvården, där man terminologiskt sett inte längre har några *fångar*: *fängelset* har blivit *kriminalvårdsanstalt*.

På det sociala området har vi fått många exempel på den här sortens ordbyten. Officiellt finns det inte längre några *fosterhem*, bara *familjehem*. I dessa bor de egna barnen och vad som tidigare kallades *fosterbarn*. I lagen finns inga beteckningar med vilkas hjälp man kan skilja dem åt. Även om man noterar

en human strävan och ett försök till positiv attitydförändring, ser man ju vad det rör sig om: styrning av tänkandet och utplåning av ord.

I linje med detta ligger utbytet av *övervakare* mot *kontaktperson*. Här har humaniseringen drivits så långt, att den valda beteckningen inte associeras med begreppet. När språkbrukaren i gemen upptäcker manövern, kan det mycket väl tänkas att *kontaktperson* i sin hittills normala betydelse blir oanvändbart. Det kommer då också att drabba ordets användning i det sociala sammanhanget. Så blir det dags att finna ett nytt ord igen, och svänghjulet snurrar vidare. Ibland vill man åstadkomma en statushöjande effekt. Man ersätter *städerska* med *lokalvårdare* osv. Det är ytterst viktigt att vara observant i sådana fall. Alltför lätt stannar det hela vid vad som har kallats verbala reformer. De är billiga för myndigheterna men i praktiken bara ett sken så länge inte de faktiska förhållandena i samhället ändras. Det är en öppen fråga om man kan förändra samhället genom att byta etiketter på yrkena.

Närmast Newspeak ligger vi nog i en serie ord och uttryck med tydligt politiskt brus. Ökänt är *friställa* som med sina associationer till *frihet* ger en cynisk version av en ofta hård verklighet. Liknande ting döljs gärna under abstrakta formuleringar som *arbetsmarknadspolitiska åtgärder* och *regionalpolitiska åtgärder*. Ett särskilt omnämnande förtjänar det poetiskt klingande *föryngringsyta* för det sakliga *kalhygge*.

Ser vi vidare ut över världen, finner vi som bekant miljöer där man bör tänka som överheten för att det skall gå en väl. Låt oss föreställa oss en motbild: gör hela världen till ett *Orators' Corner*. Låt vem som vill bildligt talat ställa sig på sin låda i sin stad och fritt plädera för sina synpunkter.

Det finns som bekant också miljöer, där historieskrivningen ändras i takt med att översåtarna växlar. Det förflutna hänger samman med *Oldspeak* som är *crimethink*, för att anknyta till Orwell igen. Kunskapen om vad som har varit – den som finns i våra hjärnor och i våra arkiv och bibliotek – har sådan kraft genom att den förser oss med jämförelsematerial och där-

igenom ger oss perspektiv på nuet och framtiden. Glöm aldrig att vi måste minnas!

Om vi ser till våra egna förhållanden igen, skall vi finna att det är mer att säga om överhetens språk. Något av det följer.

I riksdagen talas det mycket, omkring fyra miljoner ord under ett år. Det motsvarar ungefär femtio romaner. En intressant aspekt på det språket gäller det antal olika ord man använder. Det visar sig att det är betydligt lägre än i t.ex. tidningsspråk. Man får lika många ordformer – omkring 100 000 – om man undersöker *en* miljon ord tidningstext. Riksdagen speglar tydligen verkligheten på ett mindre varierat, mera abstrakt sätt. Har avståndet till verkligheten blivit *för* långt?

Brist på klarhet förknippas, delvis med rätta, med översåtarnas språk. Jag återkommer strax till detta. Låt mig belysa den oavsiktliga grumligheten med följande citat ur ett riksdagsstenogram: "När vi häromsistens motionerade, höll jag på att säga, när man häromsistens fick veta att vi för ett par tre månader sedan hade motionerat, det är den rätta beskrivningen av det hela, när man kom underfund med att vi närmare bestämt i mars månad hade motionerat och upptäckte – vi motionerade den 15 mars – det tycks man ha upptäckt den 29 mars, inte i utskottet för där behandlade man motionen strax efter den väckts men i kammaren och även på regeringshåll tycks man ha kommit underfund med det här den 29 i månaden därpå, alltså efter sex veckor, då talade man om det här dels som om det vore någonting annat än vad det är."

Fastän det pågår en hel del arbete på att göra myndigheternas språk mera tillgängligt, är det lätt att finna svårartade prov på den administrativa apparatens språk. Ibland är den verklighet man behandlar mycket komplicerad, och då kan det vara motigt att finna en rättfram språklig uttrycksform. Men när verkligheten är tämligen enkel, skulle det vara onödigt att krångla till och fördunkla saker och ting: "Vad som föreskrives om företrädesrätt till ny anställning i 7 – 11 §§ och bestämmelser som avses där äger motsvarande tillämpning vid läroanstalt som avses i 1 § när läroanstalten tillsätter sådan tjänst

som lärare med förordnande tills vidare från och med den 1 juli 1975 som ersätter annan tjänst som lärare med tidsbegränsat förordnande till och med den 30 juni 1975.”

Det gäller också ”att ta tag i nya utvecklingsbitar som kommer”, för att citera jargongen. Här är några exempel på formuleringar ur en fritidsutredning: ”Hög ålderstolerans har främst aktiviteter som inte kräver alltför stor fysisk ansträngning.” Denna upptäckt exemplifieras med trädgårdsarbete, ”där andelen utövare är större i de äldre åldersgrupperna än i de yngre”. Utredningen tycker sig förstå varför: ”Det sammanhänger sannolikt med att personer i de yngre åldersgrupperna mer sällan har tillgång till trädgård.” Man garderar sig som synes med ”sannolikt”. Utredningen går också in på bilens förhållande till fritiden och presenterar följande resultat: ”Ofta utövade aktiviteter, där bilen tycks ha liten eller ingen betydelse, är cykelutflykter.”

Värre är ändå den avsiktliga oklarheten. I synnerhet finner man exempel på sådan i politikernas språk. I stället för att svara på en fråga glider man in på något annat. Man ger helt enkelt inte besked.

– Det är ett tillmötesgående som ... som har missuppfattats men som ... som ...

– På vad sätt är det ett tillmötesgående?

– Jo, därför att det är ju ett mycket enklare sätt att beräkna det här. Alltså det är ju så här. Ja, det är ju mitt favoritämne som ... som herr B. nu lockat in på här.

Denna undanlidande funktion hos språket strider mot den grundläggande förutsättningen för mänsklig kommunikation, nämligen viljan att förstå varandra.

Den avsiktliga oklarheten innebär i själva verket att man använder språket som informationshinder. I vissa kretsar anses detta vara inte bara berättigat utan lovvärt. Partitaktiska mål blir överordnade en förnuftig diskussion.

Med den genomslagskraft som våra massmedier har är risken stor, att denna språkhantering blir mönsterbildande. Konsekvenserna kan bli ödesdigra. I en ökande dimbildning upp-



slukas inte bara förtroendet för makthavarna utan också förtroendet för medmänniskorna i allmänhet.

Till bilden hör också ett många gånger onödigt braverande med svåra ord, inte sällan som sköldar för lätta tankar. Här om året kunde man läsa om vad som kallades *självkänslorelevant behovstillfredsställelse* och som vid närmare betraktande visade sig vara *arbetsglädje*.

Mest är det kanske ekonomiska termer som svingas. Innebörden i ett par vanliga ord som *inflation* och *inflationstakt* är långt ifrån uppenbar. Det ger förstås också möjligheter till skilda beskrivningar av samma verklighet allt efter syftet.

Ibland framkastas det att datatekniken leder till ett torftigare språk och mera onyanserade resonemang. Jag tror att man måste hålla isär några olika saker för att få ordning på den diskussionen. Man får åtminstone skilja på de programspråk som används för att formulera instruktionerna till datamaskinen, sådana system som har utvecklats för kommunikation människa-maskin (interaktiva system) och register eller databaser med information om sakförhållanden eller personer.

Låt mig först säga att varje förändring i samhället avspeglas i språket. Det var så när hantverksskulturen kom in under medeltiden, när industrialiseringen skedde på 1800-talet, när filmen blev verklighet på 1900-talet osv. Vi kan lugnt förutse att språket tillförs nya ord och uttryck genom datatekniken, delvis på bekostnad av äldre. En del nyheter har vi ju redan fått.

Det är obevisat och inte alltför troligt, att användningen av själva programspråken skulle försämra det mänskliga språket. Varför skulle *gosub*, *for ... next* eller *rnd* ha ett större negativt inflytande än matematikens *lim* eller *log* eller kemins *pH* eller *koltetraklorid*?

Nej, datatekniken i sig är varken ond eller god. Det är dess användning som är avgörande. Det är förekomsten av dåliga interaktiva system och växande dataregister med fyrkantiga personuppgifter som har vållat berättigad oro.

Systemen för kommunikation mellan människa och maskin bör och kan utformas med människans krav som riktpunkt. Här

behövs ett starkt utvidgat och fördjupat samarbete, där data-lingvistikens erfarenhet tillvaratas.

Myndigheternas system för registrering av personuppgifter – i den mån sådana vill till – behöver heller inte vara så onyanserade som de nu ofta är. I varje acceptabelt system skall det bland annat finnas utrymme för kommentarer och tillägg i fritt format. Här kan nyansering, tvivel på uppgifters korrekthet osv. anges och därigenom bli beaktade.

Det finns alltså åtskilligt att iaktta när det gäller överhetens språk här i landet. Vi har inget Newspeak, men det är viktigt att motverka tendenser i den riktningen. Ni minns kanske från romanen att Newspeak inte var genomfört 1984 utan först beräknades ha slagit igenom omkring 2050.

Som väl är kan man inte hantera språkbrukarna hur som helst. Det finns en viss tröghet i språkbruket: ord lever kvar, fast översåtarna föreskriver andra ord. Man hör fortfarande ofta *sjuk-kassa* för det som officiellt är *försäkringskassa*. Man talar om *fosterhem* och *fosterbarn* när man behöver det. Man kallar en spade för *spade* och ett kalhygge för *kalhygge*. I sådana fall är det inte bara fråga om tröghet utan om genomsådande, och det är en stor tröst. Ett gott tecken är också att humorn frodas, som när 'gatsopare' föreslås heta *ekologisk operatör*. Bland det viktigaste som kan ske är ett aktivt arbete på att höja den språkliga medvetenheten.

Någonstans ovanför den offentliga slagväxlingen finns allas vårt gemensamma intresse. Kalla det nationen, landet, folk-gemenskapen eller något annat med samma innebörd. Och se det som ett led i en ännu större gemenskap – den globala. När debatten i alltför hög grad präglas av kortsynt partstänkande och språkinflation i stället för att vara ett offentligt samtal, är det detta vårt gemensamma intresse och därmed vars och ens situation som drabbas.

I sitt tal vid Nobelfesten kom 1983 års Nobelpristagare i litteratur, William Golding, in på situationen i världen. Det är dags att tänka om, menade han, rentav att ta ett steg tillbaka. Vad som krävs är i grund och botten generositet, djärv generositet: "Give, give, give!"

Kravet på generositet är stort och innefattar verkligen en rejäl portion djärvhet. Men jag undrar om inte den djärvheten finns innerst inne. Det är lätt att hålla sig för skratt när man försöker föreställa sig vad alternativet skulle innebära.

För Orwell var 1984 ett år någonstans långt borta i tiden. Men samtidigt stod det klart, att vi alltid kan ha ett 1984 bakom nästa krök. Därför krävs det en kritisk hållning och en ständig beredskap. Inte minst gäller det inom teknologin, den som kan vara en av våra bästa bundsförvanter. Det är värt att tänka över vilken dess roll skall vara i det stora sammanhanget.

Man kan säga att språket är människans karaktärsdrag. Det är en del av hennes personlighet. Att vaksamt, varsamt och verkligt arbeta med sitt språk blir att arbeta med sig själv. Och att göra det blir att arbeta för gemenskapen, för generositeten. Då skymtar ett mänskligt 2084.

## En forskningsstrategi för språkvetenskaplig databehandling. Perspektivskiss

Språkdata 1983.

Att det finns behov av en forskningsstrategi för en vetenskaplig institution behöver inte närmare motiveras. Syftet är att på grundval av en teoretisk grundsyn ange riktningar och utstaka mål liksom att skissera vägar att närma sig målen. Självfallet bör de övergripande idéerna ge rikligt utrymme för grundforskningsuppgifter.

I denna perspektivskiss behandlas inledningsvis under rubriken *Allmänt* några viktiga teoretiska och metodologiska grunddrag och en därur utvecklade forskningsstrategi. Därefter diskuteras i avsnittet *Speciellt* några huvudlinjer i den forskning som hittills har bedrivits vid institutionen och implikationerna därav för framtiden. I det avslutande avsnittet *Aktuellt* anges var det sett från det anlagda perspektivets synpunkt framstår som särskilt angeläget att satsa nu.

### Allmänt

Den teoretiska grundsyn som den här kortfattat presenterade forskningsstrategin bygger på har några karakteristiska inslag. Till en början skall det göras klart att ämnesområdet är språkvetenskapligt: det gäller ytterst att vinna ökad insikt i naturligt språk. Vidare framstår det som uppenbart att vi i datamaskinen har fått ett redskap som rätt utnyttjat i högsta grad kan främja språkvetenskapen. Detta kan ske bland annat på två grunder: genom de absoluta kraven på specificering och explicitet i fråga om enheter, regler och procedurer och genom den effektiva hanteringen av stora mängder språkliga data på alla nivåer. Symbiosen mellan språkvetenskap och informationsbehandling markeras i benämningen språkvetenskaplig databehandling och dess synonym datalingvistik.

Grundläggande är också idén om forskningens meanderslinga. Den innebär ett successivt framskridande inom ett givet projekt genom upprättande av en teoretiskt baserad modell, dennas applicering på ett autentiskt språkmaterial, revidering av modellen med hänsyn till de faktiskt konstaterade förhållandena, ny tillämpning på materialet osv. Det är alltså fråga om en växelverkan mellan teori och empiri. Också spiralen har använts som metafor för vad det gäller här.

Över huvud taget spelar studiet av språkmaterial, främst texter, skrivna och talade, en fundamental roll. I dessa kan man avläsa både de rekurrenta, ständigt återkommande fenomen som utgör stommen i alla texter och det kreativa utnyttjandet av språkssystemets dynamik. Jag har kallat dessa två aspekter texternas *ostinato* respektive *capriccio*. Textstudiet aktualiserar bland åtskilligt annat också frågorna om textrepresentation, lagringsform och tolkningsmodeller (där begreppet möjliga världar är av stort intresse). Det internationella forskningsläget i flera av dessa avseenden avspeglas i förhandlingsvolymen från det Nobelsymposium över temat *Text processing* som jag arrangerade i samarbete med Lars Gyllensten för en tid sedan. Vid sidan av textanalysen kan textsyntes ge viktiga inblickar i språket.

Ett betydelsefullt inslag i den teoretiska grundsynen är den centrala roll som tillmäts lexikonet i språkförmågan och språkbeskrivningen. Det gäller här den lexikaliska komponenten i vid mening: å ena sidan innefattas såväl ord som ordled och ordförbindelser, å andra sidan inbegrips utöver stavning, uttal och böjning också ordbildning, syntaktiska egenskaper, betydelseinnehåll och relationer till andra enheter i det lexikaliska systemet m.m. Rehabiliteringen av lexikonet innebär inte att grammatiken, semantiken, textlingvistik eller pragmatiken negligeras, snarare att de relativa vikterna förändras.

Till detta kommer intresset av att efterforska inte bara kvalitativa egenskaper utan också kvantitativa uppgifter som frekvens och spridning över texttyper hos de språkliga fenomenen. Särskilt skall här framhållas att kvantiteter kan studeras på alla

språkliga plan: om ett språkligt fenomen kan specificeras, låter det sig också bokföras på ett eller annat sätt.

Det skall vidare göras klart, att språket på grund av sin ytterst intrikata natur inte kan restlöst behandlas med helt automatiserade metoder. Samtidigt som utvecklingen av datalingvistiska algoritmer spelar en mycket betydelsefull roll, blir därför det interaktiva förfarandet – samverkan i realtid mellan forskaren och datamaskinen – av stor vikt.

När jag i olika sammanhang har presenterat en karakteristik av ämnesområdet, har kombinationen av automatisk och interaktiv databehandling varit en viktig aspekt. Som ram för ämnet har jag angett ett sammanfattande språkligt programsystem med inriktning på textanalys, textsyntes och dialog med en mänsklig partner. De huvudtyper av resurser som jag översiktligt har betecknat som nödvändiga för ett sådant system är språklig kapacitet (som främst avser lexikalisk, grammatisk, semantisk och textlingvistisk potential inklusive kvantitetsuppgifter eller om man så vill probabiliteter), kognitiv kapacitet (som bland annat gäller strukturering av inkommande information och slutsatsdragning) och encyklopedisk kapacitet (som hänför sig till världen och däri timade händelser). Naturligtvis finns det klara beröringspunkter mellan dessa ting. De kan sägas svara mot människans lingvistiska (och därmed förknippade stilistiska), kognitiva och encyklopediska resurser.

Så länge den första av dessa komponenter är så otillräckligt känd som fallet är, finns det anledning för datalingvistiken att lägga tyngdpunkten där med nödigt iakttagande av överlappningarna i förhållande till de båda andra. Vissa av frågorna, kanske främst de kognitiva, har man kommit att intressera sig för på området artificiell intelligens, vars syfte är att skapa program som utför olika uppgifter, t.ex. schackspelning, på ett till synes intelligent sätt. Det kan tilläggas att den principskiss jag har tecknat i sig också innefattar simulering, i tillämpliga delar, av mänsklig språkfärdighet (med de starka restriktioner som därvidlag måste göras). Av det skälet har jag ibland kallat modellen Datorn som språkbrukare. Ytterst är målet att fånga språket i kategorier som är psykologiskt relevanta. Det finns

särskild anledning att hålla detta i minnet på ett område som vårt.

Till de allmänna förutsättningarna hör givetvis att en adekvat teknisk utrustning står till förfogande. En central resurs utgör den akademiska datacentralen, och den bör bland annat utnyttjas för tunga körningar av generell art. För mera speciella uppgifter behöver den kompletteras med institutionell utrustning. Det gäller här främst att tillgodose särskilda krav på in- och utmatning i olika medier med stor teckenrepertoar och att tillhandahålla instrumentering för intensiv interaktiv bearbetning i kombination med tillgång till stort direktaccessminne. På indatasidan kan särskilt nämnas det kraftiga behovet av optisk klartextläsning och taligenkänning. I det senare fallet finns ännu inget tillräckligt avancerat system. På utdatasidan kan i synnerhet behovet av full typografisk arsenal och olika alfabet samt talsyntes nämnas. Vidare måste utrustningen ge tillgång till programspråk och andra programvaror, inklusive operativsystem, som är lämpade för forskningsuppgifterna.

Ett drag som har präglat forskningen vid institutionen hittills och som det finns all anledning att bibehålla är att forskningsprojektens vetenskapliga resultat i tillämpliga delar dokumenteras på ett sådant sätt att de kan få praktisk användning. Ett typexempel är det första stora projektet, Datamaskinell undersökning av tidningsprosa, som innefattade publiceringen av Nusvensk frekvensordbok baserad på tidningstext (NFO) 1–4 (1970–80), Tiotusen i topp (1972) och homograflexikonet Olika lika ord (1978). Andra arbeten av liknande typ publicerade av deltagare i projekten är *The morphology of Present-Day Swedish* (projektet Algoritmisk textanalys), *Svensk baklängesordbok* (projektet Riksspråkligt datalexikon) och *Förnamnsboken* (projektet Svenska namn). Skrifterna publiceras i allmänhet i serien *Data linguistica*, som utges från institutionen och som också är öppen för bidrag från andra håll.

Resultaten av projektverksamheten finns naturligtvis också i maskinläsbar form. Det kan gälla textuella och lexikaliska material i olika grader av vetenskaplig bearbetning. Sådana är ofta intressanta för andra forskare och i många fall för verksamheter

utanför forskningen: hos myndigheter och organisationer, i näringslivet – främst inom den grafiska industrin och text- och ordbehandlingen – på handikappområdet osv. Detta ställer krav på ett serviceorgan. Vid institutionen finns också ett sådant varom mera längre fram.

Till den allmänna bilden av forskningsstrategin hör utbildningen. Av naturliga skäl krävs såväl grundutbildning som forskarutbildning. När det gäller forskarutbildning på ett så övergripande fält som det institutionen företräder är det rimligt att utarbetade avhandlingar i vissa fall framläggs i angränsande ämnen (t.ex. allmän språkvetenskap, nordiska språk, svenska eller annat särspråk). Vidare kan i enstaka fall avhandlingar som inte har någon naturlig plats i det akademiska ämnes-systemet men genom sin inriktning tydligt överlappar med vårt ämne läggas fram här. Gränsdragningar får inte hindra forskningsdynamiken.

Sammanfattningsvis kan med en viss förenkling sägas, att institutionens forskning hittills har präglats av datalingsvistisk textanalys med vid lexikalisk inriktning. Det finns all anledning att fortsättningsvis vidga och fördjupa både det lexikaliska och det textuella engagemanget.

Detta låter sig på ett fruktbart sätt göra inom ramen för en filosofi som knyter sig till begreppet *språkbank*, taget i icke-trivial mening. Bakgrunden är följande. Ett klart mål för verksamheten på ämnesområdet enligt den tidigare givna karakteristiken är att med hjälp av de metoder som håller på att utvecklas komma åt innehållet i text. Olika förfaranden för att verifiera ett sådant företags resultat har på olika håll använts eller angetts. De är exempelvis att programvägen skriva en sammanfattning, göra en parafra, göra en översättning eller utföra den handling som texten inbegriper. Ett oss närliggande sätt är att skapa ett system som gör det möjligt att hänföra varje element respektive kategori i texten till dess relevanta plats(er) i en sofistikerad språkbank. Denna uppläggning rullar i själva verket upp hela fältet, eftersom den kräver insatser på alla lingvistiska nivåer: lexikon, syntax, semantik etc. Fenomen av vitt skilda slag kommer således i fråga. I ytterlighetsfallet



gäller det också sådana som ytligt sett inte är manifesterade i själva texten. Ett exempel är de modala och temporala kategorier som kan hänföras till en sekvens som "(var lugn) han gör det" i betydelsen 'han kommer med säkerhet att göra det'. Som jag har påpekat i annat sammanhang kan man också notera, att var och en av oss är ett slags levande språkbank.

## Speciellt

Det tidigare nämnda frekvensordboksprojektet (NFO) spände över flera av de områden som angetts. Bestämningen av formellt grundade enheter på ordnivå krävde bland annat en viss morfologisk och syntaktisk analys, bestämningen av ordförbindelserna krävde en viss syntaktisk och semantisk analys och bestämningen av ordleden krävde morfotaktisk och ganska ingående semantisk analys. Detta arbete har givit en mycket värdefull erfarenhetsbakgrund och genererat en rad nya forskningsuppgifter samtidigt som de praktiska tillämpningarna har varit många. Det underliggande textmaterialet – en korpus av en miljon löpande ord ur morgontidningar – har blivit genomlyst på ett sätt som kanske inget annat. Likväl genomfördes aldrig en av de uppgifter som planerades på ett tidigt stadium: en grammatisk strukturanalys med samma kvalitativa/kvantitativa förtecken. Den rymdes helt enkelt inte inom det femtonåriga projektarbetet. Det finns anledning att återkomma till den.

Den lexikaliska huvudströmmen har förts vidare framför allt i projektet Lexikalisk databas / Svensk ordbok (LDB). Inom detta projekt ligger tyngdpunkten i lexikalisk semantik kombinerad med databasteknik. Avsikten är att beskriva ett stort antal lexikaliska enheter (minst 100 000) med avseende på stavning, uttal, böjning, ordbildningsegenskaper, betydelse, bruklighet, i förekommande fall fackområde, syntaktiska egenskaper och idiomatisk fraseologi, varvid betydelsedefinitionen särskilt fokuseras. Avsikten är vidare att lagra beskrivningarna i ett nätverk, som medger sökning efter kategorier av olika typ,

innehållsliga såväl som formella. Dessutom förutses den viktiga möjligheten att från detta lexikaliska håll främja den algoritmiska textanalysen.

Hittills har arbetet främst inriktats på att sammanställa materialet och utföra de lexikaliska beskrivningarna. Den utan jämförelse mest arbetskrävande insatsen har gällt den semantiska analysen och definitionsutformningen. Siktet är nu inställt på att generera det första praktiska resultatet av projektet, Svensk ordbok. I sin egenskap av definitionsordbok blir denna av en typ som hittills har saknats för modern svenska men som anknyter till internationell lexikografi, bland annat på de engelska, franska och tyska språkområdena. Det är en rimlig tanke att svenska språkbrukares behov av en dokumentering av det levande språkets viktigaste lexem inte är mindre än på dessa andra håll.

Under det aktuella skedet har det lexikaliska stoffet lagrats i ett system av datafiler, relaterat till det fasta inkodningsformat som förutsatts från början. När sättbandet till ordboken har genererats, öppnar sig på allvar möjligheten att genomföra den ursprungliga intentionen att etablera den lexikaliska databasen som ett länkat nätverk, sökbart på ett mångsidigt tesauristiskt sätt. Det fasta inkodningsformatet och definitionernas strikta utformning är härvidlag stora tillgångar. Av de från början skissade lingvistiska uppgifterna återstår framför allt följande två, som var för sig är högtintressanta. Den ena gäller analysen av själva definitionerna, ytterst i termer av djupkasus (roller, aktanter). Den andra gäller upprättandet av de semantiska relationerna mellan de olika enheterna i termer av semantisk närhet, över- eller underordning inom lexikaliska subsystem m.m. Närmare specifikationer krävs självfallet här som på andra punkter. Också detta skede i projektet kan avkasta praktiska resultat, t.ex. en synonymordbok eller en begreppsordbok av ny typ. Inte minst den sistnämnda vore ett välbehövligt tillskott till den svenska ordboksuppsättningen.

Ett annat lexikaliskt projekt med stor potential har just startat: O·S·A, dvs. Om svar anhålles. Dess syfte i den första fasen är att läsa in Svenska Akademiens stora historiska ordbok

(SAOB) med hjälp av den nyligen installerade optiska textläsaren och lagra innehållet som en databas. Tack vare ordboksredaktionens genom åren ganska enhetliga principer skall också denna databas kunna ges en välstrukturerad form, som medger sökning efter exempelvis morfologiska, semantiska, ämnesmässiga, kronologiska och etymologiska kategorier. Som projekttiteln anger skall man kunna få svar på frågor som den tryckta ordboksversionen inte medger eller endast medger efter insatser av orimligt stora mängder manår.

Projektet är tillförsäkrat kapacitet att göra själva inläsningen. Vad som ytterligare krävs är en lingvistisk insats för att bland annat utarbeta principerna för de kontrollprogram som skall köras på resultatet av inläsningen, utforma principerna för struktureringen av materialet och utveckla modeller för bearbetningar av det. Därtill kommer uppgiften att utforma själva programmen. Efterhand bör databasen göras allmänt tillgänglig. Vissa bearbetningar kan också komma i fråga för tryckning, t.ex. en kronologiskt ordnad ordsamling.

I förlängningen ligger två lockande uppgifter. Den ena innebär inläsning av äldre texter i syfte att komplettera ord- och beläggs materialet i det (visserligen rika) urval som SAOB utgör. Den andra tar sikte på det faktum att bakre tidsgränsen för SAOB, som är 1521, visserligen inte är omotiverad men i ett större perspektiv ändå arbiträr – ordförrådet har ju utvecklats successivt utan någon särskild gräns det nämnda året. Det blir därför viktigt att utröna möjligheterna att optiskt läsa in också fornsvenskt lexikaliskt material (Söderwall, Schlyter) och textuellt material.

Vid institutionen finns 10 upplagan av Svenska Akademiens ordlista (SAOL) i maskinläsbar form. I samarbete med redaktören för den kommande 11 upplagan Sven Ekbo sker nu en uppdatering som förberedelse för tryckningen av den nya upplagan. Härigenom blir också denna tillgänglig i maskinläsbar form och kan utnyttjas för sökningar och undersökningar. Av uppenbart intresse är bland annat en studie av de förändringar som skett mellan de båda upplagorna.

Svenskans lexikaliska relation till främmande språk är av stort intresse. Det arbete som har utförts inom projektet Lexin (Lexikon för invandrare) har lett till sammanställning av ett begränsat men centralt lexikaliskt stoff (omkring 15 000 enheter), som har översatts till ett par främmande språk. Här kan således flerspråkiga databaser etableras, och det blir möjligt att på ett smidigt sätt komplettera dem med efterhand aktualiserade ord och uttryck av mer eller mindre termliknande typ. Naturligtvis är vidare det stora materialet i Lexikalisk databas ett utmärkt underlag för arbete med tvåspråkiga ordböcker av större omfattning. Intressanta tillämpningar inom språköversättningen avtecknar sig också. Detta gäller både i den mån man inriktar sig på automatisk översättning inom vissa ämnesområden och i det mer aktuella fallet att man syftar till att skapa bättre förutsättningar för mänsklig översättning med datorstöd.

De nämnda lexikaliska arbetena NFO, LDB, O·S·A och SAOL kan ses som delmängder i en *Svensk lexikalisk databas* under utveckling. Denna, respektive dess delmängder, faller naturligt in i det strategiska grundmönster som tecknades genom språkbanksbegreppet.

Som jag nämnde tidigare förutsätter begreppet också aktiviteter på andra lingvistiska plan. Främst vill jag i detta sammanhang peka på det grammatiska, enkannerligen det syntaktiska. Syntaktisk analys är bland annat ett led i identifieringen av de lexikaliska enheterna. Självfallet har den emellertid också ett värde i sig genom blottläggande av de grammatiska strukturerna i texterna. Det var detta mål som föresvävade oss i NFO-projektet men som fick vika genom att vi satsade på en bred lexikalisk offensiv. Med tanke på hur väl undersökt textkorpusen är från lexikalisk synpunkt vore det mycket tillfredsställande att också få lägga den syntaktiska brickan på plats.

Förutsättningarna är gynnsamma för att genomföra detta syntaktiska projekt. Den detaljerade lexikaliska beskrivningen kan utnyttjas. Tidigare erfarenheter har vunnits i projektet Algoritmisk textanalys. I flera projekt har påläggning av lingvistiska uppgifter på språkliga enheter, s.k. taggning, utförts. Härtill kommer erfarenhetsutbyte med andra forskare här

hemma och utomlands. Dessutom har en förundersökning till projektet SynTag utförts med mycket gott resultat. Det blir viktigt att inom projektet kombinera interaktiv taggning med utveckling av automatiska delrutiner. Publicering av en volym med uppgifter om strukturfrekvenser, inklusive valensuppgifter för verben, bör övervägas.

Också inom de lexikaliska projekten LDB och O·S·A finns det behov av att utveckla syntaktiska analysrutiner, nämligen för den syntaktiska delen av analysen av betydelsedefinitionerna. Samverkan med SynTag-projektet ter sig naturlig.

Här finns det vidare anledning att nämna den analys av vissa semantisk-kommunikativa kategorier som har beskrivits i projektet Språksociala studier i en talspråklig databas. Materialet består av fria samtal mellan par av personer boende i Göteborg och har ett omfång av en halv miljon löpande ord. Det har inkodats och analyserats interaktivt. De kategorier som främst har uppmärksammats är modalitet och tid.

De strukturella och kvantitativa data som inhämtas i SynTag-projektet och de övriga nämnda projekten kan förväntas främja vidareutvecklingen av den syntaktiska textanalysen i betydande utsträckning och framstår därigenom som högst relevanta i grundstrategins perspektiv. Detta gäller emellertid också i ett annat avseende. I själva verket vill jag se dessa studier som ett första viktigt steg i riktning mot en *Svensk grammatisk databas*.

En sådan databas bör innehålla strukturbeskrivningar av de grammatiska fenomenen, uppgifter om deras innebörd i den paradigmatiska och den syntagmatiska dimensionen och faktiska exempel. Utformningen bör vara terminologiskt enhetlig och formatmässigt sammanhållen. Liksom sin lexikaliska pendant bör databasen vara så välstrukturerad att den ger många ingångar till materialet. Den sammanställning av strukturfenomen och litteraturreferenser som görs inom Lundaprojektet En svensk referensgrammatik är en annan sorts källa som torde bli värdefull.

Tillsammans med den lexikaliska databasen skulle den grammatiska kunna ge utförliga och mångfacetterade besked

på frågor av skiftande slag. Denna sammanhållna resurs blir sålunda i själva verket en *Svensk språklig databas*. I denna, eller om man så vill dess integrerande delar, skulle man exempelvis kunna söka på ett kommunikativt intressant begrepp som *läge*. Ur den lexikaliska basen skulle man få definitionen av begreppet och dess relation till andra enheter som uppvisar semantisk affinitet, i de mest typiska fallen genom att i sina definitioner ha *läge* som genus proximum. Man skulle bland annat få fram *position*, *planera* och *på* med tillhörande uppgifter allt efter önskemål. Ur den grammatiska delen skulle man bland annat få besked om att *läge* kan uttryckas med vissa typer av prepositionsfraser, exemplifierat med fall av typen *Karlsson på taket*. En på detta sätt fyllig, välordnad och mångfaldigt tillgänglig Svensk språklig databas svarar väl mot den i forskningsstrategin förutsatta sofistikerade språkbanken. Den svarar otvivelaktigt också mot ett verkligt behov i det nutida informationssamhället med tanke på den centrala roll språket spelar i detta.

Institutionens tidigare nämnda serviceorgan har nationellt uppdrag att bygga upp ett textarkiv och en ordbank, att utföra grundläggande bearbetningar av texter och att tillhandahålla materialet. Textarkivet omfattar nu omkring 30 miljoner löpande ord och ordbanken omkring 500 000 olika enheter. Som nämnt förvaltar serviceorganet också textuella och lexikaliska material och bearbetningar från forskningsprojekten. Allt eftersom resultaten av arbetet inom ramen för det forskningsstrategiska språkbanksbegreppet utkristalliseras, omhändertas de således här. Serviceorganet blir på så sätt detta begrepps konkreta motsvarighet, vilket gör att vi nu gärna kallar det Språkbanken i stället för Logoteket, som det hittills har hetat.

## Aktuellt

Det fortsatta arbetet inom språkbanksstrategins ram bör orienteras så, att allt fler deluppgifter kan lösas, helt eller delvis, programvägen. Detta kräver vidareutveckling av teori och

metod, fortsatt uttestning av algoritmer och successiva undersökningar och bearbetningar av språkliga material.

De närmaste framstötarna bör göras inom projekten Lexikalisk databas, O-S-A och SynTag i enlighet med intentionerna i föregående avsnitt.

Det blir av stor betydelse att förstärka datautrustningen så att den svarar mot behoven med avseende på föreliggande och planerade forskningsuppgifter.

För institutionens serviceorgan bör gälla att arbetet förs vidare med såväl tal som skrift i åtanke. Som en särskild uppgift framstår omhändertagandet av Strindbergsmaterialet (text och bearbetningar) i samband med utgivningen av Samlade verk. En strävan bör vara att vidga den utåtriktade verksamheten. I flera fall kunde exempelvis databaser tillhandahållas som tjäns-ter i ett system som Teledata.

Såväl för rekryteringen till ämnet som för samhällets behov i övrigt är det slutligen viktigt att undervisningsutbudet får en betydligt ökad omfattning. För doktorsavhandlingar ger forskningsstrategin rika möjligheter.

\* \* \*

Synpunkterna i denna perspektivskiss har vuxit fram efterhand och glimtvis publicerats eller föredragits på olika håll och i olika sammanhang. Jag tackar mina vänner på institutionen för fruktbara diskussioner och idogt samarbete. Min seminariserie höstterminen 1982 ägnades åt det tema som har behandlats här. Naturligtvis har jag under årens lopp också haft många värdefulla samtal om hithörande ting med kolleger på olika fält inom och utom landet, vilket jag tacksamt noterar.

## Andas

Struktur och variation. Festskrift till Bengt Loman. 1983.

Principiellt är det möjligt att uppfatta rubriken så, att ämnet för dessa rader är ett slags sammansatta substantiv med *and-as* som typexempel eller en grupp genitivformer (*samma andas barn*) eller en sorts släktnamn (*som Mutas*) eller ytterligare något språkligt fenomen. Jag är emellertid övertygad om att de allra flesta förstår mig rätt – det gäller det fysiologiska verbet. Som rubriksättare kan man förlita sig på den psykologiska relevans som ords grundform har och den betydelsefulla roll som kvantitetsförhållandena spelar. Att det aktuella verbet visar deponentiell form i modern svenska är knappast någon störande faktor.

Om man tar en överblick över situationen sådan den framträder i SAOB:s och Söderwalls språkprov, får man en rätt intressant bild. Grundbetydelsen är belagd sedan fornsvenskan: *gether anda ey gönom näsburona som annor dywr, vthan heller gönom öronen dragha andhan; the thura ey andas* (Söderwall med supplement). Vi kan ange den som 'respirera', vilket innebär att luft dras in i lungorna och en något annan gasblandning släpps ut. Hit kan också föras den utvidgade användningen i fall som *fiskar andas med gälar* och *maskar andas genom huden*.

Det finns enstaka tidiga exempel på bildlig användning med nära anknytning till grundbetydelsen: *Gud andhas och bläs på vårt trångå hierta* (1582). Den stora vidgningen av verbets register börjar emellertid på 1700-talet med tidens ökade behov av uttrycksmöjligheter (i fortsättningen ger jag alla språkprov i modern stavning).

Inom grundbetydelsens ram kan man utnyttja verbets potential genom att låta sättet att andas markera sinnestillstånd: *andas lätt; andas tungt; andas fritt*.

Att låta andningen representera livsprocessen som helhet ligger i och för sig nära till hands: *hon andas blott för dig*. Med



tanke på det nytillkommande argumentet *för dig* är innebörden dock så avvikande att vi etablerar den med en egen parafra, 'existera'. De två huvudmomenten i andningen kan naturligtvis fokuseras var för sig. Inandningen ger vissa möjligheter. Vi får konstruktioner med objekt som *den luft jag andas; andas kanelångor*. I den bildliga förlängningen ligger *den oskuld man här andas; du andas liv och vederkvickelse i naturen*. Den övergripande parafrazen blir den medicinska 'inhalera'.

I synnerhet är fokuseringen på utandningen semantiskt fruktbar. I grunden är det fråga om en direkt knytning till det fysiologiska förloppet i fall som *andas på fönsterrutan; andas vitlök*. Denna fokusering kan betecknas med den likaså medicinska parafrazen 'exspirera'.

I bildligt utförande får vi exempel som avlägsnar sig ganska långt från fysiologin: *dottern andas kärlek; lybeckarna andas mord*. Det är här snarast fråga om en utstrålning. Ännu tydligare blir det när subjektet släpper personkravet: *kristendomen andas glädje; uppsatsen andas misstro*. Denna argumentförändring ger oss anledning att ansätta en särskild kategori. Vi väljer parafrazen 'emittera'.

En särskild innebörd får ordet när det markerar ett språkligt meddelande: *inte andas ett ord därom*. Vår beteckning blir 'kommunicera'.

Läget när vi går in i nusvenskan kan alltså kort sammanfattas så här.

<i>in och ut</i>	
respirera:	<i>andas</i>
existera:	<i>andas för ngn</i>
<i>in</i>	
inhalera:	<i>andas kanelångor</i>
<i>ut</i>	
exspirera:	<i>andas vitlök</i>
emittera:	<i>andas misstro</i>
kommunicera:	<i>andas om ngt</i>

För att skaffa sig en inblick i de nutida förhållandena kan man gå till tre stora textmaterial som finns i konkordansform (på Språkdata i Göteborg). De representerar tidningsspråk, riksdagsspråk och romanspråk under andra hälften av 1970-talet. Sammanlagt omfattar de drygt tio miljoner ord (10,8), fördelade på följande sätt: tidningar 1,2 miljoner, riksdagsprotokoll 4,0 miljoner, romaner 5,6 miljoner.

I den följande uppställningen redovisas förekomsten i materialen av de betydelsekategorier vi har urskilt. Av de båda förkortningarna står *Abs* för observerat antal och *Rel* för relativt antal (per en miljon löpord).

	Press 76			Riksdags-			Romaner 77 protokoll 78/79			Totalt		
	Abs	%	Rel	Abs	%	Rel	Abs	%	Rel	Abs	%	Rel
respirera	11	73	9	5	12	1	469	87	84	485	81	45
inhalera	2	13	2	1	2	(1)	22	4	4	25	4	2
exspirera	0			0			18	3	3	18	3	2
emittera	2	13	2	34	83	9	7	1	1	43	7	4
kommunicera	0			1	2	(1)	25	5	4	26	4	2
Totalt	15	99	13	41	99	10	541	100	97	597	99	55

Antalet belägg på olika former av *andas* är 15, 41 respektive 541. Uttryckt i relativa tal är värdena 13, 10 respektive 97. Redan nu kan vi alltså konstatera, att de som får våra texter att *andas* först och främst är de skönlitterära författarna.

Bland de sammanlagt nästan 600 beläggen på verbet (597 för att vara exakt) dominerar grundbetydelsen 'respirera' i egentlig och mer eller mindre bildlig användning. Den svarar för över 80 % av fallen (81). Därefter kommer med procenttal omkring 5 vad jag har betecknat som 'emittera' (7), 'kommunicera' (4), 'inhalera' (4) och 'exspirera' (3), medan 'existera' inte är belagt alls.

Som så ofta lönar det sig emellertid att granska totalvärdenas inre. När det gäller 'respirera' framträder därvid en klar skillnad mellan texttyperna. Koncentrationen till denna kategori är störst i romanerna (87 %), men även tidningarna ligger högt (73 %). Däremot respireras det föga i riksdagsprotokollen

(12 %, vilket innebär 5 belägg). Det är i och för sig inte oväntat att det är författarna och journalisterna som arbetar mera livsnära, medan riksdagsledamöterna opererar på ett abstraktare plan. Men skillnadens storlek är tankeväckande.

Här följer några exempel: *han kunde inte andas och det svartnade för ögonen på honom; rätten att andas utan att äventyra sin hälsa; rummen levde, som om de andades; de ljög lika naturligt som de andades; hon andades i korta, upprörda stötar; hudens förmåga att andas; han hörde björnen andas; han sträcker på sig för att andas fritt; barnen är som persikor som andas; han måste få andas i jessu namn; jag andas, jag lever, jag rör mig; i Afrika andas jag lättare; allt som andas kunde – höll andan; valvet var en käft, som dagen lång andades in och ut folk, kärror, hästar, kor, getter, grisar.*

Fokuseringen på inandningen uttrycks ungefär dubbelt så ofta med *andas in* som med *andas*. Eftersom det här är partikeln som anger riktningen, noteras verbet i kategorin 'respirera'. Som exempel kan nämnas: *andas in eventuella giftiga ämnen; andas in kraftigt; hon andades in doften av hans dyra rakvatten; jag gick, red, drack vatten, andades in luft, sov i lugn och ro; när jag andas in andas du ut; ligga hela natten och andas in honom i långa tag.*

När verbet självt anger inandning, kategorin 'inhalera', har det oftast *luft* som objekt, men det finns också andra fall: *vad gör mig några snytingar, då jag andas rosens doft; luften han andades blev kyligare; de andades luft len som honung; hon andades mögel därinne; han andades vid komman i texten; andas frisk luft; dimman kändes luddig att andas; han gick ut för att andas livsluft eller plocka granris; de har aldrig andats fabriksdamm; få något syre att andas.* Fördelningen över texttyperna är inte särskilt anmärkningsvärd.

Framhävandet av utandningen uttrycks på motsvarande sätt ungefär dubbelt så ofta med *andas ut* som med *andas*. Också dessa fall har analogt förts till 'respirera'. Exempel: *han andades ut i en suck; han andades ut rök genom näsan; när han andades ut såg det ut som om han krympte; han andades ut av lättnad; livet tycktes andas ut ur hennes hud.*

Utan partikel är verbet i kategorin 'exspirera' bara belagt i romanerna. Bland beläggen är dessa: *han andades henne i nacken; ett doftande utrymme det andades halm; de andades på det igenfrusna glaset; en berusad man andades vitlök och vin över henne; andedräkten stod som fanor ut ur munnen varje gång de andades.*

När det gäller 'emittera' finner vi en klar dominans hos riksdagsledamöterna. Relationstalet är 9 gentemot 2 respektive 1. Vi ser också att denna kategori har ungefär samma överläge inom riksdagsmaterialet (83 %) som 'respirera' inom romanmaterialet (87 %).

Vidare är det mycket tydligt att användningen präglas av debattsituationen. Oftast står ett negativt uttryck som objekt: *inlägget andades negativism och misstänksamhet; svaret andas en alltför passiv hållning; reservationen andas misstro; uttalandet andas förakt; motionen andas missnöje; flygbladet andas invandrarfientlighet; betänkandet andas osäkerhet; svaret andas samma slags tongångar som 1972; utskottsbetänkandet andas resignation; kongressrapporten andas bekymmer osv.*

Uttryckssättet blir schablon. Från språkanvändningens synpunkt kan man notera att vi har fått ett polemiskt *andas*. I själva verket har utvecklingen gått ända därhän, att det uppträder en ny konstruktionstyp, nämligen *andas att*: *anförandet andades att han talade samtidigt för två saker.*

Beroende på debattläget kan objektet naturligtvis vinklas i positiv riktning utan att det polemiska syftet försvinner: *utskottet andas tillförsikt; svaret andas öppenhet; utskottsbetänkandet andas omsorg; svaret andas en positiv inriktning; svaret andas förtröstan osv.*

Hos journalisterna finns bara två fall av 'emittera': *ingenting i berättelsen andas lidelse; miljoner arbetsplatser andas 1800-tal.* Också tidningssituationen inbjuder ju ofta till polemik, men tydligen väljer man i det fallet oftast andra uttryck.

Beläggen hos författarna har en annan karaktär. De bidrar till att känneteckna miljö eller personer: *alltsammans andades en osedvanlig frid; interiörens gamla ordning, som andades*

*hennes föräldrar; hela våningen andades lugn och ordning; hon andades sorg; brevet andades stora passioner; hon hade andats kärlek genom varje por osv.*

Den pregnant betydelsen 'kommunicera' har inte oväntat en tydlig profil. I tidningsmaterialet saknas den helt, och i riksdagsspråket finns det ett enda belägg (*inte andas om att*).

Kategorin fyller uppenbarligen en funktion i romanspråket, där verbet bland annat förekommer som sägeverb: "*Vad menar du?*" *andades Eva*; "*En automobil!*" *andades hans mor vördnadsfullt*; "*Jill, älskling*" *andades hon*, "*vet du vad: somliga män är som sobelpälsar*"; *med läpparna nära hans andades hon ett aah som nästan gjorde det till ja osv.* I de flesta fallen är det fråga om översättningar från engelska.

Flera andra konstruktionstyper inom den kommunikativa ramen är belagda i romanerna. De är: (*inte*) *andas ett ord om ngt*; (*inte*) *andas om ngt*; *andas om att*; *andas ngt till ngn*; *andas till ngn om ngt*. I det anlagda perspektivet framstår denna kategori alltså snarast som ett episkt *andas*.

\* \* \*

De semantiska övervägandena i kombination med analysen av materialet har sammanfattningsvis lett till några iakttagelser rörande *andas* idag.

Innehållsmässigt har det visat sig att materialets belägg kan fångas in i fem urskilda kategorier: 'respirera; inhalera; exspirera; emittera; kommunicera'. En sjätte historiskt belagd kategori ('existera') har inte påträffats.

Formellt har vi konstaterat en ny konstruktionstyp: *andas att* + sats.

Vad användningen beträffar har det framgått att det framför allt är kategorin 'respirera' som utnyttjas och att det framför allt är romanförfattarna som utnyttjar den. Kategorin är också den vanligaste i tidningsmaterialet.

Till romanspråkets profil hör vidare ett ganska utbrett bruk av kategorin 'kommunicera': ett episkt *andas*.

Riksdagsspråket slutligen karakteriseras i hög grad av kategorin 'emittera' med subjekt som betecknar sak och objekt som betecknar sinnestillstånd: ett polemiskt *andas*.

## Text Processing as a Theme. The Capriccio and Ostinato of Texts. Opening Address at a Nobel Symposium 1980

Text Processing. 1982.

If, as the well-known description runs, a dialectologist is a scholar who was born 50 years too late, what then is a text-processing researcher? One who was born 50 years too early? I hope not – and indeed I think not! After all, we have come together for a Symposium on Text Processing.

In the directions of the Nobel Symposia, it is stated that the Symposia should be devoted to fields emerging as scientific break-through areas or fields that are of central importance to society. The Symposia should also have an interdisciplinary character. Our theme should, in addition, be of interest to the Swedish Academy and to the field of computational linguistics.

*Text processing* is of course as widely interpretable an expression as any. On the premises of a computer company it is normally restricted to the domain of office routines. Text processing is then regarded as an aid for those working with documents in such an environment. Evidently, this is too narrow an interpretation for our purposes.

A text, written or spoken, can be seen from a large number of viewpoints. Among other things, it can be looked upon as

- a message in a communicative situation,
- an act,
- a way of organizing some intended content,
- a model of a possible world,
- a means of suppressing information,
- an experimental field for grammatical and lexical analysis,
- an application of a generative model,
- a flow of linguistic data,

a reflection of an author's temperament,  
 a genre piece,  
 a prognosis of certain reactions in the receiver(s).

The theme of Text Processing thus brings up a considerable number of the questions focused on in present-day linguistics and neighbouring disciplines. They include

grammatical problems brought to light by the widening of the range of vision to sequences of sentences;  
 lexical and semantic problems involved in a deeper understanding of a text;  
 questions of communication and speech acts, including their psycholinguistic aspects;  
 in a wider cultural context, various ethno-linguistic and socio-linguistic questions;  
 in a literary perspective, questions of style and text attribution;  
 and, last but not least, the demands for a higher degree of stringency in the direction of formalization, paving the way for increased possibilities of using the computer as an instrument in humanist studies.

The three topics, Text Representation, Text Analysis and Generation, and Text Typology and Attribution, are meant to identify some areas of fundamental importance. They are not, of course, the only prominent areas. We have, for instance, deliberately excluded the lexical field, although it seems to be basic in linguistic communicative research.

Looking into texts, I have been mainly concerned with some qualitative and quantitative linguistic phenomena. With reference to this work, which was started, on a computational basis, in the early 60's, I should like to make just two points with regard to texts in this connection. Using metaphors, like some of you, let me call them *the capriccio trait* and *the ostinato trait*.



Dealing with authentic texts, at all levels, we must learn to *expect the unexpected*. This is what I mean by the capriccio trait. The two following strings are words observed in Swedish newspaper text:

statligt-kooperativt-fackförbunds-bostadsproduktionsföretag  
'government-co-operative-trade union-housing production company'

ligga-på-sängen-och-plugga-och-höra-hur-det-regnar-in-dag  
'lying-on-one's-bed-studying-and-listening-to-the-rain-leaking-in-day'

The morphosemantic concordance of Volume 4 of the *Frequenay Dictionary of Present-Day Swedish* (1980) contains many examples in the domain of lexical morphology. The concordance provides a complete account of the morphosemantic elements in a corpus of one million words. The following fragment presents the element *fog* [fu:g] (which is not related to English *fog*).

Enhet och kontext Unit and context	Ftext	Flex
<b>fog</b>	<b>170</b>	<b>20</b>
fog 1 'skäl' ('reason')	27	3
<i>fog nn</i>	12	
<i>be fog ad av -t</i>	10	
<i>o be fog a d av -t</i>	5	
fog 2 'skarv' ('joint')	44	8
<i>fog a vb -ad</i>	12	
<i>bi fog a vb -ad</i>	5	
<i>hop fog a vb -ad</i>	1	
<i>in fog a vb -ad</i>	7	
<i>mur fog nn -en</i>	1	
<i>sats fog ning nn -en</i>	1	
<i>till fog a vb -ad</i>	15	
<i>samm an fog a vb -ad</i>	2	
fog 3 'anpassa' ('to adapt')	7	2
<i>fog a vb -ad</i>	6	
<i>fog lig t ab</i>	1	
befogenhät 'auktorisering' ('authority')	21	3
<i>be fog en het nn -en</i>	18	
<i>be fog en het s lös av -t</i>	1	
<i>makt be fog en het nn -en</i>	2	

förfog 'disponera' ('to have at one's disposal')	70	3
<i>för fog a vb -ad</i>	15	
<i>för fog a nde nn -t</i>	54	
<i>för fog a nde rätt nn -en</i>	1	
ofog 'skälmstycke' ('mischief')	1	1
<i>o fog nn -et</i>	1	

The formal morpheme *fog*, established on the basis of commutation and classification, has three semantic (polysemy) components, characterized as distinct, paraphraseable kernel senses: 'skäl' ('reason'), 'skarv' ('joint'), and 'anpassa' ('to adapt'). There are also three concatenated cases. Notice the last case: when *o-*, the prefix of negation, is affixed to *fog* we get the unpredictable sense of 'skälmstycke' ('mischief').

Now look at the short passage below, taken from a newspaper article. There is a concomitant Swenglish translation.

Vid Löderups strandbad  
 'At Loederup seaside resort  
 börjar ett planterat bälte igen:  
 (there) begins a planted belt again:  
 gran, tall, vide och ljung  
 spruce, pine, willow, and heather  
 som skyler de gamla flygsandsmarkerna.  
 which hide the old drift-sand grounds.  
 Där badgäster brukar vimla  
 Where visitors usually swarm  
 och bilar brukar vara parkerade i varje glänta,  
 and cars are usually parked in every glade,  
 med familjer avsuttna bredvid,  
 with families unseated close by,  
 på fällstolar och fällbord,  
 in folding chairs and folding tables  
 vid TV-kanne- och plastburkfrukost i det gröna;  
 at TV-thermos and plastic-pot breakfast on the grass;  
 där syntes inte en själ nu,  
 there not a soul was seen now,  
 inte ett enda av själens fyrhjulade skal ens,  
 not even one of the soul's four-wheeled shells,  
 inte ens vid Mälarhusen,  
 not even at Maelarhusen,  
 inte ens vid Abbekås.  
 not even at Abbekaas.'

The passage is about a trip along the coast. It consists of two sentences. You will have noticed the complexity of the second sentence, made up of various clauses and a number of more or less loosely connected adverbial expressions.

Among other things, the families are said to be unseated but nevertheless they are understood to be sitting – in chairs and tables, as the phrase is. The sequence "in folding chairs and folding tables" seems, in fact, to mean 'sitting in folding chairs and having their things on folding tables'. How accessible is such an interpretation to a text-processing system? And how accessible is the capricious metaphor "the soul's fourwheeled shells", suggesting cars?

Drawing the moral, the capriccio trait of texts makes it extremely difficult to apply models of various kinds to real text. That's why I am in favour of a reciprocal research process, making theory influence empirical work, this work, in turn, influencing theory, in a constantly meandering pattern.

The ostinato trait of language manifestations is very striking. What it amounts to is this: you can safely *expect the expected*.

The collocational investigation of the millionword corpus referred to revealed, among other things, the following. There were about 600,000 cases of recurrent word combinations, i.e., sequences having two or more identical instances in the material. About 200,000 of the combinations were phraseologically relevant strings, called constructions, representing some 50,000 different collocations. The constructions ranged from sentences and clauses to, e.g., nounphrases and adverbial phrases. Idioms, in the restrictive sense of the term, constituted a remarkably small proportion. The following list shows a few examples of recurrent constructions.

något annat var väl heller inte att vänta  
'nor was anything else to be expected, you know'  
det är svårt att veta hur  
'it is difficult to know how'  
det är så sant som det är sagt  
'it is quite true'

det finns tecken som tyder på att X  
 'there are signs indicating that X'  
 X är av avgörande betydelse  
 'X is of crucial importance'  
 X har en stark känsla av att Y  
 'X has a strong feeling that Y'  
 att bryta de diplomatiska förbindelserna med Storbritannien  
 'to break off the diplomatic relations with Great Britain'  
 på grund av brist på X  
 'for lack of X'  
 den psykologiska effekten av X  
 'the psychological effect of X'  
 analys av den moderna människans situation  
 'analysis of the situation of the modern human being'

The passage below is a typical example of the occurrences of constructional collocations in text. The strings in question are underlined. Overlap between recurrent phrases is quite common.

Detta sätt att lösa olika frågor ger en bild av de vägar

'This way of solving various questions gives a picture of the ways  
 man gått för att försöka skapa förutsättningar för  
 they have gone in order to try to create opportunities for  
 ett definitivt genombrott för trafikflyget i Dalarna.  
 a definite break-through for the air services in (the province of)  
 Dalarna.'

In my view, these types of collocation have not been properly attended to in linguistic presentations. Nevertheless, they play an important part in the making up of genuine fluency in a language.

If we consider recurrent word combinations irrespective of their phraseological relevance, we get an even stronger impression of the ostinato trait. The picture emerging is what I call the recurrence wave. Again, there is a Swenglish translation (next page).

The marks above the words indicate recurrent combinations. Notice again how they overlap. In two places there are three recurrent word combinations overlapping. A detail illustrating the width of the recurrence phenomenon is the fact that 1500 out of the 50,000 sentences making up the corpus studied begin with the same sequence, namely *det är* 'it is' / 'there is'.

När jag nu kommer till en främmande stad,  
 .....  
 When I now arrive at an unfamiliar town,

och det behöver inte alls vara någon utländsk,  
 .....  
 and it need not at-all be some foreign,

vill jag ogärna gå hem till folk.  
 want I reluctantly to-go home to people.

Det beror inte på att jag inte vill träffa folk.  
 .....  
 It is-due not to that I not want to-meet people.

Men om jag kommer hem till människor,  
 .....  
 But if I come home to people,

tycker jag det känns ännu mera ledsamt att vara hemifrån.  
 .....  
 think I it feels still more depressing to be from-home.

The overwhelming recurrence of these collocations (or, from another point of view, lexical blocks) reflects, I suspect, the linguistic behaviour of language users in general. Thus, the

ostinato trait should be taken into account in the design of a plausible model of linguistic activity.

One of the problems we are faced with is how to *represent* this varied picture of collocations, forming a scale from idioms to more loosely connected but nevertheless stylistically relevant sequences. Furthermore, they are certainly to be considered in *text generation* of good quality. They offer an interesting criterion in specific *text typology* and, consequently, a sensitive instrument in *text attribution*.

A fascinating aspect of texts is the very interplay between the steady stream of ready-made blocks and the creative use of the dynamics of the language system. This is not an entirely formal matter but is indeed semantically relevant, too. We are even approaching the stronghold of artistic writing.

## Om förändringar i svenskan sedan 1950

Svenskans beskrivning 13. 1982.

Utlandssvenskar som kommer hem känner inte riktigt igen sig i språket, säger de. Det konstaterade till exempel Assar Janzén, som ju var professor i USA i många år och kom hem på sabbatsbesök då och då. Detsamma har jag hört från Sven Åhman, en journalist som länge var utomlands. Och jag måste säga att jag tycker man kan förstå dem. Nuspråket skapar sin historia, det också.

Om förändringar i språket sedan 1950 borde jag egentligen kunna ge besked, ty jag var med. Men två ögon och två öron är inte mycket i det här sammanhanget. Många miljarder ord sägs och skrivs varje år på svenska, och därför blir studiet av språkmaterial och tillgång till andras iakttagelser mycket viktiga faktorer.

Det huvudperspektiv som jag skall försöka anlägga har kommit fram ganska fint i en liten dikt av Sandro Key-Åberg. Den står som inledningsdikt till Svenska Akademiens debattbok *Vad händer med svenska språket?* (1976), som ju i och för sig är en relevant bok i sammanhanget. Så här står det i dikten:

Språk är mylla.  
Det är mjukt och levande som ditt eget liv,  
bär i sig avtrycken av det levande,  
av dess rörelser och strävanden.

Vad det kan bli fråga om här är att antyda några drag i dessa avtryck av liv.

Språket förändras kanske snabbare nu än förr. Utvecklingen hänger samman med samhällets förändring på områden som utbildning och teknik, invånarnas flyttning mellan olika områden och mycket annat men främst, skulle jag vilja säga, med informationsförhållandena. Vårt mångfacetterade samhälle är i grunden ett informationssamhälle.

Jag skall börja med att citera texten i några annonser i Dagens Nyheter den 11 oktober 1950, idag för 31 år sedan. Den första lyder: "Bli Du med Litteraturen genom All världens Berättare." Här skrivs *Du* med stor bokstav och *Litteraturen* likaså. Den andra visar en liten flicka med rosett i håret och ställer frågan: "Vems lakan är vitare – tant Elsas eller tant Karins?" Man ser också dessa båda damer och deras lakan, och man får veta att bådas är vitast, för båda använder Persil. I en tredje annons läser vi: "Florodol tvålar passa husmors kassa och familjens smak för kvalitet och vällukt." Jag behöver kanske inte peka på några detaljer för att våga påstå att de här annonserna visar lite av en annan värld.

Vi kan jämföra detta med en annons häromdagen i samma tidning: "Lektys värsta! 100 otroliga sidor ... Läckraste tjejnerna. Galnaste skämten. Bilderna vi aldrig vågat visa." Det man säljer på här är inte det *bästa* utan det *värsta*. Ordet *sex* nämns inte, men exploateringen är solklar och framhävs av en tillhörande bild. Vi noterar användningen av ordet *tjej* osv .

Annonserna ger många utgångspunkter för kommentarer till språkutvecklingen. Låt oss börja med en företeelse som berör få ord men som är mycket genomgripande i alla fall, både språkligt och socialt, nämligen tilltalskicket. Sedan några år har vi för övrigt en monografi över tilltalsordet *Ni* (Ahlgren, 1978). *Du* har ju alltid varit det naturliga tilltalsordet inom flera samhällsgrupper. Genom inverkan från idrottsrörelsen och andra folk-rörelser, värnpliktstjänstgöring, bilism och annat har *du* nu slagit igenom på bred front: på arbetsplatserna, i hemmen, i fritidsmiljön och även vid tilltal till okända. Vi är alltså mycket nära ett läge som motsvarar engelskans *you*, även om vi inte är där. Att lägga bort titlarna med litteraturen är poänglöst i dagens samhälle. Liksom i engelskan kan vi naturligtvis fortfarande markera olika grader av intimitet med andra medel.

Man kan också konstatera att *ni* inte är helt utslaget. Man möter det till exempel i en del skrivelser och enkäter. Jag blir själv tilltalad *ni* ibland, både i telefon och i levande livet, både i Sverige och i Finland. Till detta kan läggas att man är mera restriktiv med *du* i Finland än i Sverige. I anslutning till änd-



ringen i tilltalsskicket ändras också hälsningsorden. *Hej (då)* lever ett rikt liv. Det som tidigare kunde formuleras *Säg adjö till tant Elsa* blir nu snarare *Har du sagt hej till Elsa?*

Om man talar med andra om vad som är viktigt när det gäller förändringar i språket efter 1950, kommer man snart in på stilförskjutningarna. Här har verkligen stora förändringar skett. I fråga om skönlitteraturen som helhet tror jag man nu kan säga att intet är den främmande. Där finns knarkarjargongen, där finns sköra stämmningsstycken, där finns filosoferande utläggningar, där finns de tidigare tabuorden osv.

Ett sätt bland många att känna på stilläget då och nu är att gå in i en genre. Genom tillmötesgående från Arkivet för ljud och bild och Sveriges Riksradio har jag fått kopior på kassettband av två radioföredrag över samma ämne, nämligen Dödahavsruellarna. Det ena hölls 1949 och det andra 1979. Så här formar sig inledningarna:

Ibland kan viktiga upptäckter göras av en slump som när Alexander Fleming händelsevis råkade lägga märke till penicillinets bakteriedödande effekt. Och det är inte bara inom naturvetenskaperna som sådant sker. Ett av de märkligaste handskriftsfynd som någonsin gjorts var också slumpens verk. En arabisk herdepojke gick tidigt på våren 1947 och letade efter en försvunnen get och hittade istället en grotta med urgamla dokument, gömda i lerkrutor. Det hände i Juda öken, inte långt från Döda havets västra strand. Platsen hette Qumran.

Grottan, det hemlighetsfullt undanstuckna hudvecket vid moderjords barm, har i alla tider varit ett kärt tillhåll för människors barn som ville gömma undan någon hemlighet, och den har nästan aldrig svikit förtroendet. Den har hållit hårt på hemligheten. Om den någonsin släpper den ifrån sig, så har det ibland dröjt årtusenden. Det var en ren slump, en liten släppa i en terrasslutning, som någon gång under år 1947 lät en beduinsk herde kika in i en mörk grotta. Han gick och vallade får strax nordväst om Döda havet, söder om det gamla Jeriko. Han hade en smula tråkigt, och en smula pojkaktig äventyrslystnad hade han nog också i blodet, så han ålade sig in genom det lilla hålet, och då han väl var där nere i mörkret, tog han en sten och slängde den inåt grottan.

Utan att gå djupare in i någon stilistisk analys kan vi kanske enas om att anslagen är olika. Vilket av de två citaten är äldst? Jag har inte givit dem i kronologisk ordning. Bakom det längre,

mer elaborerade, lite sagobetonade står H.S. Nyberg. För den kortare, lite rakare presentationen svarar Bertil Albrektson. Jag tror inte att den faktiska tidsföljden överraskar er. Till en del rör det sig nog om en förskjutning av attityden i förhållande till lyssnarna.

Av mycket stor betydelse för stilsituationen är naturligtvis det språk som man möter i massmedierna. I stor utsträckning är det ett bra språk, menar jag, fast många kanske tycker att jag inte borde säga det. Men nog har det sina avigsidor, och som en orsak till det ser jag själva grundförutsättningen för massmedierna, nämligen den att de skall sälja nyheter. Numera säljer de dessutom underhållning. En bra, dvs. säljbar, nyhet skall helst vara sensationell, och det leder till språkliga tånjningar och tøjningar i olika riktningar och inte minst till ordinflation. Man behöver förstärkningar: *atombillig*, *drömmesa*, *idiotsäker*, *pangnyhet*, *stenkul*, *toppmodern* osv. Hit hör naturligtvis också superlativerna i Lektörs tappning, *värsta*, *läckraste*, *galnaste*, liksom användningen av ord som *otrolig*. Och till detta kommer det saftigare ordvalet. Man *avskedas* eller *avsätts* inte utan man *sparkas*. Man *utesluts* inte utan man *spolas*, och en verksamhet *läggs* inte *ner* utan den *skrotas*. Det är alltså ett hårdare språkklimat, och i det hårdare språkklimatet hör skjutjärnsjournalistiken hemma. Det finns ett belägg i textunderlaget till Nusvensk frekvensordbok (Press 65), där det står om en journalist att han odlade en intervjukonst som var ”distinkt och klar utan de enfaldiga självförhåvelserna i skjutjärnstekniken”.

Det förefaller mig uppenbart att denna teknik ibland driver vinklingen till det absurda, men den är också intressant på ett annat sätt. Den har effekten att ställa journalisten i centrum, journalisten som egentligen skulle vara förmedlaren. Den aspekten har fångats på ett träffande sätt av Lars Gyllensten i ett debattinlägg i tidskriften *Artes* (1981). Till artikeln har han fogat ett exempel, vad han kallar en illumination, som jag gärna unnar er.

Vi ber våra lyssnare om ursäkt! I den nyss lästa programredogörelsen över programmet *Mitt i Prick i Västerlandet*, med Nicke Pappskalle som producent, Miriam Snorum-Larsson och P.O. Vin-

terquist som reportrar, Bengt Allvrång och Sven Yxskäft som ljudtekniker, Sven-Berit Elgehorn som sminkös, Fiffra O'Jönsson som telekommunikant, Tutta Janthe som rumsauspicient, Per O. Jönsdotter som klippare, Råkan Platthammar och Sylvia Munsår som musikunderhållare, Ynka Rumsren, Fjante Fusk, Israel Gedigen och Ante Jonsäter som arkivforskare, Lennart Öjedrång och Hans-Isak Gahm som belysningstekniker, Karl Vändbar som scripta, Peter Bjatt som sanitetskontrollör, Göran Myrberg, Jan Palme, Lill-Zachris Gustafsson, Sven Lindflopp, Bintje Gärsgård, Birgitta Volm och Theodor Kalvdans som handräkningsamanuenser och med Wolfram Oljeblick som praktikant, glömde vi att nämna att de som intervjuades var presidenterna Ronald Reagan och Leonid Bresjnev och att intervjuerna behandlade de atombombfällningar som i torsdags av misstag gjordes över Berlin, London, New York, Moskva och Leningrad.

— — —  
Och själv heter jag Einar Bassalorum.

Frågan är i vilken utsträckning massmediernas folk är beredda att ta det språkliga ansvar som deras position egentligen fordrar. Jag har diskuterat detta med journalister i olika sammanhang, och det är ett problem som många är medvetna om. De kan emellertid inte alltid själva styra sin verksamhet.

Det offentliga språket har kommit i förgrunden under senare år. Det har faktiskt inträffat en del på det området, fastän inte alla kanske vill kännas vid det. Arbetet på att förenkla språkformen är väl det som särskilt skall nämnas. En viktig faktor var Statsrådsberedningens och Bertil Moldes promemoria Språket i lagar och andra författningar, som kom 1967. Sedan dess har man satt språkkonsulter i arbete i olika sammanhang: i Statsrådsberedningen och på åtskilliga andra håll bland myndigheterna.

Nu kan man till och med se avsnitt om själva språkformen i statliga betänkanden, en klar nyhet. Det är viktigt att slå fast en sak när det gäller strävandet mot förenkling. Man skall vara observant på att den verklighet som skall skildras ofta är utomordentligt komplicerad och att detta kan avspeglas i språket. Det är alltså inte någon enkel uppgift man står inför.

Jag skall ta ett par exempel på vad som har skett på det här området. Ture Johannisson (1979) har visat hur bruket av modala hjälpverb ändrats på ett överraskande sätt i lagspråket, närmare bestämt mitt inne i universitetsförordningen 1975. Det är dess-

utom inkonsekvent gjort. Man har ju tidigare använt *må* och *äger*, men nu går man över till *får* i stället. Och i denna universitetsförordning kan man å ena sidan läsa: "Ledamot [nämligen av konsistoriet som det då hette] må ej utan laga förfall utebliva från sammanträdet". Å andra sidan står det: "Ledamot i tjänsteförslagsnämnd får ej utan laga förfall utebliva från sammanträdet". Som Ture Johannisson påpekar har man använt universitetsförordningen som ett språkligt experimentalfält.

I högskoleförordningen från 1977 är *får* fullständigt genomfört. Carl-Eric Thors (1979) har påvisat en inte helt identisk men liknande utveckling i det finlandssvenska lagspråket.

Jag skall ge en glimt till från det offentliga språket. Det gäller några rader ur en rapport från Statskontoret. Där har man utvecklat ADB-system för löneuträkning (SLÖR), matrikelföring och personaladministrativ information (PI). Statens löne- och personalpensionsverk (SPV) förbereder införande av systemet osv. Här har vi alltså beteckningarna *ADB* (automatisk databehandling), *SLÖR* som är lönerutinerna, *PI* som är den personaladministrativa informationen och *SPV* som är ett verk, och hela saken utmynnar i det mycket tilltalande ordet *MYS-projektet!* Detta döljer en torr verklighet, lindrigt sagt, nämligen myndighetsrutiner för *SLÖR* och matrikelsystem.

Spelet med associationer är en sak. Det verkligt bekymmersamma med den ökade användningen av detta slags förkortningar är att den gör en svårtillgänglig språkform svårtillgängligare. Det ligger inget paradoxalt i att ett längre uttryck kan vara effektivare från kommunikativ synpunkt. Naturligtvis finns det lägen där förkortningar är funktionella. Fältropet blir alltså inte *Död åt förkortningarna utan Håll förkortandet kort.*

I och med detta är jag inne på formella språkkategorier, och jag fortsätter lite med dem, innan jag går in på några innehållskategorier. Först några ord om stavning. År 1969 motionerade man i riksdagen om en stavningsreform i ljudenlig riktning. Motionen avlogs. Nästa år kom nya motioner. De avlogs också. Utskottet som behandlade ärendet 1970 hade den motiveringen, att ändringar i stavningen bör "anknytas till ett långsiktigt, fortlöpande utvecklingsarbete i intim samverkan mellan

organ, där skol- och undervisningsväsendet, språkvården och massmedia m.fl. är företrädare" (Ståhle 1970). Den linjen gick alltså riksdagen på. Jag menar att riksdagen gjorde det med rätta.

Det är många argument för varsamhet som kan föras fram, och det har skett i åtskilliga sammanhang. Ett är att det bara är omkring 15 % av stavfelen som är hänförliga till avvikelserna från ljudenligheten. Ett annat gäller kontinuiteten i det skrivna och det tryckta språket. Den stavningsdebatt som jag har tagit upp nu är ju inte exempel på en förändring. Det intressanta med denna icke-förändring är att det var fråga om en ganska kraftig framstöt.

Under perioden har det kommit två upplagor av Svenska Akademiens ordlista, den nionde upplagan 1950 och den tionde 1973. I dem finns det en tendens som man kan nämna i detta sammanhang, nämligen en viss återgång till långivande språks stavning: *kliché*, inte med *sch* utan med *ch*, *resumé*, inte med *y* utan med *u*, och *wienerbröd*, inte med *vi* utan med *wie*.

På uttalssidan tycker jag mig kunna konstatera en fortsatt ömsesidig påverkan mellan tal och skrift. Den är till stor del förmedlad av uppläsningsspråket i radio och TV. Ett slående drag i en utveckling åt annat håll är spridningen av uttalet *bucker* för *böcker* (Ståhle 1971). Det ger oss en sämre distinktionsmöjlighet i vissa fall och är därför inte någonting som man vill stödja.

Gun Widmark (1973) har påpekat att uttalet av kort *o* framför *rt* i *fort*, *port*, *skjorta* osv. är expansivt i östra Sverige. Vi har vidare uttalet *dom* av *de* som en expansiv företeelse av annat slag. Låt oss notera att alternativa former, t.ex. *di*, lever kvar.

Inom morfologin finner vi en genomgripande förändring under den här perioden, också med talspråklig bakgrund, nämligen generaliseringen av singularformen hos det finita verbet. Redan under decennierna före 1950 hade många författare gått över till enhetsformen. Pär Lagerkvist har den, Artur Lundkvist har den osv. Harry Martinson var en av dem som växlade mellan de två formerna. Veckopressen hade också gått över före 1950 till största delen. På 1940-talet hade skolorna en varierande inställning. Jag minns att jag fick välja om jag ville skriva *sing* eller *plur* i marginalen i uppsatsboken. Jag valde från början att skriva

*sing*, så jag har aldrig behövt lägga om. Just vid den tiden inträffade den avgörande händelsen att TT införde enhetsformen. Det skedde 1945. Intressant är att se hur den rekommendation som utfärdades av Nämnden för svensk språkvård, som Svenska språknämnden då hette, egentligen ter sig. Sammanfattningen lyder (Wellander 1947):

I material, som icke är av strängt officiell natur, användes i regeln den förenklade verbböjningen, utan särskilda pluralformer av predikatet. I officiella meddelanden och liknande material, som återges i direkt anföring, bibehållas pluralformerna. De böra även bibehållas, då en författare särskilt begär detta.

Nämnden själv formulerade sig 1945 med pluralformer.

I fortsättningen förlöpte utvecklingen snabbt. Riksdagen gick över på 50-talet, i lagspråket gick man över på 60-talet osv. En språkforskare som Carl Ivar Ståhle gick över i sina skrifter omkring 1960. Nu är enhetsformerna allena rådande, och för den yngre generationen är *vi gå* på sin höjd en litterär form – *Vi gå över daggstänkta berg*.

En annan intressant aspekt på verbet är temaböjningen. För rätt många år sedan skrev jag en uppsats som handlade om verbal allomorfi, baserad på korpusen Press 65 (Allén 1967). Jag konstaterade där bland annat att det finns ganska många verb som har saxad böjning, där alltså olika konjugationers böjningssystem växlar med varandra. Nu finns det ett jämförelsematerial från 1976, och det indikerar att tre verb är i rörelse. De är *sprida*, *ge/giva* och *betala*.

Det är formerna i preteritum och supinum som är intressantast. För *sprida* gäller att verbet växlar mellan svag och stark konjugation: *spridde* eller *spredd* respektive *spritt* eller *spridit*. Normalformerna i det äldre materialet var *spredd*, *spritt*. I det yngre har den starka supinumformen seglat upp jämsides, så att vi får normaltemat *spredd*, *spritt/spridit*.

Också *ge/giva* har växling mellan svag och stark konjugation. I preteritum har vi inga belägg på *gedde*, som man dock hör i tal då och då, utan enbart på *gav*. Medan däremot *gett* och *givit* var jämbördiga tidigare, har nu *gett* blivit kraftigt överlägset. Vi får den saxade normalböjningen *ge*, *gav*, *gett* (jämför *se*, *såg*, *sett*

m.fl.). En viss betydelse kan det ha haft, att infinitivformen *ge* har fått en starkare ställning.

Verbet *betala* varierar mellan två svaga paradigm: *betalade* eller *betalte* respektive *betalat* eller *betalt*. Den tidigare normalböjningen *betalade*, *betalt* har fått vika för den osaxade *betalade*, *betalat*.

Ett utslag av det morfologiska tvång, som ligger i sådana här system, såg jag nyligen i Dagens Nyheter. Där stod det en dag på första sidan: *Ingendera av böckerna förståddes av recensenterna*. Vi konstaterar ren tredje konjugation: *förstå*, *förstådde*, *förstått*.

Ett annat påfallande fenomen är våra pronomen i tredje person. Vi får ofta *han*, där standardspråket kräver *honom*. Vi får *de* för väntat *dem* osv. Lärare frågar gärna hur man skall ställa sig till dessa former. Mitt svar är att störningsfri kommunikation inom språksamfundet som helhet kräver att distinktionerna upprätthålls.

I fråga om syntaxen skall jag peka på ett fenomen som ger utslag i konstruktionstypen hos verb av det slag som Martin Gellerstam har behandlat (1980): *uppleva något som positivt* kompletteras med konstruktionen *uppleva att något är positivt*. Jag har tittat på ett annat verb, nämligen *uppskatta* 'beräkna', som har genomgått en liknande utveckling. *Uppskatta något till något* har nu kompletterats med *uppskatta att något är något*. Det finns inget belägg på det i Press 65. I ett likartat material från 1976 (Press 76) får vi *Källor i Washington uppskattar att det finns 100 000 vietnameser kvar i stället för Källorna uppskattar antalet vietnameser till 100 000*. Ett annat exempel är *Man har uppskattat att det kan ta allt mellan 10 och 200 år*. Och häromdagen stod det i Svenska Dagbladet: *Polisen uppskattar att kring 50 000 bilar körs olagligt*. Det är alltså ett mycket starkt formellt mönster som vi har här. Vi kan säkert vänta oss fler verb som tenderar åt samma håll.

Detta om de mera formella tingen. En del av de uppgifter jag har gett och kommer att ge bygger på frekvensstudier och kon-kordanser över stora textmaterial. Jag utnyttjar här tre textkorporusar med tillhörande bearbetningar, framtagna vid Språkdata

i Göteborg och tillgängliga på mikrokort: Press 65, Press 76 och Riksmötet 1978/79. Dokumentation om dessa (och andra) material finns i Allén et al. 1970–80, Allén 1981 och den årligen utkommande Språkdata Research Report.

Om konkordanser och deras användning har Lars Huldén skrivit en dikt, som kommer väl till pass.

Du vet ju, min käre Teophilus,  
att det i himmelen finnes en konkordans  
över alla ord som någon nån gång  
har sagt och/eller tänkt jämte  
tillräcklig kontext för bedömning  
av andemeningen.

Men vad du kanske inte vet  
och inte heller jag visste förr  
är att helvetet har fått en  
mikrofilm av alltihop för  
djävlarernas studiers skull.

Nyligen lär den Fule själv ha  
hållit ett seminarieföredrag  
om orden *fosterland*, *ära*, *svenskhet*  
och vad man kan göra med dem.

Dikten är naturligtvis mångbottnad, men om vi skall hålla oss till vad svenskarna gör med de nämnda orden, så tycks det vara detta. Vid en jämförelse mellan materialen från 1965 och 1976 visar det sig att *fosterland* har gått ner från 7 till 1, *ära* från 46 till 24 och *svenskhet* från 3 till 0. Sammanlagt har frekvenstalet alltså halverats. Det tycks vara så, att man börjar dra sig för att använda sådana ord, och därom vore något att säga i och för sig. Men här får det snarast tjäna som introduktion till en politisk aspekt på språket.

Det är så att Eva Block vid historiska institutionen i Lund har skrivit en rapport (1979) om begreppen 'frihet' och 'jämlighet' under den period vi betraktar. Det material som hon har undersökt är fem ledande morgontidningar, i princip fördelade på samma sätt som i pressmaterialen samlade i Göteborg. Resultaten ges som procentandel ledare för varje år, som med olika uttrycksformer – enskilda ord eller längre uttryck – har nämnt begreppen i fråga. Poängen från projektets synpunkt är att



'frihet' tolkas som ett liberalt honnörsord och 'jämlighet' som ett socialistiskt. Hon får ett resultat som innebär att kurvan för 'frihet' sjunker medan kurvan för 'jämlighet' stiger. Mot slutet av perioden etablerar de kontakt. Detta förlopp tolkar alltså Eva Block som en generell svängning åt vänster.

När jag såg detta, tyckte jag att det var intressant inte minst därför att undersökningen bekräftade en iakttagelse, som jag hade gjort i våra material och som jag bland annat hade nämnt för Lars Westman i en intervju för Vi (1981), nämligen att man tenderar att använda *frihet* mindre nu än tidigare. År 1965 hade vi 231 belägg, 1976 bara 70. *Jämlighet* visar i stort sett oförändrad frekvens, och det innebär att de två talen har närmat sig varandra mycket starkt i materialet.

En aspekt som jag tycker hör hit, men som inte ingår i Eva Blocks undersökning, gäller den könsrelaterade jämligheten, *jämställdhet*. Antalet fall är i stigande, från 3 till 13. Som jämförelse kan man kasta en blick på en bearbetning av riksdagens snabbprotokoll 1978–79. Där ser man att *jämlighet* och *jämställdhet* tillsammans gott och väl kommer upp till den nivå som *frihet* har.

Låt oss betrakta några andra ord och därmed begrepp som speglar viktiga ting i samhällslivet. Jämför man pressmaterialen från 1976 och 1965, finner man – som väntat – att vissa ord används ungefär lika ofta. Till dessa oförändrat aktuella ord hör *demokrati*, *problem* och *öl*.

Till dem som har ökat i aktualitet hör energiord som *olja* och *bensin*, ekonomiska ord som *inflation* och *kris* (som dock har en vidare användning), informationssamhällets ord *information* och *data*, det politisk-administrativa *kommun*, könsrollsdebattens *kvinnor* och det eviga *vädret*.

Ett bistrare intellektuellt klimat indikerar ord på nedåtgående som *forskning*, *intellektuell*, *kunskap*, *litterär*, *teori*. Minskar gör också *kyrka*, en reflex av sekulariseringen. Bekymmersamt nog är likaså de existentiella orden *liv*, *död* och *kärlek* på väg nedåt. Om *frihet* har jag talat tidigare.

Beträffande stilmarkörer som befinner sig i rörelse kan jag nämna, att *ska*, *dom* och *mej* går framåt, medan *icke*, *ej* och

emellertid, jag höll på att säga naturligtvis, går bakåt. *Jobb* ökar på *arbetes* bekostnad (orden är inte helt ekvivalenta). *Fru* ökar, *hustru* minskar.

Här är så några exempel på nyord under perioden. Belägg kan man finna i skrifter från Svenska språknämnden (Mattson 1974, Allén 1982 m.fl.). Bland 50-talsorden kan jag nämna *bandspelare*, *datamaskin*, *diskotek*, *knark*,  *kreativ*, *kärnvapen*, *massmedium*, *moped*, *naturvård*, *skattesmitare*, *spökskrivare*, *stress*, *svartbygge*, *terrorbalans*, *tittarstorm*, *transistor*, *underutvecklad*, *utvecklingsstörd* och *väskryckare*. På 60-talet kom bland andra: *akutintag*, *arbetsmiljö*, *bingo*, *dagmamma*, *extraknäck*, *flexibilitet*, *gatukök*, *grundskola*, *hurtbulle*, *kortkort*, *könsroll*, *ljudvallen*, *lokalvårdare*, *miljövård*, *mobba*, *pendlare*, *postnummer* och *strippa*. Till 70-talsorden hör *akupunktur*, *barnomsorg*, *historielös*, *omlokalisera*, *personbevis*, *pilotstudie*, *ångervecka*, *återanvändning* och *överfiskning*. Jag kan tillägga att Svenska språknämnden har kommit långt med arbetet på en nyordsordbok för just den här perioden.

Också i beteckningarna för de båda könen kan man avläsa förskjutningar. Om man tar pressmaterialen från 1965 och 1976 som underlag, finner man att paret *kvinn*a och *man* håller positionerna, medan *dam* och *herre* är på klar nedgång och *flicka* och *pojke* också sjunker något. Däremot stiger *tjej* och *kille*.

I materialet från 1976 talas det om att det inte finns särskilda *tjejjobb* och särskilda *killjobb*. Ett bildningsförbund i Mölndal ordnade nyligen en kurs som hette "Data för tjejer". Häromdagen var det en intervju i Svenska Dagbladet med en skådespelerska som sade: "Det är ett privilegium att få spela en jätteroll som Karin Månsdotter. Det finns ju så få riktigt stora *tjejr*oller i teaterlitteraturen." Ordet *tjejr* har många olika laddningar, och det är inte normalordet för 'kvinn', men det rör sig i den riktningen.

En annan aspekt på könsrollerna visar frekvenslistan härintill. Den gäller riksmötet 1978/79. Materialet omfattar ungefär fyra miljoner löpande ord, vilket alltså är vad man säger på ett år i riksdagen. Orden i listan är ordnade efter fallande frekvens. Det finns flera intressanta iakttagelser att göra, men låt oss i detta

sammanhang inskränka oss till fördelningen mellan män och kvinnor. Lägg märke till luckan där det saknas en tusentalssiffra. Den drastiska minskningen på manssidan motsvaras naturligtvis av en lika drastisk ökning på kvinnsidan. Frågan blir vad det är för ett ord som visar denna utpräglade snedfördelning. Svaret är *barn*. Tittar vi sedan på fördelningen över de politiska partierna, finner vi att siffrorna är väl spridda, dvs. alla partierna är i stort sett lika goda.

	Tot.	Män	Kvinn.	A	V	S	C	F	M	Opol.
verksamhet	1877	1482	395	32	301	461	337	438	270	38
betänkande	1871	1523	348	47	222	508	347	392	336	19
land	1868	1571	297	148	190	461	322	344	402	1
flera	1850	1509	341	59	268	467	256	505	275	20
ligger	1846	1505	341	65	259	462	270	493	282	15
menar	1837	1472	365	46	334	472	278	379	324	4
dess	1830	1538	292	65	315	470	278	383	307	12
borgerliga	1823	1532	291	50	462	1141	32	21	117	0
sådant	1815	1509	306	29	290	397	302	480	309	8
behöver	1805	1389	416	29	258	505	331	418	255	9
heller	1805	1474	331	55	339	401	313	443	246	8
skäl	1798	1433	365	33	244	420	304	487	297	13
kunde	1755	1459	296	30	242	547	287	405	235	9
emellertid	1734	1447	287	25	147	401	290	524	306	41
viktigt	1734	1282	452	11	171	487	361	477	208	19
debatten	1723	1461	262	53	229	458	313	388	280	2
samhället	1719	1397	322	46	337	560	326	240	195	15
svaret	1705	1379	326	54	206	599	359	227	251	9
stället	1687	1371	316	53	350	441	264	293	283	3
mitt	1686	1372	314	28	169	449	234	565	228	13
verkligen	1673	1300	373	61	247	462	290	260	347	6
talman	1664	1396	268	24	211	493	227	273	435	1
nuvarande	1663	1364	299	82	250	425	244	383	259	20
samtidigt	1654	1335	319	44	259	420	287	393	239	12
klara	1649	1378	271	25	221	534	288	342	233	6
barn	1647	764	883	22	325	310	243	446	297	4
grund	1631	1266	365	27	217	436	267	420	248	16
dock	1611	1316	295	14	195	335	311	508	236	12
sammanhang	1610	1358	252	71	246	353	296	412	209	23
förslaget	1609	1317	292	23	161	461	276	427	230	31
milj	1609	1359	250	17	103	484	218	519	265	3
ni	1609	1484	125	49	344	685	167	179	183	2
möjlighet	1602	1258	344	16	207	367	322	426	256	8
ur	1596	1298	298	50	272	455	265	281	269	4

*Ordfrekvenser i snabbprotokollen från riksmötet 1978/79.*

Ordet *barn* är inte ensamt i sitt slag. Här är några andra ord som är kraftigt överrepresenterade på kvinnosidan och kraftigt underrepresenterade på manssidan: *kvinnor, män, föräldrar, elever, skolan, läkare* och *jämställdhet*. Det motsatta förhållandet gäller för ord som *produktion, sysselsättning, företag, näringslivet, politiken, majoritet, skatter* och *miljarder*. Lägg märke till att denna bild stämmer väl med det som kom fram vid undersökningen av kvinnliga och manliga journalisters språkbruk i materialet från 1965 (Allén 1977). Det man kan fråga sig här är: Skall det bli någon ändring?

En annan sak som hör hit är utvecklingen av könsneutrala ordbildningssuffix: *hon är språklärare, hon är direktör*. Fall som dessa är nu allmänt accepterade. Det intressanta idag är om ursprungligen feminina suffix kan utvecklas på motsvarande sätt. Detta har Thorsten Andersson skrivit om i festskriften till Gösta Holm (1976). Ordet *sjuksköterska* är på god väg. Vi börjar få belägg som *manlig sjuksköterska* och *kvinnlig sjuksköterska*. Och liknande gäller *barnsköterska* och *avdelningssköterska*. I konsekvens härmed får vi uttryck som *syster Hans*. De känns kanske lite ovana men tycks fungera i praktiken. Vi kan i och med detta göra en ny anteckning i språkets annaler.

Så har vi pronomenfrågan. Sedan gammalt är *han* dels könsneutralt, dels maskulint. Något alltigenom könsneutralt personpronomen har vi aldrig haft, och vi har inte fått det nu heller trots flera olika förslag. Vad diskussionen har lett till är uttryck som *han eller hon, han/hon* och kanske en ökad användning av *man, den* och *de* i allmänhet uttalat *dom*. Också *vederbörande* har då och då dammats av. Ibland används *hon* i den aktuella funktionen, men det är ju knappast någon lösning. Som en speciell markering har det förts fram att Gud borde få pronomenet *det*: Vårt Gud är oss en väldig borg, det är vårt vapen trygga. Att skapa ett könsneutralt pronomen vore att ge språket en ny uttrycksmöjlighet som säkert skulle bli mycket utnyttjad.

Under den här perioden gick man långt bland poeterna när det gällde att knåda språket. Bland annat fick vi se ett slags grafisk lyrik, där skrivtecknens formationer på sidorna spelade en väsentlig roll, och vi fick kombinationen diktsamling-grammo-

fonskiva. Ett exempel på detta är Bengt Emil Johnsons Gubbdrunkning från 1965. Där är brottningen med språket långt driven. Bengt Emil Johnson säger att han är ledsen att han inte har kunnat utnyttja den typografiska tekniken så som han egentligen hade velat. Men han gör exempelvis så här:

	Inom förtöjningsområdet, valfria i rymden i himlen:	fiskvatten badvatten sköljvatten tvättvatten dricksvatten sjövatten saltvatten smältvatten mineralvatten bäckvatten
än du då, va?		
du då, va, va, du?		
du, va, du, du, va, du, va, va?		
va, du, va, va, du, va, va, va, du, va?	(tydligt: "gång på gång eftersom förhållandet mellan de olika luc- korna – gluggarna – förskjuts")	
	Stampa fram genom vildmark. Vidbrända, resliga. TÖRSTIGA. Varenda blodbild avtappad, i själva verket tarvliga, kvalitetsmedvetna. "Ja pressa honom mot väggen me domkraft" Ja presentera mej hoppkraft	
prussu dumkruft prusunturu huppkruft		

Många poeter står ju i ett slags hat-kärleksförhållande till språket. Och det tar sig ibland uttryck i att man misströstar om språket som medium. Torgny T. Segerstedt har i en artikel i Artes nyligen tagit upp frågan (1981). Han behandlar nya diktsamlingar av Lars Forssell, Göran Tunström, Bengt Emil Johnson, Lars Norén och Werner Aspenström. I sin sammanfattning skriver han:

Gemensamt för de fem jag studerat är att de alla visar en stark misstro mot språket, som ett medel att förbättra världen. Därför drar de sig undan den sociala verkligheten och söker en annan verklighet. Men inte heller denna andra verklighet förmår språket att fånga, anser de.

Segerstedt slutar med att fråga om detta är något typiskt endast för dessa fem diktare eller om det är ett uttryck för en generations besvikelse. Jag låter frågan gå vidare till er med den enda kommentaren att jag inte tror, att hans fråga har något enkelt svar.

Så är det dags för det som är ofrånkomligt, det engelska inflytandet på svenska språket. Det är ju så uppenbart och omvittnat, att jag bara markerar det med ett litet urval exempel. Vi har *drive*, *festival*, *jeans* och *juice*. Dessa ord stavas i allmänhet på engelskt sätt, även om *juice* kan stavas *jos*. I en del fall har vi ändrat stavningen: *tejp* och *tuff*. Adjektivet *tuff* bildar nu en homograf tillsammans med ett annat ord i språket. Ganska nytt är att *papper* också har kommit att betyda 'föredrag' efter engelskans *paper*. Ordet *topp* har likaså fått ett betydelse tillskott i och med att det betecknar ett klädesplagg. *Leva upp till* och många andra sådana uttryck har också kommit in (*to live up to*). Vi skall inte heller glömma de analytiska sammansättningarna. *Florodol tvålarna* stod det i den gamla annonsen. Typen fanns redan då.

Det är de många anglicismerna som ligger bakom beteckningar för svenska språket som *svengelska*, *Swenglish* och *riks-Swedish*. Förhållandet är som bekant inte unikt. I Frankrike talar man om *franglais*, och i Indien talar man om *Hinglish*.

Men svenskans förhållande till främmande språk har blivit mycket mera komplicerat än så. Kontaktytan har vuxit enormt under perioden, inte bara genom att de internationella kommunikationerna har förbättrats utan också genom att vi har fått omkring 150 invandarspråk. En tidning har berättat om en femårig flicka, som hade finsk mamma och indisk pappa, som gick på svenskt daghem (dagis), och som hörde mamma och pappa tala engelska med varandra i hemmet. Den fråga som ställdes i artikeln var: Vilket språk skall hon lära sig?

Situationer som vetter åt det här hållet är vanliga och naturligtvis ofta besvärliga i det enskilda fallet. Hit hör exempelvis också det förhållandet att svenskan används som kommunikationsspråk mellan invandrare av olika nationalitet, låt oss säga grekisk och jugoslavisk.

Om vi ser det från en mer övergripande synpunkt, anmäler sig frågan: Vad betyder detta för svenska språket? Kan en bristfällig

språkbehärskning få någon inverkan på det svenska språksystemet?

För egen del tror jag inte att inflytandet blir alltför stort. Några mer djupgående grammatiska förändringar i stil med vårt kasus-systems sönderfall under medeltiden verkar inte sannolika. Där-  
emot kommer säkert ordförrådet att tillföras vissa ord, en tendens som givetvis stöds av ett livligt internationellt umgänge.

Särskilt intressant blir det att observera i vad mån nya idiomatiska uttryck på främmande substrat dyker upp. Jag syftar alltså på fall som motsvarar engelsktalande svenskers användning av *it's no idea* (byggt på svenskans *det är ingen idé*) i stället för det korrekta *it's no use*.

Vi hamnar ju principiellt i samma situation som engelskan. I det fallet talar man nu om att engelsmännen och amerikanerna inte längre äger sitt språk. Det fylls av uttryck från alla möjliga håll, och det har ju fyllts på under långa tider. Man har t.o.m. börjat tala om att det finns ett slags allmän europeisk engelska med särskilda idiomatiska drag. Och i den meningen kanske rikssvenskar och finlandssvenskar kommer att få avstå en del av äganderätten till sitt språk.

Nu för tiden gör också andra avtryck i språkets mylla.

## Litteratur

Ahlgren, Perry, Tilltalsordet ni. Diss. Uppsala. 1978.

Allén, Sture, Dator och datamaskin, fackspråk och allmänspråk. (Tankar om språket. Red. Bertil Mode. 1982, s. 104–108.)

Allén, Sture, Herr talman tolv tusen gånger. (Dagens Nyheter 30/9 1981.)

Allén, Sture, Några könsskillnader i journalistspråk. (Könsroller i språk. Utg. av Forskningskommittén i Uppsala för modern svenska. 1977, s. 9–12.)

Allén, Sture, Verbal allomorfi i modern svenska. Göteborgs universitet. 1967. (Dupl.)

- Allén, Sture, et al., Nusvensk frekvensordbok baserad på tidningstext. 1–4. 1970–80.*
- Andersson, Thorsten, Manlig sjuksköterska. (Nordiska studier i filologi och lingvistik. 1976, s. 1–11.)*
- Block, Eva, Frihet och jämlikhet – indikatorer på politisk förändring i Sverige 1945–75. (Utveckling av innehållsanalytisk metod. Red. Lowe Hedman. 1979, s. 397–405.)*
- Gellerstam, Martin, Uppleva som positivt. (Nysvenska Studier 59–60 (1979–1980), s. 86–94.)*
- Gyllensten, Lars, Konst och demokrati. (Artes 1981: 2, s. 126–128.)*
- Huldén, Lars, Långdansen. Dikter 1958–1976 i urval. 1976.*
- Johannisson, Ture, Författningsspråkets förnyring. (Språkform och språknorm. Utg. av Svenska Akademien och Svenska språknämnden. 1979, s. 134–155.)*
- Johnson, Bengt Emil, Gubbdunkning. 1965.*
- Key-Åberg, Sandro, Vem äger språket? (Vad händer med svenska språket. Red. Inge Jonsson. 1976, s. 11–12.)*
- Mattsson, Gösta, Ordlistans tionde upplaga – riktlinjer och kommentarer. (Ture Johannisson & Gösta Mattsson, Svenska Akademiens ordlista under 100 år. 1974, s. 55–116.)*
- Myndighetsrutiner för SLÖR- och matrikelsystem. Statskontoret. 1981.*
- Segerstedt, Torgny T., Ord, jag och verklighet. (Artes 1981: 4, s. 109–122.)*
- Språkdata Research Report. Göteborgs universitet. Utges årligen.*
- Språket i lagar och andra författningar. Statsrådsberedningen. 1967.*
- Ståhle, Carl Ivar, Det nordiska rättstavningsmötet 1869 och hundra års svensk rättstavning. (Språk i Norden 1970. 1970, s. 5–36.)*
- Ståhle, Carl Ivar, ”Mötet uppnas på söndag”. (Studier i dagens svenska. Red. Bertil Mode. 1971, s. 1–15.)*
- Svenska Akademiens ordlista över svenska språket. 9 uppl. 1950. 10 uppl. 1973.*



*Thors, Carl-Eric*, Några iakttagelser om svenskt lagspråk i Finland. (Språkform och språknorm. Utg. av Svenska Akademien och Svenska språknämnden. 1979, s. 246–251.)

*Wellander, Erik*, Vi gå eller Vi går? 1947.

*Westman, Lars*, Nu är kärlek mindre vanligt. (Vi 1981: 14, s. 10–11.)

*Widmark, Gun*, En blick in i dagens språkhistoria. (Svenska studier från runtill till nutid. 1973, s. 221–234.)

## The Lemma-Lexeme Model of the Swedish Lexical Data Base

Empirical Semantics II. 1981.

As a background, some general outlines of the Swedish project Lexical Data Base are presented. This leads up to a discussion of the problem of lexical entry versus linguistic sign. Then the lexical principle developed for the project – the lemma-lexeme model – is sketched. In this model, the lemma covers the form aspects of the lexical entry, i.e., expression and function, whereas the content aspect is covered by the lexeme. Finally, some alternative ways of presentation in dictionaries are commented on.

The paper is based on a paper read at the 7th International Conference on Computational Linguistics, Bergen 1978.

### 1. Lexical data base

The aim of the project, which is called Lexical Data Base, is to establish a well-structured data base of Swedish lexical material, from which dictionaries can be generated and printed and computer dictionaries for use in algorithmic text analysis can be developed. The work is carried out within the framework of modern linguistic theory.

We want the data base to be as systematic a representation as possible. In fact, what I am going to indicate is the outline of a linguistically motivated apparatus for the handling of authentic lexical data on a fairly large scale by means of interactive processing.

The first sketches were made in the late sixties. The development of the project has been influenced by a number of factors. One is the investigation presented in *Frequency Dictionary of Present-Day Swedish 1–4*, which concerns words, collocations, morphemes, and meanings.

I should also like to mention our experiences from the project Algorithmic Text Analysis, particularly from the work on the morphological system. A book on the Morphology of Present-Day Swedish, based on this project, has been published by Staffan Hellberg in our series *Data linguistica*.

A third factor influencing the project is the increasing emphasis on the lexical component in linguistic theory. This approach seems to us to be fruitful. It may also promote the study of other linguistic fields.

A fourth point is the fact that we can count on support from the service organ of our Department. The Swedish Logothèque is responsible for the establishment of text and word banks in machine-readable form. Thus, we have access to an increasing amount of linguistic material.

Fifthly, the accumulation of new knowledge concerning data base problems should be mentioned. The elaboration of a data base in the technical sense of the word provides for multidimensional organization of data. This has undoubtedly opened up new perspectives for linguists.

The project profits from contacts with several other organizations and groups. Among them, the following may be mentioned: The editorial office of the Swedish Academy, the Swedish Language Committee (office headed by Bertil Molde), the Swedish National Board of Education, the Bielefeld group led by János Petöfi, and the Bergen group led by Kolbjørn Heggstad.

The lexical data base should be linguistically dynamic. One aspect of this is that it should cover the ample possibilities of compounding and derivation that are so typical of Swedish and some other Germanic languages. This requires an explicit morphological model. Such a model has been provided by the earlier projects mentioned. There is also a dynamic aspect to the semantic definitions which I shall touch upon later.

The syntactic model decided on is an adapted version of case grammar. It seems to be favourable because it is sufficiently surface-oriented and suitably detailed, from a semantic point of view, to be used in actual analysis. Notice also that an argument frame is a mini-version of a semantic frame. We take advantage

of the grammatical model in the design of the types of definition, too.

The lexical entries of the data base are referred to three groups: the defining vocabulary (DV), the fully defined vocabulary (FDV), and the paraphrased vocabulary (PV). This division does not imply any difference in status between the groups, just a difference with regard to the amount of information given.

The definitions are made up of words from the defining vocabulary or of words reducible to this vocabulary. The units of the defining vocabulary are arrived at, chiefly, by the paraphrasing of senses. These units are treated as axiomatic elements. They are used in order to avoid circularity in the FDV definitions. Notice that the fact that the definitions are non-circular does not exclude the possibility of representing various types of semantic connexions between different units by means of linking.

The linguistic information aimed at includes graphonomic and phonetic data, morphotactic properties, semantic definitions, paraphrases (including synonyms), grammatical constructions, phraseological and stylistic data.

The lexical data base is conceived of, basically, as a linked network. It is being tailored for the specific purpose of the project and implemented on the medium-size computer of the Department by successive development.

Interactive processing is the technique most extensively used in the project. The linguistic information encoded, using a fixed format, is temporarily stored in working files. At regular intervals they will be loaded into the data base proper, which is then primarily used only for inspection and comments. The problems of security and back-up are reduced in this way. The most important type of elements to be linked together is the canonical forms of the morphemes of the lexical entries.

In addition to the linking of elements, there are category lists produced by batch processing. These lists contain classes like parts-of-speech, which need not be immediately available through linking. A third type of access is free search for character strings in explicit parts of the data base.

## 2. Linguistic sign

A fundamental question is the character of the lexical entries. In an ideal world, the linguistic sign – the solidarity between an entity of expression and an entity of content – would be the obvious unit. It seems, however, that this concept loses some of its usefulness when confronted with the intricacy of actual lexical material. This is due not least to the non-trivial interplay between expression and function, on the one hand, and content, on the other.

Ever since Saussure, the linguistic sign has been a recurrent topic of discussion. It has been discussed in terms of the nature of concepts like sign, symbol, signal, icon, etc., by Malmberg, Eco, and others; in terms of the arbitrariness of the sign by Bloomfield, Roman Jakobson, and others; in terms of the form and substance of expression and content, respectively, by Hjelmslev and others; in terms of texts as signs by scholars in text linguistics, and so on.

To put it somewhat bluntly, my general impression is that through much of this discussion the linguistic sign has been looked upon as a fairly neat and well-defined little thing.

Everybody is probably aware of the fact that real language is not like this. It is extremely complicated, intermingled, full of surprises, irregular, and what not. A common example like *tree* looks simple enough. But what about a verb like *go* or *put* or even *count*, etc., or *die*, if you like? A closer look at *tree* would, in fact, reveal a fair amount of problems, too.

Nevertheless, the linguistic sign is a useful theoretical notion. You had better keep an eye on it in your lexicological work. My point is that the linguistic sign and the lexical entry do not generally coincide. I shall try to give some reasons for this view.

## 3. Lexical entry

In e.g. Malmberg's model (1973, p. 27) the linguistic sign is the morpheme, which is also the lexical unit. Morphology and syn-

tax are descriptions of the behaviour of the morpheme/linguistic sign/lexical unit. But it seems clear to me that the lexical entry, at least in the class of languages I have in mind, is, normally, a word. The native speaker expects to find, and is entitled to find, e.g. *decision* as an entry in the lexical data base, not some linguistically motivated medial segment of it like *-cis-*, perhaps under *-cide* from *decide*.

In the lexical unit, function constitutes an integral part. Whereas it is possible – evidently – to discuss the linguistic sign as a binary combination of expression and content, it is not to the point to discuss the lexical unit within such a framework. Inherent in the lexical entry is its potential in word-formation, phraseology, syntactic behaviour, and inflexion.

There is also an important associative influence of expression on content in the lexical unit. A case in point is Swedish *sticka*. There are two verb types, one with strong and one with weak inflexion. Together they have four basic senses. The senses are, in a reasonable order, 'to prick; to knit; to contrast; to go off'. They are, however, never presented in this order in dictionaries. Why? Because the second sense alone goes with weak inflexion and, accordingly, is always kept apart. This phenomenon seems to be relevant from a psychological point of view. The solidarity of expression and content is thus not lexically autonomous in cases like this. In other words, you do not just pick an item of expression and an item of content and treat them independently of other such combinations.

There is, in addition, the more or less accidental difference in parallel cases between even closely related languages. As an example, consider *deep* and *depth*. In English, there are by definition two linguistic signs in this case, differing with respect to expression but not with respect to content. There are also two lexical entries. In Swedish there is just one sign *djup*, covering the two English cases. Morphology and syntax tell us, however, that there are two lexical entries in Swedish, too.

Having established the lexical entry as a unit in its own right, let us note that traditional lexicography has, generally speaking, not devoted itself to a sufficient analysis of this fundamental

question. On the whole, I think it can be characterized as quasi-etymological. The dilemma of this approach is the difficulty of applying etymological criteria in a consistent and psychologically reasonable way to an actual lexical system.

A few examples from the Glossary of the Swedish Academy will illustrate my point.

One lexical entry, two senses:  
*bankett* n. banquet; verge (of road)  
*bestrida* v. deny, contest; defray, pay

In each case, the two senses are very different indeed. If the verb is followed by *cost*, the senses are contradictory.

Two lexical entries, two senses:  
 1 *balja* n. tub, bowl  
 2 *balja* n. kind of seed-vessel, pod

Although these items are treated as two entries, you can easily see the relation: various kinds of vessels.

Three lexical-entries, four senses:  
 1 *kapell* n. cupboard for vaporization  
 2 *kapell* n. cover  
 3 *kapell* n. chapel; band of musicians

Why three entries?

#### 4. Lemma-lexeme model

I think that two linguistic tools are helpful in handling the problem of lexical entries in defining dictionaries: one to take care of expression and function, one to take care of content. Using them and presenting the interplay between them, the linguist is in a position to arrive at and provide some basic insights into a lexical system.

As regards expression and function we have already the *lemma*. In my definition, given in the mid-sixties (1967, p. 86), it is based precisely on criteria from morphology, syntax, phono-

logy, and graphonomy (1970, p. XXXV): "A lemma is defined as a group of forms within a word-class which are assignable either to one and the same series of inflection (in the case of indeclinable words comprising only one form) or to several series of inflection that converge in speech and/or in writing, the divergences of which show purely facultative variation (free variation)." Figure 1 shows how the lemma status of inflectional series is established (1970, p. XXXVI).

An example indicating the relation between the levels of graphic words, homograph components, and lemmas, respectively, is shown in Figure 2 (1971, p. XXXII). This lemma concept has been used, in certain cases with some modification, in a number of text-oriented quantitative investigations.

On the content side we establish the *lexeme*. It can be defined as a kernel sense and the semantic domain around it specified by a system of semantic modification rules. Thus, the picture looks like Figure 3.

The lexical graphic word *sticka* is involved in three lemmas: one noun and a strong and a weak verb. These three lemmas are involved in six lexemes, i.e. we are faced with six lexical entries. Notice that in this simplified example the lexemes are not defined but indicated through paraphrases (synonyms). The borderline between homography and polysemy is also drawn in the picture. In my view, the former is based on expression and function, the latter on content.

Some types of kernel sense modification within a semantic domain should be mentioned as examples of this rule system. One type is analogical expansion in a metaphorical direction. By way of example, *veck* 'fold' is recognised in *invecklad* 'complicated'.

Another type of modification is generalization from a certain material to an object made of or consisting of that material. The unit *järn* thus includes the related senses of 'the metal iron' and 'tool made of iron'.

A third type is expansion through the addition of specialized shades of meaning. To this type belongs, for example, *väg* 'way' in *järnväg* 'railway'.



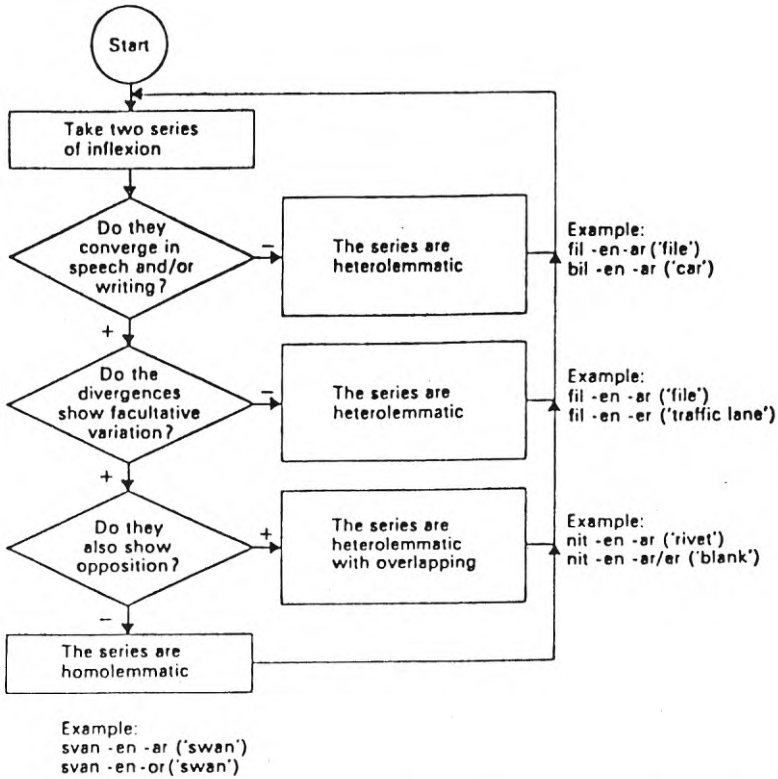


FIGURE 1. *Determination of lemma status of inflectional series.*

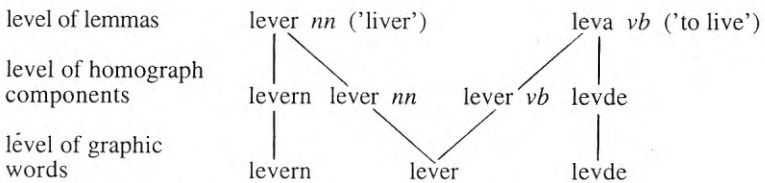


FIGURE 2. *Three levels of description.*

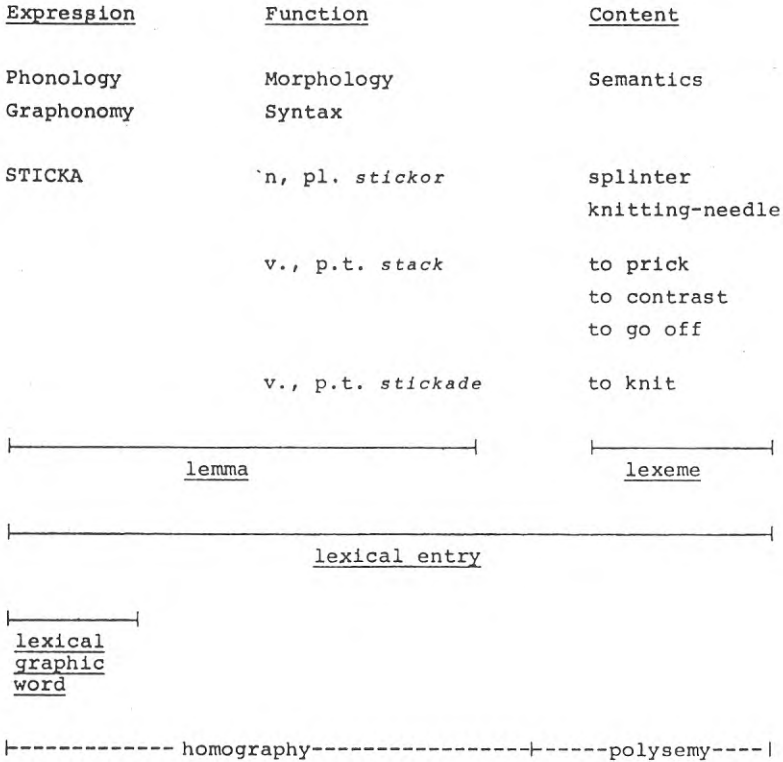


FIGURE 3. *Lemma, lexeme, lexical entry.*

By now I have suggested the outline of what I call the *lemma-lexeme model*. It enables us to specify the role of expression and function, on the one hand, and the role of content, on the other, in a fairly consistent and, at best, psychologically motivated way. It is also easy to stress different aspects of the framework for different purposes.

Thus, the presentation of those parts of the lexical data base that I have been concerned with here can be designed in, e.g., one of the alternative ways illustrated in Figure 4.

- 
- (1) 1 STICKA n., pl. *stickor*  
       1 splinter  
       2 knitting-needle
- 2 STICKA v., p.t. *stack*  
       1 to prick  
       2 to contrast  
       3 to go off
- 3 STICKA v., p.t. *stickade* to knit
- (2) 1 STICKA n., pl. *stickor* splinter  
       2 STICKA n., pl. *stickor* knitting-needle  
       3 STICKA v., p.t. *stack* to prick  
       4 STICKA v., p.t. *stack* to contrast  
       5 STICKA v., p.t. *stack* to go off  
       6 STICKA v., p.t. *stickade* to knit
- (3) STICKA  
       1 n., pl. *stickor*  
       1 splinter  
       2 knitting-needle
- 2 v., p.t. *stack*  
       1 to prick  
       2 to contrast  
       3 to go off
- 3 v., p.t. *stickade* to knit
- 

FIGURE 4. *Presentation of lexical entries.*

First of all it should be observed that there are six lexical entries in each of the three cases. The mode of presentation is a matter of typography and layout in addition to linguistics. Personally, I prefer the first alternative illustrated because it takes advantage of the potential of both the lemma and the lexeme and presents these two aspects in a clear way.

We think that a data base established along these lines may be useful in several respects. The lexical information can be approached from different angles. It can be processed for different purposes. Interesting experiments can be performed. A dictionary of contemporary Swedish can and will be generated from the data base.

**References**

- Allén, Sture, *Frequency Dictionary of Present-Day Swedish*. 1. Graphic Words, Homograph Components. Stockholm 1970. (Data linguistica. 1.)
- Allén, Sture, Förhållandet mellan tal och skrift. (The relation between speech and writing.) *Språk, språkvård och kommunikation*. Uppsala 1967, pp. 76–92 .
- Allén, Sture, Text-Based Lexicography and Algorithmic Text Analysis. *Association for Literary and Linguistic Computing Bulletin* 5:2 (1977), pp. 126–131.
- Hellberg, Staffan, *The Morphology of Present-Day Swedish. Word-Inflection, Word-Formation, Basic Dictionary*. Stockholm 1978. (Data linguistica 13.)
- Malmberg, Bertil, *Teckenlära. En introduktion till tecknens och symbolernas problematik*. (Semiotics. An introduction to the problems of signs and symbols. ) Stockholm 1973.

## Tjugohundratalet och några andra tidsord

Språkvård 1981: 2. Uppsatsen har tidigare publicerats i Festskrift till Carl-Eric Thors. 1980.

Vad skall vi kalla nästa århundrade? Och vad skall första årtiondet av varje århundrade kallas? Dessa och närliggande frågor tas upp i denna artikel.

### Första årtiondet

Upprinnelsen till den här uppsatsen var en fråga som anmälde sig under arbetet på Förnamnsboken. I den ges bland annat uppgifter om förnamnens fördelning över decennierna från 1800-talets slut till 1970-talet. Beteckningarna för dessa årtionden är i allmänhet oproblematiska: *90-talet*, *30-talet*, *70-talet* osv. Men vad skall man kalla decenniet 1900–1909?

I ett sådant sammanhang kan man ju inte gärna lista sig ut genom de vanliga kryphålen: *det första årtiondet under 1900-talet*, *1900-talets första decennium*, *det första decenniet av tjugonde århundradet* osv. Det behövs andra tag.

Varför då inte ge detta undanskuffade årtionde samma villkor som andra årtionden? De har fått namn efter de två sista siffrorna i beteckningen för respektive årtiondes första år. För det aktuella decenniet ger principen alltså *00-talet* med alfanumerisk skrivning, *nollnolltalet* med alfabetisk skrivning. Det är så decenniet har fått heta i Förnamnsboken.

Systematiken i beteckningarna för decennier är sålunda fullt klar. Tabell 1 ger en översikt. Själva året 1900 kallas ju ofta *00*. Detta gäller nog mest i förbindelsen "(någon är) född 00". Också från detta håll finns det med andra ord en inkörsport till *00-talet*.

Naturligtvis har problemet känts av tidigare. Litteraturvetenskapens dilemma är exempelvis ingen nyhet. Låt oss se hur E. N. Tigerstedt behandlar tiden kring sekelskiftet.

TABELL 1. Årtionden.

<i>Alfabetiskt</i>	<i>Alfanumeriskt</i>	<i>Numeriskt</i>
(artonhundra)- femtioalet	50-talet	50–59
sextioalet	60-talet	60–69
nittioalet	90-talet	90–99
(nittonhundra)- nollnolltalet	00-talet	00–09
tioalet	10-talet	10–19
femtioalet	50-talet	50–59

Först skall vi kanske vara överens om att decennierna som termer mera avser litterära strömningar än precisa tidsavsnitt. Men samtidigt är det ovedersägligt att anknytningen till årtionden är stark. En strömnings genombrottstid blir faktiskt namngivande.

Några citat ur Svensk litteraturhistoria (4 uppl., 1971) får belysa frågan. Den aktuella kapitelrubriken lyder (s. 337): "Europeisk realism och nationell romantik: åttiotal, nittioal, tioal". Perioden omfattar tiden 1879–1919, alltså fyra decennier. Rubriken anger tre strömningar/årtionden.

"Vad beträffar tioalet som den nya litterära generationen av praktiska skäl oegentligt döpts av K.O. Bonnier och Ludvig Nordström – kom det överhuvud aldrig till någon öppen brytning mellan det och nittioalet ..." (s. 340). Vi får alltså i själva verket en grupp oegentliga tionalister vid sidan av de senare, egentliga tionalisterna. De praktiska skäl som nämns bör vara språkliga.

Om Hjalmar Söderberg och Bo Bergman, "stadens sångare", heter det (s. 407): "Ofta ha de räknats till tionalisterna, den författargeneration som efterträdde nittioalet, men både kronologiskt och ideologiskt höra de till en äldre tid." Här används sålunda beteckningen *tionalist* utan reservation i sin vida mening.

Med denna användning av termerna blir det möjligt att skriva följande sammanfattning (s. 346): "Åttiotalet, nittioalet och tioalet utgöra tillsammans den första inledande och grundläggande perioden i modern svensk kultur och litteratur."

En poet som Anders Österling blev ju mycket uppmärksammad under 1900-talets första decennium för sina naturlyriska dikter. Bland verken från detta årtionde är Preludier 1904, Offerkransar 1905 och Årets visor 1907. För Vilhelm Ekelund gäller på motsvarande sätt att han bland annat publicerade Vårbris 1900, Melodier i skymning 1902 och Dithyramber i aftonglans 1906.

Motiveringen till att författare som dessa placeras under rubriken "Tiotalet" är denna (s. 419): "De författare det här närmast är fråga om debutera visserligen i de flesta fall redan under det föregående decennium. Men det är först omkring 1910 som de ha funnit sig själva ..." Enligt mitt synsätt är det diskutabelt att på litterära grunder hävda en sådan mening om författarna till Årets visor och Dithyramber i aftonglans. Vad Tigerstedt här försöker sig på är att göra författare som av språkliga skäl oegentligt har fått beteckningen tiotalister till egentliga tiotalister. Systemtvånget är tydligen mycket starkt.

Det blir nu intressant att se på ett aktuellt uttalande av Österling själv. Av en presentation av ett TV-program, skriven av Per Wästberg, framgår det att han betecknar sig som "en gengångare från generation 00" (DN 20/11 1978). Det är väl uppenbart att han syftar på det första årtiondet, som ju faktiskt var hans genombrottstid. Han placerar sig i 00-talet och inte i 10-talet.

Det finns också enstaka belägg på beteckningen *00-talet*. Carl G. Laurin skriver i sin bok *Minnen 1898–1908* (1931, s. 7): "Mina 'Minnen' från det tiotal, som Strindberg, vilkens världsvy helt och hållet skilde sig från min, något ohövligt kallat 00-talet, behandla början av min författarbana ..." Språkprovet återfinns i förkortad form i Östergrens ordbok under *tiotal* som benämning på 1910-talet, alltså av allt att döma felplacerat. Vi kan konstatera att Strindberg har iakttagit tidssystematiken.

Laurin antyder också möjligheten att väcka de negativa associationer som ordet *nolla* och den dåtida beteckningen *nollnoll* för 'klosett' erbjuder. Denna principiella möjlighet tycker jag är en ringa belastning för *00-talet* också i en funktion som uttryck för en litterär strömning. Värderingen av denna kommer att växla genom åren precis som fallet är med andra strömningar. Ett känt

exempel från målarkonsten är det ursprungligen förlöjligande *impressionist* som etikett på färgkonstnärer i en tid då linje och färg skulle hållas isär.

Intressant nog har jag alldeles nyligen fått *00-talet* belagt också hos en nutida Strindbergkännare. I en promemoria inför utgivningen av Strindbergs samlade skrifter i Statens Kulturråds regi skriver huvudredaktören Lars Dahlbäck: "Det gäller bland annat ungdomsdramerna från 70-talet ... broschyrerna och artiklarna om teatern från slutet av 00-talet ... samt de språkvetenskapliga skrifterna från de sista levnadsåren." Ordet fungerar bra.

När beteckningen *00-talet* väl har etablerat sig, är marken beredd för avledningar som *00-talist* och *00-talism* att användas vid behov.

Det finns naturligtvis anledning att ta ett stickprov på den historiska sidan också. I Alf Henrikssons och Björn Bergs Alla Tider (1978) är *1900-talet* och motsvarande beteckningar systematiskt tvetydiga. Ett särskilt belysande fall är *900-talet e.Kr.* Som sidrubrik betyder detta århundradets första decennium. Texten under rubriken börjar: "Från 900-talets början till 1200-talets slut är *Kina* ett politiskt splittrat land." Här betyder *900-talet* hela seklet. Konsekvensen blir att liksom rubriken på uppslagets högersida *910-talet* utläses *niohundredratiotalet*, så bör rubriken *900-talet* på vänstersidan för klarhetens skull utläsas *niohundraollnolltalet*.

Vi har ingen konvention för att klara denna ambiguitet i den alfanumeriska skrivformen. Eventuellt skulle man kunna använda en apostrof som markering, när man syftar på 00-talet i ett århundrade: *9'00-talet* för första decenniet, *900-talet* för hela århundradet.

I Henrikssons fall kompliceras bilden något av att 900-talet i boken omfattar perioden 901–910, inte 900–909 som vanligen är fallet i modernt språkbruk. Detta hänger kanske samman med att vi ju som han påpekar (på sidan med rubriken Första årtiondet e.Kr.) inte räknar med något år 0. De första tio åren blir därför 1–10, och tiotalet börjar med år 11. För egen del vill jag i stället betrakta det första decenniet som ett kortdecennium omfattande



åren 1–9. Redan i och med tiotalet 10–19 är vi då inne i det rådande språkbrukets rytm.

## Första århundradet och tjugoförsta

Liksom ett århundrade inleds med ett årtionde inleds ju ett årtusende med ett århundrade. I detta större perspektiv avtecknar sig liknande frågor.

Vi kan ta 1984 som utgångspunkt. Som tal betraktad har denna sifferkombination den vedertagna benämningen *ettusen-niohundraåttiofyra*. Icke så som årtal betraktad. I det fallet heter det, tre stavelser kortare, *nittonhundraåttiofyra*. Vi kan alltså iakta en talprincip och en årtalsprincip.

Med århundradena är det på samma sätt som med årtalen. Vi lever på *nittonhundratalet* och inte på *ettusen-niohundratalet*.

Om vi så startar en nerräkning från nittonhundratalet, kommer vi via artonhundratalet osv. till elvahundratalet och – tiohundratalet. Principen ger oss alltså möjligheten att skilja mellan *tiohundratalet*, åren 1000–1099, och *(et)tusentalet*, åren 1000–1999. Det finns all anledning att utnyttja denna möjlighet. Så sker också i viss utsträckning.

Den logiska motsvarigheten på årtalssidan är att använda beteckningen *tiohundra* för år 1000. Man ser ju runda tal som 1000, 1100, osv. normalt såsom inledningar till århundraden. I den mån man anlägger årtusendenas perspektiv blir beteckningen *(et)tusen* mer motiverad.

Räknar vi vidare baklänges, kommer vi snart ner till perioden år 1–99, det första århundradet i det första årtusendet.

Den konsekventa benämningen blir *nollhundratalet*. Det är hela tiden siffran eller siffrorna före de två sista siffrorna i årtalen inom seklet som är namngivande. I det här fallet tar vi alltså fasta på att 1–99 är ekvivalent med 001–099.

Går vi nu i stället framåt i tiden, ställs vi inför den aktuella frågan vad det tjugoförsta århundradet skall kallas. Mot den bakgrund vi har skaffat oss är det rimliga svaret *tjugohundratalet*.

Systematiken i beteckningarna för århundraden sammanfattas i tabell 2.

TABELL 2. Århundraden.

<i>Alfabetiskt</i>	<i>Alfanumeriskt</i>	<i>Numeriskt</i>
nollhundralet	000-talet	001–099
etthundralet	100-talet	100–199
niohundralet	900-talet	900–999
tiohundralet	1000-talet	1000–1099
elvahundralet	1100-talet	1100–1199
nittonhundralet	1900-talet	1900–1999
tjugohundralet	2000-talet	2000–2099
tjuogoetthundralet	2100-talet	2100–2199
tjugoniohundralet	2900-talet	2900–2999
trettiohundralet	3000-talet	3000–3099

För fullständighetens skull anger jag systematiken för årtusendena i tabell 3.

TABELL 3. Årtusenden.

<i>Alfabetiskt</i>	<i>Alfanumeriskt</i>	<i>Numeriskt</i>
nolltusentalet	0000-talet	0001–0999
(et)tusentalet	1000-talet	1000–1999
tvåtusentalet	2000-talet	2000–2999
niotusentalet	9000-talet	9000–9999
tiotusentalet	10000-talet	10000–10999
elvatusentalet	11000-talet	11000–11999

De alfanumeriska beteckningarna för *nollhundralet* och *nolltusentalet*, *000-talet* respektive *0000-talet* verkar kanske inte omedelbart anslående. De finns där som en systematisk möjlighet för den som söker.

Precis som när det gäller *tiohundralet* respektive *(et)tusentalet* bör vi skilja på *tjugohundralet*, 2000–2099, och *tvåusen-*

talet, 2000–2999. Årtalet 2000 kan kallas *tjugohundra* eller *två-tusen* enligt det resonemang som fördes om år 1000.

Vad tjugohundratalet skall heta blev aktuellt för de danska, norska och svenska språknämnderna 1968 genom en fråga till den danska nämnden. Jag har fått kopior av de resulterande breven genom de svenska och danska nämndernas försorg, vilket jag tackar dem för.

Den danska nämnden framhåller, att *2000-talet* ”vistnok alltid” utläses *totusindtallet* men säger sig kunna acceptera att det utläses *tyvehundredetallet* (9/5 1968). Den norska nämnden skriver, att man har samma praxis i norskan som i danskan, vilket skulle ge *totusentalet*, men att man också kan tänka sig *tjuehundretalet* (20/3 1968; också Årsmelding 1968, s. 15). Den svenska nämnden rekommenderar i sitt svar *tjugohundratalet* med hänvisning till dels att *1000-talet* alternativt utläses *tiohundratalet*, dels att *tvåtusentalet* kan beteckna hela tredje årtusendet (13/3 1968). Varför inte försöka samlas kring *tyvehundredetallet/tjuehundretalet/tjugohundratalet* i Norden?

Jag har i olika sammanhang noterat både benämningen *tjugohundratalet* och benämningen *tvåtusentalet* för nästa århundrade. Man kan säga att språkbruket står och väger för närvarande. Det finns anledning att fästa språkbrukarnas uppmärksamhet på systematiken, så att den enklare och klarare benämningen *tjugohundratalet* får sin rättmätiga chans att slå igenom. Och det bör göras nu. Tiden går.

## Litteratur

- Allén, Sture, & Wåhlin, Staffan, Förnamnsboken. 1979.  
 Dahlbäck, Lars, Utgivningsplan för August Strindbergs Samlade skrifter. Statens Kulturråd 1979-05-15. (Dupl.)  
 Henriksson, Alf, Alla Tider. Kronologisk översikt av världshistorien med teckningar av Björn Berg. 1978.  
 Laurin, Carl G., Minnen 1898–1908. 1931.  
 Norsk språknemnd. Årsmelding 1968.  
 Tigerstedt, E. N., Svensk litteraturhistoria. 4 uppl. 1971.

Wästberg, Per, Österling, lyckligt produktiv. Dagens Nyheter  
20/11 1978.

Östergren, Olof, Nusvensk ordbok. 1919–1972.

## Herr talman tolv tusen gånger

Dagens Nyheter 30 september 1981.

Utomlands möter man ofta den uppfattningen att Ingmar Bergmans filmer ger en sann bild av Sverige. Som allmänt påstående är det naturligtvis diskutabelt. Lika litet kan man säga att riksdagens språk ger en sann bild av svenskan. Men det hindrar ju inte att båda fenomenen är högst intressanta.

Under riksmötet 1978/79 hann man med att säga ungefär fyra miljoner ord. Det är en ordmassa som gott och väl svarar mot femtio romaner. Anmärkningsvärt är att man inte använde mer än 100 000 olika ordformer. Samma antal får man om man tar en textmängd av en miljon löpande ord ur tidningsspråk. Det ser alltså ut som om tidningsnumren på ett mera varierat sätt speglar verkligheten än vad riksdagsdebatten gör. Det är i och för sig rimligt.

Den undersökning som uppgifterna bygger på gäller språket i riksdagens snabbprotokoll. Den utförs vid institutionen för språkvetenskaplig databehandling (Språkdata) vid Göteborgs universitet. Texten har kommit till institutionen i datamaskinellt läsbar form från tryckeriet i Riksdagshuset.

Gunnar Ljusterdal och Gunnar Richardson har från olika håll – riksdagsstenografens respektive historikerns – pekat på att det finns skillnader mellan vad som faktiskt sägs i riksdagen och vad som står i protokollet. Principen är dock, att ändringar skall göras så varsamt som möjligt och bara när de är ”uppenbart nödvändiga” som det står i direktiven. De berör på det hela taget inte ordförrådet i någon större utsträckning. Under alla förhållanden är den språkform man finner i protokollen mycket betydelsefull. Det rör sig om en huvudtyp av offentligt språk.

## Jobb och arbete

Man kan förstås ställa många olika frågor angående materialet. Vad finns det för politiskt betingade språkliga skillnader mellan olika grupper? Finns det några könsbetingade skillnader? Vad är allmänt sett vanligt och ovanligt? Vad talade man egentligen om? Osv. Några glimtar följer.

Bland de politiska partierna yttrade vänsterpartiet kommunisternas och folkpartiets representanter överlägset mest i förhållande till partiernas andel av riksdagen. Det hänger förmodligen ihop med att vänsterpartiet kommunisterna har en liten andel av riksdagens ledamöter och att folkpartiet var ensamt regeringsparti under riksmötet. Det är också att märka att vänsterpartiet kommunisterna inte satt i utskotten.

Moderaternas andel av protokollstexten svarar mot deras andel av riksdagen. Centern och socialdemokraterna har värden som ligger ganska långt under andelen riksdagsledamöter. Den klart flitigaste talaren var Jörn Svensson.

Många ord är snedfördelade över de båda politiska huvudgrupperna. Medan exempelvis *privata* (*de privata företagen* m. m.) är dubbelt så vanligt på den socialistiska sidan som på den icke-socialistiska, förhåller det sig tvärtom med *enskilda* (*de enskilda människorna* m. m.). Ordet *jobb* (t. ex. *flera tusen nya jobb*) är också dubbelt så vanligt på den socialistiska sidan som på den icke-socialistiska, medan det förhåller sig tvärtom med *arbetet* (t. ex. *arbetet med att ...*). En klar markering innebär ordet *borgerliga*. Det används ungefär tio gånger så mycket på den socialistiska sidan som på den icke-socialistiska.

Andelen kvinnor i riksdagen under riksmötet var 23 procent. Av ordmässan i anförandena svarade kvinnorna för 19 procent. De kom alltså inte riktigt upp till sin proportionella del.

Ord som är kraftigt överrepresenterade bland kvinnorna är t. ex. *barn, kvinnor, män, föräldrar, elever, skolan, läkare och jämställdhet*. På motsvarande sätt är ord som *produktion, sysselsättning, företag, näringslivet, politiken, majoritet, skatter* och *miljarder* överrepresenterade på den manliga sidan. Är vi en majoritet här i landet som hellre skulle ha sett en annan bild?

Det obesträtt vanligaste ordet är *att* som förekommer 160 000 gånger. Den kraftiga dominansen signalerar ett språk rikt på bisatser. Man kan jämföra med att *och* ligger främst i det material som presenteras i "Nusvensk frekvensordbok", baserad på tidningstext. De mest använda substantiven är inte oväntat *herr*, *talman*, *regeringen* och *fråga*.

Ett ord som *tyvärr* är också mycket vanligt, och det gäller inte minst förbindelsen *tyvärr inte*. Riksdagsledamöterna tänker i hög grad för oss: ord som: *tänka*, *tänkande*, *tänkare*, *tänkbar*, *tänkesätt* och *tänkvärd* förekommer omkring 2 000 gånger sammanlagt. Man *tänker efter*, *tänker igenom*, *tänker om*, *tänker till*, *tänker ut* och *tänker över*.

Engångsorden bildar som vanligt i sådana här sammanhang en mycket stor grupp. Dit hör *tältmiljö*, *Tännäsområdet*, *tättningslister* och *tävlingshysteri*. Man har också varit inne på en *tyghund* (av äkta plysch) och talat om (planverkets) *typgodkännande-verksamhet*.

Det politiska blomsterspråket vore värt ett eget kapitel. Låt vara att det ofta rör sig med välkända figurer: *det har ingen talare vågat ta i ens med tång*; *Han blir trampad på tårna*. Ibland blir det riktigt uppfriskande: *trots detta staplar utskottet de goda tårtorna ovanpå varandra*.

En systematisk listning av de fyra miljonerna ord med tillhörande kontexter (med andra ord en konkordans) är tillgänglig på mikrokort. Det är inte bara ett komplett register till riksdagsdebatten. Tillsammans med andra bearbetningar utgör den ett material som lockar till framstötter på många olika områden: språkvetenskap, samhällsvetenskap, argumentationsanalys, informationssökning osv. Till grund för användningen ligger ett beslut av datainspektionen.

## Hur djupt går SAOB?

Nysvenska studier 59–60. 1980.

När man arbetar tätt in på en text, gör man observationer av många olika slag. Naturligtvis gäller de i första hand texten själv med alla dess egenskaper och egenheter. Också den bakomliggande verkligheten, den kulturella och den materiella, träder efterhand fram – och måste fås att träda fram för att texten skall kunna avläsas. Observationerna gäller likaså de instrument man utnyttjar: ordböcker, grammatiker, monografier och mycket annat.

Det ojämförliga redskapet när man arbetar med text från senare århundraden är Ordbok över svenska språket utgiven av Svenska Akademien, SAOB. Ordboken är inte bara ojämförlig utan också outhärlig. Noggrannheten i redaktionens arbete i fråga om både källkontroll och lexikalisk analys är omvittnad.

Emellertid kan man inte begära att ens SAOB skall ha svaret på alla frågor inom sitt område. Med tanke på de datamängder det gäller kan man knappast heller vänta att varje enskilt besked skall vara helt korrekt. Det kan därför vara av intresse att lite mera ingående relatera SAOB till en text från äldre nysvensk tid.

Vid mitt arbete med Johan Ekeblads brev till brodern Claes Ekeblad 1639–1655 gjorde jag åtskilliga iakttagelser av SAOB:s förhållande till texten. I inledningen till editionen, som publicerades 1965, redovisade jag i korthet några.

Det allmänna läget karakteriserades så här (s. XLI): ”SAOB har låtit excerpera Ekeblads brev efter Sjöbergs utgåva. I efterhand har viss kontroll i originalen företagits. Denna har dock inte omfattat alla språkproven och har där den skett främst inriktats på uppslagsorden. Ordbokens exempel från Ekeblad företer på grund härav påtagliga brister. Tillägget ’rättat efter hskr.’ innebär i allmänhet inte, att språkprovet överensstämmer med originalet. Sedan några år kollationeras emellertid genomgående det Ekebladmaterial som utnyttjas.”



Eftersom den nämnda utgåvan av N. Sjöberg (1–2, 1911–15) tillhör textsituationen, kan en kort kommentar vara på sin plats. När första delen kom – det är den som är aktuell här – hade SAOB hunnit till B. Textutgåvan mottogs mycket negativt och betecknades bland annat som inkonsekvent och godtycklig av Gustaf Jacobson i *Historisk tidskrift* (1913, s. 98). Här följer tre exempel på Sjöbergs textbehandling.

I ett brev skriver Johan Ekeblad om drottningens svårighet att tåla mat och anger enligt Sjöberg som orsak ”dessa många och kloka doctorerna, som hennes majestät med deras myckna intagande bortskämt hafva” (del 1: s. 47). Vad Ekeblad i själva verket pekar på är ”thesze många och okloke doctorna som hennes Maga med theras mykra intagande bortskämt haffua” (editionen s. 20: rad 23). Sjöberg har gjort oklok till klok och mage till majestät. De många detaljvikelserna tillkommer.

När Ekeblad på sitt lättsamma sätt talar om att göra en resa till ”*signor* mörner i Småland” (44:18), blir det hos Sjöberg en resa till ”Sigurd Mörner i Småland” (1: s. 104). Bekantingen hette Hans Georg Mörner af Morlanda.

I samma anda skriver Ekeblad i ett annat sammanhang att han skickar brevet med ”*Mons[ieu]r* taub och hans oxen” (163: 16). Här ansätter Sjöberg ”monsieur Taub och Hans Oxen” (1: s. 364). Oxen var i verkligheten Margareta Oxen, hustru till Otto Reinhold Taube af Karlö. För Hans Oxen liksom för Sigurd Mörner har det aldrig funnits någon värld utanför textutgåvan. De är inte ensamma om det ödet.

När jag nu går in på frågan om SAOB:s täckning av Ekeblads text, ger jag till belysning ett urval exempel av olika typer. Jag vill gärna framhålla, att det är principfrågan jag ser som intressant.

### **Inte belagda belägg**

Av en typ för sig är naturligtvis fall som har lästs eller tolkats felaktigt. Man talar ibland om spökord. I ett brev står det (82: 8):

eder lust med iagten ähr migh kär att förnimma, och ähr iagh nu nästan glad att iagh inthet ähr där tillstädes efter iagh inthet skulle orka fölia med. doch i karet orka iagh fulle sittia,

I ordbokens formparentes åberopas detta ställe som belägg på *karet* 'stor, helt sluten (gala)vagn dragen av två l. flera hästar' (K 565). Men Johan Ekeblad jagade inte på safarimaner utan satt stilla i sitt *kar*, dvs. sin skottkoja. Betydelsen finns hos Dalin (1: 812 a), och Ekeblads text har ytterligare två belägg (*kaar* 48: 19, *karet* 95: 2). I SAOB registreras inte denna betydelse under *kar*.

I inledningen till texteditionen har jag givit ett par andra exempel. Det ena är SAOB:s *masurvind* 'sned liksom fibrerna i masur', som återgår på Sjöbergs utgåvas läsart (1: s. 141): "Herr Späns ... den masurvinna i ögonen". Handskriften har emellertid: "h[err] spänsz ... den maszurrinna i ögonen" (62:3). Ordet betyder 'masurvuxen; rynkig'. SAOB tar upp det som biform till *masrunnen* (M 424). *Masurvind* är obelagt.

Det andra exemplet rör ordet *kanalje* 'skälm'. SAOB:s äldsta belägg anges vara från Ekeblad 1652 (K 331). Detta beror på att Sjöberg har läst *canalier* (1: s. 145) där det står *Caulier* (65: 6), som var namnet på Ekeblads hund. Det äldsta faktiska belägg som ordboken hänvisar till är från 1741.

Ett språkprov som anförs utan att återge handskriftens lydelse kan i sin helhet betraktas som ett inte belagt belägg. Följande citat ur ordboken har tillägget "rättat efter hskr.". Exemplet gäller uttrycket *vara klippt* 'vara motigt' (K 1255).

Jag begynner nu till att gå något ut, oansedt med möda ... Men i den staden är min pojke dödssjuk af hwardagssielffua, så att dätt skall ändoch vara klippt.

Detta återger följande text (101: 7):

iagh begijnner nu till att gå något wth oansedt med möda ... men i dän staden ähr min poike dödhsiuk aff hwardags sielffua, så att dätt skall ändoch wara klippt.

Ordboken följer i stort sett Sjöbergs utgåva (1: s. 230). Särskilt kan man lägga märke till att belägget på det aktuella uppslagsordet *klippa* inte har autentisk form.

## Uppslagsord saknas

En viktig grupp bildar sådana ord och uttryck som inte alls är upptagna i SAOB. De exempel jag skall ge ligger på olika ställen längs skalan mellan inhemskt och utländskt. De är också i olika grad integrerade i det svenska ordbildnings- och ordböjningssystemet.

Först kommer ett antal blandade exempel på osammansatta ord. Rak stil svarar mot textens huvudstil, den tyska, medan kursiv återger bistilen, den latinska.

*auantage* 92: 16 'fördel'  
*Consolation* 102: 12 'tröst'  
 danserij 86: 15 'dansande'  
*desplaisir* 152: 4 'missnöje'  
 durcheinander 112: 15 'om vartannat'  
 fisza 78: 11 'gnida'  
 gehirn 53: 7 'hjärna'  
 höffuelisera 84: 1 'hyfsa'  
 lupat 49: 19 'överlistat'  
*Miserablement* 91: 16 'eländigt'  
*portement* 8: 27 'hälsa'

Av dessa finns *fissa* hos Rietz (177 b) och *lupa* hos Dahlgren (516 a).

Av mera tillfälliga sammansättningar som är omedelbart förståeliga utifrån ledernas betydelse utesluts ett stort antal, påpekar Sven Ekbo i *Vägledning till Svenska Akademiens ordbok* (s. 7). Man konstaterar dock tacksamt, att mängden medtagna sammansättningar är stor. Ekebladiska fall som saknas i SAOB är bland annat följande:

brudgumewagn 118: 18 'brudgumsvagn'  
 bärgskripa 72: 15 'bergskreva'  
*Comis* bodh 125: 9 'beklädnadsaffär'  
 einsidlerståndet 94: 27 'ungkarlsståndet'.  
 Jortemördery 63: 21 'hjortdödande'

korggiffuande 33: 21 'avslag på frieri'  
 kärleksduellerna 147: 4 om fåglars spel  
*lamentationsskrift* 108: 18 'klagoskrift'  
 mordiagten 139: 16 'nyttjakten'  
 pestilensproceszen 123: 15 'epidemiförloppet'

Det kan nämnas att *bergskripa* och *skripa* finns hos Rietz (600 a) och att SAOB har sammansättningen *ensidlermunk* 'eremitmunk' från 1712 (E 571). Av särskild vikt bland de illustrerande orden är *mordiagten*, beteckning för den med fasta redskap bedrivna nyttjakten till skillnad från *lustiagten* (139: 16), den jakt som bedrevs för nöjes skull (L 1232).

När det gäller ord av främmande ursprung är det från språk-systemets synpunkt särskilt intressant med belägg som visar svenska avlednings- respektive böjningsändelser. Sådana finns det en hel del, exempelvis dessa:

*admirera* 75: 11 'beundra'  
*affligerar* 17: 15 'smärtar'  
 allamode 80: 7 ("den a. räffuen") 'moderne'  
*Contrecomendet* 6: 11 'kontraordern'  
*discontinuera* 127: 23 'upphöra'  
*entertienet* 83: 22 'samtalet'  
*esperancer* 158: 7 'förhoppningar'  
*estendera* migh på 24: 1 'utbreda mig om'  
*inCom[m]oderligt* 46: 9 'olägligt'

Man kan lägga märke till att ändelserna i en del av fallen skrivs med den vanliga tyska stilen, medan de franska ordstammarna skrivs med latinsk stil. Vidare gäller att ett par av de nämnda adjektiven har närstående former belagda från ungefär samma tid i SAOB: *alamodig* och *alamodisk* ([A] 895), *inkommod* (I 583).

## Varianter saknas

I åtskilliga fall återfinns visserligen den lexikaliska enheten i ordboken, men textens belägg representerar en variant som inte har registrerats. Exempelen visar några olika typer.

Fallen i den första gruppen har formell karaktär. De gäller i huvudsak stavning och böjning.

abgesandter 71: 14 'abgesant, envoyé'

apeteken 92: 8 'apoteken'

bär 90: 25 'ber'

*Comparaison* 147: 8 'komparation, jämförelse'

kwijter 34: 23 'kvitt, fredlig'

muszigh 104: 17 'mysig, stursk'

påste 65: 10 'post, postlägenhet'

Närmast följer några fall där belagda betydelsevarianter saknas i ordboken.

affall 33: 3 'förlust'

argwohn 100: 22 'misstroende'

bakdant 132: 21 som personbeteckning

*flatterar* 46: 21 'angenämt berör'

På den kombinatoriska sidan ger jag några exempel som rör konstruktionssätt och fraseologi.

*accorderar sigh* 25: 19 'passar ihop'

altutförett 139: 12 'ständigt'

*asseurera* 24: 13 ('a. min k[iäre] bror om min redebogna  
Tiänst') 'försäkra (ngn om ngt)'

igenom *reputation* 60: 9 'ryktesvis'

I det första fallet är det den reflexiva konstruktionen vid den iakttagna betydelsen som avses. I det andra fallet har SAOB varianterna *allt för ett* och *ut för ett* ([A] 1086).

## Yngre belägg

De ord och uttryck som vi fortsättningsvis skall betrakta är i motsats till de föregående beaktade i ordboken. Belägg hos Ekeblad ger emellertid en förskjutning av dateringen, inte sällan en avsevärd förskjutning.

Naturligtvis är det oftast så, att dateringen flyttas bakåt i tiden. Jag tar emellertid först upp tre exempel där beläggen hos Ekeblad är yngre än de yngsta beläggen i SAOB. I två av exemplen rör det sig om över hundra år. Detta är det maximala fallet med tanke på att ordboken tar vid på 1520-talet och att Ekeblads brev skrevs vid mitten av 1600-talet.

(Tuve Tuveesson Hökeflycht) behöver inte frukta att han skall dö *arfflös*, skriver Ekeblad 1652 (60: 11). Med detta menar han 'utan arvingar', det som i bibeln 1541 kallas *arffualöös* (citerat A 2194 *arfvalös*). Något yngre belägg än detta från bibeln redovisas inte i ordboken. Bortfallet av bindevokalen behandlas i en anmärkning i spalt A 2167.

Verbet *affkomma* 'komma undan, slippa undan' finns hos Ekeblad 1652 (48: 16). Det yngsta belägg som SAOB redovisar är från 1638 ([A] 282). Förbindelsen *komma (ut)av* är däremot belagd 1845 (K 1960).

I ett brev från 1654 nämner Ekeblad Christer Horn af Åminne och skriver: "han groszar eer" (137:8). Detta ger oss det yngsta belägget på *grotsa* 'hälsa'. SAOB:s sista belägg är från 1531. Formvarianten är inte heller representerad tidigare.

## Äldre belägg

Utan att ha gjort någon fullständig genomgång av materialet har jag noterat ett hundratal fall av äldre belägg. Antedateringen spänner mellan åtta månader och 271 år.

Inom 1600-talets ram ligger naturligtvis många fall. Den minsta förskjutningen gäller *diuertera sigh* 'förströ sig', som är belagd den 3 november 1652 (75: 21). SAOB:s äldsta belägg är

också från ett av Ekeblads brev till brodern, nämligen den 29 juni 1653 (D 1712). Detta är givetvis mest en kuriositet.

Andra exempel från århundradet:

*action* 82: 5 'förhandling' 1652, jfr 1660 [A] 873

obränd aska 50: 9 som beteckning för käpp eller liknande var-  
med någon agas 1652, jfr c. 1660 A 2474

*porterat* 150: 14 'hågad' 1654, jfr 1660 P 1505

rooligheet 94: 15 'förlustelse' 1653, jfr 1674 R 2388 (Colum-  
bus)

*Cornet* 4: 4, *Cårnet* 87: 2 'kornett, kavalleriofficer av lägsta gra-  
den' 1646 respektive 1652, jfr 1677 K 2450

griszar 4: 13 'latar mig' 1646, jfr 1679 G 946 (Rudbeck)

göra mina ... *recommendationer* 146: 8 'rekommendera mig (i  
ngns ynnest)' 1654, jfr 1689 R 907

*huius* 66: 3 'dennes, i denna månad' 1652, jfr 1699 H 1356  
(Carl XII)

Några ställen där det verkar troligt att Sjöbergs utgåva har lagt hinder i vägen kan också nämnas. I ett brev 1653 kommer Ekeblad in på kvinnliga prydnader och talar bland annat om "gravor och *fauor*", dvs. '(svart) slöja och sidenband(srosett)' (79: 3). Detta har hos Sjöberg blivit "gravor och fanor" (1: s. 189). Det äldsta av SAOB:s belägg på *favor* (*favör*) i den aktuella betydelsen är från 1661 (F 412). Ordbokens äldsta belägg på *gravor* är daterat 1671 (G 873).

På liknande sätt förhåller det sig med *däka* 'flicka'. Johan Ekeblad skriver den 5 oktober 1649 att han inte kan komma till ett förestående bröllop och ber brodern rekommendera honom "hoosz heela samquemmet ... och isynnerheet hosz dekorna" (6: 17). Sjöbergs utgåva har "dockorna" (1: s. 14). Formen *deka* finns visserligen i SAOB, men det äldsta belägget är från Rudbeck 1679 (D 2507).

Till fallen som rör 1700-talet hör följande:

efter alla *aparentier* 22: 14 'efter alla apparanser, efter allt att döma' 1650, jfr 1700 A 2015  
 förbriidd 112: 15 'som är i bryderi' 1653, jfr 1702 F 2434  
 bliffua *Engagerat* 80: 15 'erhålla befattning' 1652, jfr 1733 E 609 (Dalin)  
*ambassade* 154: 6 'diplomatiskt uppdrag' 1654, jfr 1734 A 1181  
*amour* 142: 17 'kärlek' 1654, jfr 1762 [A] 1226  
*quasi* 47: 16 'skenbart; nästan' 1652, jfr 1765 K 3374 (*kvasi*)  
 guardiregimentet 12: 21 'gardesregementet' 1649, jfr 1788 G 113  
 lustiagten 139: 16 'jakten för nöjes skull' 1654, jfr 1794 L 1232

Vi är så framme vid vad som är 1800-talsfall i SAOB.

granne 113: 15 'bordsgranne' 1653, jfr 1811 G 830  
*apropos* 174: 24 'läggligt' 1655, jfr 1815 A 2053  
*incommodera* meig 150: 22 'göra mig besvär' 1654, jfr 1845 I 583  
*alias* 60: 9 'annars kallad' 1652, jfr 1871 [A] 920  
*Cito* 3: 3 'angeläget (som påskrift på brev)' 1639, jfr 1871 C 218  
*audiensstol* 83: 21 1652, jfr 1881 A 2654

I fråga om *cito* kan vi dessutom lägga märke till att superlativformen *citissime* också är belagd 1639 hos Ekeblad (*Citisime* 3: 3), medan ordboken har 1729 som sitt äldsta år (C 218).

I ett fall slutligen når vi en bit in på 1900-talet. Det gäller uttrycket *reuanchera migh* 'återgälda (en tjänst)'. Ekeblad har det 1652 (63: 26), SAOB först 1923 (R 1624).

### Några kommentarer

Efter de exempel jag har givit antar jag att vi är eniga om att ordboken inte alltid går så djupt som vi skulle önska. Slutsatsen gäller givetvis i första hand den undersökta texten. I Johan Eke-



blads fall kan en del särskilda omständigheter ha bidragit till utfallet.

Handstilen får sägas vara svårforcerad. Tolkningen av texten är krävande. Detta hänger bland annat samman med Ekeblads raljanta stil och hans anspelningar på aktualiteter och litteraturställen. Stor roll spelar naturligtvis också den intrikata väv av personrelationer, politiska händelser och samhällsförhållanden som ligger bakom texten. Vi skall komma ihåg att det rör sig om privatbrev. Dessutom har Sjöbergs utgåva uppenbarligen verkat i hög grad störande.

Man kan emellertid också uppfatta resultatet av granskningen som en indikation på vad man a priori väntar, nämligen att texter normalt inte täcks fullständigt. Bilden av lexikaliska enheters historia kan med andra ord preciseras, likaså bilden av utnyttjade källors och studerade tidsepokers ordsfatt.

En precisering kräver framför allt att två villkor uppfylls. Det ena är en vidgad utgivning av opublicerade texter i pålitliga editioner. Det andra är ett effektivt tillvaratagande av det lexikaliska stoffet i utgivna (respektive från början tryckta) texter genom icke-subjektiv excerpering.

För 1600-talets vidkommande skulle redan en vidgad Ekebladfilologi ge stora bidrag till ordhistorien. Utöver de 108 breven från Johan Ekeblad till brodern har 640 brev till andra adressater, främst fadern, bevarats. Därtill kommer drygt 700 sidor annan text, varav ungefär hälften översättningar till svenska (editionen s. XXXIII). Jag tänker också på de andra personerna i kretsen, såsom Johans far Christoffer och inte minst systemen Margareta. Hennes brev är en intressant dokumentering i en tid som uppvisar få kvinnliga skribenter.

Den excerpering som ligger till grund för SAOB betecknas i själva verket av Bengt Loman som "en mera slumpmässig, extensiv excerpering" (Vägledning s. 84). I motsats härtill brukar en heltäckande genomgång kallas djupe excerpering. Loman tar upp möjligheten till datamaskinell djupe excerpering men avvisar den på vad som numera får anses vara inte fullt tillräckliga grunder.

När jag talar om icke-subjektiv excerpering har jag ett slags djupexcerpering i tankarna. Genom att utnyttja datalingvistisk metodik kan man kontrollera excerperingen i vissa avseenden. Bland annat kan den styras att inte gälla vanliga och välbelagda ord.

En tilltalande möjlighet vore att välja ut en serie centrala texter från de aktuella århundradena, särskilt de tidigare, för behandling. Resultatet kunde skrivas ut som belägg i kontext med källhänvisning. Lexikograferna skulle få ett fint material att arbeta vidare med. På detta sätt skulle man få större säkerhet mot ojämnhet i excerperingen. Ett slags ordhistorisk ryggrad skulle etableras.

Visserligen har SAOB hunnit till bokstaven S, men de återstående bokstäverna kräver flera decenniers arbete. Ett material som skulle tillkomma på det antydda sättet kunde användas för att komplettera underlaget för den återstående delen av ordboken och i övrigt tillföras samlingarna och hållas i beredskap för en framtida revision.

Sammanfattningsvis har vi iakttagit att SAOB i all sin rikedom har luckor i den lexikografiska täckningen som inte är försumbara. Tiden är kanske mogen att dra nytta av tillgängliga metoder för att komplettera excerptomaterialet.

## Litteratur

- Allén, S., Grafematisk analys som grundval för textedering med särskild hänsyn till Johan Ekeblads brev till brodern Claes Ekeblad 1639–1655. Göteborg 1965.
- Dahlgren, F.A., Glossarium öfver föråldrade eller ovanliga ord och talesätt. Lund 1914–16.
- Dalin, A.F., Ordbok öfver svenska språket. 1–2. Stockholm 1850–53.
- Ekbo, S., & Loman, B., Vägledning till Svenska Akademiens ordbok. 1965.
- Ekeblad, J.: Johan Ekeblads bref. Utg. af N. Sjöberg. 1–2. Stockholm 1911–15.

- Ekeblad, J.: Johan Ekeblads brev till brodern Claes Ekeblad 1639–1655. Utg. med inledning, kommentar och register av S. Allén. Göteborg 1965.
- Jacobson, G., [Rec. av Johan Ekeblads brev. 1. Utg. af N. Sjöberg. Stockholm 1911.] (Historisk tidskrift 32 (1912, tr. 1913), Öfversikter och granskningar s. 95 ff.)
- Jacobson, G., [Rec. av Johan Ekeblads brev. 2. Utg. af N. Sjöberg. Stockholm 1915.] (Historisk tidskrift 36 (1916, tr. 1917), Öfversikter och granskningar s. 64 ff.)
- Ordbok öfver svenska språket. Utg. af Svenska Akademien. 1 ff. Lund 1898 ff.
- Rietz, J.E., Svenskt dialekt-lexikon. 1–2. Malmö ... 1867.

## The Language Bank Concept

Data Bases in the Humanities and Social Sciences. 1980.

This paper discusses the theory and practice of a language bank in a framework for language decoding. The concept is seen in relation to man, computer, and computational linguistics.

### Framework

We have each of us at our disposal a good deal of knowledge concerning linguistic phenomena, in particular with regard to (1) the linguistic *systems* of our language, such as lexical items or conversational conventions; (2) the *use* of these systems, i.e. priorities in different situations, e.g. a private letter or a public address; (3) a number of actual specimens of linguistic *manifestations*, e.g. various sayings or poems or fragments of novels and perhaps textbooks. In other words, each of us is a living *language bank*.

Furthermore, when we are exposed to manifestations in our language we are generally able to process them, even to understand them. Successively, we are thus able to update our individual banks. We are of course also in a position to produce manifestations ourselves. All this means that each of us is a *language user*.

Analogously, the computerized language bank is, to my mind, part of a *framework for language decoding* (cf. Allén 1977). There is no point in just amassing sleeping stuff. This framework is the basic theoretical notion. Some important methodological consequences are the following. We have to develop methods for the input and storing of linguistic information in the language bank, as well as for the access to it. But we should also develop methods for making *the computer a language user*, as it were, with the capacity of updating the bank not only by adding new specimens of manifestation, but also by adding new data about

the linguistic systems and about usage. This brings us to the very heart of *computational linguistics*.

If we consider the notion from a sort of deep-case perspective, it is clear that man is the source, that man is also the agent, and that the computer is the instrument. The instrumental role of the computer is twofold: it is a book-keeping device and an analytical device.

The language bank, considered in a framework for language decoding, has a reciprocal relation to the model of the computer as language user. The development of computational methods for linguistic analysis increases the amount of data contained in the bank and, consequently, its usefulness. This, in turn, provides the possibility of improving the computer's capacity for analysis, etc. The language bank can in fact be seen as a reservoir of linguistic information to be used by the language-user system for the purposes of analysis and synthesis. This is one of the main reasons why the bank should be operated in close connection with a research department. It is also one of the reasons why I choose to call such a bank a *dynamic language bank*.

The reciprocal relation just suggested raises the question of how far it is feasible to analyse text by means of computer. Generally speaking, I think it is possible to an interesting degree. Just two comments here. Man-machine co-operation by means of interactive processing comes in very handy. Further, information on the degree of uncertainty inherent in data obtained through fully automatic procedures should be attached to the data resulting from such processing.

It is not my aim to embark on a survey of the interesting work going on in various places on both sides of the Atlantic as well as in some other places. The questions of principle are my main concern. As I go along, however, I shall give some glimpses of the work from which I have drawn most of my experience. This relates to the Swedish language bank called the *Logothèque*, which is the service organ of the Department of Computational Linguistics in the University of Gothenburg. I do not claim, however, that all good ideas have actually been realized there.

The Logothèque was established by the Swedish government in 1975 as a permanent service organ. Service on an informal basis had been provided since the sixties. The Logothèque is nationally responsible for the collecting, storing, processing, and providing of linguistic material in machine-readable form. Its emphasis is on the Swedish language, but other languages are not excluded. In fact, there is work going on related to the languages of our numerous immigrants. At present, the Swedish text material amounts to about 30 million words from newspapers, novels, legislative statutes, etc. It dates from 1965 and onwards but represents mainly the last few years. The lexical material comprises about 200 000 entries, including the vocabularies of the Word-List of the Swedish Academy and the Frequency Dictionary of Present-Day Swedish Based on Newspaper Material. The holdings of the Logothèque are accounted for in *Språkdata Research Report*, which is issued annually free of charge (address: Department of Computational Linguistics, N. Allégatan 6, S-413 01 Gothenburg).

When I talk about a language bank, I refer in the first place to a computerized bank of information concerning general-purpose language and such special-purpose material as has a bearing on general-purpose language. From my point of view, a terminological data base, for example, is a special case of the wider notion. Also, I refer primarily to a synchronic bank, but what I say is largely applicable to the diachronic or historical case, too. As time goes by, the continuous work of any language bank will make it possible to study linguistic development.

I do not refer to such various kinds of collections of linguistic information as paper archives of, say, dialects or historical lexical material. Nor do I refer to printed books like dictionaries. It would in many cases be advantageous to computerize these types of repositories in order to make them available through several profitable kinds of access. This is now being realized in some places.

## Bank structure

The basic structure of the language bank follows from the theoretical concept. There are thus two main branches: *authentic text* and *analysed material*, the latter including *systems data* and *priorities data*.

The *text* branch should contain written and spoken manifestations. As regards written text, the first choice is to collect already encoded material. This requires co-operation with the graphic industry and publishing houses. The bank has to develop a standard format for the storing of texts from different types of equipment. I am now talking about the internal work of the bank. As far as exchange is concerned, documentation is the key-word rather than standardization. The bank also has to tackle the copyright problems. They have in fact turned out not to be problems in the case of the Logothèque. Admittedly, the legal status of copyright on machine-readable material is not fully clear. The publishing houses and author associations concerned have investigated the issue, however, and consented to provide such material for scholarly purposes. If commercial purposes are involved, each case will be treated separately. At an early stage, we worked out agreement forms regulating the relation between the supplier and the Logothèque, on the one hand, and the Logothèque and the user, on the other.

If a text is not available in machine-readable form, a scholar may choose to encode it. There is thus also some co-operation with colleagues. In addition, the bank itself can of course have texts encoded. This is done only under special circumstances. One example is the new critical edition of August Strindberg's collected works under the supervision of the Swedish Cultural Council. It involves between six and seven million running words. The encoding of the texts and the processing of them for the benefit of the successive editors of the volumes will be performed at our Department. The texts as well as analysed material will of course be included in the bank.

There are very few spoken texts available in transcribed form. They are indeed welcome when they appear. The notational

system used should be documented in each case. Ideally, the user of the bank should have the possibility not only of getting the transcribed form but also of hearing the recorded form of the manifestation.

It might be added that the distinction between written and spoken text is not always clearcut. One recent example is connected with the inclusion in the Logothèque of the 1978/79 proceedings of the Swedish Parliament, involving some 4 million running words. The deliberations are encoded day by day and a preliminary version of the proceedings is printed night by night, to be presented the following morning. This material is in a way spoken, but part of it is read from manuscripts and thus ultimately written. Naturally, this particular property provides no reason for the exclusion of the material.

The second main branch contains, as I said, *analysed material*. It is important that the analytical models are specified. If not, priorities data run the risk of being useless. Linguistic entities like words on various descriptive levels, collocations, morphemes, grammatical constructions, etc. are obvious examples of interesting units. Semantic entities as well, like kernel senses, have been included in our case. The material referred to will soon be published as Volume 4 of the Frequency Dictionary mentioned earlier.

It is in the nature of things that one main source of analysed information is the set of texts stored in the bank itself. This is the working-field of the language-user system. Another main source is made up of already analysed material in machine-readable form, such as dictionaries on magnetic tape. They are obviously of great value for the important lexical section of the current branch.

Research projects can also contribute substantially. The most extensive project going on in our Department now is called a Swedish Lexical Data Base. The aim is to establish a large data base on an explicit linguistic basis taking advantage, to a great extent, of the evidence contained in authentic texts supplied by the Logothèque. This will promote our development of machine dictionaries. A printed dictionary of contemporary Swedish will



also be generated from the data base. In due course the material will be included in the language bank.

Just a few words about the *technical* aspects involved. We think it important to have our own equipment, which is designed to meet the demands of natural-language processing in a predominantly interactive environment. It is based on a Data General Eclipse, has a disc drive with a capacity of 200 million bytes, soon to be 400, versatile input and output facilities and fifteen video display terminals, i.e. roughly one for each researcher. We also use the computing centre of the university, however.

One problem is whether to use general-purpose data-base systems or specially designed systems. In our experience so far, the big commercial systems have turned out to be an expensive alternative. In any case, the data should be stored in a way that provides multiple access, making it easy to find one's way through the data base according to criteria of current interest.

### **Selection and representativity**

With regard to the *selection* of material, the bank should, at best, reflect the information contained in the body of individual human banks and the vast amount of manifestations. From another point of view, it can be regarded as an extension of the human mind. Anyhow, the problems of selection and representativity are brought to the fore.

First, let me remind you of the fact that there are other ways of establishing linguistic material than by means of text (cf. Allén 1978). Methods that belong here are experiment, inquiry, taking notes, and introspection. Information resulting from the application of such methods may also deserve a place in the bank. The most urgent problem, however, seems to be the selection of texts.

The text branch of the bank need not necessarily be as large as possible but rather as useful as possible. This does not mean that it should not be large. A lexical word may occur once in a billion running words, and the magnitude of the bank can, in fact, be in

the order of billions considering the technique at hand. The existence of such rare birds, however, does not make it reasonable to collect and store year by year, for example, the totality of texts printed in Swedish, which amounts to three billion words annually. Selection is the word.

It is necessary to distinguish a language bank from a project. The latter has a specific linguistic goal and a plan for arriving at it within a definite span of time. The aim of a language bank is much wider, and there is no time limit involved. If a project is successful, its result should be of general interest and may then be incorporated in the bank.

It is obvious that a collection of texts gives us indirect evidence of the relevant linguistic systems and that it is a sample of usage. Is it reasonable to consider such a sample representative of a language? Let us take a look at what this might mean. Some of the factors involved in linguistic materials are the following: mode of expression (speech or writing), stylistic varieties (such as those relating to genre, author, and topic), regional varieties, and social varieties. Things like the number of copies printed and the number of readers should also be kept in mind when specimens of writing are considered.

The establishment of the text branch of a language bank, representative in a quantitative sense of the population involving the above-mentioned factors, would be an immense task. In my view it would also be dubious. Who could imagine the kind of average language that would be involved? Whose language would it be? What general conclusions should we be entitled to draw? This rather sceptical view also applies to more limited corpora with a general-purpose aim. Notice that I am now talking in terms of statistical representativity.

On the other hand, the language bank should obviously be able to provide different types of text for different purposes. This is of course also highly desirable from, e.g., a lexical point of view. It requires, among other things, investigations to find out what are the relevant types of text and information in the bank concerning the text type, author, date, etc., of the specimens stored. It requires, in addition, many entries to the data bank if it

is to be useful in this respect. A bank organized along these lines will be able to provide projects of various kinds with tailored sub-corpora. This is a further reason for calling the bank dynamic.

### Access

Flexible *access* to the various kinds of information contained in the data bases of the bank is vital. Paper printout has not ceased to be of interest and never will, I think. People want to handle their data more or less physically in many cases, and rightly so, making notes, etc. However, one must be careful not to drown them, which is quite easily done.

Publication of results by means of computerized photo-composition and printing should also be mentioned. This applies when many immediately readable copies are desirable. Microfiche is often a convenient medium, particularly with regard to concordances on different linguistic levels, provided the underlying text material is not very limited. The Logothèque uses this medium when concordances of one or several million words are produced. And the users use them.

Direct access via terminals is another important technique. A prerequisite is that the problems of data security have been solved. The possibility of having language banks interconnected in a network enhances the importance of this type of access.

The *users* of the language bank are found in many quarters indeed, not only in linguistics, literature, social sciences, and other academic disciplines, but also in, e.g., information retrieval, the graphic industry, and journalism. Feedback from users is of course very helpful. It is part of the dynamics of the language bank.

If you look up the word *bank* in the *Concise Oxford Dictionary*, you will find that it means an "establishment for custody of money, which it pays out on customer's order". If the phrase "linguistic valuables" is inserted for "money", we may get a first approximation of what is involved. On closer inspection, how-

ever, it seems that there is more to a language bank than just custody and output.

### **References**

- Allén, S., Text-Based Lexicography and Algorithmic Text Analysis, *ALLC Bulletin* 5: 2 (1977) 126–131.
- Allén, S., The Role and Organization of a Corpus in the Investigation of a Language, *ALLC Bulletin* 6: 1 (1978) 10–14.

## **Carl Ivar Ståhle. Inträdestal vid Svenska Akademiens högtidssammankomst den 20 december 1980**

Svenska Akademiens handlingar 88/1980. 1981.

Carl Ivar Ståhle var fascinerad av språket, språket som kommunikationssystem, som historisk dokumentation och som uttrycksmedel. Den som en gång har dragits in i kraftspelet mellan innehåll, form och funktion kan som han fångas för alltid. Det blir som det på ett annat plan har sagts om våra mycket små: "Once you stand up and start to walk, you're committed for life!"

Han blev följdriktigt språkforskare, en av våra verkligt stora. Ett ovanligt djup i forskningen förenades hos honom med en ovanlig bredd. Han hade också ett livligt intresse för människorna bakom språkyttringarna och deras kultur och samhälle. Texternas litterära kvalitet upptog honom i hög grad, och han var själv en fin stilist. Allt detta gör hans skrifter levande och entusiasterande. Vi som hade förmånen att känna honom vet, att detta också gällde hans person.

Själva grunddraget i hans personlighet vill jag beteckna som intensitet. Det var ytterst den som satte sin prägel på hans forskning, som gjorde hans insats som akademisk lärare legendarisk och som gjorde den personliga samvaron med honom så charmfull.

En förträfflig inkörsport till Ståhles skrifter och därmed till deras författare, hans hållning och framtoning erbjuder en uppsats som heter "De främmande orden och den svenska ordbildningen". Titeln verkar kanske inte direkt upphetsande, men studien är skriven med verklig frenesi och visar åtskilligt av hans språksyn och stilkonst.

Utgångspunkten är att vi inte kan avvisa allt som vi har lånat in i språket. I stället kan vi konstatera att dagens svenskar grovt

taget har två ordbildningssystem i sitt språkmedvetande: ett inhemskt-germanskt och ett romanskt-internationellt. Mot *avfolka* (och för Grönköpings del *avhundas*) står *decentralisera* osv. Vi får också blandformer som *avbyråkratisering* eller rentav följande, som Ståhle har hittat i ett skådespel, där en kavaljer riktar en uppmaning till sin hetsige rival med orden: "Kallsinnera er, Herr styckjunkare!"

I en fin passus i uppsatsen analyserar Ståhle hur de två systemen utnyttjas av en stor språkkonstnär. Hör här: "I Midsommardalen låter Harry Martinson det svenska språket sjuda och glittra i kapp med högsommaren själv: han frigör språkljudens förmåga att suggerera öga och öra, och så länge den svenska sommaren kryllar och lyser firar hans språkkonst triumfer i fråga om ordbildning med ärvda, inhemska medel, ljudhärmingar, sammansättningar och avledningar: luften är fylld av knirrande, svirrande och risslande liv, grantopparna utför en sassande dans under stormmolnen, dygrunden pluddrar, sjön eldgлитtrar i sommarsolen ... Men i skymningstimmarna, och när det lider mot hösten, och iakttagelsen och upplevelsen får lämna plats för meditationen, då träder det andra ordbildningssystemet till: 'skymningen blir sällan något enbart faktiskt och jordrotatoriskt', sagoberättarna är 'att betrakta som folkets kompensatorer' ..."

"Språkvetenskapen kan inventera våra tillgångar", skriver Ståhle sammanfattningsvis, "teckna de mönster som är levande i språket och visa på vårt språks rikedom och möjligheter, språkkonstnärerna kan utnyttja det bästa och få det i bruk."

I det här sammanhanget vill jag också gärna nämna en studie över soluppgången som poetiskt tema. Efter några inledande observationer, där Ståhle bland annat noterar att det första stråk av lyrik som klingar fram ur en svenskspråkig behandling av motivet är ett ställe hos Birgitta, koncentrerar han sig på tre verk: Creutz' *Atis och Camilla* från 1761, Strindbergs *Stadsresan* från 1905 och Harry Martinsons *Sommarmorgon* från 1938.

När den ganska utförliga genomgången går mot sitt slut, säger han någonting som jag inte har sett i någon annan vetenskaplig skrift: "Någon sammanfattning kan det inte bli fråga om – det finns ingenting att sammanfatta. Jag har inte avsett några resultat,

bara försökt lyssna till hur tre av det svenska litteraturspråkets främsta förnyare ... har valt att i ord gestalta ett tidlöst skeende, som inte brukar lämna betraktaren oberörd ..." Denna unika avrundning säger något väsentligt om Ståhles sätt att förhålla sig till konstnärliga texter – och för övrigt inte bara till dem.

Carl Ivar Ståhle var född 1913 i den månad som också satte punkt 1980: juni, midsommarmånad. Födelseorten blev Salshult, Stenberga, Jönköpings län, men familjen var bosatt i Stockholm. Hans tre syskon, alla äldre, föddes i Stockholmsområdet. Salshult blev med tiden genom släktens försorg en mycket uppskattad tillflyktsort för honom och hans familj.

Modern Greta Maria Stålhammar var också född i Småland. Hon var dotter till översten vid Kalmar regemente Carl Adam Stålhammar och friherrinnan Naëmi Elisabet Lilliecreutz. Det berättas att hon hade starka litterära intressen och var varmt religiös.

Fadern Lars Valdemar Ståhle var ingenjör. Han ägnade sig åt hjulburen kommunikation och blev så småningom järnvägsinspektör. Han hade också tydliga konstnärliga anlag, något som bevarade målningar vittnar om.

Farfadern Lars Gustav Ståhle var likaså tekniker. En period var han chef för den tidens största pappersbruk, Nykvarn. Till de mest utmärkande dragen hos denne stålman hörde enligt skrifterna en ovanlig arbetsamhet och ihärdighet och en ungdomlig rörlighet och vigör till både kropp och själ ända upp i hög ålder. Till honom skrev Elias Sehlstedt en visa kallad "Skål för Lasse Ståhle i gröngräset 1852". Hans far i sin tur och hans farfar var båda handskmakarmästare i Norrköping. Med dem är vi nere i 1700-talet.

Åter till vårt århundrade. Carl Ivar Ståhle tog studenten 1931 på Stockholms norra latinläroverk. En av lärarna där var Erik Wellander, som redan då arbetade på vad som skulle bli språkvårdshandboken "Riktig svenska", utgiven 1939.

Ståhle blev filosofie magister 1934. Redan hans ämbetsexamen speglar vidden av hans intressen. I den ingick nordiska språk, litteraturhistoria, latin, grekiska och klassisk fornkunskap. Efter

licentiatexamen 1940 disputerade han i nordiska språk 1946. Han blev docent samma år. Under docentåren deltog han bland annat i det arbete som bedrevs av Vitterhetsakademiens diplomatariekommitté.

Av ett enhälligt lärarråd kallades han 1955 till professor i nordiska språk, företrädesvis svenska, vid dåvarande Stockholms högskola. Åt arbetet där ägnade han sig med liv och lust. Hur inspirerande hans undervisning och handledning var har ofta betygats. Kännetecknande för forskarhandledningen var hans förmåga att gå till botten med frågorna och hans respekt för dem som arbetade seriöst med sina forskningsproblem. Han var kritisk på det konstruktiva sättet.

Också som gästföreläsare var han mycket uppskattad. Jag minns hur han drog västerut till vårt seminarium och initierade högintressanta diskussioner. Ibland avbröt han sig och sträckte med en karakteristisk gest på nacken. Man tyckte sig märka hur han då rannsokade sitt formidabla minne och betänkte det senast sagda. Nya, spännande idéer var på väg att verbaliseras.

Han lämnade professuren av hälsoskäl 1971. Då tillträdde han en forskartjänst som Svenska Akademien inrättade för honom och som han uppehöll livet ut.

År 1974 valdes han till ledamot av Svenska Akademien. Samma år tilldelades han Vitterhetsakademiens stora pris.

Han var medlem av 1963 års bibelkommitté, ordförande i Svenska språknämnden (respektive dess föregångare Nämnden för svensk språkvård) från 1967 och ordförande i Svenska Vitterhetssamfundet från 1969, många andra uppdrag att förtiga.

I all denna intellektualitet var han förstas också en människa bland människor och ett personnummer bland personnummer. Gemenskapen i familjen var viktig för honom. Där hade han också en avgörande stödjepunkt. I vänkretsen var han mer än välsedd allt ifrån ungdomens studentskivor till senare års sammankomster i Sällskapet Februari-Gubbarne. Han hade en avgjort positiv inställning till stuvad lake. Han kunde hetta till, särskilt om han stötte på intellektuell ohederlighet. När han koppelade av sökte han sig gärna till naturen. Smultronställena fanns



på Lidingö, i Småland och i Dalarna. Han var också sportintresserad och ägnade sig i yngre år åt skidåkning och fäktning.

Ståhles första tunga vetenskapliga insats var hans ortnamnsstudier. Det centrala verket är doktorsavhandlingen från 1946, "Studier över de svenska ortnamnen på *-inge* på grundval av undersökningar i Stockholms län". Vad det gäller är namn av typen *Huddinge*.

Bakom ortnamnen på *-inge* ser Ståhle inbyggarnamn som i sin tur normalt är bildade till naturbetecknande ord. *Huddinge* hette på fornsvenska *Uddunge*. Ändelserna *-ing* och *-ung* var avljudsformer som kunde växla med varandra, och *h-*förslaget har tillkommit senare genom utveckling i dialekten. Namnet betyder uddingarnas, eller om man så vill uddbornas, område eller samlade menighet. Och uddingarna var de som hade sin mötesplats vid udden, nämligen Vårby-udden.

Avhandlingen innehöll flera nyheter. Viktigast var det helhetsgrepp som Ståhle tog på tolkningen av ortnamnstypen. Hans metodologiska säkerhet hade sin grund där. Han samlade också ett mycket omfattande bebyggelsehistoriskt och arkeologiskt material med vars hjälp han kunde underbygga sina tolkningar. Dessutom gjorde han noggranna undersökningar av naturförhållandena. Avhandlingen mottogs med stort gillande, och resultatet har i allt väsentligt stått sig.

I förordet tackar han på hävdvunnet sätt sin akademiske lärare Elias Wessén och andra som har stött honom i hans arbete. Men där finns också en detalj som förtjänar att lyftas fram, därför att den belyser arbetsvillkoren. Han riktar ett särskilt tack till direktören för Stockholm-Roslagens-Järnvägar för fribiljett sommaren 1944.

I en uppsats om de svaga maskulinerna, typen *gubbe*, kommer Ståhle in på de normativa 1700-talsgrammatikerna och deras betydelse för utvecklingen. Kan de ha så stor makt över ett levande kulturspråk? frågar han. Har andra krafter verkat, och i så fall vilka? Vad har egentligen hänt?

Frågan om vad som egentligen har hänt – ”wie es eigentlich gewesen”, för att använda von Rankes klassiska formulering – ställde han ofta. Och förbluffande ofta lyckades han klarlägga skeendet.

Gång efter annan underströk han hur viktigt det är att studera språkets utveckling inom varje genre för sig. Särskilt tre genrer har han själv djuplodat: lagspråket, bibelspråket och versens språk.

Lagspråket kom han att intressera sig för tidigt. Hans huvudarbete är från 1958 och har titeln ”Syntaktiska och stilistiska studier i fornnordiskt lagspråk”. I det koncentrerar han sig på kärnan i lagens språk, nämligen hur rättsregeln utformas.

Det språkliga uttrycket har två delar: förutsättningen för att lagbudet skall gälla och lagbudet självt. Ett ställe i balken om sårsmål i Äldre Västgötalagen, vår äldsta bevarade bok på svenska, kan låta så här i moderniserad form: ”Hugger någon handen av en man, böte nio marker för sår och tre marker för lyte.”

Stähle fokuserar särskilt förutsättningens form. Det finns flera andra uttrycksmöjligheter, t. ex. ”Någon hugger handen av en man” eller ”Om någon hugger handen av en man”. Han ser de olika formerna i ett internationellt perspektiv som omfattar romersk, kyrklig, kontinental, engelsk och västnordisk rätt.

Det visar sig att det naturliga uttrycket för förutsättningen på nordisk botten var den rättframma berättande formen. Som ursprungligen narrativ räknar han med fog också typen ”Hugger någon handen av en man”. Den var särskilt utmärkande för just fornsvenskt lagspråk och tillhörde där framför allt de äldsta delarna av straffrätten. Formuleringen stod den folkliga berättelsen nära, var ofta rytmiskt anlagd och är lätt att minnas.

I flera uppsatser behandlar Stähle olika problem som rör de gamla lagarnas handskrifter och andra textfilologiska frågor. I en uppsats som är ett mästerstycke av kritisk analys och kringsyn vederlägger han ett försök till rekonstruktion av ett fornsvenskt lagmanskväde, som skulle ligga bakom den så kallade västgötska lagmanskrönikan. Han har också skrivit ett betydelsefullt arbete om Östgötalagen, som kommer att tryckas. Dessutom klarlägger han de första tryckta lagarnas tillkomsthistoria. Märkligt nog

trycktes Upplandslagen och Östgötalagen före landslagen. Båda utkom 1607.

Studierna av bibelspråket ingår i Statens offentliga utredningar 1968. Det året lämnade bibelkommittén ett betänkande som gällde nyöversättning av Nya testamentet. Ståhle bidrog med en översikt över bibelspråkets historia från 1526 till 1917 och dessutom med en presentation av sådana drag i svenskans förändring under 1900-talet som var av betydelse för nyöversättningen.

Reformatoreernas utgångsläge i början av 1500-talet var inte lätt. Det fanns ännu inget riksgiltigt svenskt språk, vare sig i tal eller i skrift. Det fanns i många fall heller inga ord som svarade mot dem som förekom i den latinska eller grekiska förlagan.

Huvudpunkterna i den utveckling som Ståhle följer utgörs av Nya testamentet 1526 med dess något omständliga talspråkighet, Gustav Vasas Bibel 1541 med dess stramare, Lutherinspireerade stil, normalupplagan 1883 med dess lite glanslösa originaltrohet och Gustaf V:s bibel 1917 med dess försök att förena verklig bibelstil med naturligt talspråk.

Bibeln 1917 var slutresultatet av den kommissions arbete som tillsattes 1773. Efter 144 år hade man ändå inte nått fram till en version som väckte allmän anklang. Som Ståhle nämnde i sitt inträdestal här var hans företrädare H. S. Nyberg en av de hårdaste kritikerna. Han fann översättningen pratsam och platt.

Den avgörande frågan för Ståhle blir den som restes i direktiven till bibelkommittén, nämligen om det är möjligt att förutse efter vilka linjer språkets fortsatta utveckling kommer att ske. I så fall skulle man ju kunna anpassa en nyöversättning efter prognosen.

Ståhles svar är klart och som jag ser det evident. Vi vet att förändringar kommer att inträffa, men deras arter och riktningar kan vi inte förutse, bland annat på grund av utomspråkliga faktors inverkan. Vi kan aldrig garantera att bibeln får en hållbar språkdräkt.

Ståhles arbete om versens språk, som blev hans sista prestation i det stora formatet, heter "Vers och språk i Vasatidens och stor-

maktstidens svenska diktning” och kom ut 1975. Här behandlar han förändringarna i versens villkor under den omvälvande tiden från 1500-talets början till omkring 1700.

Stähle kan naturligtvis utnyttja tidigare forskning. Bland annat hänvisar han till sin företrädare på stol 3, Henrik Schück. Han har emellertid själv gett sig i kast med hela det stora primärmaterialet, tryckt och otryckt. Ur detta myller letar han fram de stora linjerna, samtidigt som han gör rader av enskilda iakttagelser.

Undersökningen var från början tänkt att ingå som ett led i en framställning av svenska språkets historia från 1500-talets början till 1700-talets slut. Det var detta arbete han ägnade sig åt på den forskartjänst som Svenska Akademien inrättade åt honom. Han fann snart att framställningen måste komma att gälla det svenska skriftspråkets historia – skrivna och tryckta texter är ju vad som finns kvar. När versstudierna växte fram, var det bland annat med tanke på att versen var ett viktigt hjälpmedel för att minnas språkliga formuleringar i en tid då läskunnigheten var förbehållen ett fåtal. Utforskning av principerna för versens byggnad ger också inblickar i diktarnas material, språket.

Uppgiftens komplexitet är stor inte minst därför att 1600-talet så att säga var den svenska poesins pubertetsålder. Vid mitten av århundradet skrevs vers av fyra helt olika slag.

Den medeltida knitteln med sitt enkla krav på fyra höjningar per versrad levde kvar. Västgötaöversten Christoffer Ekeblad skriver (skanderingen markeras här med apostrofer):

En frå'ga till di'g kiäre lä'sare go'dh,  
Ja'gh är ko'mmen aff kiö'tt och blo'dh ...

Den bara några decennier gamla stavelseräkande versen, som visserligen också den började bli omodern, var den andra typen. Därtill kom den efter antikt mönster kvantiterande versen. Dessa båda typer ger ofta ett skralt resultat när man läser högt. De är på sin höjd en vers för ögat.

Det som lade grunden till de följande århundradenas diktning var framväxten av den accentuerande versen, den naturliga diktionens poesi. Lyssna till alexandrinen hos Stiernhielm:

Een liu'f Voca'l-Musi'c, Basu'ner, o'ch Corne'tter,  
The gå' i fu'llan swa'ng bland Pu'kor o'ch Trompe'tter.

Och hör på hans hexameter:

Bo'rt med pa'pper, och ble'ck; bort bö'ker C'irklar, och på'nnor.

Ett särskilt avsnitt ägnar Ståhle just åt den som nog var hans älsklingskald, Georg Stiernhielm. Till honom kom han tillbaka i många sammanhang, och han lade ner ett mycket stort arbete på att tränga in i hans språk- och tankevärld.

En grundtanke hos Stiernhielm är att svenskan eller "götiskan" återgår på Adams språk, urspråket, och därmed är givet av Gud. Ordens form svarar därför mot tingens sanna väsen. Ordens rötter ger kunskap om universum.

När Stiernhielm hämtar upp ord ur fornspråket och slipar ljudklangerna, har det alltså en dubbel motivering: estetisk och filosofisk. Ståhle urskiljer på detta sätt nya, intressanta övertoner i "Hercules".

Med alldeles särskild inlevelse har han närmat sig Stiernhielms dikt "Emblema Authoris", den första kända sonetten på svenska. Den handlar om silkesmasken, som genom anspråkslöst men idogt arbete förvandlar blad till silke. Dikten blir till en symbol för Stiernhielms humanistiska livshållning. Man anar en själsfrändskap mellan 1600-talsskalden och 1900-talsforskaren.

Undersökningar som Ståhles måste få konsekvenser på flera plan. Viktig är inte minst deras återverkan på periodindelningen i svenskans historia. Ett vidgat studium av utvecklingen inom olika genrer och av relationen till kulturella och politiska skeenden i Ståhles anda skulle troligen komma att leda till en omprövning av de nuvarande gränserna.

Filologen Ståhle gjorde naturligt nog också texteditioner. Tillsammans med forskarkolleger gav han ut "Rannsakingar efter

antikviteter". Dessa texter är resultatet av en inventering av rikets fornminnen, som genomfördes på 1600-talet med riksantikvarien Johan Hadorph som initiativtagare. Givetvis, frestas man att säga, har Ståhle också gjort en läcker utgåva av Stiernhielms "Hercales".

Han gav ut del 1 i serien "Sveriges litteratur". Boken är en antologi som innehåller de viktigaste verken under de första 700 åren av den svenska litteraturens historia, från Rökstenen på 800-talet till slutet av 1500-talet. Dessförinnan hade han skrivit den förnämliga översikt av medeltidens profana litteratur som ingår i första delen av "Ny illustrerad svensk litteraturhistoria".

För den som arbetar mycket med texter ligger det nära till hands att intressera sig för författarfrågor. Ståhle har mycket riktigt lämnat flera bidrag på textattributionens område. En uppsats ägnar han åt ett stycke rim som har tillskrivits den inte alltför talangfulle 1400-talsbiskopen Henrik Tidemansson. Med en karakteristisk formulering säger Ståhle, att uppsatsen endast så-tillvida avser att öka Linköpingsbiskopens litterära berömmelse, som den vill söka skilja honom från författarskapet till ett av de obetydligare av de rimverk, som går under hans namn. Uppsåtet torde ha lyckats.

I boken om versen och dess språk tar han upp två intressanta författarfrågor. En av dem gäller vem som skrev texten till de båda baletter som dansades vid drottning Kristinas hov 1643 och 1645. Det är i de texterna som svenska språket debuterar som litteraturspråk med europeiska ambitioner, som Ståhle uttrycker det. Starka skäl dras fram för att tillskriva den skald som kallar sig Skogekär Bergbo dikterna. Härigenom har, som han påpekar, en anonym identifierats med en pseudonym. Nästa steg blir att ställa frågan vem Skogekär Bergbo var. Ståhle ansluter sig till ståndpunkten att det rör sig om Gustaf Rosenhane och ger ytterligare skäl för den.

Ett klassiskt problem är om den mest kända av 1600-talets bröllopsdikter, "Bröllopsbesvärs Ihugkommelse", är skriven av Stiernhielm eller inte. Olika forskare har här kommit till olika resultat. Ståhle tar sig an frågan genom att i första hand analysera versmått, syntax, ordval och ordbyggnad. Efter att ha provat

argumenten för och emot kommer han fram till att de inte ger något pålitligt utslag. Man får vara tacksam, menar han med en elegant vändning, att tiden frambragte en bröllopsdikt som står så nära Stiernhielms diktning att det är svårt att se om den hör dit eller inte.

Stähle anförtroddes också det offentliga uppdraget att skriva ett yttrande över den uppmärksammade språkutredningen 1960 rörande de anonyma breven vid biskopsvalet i Strängnäs 1952. Granskningen skulle gälla hållbarheten från språklig synpunkt av den tillämpade metoden. Han gjorde en grundlig genomgång och förde en ingående diskussion. Slutsatsen blev, att metoden är hållbar och genomförd med skärpa och omsorg och att den framförda kritiken inte är berättigad.

Något som låg Stähle varmt om hjärtat var Stockholmsspråket. Han påpekar att det är en av våra minst kända dialekter trots den betydelse det har och har haft för vårt riksgiltiga språk. När han citerar forskare som från olika utgångspunkter har riktat kritik mot vissa drag i Stockholmsspråket, försummar han inte att presentera dem som "den göteborgsfödde Johan Erik Rydqvist" och "den göteborgsfödde Lundaprofessorn Fredrik Wulff".

I slutet av 1950-talet startade han och hans medarbetare på institutionen i Stockholm ett projekt för att ta reda på mer om stockholmskan. Det pågick till in på 70-talet. Med bland annat detta projekts materialsamlingar som underlag skrev han en rad uppsatser om Stockholmsspråket idag och igår. De kommer att sammanställas till en bok.

Uppsatserna ger en god inblick i de karakteristiska dragen hos en språkform, där *slagit* heter *slaji*, *legat* heter *låga* och *tryckte* heter *tröck* och där *nej* kan heta *sällan* och *ja* kan heta *var lugn för svenska folket*. Han utreder på sitt vanliga omsorgsfulla sätt hur ändelserna *-an* och *-is* har vuxit fram och hur de används: *Kungsan* för *Kungsträdgården*, *Ackis* för *Konstakademien* osv.

I uppsatsen "Mötet uppnas på söndag" tar han upp en företeelse som inte bara är stockholmsk. Han ser den från språkvårdssynpunkt och är därmed inne på ett annat av sina intresse-

områden. Det uttal det gäller kan inte ställas upp som mönster, menar han, och diskuterar varför.

Inte för att det är fult. Att säga att man har ”glumt sin mussa” är inte i och för sig fulare än att tala om ”Guds under”. Däremot är det viktigt att kunna skilja på namn som Lundberg och Lönnberg och på räkneord som 45 och 75. Ett starkt skäl är också att avståndet mellan tal och skrift skulle öka, om uttalet godtogs. Det resonemang som förs här ligger i linje med hans grundinställning i språkvårdsfrågor: målet är störningsfri kommunikation.

Språkvetenskapen har genomgått en intressant och viktig utveckling under senare decennier. Nya strömningar har gjort sig gällande, och olika uppfattningar har brutits mot varandra. Det har stundom hänt att mina tankar har gått till en fabel som jag en gång stötte på utomlands. Den handlar om den på jorden ständigt missförstådde förtroendemannen, som klagade sin nöd för Vår Herre. Den Högste förbarmade sig över honom och lät honom återvända till jordelivet med förmågan att gå på vattnet. När han så kom gående över sjön på väg mot platsen för sina tidigare bekymmer, stod folket på stranden och ropade: ”Titta, han kan inte simma!”

Stähle såg inte språkvetenskapens sak så. Han deltog med aktivt intresse i nyutvecklingar, samtidigt som han slog vakt om uppnådda resultat och färdigheter. Också härigenom kom han att spela en viktig roll.

Han var en av dem som 1963 kallade till det första symposiet med den av honom myntade rubriken ”Sammankomst för att dryfta frågor rörande svenskans beskrivning”. Avsikten var att skapa ett forum för diskussion av nya, strukturella metoder inom språkforskningen och deras tillämpning på modern svenska. Initiativet slog så väl ut, att sammankomsterna har utvecklats till det största evenemanget på sitt område. Också finlandssvenskarna deltar mycket aktivt.

Han insåg tidigt att datamaskinen – symbolbehandlaren – har en unik potential som lingvistiskt redskap. I synnerhet intresserade han sig för sådana tillämpningar som gällde teoretiskt grundade bearbetningar av stora språkmaterial. Hans klarsyn kom att



betyda mycket för utvecklingen också därigenom att han anlitas som språkexpert av forskningsfonder.

Det ligger i sakens natur att Ståhle gjorde sina största insatser på det textfilologiska-språkhistoriska området. Det har framgått att han också arbetade framgångsrikt på det strukturlingvistiska. Dessutom främjade han sålunda framväxten av den processlingvistiska inriktning som nu gör sig alltmer märkbar och som bland annat karakteriseras av kognitiva helhetsbegrepp och datamaskinell metodik.

Bilden av Carl Ivar Ståhle har här till största delen fått framträda genom hans skrifter. De rymmer så mycket av hans väsen att det har legat nära till hands att låta dem komma till tals. Essensen i hans verk är kanske till sist denna: han gjorde språkets konst till sin vetenskaps mål och sin vetenskap till konst.

## Språkvetenskaplig databehandling

Presentationer av ämnesområden utgivna i samband med Göteborgs universitets professorsinstallation lördagen den 20 oktober 1979.

Frågorna från olika håll blir många, efterhand som ett nytt ämne växer fram. De kan ha sin utgångspunkt i att det hela verkar obegripligt eller fascinerande eller överflödigt eller nyttigt eller något annat. I allmänhet bottnar de i ett intresse att få veta mer. Undringarna tar här gestalt i en kursiv samtalspartner. Det utspinner sig ett belärande samtal eller som det en gång hette en dialogus.

### *Hur började det?*

I vårt fall har ämnet spirat ur forskning i nordiska språk med intresse för frågor av allmänt språkvetenskaplig natur. Konkret gällde det i början att utveckla metodik för att snabbt få fram information om vilket element som helst i en text, t ex beläggen på ett ord eller en bokstav. Behovet kunde tillgodoses av en datamaskinell konkordans. Konkordansen till Johan Ekeblads brev togs fram 1961–62.

Sedan dess har en rad mer raffinerade konkordanstekniker skapats. Konkordanser av växlande slag är fortfarande det mest efterfrågade forskningshjälpmedlet både utom och inom institutionen. Här är några rader ur en konkordans över ordledet *log* med de fulla orden som kontext. Markeringarna till höger har att göra med klassifikation och beläggställe.

bi o	log	isk	AV -T	7306
bok kata	log		NN -EN	7944
chef s id eo	log		NN -EN	10224
Chicago soci o	log	i	NN -N	10262
dia	log		NN -EN	11807
dia	log	is er a nde	AV =	11811
dia	log	scen	NN -EN	11815
djup psyk o	log	i	NN -N	12211
djur fys io	log	i	NN -N	12231
eko	log		NN -EN	13592
ent myt o	log	is er a	VB -AD	14182
epi	log		NN -EN	14245

Det finns också andra utgångspunkter för ämnet. På 1950-talet hade datamaskinen bland annat använts på olika håll för försök med språköversättning och informationssökning och för beräkningar i samband med textattribution eller författarbestämning.

*Frekvensundersökningar nämns ofta. Vilken roll spelar de?*

De är inte så dominerande som många tror. Jag menar då frekvensundersökningar som går utöver den rena bokföringen av teckensträngar mellan blanka typrum i en text, i den här meningen alltså "Jag", "menar", "då" osv. Mer avancerade frekvensundersökningar kräver utveckling av väl-specificerade beskrivningsmodeller, något som leder rakt in i centrala lingvistiska problem. Märk väl att frekvensstudier inte alls nödvändigtvis gäller ord – som redan det är ett svårfångat begrepp – utan också ordförbindelser, ordled, grundbetydelser eller kärnbetydelser, syntaktiska konstruktioner och åtskilligt annat.

Det är fullt klart att en frekvensundersökning av modernt svenskt material, bedriven på det antydda sättet, har varit av grundläggande betydelse för ämnets framväxt här. De teoretiska och metodologiska övervägandena och tillämpningen av de utarbetade modellerna har kommit konturerna till ämnesområdet att framträda. Naturligtvis har internationella kontakter genom konferenser, arrangemang av forskarkurser och arbetsseminarier, samarbetsprojekt, publikationer och annat också givit betydelsefulla impulser.

Det projekt som har resulterat i Nusvensk frekvensordbok baserad på tidningstext skisserades 1964. Materialet samlades in 1965. Det bestod av färdigstansade håltremsor från tidningssätterier. Bearbetningarna började 1966 och utfördes inom den då bildade forskningsgruppen för modern svenska. Det är alltså i hög grad fråga om ett lagarbete. Den fjärde och sista ordboksdelen, som just gäller ordled och kärnbetydelser, produktionskörs hösten 1979. Formellt upplöstes forskningsgruppen i samband med att institutionen bildades 1977.

### *Hur kan man då karakterisera ämnet?*

Språkvetenskaplig databehandling är som namnet anger ett språkvetenskapligt ämne. Det allmänna målet för ett sådant är att vinna insikt i naturligt språk och det vetenskapliga studiet av naturligt språk. Inom vårt speciella område är det grundläggande att betrakta språket som en process. Allmänt sett är processer mer komplicerade än strukturer genom att man måste ta hänsyn till ordningsföljd. Exempel på frågor som aktualiseras är hur man skall finna en satsdels gränser; hur man skall komma åt en diskontinuerlig förbindelses delar – ett exempel är på ett på förhand givet sätt; hur man skall identifiera och sammanföra böjningsformer och variantformer; hur man skall känna igen en lexikalisk huvudbetydelse – kursiv adjektiv 1 lutande (om stil) 2 mindre noggrann, oförberedd (om läsning och översättning) 3 fortlöpande, obegränsat pågående (om verbhandling); i vilken ordning man skall generera och presentera de olika konstituenterna i en struktur osv.

Denna dynamiska inriktning kräver lexikaliska och grammatiska komponenter som är långt mer detaljerade, täckande och välstrukturerade än vad som nu står till buds. De måste med ett ord vara explicita. Processynen leder enligt min tanke till uppställande av ett långsiktigt mål för forskningen inom ämnet som kan sammanfattas i formuleringen datorn som språkbrukare.

### *Datorn som språkbrukare?*

Låt oss försöka undvika missförstånd. Den oförlikneliga språkbrukaren är människan. Det är vi överens om. Analogin med datamaskinen innebär inte att människan betraktas som en automat, inte heller att datamaskinen någonsin kan förväntas nå människans kommunikativa höjder. Genom att placera datorn på språkbrukarens plats ställer vi de lingvistiska frågorna på sin spets. Det är det som är den vetenskapliga poängen.

Det finns en viktig poäng till. Datorn får ju allt fler användningar i samhällslivet. Därigenom ökar behovet av kommunikation med den. Det är väsentligt att denna kommunikation sker på människans villkor.

### *Var ligger tyngdpunkten?*

Låt oss först kasta en blick på modellen. Den ser ut ungefär så här. Språkliga yttringar kommer in till datamaskinen. Det gäller att avkoda dem – jag undviker ordet *förstå*. De skall i sin tur ge upphov till utdata, som kan vara av många slag, t ex en omskrivning, en konkret åtgärd, placering av språkgodset på rätt plats i en språkbank, ett svar på en fråga, en sammanfattning eller ett utkast till översättning.

Allt detta kräver naturligtvis kommunikativa resurser. Dem kan man ordna i fyra grupper. Med din tillåtelse betecknar jag dem med några facktermer och kommer strax in på vad de står för: lingvistisk kapacitet, encyklopedisk information, dynamiska faciliteter och algoritmer. Den lingvistiska kapaciteten gäller det lexikaliska systemet, det grammatiska systemet och språkbruket. Observera att data om språkbruket, användningen av språk-systemet, inte är mindre viktiga än data om systemet självt. Man kan varken analysera eller generera naturligt språk på ett tillfredsställande sätt utan en uppsjö av information om hur och i vilken utsträckning språkssystemets olika delar används.

Den encyklopediska information som krävs är systematiserad kunskap om världen. Vad kan exempelvis befinna sig ovanpå något annat? Fanns det verkligen bussar på Karl XII:s tid? Uppgiften är lindrigt talat stor och långsiktig. Den inbegriper flera olika vetenskapsgrenar. Detsamma gäller de dynamiska faciliteterna, som innefattar sådant som modeller för uppbyggande av kunskap genom perception och slutsatsdragning och strategier för konversation. Den forskning som bedrivs inom området artificiell intelligens är av stor betydelse härvidlag. Algoritmer slutligen behövs naturligtvis över hela fältet. Deras roll är ju att i detalj ange de procedurer som programmeringen bygger på.

Svaret på frågan om var tyngdpunkten ligger blir efter detta, att utforskningen av den lingvistiska kapaciteten och de därmed oskiljaktigt förenade algoritmerna är det centrala.

*Människan har ju också en psykologisk dimension. Vad blir det av den?*

Analogin med språkbrukaren framhäver i själva verket den aspekten. Att en modell har vad som brukar kallas psykologisk relevans får allt större betydelse. Den psykologingvistiska forskningen är också livaktig och följs med uppmärksamhet. Resultaten är emellertid ännu ganska motsägelsefulla och därför inte utan vidare tillämpliga. Men om det kan göras troligt att den ena av två i övrigt likvärdiga modeller är psykologiskt mera träffande än den andra, är det naturligtvis att föredra.

I vårt arbete på att komma åt de operativa enheterna i det lexikaliska betydelsesystemet har vi försökt att fånga de kategorier som kan antas vara relevanta för en observant och intresserad språkbrukare med tillgång till några vanliga nuspråkliga ordböcker.

*Har detta synsätt stått klart från början?*

Det har vuxit fram efterhand. Några saker har stått klara från början. Det hör betydelsen av att undersöka autentiskt språk, att göra det på basis av teoretiskt grundade modeller och att utveckla datamaskinell metodik för det.

*Vad kan datamaskinen redan nu göra av det vi har varit inne på?*

Jag kanske får precisera din fråga på en viktig punkt. Data-maskinen gör inget annat än exakt det som vi programvägen har gett den möjlighet att göra. Det är alltså vi som verkar genom den. När detta är sagt, vill jag svara att vi under överskådlig tid får räkna med samarbete mellan maskinens programsystem och människan genom interaktiv databehandling. Successivt kan behovet av mänskligt ingripande under körningarna av allt att döma bli mindre. I flera avseenden kommer det förmodligen att bestå, såsom vid lösning av besvärliga flertydigheter, hantering av avsiktliga ordlekar av typen "Niagara är ett gränsfall", analys av texter där olika sociala och religiösa system interfererar med

varandra, översättning av många slags texter osv. Å andra sidan finns det redan nu programsystem som kan utföra ganska goda morfologiska, syntaktiska och lexikaliska analyser, svara på frågor inom vissa områden osv.

Sedan hör det till saken att datorn också är ett redskap i själva forskningsprocessen. Den används nämligen för att bygga upp de kunskapsmängder och utveckla de system som krävs enligt språkbrukarmodellen. När det gäller den lingvistiska kapaciteten kan datamaskinen bland annat användas för experiment, t ex testning av språkregler, och insamling och bearbetning av lexikaliskt material. Den kan också användas som en stor och flexibel informationsbevarare.

*Vilken roll spelar datateknikens utveckling?*

Den är viktig på det sättet att den successivt gör tidigare omöjliga ting möjliga och tidigare tidsödande ting snabbt avklarade. Och utvecklingen har gått raskt. En stordator från mitten av 1950-talet som kostade 3 000 000 kronor kunde utföra 2000 instruktioner per sekund och hade en vikt av 7 ton. En mikrodator från mitten av 1970-talet som kostade 3000 kronor kunde utföra 200 000 instruktioner per sekund och hade en vikt av 7 kilogram.

För oss är minidatorn inte minst viktig. Den första modellen av en sådan lanserades 1960. Man har jämfört utvecklingen av minidatorn med utvecklingen av en bil som Volkswagen för att konkretisera förhållandena. Om bilen hade haft motsvarande utveckling av prestanda och pris, så skulle den nu ha haft en topphastighet av 100 000 kilometer i timmen och kostat 10 kronor. Den avgörande slutsatsen av detta är, att vi inte skall låta oss hindra av för tillfället rådande teknologiska begränsningar vid utvecklingen av forskningsmetodikerna på området. Datamaskinen är trots allt bara ett tredjedels sekel gammal.

*Vilken utrustning har institutionen?*

Vi har vad man kan kalla ett stort minidatorsystem som är speciellt utformat för behandling av språkligt material. Det kräver bland annat stor minneskapacitet, god strängbehandling, obegränsad typuppsättning och terminaler för interaktiv bearbetning. Forskarnas närkontakt med datamaskinen via textskärmsterminaler är av mycket stor vikt. Vi utnyttjar också flitigt ett par terminaler som är kopplade till Göteborgs Datacentral, där vi får tunga bearbetningar utförda.

*Vilken inriktning har institutionens forskning för närvarande?*

Frågan förutsätter helt rimligt att vi inte försöker täcka hela fältet. Man kan säga att datalingvistiska undersökningar av det lexikaliska systemet och av språkbruket utgör huvudområdena. Efter frekvensundersökningen som nu avslutas är det största projektet det som syftar till att lägga upp en omfattande svensk lexikalisk databas på modern lingvistisk grund och med det levande språket som källa. Genom denna förs utvecklingen av lexikon för datamaskinell analys och syntes vidare. Från databasen skall vi också generera en ordbok över modern svenska för allmänt bruk. Andra lexikaliska arbeten gäller ordlistor för invandrare och en studie av alla svenskers namn. Namn är ett rätt stort inslag i språkets lexikon. Under hösten 1979 utkommer Förnamnsboken.

Vi strävar efter att också arbeta med projekt som rör system och procedurer på den mera renodlat grammatiska sidan inom ramen för algoritmisk textanalys. En studie i program för morfologisk och syntaktisk analys är just avslutad.

Av flera skäl är språkmaterialet oftare skrivet än talat språk. En del av undersökningarna tar emellertid hänsyn till talspråk. Ett projekt arbetar helt med det. Undersökningen gäller vissa kommunikativa drag i talat vardagsspråk i Göteborg och deras relation till sociala faktorer.



*Vad får forskningen för konsekvenser för teoribildningen?*

Tillkomsten för första gången i historien av en symbolbehandlande maskin är naturligtvis en utmaning för lingvistik. Detta understryks av att minneskapacitet, hastighet och kompaktitet nu närmar sig den mänskliga hjärnans. I stället för att begränsa oss till spekulationer om språkbeskrivningens utformning kan vi använda datamaskinen som redskap. Det visar sig att också starkt uppmärksammade teoretiska inriktningar råkar i svårigheter. Det som skymtar är en lingvistisk teori som en teori om analys och syntes av naturligt språk med hjälp av en uppsättning kommunikativa resurser inklusive algoritmer.

*Det talas ibland om mjukdata. Vad blir det av dem?*

I vissa sammanhang talar man mycket riktigt om mjukdata, som då gärna ställs i motsats till hårddata. Detta väcker lite av samma föreställningar som mjukvara och hårdvara inom databehandlingen, dvs. program respektive maskinutrustning. Intressant är då att gränsen mellan dessa håller på att mjukas upp. Med hjälp av mikroteknik kan program nu monteras in som maskinkomponenter. Vi får hård mjukvara.

Som mjukdata betecknas exempelvis uppgifter om människors attityd till olika språkliga uttryck, t ex yrkesbeteckningar, människors upplevelse av sjukdom eller åldrande osv. Som hårddata betecknas då observerade språkliga frekvenser, resultat av laboratorieprov osv. I båda fallen gäller emellertid att uppgifterna måste svara mot vetenskapliga krav på noggrannhet i dokumentationen. De kan då principiellt också behandlas med likartade metoder. Orden mjukdata och hårddata representerar därför snarare olika vetenskapliga forskningsinriktningar än fundamentalt olika slag av data.

*Kan man då också säga att epistemiska, fenomenologiska, holistiska och ontologiska frågeställningar inom språkvetenskapen är förenliga med forskningen på området?*

Naturligtvis.

*Är det riktigt att beteckna ämnet som tvärvetenskapligt?*

Ja och nej. Beteckningen tvärvetenskap är inte alldeles klar. Vid dess sida står bland annat mångvetenskap som benämning på forskningsföretag som bygger på samverkan mellan olika vetenskaper. Sann tvärvetenskap innebär en sammansmältning av flera ämnen till en ny helhet. Detta är den tvärvetenskapliga paradoxen. När ett nytt ämne har bildats på det sättet, kan det ju inte längre med rätta kallas tvärvetenskapligt annat än från historisk synpunkt.

Vårt ämne har emellertid livliga förbindelser med andra ämnen. Det är också hjälpvetenskap för ämnen inom och utom språkvetenskapen och utnyttjar i sin tur andra ämnen som hjälpvetenskaper.

*Vad gäller för utbildningen?*

Som den är upplagd nu riktar den sig enbart till doktorander och utgår från grundexamen med tre terminers studier i allmän språkvetenskap eller motsvarande kunskaper. Studier i informationsbehandling är inte ett förkunskapskrav. Den som saknar sådana kunskaper får tillägna sig dem under utbildningen genom kursläsning och laborationer. Utöver doktorandutbildningen ger vi ibland orienteringskurser för olika grupper. Den verksamheten borde nog vidgas.

*Vart vänder man sig om man vill utnyttja institutionens språkmaterial?*

Då tar man kontakt med institutionens serviceorgan Logoteket. Det är en dynamisk språkbank som inrättades 1975 som en permanentning av en länge bedriven serviceverksamhet. Logoteket har vissa nationella uppgifter, nämligen att samla in och bevara datamaskinellt läsbara texter, att tillhandahålla sådana texter och vissa bearbetningar av dem, att bygga upp en ordbank och att ge råd på sitt område, allt i mån av resurser.

Textsamlingarna omfattar nu omkring 30 miljoner ord ur bland annat skönlitteratur, tidningar och författningar. Två spe-

ciella textmängder är särskilt aktuella. Den ena är riksdagens snabbprotokoll från arbetsåret 1978–79, tillsammans omkring fyra miljoner ord i en språkform någonstans mellan talspråk och skriftspråk. Den andra är Strindbergs samlade skrifter som skall kodas in och bearbetas här som ett led i kulturrådets nyttigivning. Omfånget uppskattas till mellan sex och sju miljoner ord.

Ordbanken innehåller ett par hundra tusen ord försedda med varierande uppgifter. Huvudinslag är frekvensordbokens olika bearbetningar och Svenska Akademiens ordlista. De inkommande texternas ordförråd tillförs successivt.

Konkordanser på mikrokort framställs systematiskt. Belägg på olika typer av ord och konstruktioner efterfrågas också. Avnämarna finns inom språkforskning och samhällsforskning, språkvård och journalistik, informationsbehandling och informationssökning, olika grenar av grafisk industri osv.

*Hör arbetet kring språkbanken till den vetenskapliga verksamheten?*

På flera sätt. Urvalet av språkmaterial och utformningen av databaserna, innehållsligt och tekniskt, är exempel på forskningsproblem. Men det finns skäl att anlägga ett vidare perspektiv. Inom en vetenskapsgren, och då inte minst språkforskningen, står vad man kunde kalla vetenskapsvården i centrum. Vad det gäller är teoribildning, metodutveckling och kunskapsgenerering. Ett annat betydelsefullt område är utbildningen, som ju är en förutsättning för att den vetenskapliga genomlysningen skall föras vidare. Ett tredje område har jag kallat resultatvård. Här hör servicen hemma tillsammans med praktiska tillämpningar inom olika samhällssektorer. Forskarnas del i ansvaret för resultatens användning kommer också in i bilden.

*En sak till. Intuitionen, vart tar den vägen?*

I grunden finns det ingen motsättning, menar jag, mellan ett datalingsvistiskt betraktelsesätt och en intuitiv uppläggning. Ytterst bör det nämligen vara så, att intuition är grundad på skarp ob-

ervation. Förhåller det sig på det sättet, och min intuition får mig att tro det, blir ämnesområdets framtid inte mindre intressant.

Efter avslutat samtal förfogar man sig på klassiskt maner till caldarium, där armhävning och hett bad väntar.

## Skriva som vi talar?

Språkform och språknorm. Studier tillägnade Bertil Molde. 1979.

Det drog ihop sig till val 1968. Frågan om andrakammarparlamentarismen stod i förgrunden. Här är ett brottstycke ur ett telefonväktarprogram. Den första repliken är programledarens.

P: Om ni följer författningen, var kommer då den här 50-procentiga ...

E: (*kort och bestämt*) Nä, alltså, nä, det är ett tillmötesgående mot de borgerliga kraven för att man skall på enklare sätt skall kunna skapa förviss ... visshet om att (*med sjunkande, gravlik stämman*) det är omöjligt för att föra en socialdemokratisk politik. Det är riktigt. Det är tillmötesgående som ... som har missuppfattats men som ... som ... ehh ...

B: På vad sätt är det ett tillmötesgående?

E: Jo, därför att det är ju ett mycket (*med lätt skratt*) enklare sätt att beräkna det här. (*undervisande*) Alltså det är ju så här. (*paus, därefter förtjust skratt*) Ja, det är ju mitt favoritämne som ... som hr Borgström nu lockat in på här.

B: Ja, det går bra, bara fortsatt.

E: Ja, ha, ha (*mer hjärtligt skratt*).

B: (*hjärtlöst*) Bara fortsatt. Det är mitt favoritämne också.

E: (*låter sig ej bekommas*) Var ... var bor hr Borgström. I Stockholm?

B: I Halmstad.

E: Ojojjoj ... Och där som jag är på somrarna ... (*hjärlöst*) Men varför kunde vi inte ha gjort upp det här där då. Alltså, jag vill kort och gott säga följande. Vi har tyvärr (*skratt*) ett ett ett valsysteem som gör att de borgerliga sannolikt blir överrepresenterade i andra kammaren.

B: Ett ögonblick där. Är inte socialdemokraterna också överrepresenterade där?

E: Inte i andra ... jo, alltså men inte ... inte ... borgerligheten räknar ju alltid (*skratt*) som om det skulle finnas ett block ... ett socialistiskt block såväl som ett borgerligt block ...

B: Ja, och det gör ju inte hr Erlander, va?

Samtalet mellan dåvarande statsministern Erlander och den skeptiske hr Borgström från Halmstad fördes sex dagar före valet. Det återgavs i Dagens Nyheter den 12 september under rubriken Vi träffas vid Laxvik.

Det är en van talare vi möter i huvudpersonen. Likväl är meningsbyggnaden med alla sina lösa ändar – talspråklig. Inget ont i det i och för sig, men är det så vi skall skriva? Postverkets reklamkampanj Skriv som du talar kan uppfattas så.

”Han ä ssäja försikti” är en replik ur dagligt tal som jag har lagt på minnet, mest på grund av det förlängda s-et. ”Han e så säja försikti” kunde vara en återgivning av ett annat belägg på uttrycket. Eller ”Han är so at säja försiktig” eller rentav ”Han är så att säga försiktig”. Skall man ta slagordet på orden, ger det oss en stavningväxling åt det här hållet. En sådan skulle vara mycket opraktisk. Argumenten för den bedömningen har blivit många, sedan parollen fördes fram i stavningsdiskussionen för några hundra år sedan. En del av dem tas upp i min uppsats Stavning 1973.

Konsekvenserna är fler. De gäller ju inte bara struktur utan också stil och stringens. Därtill kommer i själva verket att det är principiellt omöjligt att skriva som man talar. Orsakerna ligger i de båda kommunikationssituationernas olikhet och i de båda uttrycksmediernas olikhet. Detta är välbekant för språkvetare. Ett sätt att få frågan belyst utöver vad det inledande exemplet ger är att tänka igenom vad en paroll Tala som du skriver skulle innebära.

Det står med andra ord klart att postverket inte menar vad slagordet säger. Det är fråga om en vinkling. Enligt Carl-Otto Ohm vid postverkets marknadsavdelning, där ansvaret ligger, förstår 95 % av svenska folket att det förhåller sig så. Han gjorde skattningen vid den språkvårdskonferens som arbetsgruppen för språkvårdssamverkan anordnade i mars 1978. Om vi leker med procenttalet, finner vi att 400 000 svenskar lämnas åt sitt öde. Den fråga som främst anmäler sig är emellertid, vad som föranleder postverket att på egen hand göra denna stora satsning under denna paroll.

## Motiveringar

Vi kommer in på olika sätt att motivera insatser på språkvårdens område. Det finns en grupp motiveringar som jag kunde tänka mig att sammanfatta som *språkvetenskapliga*. Ett par tar sikte på själva det språkliga mediet. Hit hör i mångt och mycket purismen (Swedberg, Collinder och andra). Hit hör också inriktningar som ställer upp äldre ("klassiska") skeden som mönster (Rydqvist). I ett vidare språkvetenskapligt synsätt vidgas den kommunikativa ramen genom att avsändaren också tas med. Man hänvisar då till språkbruket (Sahlstedt, Leopold). När man så tar hänsyn också till mottagaren, är man framme vid ändamålsenlighetsprincipen, den funktionella inriktningen. Det är den som har dominerat tänkandet under de senaste hundra åren (E. Tegnér d.y., A. Noreen, Wellander osv.).

Vid sidan härav har det vuxit fram en *yrkesmässigt* motiverad språkvård. Den svarar mot yrkeslivets behov att precisera och förenhetliga. Här är det således fråga om att auktorisera fackspråkliga termer och att fastställa termer i samband med standardisering. En drivkraft har givetvis varit en strävan att höja effektiviteten. Arbetet gagnar också undervisningen och förvaltningen, och staten har gått in med ekonomiskt stöd.

Under senare år har också *politiska* motiveringar förts fram. Jag bortser då från det politiska inslag som purismen kan ha. En av de politiska inriktningarna är naturligtvis demokratismen med honnörssord som jämlikhet och tolerans. Se t.ex. Karl-Hampus Dahlstedts väldokumenterade uppsats Demokrati och språkvård 1977. I denna anda har en antibyråkratiutredning tillsatts, och tjänster som språkkonsult har inrättats vid vissa myndigheter. En annan politisk motivering utgår från en samhällsuppfattning av konfrontationstyp. Det samlande ordet är maktspråk. Se t.ex. volymen med detta namn från 1973 (Pockettidningen R). I stort sett varje dokument från myndigheterna uppfattas i detta fall som ett utslag av förtryck.

För egen del skulle jag vilja lägga till en fjärde huvudtyp. Som humanist är man intresserad av människan som människa. Språket – människans specifika uttrycksmedel – får därvid en

särskild dignitet. Att tränga in i språket blir att tränga in i människan. Det blir också angeläget att dela med sig av den insikt man kan vinna om språket i dess olika dimensioner. Vad jag vill ange är sålunda en i djupare mening *humanistiskt* motiverad språkvård.

Postverket har inte till uppgift att driva någon språk-vetenskapligt, politiskt eller humanistiskt motiverad språkvård. Och en yrkesmässigt motiverad språkvård skulle i dess fall ha gällt post-termer.

Vad vi iakttar är något nytt i språkvårdens historia, en *kommersiellt* motiverad språkvård. Det rör sig alltså om en femte huvudtyp. Bakgrunden är att postverket är affärsdrivande. Vill man uttrycka saken litet tillspetsat, kan man säga att kampanjens slagord Skriv som du talar i grund och botten är en vinkling av Köp fler frimärken. Detsamma gäller principiellt de brevställare som postverket tidigare har producerat. Jag vill inte alls förneka att det dessutom finns ett intresse för språkfrågan inom postverket.

## Program

Behöver en kommersiellt motiverad språkvård vara av ondo? Naturligtvis inte. All verklig hjälp är välkommen. Vad det hänger på är hur budskapet uppfattas. Tittar man närmare på postverkets material, finner man den alternativa formuleringen ”utrota krångelsvenskan”. Om det budskapet går fram, har kampanjen inte varit förgäves.

Här måste dock en kommentar göras. Det framgår av arbetshäftet för skolan, att krångelsvenskan utan reservation sätts lika med det officiella språket. Detta är en alltför grov generalisering i maktspråksanda. Det finns ingen anledning att utrota det officiella språket i de växande delar som redan har en enkel och lätt-tillgänglig prägel. Här borde man i stället satsa på uppmuntrande tillrop.

Att få myndigheterna att skriva enkelt och rakt på sak har länge varit en central uppgift för Svenska språknämnden. Några



exempel på skrifter är Erik Wellanders Kommittésvenska 1950, promemorian Språket i lagar och andra författningar 1967, som bär Bertil Moldes signatur, och Bertil Moldes och Margareta Westmans Offentlig svenska – en textbok 1975.

Det som saknar motstycke i detta långvariga arbete är den aktuella kampanjens format. Det hänger självklart ihop med den kommersiella bakgrunden. För språknämnden måste detta vara en tankeställare. Näst efter frågan om språkvårdens teoretiska underlag är frågan om språkvårdens kanaler den viktigaste enligt mitt sätt att se.

När kampanjen har klingat ut, vad kan man då vänta finns kvar i språkbrukarnas medvetande? Det bästa tipset måste vara: rätt och slätt slagordet *Skriv som du talar*. Risken är säkert inte liten att parollen kommer att uppfattas som ett mål för skriftspråket. Att detta som sagt är principiellt omöjligt är ovidkommande härvidlag.

När jag hade kommit till denna punkt i resonemanget, dök en artikel av Sven Delblanc med anknytning hit upp i *Dagens Nyheter* (1/7 1978): Den goda svenska stilen måste erövras på nytt. Den har hare, som tager honom. Delblanc kommer mot slutet in på postverkets kampanj. Artikeln har intresse i detta sammanhang, därför att den bekräftar mina misstankar om slagordets roll. För Delblanc är nämligen kampanjen helt enkelt parollen. ”Läs affischer och broschyrer: ’Skriv som du talar!’ Men det är just det vi inte får göra.” Som den verksamme språkbrukare han är tar han fortsättningsvis i med starka ord mot uppmaningen.

Det är dags att sätta in parollen som sådan i ett större program-sammanhang. Jag skulle vilja formulera ett dynamiskt program för skriftspråk med fyra huvudpunkter.

- (1) *Skriv*.
- (2) *Skriv läsvänligt*.
- (3) *Skriv insiktsfullt*.
- (4) *Skriv uttrycksfullt*.

Den första punkten ligger i linje med vad man finner i skolöverstyrelsens rapport *Basfärdigheter i svenska 1975*: det är bättre att skriva vad man tycker och tänker än att inte skriva alls.

Skriv som du talar är inte långt borta. Det verkligt viktiga är att denna första punkt inte får framställas som slutmålet.

Att skriva läsvänligt innebär flera olika saker. Det är betydelsefullt att språket är störningsfritt i den meningen att formen – stavningen, ord- och satsbehandlingen – inte drar till sig uppmärksamheten på innehållets bekostnad. Att kunna ansluta sig till en riksgiltig språkform som successivt utvecklas är från den synpunkten en ovärderlig tillgång. Det läsvänliga språket måste också vara mottagaranpassat i den meningen att det är effektivt från kommunikationssynpunkt genom att ha en genomskinlig struktur och utnyttja gemensam erfarenhet och kunskap. Skriv som du talar täcker bara delvis denna huvudpunkt.

Målet är inte nått ändå. Att skriva insiktsfullt innebär bland annat att inte ta t.ex. annonsörernas och politikernas språk med alla de semantiska fallgroparna för gott och utan närmare eftertanke göra det till sitt. Här krävs en djupgående insats i riktning mot höjd språklig medvetenhet. Vi är inne på en verklig kärnfråga. Denna punkt ligger klart över den diskuterade parollens nivå.

Min fjärde huvudpunkt gäller uttrycksfullheten i språket. Att den har stor betydelse för effektiviteten i kommunikationen menar jag är fullt klart. På ett allmänt plan gäller det här exempelvis åskådlighet i framställningen, slående exemplifiering, omsorg i ordvalet och finess i uppläggnigen. Mera speciellt närmar vi oss här den konstnärliga gestaltningen av språket. Visst finns det uttrycksfullt tal, men Skriv som du talar tar knappast sikte på det.

Sett i hela det nu skisserade programmets perspektiv blir det tydligt att parollen som den står tenderar att ge ett intryck av ett alltför torftigt slutmål. Detta skulle även på ganska kort sikt kunna leda till förfång för vår språkgemenskap.

I stora drag ligger det alltså till så, att postverket har gått ut på ett länge bearbetat område, med tidigare oanade resurser, under en paroll som inbjuder till missförstånd. Låt oss hoppas att resultatet blir övervägande positivt för språkbrukarnas del. I synnerhet som vi är överens i sak: vi skall inte skriva som vi talar.

## Litteratur

- Allén, Sture, Stavning. (Ord och stil. 5. 1973, s. 46–55.)  
Basfärdigheter i svenska. Skolöverstyrelsen. 1975.
- Dahlstedt, Karl Hampus, Demokrati och språkvård. (Kungl. Vitterhets Historie och Antikvitets Akademiens Årsbok 1977, s. 121–140.)
- Delblanc, Sven, Den goda svenska stilen måste erövrans på nytt. Den har hare, som tager honom. (Dagens Nyheter 1/7 1978.)  
Maktspråket. Om språket som medel för förtryck. (Pockettidningen R. 3–6. 1973.)
- Molde, Bertil, & Westman, Margareta, Offentlig svenska – en textbok. 1975.
- Skriv som du talar! Arbetshäfte för skolan. Utg. av Postens Marknadsavdelning. 1977.
- Skriv som du talar! Uppmuntrande läsning för alla som vill utrota krångelsvenskan. Utg. av Posten. 1977.
- Språket i lagar och andra författningar. Statsrådsberedningen. 1967.
- Wellander, Erik, Kommittésvenska. (SOU 1950: 26.)  
Vi träffas vid Laxvik. (Dagens Nyheter 12/9 1968.)

## Om språkvårdens principer

Språk i Norden 1979.

Vid det nordiska språkmötet i Helsingfors 1975 tog jag upp frågan om inte språkmötena borde behandla språkvårdens teoretiska och principiella grundfrågor. Förslaget fick stöd av bland andra Erik Hansen. Vid det följande språkmötet i Húsavík på Island beslöts att principfrågorna skulle bli en huvudpunkt vid språkmötet 1977 i Trondheim, vilket också blev fallet. Vid detta möte framlade Erik Hansen, Henrik Galberg Jacobsen och Allan Karker material och resultat beträffande den danska språknämndens rådgivningsarbete. I ett inlägg i den följande diskussionen angav jag i åtta punkter några viktiga principfrågor för att understryka, att vi hade åtskilligt kvar att diskutera. Det överenskomms att de principiella frågorna skulle bli huvudämnet också vid det 25:e nordiska språkmötet i Visby 1978.

I den översikt som jag nu presenterar är de nyssnämnda punkterna insatta tillsammans med annat stoff i en större ram. Jag gör naturligtvis inte anspråk på att täcka alla frågor. Det allmänna förhållningssätt som jag efterhand har kommit fram till vill jag emellertid försöka förmedla. Min avsikt är att kommentera tio olika aspekter på språkvårdens principer. Det är i detta sammanhang naturligt för mig att sammanfatta och vidareutveckla synpunkter som jag har fört fram i några av mina uppsatser i språkvårdsfrågor sedan 1964.

### Förutsättningar

En grundförutsättning är det faktum att språknämnderna har vissa bestämda uppdrag. Svenska språknämndens stadgar är fastställda av regeringen. I ändamålsparagrafen står det att nämnden har till uppgift att följa det svenska språkets utveckling i tal och skrift samt utöva en språkvårdande verksamhet. Det står också att nämnden skall verka för nordiskt samarbete på språk-

vårdens område i syfte att vidmakthålla och stärka den nordiska språkgemenskapen. Detta kan stå som exempel på hur en språknämnds uppdrag har formulerats.

Det måste också konstateras att det råder olika förhållanden i de olika länderna. Vi har exempelvis Finland med dess officiella tvåspråkighet och relationerna till rikssvenskan och sverigefinskan. Vi har Norge med dess speciella tvåspråkighet. Vi har Island med vad man kunde beteckna som en utpräglad enspråkighet (med visst tryck från amerikansk engelska). Vi har det samiska språkområdet som sträcker sig över nationsgränserna. Uppräkningen kunde fortsätta, men det sagda ger en tillräcklig antydan om komplexiteten.

Mångfalden får emellertid inte tillåtas att skymma de gemensamma problemen. Tvärtom bör vi dryfta grundfrågorna med utnyttjande av den stora och växlande erfarenhet som sålunda finns.

## Mål

Enligt mitt synsätt har språkvården att beakta både språket och språkbrukarna. Språkvården bör sträva efter att bidra till att göra det berörda språket till ett smidigt och uttrycksfullt redskap i kommunikationen. Språkvården bör också sträva efter att hjälpa språkbrukarna att komma till insikt om de språkliga medlen och därigenom nå en bättre behärskning av dem.

Med smidighet menar jag bland annat störningsfrihet i den meningen att uppmärksamheten inte dras från innehållet till formen. Genomskinlighet i strukturen är en annan faktor.

Uttrycksfullheten gäller både den lexikaliska och den grammatiska behandlingen av språket. Den gäller också sådana ting som åskådlighet, exemplifiering och liknande. Den här punkten vetter åt konstnärernas umgänge med språket.

Med insikt menar jag bland annat att inte falla till föga för förföriskheterna i exempelvis kommersiellt, politiskt eller teknologiskt språk. En höjd språklig medvetenhet borde kunna ge ett

bättre grepp om både perception och produktion av språk. Från en synpunkt sett är detta en språksocial inriktning.

## Område

Först konstaterar vi att fiktionslitteraturen naturligtvis faller utanför språkvårdens verksamhetsområde. Det centrala är allmänspråket i övrigt. Som så ofta när det gäller mänskliga aktiviteter är det svårt att ge en skarp definition. Låt oss här framför allt se allmänspråket i relation till fackspråk och notera, att det medger kommunikation över yrkesgränserna.

Fackspråk däremot hänför sig just till yrkeskategorier. I terminologioorganens fall rör det sig om auktorisering av en speciell vokabulär. Självfallet finns det emellertid också spontant fackspråk. På allmänspråkets vägnar har språknämnderna viktiga intressen att bevaka även när det gäller fackspråk.

## Motiveringar för språkvård

Språkvård kan motiveras och har motiverats på åtskilliga sätt. Jag betecknar en första grupp som språkvetenskapliga motiveringar, därför att de har förts fram och debatterats i språkvetenskapliga kretsar. Direkt inriktning på det språkliga mediet har purismen, vars syfte är att slå vakt om ett i viss etymologisk mening renare språk. Purismen har en del gemensamt med den historiska skolan, som ställer upp äldre språkskeden som mönster. En annan språkvetenskaplig motivering inbegriper även avsändaren. Man vill värna om språkbruket. Ytterligare en annan inriktning tar sikte också på mottagaren. Det samlande begreppet är ändamålsenlighet. Det som är bra för både avsändaren och mottagaren är funktionellt. I viss mån kan man jämföra med vad som i Ferguson 1977 kallas *efficiency*. Detta är dock ett vidare begrepp, som också omfattar olika praktiska åtgärder.

Bakom auktorisering och standardisering av fackspråkliga uttryck ligger en yrkesmässig motivering. Den precisering som eftersträvas har till syfte att höja effektiviteten i yrkeslivet.

Politiska motiveringar har blivit mera framträdande än tidigare. Det ligger nära till hands att ställa upp det demokratiska idealet också för språkvården (se Dahlstedt 1977). Honnörsord som jämlikhet och tolerans kommer i förgrunden. På svenskt regeringshåll gäller intresset till mycket stor del myndigheternas språk. Man tillsätter språkkonsulter och vidtar andra åtgärder för att förbättra förståeligheten. En annan politisk motivering samlas kring begreppet maktspråk. Skrivelser från myndigheterna uppfattas till synes generellt som utslag av förtryck, åtminstone i den mest utpräglade formen av denna åsikt. Till gruppen politiska motiveringar för jag också sådana nationella strävanden som syftar till frigörelse från ett annat lands inflytande i politiskt, kulturellt och språkligt avseende. Också arbete på att bevara en befolkningsgrupps språk och kultur utan att syfta till nationell självständighet kan sägas höra hit.

På senare tid har man också kunnat iaktta en kommersiellt motiverad språkvård. Det tydligaste exemplet är det svenska postverkets kampanj *Skriv som du talar*. Slagordet är olyckligt valt, men syftet är i och för sig gott: att stimulera språkbrukarna att skriva enklare. Skriver de enklare, kan de antas skriva oftare och därigenom öka postverkets intäkter.

I anslutning till vad jag sade om språkvårdens mål vill jag också ange en i vidare mening humanistisk motivering för språkvård. Jag tänker då inte på vad som ibland lite diffust kallas de humanistiska värdena utan på ett intresse för medmänniskorna i språkgemenskapen och deras språkliga aktivitet. Tolerans med språkbrukares uttryckssätt anger inte den inställning jag avser. Drift med enskildas uttryckssätt ligger långt ifrån. Centralt blir just intresset för användningen av språket och för möjligheten till vidgning av registret.

### **Språkvårdens komponenter**

Givetvis är stil en viktig komponent i språkvården, även om långt ifrån allt är stilfrågor (Allén 1966, 1968). Att det är angeläget att försöka förmedla en känsla för stillägen är fullt klart. Vi kan se

detta som språkvårdens estetiska komponent. Uttryck kan vara bra från lexikalisk, grammatisk och logisk synpunkt men dåliga från stilistisk. Jag tar ett känt fall av inbäddning som exempel: *Hela skogen, som björkarna, som gulnade, växte i, dog*. Normalt är detta skralt från stilistisk synpunkt alldeles oavsett att korttidsminnets restriktioner också spelar in.

En annan av komponenterna gäller stringensen. Här är det fråga om den logiska aspekten, klarheten. Det är naturligt att språkvården avråder i fall som detta (efter Wellander): *Yemens invånarsiffra är nu uppe i två och en halv miljon, varav 70 000 är judar, som utvandrade för 2 300 år sedan*.

Den betydelsefullaste komponenten är enligt min mening otvivelaktigt den som rör strukturen. Här gäller det lexikon och grammatik i vid mening, språkets kärnområden.

## Norm och språkbruk

Norm kan sägas vara ett system av språkregler som styr det sätt på vilket man uttrycker respektive förstår ett givet innehåll. Mina synpunkter tar mera sikte på arbete med en befintlig norm än på etablering av en norm, som fallet är exempelvis i en rad u-länder.

Vi har alla erfarenhet av olika normer. I grova drag kan vi se dem som motsvarigheter till olika roller i vår tillvaro. Vi har varierande normer inom familjen, bland vännerna, bland idrottskamraterna, i yrkeslivet, i den regionala språkgemenskapen, i den nationella språkgemenskapen osv. Det är viktigt att slå fast att de olika normerna har utvecklats för och fungerar för olika syften.

Det är inte språkvårdens uppgift att verka för en nivellering av dessa varierande normer med all deras språkliga och kulturella rikedom. Huvuduppgiften måste i stället vara att sätta in krafterna på den nationella normen. Vi behöver ett redskap för kommunikationen mellan invånarna i samhället. En norrbottning och en bohusläning måste kunna komma till tals med varandra. Det samma gäller sociala grupperingar.

Vi behöver också en gemensam språkform som vi kan erbjuda dem som lär sig språket som främmande språk. Det är främst



fråga om invandrare men även om svenskstuderande på andra håll i världen. Lika litet som vi lär oss t.ex. Yorkshiredialekt när vi läser engelska bör vi själva tillhandahålla en eller annan dialekt med begränsad räckvidd.

Det är också av grundläggande betydelse att kunna ta del av en gemensam fond av språkfäst erfarenhet och kunskap på livets olika områden utan att behöva forcera ett stort antal växlande språkformer. Jag syftar inte bara på faktamängder utan också på (skön)litterärt dokumenterad livserfarenhet.

Ulf Teleman har tagit upp normfrågan i en skrift om språkriktighet 1978. Han gör där i anslutning till Habermas offentlighetsteori en uppdelning i två normer: privat språk och offentligt språk. Här är det privata språket tänkt att vara det vardagliga språket och det offentliga språket det som ligger utanför privat-sfären.

Genom en glidning i resonemanget gör emellertid Teleman en karakteristik av det icke-privata språket som i stort sett inskränker det till överhetens språk (s. 6): "Offentlighetens språk är alltså i hög grad envägskommunikationens och masskommunikationens språk. Sändaren talar eller skriver ofta inte ens å egna vägnar utan för en organisation, en grupp, en myndighet, staten eller samhället självt med stort S." Härigenom blir följande beskrivning möjlig (s. 6): "Det offentliga språket är därför det opersonliga språket. Vi letar där förgäves efter privatsfärens känslor, bilder, lukter eller smaker."

Teleman nämner själv att framställningen är starkt schematiserad och påpekar, att fackligt och politiskt aktiva arbetare över-skrider privatsfären i sitt språkbruk (s. 5). I själva verket är schematiseringen kraftigare än så. Den övergripande normen är långt ifrån begränsad till det offentliga språket i inskränkt mening (motsvarande offentlig svenska hos Molde & Westman 1975). Tag som exempel från Sverige ett nummer av Aftonbladet eller Göteborgs-Posten, Ivar Lo-Johanssons Romanen om Måna eller Linnea Fjällstedts Hungerpesten eller Henrik Tikkanens 30-åriga kriget, David H. Ingvars Rapport från hjärnan eller Åke Ortmarks De okända makthavarna. Exempelen domineras inte av någon planläggningens opersonliga fyrkantighet.

Vad jag vill argumentera för är inte att den nuvarande riksomfattande normen i ett fall som svenskan är idealisk. Den är naturligtvis som alla andra normer präglad av sin tillkomst-historia. Och denna i sin tur är inte idealisk.

Det jag menar är att riksnormen har uttrycksmöjligheter av helt andra dimensioner än Teleman antyder. Jag menar också att den har goda utvecklingsmöjligheter. Det finns därför anledning att arbeta vidare med den i öppen och konstruktiv anda.

Relationen mellan norm och språkbruk är vidare av stort intresse. Man kan se en viss norm som en abstraktion ur motsvarande språkbruk. Normen kan inte förbli opåverkad av förändringar i bruket. Detta är den dynamiska aspekten på normen, ett mycket besvärligt problem för språkvården.

Samtidigt som språkvårdare måste vara snabbt ute när någonting händer, är det rimligt att de är något återhållsamma i sin bedömning av nyheter. En språklig nyhet kan inte utan vidare anses tillhöra normen. En betydelsefull diskussionspunkt är därför vilka riktlinjer man skall ha för värdering av nyheter i språkbruket. Jämför i detta sammanhang också Erik Hansens diskussion 1975 av den hemliga premissen.

## Argument

I min framställning av argumenten eller kriterierna inom språkvården kan jag gripa tillbaka på ett synsätt som jag först publicerade 1966 och sedan har återkommit till i några olika sammanhang. Jag räknar med två typer av kriterier: strukturvillkor och plusvärdesvillkor. Strukturvillkoren gäller anslutningen till den rådande normen. Plusvärdesvillkoren ställer kravet att den betraktade nyheten skall innebära en vinst av något slag för språket.

Jag har uppställt fem strukturvillkor. De rör uttal, stavning, böjning, bildning och icke störande sammanfall (homonymi). Strukturvillkoren skall uppfyllas obligatoriskt. Det är de som bevakar att språkets karaktär bevaras, och de är därför grund-

läggande. I denna modell är alltså normen i detta avseende explicit angiven.

När det gäller plusvärdesvillkoren har jag tyckt mig kunna göra en reduktion till fyra huvudpunkter. Jag räknar med begreppsmässigt, internationellt, funktionellt och socialt plusvärde. Ett ord eller uttryck som betecknar en ny företeelse har naturligtvis ett begreppsmässigt plusvärde. En aktuell fråga är om motsvarigheten till engelska *viewdata* och liknande system skall heta *data-TV* (*datateve*), *datavision* eller något annat. De nämnda alternativen är naturligtvis jämställda i fråga om begreppsmässigt plusvärde.

Det internationella plusvärdets roll är en kardinalfråga för det nordiska samarbetet på språkvårdens område. Här strider ju ofta det rent nordiska kriteriet mot det i vidare mening internationella. Om vi i Sverige hade stannat för *fjärrsyn* i stället för *television* eller *TV*, hade vi gjort gemensam sak med Danmark och Norge (och Tyskland) men avlägsnat oss från England, Frankrike osv. I den mån det blir aktuellt att införliva nya internationella ord och uttryck, vilket varierar från fall till fall och från språk till språk, kunde Norden med fördel gå på en gemensam linje. Den internt nordiska aspekten återkommer jag till.

I det funktionella plusvärdet sammanfattar jag sådana egenskaper som ökad smidighet, tydlighet o.d. För att ta ett enkelt och välkänt exempel: det är naturligtvis funktionellt att kunna skilja på *massproduktion* 'produktion i massor' och *massaproduktion* 'produktion av massa'.

Det sociala plusvärdet är självfallet viktigt. Det är på den här punkten som samhällssynen avspeglas. Vi har fått en ganska lång rad statushöjande beteckningar av typen *lokalvårdare* i stället för *städerska*. Vi använder numera *lärare* också som beteckning för kvinnliga pedagoger. Hela den språkliga delen av könsrollsdebatten hör givetvis hit, t.ex. frågan om huruvida det allmänna pronomenet *man* skall ersättas och i så fall med *han* eller *hon*, *vederbörande* eller någonting annat. I det här sammanhanget bör vi väl notera, att nya språkliga etiketter knappast förändrar samhällsförhållandena i djupare mening. Men också om de bara gör temporär nytta, kan de vara berättigade.

## Rekommendationer

I vissa fall visar de betraktade alternativens plusvärden på en bestämd rekommendation. Ett av alternativen har då fler plusvärden än de andra (plusvärdena viktas inte i detta resonemang). Från praktisk synpunkt har man i sådana fall ett bra läge.

Den modell jag antyder medger emellertid också att flera alternativ rekommenderas, nämligen om de har samma antal plusvärden. Detta ger bland annat möjlighet för synonymer att etablera sig. Efterhand kommer kanske ett val mellan dem att träffas i språkbruket. Modellen svarar här emot vad jag anser vara ett viktigt krav, nämligen att klargöra för rådfrågande att det kan finnas olika likvärdiga uttryckssätt, att inte alltid en sak är rätt och alla andra fel (se senast Allén i *Språkvård* 1978: 4).

Till rekommendationerna bör som jag ser det också fogas vad man kan kalla en varudeklaration. Vad jag menar är sådant som att informera om ett uttrycks stilvärde och dess konnotationer i politiskt, socialt osv. avseende när någon frågar om råd. Detta skulle i svenskans fall gälla exempelvis användningen av *dom* (pronomen), *teknokratiskt språk* (jämfört med *teknologiskt*, *tekniskt*) och uttryck som *det är upp till dig*.

## Behov

En viktig förutsättning för en fungerande språkvård är en ganska omfattande forskning. Det krävs bland annat forskning rörande språkvårdsteori, språkssystem (lexikon och grammatik) och språkbruk. Språkvårdsteorin bör uppmärksammas av både språknämnderna och universiteten. När det gäller språkssystem och språkbruk är det i allt väsentligt universitetens forskning man har att stödja sig på.

Kodifiering av normen är också ett viktigt krav. Det gäller å ena sidan ordlistor och ordböcker, å andra sidan grammatiker. Både normativa och deskriptiva arbeten är behövliga. I detta sammanhang kan det vara på sin plats att erinra om det problem

som ligger i att även deskriptiva verk tenderar att uppfattas normativt.

Tillgång till stora aktuella språkmaterial är också ett viktigt behov. Konkordanser på mikrokort har visat sig vara ett värdefullt hjälpmedel. Det är numera även möjligt att med hjälp av dataterminaler ha direktaccess till en språkbank. En sådan kan i princip tillhandahålla språkmaterial och ge kvalitativa och kvantitativa upplysningar om lexikaliska och grammatiska fenomen (se Allén 1978 i ALLC Bulletin). Logoteket i Göteborg byggs upp enligt en sådan strategi. Även i ett kort perspektiv får språkvården härigenom nya möjligheter med stor potential. Också detta bidrar till att göra språkvårdsdebatten angelägen.

### Konsekvenser

Det skulle inte vara orimligt att satsningen på principfrågorna fick en viss återverkan på språknämndernas arbetssätt. Jag menar naturligtvis inte att varje akt av rådgivning skall ackompanjeras av de argument som ligger bakom. Vad jag menar är att teoretiskt grundade överväganden skall ha gjorts, så att argumenten kan anges när språkbrukaren frågar eller när det annars förefaller berättigat. Det ligger utanför mitt ämne att gå närmare in på detta. Jag är övertygad om att det är något som de som är verksamma inom den praktiska språkvården har för ögonen.

I ett nordiskt sammanhang som detta finns det dock anledning att se litet närmare på de praktiska problem som knyter sig till en av principfrågorna. I instruktionerna för Nordiska språksekretariatet och Svenska språknämnden, för att ta dem som exempel, står det ju att de skall vidmakthålla och stärka den nordiska språkgemenskapen. Regeringarna har alltså slagit fast att språkgemenskapen *skall* inte bara bevaras utan också vidareutvecklas. Vad detta innebär är inte alldeles lätt att säga.

Låt oss betrakta språksituationen en gång till. Egentligen har vi ett myller av språk i Norden, om vi nämligen tänker också på invandrarspråken. Dem kan vi emellertid avföra från denna speciella diskussion. Vad som återstår är brokigt nog.

Här finns danska, norska med de båda officiella språken bokmål och nynorska och svenska med de båda officiella språken (riks)svenska och finlandssvenska. Mellan dessa språk är förståeligheten god, delvis mycket god.

Vi har de ö nordiska språken isländska och färöiska. Mellan denna grupp och föregående grupp är förståeligheten starkt begränsad. Båda grupperna tillhör emellertid de germanska språken och är därigenom nära besläktade.

Helt obesläktade med de föregående är de båda finsk-ugriska språken finska och samiska. För finskans del kan man tala om en (riks)finsk och en sverigefinsk varietet. Också samiskan har vissa varieteter. Endast (riks)finskan har officiell status i denna grupp.

Grönländskan slutligen är ett eskimåiskt språk och som sådant obesläktat med såväl den germanska som den finsk-ugriska gruppen.

Hur skall de språkvårdsorgan som är ålagda att stärka språkgemenskapen inom denna krets handla? Skall man för att ta ett extremfall sträva efter att närma svenskan och grönländskan till varandra och i så fall i vilka avseenden? Frågorna är inte helt retoriska. Alla språken är representerade i Nordiska språksekretariatet.

Beroende på från vilket språks synpunkt man ser saken måste inriktningen bli olika. När det gäller de danska, norska och svenska språkformerna är det naturligt att söka en rätt bred samverkan. För finlandssvenskans vidkommande blir avstämningen i förhållande till finskan också betydelsefull. För att ta ytterligare ett exempel är det viktigt inte minst för islänningar och färöingar att främja den s.k. skandinaviskan, det talspråk som fungerar som ett slags lingua franca.

Lingvistiskt sett kan ett samarbete i språkvårdsfrågor ske på olika plan. Kraftigast blir effekten om ett uttryck till både form och innehåll överensstämmer (inom ramen för vad respektive språks strukturvillkor medger).

Emellertid kan en hel del vara vunnet om det går att nå en viss samklang i uppsättningen av nya begrepp även om motsvarande uttrycksformer växlar. Detta gäller inte bara fackspråk utan i betydande utsträckning också allmänspråk. Ett närliggande

exempel är begreppet 'språksekretariat'. Vi vet att det kommer att heta olika på olika håll i Norden, men innehållet har vi en tämligen överensstämmande uppfattning om. Det är på detta plan som svenskan och grönländskan i första hand kan tänkas mötas.

Ett tredje lingvistiskt samarbetsplan gäller språkvårdsteorin. En fortlöpande diskussion på det planet fäster jag stort avseende vid.

## Litteratur

- Allén, Sture, Dator, fackspråk och allmänspråk. (Språkvård 1978: 4, s. 21–23.)
- Allén, Sture, Dynamisk språkvård. (Dagens Nyheter 11/6 1968.)
- Allén, Sture, Massmedium, etermedium, mediatek. (Svenska Dagbladet 6/10 1964.)
- Allén, Sture, Språkvård i teori och praktik. (Antennen. Personal-tidning för Sveriges Radio. Juni 1966, s. 18–20.)
- Allén, Sture, The Role and Organization of a Corpus in the Investigation of Language. (Association for Literary and Linguistic Computing Bulletin 6:1. 1978, s. 10–14.)
- Dahlstedt, Karl Hampus, Demokrati och språkvård. (Kungl. Vitterhets Historie och Antikvitets Akademiens Årsbok 1977, s. 121–140.)
- Ferguson, Charles A., Sociolinguistic Settings of Language Planning. (Language Planning Processes. Ed. Joan Rubin et al. The Hague 1977, s. 29.)
- Hansen, Erik, Den hemmelige præmis. (Analyser af sprogbrug. Red. Klaus Kjøller. København 1975, s. 88–102.)
- Molde, Bertil, & Westman, Margareta, Offentlig svenska – en textbok. 1975.
- Teleman, Ulf, Språkriktighet i och utanför skolan. (ROLIG papir 11, Roskilde Universitetscenter. 1978.)
- Wellander, Erik, Riktig svenska. 2 uppl. 1941.

## Dator och datamaskin, fackspråk och allmänspråk

Språkvård 1978: 4. Föreliggande lätt reviderade version har publicerats i Tankar om språket, 1982.

I början talade man om *matematikmaskin*. Det var inte bra, eftersom apparaten långt ifrån bara används som räknare. Man kallade den också *elektronhjärna*. Det var naturligtvis ändå mycket sämre.

År 1956 lade Olle Karlqvist fram ett förslag till terminologi på området. Tekniska nomenklaturcentralen sände ut det på remiss till en liten grupp personer och fick in en del kommentarer. Diskussionen fördes vidare i en arbetsgrupp inom Svenska Elektriska Kommissionen. Vid ett sammanträde 1957 kom det fram åtta olika förslag: *automatisk databehandlingsmaskin*, *databehandlingsmaskin*, *automatisk datamaskin*, *datamaskin*, *databehandlingsautomat*, *dataautomat*, *datomat* och *räkneautomat*. Senare samma år enades man om att föra fram förslaget *automatisk datamaskin*.

Gruppens förslag ledde till att termen *automatisk datamaskin* fastställdes i Svensk Standard 1961. Där upptogs också *datamaskin* som "förkortat uttryck för automatisk datamaskin". På 1960-talet var det normalt så apparaten hette. I Grönköpings Veckoblads narrspegel blev det förstås *datummaskin*.

En vacker dag stod det i DN (29/3 1969) att tidningen skulle börja använda *dator* i stället för *datamaskin*. Varför? Den huvudsynpunkt som fördes fram den gången var att man ville ha ett kortare ord. Det var det som var Börje Langefors' argument i den fotnot i boken *Introduktion till informationsbehandling* 1968 där ordet lanserades. Man nämnde också att det var lätt att göra sammansättningar som *datorstyr*. Man hade rentav vissa tankar om att *dator* skulle kunna få internationell spridning.

Langefors var emellertid inte den som hade skapat ordet. Det är belagt redan 1956 i ett svar på den nämnda rundfrågan. B. Andrén, ASEA, skriver: "Ett sätt att finna ett lämpligt begrepp



för det engelska uttrycket 'data processing machine' är att göra en helt ny konstruktion. Kanske är Dator acceptabelt." Ingen kommentar och heller ingen effekt i det läget. (Jag tackar Lars Algotsson, TNC, och Hans Svensson, SEK, för hjälp med en del av det äldre materialet.)

Beteckningen för den här maskinen har blivit ett ofta nämnt exempel i diskussionen om nuspråket. Detta hänger inte bara samman med att det gäller ett tydligt behov utan också med att man snabbt kommer in på flera viktiga principfrågor. Det finns anledning att titta lite närmare på saken.

## Nuläge

Hur är läget idag? I tidningsspråket som genom sitt inflytande är ett av allmänspråkets viktigaste områden används både *datamaskin* och *dator*. Det finns ungefär lika många av dem i ett material på en och en halv miljon ord från 1976. I riksdagens snabbprotokoll från 1978/79, som omfattar fyra miljoner ord, finns också båda, även om *dator* överväger.

Däremot har man satsat enbart på *dator* i fackspråket. I Dataordboken (1977) finns *datamaskin* inte alls upptaget. En viss uppmjukning kan emellertid skönjas. Språkvårdsgruppen – samarbetsorganet med representanter för Svenska Akademien, Svenska språknämnden, Tekniska nomenklaturcentralen, Standardiseringskommissionen och de statliga språkkonsulternas förening Samspråk – har rekommenderat "termerna *dator* och *datamaskin* för engelskans *computer*" vid ett sammanträde 1981.

På ett tidigt stadium började en språklig förenkling komma in. Precis som vi har fått (en) *radio* genom att kapa *radioapparat* fick vi (en) *data* genom att kapa *datamaskin*. Den korta formen mottogs med blandade känslor, och det var en bidragande orsak till att *dator* fick spridning. I Olof Johannessons (dvs. Hannes Alfvéns) uppmärksammade bok från 1966, Sagan om den stora datamaskinen, talas det på många ställen om *datan*. Svenska språknämnden och Tekniska nomenklaturcentralen har båda konstaterat att ordet är korrekt bildat.

I det tidningsmaterial och det riksdagsmaterial som jag nämnde finns det flera fall av *data* som beteckning på apparaten: *datan* skojade med småföretagare, ett minne hos *datan*, alla svar körs på *data* osv. I Aktuellt i teve har jag bland annat noterat: En del upplever *datan* enbart som en hjälp. I en annons frågas det: Vad vet *Datan* om *Dig*? Är *datan* på väg att ta hem spelet mellan *datamaskin* och *dator*?

### En ordgrupp behövs

Det som ligger *dator* i fatet är språkligt sett inte minst otympligheten i ordbildningen, trots *datorstyrd* och *mikrodator*. Vi behöver ju inte bara ett ord för själva apparaten utan också ord för närliggande begrepp som gäller verksamhet och egenskap. Se på uppställningen här nedan.

Apparaten	Verksamheten		Egenskapen
	1	2	
datamaskin	att data-	data-	datamaskinell
flera datamaskiner	behandla	behandling	
data	(att data)	(datning)	(datlig?)
flera datorer			
dator	(att dator-	(dator-	datorbaserad
flera datorer	behandla)	behandling)	(datorisk?)

Den första raden upptar en välkänd serie som utgår från *datamaskin*; här kan *data/behandla*, *-behandling* betraktas som ellipser för *datamaskin/behandla*, *-behandling*. Den andra innehåller verbet *data*, som väl hittills bara har framkastats tillfälligt. Det är emellertid regelbundet bildat, jämför t.ex. (en) *kratta*, (att) *kratta*. I och för sig kan *databelhandla* också fungera som verb till substantivet *data*. Ordet *datlig* har jag visserligen aldrig stött på, men också det är bildat efter känt mönster: *skriflig*, *statlig*, *vanlig* osv. Rad 2 har alltså utbyggnadsmöjligheter. För den tredje raden gäller att *datorbehandla* visserligen finns i senaste upplagan av Svenska Akademiens ordlista men enligt min erfarenhet används mycket lite (exempelvis finns inget belägg i tidnings-

materialet). Lagg märke till att verbet *datorisera* används i betydelsen 'utrusta med datorer'. För att uttrycka egenskapen tar man till en sammansättning av typen *datorbaserad*. *Datorisk* vore möjligt.

Vad speciellt *data* angår ser vi att det blir sammanfall mellan beteckningen för maskinen och den plurala beteckningen *data* för de uppgifter som bearbetas i maskinen. Detta är i och för sig inte eftersträvansvärt. Det bör likväl nämnas, att sammanhanget normalt avgör om man syftar på det ena eller det andra. I sådana fall där man vill ha garantier för att det uttryck man använder skall uppfattas som beteckning för apparaten kan man utnyttja den längre formen *datamaskin*.

I det här sammanhanget anmäler sig också frågan vad *Data* i *Dataordboken* betyder. Gör man en omskrivning av titeln, blir resultatet ungefär Ordbok över termer på databehandlingens område, vilket också svarar mot den engelska undertiteln *Data processing glossary*. Vi ser alltså hur *data* med utgångspunkt i *databehandling* har utvecklats till en ganska behändig allmän beteckning för området, använd också i fackspråket. Det är mot den bakgrunden man skall betrakta ord av typen *dataanläggning*, *datacentral*, *datateknik* osv., där *data* svarar mot *databehandlings-*. Det är nog också därför som *datorbehandla*, *datorcentral* osv. känns som alltför snäva beteckningar.

Man har nu börjat experimentera med olika system där våra tevear skall kunna användas som ett slags dataterminaler. Just den allmänna innebörden i *data* skulle vi kunna utnyttja här. Varför inte ge sådana system den samlande beteckningen *datateve* (*data-TV*)?

## Andra språk

Vi har också anledning att se ut över gränserna. I Norge dominerar *datamaskin*. Där har man emellertid debatterat ett förslag att gå över till *dator* under appellen "Norge inn i datoralderen". Inlägg till försvar för *datamaskin* har inte saknats. I debatten har också ett alternativ *dat* dragits in. Ett argument som har förts fram

är att *dat* är den minsta gemensamma nämnaren för den ordgrupp det gäller. Uppfattningarna i debatten har gått rätt starkt isär.

I Danmark används *datamaskine* men framför allt *datamat*. Detta ord är lika speciellt danskt som *dator* är speciellt svenskt.

Engelskan har sedan gammalt *computer*, och det ordet är nu på frammarsch i tyskan: *der Computer*. Där har man också *Datenverarbeitungsanlage*, *EDV-Anlage* och *Rechenanlage*. I franskan används *calculateur* och *ordinateur*.

### Variation i allmänspråket

I den situation som råder blir det viktigt att göra ett påpekande som egentligen är alldeles odramatiskt. Poängen ligger i att det faktiskt måste göras. Alltså: en viss sak behöver inte ha endast en beteckning i allmänspråket. Tvärtom är det ofta en tillgång att vi kan uttrycka (ungefär) samma sak på flera olika sätt. Allmänspråket har med andra ord inte samma behov av en enhetlig terminologi som fackspråket (som emellertid även det kan vara betjänt av synonymer). Men vi skall naturligtvis inte skapa onödiga skillnader. Här finns i själva verket utrymme för ett samarbete på flera fronter till fördel för språket och språkbrukarna.

I fråga om beteckningen för den aktuella apparaten ter det sig då så här. Det finns ingen anledning att motverka *datamaskin* i allmänspråket. Det internationellt sett isolerade och språkligt sett ganska oviga *dator* bör också i fackspråket få *datamaskin* vid sin sida. Det ordet med tillhörande ordgrupp är sedan länge väl-etablerat i språket och är det bästa i relation till danskan och norskan. Skulle *datamaskinen* och *datorn* i större utsträckning få dela plats med *datan* i allmänspråket, så bör vi kunna bära det också.

## **Sensmoraler**

Fackspråkets termer är inte alltid bra för allmänspråket – och inte för fackspråket heller.

Språkbrukares böjelse för schablonlösningar är inte alltid bra för språkbruket – och inte för språkbrukarna heller.

Svenska ord är inte alltid bra för norskan – och inte för svenskan heller.

## Terminologi i nordiskt samarbete

Språk i Norden 1978. Föredrag vid konferensen Nordterm 1978.

Kort sagt är min uppgift att ge synpunkter på samarbetet mellan olika organ i Norden som har att göra med frågor om terminologi. Jag har blivit ombedd att bland annat anlägga språkvetenskapliga synpunkter och utnyttja erfarenheter av samarbete på språkvårdsområdet i Sverige. Jag vill gärna säga att detta får göras med grova penseldrag.

### Allmänspråk och fackspråk

Konferensens ämne leder oss in på frågan om allmänspråk och fackspråk. Först måste då sägas att det inte är fråga om skarpt skilda språkarter. Man kan emellertid urskilja två olika tyngdpunkter. Ett sätt att karakterisera allmänspråket är att ange det som det på naturlig väg inlärd språk som man tillägnar sig genom att ansluta sig till en grupps språkpraxis (utan egentliga föreskrifter). Med den bestämningen kommer allmänspråket att innefatta barnspråket. Fackspråket å sin sida är ett mer eller mindre explicit normerat (auktoriserat) språk. För en del betraktare framstår fackspråket som ett slags degenererat allmänspråk, som har förlorat mycket av sin friskhet och uttrycksförmåga, som har stelnat i vissa former osv. Andra framhåller att det fullt utvecklade fackspråket är ett språk på högsta nivå. Det är det som används i forskning och utveckling och som gör det möjligt att överföra verklig insikt. I detta synsätt representerar allmänspråket den lägsta nivån. När det används i informerande syfte ökar risken för feltolkningar.

Ingendera skissen fångar naturligtvis det hela. Vi kan försöka lägga till några drag. Allmänspråket är äldst och utgör ännu idag grunden för fackspråket. Det är med andra ord stor överlappning mellan de båda. Allmänspråket är vidare det mest använda språket. Fackspråket å andra sidan är nödvändigt för många aktiviteter i samhället.

Allmänspråket kan vara vagt, vilket i många sammanhang är en styrka. Genom ömsesidig anpassning kan vi söka oss fram till ett bra förståelseläge i kommunikationen. Fackspråket måste vara mera exakt och helst omedelbart överföra den rätta informationen. Frågar man efter vägen till postkontoret, gör det oftast inte så mycket om man kommer dit några minuter senare än man tänkt sig. Leder å andra sidan en arbetsorder till att man sätter en ventil upp och ner, kan resultatet bli en utblåsning.

## Terminologi

Låt oss så titta närmare på terminologiarbetets plats i lingvistisk verksamhet. Figur 1 är en grov och bara delvis utförd illustration. De två huvudkolumnerna representerar språksystemet och språkbruket. Språksystemet kan man tänka sig som den uppsättning av lexikaliska enheter med tillhörande karakteristika och grammatiska regler av olika slag som krävs för att man skall kunna producera språkliga yttranden. Sett från den ideala språkbrukarens synpunkt kan detta ungefärligen kallas lingvistisk kompetens. Språkbruket (endast antytt genom streckning) representerar användningen av systemet. Här kan man tala om i vid mening stilistisk kompetens. I synnerhet är detta motiverat om man har funktionsstilistikens bestämning av begreppet stil i tankarna: stil är det sätt varpå man löser problemet att i en given situation ge sina ord avsedd effekt. Här skulle då de pragmatiska och sociolingvistiska aspekterna kunna sägas vara inbegripna.

För lexikonet, som antas innehålla information om de lexikaliska enheternas (ordens, ordförbindelsernas, ordledens) form, funktion och innehåll, har jag angivit distinktionen spontant: normerat. Samma distinktion är gjord för grammatiken (som inkluderar reglerna för texters uppbyggnad – textdynamik).

De två korsande raderna gäller allmänspråk respektive fackspråk med de reservationer som jag tidigare har gjort.

Jag inbjuder betraktaren att tänka ut fall som passar i de tomma rutorna som uppkommer. Som exempel på spontana lexikaliska enheter i fackspråk kan jag ge snickarnas beteckningar

på olika företeelser: hammarband, regel, röste, syll, vindskiva osv. Den terminologi som är konferensens ämne är den normerade aspekten på fackspråkets lexikon (rutan T). Om någon tycker att rutan verkar liten, vill jag gärna replikera att varje ruta här i själva verket är stor. Språket är ett oerhört rikt och mångfasetterat redskap.

	SPRÅKSYSTEM				SPRÅKBRUK
	LEXIKON		GRAMMATIK		
	Spon- tant	Nor- merat	Spon- tant	Nor- merat	
Allmän- språk					
Fack- språk		T			

FIGUR 1. Terminologins språkliga område.

Det bör understrykas att den distinktion jag har illustrerat i fråga om språkssystemet gäller vad som är spontant i förhållande till vad som är normerat. En annan viktig distinktion där normering är ena ledet har deskription eller beskrivning i andra ledet. Deskription är en viktig fas i ett arbete som syftar till normering.

Eftersom sålunda terminologiskt arbete från språkvetenskaplig synpunkt sett är en del av lexikografin, krävs av dess utövare lingvistisk skolning eller tillgång till lingvistisk expertis. Detta i förening med bland annat den stora överlappningen mellan allmänspråk och fackspråk gör ett samarbete mellan olika organ inom ett land mycket angeläget. För undvikande av dubbelarbete och för undvikande av missförstånd på grund av onödiga avvikelser i betydelse och form, för stärkande av det kulturella sambandet och av andra skäl är vidare ett nordiskt samarbete mycket angeläget. Av delvis samma skäl är det också viktigt med ett vidare internationellt samarbete.



## Organ

Vilka är de organ som arbetar med terminologi? En viktig grupp är naturligtvis de olika centralerna för teknisk terminologi. Det rör sig här om Rådet for teknisk terminologi i Norge, Centralen för teknisk terminologi i Finland och Tekniska nomenklaturcentralen i Sverige. Också den isländska ingenjörsföreningen med dess mångåriga arbete på områden inom teknisk terminologi måste nämnas här. Över huvud taget vill jag i detta sammanhang framhålla vikten av att inte bara centralnordiska organ beaktas.

En annan grupp bildar de nationella nordiska språknämnderna: Norsk språkråd, Dansk Sprognævn, Sverigefinska språknämnden osv. Språknämndernas arbete med hithörande frågor betingas främst av överlappningen mellan allmänspråk och fackspråk och det därmed sammanhängande flödet av termer från fackspråk till allmänspråk. Om en svensk medicinsk språknämnd kommer till stånd enligt planerna, blir den givetvis betydelsefull också från termsynpunkt.

En tredje grupp bildas av ländernas organ för standardisering. Hit hör Norges Standardiseringsforbund, SIS – Standardiseringskommissionen i Sverige m.fl. Fastläggandet av termer är i detta fall ett led i standardiseringen i vidare mening.

Det finns också en grupp ämbetsverk som måste nämnas i detta sammanhang. Ett norskt exempel är Forsvarets felles materielltjeneste, som utnyttjar Norsk språkråds tjänster när det gäller genomgång av termförslag. För Sveriges del kan bland annat nämnas statens invandrarverk, som främst har behov av sociala och administrativa termer.

Begreppet terminologiorgan är alltså inte skarpt avgränsat. I själva verket finns det ytterligare ett antal organ av olika typer som bör delta i ett samarbete. Inte minst gäller detta institutioner och institut som är verksamma inom datalingvistik och tillämpad lingvistik. Sådana är Prosjekt for datamaskinell språkbehandling (PDS) i Bergen, Institut for anvendt og matematisk lingvistik (IAML) i Köpenhamn, och Institutionen för språklig databehandling (Språkdata) i Göteborg. Skälen är flera. Dessa insti-

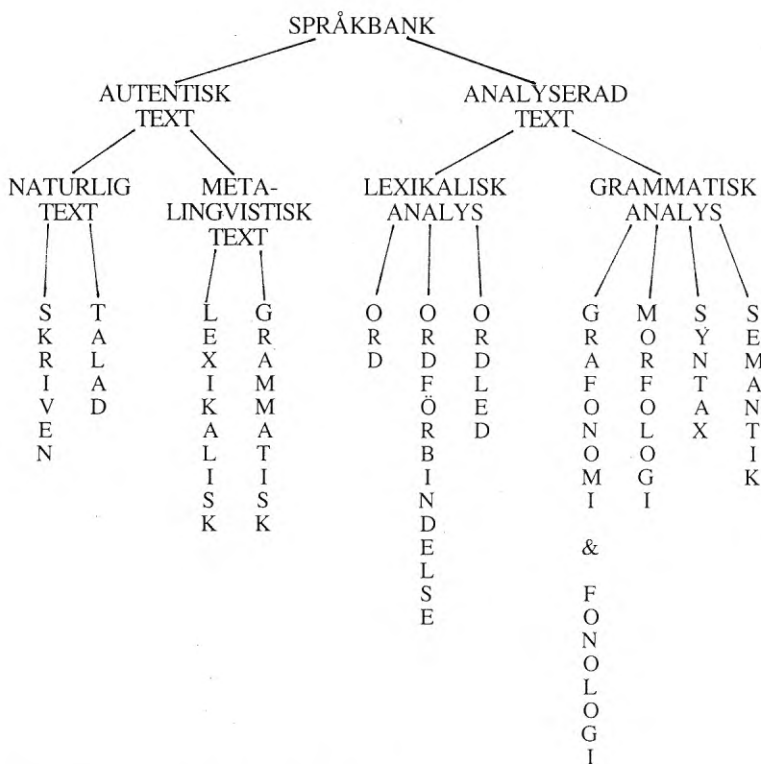
tutioner har lång erfarenhet av här aktuell teknik och metodik. De kan bland annat också ge möjligheter att undersöka hur terminologiska rekommendationer slår i allmänspråket. I synnerhet gäller detta naturligtvis i den mån språkbanker byggs upp. Som exempel kan jag nämna att Logoteket vid Språkdata har en textbank som nu närmar sig tjugo miljoner löpord och en ordbank som innehåller omkring 200 000 ord, vartill kommer omkring 300 000 personnamn. Vid institutionen drivs också ett forskningsprojekt som bygger upp en omfattande lexikalisk databas.

En språkbank kan se ut som figur 2 visar. Den innehåller alltså autentisk (obearbetad) och analyserad text. Den autentiska texten kan vara naturlig text eller metalingvistisk text, det vill i senare fallet säga ha formen av lexikon eller grammatik. Den analyserade texten kan ha underkastats lexikalisk eller grammatisk analys med dessa båda kategoriers underavdelningar. En fråga blir var en termbank hör hemma i denna strukturskiss. Den har sin plats i den lexikaliska slutpunkt som man kommer till via noderna autentisk text och metalingvistisk text. I detta övergripande perspektiv är en termbank sålunda en delmängd av en språkbank. Detta hindrar givetvis inte att den fysiskt sett kan vara en självständig delmängd.

Också sådana organ som arbetar med dokumentation och klassifikation har en plats i ett vidare samarbete. Jag tänker på organ av typen Danmarks Tekniske Bibliotek. Bland annat är tesaurusbegreppet en förmedlande länk.

Härutöver finns ett antal samarbetsgrupper av olika slag. Som exempel på nationella sådana kan jag nämna den danska terminologigruppen och den svenska arbetsgruppen för språkvårds-samverkan. Som exempel på en nordisk samarbetsgrupp kan jag nämna den arbetsgrupp som behandlar datanätsfrågan.

Till de nu nämnda organen kommer så Nordiska språksekretariatet, som börjar sin verksamhet den 1 juli 1978 och därmed inleder en ny fas i det nordiska samarbetet på det språkliga området.



FIGUR 2. En språkbanks struktur.

Vi får sålunda åtminstone nio olika typer av organ, som på olika sätt borde ingå i ett terminologiskt samarbete. De är sammanfattningsvis följande.

- Centraler för teknisk terminologi
- Språknämnder
- Standardiseringsorgan
- Vissa ämbetsverk
- Datalogvistiska organ
- Dokumentations- och klassifikationsorgan
- Nationella samarbetsgrupper
- Nordiska samarbetsgrupper
- Nordiska språksekretariatet

## Samarbete

Allmänt sett kan man urskilja minst fem grundläggande samarbetsuppgifter. Centralt är naturligtvis att normera termer med avseende på i första hand innehåll och i andra hand form. Vidare är det betydelsefullt att efterhand försöka arbeta sig in på nya fackområden av gemensamt intresse. En uppgift med ökande vikt är den att bygga upp kompatibla termbankar, exempelvis knutna till ett datanät. Hit hör bland annat frågan om i vad mån Mater är lämpligt som format för utbyte av terminologisk (och lexikografisk) information. En fjärde uppgift är att tillvarata allmänspråkets intresse i fråga om termer och i samband därmed studera termrekommendationernas öde i allmänspråket. Som en femte samarbetspunkt vill jag framhålla utvecklingen av internationella kontakter.

Min syn på samarbetsformerna är denna. Grundläggande är att varje enskilt organ skall kunna ha ett fritt och obundet samarbete med varje annat organ. Detta ger principiellt ett mycket tätt nät av förbindelselinjer. Om man så vill, skall man också kunna kanalisera samarbete via nationella eller nordiska samarbetsgrupper i den mån sådana finns.

Återstår Nordiska språksekretariatets roll. Jag har ställt samman de avsnitt i olika dokument som närmast berör nordiskt terminologiskt samarbete. I betänkandet Nordiskt språksekretariat (framlagt av Arbetsgruppen för utredning av språkvårdssamarbetet i Norden, 1977) heter det (s. 27):

Nordiskt språksamarbete borde bl.a. syfta till att så långt det är möjligt ge nya benämningar en likartad gestalt i de nordiska språken, och då det rör sig om fasta termer försöka åstadkomma definitioner med samma innehåll inom de olika språken.

Punkterna 4 och 6 i Riktlinjer för Nordiska språksekretariatet lyder:

4. Språksekretariatet skall kontinuerligt följa behovet av och upplysa om värdet av språk- och termsamarbete inom de nordiska länderna.

6. Språksekretariatet skall också undersöka och kontinuerligt följa det nordiska samarbete med språklig anknytning som förekommer eller etableras utanför språksekretariatet. Språksekretariatet bör kunna fungera som en central informationsinstans för alla former av nordiskt språksamarbete.

De två första momenten i paragraf 2 i Instruktion för Nordiska språksekretariatet har följande lydelse:

Språksekretariatet har särskilt till uppgift att

- a) vara rådgivande och koordinerande organ i språk- och termfrågor som rör mer än ett nordiskt språk,
- b) ta initiativ till utredningar och forskningsprojekt av gemensamt nordiskt intresse rörande t.ex. språkvård, terminologi, nordisk språkförståelse och grannspråksundervisning samt verka för att dessa frågor uppmärksammas vid universitetsutbildningen.

Jag tycker mig kunna strukturera dessa uppgifter ungefär så här. När det gäller termfrågor på nordiskt plan har Nordiska språksekretariatet till uppgift att

- (1) fungera som central informationsinstans
- (2) vara rådgivande och koordinerande
- (3) ta initiativ till utredningar och forskningsprojekt
- (4) verka för att frågorna uppmärksammas i universitetsutbildningen
- (5) upplysa om värdet av samarbete.

Att vara central informationsinstans bör innebära att samla in och sända ut information. Det kan gälla översikter, nyheter osv. Det blir viktigt att berörda organ kontinuerligt håller språksekretariatet underrättat om vad som sker, t.ex. genom att översända kopior av relevanta handlingar.

Den rådgivning som språksekretariatet kan komma att bedriva bör givetvis basera sig på uppgifterna i auktoriserade termlistor. Ett exempel på koordinerande verksamhet ger jag i kommentaren till nästa punkt. Till koordineringen räknar jag också förmedling av kontakter. Man skall kunna vända sig till språksekretariatet för att få veta vem man fördelaktigast skall vända sig till.

Initiativ till utredningar och forskningsprojekt bör väl främst gälla nya områden (arbetsmarknad, administration, utbildning, socialvård osv.). Det kan naturligtvis också gälla nya tekniska förfaranden. Man kan tänka sig att ett initiativ får den formen att språksekretariatet kallar en grupp fackexperter till en nordisk konferens. Där kan man tillsätta en kommitté som samarbetar med nationella organ på lämpligt sätt. Språksekretariatet kan koordinera samarbetet. Flera andra modeller kan tänkas, och initiativ kan som jag tidigare har sagt tas från vilket organ som helst. Ett samarbete är till stor del beroende av gemensamma nordiska anslag. När det gäller att utverka sådana medel bör språksekretariatet ha vissa möjligheter.

I fråga om universitetsutbildningen finns det plats för initiativ från språksekretariatets sida. Sekretariatet bör därvid dra nytta av erfarenheterna från den terminologiska utbildning som äger rum vid Handelshøjskolen i Köpenhamn. I detta sammanhang har jag också anledning att nämna den terminologikurs som anordnas inom ramen för Nordiska forskarkurser 1978.

Uppgiften att upplysa om värdet av samarbete ser jag som ett led i stärkandet av den nordiska kulturella samhörigheten.

Detta är viktiga och omfattande uppgifter. Det är uppenbart att språksekretariatet snarast behöver stöd från de tillkommande medarbetare som enligt planerna skall placeras i anslutning till de nationella språknämnderna. Vi skall också komma ihåg, att vad jag här har talat om rör terminologifrågorna. Bland annat tillkommer de allmänspråkliga och de s.k. språkpolitiska uppgifterna.

Låt mig till sist anknyta till språkvårdaren Erik Wellander. När han skulle sammanfatta sina rekommendationer, gjorde han det som bekant i form av fyra uppmaningar:

Skriv klart!

Skriv enkelt!

Skriv kort!

Skriv svenska!

För samarbetet i fråga om termer skulle jag vilja tänka mig tre imperativer:

Tänk fackspråkligt!

Tänk allmänspråkligt!

Tänk nordiskt!

## Litteratur

Allén, Sture, Den svenska språkvårdens organisation. (Språkvård 1975: 4, s. 7–11.)

Fackspråk. En antologi red. av Bertil Molde. (Skrifter utg. av Svenska språknämnden. 57. 1976.)

Grünbaum, Catharina, Ett nordiskt språksekretariat. (Språk i Norden 1977, s. 23–31.)

Nordiskt språksekretariat. Betänkande avgivet av arbetsgruppen för utredning av språkvårdssamarbetet i Norden. (Nordisk utredningsserie A 1977: 1.)

Nordterm 1976. Terminologiskt samarbete i Norden. (Tekniska nomenklaturcentralen. 1976.)

Petöfi, János S., & Podlech, Adalbert, & von Savigny, Eike (Hrsg.), Fachsprache–Umgangssprache. Kronberg 1975.

Schwarz, Stephan, Om fikonspråkets nödvändighet. (Språkvård 1975: 1, s. 4–8.)

The Termdok Bulletin 34. (Tekniska nomenklaturcentralen. 1977.)

## Text-Based Lexicography and Algorithmic Text Analysis

Association for Literary and Linguistic Computing Bulletin 5: 2. 1977.

This paper reports on work done in cooperation with Sture Berg, Mats Eeg-Olofsson, Rolf Gavare, Martin Gellerstam, Staffan Hellberg, Jerker Järborg, Jonas Löfström, Bo Ralph, Christian Sjögreen, and Staffan Wåhlin at the Department of Computational Linguistics (Språkdata), University of Göteborg.

My purpose is to take up some aspects of the coordinated work of three groups in the Department. The first of these is the Research Group for Modern Swedish, which carries out the work underlying the volumes of the *Frequency Dictionary of Present-Day Swedish*. The second is the group developing the Swedish Logotheque, a text and word bank. The third is the group working on the project Algorithmic Text Analysis.

The general framework for the projects is the following: analysis of authentic text on a theoretical basis in order to allocate the instances of a range of linguistic units to their appropriate places in a data bank. The term "linguistic units" refers to lexically and grammatically (including semantically) relevant phenomena, such as word collocations, word elements, word-classes, grammatical types (morphological and syntactic), principal senses of words, etc.

The framework just outlined offers one way of establishing that the computer, or rather the cerebrally developed program system, has attained a certain ability with respect to language decoding. The power of the framework may seem modest, as compared with the goals of undertakings like paraphrasing, conversation, and translation. The actual level of achievement, however, depends on the sophistication of the data bank. In fact this bank can be designed to meet requirements of ever-increasing complexity.

A basic idea is that theoretical considerations and data observations are dependent on each other. When a linguistic model, theoretically founded, is applied to a corpus of authentic text, a



considerable number of problems inevitably present themselves. Works of reference naturally tend to disregard many of these things. Studies of special issues, furthermore, often presuppose that the solutions needed in the particular case are in fact at hand, even if they are not (which is of course, as a matter of principle, justifiable). Thus the linguistic model has to be gradually modified, as new insights are gained from language data and from linguistic theory. This might be called the reciprocity or spiral approach.

### Text-Based Lexical Investigations

The dictionary project aims at a qualitative and quantitative lexical analysis (in a wide sense) of a corpus of one million running words from five morning papers, representing Present-Day Standard Swedish in its written form. The general strategy developed for the investigation is shown in the scheme on the next page, which includes some very rough indications of criteria.

The phases "segmentation" – "classification" are based on formal criteria, combined to a certain extent with a negative semantic criterion (semantic incompatibility). The necessity of a semantic bar is evident when commutation is used. If no such criterion is established, the segments arrived at will be the graphemes. Ideally, the semantic bar should approximate the intuition of the speakers. An example: in *tjänst* 'service', *vinst* 'winnings', etc. there is a suffix *-st* (cf. *tjäna* 'to serve', *vinna* 'to win', etc.); the semantic bar prevents a corresponding analysis of *könst* 'art' based on the word *kon* 'cone'.

In the phase "differentiation" positive semantic criteria are applied: the result is a division into principal senses. Because of the structure of the words (compounding, derivation), this phase is best carried out on word elements (disregarding, for the moment, idioms). Residual morphemes as well as certain affixes are concatenated before differentiation. This also applies to the compound type *smörgås*, which means 'sandwich', although (*smör* means 'butter' and) *gås* means 'goose'.

<i>Main Phases</i>		<i>Descriptive Levels, Criteria</i>		
		(1) Words	(2) Collocations	(3) Word Elements
(I)	Segmentation	<i>Graphic Words</i> Syntagmatic Contrast	<i>Combinations</i> Recurrence	<i>Graphic Morphs</i> Commutation
(II)	Separation	<i>Homograph</i> <i>Components</i> Syntax, Inflection	<i>Constructions</i>  Syntax	<i>Homomorph</i> <i>Components</i> Morphotactics, Phonology
(III)	Classification	<i>Lemmas</i>  Syntax Inflection	–	<i>Formal</i> <i>Morphemes</i> Morphotactics, Phonology
(IV)	Differentiation	–	<i>Idioms</i>  Semantics	<i>Polysemy</i> <i>Components</i> Semantics

### *Strategy of the dictionary project.*

The units obtained by the classification of words and word elements do not necessarily correspond to the common conception of the linguistic sign. The lemma *golf*, for instance, comprises both the instances meaning 'gulf' and those meaning 'golf' (the inflection being the same).

Strictly speaking, separation contains a kind of classification inasmuch as instances of the same unit are cumulated. As regards collocations, the absence of a classification in a proper sense is due to the fact that the units make up strings of words, a varying number of which are inflected. Beyond the above scheme, words and collocations are also classified according to the abstract linguistic categories "word-classes" (as well as inflectional classes) and "grammatical types", respectively.

Connected with the phases "separation" and "differentiation" of collocations, there is a selection based on grammaticality and semantic unpredictability, respectively. A certain selection takes place in the treatment of word elements, too: abbreviations, proper names, and non-Swedish units are not analysed.

Data on the various levels – as well as more detailed accounts of principles and criteria – are presented in the *Frequency Dictionary of Present-Day Swedish* according to the following plan: Part 1 (1970): graphic words, homograph components; Part 2 (1971): lemmas; Part 3 (1975): combinations, constructions, idioms; Part 4 (in preparation): graphic morphs, homomorph components, formal morphemes, polysemy components. In addition, the central vocabulary is published in *Tiotusen i topp (Top Ten Thousand)* (1972). The work on Part 4 is carried out in an interactive mode at display terminals connected to the computer of the Department. An adaptive program is employed.

The Swedish Logotheque is, since 1975, a regular service agency on a national basis. Its purpose is to collect and store machine-readable texts (written and spoken) and processed material based on them, i.e. to build up a Swedish language bank, and to furnish users with texts and other data from the bank. This means that problems concerning input, storing, and output of text in a large character set and automatic and interactive text processing are of particular importance.

Texts are collected continuously according to a plan based on the annual print in Swedish. The intake for 1976 is two million running words. A gradual increase is aimed at. The words of the texts are stored in a graphic-word bank. A lemma bank contains a more elaborated selection. For further information see my paper "The Swedish Logotheque. A Computer-Based Text and Word Bank" (to appear in *Festschrift F. de Tollenaere*).

The goal of the project Algorithmic Text Analysis is the development and application of formalized procedures for, in the first place, morphological and syntactic analysis of authentic text. An obvious way of developing such a system is to draw on the knowledge gained from the text-based lexical studies.

One problem in text analysis, seemingly trivial but in practice tricky enough, is the normalization of the words in the text, for lexical and look-up purposes, with respect to the occurrence of capital letters and different kinds of junctures and special characters. This represents the step from textual words to graphic words. Among the difficulties encountered are those caused by

the relation between capital letters and proper names, the relation between junctures and abbreviations, and the ambiguity of hyphens. The mapping of the area by the dictionary group and the Logothèque as well as the elaboration of principled solutions, is of great value. Needless to say, the information provided by capital letters and junctures is kept for use in the overall analysis.

### The Dictionary in Text Analysis

At word levels there are several ways of making use of lexical results. In the million-word corpus there are 100 000 different graphic words manifesting 70 000 lemmas. On the basis of observed frequency and distribution, a basic vocabulary of 10 000 lemmas has been defined. This list covers 90 % of the corpus. It is now being used as the dictionary basis in the text-analysis project.

Word homography is a very characteristic feature of Swedish text. Even if it is defined in formal syntactic and inflectional terms, about two thirds of the running words in a text meet the criteria. This figure applies, however, to internal homography (i. e. between different forms of a lemma) as well as to external homography (i. e. between forms of different lemmas). When only external homography is considered, the proportion is very much reduced. The *Dictionary of Swedish Homographs* by Sture Berg, which is now in print, contains 9800 external homographs, including 4200 forms in *-s* (genitive ending and passive ending, respectively). The dictionary is based on an investigation of all the forms of more than 200 000 lexical words.

From the point of view of text analysis, heterographs are of course very attractive. The homography-heterography complex is, however, a typical scale phenomenon. There are clear heterographs and clear homographs and in between quite a number of more or less (semantically) plausible or (statistically) probable homographs. Examples: *nattliga* is the definite form of an adjective meaning 'nocturnal', but it can also be interpreted as a noun compound meaning 'gang operating in the night'; *kan* and *är*

are the auxiliaries 'can' and 'is, are' but there is a remote possibility of finding them in the senses of 'khan' and 'countenance, air', respectively. In text analysis near-heterographs as well as absolute heterographs should be used. It is important, however, to keep them distinguished from each other.

In the illustration on the next page – the beginning of a newspaper article chosen at random – the absolute heterographs are indicated by double broken lines (a literal "Swenglish" rendering of the words is given). The near-heterographs are marked by single broken lines. The words underlined cover a considerable proportion of the text specimen.

Heavy compounding and a great deal of inflection are characteristic features of Swedish. The handling of these problems in the text-analysis system has required the working out of a morphological grammar, containing well over 200 paradigms. This work has been based on the results of the lemmatization and the morphological analysis of the dictionary project. As regards the machine implementation, the starting-point was provided by Martin Kay's algorithm for nondeterministic analysis (in *Morphological and Syntactic Analysis*, 1974).

Data concerning the occurrence of inflectional forms in the corpus of the dictionary study are used in lexical searching strategies. Among the main noun paradigms, for instance, the indefinite singular form covers about 40 % of the instances, whereas the definite plural form covers about 10 %. The ratio between the four singular forms and the four plural forms is roughly 2:1.

The current phase of the dictionary project shows that the whole corpus is generated from a set of 8000 formal morphemes (exclusive of proper names), about 550 of which are affixes. The forms and morphotactic features of these units are naturally of great interest to a text-analysis system.

Not least the result of the lexicographic investigation of collocations seems to be useful in text analysis. This investigation was based on the idea of recurrence. Units that belong to a language system may be expected to turn up over and over again in manifestations of that language. Collocations having at least two identical instances in the million-word corpus were collected by the

När jag nu kommer till en främmande stad,  
 .....  
 When I now arrive at an unfamiliar town,

och det behöver inte alls vara någon utländsk,  
 .....  
 and it need not at-all be some foreign,

Vill jag ogärna gå hem till folk.  
 want I reluctantly to-go home to people.

Det beror inte på att jag inte vill träffa folk.  
 .....  
 It is-due not to that I not want to-meet people.

Men om jag kommer hem till människor,  
 .....  
 But if I come home to people,

tycker jag det känns ännu mera ledsamt att vara hemifrån.  
 .....  
 think I it feels still more depressing to be from-home.

*Collocations and heterographs.*

computer. They are the combinations. The list of combinations comprises about 600 000 instances (inclusive of included and overlapping collocations).

If the recurrent combinations are applied to the text, they cover it to a very large extent. See the illustration above where these combinations are marked above the line. Notice the wave contour of the markings, which approximates, in an intriguing way, the clause structure. In passing, let me also point out the power of the recurrence wave as a style marker and as a text-attribution criterion.

The combinations obtained are of different kinds. Those which met a set of grammatical (phraseological) criteria were listed as constructions (it should be observed that nexus strings were not included). We got 50 000 constructions representing 200 000 instances. Those occurring in the illustrative example are marked below the line. It will be seen that a large proportion of the text is covered and that verb phrases are often included.

The beginning of a sentence is very often formed by the members of a small group of combinations. There are 600 combinations occurring 10 times or more in this position. About every fourth sentence is introduced by one of these units. The instances occurring in the text example are marked by dotted lines. The collocation *det är/var* 'it is/was' alone opens every twentyfifth sentence.

Taking words in addition to stems into account, and collocations in addition to words, means that the use of a wider context is emphasized. It is a matter of course that, in principle, words are less ambiguous than stems and collocations less ambiguous than words. The trend is thus from homography towards heterography.

The collocational approach seems to be in accordance with linguistic intuition. It probably reflects at least part of the sentence planning of the sender and the expectations of the receiver in the communicative situation.

The lexicographic results sketched here, and partly illustrated in the example given, suggest that the dictionary of a system for algorithmic text analysis should contain, in addition to data on

word stems, information about heterographs, such as *kommer* 'arrive', *och* 'and', *behöver* 'need', *inte* 'not', *någon* 'some' etc., in the illustration; near-heterographs, such as *jag* 'I', *nu* 'now', *en* 'a', *alls* 'at all', etc.; frequent disambiguating collocations, such as *behöver inte (alls) vara* 'need not (at all) be' (solving the noun/verb homograph *vara*), *gå hem (till)* 'to go home to' (solving the noun/adverb homograph *hem*), etc.; and frequent sentence-initial collocations, such as *det beror* 'it is due' (solving the pronoun/article homograph *det*), *men om* 'but if, but about' (solving the noun/conjunction homograph *men*), etc. A more extensive use of frequent collocations (chiefly constructions and nexus combinations) seems also to be profitable. The text-analysis project is working along these lines.

In the procedure for syntactic analysis the use of relevant types of recurrent collocations is, as a matter of fact, the most important check of the internal structure of tentative surface constituents. Furthermore, collocations recurrent in a particular text under analysis, can be very informative. It is possible to obtain them nicely if the text is stored as a linked network (a data-base). Such a format is currently being considered by the Logotheque.

### Some References

- Allén, Sture, "On Phraseology in Lexicology." (To appear in *Cahiers de lexicologie*.)
- Allén, Sture, "The Swedish Logotheque. A Computer-Based Text and Word Bank." (In *Festschrift F. de Tollenaere* 1977.)
- Allén, Sture, *Tiotusen i topp (Top Ten Thousand)*. Stockholm: Almqvist & Wiksell 1972. (Data linguistica. Ed. S. Allén. 6.)
- Allén, Sture, "Travaux du Centre de recherches pour le suédois contemporain de Göteborg." (*Linguistica matematica e calcolatori*. Firenze: Leo S. Olschki 1973, pp. 3–18.)
- Allén, Sture, "Vocabulary Data Processing." (*The Nordic Languages and Modern Linguistics*. Ed. H. Benediktsson. Reykjavík 1970, pp. 235–261.)



- Allén, Sture, et al., *Frequency Dictionary of Present-Day Swedish*. 1–3. Stockholm: Almqvist & Wiksell 1970–75. (Data linguistica. Ed. S. Allén 1, 4, 10.)
- Berg, Sture, *Dictionary of Swedish Homographs*. (In print.)
- Hellberg, Staffan, "Computerized Lemmatization without the Use of a Dictionary: A Case Study from Swedish Lexicology." (*Computers and the Humanities* 1972, pp. 209–212.)
- Kay, Martin, *Morphological and Syntactic Analysis*. 1974. (Dupl.)
- Språkdata Research Report*. (Issued annually.)

## Morf-o-log-i och poly-sem-i. Om NFO4-gruppens forskningsarbete

Nysvenska studier 57. 1977.

Lingvistiska undersökningar är nästan alltid detaljstudier i den meningen, att de behandlar ett fragment av eller en aspekt på ett språkligt system. Det behöver egentligen inte sägas att denna typ av forskning är nödvändig.

Det är emellertid också nödvändigt att ta samlande grepp, att undersöka stora material som i väsentliga avseenden representerar systemet som helhet. På grund av språkets utomordentliga variationsrikedom ställer detta särskilda krav på teori och metod. Omvänt möjliggör den överblick och den materialkänedom som efterhand vinnes just en utveckling av teori och metod. Båda erfarenheterna har gjorts i det forskningsarbete som ligger till grund för de olika delarna av Nusvensk frekvensordbok baserad på tidningstext (NFO). Det gäller inte minst den undersökning av ordleden och deras betydelser som utgör den sista fasen i denna i vid mening lexikaliska undersökning med kvalitativ-kvantitativ inriktning.

Ordledsundersökningen har utförts av en forskningsgrupp bestående av Sture Berg, Jerker Järborg, Jonas Löfström, Bo Ralph, Christian Sjögreen (programmering) och mig själv. Det är gruppens arbete som avspeglas i detta föredrag.

Av tidigare morfologiska arbeten skall jag här nämna två. Det ena är den preliminära morfemsegmentering av orden i nionde upplagan av Svenska Akademiens ordlista, som utfördes inom forskningsgruppen KVAL vid mitten av 1960-talet. Resultatet har aldrig publicerats. Dock föreligger en datorutskrift, som vi emellertid inte har konsulterat annat än i enstaka fall. Detta hänger samman med att den som nämnt har en preliminär karaktär, att den inte är kombinerad med någon redogörelse för principerna, att den innehåller rätt många fel (avvikelser från ordlistan) och att den blev tillgänglig för oss först i slutskedet.

Ett av de få publicerade arbeten som behandlar ett material av större omfattning är Magnus Ljungs A frequency dictionary of English morphemes (1974). Korpusen omfattar i detta fall de 8000 vanliga orden i B. Thoréns 8000 ord för 8 års engelska (1959). Ljung för i inledningen en intressant diskussion om principerna för analysen. Hans material och syfte leder till en annan uppläggning av undersökningen än den vi har följt.

### **Allmänt om metoden**

På tillvägagångssättet vid projektarbetet kan flera allmänna synpunkter läggas. Viktigt är att uppmärksamma relationen mellan undersökningens utförande och presentationen av resultaten. Det råder självfallet ett nära samband mellan de båda, men medan presentationen av resultaten sker i en genomarbetad form, beskrivningsnivå för beskrivningsnivå med angivande av tillämpade kriterier o.s.v., får utförandet av undersökningen en tentativt framskridande karaktär. Utgångspunkterna på den lingvistiska sidan är den teoretiska bakgrunden, metodidéerna, den praktiska erfarenheten och det uppställda målet. Teori- och metoddiskussioner varvas med tillämpningar på materialet, vilkas resultat successivt kumuleras. Efterhand klarnar konturerna tillräckligt för att uppläggningsens huvudlinjer skall kunna anges. Vid genomförandet uppkommer likväl hela tiden nya problem som kräver modifikation av modellen. Kriterier som tillämpas vid en viss genomarbetning preciseras inför nästa och kombineras med tilläggs-kriterier och tillämpningsregler. Denna växelverkan mellan å ena sidan teori och metod och å andra sidan material kan betecknas som ett slags forskningsspiral.

Det datatekniska utförandet skiljer sig också från presentationen. Ibland kan flera moment på angränsande beskrivningsnivåer genomföras samtidigt, ibland är det fråga om ett delmoment på en nivå. Den maskinella analysen kan inte sällan med fördel baseras på grövre kriterier än den slutgiltiga. Interaktiv kontroll och modifikation spelar därför en viktig roll. Automatiska kon-

troller baserade på rent formella ting såsom format och ordningsföljd blir likaså en betydelsefull del av det hela.

Det skall också göras klart att den undersökning vi har utfört är synkronisk till sin grundläggande natur. Detta innebär att vi har låtit aktuella relationer mellan ordled slå igenom även i sådana fall, där vår kännedom om språkets historia i och för sig hade kunnat låta andra relationer blottläggas. Vi ser exempelvis ett så nära formellt och innehållsligt samband mellan *front* (ytterst av latinets *frons* 'panna') och *frondera* (ytterst av latinets *funda* 'slunga'), att de får bilda en sammanhållen enhet *frond* med varianterna *frond/front* på högre beskrivningsnivåer. Å andra sidan hålls *god* och *göd* (ur *gödkalv*, *gödset*) isär på grund av det numera stora semantiska avståndet och trots det gemensamma ursprunget.

Den övergripande strategin i NFO-projektet kan sammanfattas i uppställningen på s. 383 (se för en närmare kommentar min uppsats *Text-based lexicography and algorithmic text analysis*, 1977).

För ordledens vidkommande rör det sig om en analys som går från huvudsakligen på formella grunder framtagna grafiska morfer till semantiskt baserade polysemkomponenter. Resultaten på de fyra berörda beskrivningsnivåerna kommer att redovisas i NFO 4. Ordförbindelserna (kombinationer, konstruktioner, idiom) har redovisats i NFO 3 och orden i NFO 1 (graford, homogرافkomponenter) och NFO 2 (lemman).

## Grafiska morfer

Som material för undersökningen har vi valt lemmanas uppslagsformer i NFO 2 med vissa uteslutningar. De uteslutna ordgrupperna är huvudsakligen proprier, förkortningar, utländska enheter och med siffror skrivna räkneord. Av de 71 000 lemmarna återstod efter denna reduktion 55 000.

Ett ord, nämligen ett lemmas uppslagsform, antas i enlighet med gängse uppfattning vara uppbyggt av ett litet antal typer av ordled. De ledtyper vi räknar med är prefix, grundmorf, fog,

ordbildningsuffix och i vissa fall ordböjningsändelse (som i *flerveckorsskägg*). Segmenteringen går nu ut på att finna gränser mellan sådana element. Detta sker med kommutation som främsta metodologiska hjälpmedel. Ett litet axplock ur mängden av belägg som rör *dom* får tjäna som exempel: *chefs-dom-are*, *dom-edagsfigur*, *fransk-dom-inerad*, *skogs-dom-än*, *Nidaros-dom*, *lär-dom-saptit*, *folksjuk-dom*, *ung-dom-lig*. Lägg märke till att pronomenformen *dom* inte är en lemmatisk uppslagsform och därför inte ingår i analysmaterialet.

Huvudmoment	Beskrivningsnivåer, huvudkriterier		
	(1) Ord	(2) Ordförbindelser	(3) Ordled
(I) Segmentering	<i>Graford</i> Syntagmatisk kontrast	<i>Kombinationer</i> Rekurrens	<i>Grafiska morfer</i> Kommutation
(II) Separering	<i>Homografkomponenter</i> Syntax, böjning	<i>Konstruktioner</i> Syntax	<i>Homomorffkomponenter</i> Morfotax, fonologi
(III) Klassificering	<i>Lemman</i> Syntax, böjning		<i>Formella morfer</i> Morfotax, fonologi, grafonomi
(IV) Differentiering	–	<i>Idiom</i> Semantik	<i>Polysemkomponenter</i> Semantik

Kommutationen kan emellertid inte släppas helt fri – resultatet skulle i så fall bli språkets uppsättning av grafem. Två kontrollfaktorer har därför införts. Den ena är vad vi kallar en semantisk spärr, ett negativt kriterium av mycket grov karaktär som avser att hindra att alltför våldsamma semantiska klyftor överbryggas. Ett par exempel är *fänrik* och *kondor*, som hålls samman trots *fän(kål)* och *rik* respektive *kon* och *dor(isk)*.

Den andra kontrollfaktorn tar fasta på att ett ordled normalt utgörs av en stavelse. Undantag åt ena hållet är fall som de båda

nyssnämnda *fänrik* och *kondor*. För undantag åt andra hållet, d.v.s. under stavelsens omfång, krävs mycket starka skäl. I synnerhet gäller detta naturligtvis enskilda syngrafem. Faktiska exempel är *s* i *eldsläckning-s-attiralj*, *kött-s-lig* och *t* i *jäs-t*, *lös-drif-t-sladugård*.

Betydelsefull för segmenteringen är också frågan om vilka suffix som kan urskiljas. Här finns hela skalan från uppenbara enheter som *-ing* och *-lig* till kandidater som kräver särskilda utredningar. Återkommande genomarbetningar av det slag som jag nämnde tidigare ledde så småningom till att vi ställde upp följande huvudprinciper (vartill kommer vissa tillämpningsregler som jag inte tar upp här).

Ett ifrågasatt segment kan etableras som ordbildningsuffix, om det dels kan isoleras genom kommutation i minst två fall, dels är ordklassavledande. Genom belägg som *balsamin* och *blondin* (jämför *balsam* och *blond*) etableras alltså ett substantivbildande suffix *-in*. Två fall ger emellertid bara status som svagt suffix. Vi anser oss inte kunna ta detta som intäkt för segmentering som ger upphov till restled, d.v.s. grundmorfer som inte uppkommer genom fullständig kommutation. Om vi inte hade haft några fler kommuterande fall i det aktuella exemplet, skulle alltså *giljotin*, *magasin*, *poplin*, *vaccin* och en rad andra ha förblivit osegmenterade. (Som ett faktiskt exempel på osegmenterade lemmar med suffixliknande ordslut inskjuter jag här paret *kaross* och *koloss*.)

Vi har emellertid också preciserat vad vi kallar starka suffix. Som huvudprincip gäller att ett starkt suffix skall ha minst fem kommuterande fall. Att ett sådant suffix betecknas som starkt beror på att vi anser det så väletablerat, att det kan urskiljas också när restled uppkommer. I fråga om det substantivbildande *-in* gäller att vi utöver *balsamin* och *blondin* också har kommuterande fall som *alexandrin*, *disciplin*, *jakobin*, *termin* m.fl. Eftersom då *-in* visar sig vara starkt, segmenteras också *giljotin* i restledet *giljot-* och suffixet *-in* o.s.v.

Det skall tilläggas att principkraven på två fall respektive fem fall naturligtvis är arbiträra. Bakom dem ligger dock många överväganden och åtskilliga provborrningar. Vi tycker oss ha funnit att de träffar intuitionen rätt väl. Genom vårt sätt att arbeta ut-

kristalliserade sig efterhand en lista över prefix och en lista över suffix. Dessa användes som praktiska hjälpmedel vid den slutgiltiga genomgången.

Jag vill gärna tillägga, att mycket av det jag har berört i denna summariska översikt är av typen isberg. Bland större problemkomplex som inte ryms inom min ram kan nämnas den klassiska delmängden av lemmana, de moderna – mest engelska – lånorden, frågan om avgränsningen mellan propriella och icke-propriella segment och frågan om förekomsten av suffix som inte entydigt avleder en bestämd ordklass.

På det sätt som jag emellertid grovt har skisserat arbetade vi successivt fram ett analysstadium, där ordmaterialet hade uppdelats i segment, som också hade försetts med grundläggande ledtypsmarkeringar. Den första beskrivningsnivåns enheter, de grafiska morferna, baserar sig som beteckningen anger på de erhållna segmentgränserna (och endast på dem).

## Homomorfkomponenter

Det viktigaste kriteriet för separeringen i homomorfkomponenter på den andra beskrivningsnivån är den nämnda fördelningen på ledtyper: grundmorfer, suffix o.s.v. Härigenom kommer exempelvis vår grafiska morf *dom* att motsvara två enheter, grundmorfen *dom* och det substantivbildande suffixet *-dom*.

Härutöver gäller ett uttalskriterium. Eftersom vi rör oss med ordled, som ju har en mycket varierande ställning i orden och därmed kan ha ett växlande uttal, har vi valt att ge detta kriterium en grov karaktär eller om man så vill en försiktig tillämpning. Vi kräver med andra ord, att en uttalsgrundad separering skall ha en stark motivering. Som en sådan motivering ser vi förekomsten av uttal, som från strukturens synpunkt kan betecknas som mera centrala (mera normalt svenska) än övriga hos en given grafisk morf. Detta ger oss i tillämpliga fall – och de är inte särskilt många – vanligen en binär separering. Ett exempel är *pol* i *nord-pol* gentemot *pol* i *polera*, *poliklinik* och *metropol*. Också fallet

*dom* berörs i själva verket. Grundmorfen fördelas på *dom* i *domare*, *dominerad* och *dom* i *Nidarosdom*.

Kriteriet uttalsskillnad gäller mera generellt tillämpat också för avledningssuffixen. I *klientel* och *bindel* har vi alltså två olika suffix *-el*, ett betonat med lång vokal och ett obetonat med kort vokal. I *miljon*, *elektron* och *lexikon* har vi tre olika suffix *-on*, två betonade med lång vokal av olika kvalitet och ett obetonat med kort vokal.

Sambandet mellan de båda första beskrivningsnivåerna är, som väl har framgått, starkt. Segmenteringen förutsatte ju en allmän ordstruktur med ordled såsom prefix, grundmorfer o.s.v. Sambandet har också tagit sig uttryck i att en stor del av arbetet på de båda nivåerna har utförts vid gemensamma körningar. Segmenteringen och ordledsseparatoringen skedde sålunda genom interaktiv bearbetning med hjälp av ett adaptivt program, som var baserat på en stavelseapproximerande grafotax.

Analysen ha gett cirka 7700 olika grundmorfer och cirka 570 olika affix. De senare fördelar sig på ett 60-tal prefix, cirka 470 avledningssuffix, 10 fogar och ett 30-tal böjningsändelser (alltså i medial ställning).

Som exempel på ordledens frekvens i lemmanas uppslagsformer kan nämnas, att prefixet *be-* har i runt tal 1500 belägg, avledningssuffixet *-dom* cirka 200, fogen *-s-* 9700 och böjningsändelsen *-or-* 18. Några exempel på grundmorf-frekvenser är *far* 233, *fas* 45, *fat* 19, *fjomp* 1.

Genomsnittligt innehåller ett lemma 3,57 ordled på denna beskrivningsnivå. Eftersom de 55 000 analyserade lemmarna har cirka 930 000 belägg i texten, har ordleden sålunda cirka 3 320 000 textbelägg. Ordledens textuella genomsnittsfrekvens är följaktligen så hög som cirka 400. Jämför härmed att de analyserade lemmanas motsvarande frekvens är 17. Siffrorna får sägas ge en intressant aspekt på vokabulärens och textens makrostruktur: en liten kärntrupp av språkliga element bygger upp bådadera.



## Formella morfem

Den fortsatta analysen ägnas åt grundmorferna. Den tredje beskrivningsnivåns formella morfem framkommer genom en klassifikation som i huvudsak grundar sig på regelbunden växling av tre typer i kombination med semantisk spärr enligt ovan. Hur variationskriterierna närmare skulle preciseras, har vi diskuterat en hel del. Inte minst har övervägandena gällt samspellet mellan vokalväxling av omljudstyp och den semantiska spärran.

Växlingstyperna är fonologisk variation och grafonomisk variation inom en morfotaktisk ram samt sublemmatisk växling. Några fonologiska exempel som kan ge en antydning om hur variationen har utnyttjats är *bred/bret(t)*, *elv(a)/elf(te)*, *(kon)cep(tion)/(kon)cip(iera)*, *fabell/fabul(ös)*, *arv/ärf(tlig)/ärv(a)*, med något större semantisk vidd *hand/hant(langare)/händ(ig)/(höger)hänt* och med större fonologisk töjning *bliv(a)/bli*. Den första av formerna i varje grupp har betraktats som den form varifrån den eller de andra med naturliga processer kan härledas, och den har därför ansatts som uppslagsform.

Några grafonomiska fall är *ess/läss*, *camp(ing)/kamp(ing)*, *kaff(e)/kaf(é)/caf(eteria)* och de systematiska typerna *simm(a)/sim(ning)* och *till(försel)/til(läggsproposition)*.

I vissa fall – och det gäller både på den fonologiska och på den grafonomiska sidan – ger kriteriernas tillämpning på en homomorfkomponent en fördelning på två olika formella morfem. Det genomgående exemplet *dom* är av det slaget. Enheten *dom* med kort vokal är till en del hänförlig till det formella morfemet *dom/döm* (ur *chefsdomare*, *dödsdöma*), till en del till det formella morfemet *dom* (ur *fransksdominerad*). Tidigare har vi härutöver urskilt *dom* med långt å-ljud och suffixet *-dom*. Ett grafonomiskt fall är det nyss givna exemplet *kamp*, som dels går till enheten *camp/kamp*, dels går till enheten *kamp* ur *könskamp*, *Thermopylekamp* o.s.v.

Kriteriet sublemmatisk växling är tillämpligt i fall som *bär-* i *hembära*, *uppbära* visavi *bur-* i *bilburen*, *utsädesburen*. Vi får således det formella morfemet *bär/bur*. I analogi härmed etable-

ras dels enheten *nyp* ur *nypa* (jfr sublemmaformen *nöp*), dels enheten *nyp* ur *nypon*. På den här punkten kan vi jämföra med vad som just påpekades beträffande *dom* och *camp*.

## Polysemkomponenter

Den semantiska differentieringen av de formella morfemen gäller principiellt en uppdelning av polysema enheter i polysemkomponenter. Det allmänna målet kan sägas vara att urskilja de olika huvudbetydelseerna.

Vi övervägde i början bland annat den i och för sig intressanta möjligheten att utnyttja en tesaurus. Tanken var då att ett formellt morfems lemmabelägg skulle differentieras i polysemkomponenter alltefter de olika begreppsklasser som de kunde hänföras till. Av flera skäl lät vi denna metodmöjlighet förbli en intressant möjlighet. Ett var de märkbara brister som kända tesaurer har och som det skulle krävas ett mycket stort arbete att eventuellt råda bot på. Ett annat var de problem som uppkommer genom att en enhet inte sällan kan föras till olika begreppsklasser. Ett tredje var att vi gärna såg en precisare betydelspecificering.

I den hittills redovisade delen av analysen har vi utnyttjat ett negativt semantiskt kriterium. Vad det gäller nu är att komma åt positivt bestämbara betydelser. Sådana betydelser skall enligt våra krav vara möjliga att dels parafrasera, dels särskålla från betydelserna hos eventuella andra komponenter grundade på det aktuella formella morfemet.

Betydelseerna bestäms med ordets övriga delar som kontext – vid behov kringstående ord på lemmanas beläggställen. Ett enkelt exempel på differentiering är följande.

### TIPP

*tipp* 1 'spets' (i *svanstipp* m.fl.)

*tipp* 2 'gissa' (i *topptippad* m.fl.)

*tipp* 3 'stjälpa ut' (i *slamtippning* m.fl.)

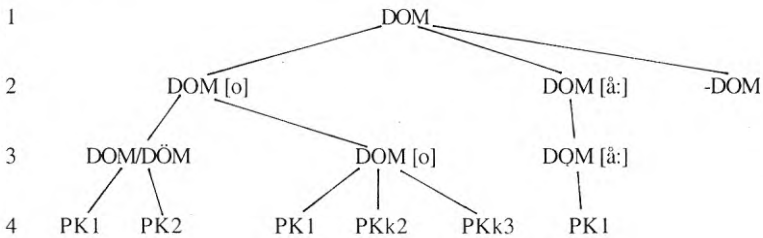
De formella morfem som differentieringen gäller är som nämnt belagda i lemman. För varje förekomst av ett betraktat morfem i

vederbörande lemman gäller det att pröva, om det kan hänföras till en urskild huvudbetydelse. I de fall där detta inte låter sig göra prövas först om en ny betydelse kan urskiljas. Går inte det, vilket inte sällan är fallet, expanderas den betraktade enheten genom konkatenering med angränsande enheter (prefix, suffix, annan grundmorf, fog) tills en positivt bestämbar betydelse som motsvarar de nyss nämnda kraven erhålles. Ett exempel:

*PAT*

*pat* 1 'sjuk' (i *patolog, patient*)  
*pat* 2 'känna starkt' (i *sympati, patos*)  
*patent* 'ensamrätt'  
*patin* 'ädelärg' (i *patina, patinering*)  
*patio* 'förgård'

I detta fall urskiljs alltså två huvudbetydelser. Beläggen i övrigt konkateneras.



FIGUR 1. *Analys av elementet dom.*

Också när det gäller *dom* blir det fråga om ett par konkateneringar. För översiktens skull sammanfattar jag behandlingen av *dom* på de olika beskrivningsnivåerna i diagrammet figur 1, där PK står för polysemkomponent och PKk för konkatenerad polysemkomponent.

De resulterande polysemkomponenterna är i tur och ordning: *DOM/DÖM* PK1 'avge kvalitativt utlåtande' (i *domstol, slutomdöme*), PK2 'välde' (i *shejkdöme*); *DOM [o]* PK1 'överkyrklig' (i *domprost*), PKk2 *domin* 'vara förhärskande' (i *domine-*

*råd*), PKk3 *domän* 'område' (i *skogsdomän*); *DOM* [å:] PK1 'katedral' (i *Nidarosdom*).

Ibland kan de belägg som visar sig kräva konkatenering ge upphov till en expanderad enhet som har distinktvt olika betydelser. Som följande fall visar gäller detta exempelvis det *passion* som uppkommer vid analysen av det formella morfemet *pass*: *passion* 1 'lidelse' (i *passionerad*), *passion* 2 'lidande' (i *passionsspel*).

De morfologiska enheternas betydelser blir av naturliga skäl oftast mindre specifika – mer abstrakta – än ordens betydelser. Abstrakthetsgraden varierar emellertid givetvis. Också inom en polysemkomponent är det både rimligt och nödvändigt att tillåta en viss semantisk vidd. Några typer av semantiska relationer inom en komponent skall nämnas.

En typ är den analogiska utvidgningen i metaforisk riktning.

#### VECK

*veck* 1 'sjudagarsperiod' (i *vecka*)

*veck* 2 'vikning' (i *veck*, *invecklad*)

Vad vi noterar här är att huvudbetydelsen 'vikning' igenkänns också i fall som *invecklad*, *utveckla*, *förveckling*.

En annan typ av utvidgning är generalisering från material till föremål tillverkat av eller bestående av materialet. Inom enheten *järn* ryms således nyanserna 'ämnet järn' och 'redskap av järn'. På motsvarande sätt förhåller det sig med *vatt* (ur *vatten*), som får innefatta både nyansen 'vätska' och nyansen 'sjö' o.s.v.

En tredje typ är vidgning genom tillkomst av specialiserade betydelsenyanser. Hit hör exempelvis *väg* i *järnväg*. Däremot spränger *gås* i *smörgås* ramen – här sker alltså konkatenering. Ett annat exempel på vad vi ser som ramsprängande vidgning gäller elementet *kre*.

#### KRE

*kre-* 'skapa' (i *kreativ*)

*kreatur* 'fä' (i *nötkreatur*)

*rekre-* 'vila upp sig' (i *rekreationsort*)

Både i  *kreatur* och i  *rekre-* har vi alltså konkatenerat. Specialiseringen har gått för långt sett från synkronisk synpunkt.

En fjärde typ är konkretisering eller faktivering, vanligen av en verbhandling. Ett fall som  *hugg* står som beteckning för både verksamheten i fråga och resultatet av verksamheten i form av ett märke av något slag.

Vokalväxlingen i t.ex.  *fall/fäll* har utnyttjats vid klassifikationen i formella morfem. Den ökade semantiska spännvidd som uppkommer i dessa och liknande fall och som kan fångas i termer som kausativ (som här), perfektiv eller inkoativ betydelse betraktar vi inte som skäl till differentiering i olika polysemkomponenter. Den kan alltså ses som en femte typ av utvidgning i semantiskt avseende.

## NFO 4

Resultatet av undersökningen kommer alltså att presenteras i fjärde delen av Nusvensk frekvensordbok. Innehållsförteckningen upptar ett 50-tal olika listningar. Dessa redovisar olika aspekter på de fyra olika beskrivningsnivåerna. Tyngdpunkten ligger på den fjärde nivån, polysemkomponenterna.

Den ojämförligt största listningen är ett initialalfabetiskt ordnat lexikon. Det är sorterat på de formella morfemen och upptar för vart och ett av dem de polysemkomponenter som har urskilts och deras parafraaser samt alla lemman vari vederbörande komponenter är belagda. I praktiken blir detta sålunda en komplett morfemkonkordans i långt driven bearbetning. Det inses att varje användare av lexikonet härigenom dels i detalj kan se hur vår analys har gjorts, dels kan göra utdrag och eventuella modifieringar efter eget behov.

## Enheternas lingvistiska status

Analysens enheter baserar sig på en överblick över hela undersökningsmaterialet. Speciellt för grundmorfernas vidkommande

gäller att detta har kompletterats vid undersökningen genom referens till Illustrerad svensk ordbok och Svenska Akademiens ordlista (10 upplagan).

Erfarenheterna från tidigare års analytiska arbete har kunnat utnyttjas. Särskilt skall här nämnas förtrogenheten med lingvistiskt materialstudium och med det faktiska undersökningsmaterialet.

Enheterna har vidare kommit fram som resultat av en genomarbetning av materialet från en rad olika utgångspunkter och i åtskilliga omgångar. Jag erinrar om de olika beskrivningsnivåernas roll och om vad som nämndes inledningsvis om arbetsmetoden.

På centrala punkter har av varandra oberoende arbetslag i kontrollsyfte arbetat med samma stoff och kommit till synnerligen väl överensstämmande resultat. Detta gäller även differentieringen i polysemkomponenter.

Vi är därför böjda att hävda, att det resultat som redovisas har relevans för svenskans ordbildningssystem i morfologisk och semantisk mening. Vi är klara över att invändningar kan göras på många enskilda punkter. En systematisk lösning med avseende på ett stort material tycks emellertid ligga i den riktning vi har gått.

## Litteratur

Allén, S., et al., Nusvensk frekvensordbok baserad på tidningstext. 1–3. 1970–75.

Allén, S., Text-based lexicography and algorithmic text analysis. (Publiceras i ALLC Bulletin 1977.)

Illustrerad svensk ordbok. Red. B. Molde. 3 rev. uppl. 1964.

Ljung, M., A frequency dictionary of English morphemes. 1974.

Svenska Akademiens ordlista över svenska språket. 10 uppl. 1973.

Thorén, B., 8000 ord för 8 års engelska. 1959.

## Några könsskillnader i journalistispråk

Könsroller i språk. 1977.

Materialet till undersökningen är hämtat ur textunderlaget till Nusvensk frekvensordbok (NFO) och utgörs alltså av artiklar ur morgontidningar i Stockholm, Göteborg och Malmö (ej ledare eller sportartiklar). Det omfattar de 150 000 löpord som har kvinnliga upphovsmän och en lika stor textmängd som har manliga upphovsmän. Den manliga delen är slumpmässigt vald ur de återstående 850 000 orden i NFO-materialet. Socialt sett får författarna sägas vara en rätt väl sammanhållen grupp.

Artiklarna i NFO-materialet är grupperade i 6 ämnessfärer, nämligen naturvetenskap med tillämpningar (NA), kulturvetenskap med tillämpningar (KU), politik och samhällsfrågor (PO), näringsliv och samfärdsel (NÄ), människa och miljö (MÄ) och konstnärlig verksamhet (KO). I det aktuella materialet råder jämvikt mellan de kvinnliga och de manliga journalisternas bidrag till ämnessfärerna KU och KO. Däremot är det kvinnlig dominans i MÄ, medan det är manlig dominans i NA och NÄ och en viss manlig övervikt i PO.

Bearbetningen gäller frekvenserna hos (de grafiska) orden i de båda könsgrupperna. Detta är självfallet bara en av en lång rad olika lexikaliska och grammatiska undersökningar som kan göras med datamaskinens hjälp. Den kan ses som ett demonstrationsexempel.

För att komma åt de eventuella specifikt kvinnliga respektive manliga orden kan man beräkna värdet på ett index baserat på den kvadrerade skillnaden i frekvens. Som könsindex (KX) uppställs

$$KX = \frac{(K-M)^2}{K+M},$$

där K är ett ords frekvens i det kvinnliga materialet och M ordets frekvens i det manliga. Detta index får värdet 0, om K och M är

lika. Står exempelvis 10 belägg mot 6, blir värdet 1, och står 10 mot 0, blir värdet 10. Skillnadens riktning framgår av täljarens tecken före kvadreringen.

Indexet är i själva verket ett specialfall av det kända spridningsmättet  $\chi^2$  (chi-kvadrat). Om indexets värde är  $KX > 3,85$ , är skillnaden mellan K och M signifikant på 5%-nivån, dvs. det är då 95 % chans att det rör sig om en verklig skillnad och inte om slumpens spel med beläggen. Är värdet  $KX > 10,83$ , har vi signifikans på 0,1%-nivån.

Med tanke på de båda könsgruppernas varierande bidrag till de olika ämnessfärerna läggs två villkor på fördelningen över dem. Ord som skall komma i fråga måste vara belagda i minst 3 ämnessfärer och ingen ämnessfär får svara för 50 % eller mer av beläggen (vid tillämpningen justeras för ämnessfärernas olika storlek).

Vi kan nu exempelvis plocka ut de substantiv som är mest karakteristiska på ömse håll. Sätter vi kravet så högt som  $KX > 10,83$  för att få en begränsad lista med typiska ord, anmäler sig följande grupp av personbetecknande substantiv på kvinnosidan: *barn, elev, flickor(na), fru, fröken, föräldrar(na), kvinnor, lärare* och *mor*. Personbeteckningarna hos de manliga skribenterna å andra sidan blir *engelsmännen, general* och *polisen*. De övriga substantiven är hos kvinnorna *betyg, huset, höst, lust* och *år*, medan de hos männen är *frihet, frågor, håll, intresse, krafter, mening, miljoner, procent* och *situationer*. Skillnaderna får tala för sig själva.

Att de villkor som ställdes på fördelningen av ett ords frekvens över ämnessfärerna verkligen ger oss ord, som inte är beroende av en viss sfär, kan belysas med ett par exempel. Fördelningen av kvinnornas belägg på *kvinnor* är (värden justerade med hänsyn till ämnessfärernas storlek ges inom parentes): NA 3 (8), KU 11 (6), PO 8 (8), NÄ 0, MÄ 5 (4), KO 9 (7). Fördelningen av männens belägg på *general* är på motsvarande sätt: NA 1 (3), KU 10 (6), PO 8 (8), NÄ 1 (3), MA 0, KO 2 (1).

Bland andra möjligheter kan vi naturligtvis också välja att pröva könsskillnader i morfologiskt bestämda grupper. En sådan grupp bildas av orden på *-ande*. Ställer vi här det mera normala



kravet  $KX > 3,85$ , får vi följande utfall. De kvinnliga journalisternas bidrag karakteriseras av de tre adjektiven *levande*, *lockande* och *spännande*. Mot dem står i den manliga gruppen dessa adjektiviska och adverbliella ord: *betydande*, *ledande*, *liknande*, *motsvarande* och *uteslutande*.

Betraktar vi med samma förutsättning ord på *-ig*, får vi ett likartat resultat. Kvinnor: *härligt*, *lyckligt*, *manliga*, *plötsligt* och *roligt*. Män: *frivilliga*, *mänskliga*, *naturligt*, *ovanligt*, *påtaglig*, *tidigare*, *troligt*, *viktiga(re)*, *väsentlig*, *ytterligare* och *årliga*.

Allmänt kan kvinnoorden på *-ande* och *-ig* karakteriseras som raka, medan mansorden då till stor del kan sägas vara mera krokiga. Med detta stämmer att ord som *bra* och *fint* också är klart kvinnliga i materialet. Jämför Gustaf Cederschiöld i Om kvinnospråk och andra ämnen 1917: kvinnor har få "tveksamma och modifierande uttryck".

Låt mig till sist visa på *hon* och *han* i det här sammanhanget. Hos de kvinnliga journalisterna är *hon* nästan 5 gånger vanligare än hos de manliga: 487 fall mot 100. Däremot är *han* mycket jämnare fördelat: 829 fall på den kvinnliga sidan mot 1139 på den manliga. Siffrorna för *hon* tycks antyda, att om inte kvinnorna talar om kvinnor, så blir det inte mycket talat om dem. Jämför Elisabet Hermodssons utrop i Synvända 1975 "Tala kvinnan in i världen!"

Texterna är från 1965. Läget kan ha förändrats mer eller mindre sedan dess: Det finns numera möjligheter att följa utvecklingen genom att utnyttja det material som samlas i Logoteket, ett nationellt serviceorgan vid avdelningen för språklig databehandling (Språkdata) i Göteborg. Mängden maskinläsbara texter av olika slag i Logotekets arkiv beräknas öka med omkring 10 miljoner löpord per år (den datamaskinella ordbanken omfattar för närvarande omkring 200 000 olika lexikaliska ord). Man kan uppenbarligen räkna med att ett sådant material kan ge intressanta upplysningar om könsroller i språk.

\*\*\*

Den här rapporterade undersökningen har utförts i samarbete med Martin Gellerstam och Christian Sjögreen vid Språkdata.

## On Phraseology in Lexicology

Cahiers de lexicologie 29. 1976.

*Les manifestations de langue naturelle peuvent être considérées comme des chaînes de symboles porteuses de données de deux sortes: données qualitatives (quelles sont les unités et les règles pertinentes?) et données quantitatives (à quelle mesure les unités et les règles reviennent-elles?). Pour caractériser quantitativement les mots, il faut donner non seulement leur fréquence et leur distribution sur les parties d'un texte, mais aussi leur tendance à former des constructions. La phraséologie, au sens large du mot, peut être regardée comme l'intersection de la grammaire et du lexique. Dans tome III (1975) de Frequency Dictionary of Present-Day Swedish sont présentés les résultats d'une étude sur les jonctions de mots dans un corpus d'un million de mots. En prenant comme critères la récurrence (c'est-à-dire la présence répétée de cas identiques dans le corpus) et les conditions grammaticales (y compris les conditions sémantiques) on a pu spécifier trois niveaux de description pour les jonctions de mots, notamment les combinaisons, les constructions et les locutions. Dans la quantité de jonctions de mots trouvée dans l'étude – plus de 600 000 combinaisons – se reflète la nature fort récurrente des langues naturelles. Ce phénomène devrait être considéré dans les discussions générales de grammaire aussi bien que dans les discussions de psycholinguistique. L'importance des jonctions de mots comme critères dans les études de style et d'attribution de textes est signalée également.*

I make a point of the first preposition in the title because what I am going to do is talk about *one* way of studying phraseology. The field is vast and complicated and there is room for many different approaches.

Being the first speaker at the Summer School, I might start by making a few comments on some notions that are relevant to

computational linguistics. They will also serve as a general background to my presentation.

Computational linguistics itself can be characterized as the elaboration and application of formalized methods of a computable kind for the purpose of gaining insights into natural language. A shorter – and vaguer – formulation would be the algorithmic study of natural language. In any case, the emphasis on natural language clearly establishes computational linguistics as a branch of linguistics, having computer science as an auxiliary science.

Natural language can be considered from many points of view: as a sequence of sound waves if you are a phonetician, as a series of linear patterns if you are a graphonomist, as a physiological phenomenon, or a kind of behaviour, or an object of learning, etc. In the particular perspective of computational linguistics it is fruitful to look upon natural language manifestations as strings of symbols. As an example, consider words, made up of graphemes, as representations of meanings.

The reason why it is profitable, in our present perspective, to view natural language in this way, is that the (digital) computer is designed to process symbols; this is what makes it such a tremendously useful tool.

The strings of symbols carry two interesting types of data: qualitative and quantitative. The purpose of studying the qualitative aspects of the flow of data is to determine what units are relevant to your description and what rules govern the use of these units. The basic question of the quantitative approach concerns the frequency of units and rules. It has to be stressed that this is by no means the only quantitative question. I shall have more to say about this in due course.

In order to study quantities you have to define the objects of your study as rigorously as possible – and please notice that words constitute just one type of a wide range of objects. However, because of the rich variation inherent in authentic language, you will always find yourself faced with a domain of cases that are difficult to settle. This is the sort of experience that is worth

carrying home from Pisa. It is one of the things at the very heart of natural language.

The quantitative question is basically a matter of style, since quantities tell you something about how language is used. In addition to what is called linguistic competence there seems to be a stylistic competence that governs the use of the former. As implicitly mentioned before, the flow of data enables you to look into both.

Two years ago I was here talking about an investigation of words defined in three basic ways – as graphic words, as homograph components, as lemmas – and words forming classes of different kinds: inflectional classes of words, parts of speech, etc. This was done within the framework of the study underlying Parts 1–2 of the *Frequency Dictionary of Present-Day Swedish* (1970–71). The type of lexical units with which we shall concern ourselves this time – phrases or, rather, collocations – is the subject of Part 3 of the dictionary (published in 1975).

An initial question is, of course: What is phraseology? I think it can be regarded as the intersection of grammar and lexicon. This means that the collocations we are interested in are, on the one hand, grammatically well-formed and, on the other, lexically selected.

Let me give you an example of lexical selection, and why not take one from Swedish? It may provide a change in the customary 'the-man-hit-the-ball' menu.

Let us take the plural noun *kretsar* "circles, quarters". Looking at the instances of this word in a million-word corpus, we find that its attributive modifiers belong to a very limited set of types. One of them comprises *akademiska* "academic", *intellektuella* "intellectual", *estetiserande* "aestheticizing", *teologiska* "theological", *kyrkliga* "ecclesiastical", and *museala* "museum". A second type includes *politiska* "political", *ekonomiska* "economic", *diplomatiska* "diplomatic", *auktoritativa* "authoritative", *informerade* "well-informed", *extremistiska* "extremist", and *kommunala* "local". Number three is represented by *danska* "Danish" and *sovjetiska* "Soviet". Number four is *interna* "internal". Number five includes three comparative forms of one

adjective: *vida* "wide", *vidare* "wider", *vidaste* "the widest". To a sixth type I assign the pronominal words *vissa* "certain" and *samma* "the same". This exhausts the list of modifiers. I think you will have little difficulty in perceiving the strong selective influence exerted by the head-word.

What I have just demonstrated implies that the study of phraseology is of interest from a general grammatical and lexical point of view. The main theoretical problem is how to incorporate collocations into an explicit linguistic description. Collocations are also of interest from a psycholinguistic point of view. The question here is focused on their role in speech production.

The third point of interest that I should like to draw your attention to is the quantitative one. As I said earlier, the frequency of a word is not enough in this respect. Another important characteristic is the distribution over sub-texts. Suppose that we study the occurrences of words in five sub-texts. One word might have the series 10, 10, 50, 0, 1, whereas another might have 15, 14, 14, 14, 14. The total is 71 in both cases but the distribution over the sub-texts is very different. This fact must be accounted for and there are different ways of doing it.

The Dioscuri frequency and distribution are not enough, however. The phraseological aspect should be added to the quantitative characterization. What is wanted is information about the extent to which a word enters into collocations. Our technical term for this phenomenon is constructional tendency. It can be calculated as the percentage of the total frequency of a (graphic) word made up by its frequency in collocations (more precisely: constructions – I shall return to this term in a moment).

One of the lists in Part 3 of the dictionary I mentioned gives the constructional tendency for graphic words with a frequency of 10 or more. There are about 9,000 such words. For some hundred of them the current value is 100 and for about one thousand it is 0. The bulk of the words have values in between. It should be clear that the value 100 may be reduced for some of the words if the corpus is widened. Among the 100-scoring words are *bortsett* in *bortsett från* "disregarding"; *enlighet* in *i enlighet med* "in accordance with"; *fjol* in *i fjol* "last year";

*loppet* in *inom loppet av* "within", *under loppet av* "in the course of", and *i (det) långa loppet* "in the long run"; and *sinom* in *i sinom tid* "in due course". Examples of 0-scoring words are *behandlar* "treats", *påven* "the pope", and *vanligen* "usually". To sum up this section, the quantitative characterization of a word should include frequency, distribution, and constructional tendency.

This brings me to an outline of the phraseological investigation of the current million-word corpus. The basic idea is the following. A necessary condition on collocations seems to be what might be called recurrence. This means simply repeated, identical occurrences. In this particular investigation recurrence is interpreted as two or more identical occurrences in one million running words. Collocations that meet this condition are called combinations. They provide the first level of description.

Language as a system can be seen as a social convention between people about what units and rules belong to the system. This convention is, so to speak, a guarantee for their occurrence in language manifestations, i. e. their recurrence.

You will easily see, however, that the condition just specified is not sufficient. We need another condition, too. One way out would be to apply a criterion used by J. McH. Sinclair and his collaborators. It is called "significant collocation" and is described as "regular collocation between two items, such that they co-occur more often than their respective frequencies and the length of the text ... would predict". The fact that this criterion is exclusively quantitative means that it gives not only natural phrases like *very nice*, *quite a few*, *old man*, etc., but also non-natural phrases like *if put*, *and to*, etc. This effect is not desirable.

It might be mentioned here that in the Swedish investigation the theoretical frequency for all recurrent combinations with an observed frequency of 10 or more has been calculated. The formula used is

$$T = \frac{F_1 F_2 \cdots F_w}{N^{w-1}}$$

where  $F_1$  is the observed frequency of the first word in the combination,  $F_2$  the observed frequency of the second word,  $w$  the number of words in the combination, and  $N$  the number of running words in the material, i. e. one million. The  $T$  value is the expected frequency in the case of a randomly ordered corpus. The ratio  $F/T$ , where  $F$  is the observed frequency of the combination, shows how many times more frequent than expected the combination in question is.

A list in Part 3 of the dictionary registers the values of the ratio  $F/T$  for those combinations which I specified a little while ago. The values vary between 180,000,000 for *på ett helt annat sätt* "in a wholly different way" and 0.03 for *en att* "one that; one to". As many as 92 % of the cases are more frequent than expected. The general trend is that the number of non-well-formed units increases with decreasing  $F/T$  value. The two extreme examples just given illustrate this trend quite well. The crucial point is, however, that cases of non-well-formed combinations crop up very early in the list. A string like *den s k* "the so-called" is 52,000 times more frequent than expected, *ut så här* "out like this" 1,000 times more frequent, etc. It is obvious that a statistical measure does not yield the collocations aimed at.

Instead, grammatical criteria have been used. On the basis of a pilot study covering a tenth of the combinations, a set of linguistic selection criteria was set up. These criteria specify seventeen types of collocations that we consider relevant to phraseology in a wide sense.

They are listed and exemplified on pp. XXXV – XXXVII in the Introduction to Part 3 of the dictionary. Roughly, they can be grouped into the following four main categories: noun-phrases, verb-phrases, connectives, and clauses. These categories may be exemplified by *den stora frågan* "the big question", *ta form* "take form", *även om* "even if", and *det är givet att* "it is evident that", respectively. The collocations obtained we call constructions. They provide the second level of description.

Constructions make up a sub-set of combinations. A sub-set of constructions can, in turn, also be distinguished: idioms. There are many suggestions concerning the definition of idioms. Bruce

Fraser takes the view, in a paper of 1970, that almost everything is an idiom. Sinclair lists a number of definitions in the publication I mentioned before. One of them comes very close to the one we have adopted.

In our investigation, an idiom is a construction whose meaning is not predictable on the basis of the meanings of its component words. An example is *stå för dörren* which means "be imminent, be at hand" although its literal meaning is "stand in front of the door". Another example is *trampa i klaveret*, meaning "make a blunder", literally: "tread on the piano".

We have applied the idiom rule quite restrictively. Simple metaphorical use has not been considered to be enough. A case in point is *lägga till grund för* "make the basis of".

Idioms provide the third descriptive level. Thus, the three levels of collocations are combinations, constructions, and idioms.

For obvious reasons there are a number of important questions that I have to ignore in this paper. Examples of such questions are the following: What exactly is meant by "identical instances of collocations" (consider, among other things, capital letters and juncture graphemes)? How should overlapping collocations be handled? How should collocations included in larger collocations be treated? How can these things be taken care of in the quantitative account? What about discontinuous collocations? For a discussion of these and other points, see the Introduction to Part 3 of the dictionary.

Now just a few words about some practical aspects of the current investigation. From our previous work we had a word concordance to our corpus, i. e. an alphabetical list giving each occurrence of each word in its context. Our first step was to generate, on the basis of the word concordance, a concordance of combinations specified, in the way I have already described, as recurrent identical collocations. Five lines from the concordance of combinations are shown in Figure 1. The first two lines contain the non-well-formed combination *den fulla* meaning "the full", followed by different nouns. The last three lines contain a well-formed combination, i. e. a construction, *den fulla sysselsättningen* meaning "full employment".



ÄR TILLÄTNA OCH DÄR ENDAST GENOM DEN FULLA FRIHETEN MARKEN KAN BEREDAS FÖR DET KOMMANDE . +  
 URERING SOM MAN ÄNNU INTE INSETT DEN FULLA INNEBÖRDEN AV . + DET EN GÅNG KULTURBÄRANDE PATRI  
 R STÅ I BOSTADSKÖ LIKA LÅNGE . + DEN FULLA SYSSLSÄTTNINGEN OCH LOKALISERINGSPOLITIKEN FÖR A  
 PROC . + DETTA BETYDER M A O ATT DEN FULLA SYSSLSÄTTNINGEN MÅSTE INNEBÄRA EN FORTSATT KAPPL  
 TIKEN , OM MAN VILL UPPRÄTHÅLLA DEN FULLA SYSSLSÄTTNINGEN . + PRISET FÖR DENNA, VILKEN BL

FIGURE 1. From the Concordance of Combinations.

The concordance of combinations proved quite voluminous, comprising over 600,000 instances. One reason for this is that included and overlapping combinations are comprehended. The really important explanation, however, is that the result reflects the vastly recurrent character of natural language.

The application of the construction criteria yielded a body of some 200,000 instances representing about 50,000 different constructions. The result of the idiom-rule application was about 1700 instances representing about 300 different idioms.

The wealth of constructions invites some comments on theoretical and psycholinguistic issues. The investigation suggests strongly that a psychologically realistic model of language production and perception should take into consideration the vast number of collocations or lexical blocks, as they might be called. Personally, I am convinced that they are, to a very large extent, part of the lexicon at the disposal of the language users. This implies, too, that they must be given much more attention in the general grammatical discussion. Up till now, idioms have been more or less the sole collocational concern of general linguists.

	F	Fx	För
FULLA 22/48			
<i>fulla</i> av	7	3	7
<i>fulla</i> förtroende	2	2	2
<i>fulla</i> hus	2	2	1
den <i>fulla</i> sysselsättningen	3	3	2
händerna <i>fulla</i>	4	2	4
händerna <i>fulla</i> av	2	2	2
sin <i>fulla</i> rätt	6	1	6
var <i>fulla</i> av	2	2	2
i sin <i>fulla</i> rätt	5	1	5

FIGURE 2. Part of the article *FULLA* in the main list.

Part 3 of the *Frequency Dictionary of Present-Day Swedish* contains 28 different lists. The main list gives all the constructions under each of their component words. Figure 2 shows part of the article *fulla* in this list. The fraction tells us that 22 out of the 48 instances of the graphic word *fulla* are found in (recurrent)

constructions. F and Fx are two types of frequency (the former including those cases in which a collocation occurs as an element in a longer collocation). "För" accounts for the number of authors having used the construction in question.

Besides in the dictionary, the data discussed here are available in machine-readable form through the Swedish Logothèque, the service organ of the Department of Computational Linguistics at the University of Göteborg.

Finally, a few words about some applications of the study of phraseological units. Literary stylistics is an obvious field; the adjacent field of text attribution is another. Language instruction should naturally be mentioned, too. The concordance of combinations, furthermore, is of course a rich mine for investigations into distributional semantics.

Thus we end up in semantics, where most things seem to meet nowadays.

## References

- Allén (S.), "Text-Based Lexicography and Algorithmic Text Analysis", to appear in *Proceedings of the 6th International Conference on Computational Linguistics*, Ottawa, 1976.
- Allén (S.), "The Swedish Logothèque. A Computer-Based Text and Word Bank", in *Festschrift F. de Tollenaere*, 1977.
- Allén (S.), "Travaux du centre de recherches pour le suédois contemporain de Göteborg", A. Zampolli (ed.) *Linguistica matematica e calcolatori*, Leo S. Olschki, Firenze, 1973, pp. 3-18.
- Allén (S.), "Vocabulary Data Processing", H. Benediktsson (ed.); *The Nordic Languages and Modern Linguistics*, Reykjavík, 1970, pp. 235-261.
- Allén (S.) et al., *Frequency Dictionary of Present-Day Swedish* 1-3, Almqvist & Wiksell, Stockholm, 1970-75. (S. Allén (ed.), *Data linguistica*, 1, 4, 10.)

- Bever (T.G.), "A Survey of Some Recent Work in Psycholinguistics", IBM, W.J. Plath (ed.), *Specification and Utilization of a Transformational Grammar*, 1968.
- Fraser (B.), "Idioms within a Transformational Grammar", *Foundations of Language*, VI (1970) 1, pp. 22-42.
- Gorcy (G.) *et al.*, "Le traitement des 'groupes binaires'", *Cahiers de lexicologie*, XVII (1970) 2, pp. 15-46.
- Hellberg (S.), "Computerized Lemmatization without the Use of a Dictionary: A Case Study from Swedish Lexicology", *Computers and the Humanities*, VI (1972) 4, pp. 209-212.
- Sinclair (J. McH.), Jones (S.), Daley (R.), *English Lexical Studies*, Birmingham, 1970.

## Alla svenskers efternamn

Dagens Nyheter 25 januari 1976.

Språkforskaren Adolf Noreen tog 1918 ett initiativ som bland annat ledde till att man gjorde en förteckning över de släktnamn som bars av landets kyrkskrivna invånare. De 48 000 namn som registrerades finner man i boken "Sveriges familjenamn 1920" (Noreen stavade rikets namn så).

Den nu aktuella undersökningen, som gjorts av det material som disponeras vid Språkdata, visar att vi i dag har inemot 200 000 olika släktnamn, från Aaben, Aaberg och Aabo till Öödik, Ööpik och Öövel. Mängden har alltså fyrdubblats på ett halvsekel.

Orsakerna är väl främst två. Den ena av dem är att myndigheterna aktivt har verkat för namnbyte, bland annat genom att ge ut böcker med namnförslag. Den andra är att de naturaliserade medborgarna, invandrargrupperna, har fört med sig rader av släktnamn precis som fallet är med förnamnen.

Låt mig påminna om de tre principerna för att bestämma vad som är ett namn: stavningsprincipen (samma stavning – samma namn), uttalsprincipen (samma uttal – samma namn) och ursprungsprincipen (samma ursprung – samma namn).

Talet 200 000 bygger på stavningsprincipen. De 120 vanligaste namnen är listade i en tabell här nedan. Det mest brukade släktnamnet, Johansson, är nästan på pricken lika vanligt som det mest brukade förnamnet, Maria.

Man kan lätt se att ungefär var tjugonde svensk heter Johansson (och var tionde kvinna Maria). Tabellens 120 namn täcker halva befolkningen. Den andra hälften av folket delar alltså på cirka 200 000 namn.

Bland de 120 namnen på listan lägger man märke till fem olika typer av namn. Son-namnen gör sig bemärkta genom att belägga de 21 översta platserna. Sedan glesnar emellertid leden, och det är trots allt bara 50 av namnen som är son-namn.

TABELL 1. *De vanligaste släktnamnen.*

380 000	Johansson	370 000	Andersson	290 000	Karlsson
230 000	Nilsson	180 000	Eriksson	170 000	Larsson
160 000	Olsson	150 000	Persson	150 000	Svensson
95 000	Pettersson	95 000	Gustafsson	75 000	Jonsson
70 000	Jansson	55 000	Hansson	50 000	Jönsson
45 000	Petersson	45 000	Bengtsson	35 000	Carlsson
35 000	Gustavsson	35 000	Magnusson	30 000	Olofsson
30 000	Lindberg	30 000	Lindström	25 000	Jakobsson
25 000	Lindgren	25 000	Axelsson	25 000	Lundberg
25 000	Bergström	25 000	Lundgren	25 000	Fredriksson
20 000	Matsson	20 000	Berglund	20 000	Berg
18 000	Sandberg	18 000	Håkansson	18 000	Henriksson
18 000	Danielsson	17 000	Fransson	17 000	Lindqvist
17 000	Forsberg	17 000	Sjöberg	17 000	Johnsson
16 000	Engström	16 000	Samuelsson	16 000	Lundin
15 000	Eklund	15 000	Mårtensson	15 000	Lind
15 000	Holm	14 000	Lundqvist	14 000	Nyström
14 000	Holmberg	14 000	Bergman	13 000	Isaksson
13 000	Söderberg	13 000	Nordström	13 000	Eliasson
13 000	Nyberg	13 000	Lundström	13 000	Björk
13 000	Arvidsson	12 000	Gunnarsson	12 000	Berggren
12 000	Ström	12 000	Månsson	12 000	Wallin
11 000	Hermansson	11 000	Nordin	11 000	Björklund
11 000	Andreasson	11 000	Jonasson	11 000	Sundberg
11 000	Hedlund	11 000	Sandström	11 000	Ekström
11 000	Holmgren	11 000	Sjöberg	10 000	Åberg
10 000	Abrahamsson	10 000	Öberg	10 000	Göransson
10 000	Hellström	10 000	Martinsson	10 000	Strömberg
9 000	Blom	9 000	Dahlberg	9 000	Norberg
9 000	Söderström	9 000	Blomqvist	9 000	Sundström
9 000	Löfgren	9 000	Åkesson	9 000	Åström
9 000	Johannesson	9 000	Bergqvist	9 000	Josefsson
9 000	Ek	9 000	Adolfsson	9 000	Börjesson
8 000	Lund	8 000	Ottosson	8 000	Hallberg
8 000	Boström	8 000	Englund	8 000	Lindholm
8 000	Nygren	8 000	Möller	8 000	Sjöström
8 000	Palm	8 000	Nyman	8 000	Viklund
8 000	Dahl	8 000	Söderlund	8 000	Ivarsson
8 000	Falk	8 000	Borg	8 000	Höglund
8 000	Vikström	8 000	Hedberg	8 000	Holmström

Flest representanter har i stället typen Lindberg, Lundgren etc. som räknar 54 stycken. Detta slags namn börjar uppträda i mera nämnvärda mängder vid mitten av 1600-talet, delvis under inflytande från adelsnamnen.

Ett dussin namn är av den enkla typen Berg, Holm, Ek, Borg etc. En del av dessa återgår på ortnamn. Andra är soldatnamn.

Trion Lundin, Wallin, Nordin tillhör en franskinspirerad typ som under 1600-talet och 1700-talet utvecklades ur latiniserade former på -inus. Av en femte typ är det enstaka fallet Möller. Namnet är inkommet från tyskan. Det är ursprungligen fråga om yrkesbeteckningen mjölnare.

Uttalsprincipen ger inte särskilt dramatiska förändringar i listan. Detta hänger samman med att en viss stavning av ett givet namn normalt dominerar kraftigt. I fallet Eriksson är det exempelvis så att just denna stavning svarar för 178 000 av sammanlagt 186 000 belägg.

Bland de tio vanligaste namnen ger uttalsprincipen den enda förändringen att Pettersson (som inte gärna kan slås samman med Petersson) passeras av Gustafsson. Detta beror i sin tur på att varianten Gustavsson mot huvudregeln är ganska vanlig.

En del namn – kanske särskilt son-namnen – har likväl en ganska varierande stavning. En av våra mest bekanta målare kallades ju X-et därför att han stavade sitt namn Sven Erixsson.

Här är tretton förekommande sätt att variera namnet, ordnade efter vanlighet: Eriksson, Ericsson, Ericson, Erixon, Erikson, Erickson, Ericksson, Erixson, Erichson, Ericzon, Erichsson, Ericxon, Erixsson.

De två sista stavningarna möter man bara en gång var. Floran är som synes rik. Det är rent av fråga om det finns något namn som varierar mer (märk väl med oförändrat uttal). Bland de vanligaste son-namnen gör det i varje fall inte det.

Ursprungsprincipen i sin tur ger en hel del förändringar i listan. Enligt denna princip sammanförs till exempel Johansson, Jo(h)nsson, Jansson, Hansson, Jönsson, Jonasson och Johanneson. Bakom alla dessa ligger Johannes.

De tio vanligaste ursprungsgrupperna bland släktnamnen blir i frekvensordning de som återgår på Johannes, Andreas, Karl, Peter, Niklas, Olof, Erik, Lars, Sven och Gustaf. Johannesgruppen kommer upp i över 650 000 fall.

Den här synpunkten ger oss en möjlighet att få en total bild av de inflytelserika förnamnens roll i den nutida svenska personnamnsskatten.

Tar vi alltså hänsyn till förekomst i både förnamn och släktnamn, blir de tio vanligaste förnamnen enligt ursprungsprincipen Johannes, Maria, Karl, Anna, Andreas, Erik, Elisabet, Peter, Ing(e)- och Margareta.

Johannes dominerar kraftigt med sina 1,2 miljoner namnbelägg. De övriga ligger i området från trekvarts ner till ungefär en halv miljon fall. Det är anmärkningsvärt att de kvinnliga namnen hävdar sig så väl trots att de inte förekommer i släktnamnen. Ursprungliga nordiska namn är bara Karl, Erik och Ing(e)-. Det kristna inflytandet är alltså massivt.

### **Stålnacke och Buskas**

Mot den frekvenslista som gäller för hela landet – jag återgår nu till stavningsprincipen – kan man ställa särskilda frekvenslistor för varje län.

Ett intressant sätt att utnyttja dessa är att matcha dem mot den riksomfattande. Man kan till exempel resonera så här. Ett namn som Hallberg har i hela landet en frekvens av ungefär 8000. Det betyder att en person på tusen heter Hallberg.

Vilka och hur många är de namn i de olika länen som där förekommer minst en gång på tusen men som är mindre vanliga än så när hela landet betraktas? Vad har vi med andra ord för regionala särdrag i översta delen av frekvensskalan.

Det är många påfallande ting som kommer fram. Jag skall göra några axplock. Fyra län ligger i särklass när det gäller antal namn med de angivna egenskaperna. Det är ytterområdena Gotland med över 90 och Norrbotten–Västerbotten–Västernorrland med vardera över 70. Som jämförelse kan nämnas att exempelvis Värmlands län bara har åtta.

De tio namn som ligger främst i det här avseendet på Gotland är Norrby, Thomsson, Gardell, Stenström, Hallgren, Hellgren, Dahlström, Ahlgren, Gardelin och Ahlström.



Motsvarande grupp i Norrbotten är Sundqvist, Öhman. Lundmark, Marklund, Granström, Öman, Granberg, Lindmark, Stålnacke och Nordlund.

Att Stålnacke är så vanligt i Norrbotten – i själva verket en på fyra hundra – är antagligen överraskande för de flesta. Också namn som Björnfot och Drugge och finskklingande namn som Niemi, Lahti, Uusitalo med flera hör till de vanliga.

Namnet Gardelin på Gotlandslistan representerar en framträdande grupp: Bolin, Kolmodin, Bendelin, Uddin, Siggelin och andra. Annars är nog Buske det namn som tillhör den mest särpräglade gruppen bland de drygt 90 namnen på listan.

Strax under frekvensgränsen ligger bland andra de samhöriga Knutas och Hoas. Det rör sig här om en speciell namntyp med östsvenska anor.

Jämför de båda tionamnslistorna från Gotland och Norrbotten med denna från Blekinge: Mattisson, Holgersson, Berntsson, Jeppsson, Gummesson, Lindell, Ingemansson, Antonsson, Enarsson och Thomasson.

Med ett undantag finner vi här bara son-namn. Kontrasten mot de två andra listorna är signifikativ. Utbredningen av son-namnen i länslistorna visar sig vara mycket klart sydvästsvensk (Blekinge–Skåne–Småland–Bohuslän–Halland–Västergötland–Bohuslän). Bland alla de över 70 namn på Norrbottens länslista finns bara ett son-namn, Olovsson.

### **Kodgge och Pissarewsky**

I statens offentliga utredningar redovisades 1960 de vanligaste släktnamnen enligt en undersökning av en del av befolkningen.

Det är den undersökningen som ligger till grund för den så kallade rikslistan. Denna upptar 154 vanliga namn. Har man sitt namn på den listan – och det har varannan svensk – är detta enligt namnlagen tillräckligt skäl för att byta namn, om man så önskar. Dessa vanliga namn betraktas nämligen inte som ”tillräckligt särskiljande”.

Det visar sig att rikslistan avviker en hel del från det facit som vi nu har till hands.

Att Andersson ligger överst i rikslistan i stället för Johansson är naturligtvis ingen stor sak, inte heller att en lång rad andra omkastningar och förskjutningar förekommer inom ramen för det antal namn det gäller.

Man skulle också kunna ha överseende med att ett tjugotal av namnen i rikslistan i själva verket inte är riktigt så vanliga att de egentligen är kvalificerade för att tas upp där.

Viktigare är att åtta namn som faktiskt har en tillräcklig hög frekvens inte finns med i rikslistan. Fyra av dem tillhör rent av de 100 vanligaste: Månsson, Åkesson, Åström och Börjesson (se tabellen). De övriga är Davidsson, Olausson, Edlund och Knutsson. Det är över 60 000 svenskar som bär de åtta namnen. Det finns all anledning att revidera rikslistan så att dessa medborgare får samma möjlighet att byta namn som de som heter till exempel Malm, Melin, Moberg och Möller och som har kommit med på listan. Hög frekvens i det egna länet enligt de beräkningar som jag tidigare nämnde bör också vara ett gott skäl.

Det är obestridligt att vi har färre fula namn än de flesta andra nationer, skriver Esaias Tegnér den yngre i en klassisk studie.

Vi har inga motsvarigheter till tyskarnas Teufel, Narr och Lumphose eller fransmännens Mauvais och Méchant påpekar han. Huvudskälet är att vi under vår korta egentliga släktnamns-historia – 200 år när han skrev, 300 år i dag – i stort sett kan sägas ha tagit oss namn, medan andra folk under betydligt längre tidens lopp har fått sina namn, till stor del som binamn eller öknamn.

När man numera tar namn här i landet sker det under överinseende av namnbyrån vid patentverket. Språklig expertis är inkopplad. Namn som Ahngwé, Kodgge och Pissarewsky, som Adolf Noreen klagade på (ingen heter numera så), skulle inte ha en chans att passera. En del av det som togförs skulle kanske ändå kännas ovant för Tegnér – men det löjeväckande kan undvikas.

## Alla svenskers förnamn

Dagens Nyheter 27 december 1975.

Om man gör en lista över de ord som används i dagstidningstext, visar det sig att vart sjunde ord i listan är ett namn av något slag. Och tre fjärdedelar av allt som trycks på svenska – bortsett från visitkort och liknande – är sådan text. Det är en av orsakerna till att man vill ha ett grepp om namnskatten när man sysslar med språkanalys. Inte minst blir personnamnen intressanta. Vilka namn finns det och hur vanliga är de? Man vill också tränga in i strukturen hos namnen. Hur är förnamnen och släktnamnen byggda? Hur kombineras förnamnen med varandra? Namngivningens variation i tid och rum drar likaså uppmärksamheten till sig. De aspekterna vetter åt samhället: kulturinflytelser, modeväxlingar, regionala skillnader.

För att kunna avgöra namnfrågor måste myndigheterna känna till vilka namn som finns och hur vanliga de är. Kunskapen om existerande släktnamn har hittills varit långtifrån fullständig och kunskapen om förnamnen ännu mera ofullständig.

Tankar av det här slaget ledde till att projektet Svenska namn startades för några år sedan som ett led i forskningen vid Språkdata, en avdelning vid Göteborgs universitet. Avsikten var att försöka göra något som aldrig hade gjorts tidigare, nämligen att undersöka personnamnen hos ett lands hela befolkning. Den uppläggningsen är nödvändig för den kompletta registreringens skull. Den är möjlig tack vare metodiken inom språklig databehandling. Projektets program har skrivits av Staffan Wåhlin.

Genom en överenskommelse med riksskatteverket fick vi tillåtelse att utnyttja en nerbantad version av de personregister på magnetband som länsstyrelserna arbetar med. På våra band finns namnen på alla svenska medborgare per den 1 januari 1973. Vi har fått datainspektionens tillstånd att utnyttja materialet. Patentverkets namnbyrå har stött projektet.

I den här artikeln håller jag mig till förnamnen. Hur många är de? Vilka är vanligast? Ja, det beror på vad man menar med namn.

Ingenting är självklart i det fallet. Tag paret Per och Pål. Uttalet Pär motsvarar stavningar som Per, Pehr och Pär. Så långt är det gott och väl. Uttalet Pål motsvarar bland annat stavningarna Pål, Pähl och Paul. Men stavningen Paul har inte bara uttalet Pål utan också uttalet Paul med diftong. Om vi helt skulle gå på uttalet, skulle vi alltså genomgående få två olika namn stavade Paul, det ena med uttalet Paul och det andra med uttalet Pål. Den effekten vill vi inte alltid ha.

Sedan lägger vi märke till serier som Per, Peter, Petter, Peder, Petrus med de kvinnliga pendangerna Pära, Petra, Petra, Pedra, Petrusina respektive Pål, Povel, Paul, Pavel, Paulus med Påla, Pavla, Paula bland motsvarigheterna på den kvinnliga sidan. Här kan hela floran sammanfattas i de båda utgångsformerna Petrus och Paulus.

För att hyfsa behandlingen kan det vara lämpligt att precisera vad som är ett namn på åtminstone tre sätt: (1) stavningsprincipen: samma stavning – samma namn (2) uttalsprincipen: samma uttal – samma namn (3) ursprungsprincipen: samma ursprung – samma namn. Enligt den första principen är Per, Pehr och Pär tre namn, enligt den andra och enligt den tredje ett namn. Jämför stavningens roll i samhället: Per Johansson är en annan person än Pehr Johansson.

När stora namnmängder skall behandlas är det lämpligt att hålla sig till stavningsprincipen. Det är för övrigt knappast möjligt att i praktiken ta reda på hur alla namn uttalas av alla namnbärare. Det visar sig att vi tillsammans har i det närmaste 17 miljoner förnamn. Vi har alltså fått i genomsnitt två namn var.

Antalet olika förnamn är förbluffande stort, i runt tal 80 000. Som jämförelse kan nämnas att Roland Otterbjörks namnbok "Svenska förnamn", som dock inte följer stavningsprincipen, innehåller 2 000. Och i almanackan har vi bara 358 (då är Luther och Petri inräknade – båda används som förnamn).

Man frågar sig var föräldrarna får allt ifrån. Saken är främst den att invandrargrupperna bland de svenska medborgarna bidrar

med stora delar av sina ursprungliga hemländers namnförråd. En liten bit av förnamnslistningen ser ut så här: Juoni, Juozapas, Juozas, Jupangda, Jupen, Jupp, Jura, Juraj, Jurane, Juraté, Jurd, Jure, Jureck, Jurek, Jurel, Jurgas, Jurgen.

Stavningsvariationen gör naturligtvis också sitt. En bidragande orsak är dessutom dubbelnamnen. Det finns till exempel långt över 100 dubbelnamn som innehåller John, från John-Aldo, John-Alex, John-Alfred etc. via John-John till John-Willard, John-Yngve, John-Åke och exempelvis Carl-John, Lill-John och Reza-John.

De vanligaste namnen är kolossalt vanliga. De 80 i toppen svarar för hälften av de 17 miljonerna namnbelägg. I tabellen på nästa sida visar jag dessa 80 namn i rangordning med avrundade uppgifter om antalet förekomster. Det finns alltså 386 000 personer som har namnet Maria osv.

Om vi nu i stället följer uttalsprincipen, blir det vissa förändringar i listan. De tio vanligaste förnamnen är då de här: Maria, Karl/Carl, Elisabet(h), Erik/Eric, Margaret(h)a, Anna, Kristina/Christina, Lars, Gustaf/Gustav, Olof/Olov. Medan Maria alltså ligger kvar överst bland de kvinnliga namnen, avancerar Karl/Carl till första platsen på herrsidan. Vanligast av alla är fortfarande Maria.

Ursprungsprincipen är inte minst intressant. Tillämpar vi den kan vi få en bild av kulturströmningar i stora drag. Vi går alltså till väga så att vi buntar ihop Maria, Mari, Marie, Marita, Mary m.fl. och ger hela gruppen etiketten Maria, som är den gemensamma utgångspunkten. Under Johannes samlar vi på motsvarande sätt Johan, John, Jo(h)nny, Jonas, Jan, Jens, Hans, Ivan m.fl.

Många gamla nordiska namn har den speciella egenheten att de har bildats av två olika namnled enligt det så kallade variationssystemet. Exempel är Ing(e)- i Ingeborg, Ingrid, Ingemar osv. och Gun(n)- i Gunvor, Gunnar, Gunvald osv. Eftersom dessa namnled har en ganska självständig karaktär, gör man den nordiska namnskatten störst rättvisa genom att gruppera ihop namn som innehåller samma led.

TABELL 1. *De 80 vanligaste förnamnen uppställda efter stavning.*

386 000	Maria	328 000	Erik	321 000	Karl
299 000	Margareta	273 000	Anna	241 000	Elisabet
197 000	Lars	183 000	Kristina	180 000	Gunnar
177 000	Nils	174 000	Karin	174 000	Sven
171 000	Eva	168 000	Lennart	167 000	Birgitta
165 000	Anders	143 000	Ingrid	142 000	Olof
141 000	Per	135 000	Johan	119 000	Åke
118 000	Elisabeth	116 000	Gustaf	114 000	Bengt
111 000	Bertil	111 000	Linnéa	110 000	Hans
107 000	Jan	103 000	Kerstin	101 000	Ingeborg
97 000	Marianne	96 000	Arne	93 000	Viola
91 000	Göran	85 000	Axel	83 000	Bo
83 000	Stig	81 000	Elsa	80 000	Inger
75 000	John	74 000	Gustav	73 000	Lena
73 000	Ulla	72 000	Ingemar	71 000	Gösta
71 000	Marie	70 000	Christina	68 000	Peter
68 000	Astrid	64 000	Sofia	62 000	Mikael
62 000	Leif	62 000	Oskar	61 000	Rut
61 000	Viktoria	61 000	Rolf	60 000	Ingvar
59 000	Märta	59 000	Vilhelm	59 000	Anita
57 000	Kjell	56 000	Roland	54 000	Ulf
54 000	Helena	54 000	Irene	53 000	Rune
53 000	Gunilla	53 000	Barbro	53 000	Margit
53 000	Ingegerd	53 000	Christer	52 000	Mats
52 000	Inga	52 000	Linnea	50 000	Carl
49 000	Britt	47 000	Signe	46 000	Bror
46 000	Magnus	46 000	Stefan		

Enligt ursprungsprincipen är de tio vanligaste namnen: Maria, Anna, Elisabet, Johannes, Ing(e)- (kvinnl.), Margareta, Kristina, Birgitta, Karl, Erik. Maria ligger stadigt kvar på sin förstaplats, men Johannes tar alltså täten bland de manliga namnen. Frekvenstalen blir nu mycket höga. Det är faktiskt trekvarts miljon kvinnor som har namn i Mariagruppen och en halv miljon män som har namn i Johannesgruppen.

Som framgick förut har många kvinnliga och manliga namn gemensamt ursprung. Till Johannes går Johanna, och med det namnet följer Hanna, Jane osv. Kvinnliga och manliga namn på Ing(e)- kan naturligtvis föras samman, för att ta ett annat exempel. En total ranglista efter ursprung får följande utseende längst

upp: Maria, Ing(e)-, Anna, Johannes, Elisabet, Kristian, Margareta, Karl, Gun(n)-, Birgitta.

Det är bara att konstatera att Maria är orubbligt vilken namnprincip man än väljer. Detta är nog inte utan skäl. Låt oss läsa listan från kulturströmningarnas synpunkt. Maria är Jesu moder. Anna är Marias moder, Johannes står för både Döparen och aposteln. Elisabet var Döparens moder och är dessutom ett helgon. Kristian är den kristne. Margareta och Birgitta är helgon. Vi ser hur sju av de tio namnen på listan dokumenterar ett genomgripande inflytande från den kristna sfären på namngivningen. Det är bara de speciella namngrupperna Ing(e)- och Gun(n)- som jämte Karl representerar det gamla nordiska stoffet.

Materialet pekar däremot på att Maria – till skillnad från exempelvis Anna – inte så ofta är tilltalsnamn. Man har funnit att det första namnet är tilltalsnamn i 2 fall av 3, men Maria är bara första namn i 1 fall av 8.

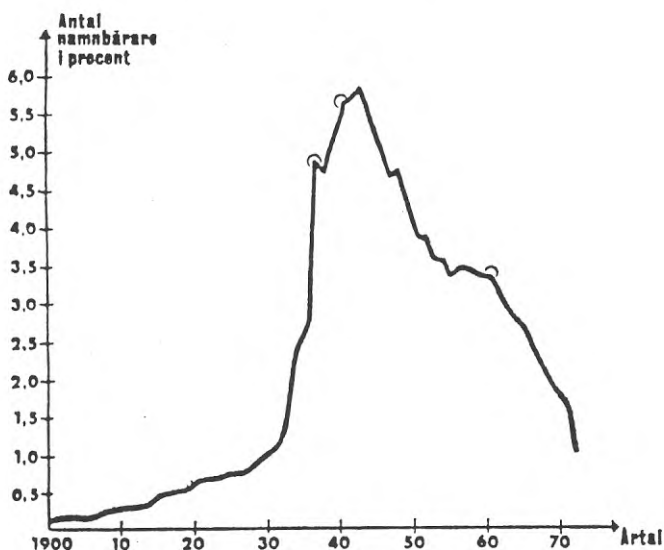
Nu har det varit mycket tal om de vanligaste namnen. Det är emellertid de ovanliga som är den stora gruppen bland de olika namnen. Det förhåller sig till och med så att hälften av de 80 000 förnamnen har en enda bärare. Några exempel är Arwida, Aryane, Ashild, Aspbjörn, Astolphus, Axelfina och Axvid.

Vad vi tidigare noterade om de kristna namnens roll är ju resultatet av ett inflytande under lång tid. Studerar man variationen under en kortare period, är det snarare modeväxlingar man iakttar. Namnmaterialet ger fina möjligheter till detaljerade observationer av förändringarna under 1900-talet. Namnbärarna kan nämligen fördelas på födelseår. Därigenom kan man följa variationen från år till år.

Som exempel har jag tagit namnet Birgitta. Hur utvecklingen från sekelskiftet har varit framgår av kurva 1. På den vågräta axeln är årtalen utsatta. Den lodräta axeln anger antalet namnbärare i procent av årsklassen (både kvinnor och män). Kurvan är ganska dramatisk. Den börjar med läget 0,1 procent år 1900. Under mitten av 30-talet stiger den brant upp mot 6 procent för att från 40-talet raskt falla ner mot läget 1 procent.

Man frågar sig varför. Aktuella personer som bär namnet kan vara åtminstone en delförklaring. Några händelser är markerade

med ringar i kurvan. Prinsessan Birgitta föddes i januari 1937. Det har säkerligen spelat in. Antalet namnbärare steg från 2,9 procent 1936 till 4,8 procent 1937. Men observera att stigningen redan då var på god på väg. Vem har ett förslag till förklaring av den plötsliga uppgången 1933? Hur aktuell var den heliga Birgitta? Till 550-årsminnet 1941 av hennes helgonförklaring gav man ut två frimärken. Den aktualiseringen kan ha bidragit till att fördröja kurvans fall. En viss uppbromsning ser man under 50-talet. Raset börjar på nytt 1961. Det råkar vara det år då prinsessan Birgitta gifte sig och flyttade utomlands.

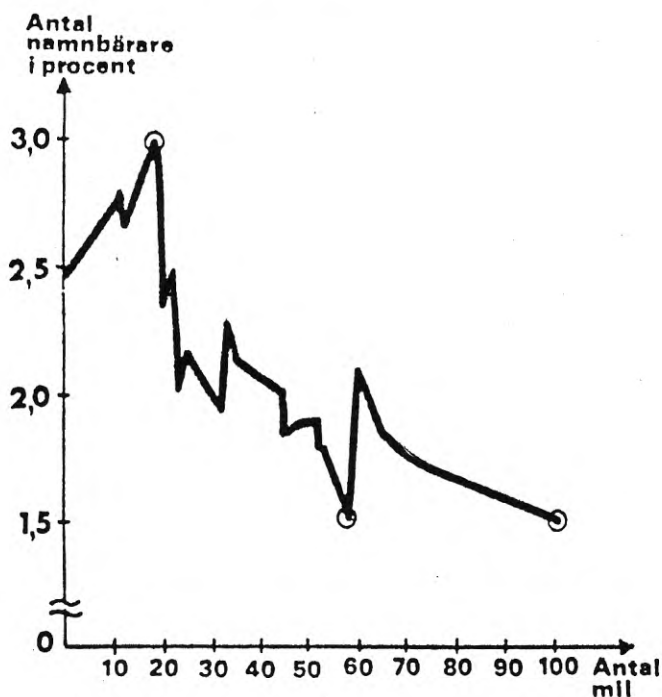


KURVA 1. *Birgittor i befolkningen under 1900-talet.*

Om det nu är så att det utgår ett inflytande från Stockholm, borde man kunna spåra det inte bara i tiden, utan också i rummet. En språkgeografisk grundtanke är att inflytandet från ett spridningscentrum blir svagare ju längre bort det når. Namnmaterialet ger möjligheter till undersökningar av detta fenomen också. Namnbärarna kan fördelas på län.



Kurva 2 visar fördelningen på län av dem som heter Birgitta. På den vågräta axeln har jag avsatt residensstädernas avstånd från Stockholm räknat i mil. På den lodräta axeln gäller det antalet namnbärare i procent av länets befolkning (av båda könen). Man ser tydligt hur regionen inom ungefär 20 mil från Stockholm har höga värden, medan övriga landet ligger betydligt lägre. Kurvan dalar med en fast trend efter hand som avståndet ökar. I Kristianstads län och i Norrbottens län är värdet nere vid 1,5 procent. I Gävleborgs län inom den markanta inflytelsezonen är andelen dubbelt så stor. De här tre punkterna är markerade med ringar i kurvan. Så blir då frågan vad man kan säga om namnlängden i almanackan mot bakgrunden av undersökningen. Först skall vi se på hur mycket almanackans namn utnyttjas. Vi håller oss därvid till uttalsprincipen.



KURVA 2. Birgittornas avstånd från Stockholm (1973).

Vi kan konstatera att det inte finns något namn i almanackan som saknar namnbärare. Däremot är det många av namnen som bärs av få. Hela 41 namn har mindre än 1 000 bärare. Av dem har 18 mindre än 500 bärare. I ordning från det ovanligaste namnet är dessa 18: Bartolomeus, Tiburtius, Moses, Nikodemus, Luther, Lukas, Sylvester, Ambrosius, Agata, Petri, Flora, Anund, Filippa, Ansgar, Matteus, Götilda, Ingemund, Signild. De 4 ovanligaste namnen förekommer mindre än 100 gånger var. Bartolomeus har bara 27 bärare.

Sammanlagt har almanackans 358 namn 11 miljoner förekomster. Detta motsvarar 65 procent av hela namnmängden, som ju omfattar 17 miljoner. Ett annat sätt att uttrycka saken är att vart tredje namn som de svenska medborgarna har fått inte återgår på almanackan.

Vad kommer fram om vi i stället utgår från namnundersökningen och ställer frågan i vilken utsträckning de vanligaste namnen förekommer i almanackan? Faktiskt saknar den inte mindre än 22 av de namn som ingår i den särskilda tabellen över de vanligaste namnen.

Dessa 22 namn är i frekvensordning: Johan, Jan, Kerstin, Marianne, Göran, Stig, Inger, John, Lena, Ulla, Gösta, Marie, Peter, Leif, Anita, Irene, Rune, Gunilla, Margit, Christer, Mats, Britt. De här namnen förekommer sammanlagt 1,6 miljoner gånger. Mot Bartolomeus i almanackan med 27 bärare står som den andra ytterligheten Johan utanför almanackan med 135 000 bärare.

Det är naturligtvis inte livsavgörande hur almanackans namnlängd ser ut. Vill man emellertid revidera den till större överensstämmelse med namnbruket, har man nu det underlag som fordras. Det räcker inte att i så fall ta fasta på namnens frekvens. Man måste också väga in stabiliteten över åren. Till detta kommer de kulturhistoriska aspekterna. Och könsfördelningen borde man också ta hänsyn till om man bestämmer sig för en ändring. De kvinnliga namnen upptar långt under hälften av namnlängden i almanackan.

## Den svenska språkvårdens organisation

Språkvård 1975: 4. Föredrag vid Svenska språknämndens årsmöte den 19 september 1975.

Naturligtvis skall jag i enlighet med rubriken tala om språkvårdens organisation, men jag vill gärna slå fast, att det egentligen är språket som står i centrum. Jag vill alltså försöka se frågan från språkets synpunkt.

Till belysning av en del av det jag tänker säga i fortsättningen kunde jag kanske helt konkret börja med några prov på mina språkliga upplevelser under den senaste tiden. Jag grupperar dem kring de tre begreppen struktur, stringens och stil.

### Språkvård och ettstrukna a

Följande tre språkprov berör strukturfrågorna stavning, uttal och böjning: *Göteborgs- och Bohuslän; modem; samtliga sökanden*. Det första uttrycket står på en stor skylt som markerar länsgränsen vid vägen mellan Kungsbacka och Göteborg. Genom bindestrecket efter *Göteborgs* och hopskrivningen av *Bohus län* sprider skylten missvisande information till rader av inhemska och utländska trafikanter. Ordet *modem* betecknar en liten låda som omvandlar signalen vid datatransmission. Efterhand som sådan överföring blir vanligare i olika sammanhang, får ordet en mera framskjuten plats. Det rör sig i själva verket om en förkortning av ordparet *modulator–demodulator*. Bland tekniker uttalas det ofta med betoning på första stavelsen, medan det på andra håll inte sällan betonas på sista stavelsen efter mönster av övriga ord på *-em*. Det senare uttalet är enligt min mening att föredra. Uttrycket *samtliga sökanden* är hämtat ur ett sakkunnigutlåtande. Flertalsform på *-n* hör inte hemma här utan bland neutrala substantiv av typen *rike*.

Här följer tre exempel på bristande stringens: *Byte av slutbleck i fönster och fönsterdörrar kommer att bytas den 15/9 kl 8; An-*

ledningen till att det är så beror på att ...; Sjukdom däremot beviljas av arbetsledaren. Att bytet skulle bytas meddelades av en byggnadsfirma. Att det fanns en anledning till anledningen nämndes i en radiokommentar. Att semesteransökan endast får tillstyrkas av arbetsledaren, medan sjukdom däremot får beviljas, påpekades av en länsarbetsnämnd.

Även om stringensen är oklanderlig oeh strukturkraven uppfyllda, kan störningar i kommunikationen uppkomma. De kan ligga på det stilistiska planet. Tre exempel: *Prod:onsavd.; Lag om ändring i lagen (1974: 359) om ändring i stats tjänstemannalagen (1965: 274); Vad som föreskrives om företrädesrätt till ny anställning i 7-11 §§ och bestämmelser som avses där äger motsvarande tillämpning vid läroanstalt som avses i 1 § när läroanstalten tillsätter sådan tjänst som lärare med förordnande tills vidare från och med den 1 juli 1975 som ersätter annan tjänst som lärare med tidsbegränsat förordnande till och med den 30 juni 1975.* Förkortningen för *Produktionsavdelning* är hämtad ur ett informationsblad från Televerket rörande extra text i telefonkatalogen. Jag har konstaterat att tankarna hos läsare rör sig i riktning mot Grönköping inför denna förkortning, och rörelseriktningen förstärks för övrigt av förkortningen på nästa rad, där det står *Tryck:i för Tryckeri*. I samma cirkulär får man dessutom veta att det tillägg beträffande telefonnummer till fritidsbostad som man kan göra i katalogen kallas *främlingsuppgift*. Blandade känslor har jag också märkt att titeln *Lag om ändring i lagen om ändring i osv.* väcker. Den långa, skilje-teckenfria meningen slutligen är ett påfallande läsarovänligt prov på offentlig svenska.

Kort sagt, det finns ett behov av språkvård i alla läger och på alla språkliga nivåer. Jag tänker då på språkvård i vid mening, språkvård som i grunden tar sikte på att höja den språkliga beredskapen. Som det stora målet framstår i det perspektivet att öka språkbrukarnas förmåga att genomskåda och därigenom behärska de språkliga uttrycken och förloppen.

Vi skall också ha svårigheten i språkvårdens uppgift helt klar för oss. Man kan jämföra med stämningen av ettstrukna a. Under

tidigare århundraden stämde det i en tonhöjd som växlade från tid till annan och från plats till annan, vilket naturligtvis vållade svårigheter. På 1880-talet enades man om en s. k. normalton eller kammarton med frekvensen 435 perioder per sekund. I praktiken stämde man emellertid högre, och 1939 var det dags att ändra normaltonen till 440 perioder per sekund. Enligt vad jag nu hör stämmer man i själva verket ändå något högre, omkring 442 perioder per sekund. Det är med språket som med tonhöjden: ytterst är det de som använder instrumentet som har avgörandet i sin hand – och de är irrationella.

## Språktyper och organ

Vad jag nu främst inriktar mig på är vården av svenskan i Sverige. Detta innebär endast en precisering av ämnet. Vården av svenskan i Finland, vården av samiskan och finskan i Sverige och den gemensamma nordiska språkvården är självfallet också områden av stor betydelse.

Vilka språktyper har man att räkna med i ett sådant här sammanhang? Viktigast för språkgemenskapen som helhet är vad vi kan kalla för *allmänspråket* (jämför uttryck som tyska *Umgangssprache* och engelska *everyday English*). Det är principiellt genom detta som vi kan kommunicera med varandra. Allmänspråket har en flytande gräns mot *fackspråken*. Dessa karakteriseras av terminologier som är skraddarsydda för speciella arbetsuppgifter. En tredje typ är *litteraturspråket*, som står relativt fritt i förhållande till de båda andra typerna. Språkets resurser används i detta fall i konstnärligt syfte med de speciella effekter och avvikelser från normen som det kan föranleda. För fullständighetens skull nämner jag, att det inom språktyperna självfallet finns regional, social och stilistisk variation. De språktyper som språkvården normalt har att arbeta med är allmänspråk och fackspråk.

En språklig kategori som kräver särskild uppmärksamhet är namnen: släktnamnen, förnamnen, ortnamnen (inklusive gatunamnen), firmanamnen, varunamnen etc. Det finns en rad organ

som ansvarar för ärenden som gäller namnreglering: namnbyrån vid patentverket, myndigheten DOVA (dvs. dialekt- och ortnamnsarkiven samt svenskt visarkiv), gatunamnsberedningarna osv. Jag koncentrerar mig ytterligare genom att inskränka mig till detta omnämmande av specialområdet namn.

Vilka är mot den bakgrund som jag har tecknat de viktigaste språkvårdsorganen? Vilka är deras uppgifter, verksamhetsformer och uppdragsgivare?

*Svenska Akademien* har enligt § 22 i stadgarna som sitt angelägnaste göromål "att arbeta uppå Svenska Språkets renhet, styrka och höghet". Litteraturspråket är ett förstahandsintresse för akademien, men allmänspråket tillhör också klart dess intresseområde. Som en kuriositet kan nämnas att akademien enligt § 46, om den finner det befogat, kan göra anmärkningar över "något i allmänhet utkommande arbete". Det sägs i paragrafen att sådana anmärkningar skall vara författade med foglighet och inte "med retande eller förolämpande ordasätt". Enligt Henrik Schück i arbetet om Svenska Akademiens historia har paragrafen aldrig tillämpats. Akademiens viktigaste insats på språkvårdens område är utgivningen av Svenska Akademiens ordlista (SAOL). Frågan om vem som är uppdragsgivare torde kunna besvaras så, att akademien är sin egen uppdragsgivare. En reflex härav är det faktum, att SAOL måste auktoriseras av myndigheterna för att bli stavningsnorm i skolorna. Detta har för övrigt ännu inte skett i fråga om tionde upplagan 1973.

*Svenska språknämnden* har som huvuduppgift "att följa det svenska språkets utveckling i tal och skrift samt utöva en språkvårdande verksamhet" (§ 1 i stadgarna). Dess område är alltså svenskan som helhet, men tyngdpunkten måste anses ligga i allmänspråket. En stor del av verksamheten består naturligt nog i rådgivning: Nämnden utger en skriftserie och dessutom ordböcker. En viss forskning bedrivs också vid sekretariatet. Nämndens främsta uppdragsgivare är staten. Också institutioner och organisationer av olika slag tillhör uppdragsgivarna, liksom i viss utsträckning företag.

*Tekniska nomenklaturcentralen*, TNC, har enligt stadgarna till uppgift "att åstadkomma en för svenska förhållanden lämpad teknisk nomenklatur" (§ 1). Dess område är alltså tekniskt fackspråk. Verksamheten är i första hand av normerande art. Man utger ordlistor och andra skrifter. Man håller också på att bygga upp en termbank med datateknikens hjälp. TNC:s uppdragsgivare är staten, institutioner och organisationer av skilda slag samt en rad företag.

Det fjärde viktiga organet är *Sveriges Standardiseringskommission*, SIS. Förkortningen, som är inregistrerat märke, har sin förklaring i att namnet tidigare var Svenska industriens standardiseringskommission. Det skall understrykas, att SIS som centralorgan för den nationella standardiseringsverksamheten endast har språknormering som en deluppgift. I ändamålsparagrafen i stadgarna nämns inte någon språklig arbetsuppgift över huvud taget. Det framgår emellertid i annat sammanhang, att den typ av standard som kallas grundstandard bland annat innefattar terminologi (jämte måttenheter, matematiska symboler, toleranser, ritregler m. m.). Språktypen är för SIS:s vidkommande uppenbart fackspråk. Genom särskilda avtal är fyra s. k. fackorgan anslutna till SIS, nämligen BST (Byggstandardiseringen), MNC (Metallnormcentralen), SEK (Svenska Elektriska Kommissionen) och SMS (Svensk Mekanstandardisering). Dessa fackorgan är i övrigt organisatoriskt och finansiellt fristående. Utöver fackorganen finns en lång rad standardkommittéer. Flera tusen experter är knutna till arbetet. Den språkliga delen av verksamheten resulterar i utgivning av ordlistor. Uppdragsgivare är i första hand näringslivet men också staten och olika organisationer och institutioner.

Mellan TNC och SIS slöts 1973 ett avtal enligt vilket en samarbetsgrupp bildades. Gruppen skall verka för informationsutbyte och vara ett organ för lösande av eventuella intressekonflikter och hindrande av dubbelarbete.

Förutom de fyra huvudorganen finns det en mängd organ som genom sin verksamhet i växlande utsträckning får betydelse för språkvården. Hit hör många verk och myndigheter, institutioner och branschorganisationer m. m. Hit hör det verksamma Språk-

vårdssamfundet i Uppsala. Hit hör på ett annat plan skolorna och på ytterligare ett plan massmedierna.

## Uppgifter

Frågan blir nu vilka en språkvårdsorganisations huvudfunktioner kan anses vara. Jag urskiljer tre viktiga områden för verksamheten: språkssystemet, språkbruket och språkvårdsteorin.

Tre deluppgifter med avseende på språkssystemet är att undersöka systemet, att utarbeta normer och att registrera normerna. Undersökningen av språkssystemet innebär, att man försöker ta reda på vilka de språkliga enheterna är och vilka de regler är som styr struktureringen av enheterna och deras kombinationer till sammanhängande språkyttringar. Den ojämförligt största delen av detta arbete utförs givetvis vid universiteten.

Tre deluppgifter med avseende på språkbruket eller språkanvändningen, som man också kan säga, är: att undersöka bruket, att registrera bruket och att påverka bruket. Undersökningen av språkbruket tar sikte på vilka enheter och regler som i praktiken används, hur ofta de uppträder m. m. Också denna uppgift utförs av naturliga skäl väsentligen vid universiteten. Undersökningen av språkbruket och registreringen av bruk och normer kan idag utnyttja den automatiska databehandlingens fördelar.

Språkvårdsteorin har som jag ser det ägnats alltför liten uppmärksamhet hittills. Tre deluppgifter på området är att undersöka principerna, att kodifiera dem och att vidareutveckla dem. Bland annat bör man här kunna utnyttja insikter inom allmän språkvetenskap och samhällsvetenskap.

Det kan också vara på sin plats att i någon mån tänka efter, vilka egenskaper en god organisation bör ha. Från organisationsteorin har jag hämtat några begrepp som jag vill trycka på här. Man kan se dem omnämnda eller diskuterade också i vissa utredningar. De fem egenskaper jag nämner är flexibilitet, integration, funktionsuppdelning, självförvaltning och projektnriktning. Kravet på flexibilitet innebär förmåga till smidig anpassning efter

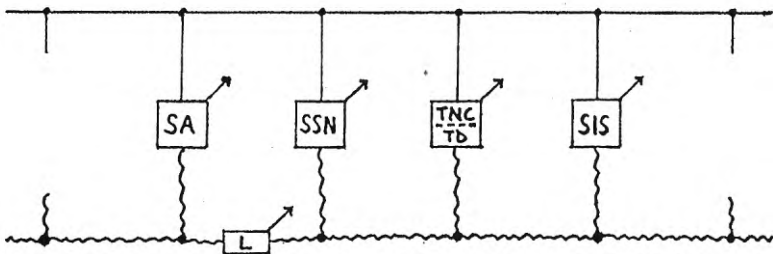


ändringar i arbetsuppgifterna. Integrationen gäller behovet av en samlad planering av verksamheten. Funktionsuppdelningen är på sätt och vis dess motsats. Den betonar vikten av att arbetsuppgifterna utförs genom specialisters försorg. Principen om självförvaltning innebär, att beslut som gäller verksamheten vid ett visst organ fattas inom organet. Projektinriktning betyder, att man försöker lösa vissa arbetsuppgifter genom temporära arbetsgrupper.

Språkvårdens nuvarande organisation är enligt mitt förmenande inte tillfredsställande när det gäller integration, funktionsuppdelning och flexibilitet. Det förekommer en viss, begränsad samverkan. Den är huvudsakligen av mer eller mindre tillfällig natur. En tydlig tendens till funktionsuppdelning kan man konstatera, men det finns inga klara riktlinjer. Hur nya behov skall tillgodoses inom organisationens ram är också oklart.

## Organisation

Övervägandena har lett mig fram till det organisationsförslag som figuren visar.



De fyra boxarna i mitten representerar de fyra huvudorganen såsom sidoställda enheter. Från vänster räknat har vi alltså Svenska Akademien, Svenska språknämnden, Tekniska nomenklaturcentralen och Sveriges Standardiseringskommission. Den lilla pil som från varje box pekar snett uppåt åt höger – om man så vill ut ur nätverket – symboliserar organens självförvaltning.

Organen arbetar alltså vidare som självständiga enheter med bevarad intern struktur och egen budget. De fortsätter naturligtvis också sin publiceringsverksamhet.

De sammanbindande linjerna representerar integrationen. Alla boxar hänger samman med alla i två avseenden. De rätta linjerna upptill representerar samarbete i fråga om språknormering, våglinjerna nertill samarbete i fråga om databehandling. Jag återkommer på den senare punkten.

Det normativa samarbetet menar jag skulle fungera så, att språknämnden genomgående konsulterar fackspråksorganen i frågor rörande terminologi på deras områden. Dessa i sin tur sänder obligatoriskt normförslag och ordlistor till språknämnden för yttrande. Det är viktigt att dessa konsultationer tas in i själva arbetsrutinerna, inte minst därför att fackspråk är potentiellt allmänspråk. Ett för språkvården betydelsefullt önskemål som kan tillgodoses inom den föreslagna organisationen är, att den centrala SAOL utarbetas på basis av ett brett samarbete. För akademien såsom utgivare kunde det också tänkas vara en fördel att på så sätt få ordlistan förankrad i de samverkande språkvårdsorganen.

Funktionsuppdelningen ger sig ganska naturligt. Språknämnden är huvudansvarig för det praktiska språkvårdsarbete som gäller allmänspråket. Fackspråksorganen har huvudansvaret för den terminologiska normeringen inom respektive områden. Arbetsfördelningen mellan TNC och SIS regleras genom särskilda överenskommelser.

En viktig del av kravet på flexibilitet tillgodoses genom att nya organ smidigt kan inordnas om så önskas. Detta markeras av de avbrutna linjerna ytterst i figuren. Som exempel på organ som kunde anslutas kan nämnas den finlandssvenska språknämnden och en eventuell medicinsk terminologicentral.

En samordnad databehandling bedömer jag också som mycket förmånlig. Det existerar redan två enheter som på ett naturligt sätt låter sig infogas i organisationen. De är systemet Termdok vid TNC och serviceorganet Logoteket vid Göteborgs universitet. Dessa representeras i figuren av förkortningarna TD respektive L.

Avsikten med Termdok är att utforma rationella metoder för insamling, lagring, bearbetning och spridning av termer och information om termer. Under 1975 har man skaffat utrustning för inkodning. Man arbetar nu på ett samlat alfabetiskt register över de svenska termerna i TNC:s ordlistor jämte motsvarigheter på engelska. Detta kan åstadkommas genom att man kodar in materialet i tidigare publicerade ordlistor och genom att man använder systemet i nyproduktionen. Utvecklingen av verksamheten kan följas i ett litet behändigt informationsblad som kallas Termdokbulletinen. Organisatoriskt sett förefaller det mycket lämpligt, att Termdok blir den enhet som tar hand om registreringen av termer på fackspråksorganens områden.

Logoteket har till uppgift att samla in, lagra och tillhandahålla maskinläsbara texter och textbearbetningar och att bygga upp en svensk ordbank. Det etablerades som reguljärt serviceorgan på nationell basis vid halvårsskiftet 1975. Verksamheten hade då pågått i mera begränsad skala i ett tiotal år inom den forskningsgrupp varur avdelningen för språklig databehandling (Språkdata), som är Logotekets moderorgan, växte fram. Logoteket förfogar redan från början över en textmängd av ett par miljoner löpande ord och ett lexikaliskt material av ett par hundra tusen ord. Datorkapacitet disponeras vid Språkdatas anläggning, som är speciellt utrustad för textbehandling. Avsikten är att årligen utöka textmängden med några miljoner löpord genom ett urval av årstrycket på svenska (texterna erhålls främst i form av hålremsor och magnetband från sätterier). Ordmaterialen i dessa texter skall infogas i ordbanken, som således kommer att registrera en betydande del av den moderna svenskans ordförråd. Det bör kanske tilläggas, att det inte är fråga om någon principiell begränsning till skrivet språk. Tillgången på transkriberade och inkodade talspråkstexter är emellertid mycket mager (här krävs förvisso en insats). Logoteket kan i sin verksamhet främja språkvården genom sitt bidrag till undersökningen av bruket och registreringen av bruket samt genom registrering av normerna (senaste upplagan av SAOL ingår exempelvis i ordbanken). Ett samarbete med fackspråksorganen erbjuder bland annat möjlig-

heten att undersöka i vilken utsträckning rekommenderade termer slår igenom.

Det är troligt att ytterligare en box skulle vara av värde i det organisationsförslag som jag nu har kommenterat. Vad jag har i tankarna är en samarbetskommitté. En sådan bör enligt vad jag kan föreställa mig vara sidoställd i förhållande till de fyra huvudorganen. Den kan lätt infogas i figuren på det sätt som jag tidigare har beskrivit.

Som ett första steg på vägen mot ökad samverkan föreslår jag, att vi tillsätter en arbetsgrupp med uppgift att behandla frågan om den svenska språkvårdens organisation och lägga fram förslag till riktlinjer för utformningen av ett närmare samarbete mellan språkvårdsorganen. Gruppen bör innefatta representanter för alla de fyra huvudorganen. Man kan exempelvis tänka sig sex ledamöter, varav tre från fackspråksorganen. Om det visar sig lämpligt, kan arbetsgruppen senare omkonstitueras som den fasta samarbetskommittén.

Ett vidgat, reellt samarbete borde öka språkvårdsorganisationens styrka och därigenom vara till gagn för språket och språkbrukarna. Det kommer naturligtvis ändå att vara långt till målet, även om man inte sätter det så högt, att alla skall kunna kommunicera friktionsfritt med alla om allting. Inom organisations teorin betonas det, att mål är ett slags ideal och att den grad i vilken man uppnår ett angivet mål kan vara ett felvisande mått på en organisations framgång. Amitai Etzioni drar i sin bok *Moder-na organisationer* en parallell med glödlampor. En lampa som omvandlar 5,5 procent av tillförd energi i ljus kan sägas vara ineffektiv. Men den är hela 22 procent effektivare än en lampa som omvandlar 4,5 procent av energin i ljus. Och den kan vara den mest effektiva som finns.

*Tillägg:* Den av Sture Allén föreslagna arbetsgruppen har på Svenska språknämndens initiativ bildats i november 1975. Den har fyra medlemmar, en från vardera Svenska Akademien, Svenska språknämnden, Sveriges Standardiseringskommission och Tekniska nomenklaturcentralen.

## I ordens ljus tonar samhället fram

Dagens Nyheter 6 februari 1975.

Varje dag fattar många tusen av oss Svenska Akademiens ordlista (SAOL) för att leta efter något på dess XVI+616 sidor. Boken är inte lätt. Den väger tre kvarts kilo.

Det ligger nära till hands att fråga sig, när man har boken i handen, varför en så tjock bok kallas ordlista och inte ordbok. Låt oss slå upp ordet i ordlistan själv och se om den har någon kommentar.

Ett första försök kan misslyckas. Bläddrar vi fram längs orden i bokstavsordningen, hamnar vi i tomrummet mellan "ordinand" med betydelsen "person som skall prästvigas eller vigas till biskop" och "ordna".

Vi måste i stället vända tillbaka till ordet "ord" och söka bland sammansättningarna där: ordaccent, ordagrann, ordagrannhet, ordalag, ordalydelse, ordasätt, ordaval, ordbehandling, ordbetoning, ordbild, ordbildande, ordbildning, ordbildningslära, ordblind, ordblindhet, ordbok osv. Så småningom hittar vi "ordlista" – utan kommentar.

I det läget kan vi gripa oss an med en ännu tyngre uppgift, två och ett halvt kilo. Vi försöker med Illustrerad svensk ordbok (ISO), som på sina X+1919 sidor ger mängder av upplysningar.

Också här hamnar vi mellan "ordinand" och "ordna" om vi inte passar oss. Men när vi har funnit "ordlista" får vi veta att det är en ordförteckning med få eller inga ordförklaringar. Vi kan också konstatera att betydelsen av "ordbok" anges som "lexikon".

### "Få" ordförklaringar

Det ligger inom möjligheternas gräns att vi nu börjar tänka på ett par saker. För det första är frågan vad som egentligen menas med få ordförklaringar. På den sida där "ordlista" står har SAOL

ett trettital. Eftersom sidan inte är unik, kommer vi fram till att antalet ordförklaringar i hela SAOL får skrivas med fem siffror.

För det andra kan vi komma att tänka på ett lexikon som Nusvensk frekvensordbok (NFO). Med sina hittills utkomna två delar (den tredje kommer ut inom kort) ligger den i en ännu högre viktklass än ISO. I NFO finns det också mängder av uppgifter om orden. De är emellertid av ett annat slag, gäller vanlighet, spridning osv. Orden preciseras med hjälp av uppgifter om ordklass och böjning. Betydelsen som sådan anges inte. Ändå är det "ordbok" som ingår i titeln.

Verkligheten stämmer alltså inte riktigt med ISO:s betydelseförklaring. Om vi tillåter oss att gå utanför de tre böcker som egentligen är på tapeten, kan vi få en annan version. Enligt band 19 av Ordbok över svenska språket utgiven av Svenska Akademien är en ordlista en ordförteckning utan någon utförligare behandling av orden med hänsyn till användning eller betydelse.

## Utförligheten

Här börjar dunklet skingras. Det är alltså inte ordförklaringarnas antal som är den springande punkten utan utförligheten i behandlingen av ordens användning eller betydelse. Men hur skall vi komma åt utförligheten?

Till all lycka kan vi utnyttja ordboksindexet Obix, som just ger ett uttryck för utförligheten i ordens behandling. Indexet beräknas i grundutförande enligt formeln

$$\text{Obix} = 100 \cdot \frac{s}{o}$$

där  $s$  är antalet behandlade ord och konstanten 100 en skalningsfaktor. Indexet är avpassat så att ordlistor får värden under 1, ordböcker värden från 1 och uppåt. Antalet rader och antalet spalter per sida liksom andra egenskaper kan naturligtvis variera från bok till bok. Vi tar detta med upphöjt lugn. Det finns mer raffinerade varianter att använda vid behov.

I SAOL behandlas 140 000 ord. Det har vi fått fram som en biprodukt vid databehandling av ordlistan inom ett forskningsprojekt. På omslaget till ISO kan man läsa att ordboken innehåller omkring 200 000 ord. I NFO behandlas ord som karakteriseras på flera olika sätt – vad ord mera exakt är kan man diskutera länge. På den närmast jämförbara nivån är antalet 70 000.

### **Som en ljuskälla**

Indexvärdena beräknar vi med uppåt avrundade tal. SAOL får  $100 \times 700 : 140\,000 = 0,5$ . För ISO blir värdet 1. Om vi räknar med de tre delar som är avslutade för NFO:s vidkommande (3200 sidor), får vi resultatet 4,6. Med hjälp av Obix har vi alltså funnit att SAOL är en ordlista. ISO har precis vad som behövs för att vara en ordbok. Det höga värdet för NFO beror på de många olika sätt som orden där redovisas på.

Som en samlande beteckning för ordlistor och ordböcker (och lexikon) använder jag här ordböcker. Efter den information om de tre ordböckerna som nu har låtit sig te är vi framme vid frågan om vad som står i dem.

Ett ord är som en ljuskälla. När strålen bryts ser man dess spektrum: stavning, uttal, böjning, (tillhörighet till) ordklass, betydelse, (uppträdande i) fraser, bruklighet. Hela spektrumet kan inte avläsas i någon av de aktuella böckerna, men tillsammans ger de en hygglig bild.

Vill man ta reda på hur ett ord stavas, går man till SAOL. Men det skall vara tionde upplagan 1973. För att kunna slå upp hur ordet stavas måste man veta (ungefär) hur det stavas, det kan inte hjälpas.

### **Bejbi eller beby**

SAOL ger också i stor utsträckning uppgifter om uttal, böjning och ordklass. Om ”baby” får vi veta att det uttalas bejbi eller bebi eller bejby eller beby, att det heter babyn i bestämd form,

babyar eller babyer i flertal, och att det är ett substantiv. En sammansättning som "ordlista" däremot har som vi har lagt märke till inga uppgifter. Meningen är att man skall titta på det enkla "lista".

När det gäller betydelsen får vi – som vi också har sett – en hel del tips i SAOL, i varje fall i fråga om osammansatta ord. Här anmäler sig emellertid ISO med hela sin bredd. Medan SAOL angett att "bahytt" är en huvudbonad, meddelar ISO att det är en slags kvinnlig huvudbonad som fästes med ett band under hakan. Till yttermera visso illustreras ordet med en bild av en dam med bahytt.

Härutöver får man en hel del fraser i ISO. Under uppslagsordet "barn" får vi förutom fem olika betydelser bland annat dessa förbindelser: hon är med barn, Barnens dag, djävulens barn, ett barn av sin tid, vara barn på nytt, klockan är bara barnet, barn och blomma.

## Stämmer inte

Vad ISO uppger om stavning, uttal och böjning kan man däremot inte utan vidare ta fasta på. Tredje upplagan 1964 bygger på SAOL:s nionde upplaga 1950. Som har framgått av en recension i DN (den 5 januari 1974) av SAOL:s tionde upplaga är ändringarna många. I fråga om vårt exempel baby anger ISO såsom uttal bäbi eller bebi och såsom alternativ flertalsform babies, vilket alltså inte längre stämmer med normen i SAOL.

Den sjunde spektralfärgen – brukligheten – knyter ordet direkt till det levande språket. I vissa fall görs denna förknippning i SAOL och ISO. Det sker på så sätt att man noterar att exempelvis "geggamoja" är ett vardagligt ord, "kadens" en musikterm och "ordasätt" ett ålderdomligt uttryck. En frekvensordning å andra sidan har som sin uppgift att presentera data om ordens användning.

Naturligtvis får man besked i frekvensordboken om antalet gånger varje ord förekommer i det material som har undersökts – i NFO:s fall en miljon ord ur modernt tidningspråk. Det är lätt



att till exempel ta reda på att ”och” är vanligast och förekommer 30 856 gånger, att ”kamp” (det som i flertal heter kamper i motsats till kampar) förekommer 116 gånger, ”humor” 33 gånger och ”schweizisk-fransk-belgisk-brittisk” 1 gång.

## Spridning

Men frekvens är bara en av de viktigaste sidorna. Spridningen över olika typer av texter är en annan. Tag ordet ”kamp” igen. I hela materialet har det rangnummer 879 (det vanligaste ordet har rangnummer 1). Men variationen mellan tidningarna är stor. I DN är ordet mycket vanligare, i Svenska Dagbladet mycket ovanligare. Rangnumren är 472 respektive 1569. Sydsvenskan ligger nära genomsnittet, har 949. Hur ställer sig detta i förhållande till tidningarnas profil?

Tag ordet ”humor”. Av sex ämnessfärer som NFO indelar materialet i är detta ord belagt i tre, kulturvetenskap, människa och miljö, konstnärlig verksamhet. Belägg saknas i de övriga tre: naturvetenskap, politik och samhällsfrågor, näringsliv och samfärdsel. Ämnessfärernas profil?

Med hjälp av data om frekvens och spridning kan man ta fram ett centralt ordförråd, en basvokabulär. En sådan ingår också i NFO. Sammanlagt finns det drygt hundra olika bearbetningar i delarna 1–3.

Också frekvensen av de fraser orden bildar – ”ha ett ord med i laget” osv. – är betydelsefulla för helhetsbilden av användningen. Det är dessa uppgifter man får i NFO 3.

## Intressanta slut

Men inte nog med detta. Också ordens roll i ordbildningen måste redas ut. Fall som ”ordlista”, ”mångordig” och ”tilltalsord” har alla något att säga om frekvensen för ”ord”. Detta skall sägas i fjärde delen av NFO. Arbetet pågår.

Svenska ord är särskilt intressanta i slutet. Där har vi till exempel den viktigaste delen av en sammansättning, och där har vi böjningsändelserna. Ett enkelt men effektivt knep att komma åt denna språkliga härlighet är att ordna orden alfabetiskt efter sista, näst sista osv. bokstaven. Vi får då vad som kallas en finalalfabetisk lista eller baklängeslista.

Bearbetningar av den sorten finns också i NFO. Vill vi veta vad för slags ord vi har omkring oss, kan vi slå upp: plankord, budord, modeord, smådeord, slagord, bibelord, nyckelord, rimord, Ibsenord, lösenord, Expressledarord, överord, dödsord, könsord, honnørsord etc.

Språket utvecklas. Det som var nusvenska blir dåsvenska. Ordlistor och ordböcker måste också utvecklas. Nya språkmaterial måste undersökas och redovisas.

Ett levande samhälle behöver och förtjänar goda ordböcker. Det första för kommunikationens skull (härunder också svenskundervisning och språkvård), det andra för dokumentationens skull. I ordens ljus tonar samhället fram.

## Tionde upplagan av Svenska Akademiens ordlista

Språkvård 1973: 4.

Ett år före hundraårsminnet av första upplagens utgivning kommer Svenska Akademiens ordlista över svenska språket (SAOL) i sin tionde upplaga. Lindrigt talat har ordlistan under de gångna åren spelat en framträdande roll på riksspråkets ordscen. Sedan sjätte upplagan 1889 är den officiell norm för stavningen i skolorna. I vidare bemärkelse fungerar den som rikslikare för språkbrukarna när det gäller såväl uttal och böjning som stavning. Det är sålunda en bok som har stor auktoritet och som måste motsvara stora krav. Jag skall här ge en ganska utförlig resonerande presentation av den nya upplagan.

### Ordförråd

Ordlistor speglar i större eller mindre grad de språksamfund som de hänför sig till. SAOL har speglat allt vidare sektorer: de senare upplagorna är mångdubbelt större än de tidigare. Från nionde till tionde upplagan har omfånget emellertid inte ökat. Det ligger kvar i storleksordningen 150 000 ord. Man har enligt inledningen berett nya ord plats genom att stryka föråldrade ord och en mängd sammansättningar och genom att gallra bland facktermer och främmande ord (man har dessutom ökat spaltlängden med några rader). Härigenom har man också fått ökat utrymme för vardagliga och regionala ord och för betydelseförklaringar.

Det är viktigt att understryka det som redan har framgått, nämligen att ordlistan är ett urval ur den principiellt gränslösa ordmängd som står oss till buds. I inledningen betecknas SAOL med korrekt försiktighet som en fyllig exempelsamling. Låt oss genom ett på måfå valt exempel se hur urvalet denna gång har

fungerat i förhållande till föregående upplaga. Exemplet utgörs av ett uppslag i den nya upplagan: sidorna 222–223 (*kavitet-kjortelväld*).

Om vi först bortser från sammansättningarna, finner vi att följande ord har tillkommit: *kelig*, *kem-* (i *kemtvätt* osv.), *kemomat*, *kenyan*, *keso* 'ostmassa', *ketchup*, *kiasm* 'korsställning', *kibbutz*, *kick* (i uttrycket *på ett litet kick*), *kicka* 'liten flicka', *kinematografi* 'filmteknik', *kinotek* 'filmsamling', *kirra* 'ordna' och *kittlare*. Exemplet tycks ge en rätt god bild. Vi finner bland annat representanter för nya begrepp (*kemomat* m. fl.), nationalitetsord (*kenyan*) och vardagliga ord (*kicka* m. fl.). Morfologiskt sett finner vi de nu så aktuella ordleden *kem-* och *-(o)tek* företrädda. (I förbigående sagt har på sistone bland många andra *klippotek* och *byxotek* skådats.)

Strukna i tionde upplagan är betydligt färre ord: *KFUK-are*, *KFUM-are* och *kindergarten*. Av dessa har väl *kindergarten* konkurrerats ut av *lekskola*, som för övrigt fanns med också i förra upplagan.

Bland de ord som står kvar är åtskilliga av klart speciell karaktär. Exempel är *kediv* 'egyptisk vicekonung' (en historisk term), *kefir* 'kaukasisk jäst mjölk', *kelim* 'ett slags österländsk matta eller bonad' och *kerreffekt* 'ett optiskt fenomen' (en fysisk term).

Det finns anledning att också jämföra det lilla ordlistestycket med motsvarande parti i Nusvensk frekvensordbok (NFO). Detta lexikon representerar ju en djupexcerpering av en miljon ord. Bortsett från en rad namn och utländska ord förekommer där fem osammansatta ord som inte finns med i den nya upplagan av SAOL. De är *kentsk* (i fågelnamnet *kentska tärnor*), *khoreut* '(kor)dansare', *kissa*, *kiton* 'forngrekisk (under)klädnad' och *kitsch* 'okonstnärlig smörja'. Inget av orden är väl egentligen mer exotiskt än den kaukasiska jästa mjölkens beteckning, men jag vill inte göra gällande att alla fem borde ha varit med. *Kentsk* är emellertid inte bara fågelord och kunde gott försvara en plats. *Kitsch* är ungefär jämförbart med det medtagna *bosch*. Restriktiviteten mot en viss ordsfär borde knappast ha behövt drabba det oförargliga verbet *kissa*.

Av sammansättningar har det faktiskt tillkommit 35 och fallit bort nästan precis lika många (siffran beror något lite på hur man räknar). Det skulle tyda på att exemplet inte är representativt i detta avseende, eftersom man enligt inledningen har gjort en kraftig reduktion. Lika fullt är det naturligtvis av intresse att se på vad som har skett. Nyttillkomna är *kedjeband*, *-byte*, *-effekt*, *-fånge*, *-företag*, *-hus*, *-index*, *-krock*, *-rökare*, *-spelare*, *-varuhus*, *kejsarkult*, *kelsjuk*, *kembar*, *-toalett*, *-tvätt*, *-tvätta*, *-tvättmedel*, *kemiingenjör*, *kemikaliebeständig*, *kemoterapeutisk*, *keramikugn*, *kerrcell*, *kikhostepidemi*, *kilrem*, *kilohertz*, *-kem*, *-pond*, *kimono-skärning*, *kioskägare*, *kiseldiod*, *-plast*, *kissemåns*, *kittelflickare* och *kjortelsäck*. En del av dem är välmotiverade, t. ex. *kilrem* och *kilohertz*, men frågan är om huvuddelen hade behövt tas upp. Argumenten för utmönstring – och de borde också gälla som spärr för intagning – har nämligen varit, ”att sammansättningen har bedömts som självklar och därmed som oväsentlig eller att sammansättningens betydelsesfär är tillgodosedd genom närstående, etymologiskt samhörig sammansättningsled”. Betrakta mot denna bakgrund exempelvis *kedjebyte*, *kejsarkult* och *kioskägare*. Också av andra skäl kan man undra över *kjortelsäck*, som betecknas som både ålderdomligt och provinsialt när det nu införs.

Utgångna sammansättningar: *kedjemätning*, *-räkning*, *kejsartiden*, *-vansinne*, *kemiprofessur*, *kemikaliehandel*, *kikarrum*, *-stativ*, *-ställ*, *kikhostpatient*, *-serum*, *kilbenshåla*, *kilskriftsurkund*, *kilvad*, *killespelare*, *kilosäck*, *kilometerhög*, *kimonoklädd*, *kimröksfabrik*, *kinaman*, *kinavin*, *kindergartenarbete*, *-lärarinna*, *-seminarium*, *kinematografbild*, *-kamera*, *kinesögon*, *kioskhandlare*, *kirunabo*, *kirurgkongress*, *kiselsyrad*, *kisthandtag*, *kittelbotten*, *-lock*, *kjolhållare*, *-våd*. Inget av dessa ord förefaller lämna något större tomrum efter sig utom möjligen *kejsartiden*, som ju har en pregnant betydelse. När det gäller *kirunabo* är att märka, att regelbundna *bo*-bildningar till stadsnamn har utgått principiellt.

NFO redovisar på motsvarande ställe ett 30-tal sammansättningar utöver dem som upptas i tionde upplagan av SAOL. Dit hör just *kejsartid*. Dit hör bland andra också *kilklack* och

*kilometerkostnad*, som borde ha slagit ut många av dem som ordlistan har med. Dit hör vidare flera sammansättningar med *kines-* som helt saknas i ordlistan.

I inledningen framhålls det särskilt, att ofullständigheten i ordförrådet blir påtagligast när det gäller terminologin på olika fackområden. För att se hur det förhåller sig på en konkret punkt i det avseendet har jag gått igenom en rad kuranta termer på databehandlingens område. Jag har inte blivit utan fångst (termerna ges här utan betydelseförklaringar): *alfanumerisk, algoritm, allokerat, bildskärm, binär, buffra, datamaskin, dator, digital, kompatibel, kringutrustning, kärnminne, mikrosekund, minidator, nanosekund, operativsystem, radskrivare, realtid, uppdatat, virtuellt*. Ett par av dessa termer (*binär, virtuellt*) används även utanför datatekniken och fanns med redan i förra upplagan.

Bland termer som saknas i ordlistan är *access, assemblering, avkoda, back-up, byte* (uttalat [bajt]), *centralenhet, databas, dataskärm, dekrement, flödesplan, hårdvara, indata, inkoda, inkrement, input, interaktiv, kurvritare, lo(o)p, maskinvara, millisekund, mjukvara, modem, off-line, oktalt, on-line, output, plotter, processor, programvara, satsvis, styrenhet, subrutin, textskärm, tidsdelning, utdata*.

Ett allmänt intryck vid genomgången var, att det var svårt att gissa rätt när man försökte förutsäga, om ett ord skulle vara medtaget i ordlistan eller ej. Förhållandevis speciella termer som *kärnminne* och *realtid* finns där men inte allmänna termer som *maskinvara* (*hårdvara*) och *programvara* (*mjukvara*), *input* (*indata*) och *output* (*utdata*). Frågan är om det inte är ord av denna senare typ som i första hand bör tas upp. Konsultationer med specialister på olika fackområden skulle vara en styrka.

## Stavning

Några stora förändringar görs inte i stavningen, däremot rätt många små. Vi finner en viss övergång från *ä-* till *e-*stavningar och från *o-* till *å-*stavningar. Nionde upplagan hade exempelvis *lättra* el. *lettra*, *späns*, *tender* el. *tänder* 'kolvagn' och *verst* el.

värst 'rysk mil'. Tionde upplagan har *lettra* äv. *lättra*, *späns* el. *spens*, *tender* respektive *verst*. Nionde upplagan hade vidare bland annat *brock* el. *bräck*, *kora*, *molla*, *spont*, *voffla* el. *våffla* och *volm* el. *vålm*. Tionde upplagan har *bräck* el. *brock*, *kora* el. *kåra*, *målla* el. *molla*, *spont* el. *spånt*, *våffla* el. *voffla* respektive *vålm* el. *volm*.

Frågan om försvenskningar eller inte i stavningen är svår och alltid aktuell. Vi finner tendenser åt båda hållen i ordlistan. C-formerna sätts främst igen i *camouflera* osv. och *ou*-formerna i *boulevard* osv. (dock inte i *budoar*). I de båda exempelfallen har NFO någon övervikt för *k*-formerna och klar övervikt för *ou*-formerna. Alternativ som *rendevu* för *rendezvous* försvinner. *Wienerschnitzel* slår ut *vinersnitsel* (men *vellpapp* står kvar framför *wellpapp*).

Å andra sidan får vi alternativ dubbelteckning i *bagg*, *foxtrott*, *jett*, *pickupp*, *popp*, *pubb*, *smogg* (men inte i *set* och inte i t. ex. *ketchup*). Vi får också former som *jos*, *musikal*, *spreja*, *teja* och *tajma*. *Manikur* el. *manikyr* har blivit *manikyr* el. *manikur* (men *pedikur* el. *pedikyr* står kvar i denna ordning).

Relationen till uttalet är också en ständig diskussionspunkt. Flera av de exempel jag har givit visar på en anpassning efter uttalet. Ett annat fall är *teve* som alternativ till *TV* och *tv*.

## Uttal

När det gäller uttalsuppgifterna berörs i första hand två ganska stora ordgrupper, orden på betonat *-ans* och *-ant*. Principen anges så, att orden på *-ans* får alternativt *n*-uttal och orden på *-ant* *n*-uttal i första hand, *ng*-uttal i andra. På dessa punkter har emellertid oredan blivit stor. *Ans*-orden har som det vill synas på ganska tillfälliga grunder i själva verket fördelats på alla de fyra olika möjligheterna: *allians*, *renässans* osv. får bara (-angs'), *kverulans*, *substans* osv. bara (-ans'), *balans*, *distans* osv. får (-angs' el. -ans') och *instans*, *vakans* osv. (-ans' el. -angs'). På motsvarande sätt får *militant*, *vigilant* osv. bara (-ant'), *frappant*, *intressant* osv. (-ant' el. -angt') och *genant*, *nonchalant* osv.

(-angt' el. -ant'). Särskilt påfallande är sådana iakttagelser som att *tolerans* får (-angs' el. -ans') men *intolerans* (-ans' el. -angs'), *assonans* bara (-angs') men *dissonans* (-angs' el. -ans'). Inte minst dessa fall antyder att det inte är fråga om något systematiskt försök att från ord till ord återge det faktiskt förekommande uttalet.

Ett par andra drag kan urskiljas. Ett närmande till stavningen finner man i ordtypen *city*, *hobby* osv., som får alternativen (-i el. -y); för *jury* anges till och med (-y äv. -i). Ett tydligt exempel på skriftenligt uttal är *sensmoral*, som nu i första hand får den bokstavstroga uttalsuppgiften (*sensmora'l*).

Ett närmande till engelskt uttal av en del ord kan man också konstatera. Där *guide* tidigare hade (gäjd el. gajd) ges nu (gajd). För *steeplechase* har uttalet ändrats från (sti'peltjes) till (sti'peltjejs), för *bacon* från (bä'ken) till (bej'ken) osv.

## Böjning

På ordböjningen har man kopplat en rad ganska kraftiga grepp. Det skall bli intressant att se vad de kommer att leda till. Jag skall kommentera några av dem. Genomgående gäller det interferensproblem, där alltså andra vokabulärsystem är inblandade (främst det engelska och de klassiska).

Pluraländelsen -s tränger på bland annat i typerna *banjo* och *speaker*. Den förra typen behandlades heterogent i nionde upplagan: *jojoar*, *banjoer* el. *banjos*, *tornados* el. *tornader*, *studior* el. *studios* osv. NFO har mycket riktigt belägg på både *studior*-(na) och *studios*. Nu gör man en stor satsning på -r: *jojor*, *banjor*, *tornador*, *studior*, *desperador*, *geckor*, *korsor*, *radior* osv. Ändelsen -s är alltså borta, men full enhetlighet har man inte siktat på; det heter exempelvis *faraoner*, *impres(s)arier*, *torsor* el. *torsor*, *cellor* el. *celli* och *cembalor* äv. *cembali*. Som inhemska förebilder till r-pluralen framstår fall som *vallmor*, *utsagor*, *nabor* och kanske *hustrur*.

På liknande sätt gör man med typen *speaker*. Nionde upplagan hade i pluralis *speaker* el. *speakers*, tionde upplagan har



*speakrar* äv. *speaker*. Till den grupp som nu går som *speaker* med alternativa pluralformer hör *bestseller*, *folder*, *globetrotter*, *outsider*, *tanker* etc. Till en annan grupp som bara får *ar*-form förs *blazer*, *blunder*, *gangster*, *spinnaker*, *vajer* etc. Till en tredje grupp med oböjd pluralform förs *copywriter*, *designer*, *manager*, *partner*, *starter* etc. Också lån från tyskan ansluter sig till mönstret och får pluralformer som *besserwissrar* äv. *besserwisser*, *schlager* el. *schlagrar*, *strebrar* respektive *rottweiler*. I NFO dominerar *s*-former: *bestsellers*, *gangsters* (även *gangstrar*), *partners*, *tankers* osv.

Det kan kanske nämnas att pluraländelsen *-s* inte är helt försvunnen ur ordlistan. Ord som har den kvar är exempelvis *clip* pluralis *clips*, *paria* pluralis *parior* el. *parias*, *spaniel* pluralis *spaniels* och de som tillhör typen *cornflakes*, *jeans*, *shorts* osv.

Tre besvärliga ordgrupper inom den klassiska sfären är de som representeras av *centrum*, *narkotika* och *tema*. Orden på *-um* behandlas enligt principen att bestämd form singularis på *-umet* och obestämd form pluralis på *-um* ställs först bland alternativa former. För just *centrum* anges *centrumet* el. *centret* el. *centrum* respektive *centrum* el. *centrer* el. *centra*, för exempelvis *plenum* anges *plenomet* el. *plenum* respektive *plenum* el. *plena*. Detta tycker jag är en naturlig väg att gå, och den kan gås längre än vad som nu har skett. Ser vi till pluralerna, finner vi fortfarande exempelvis under *fluidum* pluralis *fluider* äv. *fluida*, under *memorandum* pluralis *memoranda*, under *serum* pluralis *sera* äv. *serum* och under *spektrum* pluralis *spektrer* el. *spektra*.

I gruppen *narkotika* gäller frågan framför allt erkännandet av en bestämd singularform på *-n*. I nionde upplagan var *kontentan* och *spirituosan* etablerade och *rekvisitan* angiven som ett andrahandsalternativ. Tionde upplagan behandlar dessa tre fall på samma sätt och godtar därutöver som andrahandsalternativ *narkotikan* och *petitan*. Detta ser jag också som ett steg i rätt riktning.

Problemet i gruppen *tema* är pluralformen. Föregående upplaga hade enbart den inhemska formen i fallen *komman* och *scheman*. Däremot gavs med alternativ *stigman* äv. *stigmata* och *teman* äv. *temata*. Under *trauma* gavs ingen pluralform och

*lemma* fanns inte med. Tionde upplagan ger enbart *n*-form i alla orden utom *stigma*, som behåller sina alternativ. Tendensen är alltså densamma som i de båda föregående grupperna. NFO har belagt *schemata*, *teman* och *temata*.

Innan jag lämnar böjningsformerna vill jag dra en lans för den bestämda formen *proffset*. Nionde och tionde upplagorna ger samstämmigt *proffsen* med pluralis *proffs*. Språkbrukets kuranta form är emellertid enligt min erfarenhet *proffset*. Ordet faller in i samma kategori som bland andra *fånet* och alla *råd*-orden: *arkivrådet*, *biblioteksrådet* (ej medtaget i SAOL), *departementsrådet*, *kommunalrådet*, *lagrådet*, *statsrådet* osv.

## Redigering

Det är självfallet ett mycket stort arbete att revidera en så omfattande ordlista som SAOL. Brister i den redaktionella bearbetningen kan lätt uppkomma, i synnerhet om man inte förfogar över rationella hjälpmedel. Med tanke på ordlistans viktiga funktion är det samtidigt betydelsefullt, att den är förebildlig också i tekniskt lexikografiskt avseende.

När jag har gått igenom ordlistan har jag antecknat ett antal tekniska brister, som jag anser mig böra exemplifiera i detta sammanhang. Dubbel bokföring har ordet *courtage/kurtage* fått. De båda variantformerna inleder var sin fullständiga artikel. Enligt ordlistans principer skall den senare artikeln utbytas mot en hänvisning: *kurtage* se *courtage*. Ett annat av stickorden är feltryckt: *bookmaker* har blivit *bokkmaker*. Här kan också två andra tryckfel nämnas: den första uttalsuppgiften under *annons* saknar tryckaccent och bruklighetsuppgiften under *klarligen* skall vara *åld.*, inte *ald*.

När det gäller uttalet skall jag peka på ytterligare några saker. En hänger samman med problemen kring orden på *-ans* och *-ant*. I inledningen ges *kondoleans* som exempel på ord som har fått alternativt *n*-uttal. I ordlistan förekommer emellertid bara uttalet (-angs'). Inkonsekvens i uttalsuppgifterna visar exempel-

vis de parallella orden *irakier* och *iranier*, som har fått (-a'-) respektive (-a'n-).

Fastän *irakier* och *iranier* har identisk böjning, redovisas de oenhetligt också på den punkten. Det ena får lodstreck, upp-  
reppningstecken framför bestämda artikeln och pluraländelse, det andra får inget lodstreck, divis framför bestämda artikeln och streck i pluralis: *irak|ier ~n -ier* respektive *iranier -n -*. Andra exempel på vacklande användning av lodstreck är *partn|er ~n -er* respektive *manager -n -* och *tornad|o -on -or* respektive *desperado -n -r*.

Hänvisningar är betydelsefulla i en ordlista, och SAOL har en hel del sådana. Olika typer av fel och inkonsekvenser förekommer också. Från *miljard* och *miljon* hänvisas till *milliard* respektive *million*, men det finns ingen hänvisning från *biljon* till *billion*. Det borde exempelvis också ha funnits hänvisningar från *räls* till *räl* och från *tidsrymd* till *tidrymd*. Variantformen *resymé* till *resumé* har utgått i den nya upplagan, men kvar står raden *resymé* se *resumé*, som alltså har blivit en tom hänvisning.

## Allmänna synpunkter

Till skillnad från de allra flesta böcker, och i varje fall vetenskapliga arbeten, är ordlistan anonym. Titeln Svenska Akademiens ordlista innebär ju inte att Svenska Akademien har skrivit den. Jag tycker det vore riktigtast att ange, vilka som har utarbetat manuskriptet (och vilka som i övrigt har bidragit med insatser av olika slag). Som jämförelse kan nämnas, att detta sker i Svenska Akademiens Ordbok över svenska språket (SAOB).

En annan vedertagen princip som ordlistan inte ansluter sig till är den att ange vilket material som arbetet bygger på. Det skulle vara av betydelse att bland annat få veta hur materialet har insamlats, vilka texttyper som har beaktats, vilket omfång materialet har och vilken period det spänner över.

En teoretisk fråga av stor vikt gör sig också påmind. Vad det gäller är ordlistans inriktning. Syftet anges i inledningen vara "att förteckna ett i allmänt språkbruk gångbart ordförråd". Formu-

leringen innebär en beskrivande uppgift: att ta reda på vilka ord som främst utnyttjas i språksamfundet och redovisa dem så som de används. Om detta syfte hade genomförts, skulle man säkerligen inte ha tagit upp exempelvis ett ord som *kjortelsäck* eller en pluralform som *jojor*. Att enheterna finns med beror på att ordlistan också har ett korsande syfte, nämligen att förteckna ting som "kan anbefallas till allmännare bruk". Detta är en föreskrivande uppgift.

I denna motsättning mellan en deskriptiv redovisning av bruket och en normativ strävan att påverka bruket ligger det teoretiska grundproblemet. Man kan avläsa en av konsekvenserna däri, att ordlistan som nämnt rör sig både i riktning mot och i riktning från försvenskningar. En närmare diskussion av syftet och principerna för normeringen skulle vara värdefull. En sådan skulle också underlätta för användarna att ta ställning till nyheter som dyker upp i språkbruket mellan upplagorna.

Tidsavståndet mellan de olika upplagorna var i början ganska litet. Under perioden 1874–1889 kom sex upplagor. De fyra följande markeras av åren 1900, 1923, 1950 och 1973. Den första tidens takt med några få år mellan upplagorna skulle vara alltför hög i nuläget för en bok av SAOL:s typ. Den nuvarande takten med omkring 25 års mellanrum är å andra sidan i lägsta laget med tanke på den stora rörligheten i ordförrådet. Någonstans mitt emellan borde kunna vara lagom.

Ytterligare en sak som gäller relationen mellan de olika upplagorna är av principiellt intresse. Jag har framför mig två exemplar av nionde upplagan, det ena tryckt 1950, det andra 1965. Vill jag nu ta reda på om orden *motorväg* och *polio* finns med i nionde upplagan, blir svaret nej om jag slår i årgång 1950 men ja om jag slår i årgång 1965. Det görs alltså ändringar inom ramen för en och samma upplaga, och därom sägs ingenting i den inledande texten. Att göra ändringar av detta slag kan i och för sig vara motiverat. Det vore emellertid önskvärt att de redovisades genom en notis i anslutning till inledningen.

En känd svårighet för alla som slår i ordlistan är sorteringsordningen. Huvudprincipen är alfabetisk ordning. Från denna princip undantas sammansättningar, som bokförs under de ord

som bildar förled. Härigenom får begreppet alfabetisk ordning en flertydighet, som bland annat påverkar tolkningen av anvisningen att avledningar står på sina alfabetiska platser.

I själva verket kan man urskilja flera olika typer av alfabetiska serier i ordlistan. Högst rang har de artikelinledande huvudstickorden, som bildar en strikt alfabetisk följd: *klar*, *klara*, *klarinet*, *klarligen*, *klarning* (för att ta ett exempel från sidan 224). Därefter kommer en utvidgad serie som inbegriper vissa bisticord inuti artiklarna (med eventuella avledningar): *klar*, *klara*, *klarera*, *klarhet*, *klarinet*, *klarinet*, *klarinet*, *klarligen*, *klarna*, *klarning*. I denna utvidgade serie är så ett antal slutna serier insprängda: *klar-klarögdhet*, *klara-klarskinn*, *klarhet-klarhetskrav*, *klarinet-klarinet*. Härigenom kommer exempelvis *klar* före *klarera*, *klarhet*, *klarinet* etc. och efter *klarsynt*, *klarögd* etc. Den förvirring som lätt uppstår är enligt min mening en nackdel av större format än den fördel det innebär att hålla sammansättningar samlade.

Att märka är vidare att ordbildningssystematiken bara gäller just sammansättningarna. Avledningarna placeras i vad jag har kallat den utvidgade alfabetiska serien. I exemplet kommer därför *klarhet*, som är bildat av adjektivet *klar*, inte i den artikel som har *klar* som huvudstickord utan i den som inleds av verbet *klara*. Detta morfologiska hälftenbruk tror jag också ger upphov till svårigheter och missförstånd.

SAOL har framställts utan hjälp av databehandling. Man kan emellertid vinna väsentliga fördelar i flera avseenden genom att ta metoder från detta område i sin tjänst. En datamaskinell bearbetning av verkligt stora textmängder ökar bland annat säkerheten i urvalet av ord och bedömningen av stavningsvarianter och böjningsformer. En datorbaserad framställning av själva ordlistan kan bland annat eliminera skönhetsfläckar av typen inkonsekvenser, tomma och uteblivna hänvisningar osv. Redan en finalalfabetisk (baklängesordnad) sortering av ordlistan före tryckningen ger utmärkta möjligheter att öka konsekvensen (själv har jag vid min genomgång av ordlistan haft stor hjälp av det finalalfabetiska lexikonet i NFO 2). En mycket smidig uppdatering av ordlistan blir också möjlig.

\*

Min slutsynpunkt är att tionde upplagan av Svenska Akademiens ordlista är en god bok som kan förbättras. När det gäller en riksangelägenhet som denna bör det vara möjligt att finna en form för ändamålsenlig samverkan mellan olika krafter.

## Stavning

Tre uppsatser om stavning. Ord och stil 5. 1973.

Trots rubrikens lakoniska formulering innehåller den här uppsatsen naturligtvis långt ifrån allt som är att säga om stavning. Jag tar upp en del frågor som är av principiell betydelse och väver samman trådar som jag åtminstone till en del har tvinnat tidigare. Framträdande inslag markeras av stickord som norm, reform och språkvård.

Skriften är en del av den språkliga strukturen. Beskrivningen av skriftsystemet blir alltså en del av språkbeskrivningen. Det lingvistiska studiet av skrift betecknas som grafonomi. Området utgör en motsvarighet till talsidans fonologi. Tillsammans täcker grafonomin och fonologin språkets uttryckssystem.

För att närmare sätta in stavningen i sitt sammanhang nämner jag inledningsvis de fem olika aspekter på skriftstudiet som presenteras i min bok *Introduktion i grafonomi*. De relationella aspekterna gäller förhållandet mellan tal och skrift, de strukturhistoriska aspekterna skriftens utveckling från bildmässiga förstadier till avancerad alfabetisk skrift, de analytiska aspekterna studiet av skriften som ett system av distinktiva enheter (grafem), de generativa aspekterna skriftens plats i en explicit helhetsbeskrivning av språket och de kvantitativa aspekterna det faktiska utnyttjandet av enheterna och reglerna i skriftsystemet. Det som i första hand blir aktuellt här är de relationella aspekterna.

Man kan naturligtvis se relationen mellan tal och skrift från två håll. Utsiktspunkten kan vara talet eller skriften. Man kan vidare lägga två synpunkter på relationen, en beskrivande (deskriptiv) och en föreskrivande (normativ). Vi får följande schema:

	Tal (fonem- system)	→	Skrift (grafem- system)	Skrift (grafem- system)	→	Tal (fonem- system)
Beskrivande (deskriptiv) aspekt			<i>Fonografemik</i>			<i>Grafofonemik</i>
Föreskrivande (normativ) aspekt			<i>Ortografi</i>			<i>Ortofoni</i>

Terminologin är som synes uppbyggd kring de enkla elementen *fon* och *graf* kompletterade med *orto-* 'rätt'.

Ett fonografematiskt studium kan till exempel gälla hur en viss ljudförbindelse faktiskt återges i skrift. Förbindelsen *s* + kort *e/ä* + *t* motsvaras ju i svenskan exempelvis av *sätt*, *sett*, *set*, *Set*, *Seth*, *Zeth*, [*et*]*cet*[*era*] och [*pin*]*cett*. Det kan till exempel också gälla hur man i privatbrev eller varför inte akademiska uppsatser återger svagtonig vokal i ord av typen *genitiv/genetiv*, *medicin/medecin* eller lång konsonant i ord av typen *artikell/artickel*, *kapitell/kapittel*. Stavningsvariationen i dokument från äldre tid är ett stort fält som kan studeras från den här synpunkten.

Ett grafofonematiskt studium å andra sidan kan gälla vilka uttal ett visst skrivtecken, *g* eller *j* eller *ö* osv., eller kombinationer av skrivtecken motsvaras av. Andra frågor kan gälla förekommande uttal av en ordtyp som *patient*, *spatiös* eller enskilda ord som *acceptera* (*aksepteralasseptera*) eller *resurser* (*resusjer/resjusjer*).

Det normativa sättet att närma sig skrift-talrelationen – ortofoni – innebär, att man föreskriver att *acceptera* skall uttalas *akseptera*. På motsvarande sätt innebär det normativa sättet att nalkas tal-skriftrelationen – ortografi – att man slår fast att den rätta stavningen är *medicin*. Det är denna sistnämnda relationella aspekt vi skall ägna oss åt.

## Normen

När man föreskriver att en viss stavning är korrekt, sker det i anslutning till en norm. Det blir därför aktuellt att fråga, vilken normen är och hur den ter sig.

För de enskilda ordens vidkommande har det blivit praxis att betrakta Svenska Akademiens ordlista som normerande när det gäller svenskan. Skolorna har för sin del att rätta sig efter ett departementsbrev i saken till Skolöverstyrelsen 1951. Ett svar på frågan om normen är alltså att den är kodifierad i denna ordlista. Svaret är emellertid föga tillfredsställande. Det intressanta är vilka övergripande karakteristika normen har.



Ett viktigt drag är att det finns vissa allmänna konventioner för ljuds motsvarigheter i skrift: ett uttalat *a* återges med skrivtecknet *a* osv. Man får anta att det här rör sig om relativt stabila grundprinciper, som alltså endast långsamt förskjuts under tidernas lopp. Tvåra kast av det slaget att ett uttalat *a* vid en viss tidpunkt skulle börja återges med exempelvis *f* är mycket osannolika. Lika viktigt som det är att framhålla dessa tal-skriftkonventioner, lika viktigt är det att erinra om att det inte generellt är fråga om några enkla och lättfångade relationer. Det tidigare exemplet *sätt*, *sett*, *set*, *Set* etc. kan tjäna som illustration. Klart utanför vad det svenska systemet tillåter faller å andra sidan exempelvis en uttalsform *fitjö* e. d. samhörig med en skriven form *feature*. I och för sig skulle vägen vara öppen för ett uttal som rimmade på *skreatur* sammanhållet med en skriftform *featur*.

Besläktat är ett annat utmärkande drag: det finns ganska bestämda regler för hur ljuden respektive skrivtecknen får kombineras med varandra, fonotaktiska respektive grafotaktiska regler. Medan *drag* är bildat efter reglerna, skulle ett *dgar* falla utanför. Samspelet mellan tal- och skriftreglerna är komplicerat på den här punkten också. En skriftform som *sjung* uppfyller de grafotaktiska reglernas krav, men en uttalsform *s-j-u-n-g* står inte i överensstämmelse med de fonotaktiska reglerna. Att stavningen (tal-skriftrelationen) är korrekt hänger samman med de konventioner som nämndes i förra stycket – *sj* är en vedertagen återgivning av *sj*-ljudet, *ng* av *ng*-ljudet.

Man kan urskilja fler element i stavningsnormen. De hittills behandlade ligger på det fonologisk-grafonomiska planet. Närmast kommer det morfologiska i fråga. Skriftformen *piggt* i stället för *pikt* – jämför *sikt* eller *pikter* och *piktur* – motiveras inte av sambandet med dess uttal utan av sambandet med dess grundform *pigg*. Morfemet *pigg* får en enhetlig stavning. Det är denna princip som Esaias Tegnér d.y. på sin tid kallade samhörighetslagen. Andra exempel är *snabbt* till *snabb*, *ljuvt* till *ljuv* och *grävt* till *gräva* samt *tisdag*, *onsdag*, *torsdag* (inte *tistag* osv.) efter *söndag* etc. och *dag*. Före stavningsreformen 1906 gällde principen i större utsträckning än nu: *god* – *godt*, *kallad* – *kalladt* osv. Till den aktuella typen i vidare mening hör fall där form-

varianter sammanhålls av återkommande grafem: *skjuta – skott – skytt, lysa – ljus, jacka – jackett* osv.

Ett annat inslag i normen bildar skriftformer som i stället för samhörighet markerar åtskillnad. Vad det gäller här är homofoner, olika ord som har samma uttal. Exempel är serier som *jord – gjord – hjord* och *käck – tjeck – check* och ordpar som *jon – hjon, män – men* och *allvar – alvar*. Eftersom stavningen här skiljer ord från ord, kan vi konstatera att vi rör oss på det lexikologiska planet. I praktiken får morfologiskt betingade former som *smällt* av *smälla* gentemot *smält* av *smälta* och *skyllt* av *skyllo* gentemot *skylt* (substantiv) samma effekt. Exempel ger också vår ordgrupp *sätt, sett, set* osv.

Hit hör vidare namn visavi motsvarigheter utan versal: förnamn som *Arla* mot *arla*, *Britten* mot *britten*, *Blända* mot *blända* och *Jack* mot *jack*, efternamn som *Barmark* mot *barmark*, *Flygare* mot *flygare* och *Överman* mot *överman*, ortnamn som *Morgongåva* mot *morgongåva*, *Nacka* mot *nacka* och *Rök* mot *rök*. Genom bruket av versal undviks sammanfall i en rad exempel där tryckaccent eller tonaccent skiljer uttalsformerna åt: *Urban* mot *urban*, *Japan* mot *japan* respektive *Halvdan* mot *halvdan*, *Halvor* mot *halvor*, *Sjöman* mot *sjöman* osv. Fall som *Tom* mot *tom* och *Alvar* mot *alvar* kan också hållas isär med versalens hjälp.

Inom egennamnen som grupp kan den åtskiljande stavningen i viss utsträckning sägas övergå i dekorativ stavning. Med enkla medel kan man åstadkomma rik variation. Följande lilla uppställning genererar 324 olika former av namnet *Kronskog*.

K	r	o	n	s	k	o	g
C		oo			c	oo	gh
Ch		ou			ch	ou	
		oh					
		ooh					
		ouh					

Den mest utstuderade varianten är *Chrouhnschough*.

I utkanten av normen rör sig ofta den kommersiella stavningen. Ett av syftena med valet av firma- och varunamn kan vara att fånga uppmärksamheten genom ett eller annat knep. Bland

mycket annat har vi fått tvållösningen *Bliw*, tvättmedlet *Jör* och den grafiska industrin *Wezäta*. Dragningen mot former med *c*, *q*, *w*, *x* och *z* är tydlig.

Till normen hör bland annat också konventioner för förkortning, sammanskrivning och särskrivning, användning av logogram (ordtecken som § och +) och syntaktisk markering med hjälp av interpunktionstecken och versaler. För vårt resonemang kan det räcka att betrakta de enskilda oförkortade orden.

I stavningsnormen sådan som den föreligger kan vi sammanfattningsvis urskilja tre huvudprinciper:

fonologiskt orienterad (ljudinriktad) stavning	<i>flykt</i>	substantiv
morfologiskt orienterad (forminriktad) stavning	<i>flyggt</i>	adjektiv neutrum
lexikologiskt orienterad (ordinriktad) stavning	<i>Flyckt</i>	egennamn.

Tillståndet återspeglar tre olika behov hos språkbrukarna: att hålla ihop uttalsform och skriftform, att hålla ihop samhöriga ordled och att hålla isär olika ord. Vårt stavningsskick utgör en kompromiss som har vuxit fram under århundradenas lopp.

### Stavningsreform?

Skulle de tre principerna kunna reduceras till en genom en reform? Låt oss se på några konsekvenser.

En ljudinriktad stavnings mål skulle vara att troget återge uttalet: ett ljud, ett skrivtecken (eller snarare en beteckning). Att bestämma en beskrivningsnivå möter emellertid stora svårigheter (vad är ett ljud?): skall man återge t. ex. olika r-varianter, hur troget skall man återge assimilationer, skall man erkänna ljudreduktioner etc.? Vi skulle kunna få exempelvis *mimm bro(sj) slykkell* för *min brors cykel* och *minn sysste(sj) sjakks* för *min systems sax*. Vi skulle kunna få *natturlittvis* eller *nattutvis* eller *nattvis* eller *nattus* för *natturligtvis* osv. Det förefaller mig helt klart att en fonetisk stavning dels har stora kommunikativa nackdelar, dels inte kan genomföras konsekvent.

Men skulle inte en ljudinriktad stavning kunna baseras på en fonematisk analys i stället för på den fonetiska realisationen: ett

fonem, ett skrivtecken (en beteckning)? Man fick i så fall dels bestämma sig för en viss fonemanalys i den rad av förslag som finns, dels föra ut den till medborgarna. Man skulle komma i ett läge där envar måste kunna avgöra att exempelvis en given ljudförbindelse *brosj* skall stavas, dvs. transkriberas, *brors* (om detta är den officiella analysen), medan ett *bråsj* skall återges till exempel *bråsj* och ett *sjykkell* till exempel *sykkell*. En grundförutsättning skulle vara, att alla dessförinnan undervisats i ett gemensamt fonemsystem, eftersom variationer förekommer i olika landsdelar. Hela konstruktionen är alltså samtidigt krävande och bräcklig, delar många av den fonetiska inriktningens svagheter.

Den morfologiska principens formulering blir: ett morfem (ordled), en skriftform. Frågan om vad som skall menas med ett morfem aktualiseras i minst lika hög grad som när det gäller fonem. Konsekvent tillämpad skulle principen inte bara ge *flygg* – *flyggt* utan exempelvis också *lägga* – *läggade* (uttalat *lade* eller *la*) – *läggat* (uttalat *lakt*) och *ligga* – *liggade* (uttalat *låg*) – *liggat* (uttalat *legat*). I exemplen har principen tillämpats också på böjningsmorfemen. För substantivens pluralbildning kunde resultatet bli *stenar*, *rosar* (uttalat *rosår*), *diktar* (uttalat *dikter*), *hjärtar* (uttalat *järtan*) och *husar* (uttalat *hus*). Principen för sålunda talet och skriften mycket långt från varandra, vilket naturligtvis inte alls är önskvärt.

För den lexikologiska principen gäller: ett ord, en skriftform. Definitionen av själva det språkliga begreppet ord är inte lättare än i föregående fall. Vi skiljer på *kål* och *kol* och borde enligt den lexikologiska principen skilja också på exempelvis trädet *ask* och lådan *ask* (*asck?*). Vi skiljer på *jord*, *gjord* och *hjord* men skulle egentligen ha goda skäl att också skilja på till exempel *jord* som beteckning för planeten, *jord* som beteckning för marken (*ljord?*) och *jord* som beteckning för myllan (*djord?*). Man kan med fog ifrågasätta om principen är genomförbar annat än på mycket lösa grunder.

En reform efter någon av de angivna principerna skulle alltså stöta på stora teoretiska och praktiska svårigheter (jag har inte nämnt alla). Resultatet skulle också bryta sig mycket starkt mot

den rådande stavningen. Detta är två faktorer som pekar mot en återhållsam inställning.

Idealisk är visserligen knappast det rätta ordet att beskriva normen med. En kompromiss är en kompromiss. I den rådande situationen har den emellertid några egenskaper, som kan skrivas på pluskontot. Den är väl inarbetad och fungerar kommunikativt över hela språkområdet. Den är också väl dokumenterad i ordlistor och andra publikationer. Den är vidare bärare av mycket stora textmängder, litterära och icke-litterära. Och den är relativt stabil, vilket ger god kontakt längs tidsaxeln.

Tar man då hänsyn till de negativa konsekvenserna av en reform enligt någon av de principer vi har urskilt och till de positiva sidor som alltså finns, kommer man fram till följande slutsats. Så länge vi inte har något alternativ som vi med säkerhet vet är klart överlägset vårt nuvarande stavningsskick, bör vi inte satsa på en reform. För egen del vill jag gärna tillägga att skriftsystemet, som spelar en så stor roll i samhällets och de enskilda människornas liv, inte utan vidare bör handhas med mindre varsamhet än andra delar av den språkliga strukturen.

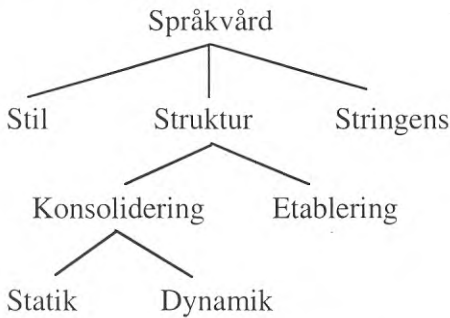
Den viktiga frågan blir alltså inte, hur vi skall gå tillväga för att göra om stavningen, utan hur vi skall förvalta den tillgång den representerar. Detta är en uppgift för språkvården i vid mening. Redan termen ortografi, det vill säga rättskrivning, anger ju för övrigt detta.

## Språkvård

Språkvårdens roll i sammanhanget kan belysas med hjälp av en analys av språkvårdsfunktionen (se min artikel Dynamisk språkvård). Man kan urskilja de komponenter som framgår av trädet på följande sida.

Skriver man *ordförandens tillfredsställande uttalande rörande hittillsvarande åtagandens förverkligande*, får man räkna med dåligt rykte som stilist. Skriver man *den tappra franska armén drog sig skyndsamt tillbaka*, blir det anmärkning i fråga om stringensen. Båda aspekterna är viktiga, men central är vården av

strukturen (från det fonologiska och det grafonomiska planet till det syntaktiska). Som har framgått av det tidigare resonemanget gäller det i vårt fall inte att etablera en strukturnorm – något som är aktuellt i många andra länder – utan att konsolidera den vi redan har. Det arbetet har både en statisk och en dynamisk del. Vi måste å ena sidan ägna oss åt att utforska och dokumentera den, å andra sidan utforma riktlinjer för hur vi skall ställa oss till nyheter av olika slag, t. ex. nya ord, nystavning av accepterade ord, nya konstruktioner. Kärnpunkten blir den dynamiska aspekten på strukturens konsolidering.



Min tanke är att man bör grunda varje resolution på ett antal strukturvillkor och ett antal plusvärdesvillkor. Med hjälp av strukturvillkoren testas nyheten mot strukturen med avseende på uttal, stavning, böjning och bildning (i vid mening) samt icke störande sammanfall med befintliga enheter. Alla strukturvillkoren måste obligatoriskt uppfyllas. För att det skall vara någon mening med att acceptera en nyhet måste den också representera ett plusvärde. Vilka plusvärdesvillkoren skall vara är en mycket viktig fråga. Tentativt har jag uppställt fyra villkor: begreppsmässigt, internationellt, funktionellt och socialt plusvärde. Här behöver kanske bara det funktionella kommenteras något. Beteckningen är en sammanfattning av sådant som exempelvis ökad smidighet (i ordbildningen) och ökad tydlighet (uppfattbarhet).

Tänkbart vore att den generativt inriktade skriftspråksforskningen efterhand skulle kunna ge oss ytterligare ett plusvärde att åberopa. Vad jag avser är egenskapen hos olika stavningar att

framstå som mer eller mindre naturliga. Resonemang av detta slag måste ha ett starkt empiriskt underlag, som vi för närvarande inte förfogar över. En svårighet är också att de grafonomiska reglernas utformning, plats och roll i en generativ beskrivning är alldeles otillräckligt utforskade. Det är viktigt att den diskussion som förs går vidare.

Det framgår att språkvårdsmodellen ger utrymme för kontrollerade modifikationer i förbättrande riktning. Detta svarar mot kravet på dynamik. Det svarar också, menar jag, mot det krav på förvaltning av normen som vi tidigare kom fram till.

När detta skrives utkommer tionde upplagan av Svenska Akademiens ordlista. I enlighet med utgivarnas strategi var det länge sedan sist, närmare bestämt 23 år. Det är alltså fråga om en lika ovanlig som viktig händelse. Mot den bakgrund som jag har skisserat är det naturligt att som exempel ta upp och diskutera en del nyheter i ordlistan.

Utan att förlora stavningen ur sikte skall jag först belysa resonemanget kring det funktionella plusvärdet med hjälp av böjning respektive uttal hos två ordtyper. Pluralböjningen av främmande ord som slutar på obetonat *-er* har länge vållat besvär. Det är bland annat en av de punkter där en tendens till en ny deklination med plural på *-s* har varit märkbar. SAOL anlägger nu moteld genom att helt utesluta formtypen *gangsters, tankers* osv. I stället anger man plural på *-ar* och/eller plural utan ändelse (som man kallar plural på *-er*). På detta sätt omhändertas för övrigt inte bara de engelska lånorden utan också de tyska av samma typ (*besserwisser, schlager* osv.). Man erhåller en huvudgrupp av ord som bara får *-ar*: *blazrar, bunkrar, gangstrar, jumprar, lasrar, libhabrar* ('älskare av ngt'), *skotrar, sluggrar, strebrar, supportrar, teleprintrar, vajrar* etc. Här har försvenskningen enligt utgivarnas bedömning gått längst. Mindre långt bedöms den ha gått i en annan stor grupp som får *-ar* äv. noll: *besserwissrar/besserwisser, bestsellrar/bestseller, blinkrar/blinkrar, bookmakrar/bookmaker* (i ordlistan har uppslagsformen råkat bli *bokkmaker*), *földrar/folder, racrar/racer, speakrar/speaker, tankrar/tanker* osv. En tredje grupp får – väl huvudsakligen av fonetiska skäl – bara noll: *designer, manager, partner, sprinkler*

m. fl. (men pluralen av *gletscher* 'jökel' ges å andra sidan som *gletschrar*). För sig självt står *schlager* som får noll el. *-ar*. De olika undergrupperna till trots skapas genom ordtypens behandling en betydande regularitet, och den kan ses som ett funktionellt plusvärde.

När det gäller substantiven på betonat *-ans* nämns det i förordet till ordlistan, att alternativt *n*-uttal har införts. Som exempel hänvisas till *balans*, *kondoleans* och *nyans*. Ordlistan ger för det första och sista av dessa ord *-angs* el. *-ans* men för *kondoleans* bara *-angs*. Nystar man upp ordgruppen, finner man i själva verket alla de fyra möjliga uttalsangivelserna väl företrädda. En stor del får *-angs* el. *-ans* (i några fall *-angs* äv. *-ans*): förutom de två nämnda exempelvis *ambulans*, *elegans*, *ordonnans*, *relevans*, *stimulans*. En nästan lika stor del får *-ans* el. *-angs*: *dominans*, *instans*, *konkordans*, *leverans*, *romans*, *vakans* osv. Till den förra gruppen förs *tolerans*, till den senare *intolerans*. Åtskilliga får bara *-angs*: förutom *kondoleans* t. ex. *allians*, *assonans*, *konvenans*, *renässans*, *resonans*. En del slutligen får bara *-ans*: *konstans*, *kverulans*, *substans* m. fl. Här tycker man sig kunna konstatera att utgivarna har försuttit tillfället att åstadkomma en funktionell regularitet av åtminstone samma omfattning som i föregående fall. Den frågan uppställer sig också, på vilka grunder fördelningen av orden på de fyra grupperna har skett.

Några exempel på ändringar i stavningen kan ställas i relation till det internationella plusvärdet. Ordlistan uppvisar, och säger sig uppvisa, en viss tendens mot stavningsformer som överensstämmer med eller ligger nära formerna i främmande språk. Man ger nu *campa* el. *kampa*, *cancer* el. *kancer* osv. med *c*-formerna som huvudformer. Man ger *boulevard* el. *bulevard*, *souvenir* el. *suvenir* osv. med *ou*-formerna främst (å andra sidan ges *budoar* el. *boudoir*). Man ger *madeira*, *kliché*, *resumé* och *sherry* utan de tidigare alternativformerna *madera*, *klisché*, *resymé* och *scherry*. Man ger exempelvis också ordningen *guide* äv. *gid* och har inte längre *vinerbröd* som alternativ till *wienerbröd*.

Å andra sidan sker en rörelse från formerna i de främmande språken. Man ger nu *foxtrot* el. *foxtrott*, *jet* el. *jett*, *pop* el. *popp* osv. med alternativ dubbelteckning. Man placerar *jos* och *musi-*



kal före alternativen *juice* och *musical*. Man ger t. ex. också *poäng* utan alternativet *point* (men samtidigt *poängtera* äv. *poin-tera*). I stavningsändringarna framträder alltså en ambivalent inställning till vad jag kallar det internationella plusvärdet.

De behandlade fallen visar bland annat på behovet av ett klart utformat teoretiskt underlag för språkvårdsarbetet. Stavningen i betydelsen rättskrivning är ett led i detta större sammanhang. Vad man skulle önska är en öppen principdiskussion. Fruktbarast kunde jag tänka mig att denna skulle bli inom ramen för ett gemensamt mål: en varsam förvaltning av språknormen.

## Litteratur

- Allén, S., Dynamisk språkvård. I: Dagens Nyheter 11.6.1968.
- Allén, S., Förhållandet mellan skrift och tal. I: Allén m. fl.: Språk, språkvård och kommunikation. 1967. S. 76–92.
- Allén, S., Grafematisk analys som grundval för textedering. 1965.
- Allén, S., Introduktion i grafonomi. Det lingvistiska skriftstudiet. Under medverkan av Staffan Hellberg. 1971.
- Hellberg, S., Graphonomic rules in phonology. Studies in the expression component of Swedish. [Avhandling (Göteborg) som beräknas föreligga 1974. ]
- Hultman, T.G., Skriftspråk och ortografi. I: Språket i blickpunkten. 1969. S. 132–149.
- Ståhle, C.I., Det nordiska rättstavningsmötet 1869 och hundra års svensk rättstavning. I: Språkvård 1969:4. S. 2–12.
- Teleman, U., Stavning och fonologi. I: Teleman, U., Tre uppsatser om grammatik. 1972. S. 7–32.

## Denominala verbalfraser

Svenskans beskrivning 5. 1972.

### 1. Strukturell bakgrund

I sin skrift *Reklamens språk* (1950) skriver Erik Wellander (s. 25): "En nyhet är ... den i reklamen ofta förekommande framflyttningen av ett adverb: *Benäget observera att detta vårt meddelande sålunda icke utgör någon uppmaning att verkställa fyllnadsinbetalning. Kemiskt tvätta i våra förstklassiga etablisserment!* Denna ordföljd är alldeles främmande för gott svenskt språkbruk och borde kunna undvikas."

Jag skall ge ytterligare några exempel. (1) *Stoppade möbler kemiskt tvättas utan att tyget avtages.* (Rikstelefonkatalogen, Göteborgsdelen 1951.) (2) *Tvärtom ökar livslängden på kläderna om man kemiskt tvättar dem ofta.* (Göteborgs-Tidningen 20/11 1955.) (3) [*Detta sker*] i samma takt som den [dvs. oljehanteringen] *ekonomiskt utvecklas.* (G. Littke i TV 24/1 1968.) (4) *Medan hans [dvs. hr Hedlunds] partivänner i regeringen äro sysselsatta med att skriva MU-brev till dr Adenauer, är han på egen hand utan något som helst anlitande av Frank Buchmans patent i färd med att moraliskt upprusta den krassliga stockholmopolisen.* (Göteborgs Handels- och Sjöfarts-Tidning 29/6 1955.) (5) *Nu kan de färdigutbildade hushållslärarna ägna sig åt att: — — — Vetenskapligt forska / På institut och laboratorier där man genom experiment och prov söker få fram kvalitetsprodukter till hus och hem.* (Idun 1960: 7.)

Närmast skisserar jag den strukturella bakgrund mot vilken konstruktionerna avtecknar sig. Jag gör detta med utgångspunkt i Natanael Beckmans *Svensk språklära*. Huvudregeln för adverbialens position i huvudsatser har Beckman placerat på en undanskymd plats, nämligen i Märk 3 sist i § 225 som behandlar hypotaktisk ordföljd, underordningsordföljd. Regeln lyder: "I prosan stå endast särskilda satsadverbial, som sluta sig till pre-

dikatet, före predikatet i huvudsats. Ex. *Han rent av häpnade. Jag bara läser. Hon till och med rodnade.*”

Vad Beckman menar med särskilda satsadverbial framgår förutom av de här givna exemplen av § 171, moment 6: ”... de sluta sig omedelbart till den satsdel, som är viktigast för sammanhanget. Denna är då tryckstark, medan adverbet självt oftast är trycksvagt.”

Enligt § 225 gäller vidare att vissa bestämmningar som i huvudsats står efter ett finit verb placeras före verbet i bisats. Sådana bestämmningar kallar Beckman rörliga bestämmningar. Han skiljer därvid mellan alltid rörliga, nästan alltid rörliga och oftast rörliga adverb. Alltid rörliga är nekande adverb. Exempel: *Du får inte gå ännu. Vad man inte kan vrida, måste man lida.* Nästan alltid rörliga är allmänna satsadverbial, särskilt adverb som antingen betecknar satsens visshetsgrad (t.ex. *antagligen, bevisligen*) eller uttrycker ett omdöme av annat slag om satsens innehåll (t.ex. *dessbättre, tyvärr*). Exempel: *Så långt kan bladet väl antagligen hava rätt. Jag kan det inte nu, ehuru jag möjligen kunde det i går.* Nästan alltid rörliga är också adverb som betecknar en yttre eller inre egenskap hos det personliga subjektet för verbets handling (t.ex. *andäktigt, artigt, beslutsamt, blygt*). Exempel: *Han var omgiven av en skara, som andaktsfullt lyssnade på hans ord.* Oftast rörliga är vissa mera obestämda tidsadverbial (*alltid, dagligen, ständigt*). Exempel: *Ord, som alltid varit tvåstaviga. Hans dikter äro oftast skrivna för bröd.*

För infinitiven gäller enligt § 129 att den antar samma slags bestämmningar som det finita verbet. Vidare gäller att dessa bestämmningar placeras på samma sätt i förhållande till infinitiven som det finita verbets bestämmningar i bisats. Om infinitivmärke förekommer, står detta i sin tur framför framförställda adverbial. Därvid uppkommer alltså den ordföljd som kallas omfattning (§ 129, § 222). Exempel (från Beckman): *Att inte försäga sig. Att aldrig tala osanning. Att alltid och gärna hjälpa sin nästa.* När adverbialet inte är av typen rörlig bestämmning slutligen, får vi efterställning vid infinitiv likaväl som normalt i huvudsats. Exempel (§ 126:3): *Att skriva tydligt är att skriva väl.*

Enklast kan man kanske sammanfatta läget genom att sätta in Beckmans, Wellanders och mina exempel i positionsscheman av det slag som återgår på Paul Diderichsens *Elementær dansk Grammatik* (s. 186). I tabell I förtecknas huvudsatserna, bisatserna och infinitivsyntagmerna var för sig. Några kommentarer: i huvudsatsernas nexusfält har jag kompletterat grundschemat med en adverbposition före det finita verbet; adverb av typen B 2b, som Beckman inte har någon beteckning för, benämner jag subjektorienterade adverb; att det inte föreligger några exempel på typerna I 2a och I 2b betyder givetvis inte att typerna saknas.

Det framgår av tabellen att typerna H 2, B 4 och I 4 inte täcks av Beckmans beskrivning. Avvikelsen i fråga om H 2 är särskilt markant. Avvikelserna i fråga om B 4 och I 4 framstår emellertid också klart, även vid en konfrontation med B 2b, den mest närstående typen: medan det är *skaran* som är andaktsfull, så är det inte *man* som är kemisk utan tvätten, inte *den* (oljehanteringen) som är ekonomisk utan utvecklingen, inte *han* (hr Hedlund) som är moralisk utan upprustningen och inte *hushållslärarna* som har vetenskaplig karaktär utan forskningen. Adverbtypen i fråga är med andra ord inte subjektorienterad. Att den saknas hos Beckman är ett tecken på att det är fråga om en nyhet i språket.

## 2. Relationer mellan nominaluttryck och verbaluttryck

För att söka komma syntaktiska bildningar av typen *kemiskt tvätta(r)*, *vetenskapligt forska(r)* in på livet skall vi företa en granskning av relationerna mellan vissa nominaluttryck och i formellt och semantiskt avseende samhöriga verbaluttryck. Vi skall närmare bestämt ta i betraktande hur nominaluttryck som innehåller flera semantiskt tunga element uppträder i verbalversion.

När nominaluttrycket är ett kompositum har svenskan mycket goda möjligheter att tillgodose behovet av en verbal motsvarighet. De båda viktigaste medlen är retrogradering och avledning: till *landsättning* m.m. bildas *landsätta*, till *landskamp* bildas *landskampa*. Hithörande företeelser har jag behandlat i en upp-

TABELL 1. Adverbpositioner.

Huvudsatser					Adverbets beteckning (H 1) Särskilt satsadv. (H 2) - (H 2) -
(Jag)	a bara kemiskt	v läser tvätta		(i våra förstklassiga etablissemang)	
(Stoppade möbler)	kemiskt	tvättas		(utan att tyget avtages)	
<i>Bisatser</i>					
(Vad man) (ehuru jag)	inte möjlig	kan kunde		(vråda) (det i går)	(B 1) nekande adv. (alltid rörligt) (B 2a) Allmänt satsadv. (nästan alltid rörligt)
(som)	andaktsfullt	lyssnade		(på hans ord)	(B 2b) [Subjektorienterat adv.] (nästan alltid rörligt)
(som) (om man) (som den)	alltid kemiskt ekonomiskt	[har] tvättar utvecklas		(varit tvåstaviga) (dem ofta)	(B 3) Obestämt tidsadv. (oftast rörligt) (B 4) - (B 4) -
<i>Infinitivsyntagmer</i>					
(Att)	inte	försäga		(sig)	(I 1) Nekande adv.
(Att)	aldrig	tala		(osanning)	(I 3) Obestämt tidsadv.
(att)	moraliskt	upprusta		(den krassliga stockholmspolisens)	(I 4) -
(att)	vetenskapligt	forska		(på institut och laboratorier)	(I 4) -

sats om nominalkomponerade verb, som kommer att publiceras i Nysvenska studier.

Nästa steg blir att undersöka vad som händer när utgångspunkten inte är ett nominalkompositum utan en nominalfras. Först kan vi konstatera att det i vissa fall inte uppstår några större problem. Detta gäller t.ex. frasen *svarta listan*: *Expressen tänker framhärda och varje år svartlista de försumliga* (Göteborgs Handels- och Sjöfarts-Tidning 4/6 1951). Ordbildningen *svartlista* ansluter sig till föregående typ. Den har för övrigt säkerligen fått stöd av engelskans *blacklist* som enligt Hans Marchand i *The categories and types of present-day English word-formation* (1960) är belagt sedan 1884 (s. 61).

En annan morfologisk lösning föreligger i följande fall: *När jag huttrat färdigt vid vassruggen smög jag in på Grand restaurant Saltsjöbaden, där chefen Holger Lindblom (ni vet han från Cecil) trummat fram värmeperioden genom att annonsbraska med vidöppet fönster och fladdrande gardin* (Expressen 9/6 1951). Bakom verbet *annonsbraska* ligger uppenbarligen nominalfrasen *braskande annonser*. Krönikören har alltså kastat om basmorfemens ordning. Möjlig vore också bildningen *braskannonsera* med verbal förled som i *brännmärka, glidflyga, sprutlackera* m.fl.

Ytterligare en lösning exemplifieras i följande språkprov: *Vi bilintresserade vet att Volvo sedan dess stigit i pris och att det t.o.m. svarta börsas på den* (Expressen 22/1 1951). Nominalfrasen *svarta börsen* har här konverterats till verbaluttryck med bibehållet spatium mellan de båda orden. Uttrycket kan sägas inta en mellanställning mellan ett nominalkomponerat verb och en egentlig verbalfras. I och för sig vore ett nominalkomponerat verb möjligt. Likaväl som *svarta listan* motsvaras av *svartlista* kunde *svarta börsen* motsvaras av *\*svartbörsa*.

Mera komplicerat är läget när nominalfrasen är av typen *kemisk tvätt*. Här föreligger ju direkta morfologiska hinder för nybildning av ett nominalkomponerat verb *\*kemisktvätta* (dock framträder en tendens till bildning av en motsvarande substantiv-sammansättning *kemisktvätt* i fallet *Kemisk-, Vit- och Industri-*

*tvätt*, hämtat ur en annons i Göteborgsdelen av Rikstelefonkatalogen 1955).

Språket söker sig då andra vägar, och jag skall närmast försöka följa dem. Därvid betraktar jag omskrivningar av typen *låta undergå kemisk tvätt, bedriva vetenskaplig forskning* som perifera från den anlagda synpunkten, likaså avledningar som *vetenskap*. Vad dem anbelangar inskränker jag mig därför till detta omnämnande.

En av vägarna är densamma som har lett till uttrycket *svarta börsa*. Den vägen trampas av naturliga skäl inte ofta. På en prydligt målad affisch i ett tvätteris fönster läste jag emellertid 1951 följande: *Kemisk tvätta här*. Bildningen kan ses som ett utslag av det kraftiga behovet av en verbal motsvarighet till den semantiskt enhetliga nominalfrasen *kemisk tvätt*.

En annan väg leder fram via en modifikation av nominalfrasen. Den viktigaste modifikationen har i fallet *kemisk tvätt* utan tvivel ett skriftspråkligt ursprung. Jag tänker naturligtvis på den vanliga förkortningen *kem. tvätt*. Denna förkortade nominalfras ger i sin tur upphov till en parallell verbalfras såsom i följande exempel: *Reparerar, ändrar, syr, pressar, kem. tvättar och färgar allt* (Rikstelefonkatalogen, Göteborgsdelen 1951). Verbalfrasens denominala karaktär framträder tydligt. Hur förkortningen *kem.* i detta fall utlästes är en öppen fråga. Analogt med *kem. tvätta* har man prövat *kem. rena* (jfr *kemiskt ren*): *Kem. rena kostymen hos Saltsjöbadstvännen* (Aftonbladet 25/5 1951).

Modifikationen av nominalfrasen kan föras vidare så att sambandet mellan de båda elementen blir intimare. Närmast får vi då formen *kem.-tvätt*. Också i detta fall finns ett parallellt verbaluttryck belagt: *Poplinrockar och andra sommarplagg / Kem.-tvättas på 8 dagar* (Göteborgs Handels- och Sjöfarts-Tidning 26/7 1951). I det här fallet kan man anta att uttalet har varit *kem-tvättas*.

Nästa stadium i nominaluttryckets sammansmältning föreligger i formen *kem-tvätt*. Motsvarande verbform har jag belagd i följande språkprov: *Nu är det dags att kem-tvätta vår- o. som-margarderoben* (Göteborgs-Tidningen 3/3 1963).

Den slutliga sammansmältningen av elementen har vi i formen *kemtvätt* utan såväl punkt som bindestreck. De båda äldsta be-läggen i mitt material på det motsvarande verbaluttrycket *kem-tvätta* är från 1951: *Varje år får hundratals kvinnor sätta livet till eller skadas svårt då de kemtvättar hemma med bensin, bensol eller andra högeligen eldfängda vätskor* (Det Bästa okt. 1951, s. 64); *När ni kemtvättar hemma, använd ett tvättmedel med hög flampunkt* (Det Bästa okt. 1951, s. 65). Dessa språkprov har jag kommit åt via Svenska Akademiens ordboksredaktions sam-lingar. I en årligen återkommande annons i Norrköping–Lin-köpings-delen av Rikstelefonkatalogen har jag funnit ett annat tidigt belägg. Det hänför sig till år 1954: *Anlita fackmannen när Edra kläder behöver kemtvättas eller färgas*. De närmast före-gående åren löd meningen: *Anlita fackmannen när Edra kläder behöva tvättas och färgas*. Man kan i förbigående notera att den finita verbformen har ändrats från *behöva* till *behöver*. Denna modernisering går väl ihop med nybildningen *kemtvättas*. Jag skall nämna ytterligare ett tidigt belägg, närmast därför att det hänför sig till talat språk. Det gäller ett radioprogram 1955 (24/5) där en av deltagarna yttrade: *Ett genomsnittligt svenskt hushåll kemtvättar för närvarande ...* För att fullständiga bilden nämner jag att participet *kemtvättad* också har medverkat vid upp-komsten av *kemtvätta*. Det är här liksom i så många andra fall alltså fråga om multipel bildning. Ett exempel på participet före-ligger i följande mening: *Om några dar har Ni dem [dvs. vinter-kläderna] åter, perfekt kemtvättade och pressade* (Svenska Dag-bladet 10/9 1951).

När jag anför de olika skrivformerna av *kemtvätt* och *kem-tvätta* menar jag naturligtvis inte att vi har att göra med en strikt kronologisk serie. Vad jag har velat visa är hur de från början två grafiska orden efterhand har börjat hopfogas med växlande grader av intimitet. ”Kortformen *kemtvätt*, liksom det därtill bildade verbet *kemtvätta* förutsätter utan tvivel skrivformen *kem-tvätt*”, står det i Tekniska nomenklaturcentralens publikation *Ord och uttryck* 1959 (s. 28: 14). Enligt min mening är det historiskt sett i själva verket fråga om en växelverkan mellan skrift och tal. Förkortningarna *kem. tvätt* och *kem. tvätta* har kommit att ut-



läsas *kemtvätt* respektive *kemtvätta*, vilket så småningom har slagit igenom i skriften. Vi har alltså att göra med en parallell till ordet *behå* (se min uppsats Förhållandet mellan skrift och tal 1967, s. 89 f.). I detta fall har den ursprungliga stavningen *bh* via uttalet *behå* ersatts av den nu gängse stavningen *behå*.

Morfemet *kem-* i *kemtvätta* ter sig ytligt sett som en retrograd bildning till adjektivet *kemisk*. Det skulle då vara en parallell till exempelvis *balt* av *baltisk*, *kuf* av *kufisk*, *spänst* av *spänstig* (se om dylika bildningar Ture Johannissons uppsats *Über Rückbildung im heutigen Schwedisch* 1951, s. 258 ff.). Snarare förefaller det mig emellertid böra sättas på linje med kortformer som *el-* i *elraka*, *rep-* i *repöva*, *gen-* i *genrepa* och *porr-* i *porrätala* (Sydsvenska Dagbladet Snällposten 8/5 1968) liksom med *tele-* i *televerket* och *polit* i *politbyrån*.

Det är intressant att se hur det morfologiska elementet *kem-* efter vad man kan kalla genombrottet med substantivet *kemtvätt* och verbet *kemtvätta* har vunnit spridning i en rad nybildade ord som grupperar sig kring de båda pionjärorden. Jag skall ge några exempel. Vi har fått substantivet *självkem* (belagt t.ex. i Göteborgs-Tidningen 3/3 1963), som betecknar ett ställe där man själv kan utföra sin kemtvätt eller, som det också heter, *kem-självtvätta*: *Kem-självtvätta och betala pr kilo* (Göteborgs-Tidningen 3/2 1963). Omkastningen av morfemen i förhållande till *självkem* skall ses mot bakgrunden av det äldre verbet *självtvätta*, belagt bl.a. i följande språkprov: *Stortvätta på 4 timmar genom att självtvätta i våra tvättcentraler ...* (Göteborgs Handels- och Sjöfarts-Tidning 13/8 1951). På liknande sätt som en självkem fungerar en *kembar* (Dagens Nyheter (B) 26/3 1966). *Kemtvätterierna* eller *kemfirmorna* som de också kallas (orden belagda i samma artikel som *kembar*) utför olika slag av kemtvätt. Ett kallas *fullkem* och innebär fullständig kemtvätt. Ett annat kallas *enkelkem* och utgör som beteckningen antyder en enklare variant. Ett tredje kallas *kemrent* och avser en ytterligare förenkling (alla tre termerna belagda i den nyssnämnda artikeln). Man kan själv tvätta kemrent på en kembar. Om ett tvätteri tvättar kemrent efter vikt, kallas det för *kilokem* (ordet belagt i de båda ovan nämnda artiklarna 3/3 1963 och 26/3 1966). Man kan

också få sin kemtvätt utförd i en *kemomat* (Dagens Nyheter (B) 14/8 1963), som är en myntmanövrerad kemtvättautomat. Det vore lätt att fortsätta exemplifieringen. Rikedomerna på termer och metoder antyds i ingressen till den flera gånger nämnda artikeln i Dagens Nyheter 1966: "Förr räckte det att en flicka var hel och kemiskt ren, nu ska hon helst vara kemklok också." I flera av de ovannämnda orden, t.ex. *självkem*, *kemfirma* och *kemklok*, uppstår morfemet *kem* ensamt betydelsen 'kemisk tvätt'. Detta antyder möjligheten av ett såvitt jag vet ännu inte belagt verb \**kema* 'låta undergå kemisk tvätt'.

Av de vägar som går från nominalfrasen *kemisk tvätt* till verbaluttryck har vi nu följt den endast tillfälliga led som för till frasen *kemisk tvätta* och den av allt att döma numera mest frekventerade, nämligen den som för till verbet *kemtvätta*. Det finns emellertid fler vägar.

En väg är den där man löser upp den morfologiska och semantiska enheten *kemisk tvätt* respektive *kemiskt tvättad* och använder konstruktionen *tvätta något kemiskt*. Konstruktionen är alltså ytligt sett parallell med exempelvis *tvätta något luktfritt*: *Vi besöka hemmen och tvätta stoppade möbler kemiskt och luktfritt* (Rikstelefonkatalogen, Göteborgs-delen 1951).

Vi är så framme vid den fjärde vägen, som är den i mitt tycke intressantaste. Det är den som leder till den syntaktiska nybildning som Erik Wellander exemplifierade med *Kemiskt tvätta i våra förstklassiga etablissement!* och som inledningsvis exemplifierades utförligare.

Jag har i och med detta sagt att vad Wellander kallar framflyttningen av ett adverb enligt mitt förmenande skall förklaras som inflytande från nominalfraser på motsvarande verbalfrasers struktur. Man skulle kunna säga att nominalfrasen projiceras på verbalfrasen. Bakom andra exempel på konstruktionen, sådana som de anförda fallen *ekonomiskt utveckla*, *moraliskt upprusta* och *vetenskapligt forska*, ligger nominalfraserna *ekonomisk utveckling*, *moralisk upprustning* och *vetenskaplig forskning*, respektive i förekommande fall samhöriga participiella uttryck.

Gemensamt för alla fallen är att morfologiska förhållanden lägger hinder i vägen för bildandet av ett sammansatt verb, som

skulle kunna uttrycka betydelseinnehållet i sammanhållen form. I fråga om *kemtvätta* har de morfologiska hindren raserats, likaså i fråga om den tillfälliga bildningen *moralrusta* (Expressen 9/6 1951) med samma innebörd som *moraliskt upprusta*. Förhållandet är här i så måtto gynnsammare som substantivet *moral* har kunnat utnyttjas som förled. Från en viss synpunkt kan *kemtvätta* och i någon mån *moralrusta* ses som symptom på ett intuitivt motstånd mot den syntaktiska nybildningen.

Som beteckning för typen *kemiskt tvätta* föreslår jag *denominal verbalfras*. En beteckning av detta slag har fördelen att samtidigt vara beskrivande och förklarande. Den synes mig därför vara att föredra framför en formulering som tar sikte på framflyttning av adverb. Jag har stannat för termen *denominal* snarare än *postnominal*, som kan ge intryck av att hänföra sig till syntaktisk position, eller *denominativ*, som används i fråga om enskilda ord (inte fraser).

### 3. Diakroniska och synkroniska aspekter

I sin uppsats Om sammansatta verb i svenskan i Nysvenska studier 34 (tryckt 1955) har Ture Johannisson visat att den i nysvenskan uteslutande fasta sammansättningen vid en rad partikelkomponerade verb, t.ex. *anklaga*, får sin förklaring av anslutning till nominalformer: *anklagad*, *anklagelse*, *anklagare* (s. 174). Han framhåller också att det inte råder någon avgörande principiell skillnad mellan denna verbtyp och typen *maskinskriva* (s. 178). Åke Åkermalm har belyst frågan ytterligare något i uppsatsen Fast sammansättning och lös förbindelse i årgång 41 (tryckt 1962) av samma tidskrift. Han uppehåller sig framför allt vid typen *bortföra*, som han menar har samma förklaring.

Man kan åskådliggöra nominaluttryckens inflytande i ett schema, där vi har verbalfrasen i utgångsläget till vänster, nominaluttrycken i mitten och de resulterande verbaluttrycken till höger (se nedan). Verbaluttrycket *klaga an* har under trycket från *anklagelse*, *anklagad* m.m. fått ge vika för *anklaga*. Verbaluttrycket *föra bort* har vid sin sida fått *bortföra*, och på samma

sätt har *göra ren* vid sin sida fått *rengöra* och *stöta på grund* vid sin sida *grundstöta*. Nu menar jag att de verbalfraser som här har diskuterats också ansluter sig till schemat. *Tvätta kemiskt* har sålunda under trycket från *kemisk tvätt* och *kemiskt tvättad* vid sin sida fått *kemiskt tvätta*:

klaga an	anklagelse, anklagad	anklaga
föra bort	bortföring, bortförd	bortföra
göra ren	rengöring, rengjord	rengöra
stöta på grund	grundstötning, grundstött	grundstöta
tvätta kemiskt	kemisk tvätt, kemiskt tvättad	kemiskt tvätta

Det är alltså samma tendens som har verkat bland såväl de partikelkomponerade verben som de nominalkomponerade verben och vad jag här har kallat de denominala verbalfraserna.

Att *kemisk* får *t*-form i uttryckstypen *kemiskt tvätta* är en automatisk följd av att huvudordets ordklass ändras från substantiv till verb (från den alternativa utgångspunkten *kemiskt tvättad* bortses här). För en jämförbar företeelse har Erik Wellander använt termen "kategorial förskjutning". Jag tänker på uppsatsen *Varm korvgubbe – vandrare i dimma* i festskriften till Elias Wessen 1954. Wellander gör där gällande att uttrycket *vandrare i dimma* i betydelsen 'en som vandrar i dimma' är en avledning till *vandra i dimma*, vars adverbial *i dimma* har undergått kategorial förskjutning till attribut i uttrycket *vandrare i dimma* (s. 244). Som en parallell anför han bl.a. *långsam vandrare* som enligt hans mening är en avledning till *vandra långsamt*. Här har då adverbet *långsamt* kategorialt förskjutits till adjektivet *långsam*. Till Wellanders uppfattning har Åke Åkermalm med vissa modifikationer och tillägg anslutit sig i uppsatsen "Utrensare utan blod" i Modersmåslärarnas förenings årsskrift 1961 (omtryckt i omarbetat skick i författarens volym *Rubriksvenska och andra studier* 1965).

Innan jag går vidare vill jag framhålla att det av Erik Wellander anförda exemplet *Benäget observera att* osv. (citerat ovan) är att betrakta som ett specialfall. Det är ekvivalent med uttryck som *Godhetsfullt observera att* och *Vänligen observera att*, vilka i sin tur är nära förknippade med de grammatiskt fylligare men jämn-

långa standardfraserna *Var god observera att* respektive *Var vänlig observera att*. Förutom detta mönster kan en strävan efter ett slags hövlighetsemfas ha spelat in vid uttryckens bildning.

För att ytterligare belysa de denominala verbalfraserna vidgar jag nu exemplifieringen och anför också vissa prov på subjekt-orienterade adverb. Ett tydligt exempel på en denominal verbalfras ger Erik Wellander i annat syfte i sin uppsats *Ett par produktiva typer av skriftspråklig nybildning i Språk och stil 15* (1915, s. 48): *kommunalt mellanskolebilda hela vårt folk*. I detta fall kan det ju inte vara fråga om någon framflyttning. Det kan inte ha gällt att mellanskolebilda folket kommunalt utan att göra folket kommunal mellanskolebildat. Uttrycket uppvisar alltså s.k. förledsanslutning, och det är detta som förklarar det omöjliga i adverbets efterställning.

Tydligt framträder vidare vissa verbalfrasers denominala karaktär i lexikografiskt metaspråk. Ett exempel: *Vi behöver faktiskt ett verb också, i betydelsen "automatiskt databehandla", och för övrigt ett substantiv med betydelsen "automatisk databehandling, ADB"* (Modern datateknik 1968:5, s. 53). Den nominalfras som bildar utgångspunkt, *automatisk databehandling*, nämns som synes också i språkprovet.

I följande fall har adverbet trängt in i innehållsfältets infinitivposition: *Man borde i stället ha mera konkret diskuterat hur utbildningen till psykoterapeuter inom läkar-, psykolog- och socionomkårererna skall se ut* (Dagens Nyheter 11/10 1968). Bakom verbaluttrycket förnimmer man nominalfrasen *konkret diskussion*.

Inte heller i följande tre fall, som gäller bisatser, är adverbena subjektorienterade. *Detta* [dvs. att ungdomen visade stort intresse för valet] *tolkade gamla valstrategier så att det eventuellt är de drygt 429 000 förstagångsväljarna som starkt påverkar tendensen i landets politik* (Dagens Nyheter 7/10 1968). Jämför här *stark påverkan, starkt påverkad*. *Utom av New York Times har Humphrey detta veckoskifte fått offentligt stöd av organisationen Americans for Democratic Action eller ADA, som han var med om att grunda men som tidigare i år kraftigt stödde senator Eugene McCarthy* (Dagens Nyheter 7/10 1968). Jämför

i detta fall *kraftigt stöd, kraftigt stödd*. Titeln är inte riktigt adekvat, eftersom det snarare torde vara ovanligt att en fattig landsortsflicka ... blir så illa berörd av en fosterfördrivning att hon obotligt insjuknar i sinnessjukdom (Svenska Dagbladet 25/1 1965). Jämför *obotligt sjuk, obotlig sjukdom*.

Subjektorienterade är däremot adverbena i bisatserna i följande båda språkprov. *Det måste dock understrykas att de Fin-Gal [ett slags havskappseglare] som framgångsrikt tävlingsseglar med halvprofessionella besättningar som regel är extremt trimmade* (Dagens Nyheter 26/5 1968). Skutorna är framgångsrika. *Dagen därpå fick Bomboko också bakläxa av regeringspartiets styrelse som enhälligt slöt upp bakom presidentens beslut att ställa Mulele inför krigsrätt* (Dagens Nyheter 11/10 1968). Styrelsen var enhällig.

I infinitivsyntagmerna i följande tre exempel åter är adverbena inte subjektorienterade. *Ibland kanske man kan få fast den [dvs. den oljeutsläppande båten] genom att kemiskt analysera oljan* (Nils Linnman i Dagens eko 29/12 1956). Jämför *kemisk analys, kemiskt analyserad*. *På grund av de tjeckoslovakiska ledarnas stora popularitet skulle ett avsked i protest allvarligt skada de sovjetiska ansträngningarna att återfå kontrollen över Tjeckoslovakien* (Dagens Nyheter 7/10 1968). Det är inte avskedet som är allvarligt; jämför i stället *allvarlig skada, allvarligt skadad*. *Trots detta avvisar man tanken på att radikalt förbättra standarden hos de kollektiva trafikmedlen och bygger i stället bara för bilen* (Dagens Nyheter 7/10 1968). Jämför i detta fall *radikal förbättring, radikalt förbättrad*.

Som jämförelse ger jag också för infinitivsyntagmernas vidkommande några exempel på subjektorienterade adverb. *Till understrykandet av varje partis och kommuniststats rätt att suveränt bestämma över den egna politiken hör ...* (Aftonbladet 11/10 1968). Partiet och staten är suveräna. [Det väsentliga i den recenserade författarens strävan efter universalism] *tar sig ändå uttryck i hans förmåga att initierat genomlysa vanligtvis så skilda områden som affärsliv, ekonomisk historia, psykoanalys, poesi, mytologi, sociologi osv* (Dagens Nyheter 7/10 1968). Författaren är initierad. *Att å ena sidan öppet och reservationslöst*

*notera alla iakttagelser och infall, och å andra sidan sakligt och systematiskt undersöka dem, sortera, bygga av deras innehåll* (Aftonbladet 11/10 1968). Den recenserade författaren är öppen och reservationslös, saklig och systematisk.

Naturligtvis måste man räkna med att andra faktorer än trycket från samhöriga nominalfraser kan ha tjänat som bidragande orsaker till de här studerade icke subjektorienterade adverbens syntaktiska uppträdande. En viss allmän betydelse kan det ha haft att tunga subjektorienterade adverb av typen *andäktigt, beslutsamt, framgångsrikt, initierat* intar den aktuella positionen i bisatser och infinitivsyntagmer. Härutöver torde man ha att räkna med två faktorer: emfas och rytm (jfr Elias Wessen, *Svensk språkhistoria* 3, 2. uppl., s. 205 f.). Båda faktorerna hör ju mera hemma i tal än i skrift. I de talspråksexempel som jag ovan har givit – *ekonomiskt utvecklas, kemiskt analysera* – uppfattade jag ingen emfas, och de rytmiska förhållandena förefaller mig inte lägga hinder i vägen för adverbiell efterställning. Detta utesluter självfallet på intet sätt att de nämnda faktorerna kan ha spelat in i andra sammanhang.

Det är emellertid uppenbart att satsorganisatoriska faktorer av motsvarande slag gör sig gällande också i skrift. Främst ifrågakommer härvid mängden och arten av övriga verbbestämningar i det enskilda fallet. I den ovan anförda infinitivsyntagmen *allvarligt skada de sovjetiska ansträngningarna att återfå kontrollen över Tjeckoslovakien* skulle avståndet mellan *skada* och *allvarligt* bli betydande om adverbet placerades efter det långa objektet (således efter *Tjeckoslovakien*). I följande fall kan det ha inverkat dels att verbet är utbildat med objekt dels att tidsadverbialet *oftare* står i den position som *kemiskt* alternativt kunde inta: *Det är konstigt att svenskarna som för övrigt är så noga med sin hygien, inte kemiskt tvättar sina kostymer oftare* (Göteborgs-Tidningen 20/11 1955).

Det måste också framhållas att det föreligger gradskillnader i fråga om semantisk enhetlighet hos de underliggande nominalfraserna. Från denna synpunkt framstår *kemisk tvätt* som en fastare enhet än exempelvis *radikal förbättring*. Det är karakteristiskt att det är *kemiskt tvätta* och *kem[iskt] rena* som har trängt

in i huvudsatser. Man kan anta att stödet från satsorganisatoriska faktorer får större vikt i de fall där den semantiska enhetligheten är mindre utpräglad.

En intressant fråga är hur fall som de här behandlade skall beskrivas i en generativ grammatik. Som en normal härledning (grovt angiven) skulle följande kunna uppställas (jämför t.ex. Owen Thomas, *Transformational grammar and the teacher of English* 1966, s. 169, och Emmon Bach, *Nouns and noun phrases* [i: *Universals in linguistic theory*, utgiven av Emmon Bach och Robert T. Harms 1968], s. 102):

- (1) (någon) granskar (något) – granskningen är systematisk
- (2) (någon) granskar (något) systematiskt
- (3) systematisk granskning

Det utmärkande för denna härledning – så långt den räcker – är att nominalfrasen (*systematisk granskning*) utvecklas ur verbalfrasen (*granska systematiskt*).

Ett alternativ presenteras av Alvar Ellegård i *Svensk-engelsk transformationsgrammatik* (1969, s. 107 f., se också s. 174), här återgivet efter samma mönster som ovan:

- (I) (1) (någon) granskar (något) – granskningen är systematisk
- (2a) (någon) granskar (något) systematiskt
- (2b) systematiskt granskar (något)
- (II) (1) (någons) granskning (av något) – granskningen är systematisk
- (2) systematisk granskning

Nominalfrasen och verbalfraserna har enligt detta förslag alltså skilda härledningar.

Man skulle också kunna tänka sig en annan möjlighet i anslutning till mitt tidigare resonemang:

- (1) (någon) granskar (något) – granskningen är systematisk
- (2) (någon) granskar (något) systematiskt
- (3) systematisk granskning
- (4) (att någon) systematiskt granskar (något)

Det nya i detta förslag till härledning är att den ur verbalfrasen (2) utvecklade nominalfrasen (3) i sin tur bildar utgångspunkt för



en verbalfras (4). Under (4) innefattar jag förutom bisatser även infinitivsyntagmer (däremot inte de ytterst sällsynta fallen av huvudsatser, till vilka jag återkommer längre fram). För regel-följden (2) → (3), (3) → (4) snarare än (2) → (4), (4) → (3) talar att nominaliseringen bör grunda sig på den generella strukturen (2) – som gäller såväl huvudsatser som bisatser och infinitivsyntagmer – och inte förutsätta den strukturellt inskränkta typen (4).

Emellertid skulle det vara önskvärt att härledningen vore så generell, att den täckte olika typer av denominala verbalfraser. Prövar man nu med exempelvis *organisatoriskt utveckla*, erfar man ett visst motstånd när det gäller moment (1):

(någon) utvecklar (något) – utvecklingen är organisatorisk

Som mera rättvisande än ett predikativt *organisatorisk* framstår ett adverbialt *i organisatoriskt avseende*, vilket bör ge följande serie:

- (1) (någon) utvecklar (något) i organisatoriskt avseende
- (2) (någon) utvecklar (något) organisatoriskt
- (3) organisatorisk utveckling
- (4) (att någon) organisatoriskt utvecklar (något)

Som underliggande struktur ansätts här sålunda verb plus vad man kunde kalla avseendeadverbial.

Ännu tydligare blir kanske utslaget i fallet *kemiskt tvätta*. Här vill man ha vad man på motsvarande sätt kunde kalla ett medelsadverbial (instrumentalt adverbial) i stället för ett predikativ (*tvätten är kemisk*):

- (1) (någon) tvättar (något) med kemiska medel
- (2) (någon) tvättar (något) kemiskt
- (3) kemisk tvätt
- (4) (att någon) kemiskt tvättar (något)

Det är kombinationen av ett medelsadverbial och ett sättsadverbial som ger den speciella effekten åt det ovan anförda exemplet *Vi besöka hemmen och tvätta stoppade möbler kemiskt och luktfritt*.

Härledningsmodellen passar också för sättsadverbialen. Vårt tidigare exempel ger:

- (1) (någon) granskar (något) på ett systematiskt sätt
- (2) (någon) granskar (något) systematiskt
- (3) systematisk granskning
- (4) (att någon) systematiskt granskar (något)

Även de av Beckman nämnda exemplen på subjektsoorienterade sättsadverbial låter infoga sig:

- (1) (någon) lyssnar på ett andäktigt sätt
- (2) (någon) lyssnar andäktigt
- (3) andäktigt lyssnande
- (4) (att någon) andäktigt lyssnar

Detsamma gäller icke subjektsoorienterade sättsadverbial:

- (1) (någon) förbättrar (något) på ett radikalt sätt
- (2) (någon) förbättrar (något) radikalt
- (3) radikal förbättring
- (4) (att någon) radikalt förbättrar (något)

Gradadverbialen kan också behandlas på detta sätt:

- (1) (någon) förbättrar (något) i avsevärd grad
- (2) (någon) förbättrar (något) avsevärt
- (3) avsevärd förbättring
- (4) (att någon) avsevärt förbättrar (något)

Vid sidan av nominaliseringen i (3) bör vidare participkonstruktionen få sin plats:

- (1) (någon) granskar (något) på ett systematiskt sätt
- (2) (någon) granskar (något) systematiskt
- (3a) systematiskt granskad
- (3b) systematisk granskning
- (4) (att någon) systematiskt granskar (något)

Den generella härledningen i fall med sätts-, grad-, avseende- och medelsadverbial skulle följaktligen bli:

- (1) verb + adverbialfras
- (2) verb + adverb
- (3a) adverb + particip
- (3b) adjektiv + substantiv
- (4) adverb + verb

Till de tidigare givna belägg som återspeglar regel (3) → (4) och som innehåller sättsadverbial kan jag föga bl.a. *hjälpigt skydda*, *katastrofalt inskränka*, *kraftigt öka*, *skarpt angripa* och *varmt understödja*.

Ytterligare belägg innehållande avseendeadverbial är *socialt och ekonomiskt jämställa* och *intellektuellt delta* (i kulturlivet). Ibland stöter man på uttryck av en friare typ, som får omhändertas på särskilt sätt: *socialt vara på väg uppåt*, *ekonomiskt stå och stampa*. Avseendeadverbial ingår också i hypotagmer av typen *ekonomiskt möjligt*, *ekonomiskt drägligare*, *moraliskt sämre*, *vetenskapligt trivialt*. De nämnda typerna får antas stödja varandra när det gäller avseendeadverbialen som uttrycksmöjlighet. Denna tycks för övrigt inte vara stort mer än hundra år gammal i svenskan (se Ragnhild Söderbergh, Suffixet *-mässig* i svenskan 1964, s. 366 f.).

Det faktum att två fraser – *kemiskt tvätta* och *kem[iskt] rena* – har trängt in i huvudsatser ser jag som en indikation på att vi här har att göra med lexikalisering. Det vore orimligt att låta regeln (3) → (4) gälla också huvudsatser för att sedan genom restriktioner eliminera dess produkter i alla fall utom några enstaka: *\*ett avsked i protest allvarligt skadar de sovjetiska ansträngningarna*, *\*förstagångsväljarna starkt påverkar tendensen i landets politik*, *\*styrelsen enhälligt slöt upp bakom presidentens beslut*, *\*oljehanteringen ekonomiskt utvecklas i samma takt* etc. men *stoppade möbler kemiskt tvättas utan att tyget avtages*.

Att *kemiskt tvätta* utgör en lexikalisk enhet är vid närmare betraktande föga mera anmärkningsvärt än att *kemtvätta*, *skumtvätta*, *hemställa* osv. gör det. Att enheter av den aktuella typen såvitt jag vet inte tidigare har föreslagits för lexikonets vidkommande hänger samman med att de deriveringsproblem jag har påtalat inte har uppmärksamrats. Lexikalisering av idiom är i och för sig naturligtvis ingen nyhet. Jag kan exempelvis hänvisa till Uriel Weinreichs uppsats *Explorations in semantic theory* i volymen *Current trends in linguistics 3* (utgiven av Thomas A. Sebeok 1966). Weinreich diskuterar där idiom av typerna *by heart* 'utantill' och *shoot the breeze* 'prata strunt' (s. 450 ff.).

Behovet av lexikalisering framträder i ytterligare ett par typfall. Det ena gäller sådana verbalfraser som har antagit en specifik betydelse. Uttrycket *moraliskt upprusta* ovan hänför sig till den etablerade benämningen *Moralisk upprustning* och avser sålunda inte *moralisk upprustning* i allmänhet.

Det andra typfallet gäller sådana fraser som återgår på uttryck med förledsanslutning. Det exempel jag har givit är *kommunalt mellanskolebilda*.

Av undersökningen har framgått att Beckmans restriktioner beträffande framförställning av adverb i bisatser och infinitivsyntagmer har upphört att gälla i fråga om icke subjektorienterade adverb i uttryckstypen *vetenskapligt forska*. I synkroniskt avseende har jag kommit in på frågan om hur reglerna för sådana verbalfrasers generering skall ordnas och i anslutning därtill frågan om i vilken utsträckning man får räkna med lexikalisering av fraser. I diakroniskt avseende innebär resultatet en accentuering av nominalfrasernas genomslagskraft med avseende på verbalfraserna. Det är inte bara så att nominalkomposita i mycket omfattande utsträckning är ansvariga för nominalkomponerade och partikelkomponerade verb, typerna *grundstöta* respektive *anklaga*. Det är också så att nominalfraser kan bryta igenom å ena sidan morfologiska hinder och ge upphov till nybildningar som *kemtvätta* och å andra sidan syntaktiska hinder och ge upphov till denominala verbalfraser, nybildningar som *kemiskt tvätta*, *vetenskapligt forska*.

## Litteratur

- Allén, S., Förhållandet mellan skrift och tal. (Språk, språkvård och kommunikation. Utg. av K.-H. Dahlstedt & B. Moide. 1967, s. 76–92.)
- Bach, E., Nouns and noun phrases. (Universals in linguistic theory. Ed. by E. Bach & R.T. Harms. New York ... 1968, s. 90–122.)
- Beckman, N., Svensk språklära för den högre elementarundervisningen. 9. uppl. 1945.

- Diderichsen, P., *Elementær dansk Grammatik*. 2. Udg. København 1957.
- Ellegård, A., *Svensk-englisk transformationsgrammatik*. 1969.
- Johannisson, T., Om sammansatta verb i svenskan. (*Nysvenska studier* 34 (tr. 1955), s. 162–181.)
- Johannisson, T., Über Rückbildung im heutigen Schwedisch. (Erbe der Vergangenheit. Festgabe für Karl Helm. Tübingen 1951, s. 257–263.)
- Marchand, H., *The categories and types of present-day English word-formation. A synchronic-diachronic approach*. University of Alabama press. U. å.
- Ord och uttryck. Korta uppsatser i aktuella frågor inom teknikens språkvård 1941–1956. 1959.
- Söderbergh, R., Suffixet -mässig i svenskan. En historisk-semantisk ordbildningsstudie. 1964. (*Stockholm studies in Scandinavian philology*. New series 5.)
- Thomas, O., *Transformational grammar and the teacher of English*. New York ... 1966.
- Weinreich, U., *Explorations in semantic theory*. (Current trends in linguistics. Ed. by T.A. Sebeok. 3. The Hague & Paris 1966, s. 395–477.)
- Wellander, E., Ett par produktiva typer av skriftspråklig nybildning. (*Språk och stil* 15 (1915), s. 44–52.)
- Wellander, E., *Reklamens språk*. 1950.
- Wellander, E., Varm korvgubbe – vandrare i dimma. ([Festskrift till] Elias Wessen 15 april 1954. 1954, s. 236–250.)
- Wessén, E., *Svensk språkhistoria*. 3. Grundlinjer till en historisk syntax. 2. uppl. 1965.
- Åkermalm, Å., Fast sammansättning och lös förbindelse. (*Nysvenska studier* 41 (tr. 1962), s. 174–196.)
- Åkermalm, Å., "Utrensare utan blod". (*Modersmållärarnas Förening*. Årsskrift 1961. 1961, s. 71–91. Omarbetad version i Åkermalm, Å., *Rubriksvenska och andra studier* 1965, s. 135–160.)

## Ordströmmen som dataflöde. En aspekt på språklig databehandling

Humanistisk forskning 1972: 2.

Det finns många sätt att se på språket. Man kan betrakta det som ljudvågor (tal) eller linjemönster (skrift), som ett fysiologiskt förlopp, som en form av beteende, som ett kommunikationsmedium, som ett inlärningsobjekt och mycket annat. Ett synsätt som det blir tal om här gäller språket som strängar av symboler. Med symboler menas då representationer av karakteristika hos (eller relationer mellan) idéer eller saker. Som exempel kan man tänka på ordkroppar såsom representationer av begrepp – de filosofiska implikationerna eller komplikationerna får ligga ogilla i det här sammanhanget.

Varför strängar av symboler? Synsättet är fruktbart i första hand därför att vi nu förfogar över ett utomordentligt kraftfullt redskap för just symbolbehandling. Vi har den digitala datamaskinen. Den är konstruerad för bearbetning av vad man kallar diskreta data, dvs. från varandra åtskilda symboler (detta till skillnad från kontinuerliga data, exempelvis fysikaliska mätvärden, som analogidatorn tar hand om). Datamaskinens potential är så stor att dess tillkomst har jämförts med de två tidigare stora genombrotten i mänsklighetens historia, uppkomsten av talet respektive skriften. Det finns alltså all anledning att konfrontera denna symbolbehandlande maskin med språkets symbolsträngar.

De symboler vi betraktar kan hänföra sig till olika språkliga nivåer: de enskilda tecknen, betydelsebärande element, grammatiska strukturer m. m. För enkelhetens skull nöjer vi oss med att betrakta vad vi från språklig synpunkt kan beteckna som ordströmmen. Vi gör det i medvetande om att detta har klar relevans för svenskan och många andra språk men inte för alla. Vi kan också för enkelhetens skull koncentrera oss på strömmen av tryckta ord (som är den mest åtkomliga) med fullt erkännande av den viktiga roll som exempelvis handskrivet och stencilerat

material spelar liksom naturligtvis i högsta grad strömmen av talade ord.

Från informationsbehandlingssynpunkt kan vi se ordströmmen som ett dataflöde. Vi har möjlighet att avlocka detta flöde uppgifter av i princip två slag: kvalitativa och kvantitativa. Den kvalitativa aspekten gäller vilka de språkliga enheterna är på olika nivåer och vilka regler som bestämmer deras uppträdande. Den kvantitativa aspekten gäller hur ofta de olika enheterna förekommer respektive de olika reglerna tillämpas. Denna aspekt är i grund och botten stilistisk. Det är fråga om hur språkbrukarna styr sin språkproduktion. När vi har kommit så här långt har vi i själva verket antytt en modell för studiet av ordströmmen. En modell måste vara baserad på en vetenskaplig teori. I det här fallet sker teoribildningen i nära kontakt med de språkliga manifestationerna. Idéer från allmän och särspråklig lingvistik, informationsbehandling, logik, matematik och statistik tas upp och relateras till det naturliga språkets enorma komplexitet. Maskinens absoluta krav på stringens ger en utslagsgivande testsituation.

Praktiskt omsätts modellen i algoritmer, dvs. formaliserade procedurer, som kodas i programmeringsspråk. Dessa system har karaktären av abstrakta maskiner som utför sina uppgifter genom datamaskinen.

Sammanfattat i ett andetag: ordströmmen ses som strängar av symboler, som bär kvalitativa och kvantitativa data, som studeras med hjälp av teoretiskt baserade modeller, som omsatta i program körs på datamaskin.

### **Från indata till utdata**

Ser vi så på den praktiska sidan av ordströmmen, avtecknar sig konturerna av ett integrerat system. Indata i detta är texter i maskinläsbar form. Analysen av dem i de för tillfället aktuella avseendena sker i samarbete mellan människa och maskin. Den utan jämförelse mest rationella formen därvid är interaktiv bearbetning med forskaren vid en bildskärmsterminal. Utdata är i

många fall lagrade på magnetband som kan tjäna som "manuskript" vid datorstyrd fotosättning av de publikationer som redovisar forskningsresultaten.

Ordströmmens väg från indata till utdata förtjänar kommentarer på flera olika punkter. Låt oss börja med de ingående texterna. De måste som sagt ha maskinläsbar form, vilket innebär att de föreligger på håltremsa, magnetband eller jämförbart medium. Optisk inläsning av klartext är under utveckling men kommer knappast att vara generell användbar och ekonomiskt överkomlig under de närmaste åren.

Av högsta intresse är det då att mängder av texter årligen kodas i våra sätterier. Vi har alltså möjlighet att på ett rationellt sätt bygga upp ett arkiv av maskinläsbara texter.

Vikten av att ha texter i maskinläsbar form kommer efterhand att bli uppenbar för allt fler. Samtidigt har vi möjlighet att med textarkivet som en av hörnstenarna bygga upp en ordbank, ett referenslexikon för vårt ordförråd. Textarkivet och ordbanken tillsammans kallar vi logotek.

Strategin för logoteket är viktig. I varje fall inom överskådlig tid kan man knappast tänka sig att lagra alla texter. Man frågar sig då, vad som egentligen trycks under ett år. Vi har tagit reda på detta i Göteborg genom att undersöka årstrycket på svenska 1970. Undersökningen omfattar allt som tryckts på svenska i Sverige och Finland under året utom accidenstrycket (broschyrer, visitkort o. d.). De framtagna uppgifterna gäller bland annat titlar och beräknat antal löpande ord totalt och per signum (på olika nivåer) i klassifikationssystemet för svenska bibliotek. Det rör sig sammanlagt om drygt 11 000 titlar med en beräknad ordmassa av 3,2 miljarder. Härav svarar böcker och tidskrifter för 0,7 och dagstidningarna för 2,5 miljarder. Tidningarnas dominans är alltså förkrossande. Andelen överstiger tre fjärdedelar. Jämför att skönlitteratur (H), som är den näst största huvudgruppen under böcker och tidskrifter kommer upp till 0,09 miljarder (varav utländsk skönlitteratur i svensk översättning (Hce) svarar för över 50 procent). Större än skönlitteratur är samhälls- och rättsvetenskap (O) med 0,1 miljarder löpord. Undersökningen av års-



trycket ger underlag för texturvalet i logoteket. Den är också av stor betydelse för uppläggnings- och frekvensundersökningar.

Den interaktiva bearbetningen i sin tur kan gälla många olika uppgifter. Det kan vara fråga om kontroll och korrigerings- och själva texten, genomgång av datamaskinella bearbetningar av den (t. ex. morfologiska, syntaktiska, lexikologiska), tillfogande av särskild information (t. ex. semantisk) osv. Tag som exempel att man kört ett program för homografseparering, vars syfte är att rätt bestämma grammatiskt flertydiga former som *råd* (av *råd* eller *råda*), *kolugn* (*ko-lugn* eller *kol-ugn*), *tvättade* (preteritum som *i han tvättade halsen*, particip som *i hans tvättade hals*) osv. Homografi är – kan jag inskjuta – en kolossalt utbredd företeelse i naturligt språk. För svenskans vidkommande rör det sig om två tredjedelar av alla löpande ord i en text, en storleksordning som jag tror de flesta finner överraskande. Vare sig det gäller en kvalitativ eller kvantitativ undersökning måste forskaren gå in och kontrollera, korrigera, komplettera. Tag som ett annat exempel att man kört ett syntaktiskt analysprogram. Som bäst ger ett sådant för närvarande kanske analyser av 90 procent av meningarna, och av dessa är kanske 90 procent korrekta. Forskarens medverkan krävs. Till detta är tre saker att säga. Det är maskinen som gör det stora arbetet, forskaren som finputsar. Också maskinens ”felaktiga” resultat ger djupa inblickar i språkets väsen. Utveckling av allt kraftfullare algoritmer pågår ständigt.

Inte minst när det gäller lexikologiska forskningsuppgifter är vägen ut viktig. Bearbetningarna och urvalet av data för publicering görs i maskinen och resulterar i en datafil på magnetband. Detta kan sedan skickas till sätteriet, där en annan dator tar hand om det. I den utförs redigering av stoffet i sidor, spalter och kolumner. Därefter sker den likaså datorstyrda optiska projiceringen av trycktyperna på underlaget för offsetplåtarna. Tre stora fördelar är att man härigenom kan utnyttja hela det typografiska registret, vilket ju är särskilt viktigt när det gäller språklig databehandling, att det hela går mycket snabbt (det är en fråga om timmar) och att man slipper korrekturläsning, som ersätts av generella kontroller rörande siddispositionen, sorteringsordningarna, stilsorterna och liknande. Nusvensk frekvensordbok 1–

2, som sammanlagt innehåller över 50 olika listningar, har satts med den här metoden.

### Data linguistica

Huvudintresset inom arbetsgruppen i Göteborg har varit inriktat på lexikologiska undersökningar baserade på text. De tyngsta resultaten är de båda hittills utkomna delarna av Nusvensk frekvensordbok. Här presenteras data om ord (definierade på olika sätt), ordklasser, böjningsklasser m. m. i ett tidningsmaterial som omfattar en miljon ord. Valet av material kan betraktas mot bakgrunden av resultatet av undersökningen rörande årstrycket.

Volymerna ingår i serien Data linguistica, som är ett allmänt forum för arbeten på området språklig databehandling. Serien omfattar såväl metodologiskt och teoretiskt orienterade bidrag som materialorienterade undersökningar. Av de sex nummer som hittills har kommit ut är ytterligare två av den senare typen. Det ena är Tiotusen i topp, som presenterar det centrala ordförrådet i det nämnda tidningsmaterialet. Det andra är Jan Thavenius' Konkordans till Hjalmar Gullbergs lyrik. Den förra typen av arbeten representeras av min Introduktion i grafonomi, det lingvistiska skriftstudiet (under medverkan av Staffan Hellberg) och Rolf Gavares Graph Description of Linguistic Structures.

För närvarande undersöker vi fraseologin i tidningsmaterialet. Resultaten skall publiceras som tredje delen av frekvensordboken. Uppläggningsen är i korthet den här. Först gjorde vi en fullständig konkordans till texten. Den omfattar alltså en miljon belägg med kontext. Därefter körde vi fram alla belägg på ordkombinationer som förekommer minst två gånger. Vi fick hela 600 000 sådana belägg. Dessa har vi gått igenom för att få fram dem som enligt vissa uppställda kriterier är fraseologiskt relevanta. Genomgången har givit oss i runt tal 200 000 sådana fall (*det finns ingen anledning att, i form av, bära frukt, fula ord, den fulla sysselsättningen* osv). Det är slående vilken viktig roll dessa spelar i ordströmmen. Vi fortsätter nu analysen av dem.

Bland andra aktuella forskningsuppgifter kan nämnas logoteket (med de två grenarna), ett svenskt homograflexikon (som närmar sig sin fullbordan) och en definitionsvokabulär (i samarbete med den textlingvistiska arbetsgruppen i Konstanz). Det sistnämnda delprojektet syftar till att få fram ett begränsat antal begrepp, med vilka ett stort antal lexikaliska enheter kan definieras. Definitionsvokabulären blir alltså av grundläggande betydelse för den successiva utbyggnaden av ett formaliserat lexikon med hög kapacitet. Lexikonets roll i analys och beskrivning av språk blir alltmer central efterhand som de semantiska frågorna visar sig få allt större vikt. Detta att vi har material och bearbetningar lagrade på magnetband och förfogar över ett utbyggt programsystem har lett till en rätt stor serviceverksamhet. Vi har sålunda utfört uppdrag för forskare och institutioner på mycket växlande områden inom och utom landet, förutom språkforskning bland annat psykologi, sociologi, religionshistoria, biblioteksteknik, handikappforskning, arkitektur och läromedelsutveckling. Man kan se detta faktum som en bekräftelse på språkets centrala roll i samhällslivet men också som en indikation på behovet av en enhet för språklig databehandling.

Den typiska reaktionen när en forskare får ett stort antal systematiserade belägg (till exempel i konkordansform) ur ett omfattande autentiskt material är att det upplevs som att en dammlucka öppnas. Enbart introspektion – där forskaren själv så att säga fungerar som informant – förslår inte. En annan återkommande reaktion är frågan, om materialet och metoderna har utnyttjats i arbetet på den förestående nya upplagan av Svenska Akademiens ordlista. Svaret är nej.

\* \* \*

Vad blir så perspektivet för ämnesområdet språklig databehandling, kortfattat bestämt som studiet av naturligt språk med hjälp av algoritmiska metoder? Fortsatt bevakning, avtappning, bearbetning av ordströmmen. Vidareutveckling av teori och metod genom konfrontation med språkliga data. Gradvis inkorporering

av effektiva algoritmer. Nyttiggörande av den tekniska utvecklingen.

Som jag ser det är det alltså inte längre fråga om huruvida språkforskaren och hans avnämare kan ha nytta av datamaskinen eller inte, utan om hur han bäst skall utnyttja den. Många slags uppgifter är det i dag egentligen inte försvarligt att utföra utan maskinens hjälp. Många kan över huvud taget inte utföras den förutan. Och för egen del är jag övertygad om att vi inte behöver rädas: det är forskaren som bestämmer vad maskinen skall göra, och den är flexibel nog att utföra mycket komplicerade uppgifter.

## Om textattribution. Kring en avhandling av Marina Mundt

Arkiv för nordisk filologi 86. 1971.

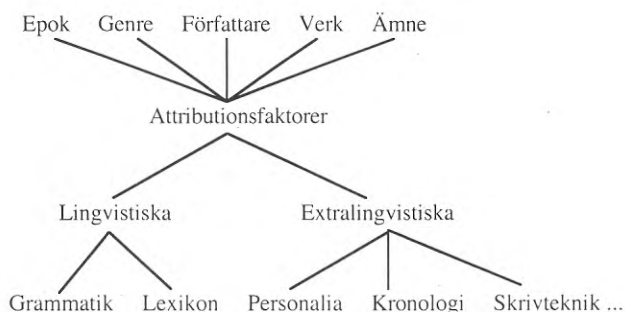
Textattribution låter litet försiktigare än författarbestämning. Detta är ett av skälen till att rubriken lyder som den gör. I själva verket vill det till att omständigheterna är mycket gynnsamma för att en textattribution skall bli en faktisk bestämning av författare.

Normalt krävs det att lingvistiska och extralingvistiska attributionsfaktorer samverkar. I figur 1 har jag antytt detta. De lingvistiska faktorerna kan hämtas från hela det språkliga fältet. De extralingvistiska kan också vara av en mängd olika slag. De kan gälla personliga, kronologiska, skrivtekniska eller andra förhållanden. I idealfallet är dessa faktorer så starka, att kandidaterna till författarskapet med säkerhet kan inringas. Figuren åskådliggör också att studerade attributionsfaktorer kan vara specifika för epok (vari jag inkluderar även en kort tidsperiod), genre, författare, verk eller ämne. Det gäller att vaska fram dem som är författarspecifika.

Ett känt bidrag på området inom äldre nordistik är Bengt Hesselmanns attribuering av Giöta Kiämpa-wisa till Gunno Dahlstierna i en "språkestetisk" undersökning från 1908 med samma namn som visan. Ett annat är Erik Noreens arbete *Författarfrågor i Lejonkulans dramatik* 1938. Det är också Erik Noreen som har lanserat termen favoritor, nämligen i *Från Birgitta till Piraten* 1942.

Till undersökningar från senare år hör Ture Johannisson språkutredning i Helander målet, från början utförd i samarbete med Erik Wellander. Den första utredningen inlämnades 1953. I samband med resning i målet inlämnades 1960 en ny och starkt utvidgad utredning. På anmodan av Dagens Nyheter sammanfattade Johannisson metodik och resultat i två artiklar som publicerades i tidningen sommaren 1961. Artiklarnas titlar, *Det dolda stilmönstret* och *Ett språkligt signalement*, ger metoden i ett nötskal. Det dolda stilmönstret framträder vid studium av vad

Johannisson kallar de särskiljande kriterierna, dvs. de val som författare gör i florán av synonyma formord och alternativa förkortningar (t.ex. *icke – ej – inte* respektive *dr. – dr – d:r*). Kravet på en kandidat gäller överensstämmelse med den anonyme skribenten i fråga om en kombination av sådana kriterier. Det språkliga signalementet framkommer genom studium av individualiser, avvikelser från språknormen. Dessa benämns identifierande kriterier. Utöver det nu sagda kan Johannisson åberopa olika slag av extralingvistiska faktorer.



FIGUR 1. *Attributionsfaktorer.*

Peter Hallberg har i en rad arbeten tagit upp författarfrågor i den isländska sagalitteraturen. År 1962 kom Snorri Sturluson och Egils saga Skallagrímssonar. Ett försök till språklig författarbestämning. Följande år kom Ólafr Þórðarson hvítaskáld, Knýtlinga saga och Laxdæla saga. Ett försök till språklig författarbestämning. Den senaste större skriften är Stilsignalement och författarskap i norrön sagalitteratur (1968). Hallbergs undersökningar bygger på mycket omfattande material, och han arbetar med flera olika attributionsfaktorer. Den metodik han främst tillämpar – och han har själv utvecklat den – är emellertid parordstestet. Parord är sådana ord som är gemensamma för två och endast två av en serie undersökta texter. Om antalet parord mellan den fokuserade texten och en av de övriga texterna är avsevärt högre än mellan den fokuserade texten och återstående

texter, antas detta visa på en affinitet mellan de båda parordsrika texterna.

En delvis likartad metod används av Alastair McKinnon och Roger Webster i en undersökning av dels den inbördes textrelationen mellan åtta av Kierkegaards pseudonymer, dels förhållandet mellan denna textgrupp och skrifter utgivna i Kierkegaards eget namn. Härom kan man läsa i deras korta uppsats *A method of 'author' identification* (1969). De finner att texterna i Kierkegaards namn är mer homogena än de som har pseudonymer och att dessa senare i stort sett kan ordnas i en hierarki efter graden av affinitet med de icke-pseudonyma.

Utanför nordistiken föreligger flera viktiga arbeten. Alvar Ellegård har i sin undersökning av Juniusbrevet, *A statistical method for determining authorship* (1962), tagit fasta på vad han kallar plusord och minusord. Han noterade vid en genomläsning av breven och ett stort jämförelsematerial sådana ord och uttryck som tycktes vara vanligare eller ovanligare i breven än i jämförelsematerialet. Han räknade därefter fram den relativa frekvensen i breven respektive jämförelsematerialet och dividerade dessa värden. Härigenom erhöll han ett mått på deras plusords- respektive minusordskaraktär, kallat *distinctiveness ratio*. Han fick på så sätt fram en testlista på 458 ord som lades till grund för undersökningen. För att motverka samplingfluktuationen – dvs. den slumpmässiga variationen i observerade frekvenser – delade han in testorden i grupper efter storleken på *distinctiveness ratio*. På basis av en statistisk analys av dessa gruppers frekvensmässiga uppträdande byggde han så sin slutsats att Sir Philip Francis hade skrivit breven.

År 1964 utgav Frederick Mosteller och David L. Wallace boken *Inference and disputed authorship: The Federalist*. En resonerande presentation av denna starkt matematisk-statistiskt inriktade undersökning ger Ivor S. Francis i uppsatsen *An exposition of a statistical approach to the Federalist dispute* (1966). Den språkvetenskapliga huvudfrågan i undersökningen gällde vilka av ett dussin omstridda bidrag till *The Federalist* som hade författats av Hamilton och vilka som hade författats av Madison. Mosteller och Wallace genomför en s.k. diskriminantanalys. De

betonar att diskriminatorerna (dvs. attributionsfaktorerna) skall vara icke ämnesbundna och icke imiterbara. Efter åtskilliga prov valde de ut en lista på 30 ord med hög frekvens: *also, consequently, matter(s), to, while* osv. Undersökningen visade att starka odds talade för att Madison hade skrivit alla de tolv bidragen. Det svagaste oddset var 90 mot 1 till Madisons förmån. I just det fallet uttalar de sig likväl med en viss försiktighet med tanke på att det näst svagaste oddset var 800 mot 1.

Dieter Wickmann publicerade 1969 en undersökning med titeln *Eine mathematisch-statistische Methode zur Untersuchung der Verfasserfrage literarischer Texte*. Syftet var att utreda vem i en krets av föreslagna författare till den 1805 publicerade anonyma romanen *Nachtwachen von Bonaventura* som kunde komma i fråga. Som lingvistisk attributionsfaktor använder Wickmann främst ordklassdigram, dvs. ordklassstillhörigheterna hos par av i den löpande texten på varandra följande ord. Genom statistisk hypotesprövning finner han att Wetzel, Jean Paul och Brentano kan uteslutas (på 5 %-nivån), medan detta inte kan ske beträffande E. T. A. Hoffmann. Schelling kan inte prövas, då inget verk av hans hand inom den aktuella genren med säkerhet är känt. Undersökningarna av Wickmann, Ellegård och Mosteller och Wallace är av allt att döma de statistiskt mest genomarbetade på området.

Delvis i samarbete med Wickmann har Yehuda T. Radday undersökt autenticiteten hos kapitlen 13–66 i profeten Jesajas bok (dvs. allt utom de 12 inledande kapitlen, som allmänt anses vara äkta). Vad han hittills har publicerat är en kort sammanfattning med titeln *Isaiah and the computer: A preliminary report* (1970). Som tre av de mest intressanta testen anger han dem som gäller ordklasstransitioner (jämför Wickmann), relativ entropi och krigsvokabulär. Hans resultat är att kapitlen 1–23 kan hänföras till samma författare, att kapitlen 24–35 är odeciderade, att kapitlen 40–48 introducerar en ny författare (kapitlen 36–39 utelämnas av vissa skäl) och att kapitlen 49–66 introducerar ytterligare en författare.

En uppsats av Ernst L. Moerk är också av intresse i detta sammanhang. Den heter *Quantitative analysis of writing styles* och



publicerades 1970. Syftet är att komma åt objektiva stilvariabler. Moerk definierar nitton stycken: antal huvudsatser, antal underordnande konjunktioner, antal substantiv i genitiv osv. Dessa studeras i samspel om 1000 löpande ord. Signifikansberäkningar utförs med  $\chi^2$ -metoden. När Moerk prövar kriterierna på dels Herodotos, Thukydidés och Xenofon, dels Tacitus, Cesar och Livius, finner han signifikanta skillnader mellan författarna och med något undantag icke-signifikanta skillnader inom författarna. Undersökningar av den typen innebär en fruktbar utvidgning av området för textattribution. Facit är känt, och man kan alltså avläsa värdet hos olika attributionsfaktorer och pröva olika analysmodeller. Denna utvidgning av fältet är ett annat av skälen till mitt val av term i rubriken.

## Laxdæla

En av de texter som diskussionen har gällt under senare år är Laxdæla saga. Peter Hallberg har presenterat resultat i de ovan nämnda skrifterna från 1963 och 1968 liksom i en uppsats från 1965, Íslandinga saga och Egla, Laxdæla, Eyrbyggja, Njála, Grettla. *Ett spåktest*. Rolf Heller har publicerat en rad inlägg från Studien zu Aufbau und Stil der Laxdæla saga 1960 till Der Verfasser der Laxdæla saga und sein Verhältnis zur Sturlubók 1969. Marina Mundt (f. Meier), som 1965 kom med uppsatsen Til spørsmålet om Laxdælas forfatter, har lagt fram en doktorsavhandling med titeln Sturla Þórðarson und die Laxdæla saga (1969). Disputationen ägde rum vid universitetet i Bergen vårtidningen 1970.

Starkast är polariseringen mellan Peter Hallberg och Marina Mundt. Hallbergs ståndpunkt är att Ólafr Þórðarson hvítaskáld är sagans författare. Som antyds i titeln till avhandlingen hävdar Marina Mundt i motsats härtill att brodern Sturla Þórðarson bör tillskrivas författarskapet.

Avhandlingen inleds med ett kapitel om problemställningen. Det sägs där ganska anspråkslöst att syftet är att komma ett stycke vidare på vägen mot ett svar på frågan om Laxdælas

ursprung. Tid och ort visar på Ólafr eller Sturla – eller, tillägger författarinnan, någon som de har känt. Av vissa skäl begränsas emellertid kandidaterna till de båda bröderna, och frågan blir vem som är sannolikast.

Undersökningen utgår från ytterligare två antaganden. Det ena är att Knýtlinga saga, som Peter Hallberg har gjort gällande i sitt arbete från 1963, är ett autentiskt verk av Ólafr. Det andra är att Laxdæla vad textgestalten beträffar är ett enhetligt verk.

I nästa avsnitt, som behandlar Laxdælas texthistoria, framhåller författarinnan att undersökningen förutsätter att en bestämd textversion av varje indragen saga läggs till grund för analysen. Denna vilar med andra ord på textkritikens resultat. För Laxdælas vidkommande väljes Einar Ól. Sveinssons utgåva i Íslenzk fornrit.

Med ett undantag görs inte någon motsvarande utredning för någon av de övriga texterna, vare sig för dem som representerar Ólafr respektive Sturla eller för dem som får utgöra jämförelsematerial. Undantaget gäller Íslendinga saga, som ägnas en särskild exkurs (s. 119 f.). Handskriftsförhållandena är i detta fall utomordentligt ogynnsamma. Texten finns varken i original eller i god avskrift. Textutgivarna får långa stycken hålla sig till tämligen otillförlitliga pappershandskrifter från 1600-talet. Författarinnan framhåller med rätta att en textedition i ett sådant läge måste betecknas som ett djärvt rekonstruktionsförsök.

Det tredje kapitlet behandlar metodfrågor. Där framförs som en grundsats att likheter och olikheter som skall tillmätas betydelse i textattributionsfrågor måste kunna uttryckas med siffervärden. Författarinnan tar sålunda avstånd från s.k. pseudokvantitativa uppgifter av typen ”ofta”, ”sällan” osv.

Av föregångare på textattributionens område nämns Erik Noreen, Ture Johannisson, Peter Hallberg och Alvar Ellegård. Av dessa behandlas Hallberg och Ellegård mera utförligt.

I Peter Hallbergs parordsundersökning rörande Laxdæla saga saknar författarinnan en text av Sturla. Hon finner att om man sätter in ett parti ur Sturlas Íslendinga saga av samma omfång som Knýtlinga saga i parordstestet, så blir antalet parord mellan Íslendinga saga och Laxdæla saga av samma storleksordning

som mellan Knýtlinga saga och Laxdæla saga (Hallberg har gjort metodologiska invändningar i uppsatsen från 1965). Hennes slutsats blir att parordstestet inte kan skilja bröderna åt, att alltså Ólafr inte framstår som Laxdælas upphovsman. Också ett par andra test som Peter Hallberg har genomfört kritiseras, främst därför att det i dessa är fråga om ett begränsat antal speciellt utvalda ord.

Författarinnan anför att hennes undersökning väsentligen är en undersökning av plus- och minusord. Hon påpekar emellertid att hon på några punkter avviker från Ellegårds metod. För det första utgår hon i princip från Laxdælas hela ordförråd. Hon utesluter dock de 140 vanligaste orden enligt Einar Haugens undersökning *Norwegian word studies 2*. Vidare utesluter hon vissa speciella typer av ord, sådana som namn, nationalitetsbeteckningar och juridiska termer. För det andra bildar hon inte grupper av testorden utan behandlar dem var för sig. Detta sker i anslutning till kritik mot Ellegård från Gustav Herdans sida. Tanken är att grupperingen skulle dämpa de enskilda ordens förmåga att ge utslag. För det tredje, och denna punkt sammanhänger med den föregående, arbetar hon inte med distinctiveness ratio.

Författarinnans jämförelsematerial omfattar de texter som ingår i Haugens undersökning, nämligen Heimskringla, Eigla och Njála. Härtill kommer de Laxdæla på olika sätt mer eller mindre närstående texterna Eyrbyggja saga, Bjarnar saga Hítðalakappa, Hrafnkatla och Konungs-skuggsjá. Frekvenslistor för några av undersökningens texter har upprättats med hjälp av datamaskin. Erfarenheten av språklig databehandling vid universitetet i Bergen har härvid kommit henne till godo.

I fjärde kapitlet redovisas de test som har genomförts. I text 1 omfattar jämförelsematerialet 454 000 löpande ord. Det är sammansatt av de nyss nämnda texterna med undantag av Konungs-skuggsjá. Sturla representeras av två texter, Hákonar saga och Íslendinga saga, som tillsammans omfattar 201 000 löpande ord. Knýtlinga saga, som representerar Ólafr, omfattar 48 000 löpande ord. Antalet löpande ord i Laxdæla slutligen är 57 000.

Med utgångspunkt i den observerade frekvensen i jämförelsematerialet beräknas den teoretiska frekvensen hos Sturla, hos

Ólafr och i Laxdæla genom enkel proportionsräkning. Den väntade frekvensen i Knýtlinga erhålles exempelvis genom att den observerade frekvensen i jämförelsematerialet divideras med 9,5, eftersom jämförelsematerialet just är 9,5 gånger så stort som Knýtlinga.

Ett ord betraktas som plusord i Laxdæla, om det har minst 1,9 gånger så många belägg i sagan som det teoretiska antalet. Vidare krävs att det skall finnas minst 3 belägg på vederbörande ord i Laxdæla. På motsvarande sätt krävs av ett minusord att det skall ha högst hälften så många belägg som väntat i Laxdæla. Vidare krävs att om Laxdæla saknar belägg, jämförelsematerialet måste ha minst 9. På de angivna grunderna får författarinnan fram en lista på 230 plusord och en lista på 186 minusord. Plusorden förtecknas i boken på s. 60 ff., minusorden på s. 68 ff.

Författarinnan inför så begreppet pluspunkt. Om ett av Laxdælas plusord hos Sturla överträffar den teoretiska frekvensen, oavsett med hur mycket, och om motsvarande inte gäller för Ólafs text, noteras en pluspunkt för Sturla, förutsatt att minst 2 belägg förekommer. I fråga om Ólafr resoneras på analogt sätt. Kravet på minst 2 belägg utgår. När plusorden genomgås efter de här principerna, blir resultatet 51 pluspunkter för Sturla och 30 för Ólafr. De övriga 149 orden ger ingen pluspunkt för någondera. Om kraven på plusorden i Laxdæla skärps så att det krävs minst 4 i stället för minst 3 belägg, reduceras plusordslistan till 164 enheter. Testresultatet blir i detta fall 37 pluspunkter för Sturla och 26 för Ólafr. Antalet odeciderade fall blir alltså 101.

I fråga om pluspunkter bland minusorden förs ett resonemang som är parallellt med det som gäller plusorden. Resultatet av testet blir 52 pluspunkter för Sturla och 22 för Ólafr, medan 112 enheter inte ger något utslag. Författarinnans slutsats av detta första plus- och minusordstest blir, att den större sannolikheten talar för Sturla som författare till Laxdæla (s. 68).

I test 2 företas en reduktion av plus- och minusordslistorna. Detta sker genom att det kravet uppställs att de undersökta orden skall hävda sig som plus- respektive minusord i Laxdæla dels speciellt gentemot Eyrbyggja och Bjarnar saga, dels också speciellt gentemot Konungs-skuggsjá. Dessa båda filter motive-

ras av de särskilda beröringspunkter som finns mellan Laxdæla och de nämnda texterna. Kvar står efter filtreringen 113 plusord. Dessa ger Sturla 32 pluspunkter och Ólafr 21; 60 enheter är odeciderade. Av minusorden kvarstår bara 28, varav 10 ger pluspunkter för Sturla och 3 för Ólafr, medan 15 hamnar bland de odeciderade. Författarinnan uppberar sin slutsats (s. 87).

Test 3 gäller en speciell ordgrupp, nämligen personbeskrivande uttryck. Avsikten är bland annat att ådagalägga eventuella skillnader mellan Íslendinga saga och Hákonar saga, respektive Hákonar saga och Knýtlinga. Författarinnan söker en neutral punkt utanför de behandlade texterna att utgå ifrån och väljer de s.k. porträtten i Trójumanna saga. Här används 118 personbeskrivande uttryck. Det befinns att 41 av dessa också förekommer i Laxdæla och att de används på ett likartat sätt i Knýtlinga, vilket alltså talar för Ólafr som författare. Å andra sidan befinns det att de 77 uttryck som inte används i Laxdæla förekommer sparsamt i Íslendinga saga och Hákonar saga, vilket talar för Sturla. Författarinnan gör ytterligare några sammanställningar av hur de personbeskrivande uttrycken uppträder i de betraktade texterna och kommer fram till att Laxdæla i detta avseende står närmare Hákonar saga än Knýtlinga. Hon finner att den tidigare dragna slutsatsen beträffande författarskapet till Laxdæla har stärkts (s. 101).

I ett femte kapitel tar författarinnan upp sex uttryck som hon betecknar som påfallande egendomligheter (Besonderheiten) i Laxdæla: *þótt eigi sé hér nefndir, honum líkar heldr þungrliga, ok fagnaði hvárr qðrum blíðliga, ok svá bar til litlu síðar, ok lýsti, því fyrir mǫnnum* och *var ástúðigt með þeim bræðrum*. Hon finner att Sturla använder dessa uttryck flitigare än Ólafr och menar att Sturla får ges företräde också på dessa punkter (t.ex. s. 110).

I slutkapitlet anför sammanfattningsvis att de argument som har framkommit vid undersökningen övervägande talar för att Laxdæla tillskrivs Sturla Þórðarson. Resultatet bekräftas, menar författarinnan, av att sagan fyller en lucka i hans produktion. Vi känner honom som historiker och skald – här möter vi honom som författare av konstprosa.

Författarinnan skall tillerkännas modet att ha tagit upp ett krävande ämne. Hon arbetar med ett stort material och intar i varje fall i vissa avseenden en kritisk hållning till detta. Under alla omständigheter bearbetar hon det målmedvetet enligt de riktlinjer hon har dragit upp. På metodens hållfasthet och resultatets tillförlitlighet skall jag strax lägga en del synpunkter.

En första svårighet ligger i begränsningen av antalet kandidater. Författarinnan nämner själv att vi endast känner till ett fåtal av 1200-talets skrivande islänningar. Hon citerar i det sammanhanget Per Wieselgren som har påpekat, att det aldrig har blivit motiverat att endast de kända författarnamnen skulle behöva beaktas (s. 13, not 14). De argument som författarinnan drar fram för att utesluta alla andra kandidater än de båda bröderna gäller främst tid, ort och bildningsnivå. Med tanke på omständigheterna är det naturligt att de inte har karaktären av något bevis. Här har vi alltså en osäkerhetsfaktor som inte får underskattas.

En annan osäkerhetsfaktor hänför sig till själva texterna. För det första råder ovisshet om upphovsmännens arbetsätt. Vad är i egentlig mening författat och vad är redigerat? För det andra råder en större eller mindre osäkerhet beträffande originalens form. Som jag tidigare har nämnt är detta förhållande särskilt utpräglat i fråga om *Íslendinga saga*.

Mina huvudinvändningar riktar sig emellertid mot den statistiska behandlingen. Allmänt gäller att det som framläggs är att hänföra till statistisk deskription. Det förekommer däremot ingen statistisk analys. Det förs inga signifikansresonemang, görs ingen hypotesprövning. De krav som exempelvis uppställs på plus- respektive minusord i *Laxdæla* är helt arbiträra. De replierar inte på någon känd statistisk modell.

Presentationen av det statistiska material som framläggs är för övrigt inte särskilt läsarvänlig. Som exempel kan tabell 1a tjäna (s. 60 ff.). Denna tabell som omfattar sju sidor ger data för plusorden i *Laxdæla*. Man får de observerade frekvenserna för varje ord i var och en av jämförelsematerialets texter, i *Laxdæla* och i de texter som representerar *Sturla* och *Ólafr*. Läsaren får själv räkna fram summan av antalet belägg i jämförelsematerialet, ett värde som är nödvändigt för de fortsatta beräkningarna. En stor

brist är också att de teoretiska värdena för Laxdæla, Sturla och Ólafr saknas. Vill man kontrollera de utfallande pluspunkterna, får man själv räkna fram de teoretiska värdena, vilket är ett rätt drygt arbete. Man saknar också uppgifter om längden på två av deltexterna i jämförelsematerialet, nämligen Bjarnar saga och Hrafnkatla.

En sak som man omedelbart frågar sig är varför inte Íslendinga saga och Hákonar saga genomgående behandlas var för sig. Att en uppdelning sker i test 3 är av underordnad betydelse. En uppdelning i de båda huvudtesten skulle ha gett en värdefull för att inte säga nödvändig kontroll av variationen inom Sturlamaterialet. Finns det en speciell Sturlastil, bör den ju slå igenom i båda texterna. I synnerhet kan man fundera över hur många pluspunkter som skulle falla på vardera texten. Om vi tar plusordslistan som exempel, är det helt säkert att vi skulle ha fått en betydligt lägre proportion än 51:30 för vardera texten gentemot Ólafr.

Låt oss emellertid provisoriskt acceptera författarinnans förutsättningar. En allmän aspekt på läget kan man få genom att exempelvis ta reda på vem av de båda kandidaterna som sammanlagt utnyttjar Laxdælas plusord mest. Man får då räkna fram summan av antalet belägg hos Sturla respektive Ólafr enligt lista 1a och sedan justera för textlängden.

Summan av de observerade frekvenserna har jag fått till 2154 för Sturla och 563 för Ólafr. Vid justeringen för textlängden skall Sturlas värde divideras med 4,2. Hans jämförelsetal blir då 513. Vad man kan konstatera är sålunda att Ólafr utnyttjar dessa ord mer än Sturla enligt författarinnans egen proportionsräkning. Redan detta är en varningssignal.

Vad händer om man för att motverka samplingfluktuationen genomför en gruppering av orden och beräknar utfallet? Jag har gjort experimentet att dela in de 230 plusorden i 23 grupper à 10 ord enligt den i stort sett arbiträra ordning som alfabetiseringen i lista 1a ger. Resultatet blir det man kunde vänta med hänsyn till de nyss givna summauppgifterna. Enligt författarinnans sätt att räkna får vi 3 pluspunkter för Sturla och 2 för Ólafr, medan 18 grupper blir odeciderade. En annan möjlighet är att ställa frekvenstalen i grupperna hos Sturla och Ólafr direkt mot varandra,

nämligen efter att först ha dividerat Sturlas frekvens med 4,2. Man får då överbikt i 12 av grupperna för Ólafr och i 11 av dem för Sturla. Också olika grupperingsprov indikerar alltså att det kan vara slumpens spel med beläggstalen för de enskilda ord författarinnan utnyttjar som har gett det utslag hon har fått.

Man kommer till frågan hur ofta de studerade texterna över huvud taget använder Laxdælas plusord. Jag bortser vid beräkningen från de båda korta texterna Bjarnar saga och Hrafnkatla, vilkas numerär författarinnan inte uppger. Efter att ha justerat proportionellt till Eyrbyggjas omfång av 38 000 löpord, har jag fått jämförelsetalen 457 för Eyrbyggja, 433 för Ólafr, 406 för Sturla, 342 för Njála, 328 för Eigla och 277 för Heimskringla. Också den här frågeställningen är alltså tänkvärd.

Den slutsats jag grundar på dessa i och för sig enkla överväganden är att metodiken och resultatet måste omprövas.

### Kumuleringsmetoden

En av de viktiga punkterna gäller samplingfluktuationen, frekvensernas slumpmässiga variation i texterna. Jag skall belysa detta något. I sådana fall där probabiliteten är liten och antalet händelser stort är Poissonfördelningen tillämplig. Med liten sannolikhet menas i det här sammanhanget en sannolikhet som är mindre än 0,1. Man kan då lägga märke till att inget ord i den nordiska språkkretsen torde ha en probabilitet av den storleksordningen. Enligt Nusvensk frekvensordbok 1 förekommer *och*, det vanligaste ordet, i runt tal 30 000 gånger på 1 miljon löpande ord, vilket alltså svarar mot probabiliteten 0,03. Kravet att antalet händelser skall vara stort svarar mot att textmassan skall vara omfattande. Detta villkor är ju också väl uppfyllt i föreliggande fall.

För att underlätta resonemanget har jag här klätt formeln för Poissonfördelningen i lingvistisk dräkt:

$$\text{Pr} = \frac{t^f}{f!} \cdot e^{-t}.$$



I detta uttryck betecknar  $f$  en viss (betraktad) frekvens,  
 $p_f$  sannolikheten för frekvensen  $f$ ,  
 $t$  den förväntade (teoretiska)  
 frekvensen,  
 $e$  basen i det naturliga  
 logaritmsystemet ( $\approx 2,7$ ).

Med hjälp av formeln kan man exempelvis beräkna sannolikheten för att ett ord som har frekvensen 5 i en text som omfattar 250 000 löpande ord skall uppträda 0, 1, 2, 3 osv. gånger i ett sampel av storleken 50 000 löpande ord. Man finner att sannolikheten för att det betraktade ordet skall uppträda 0 gånger är

$$p_0 = \frac{1^0}{0!} \cdot 2,7^{-1} = \frac{1}{1} \cdot \frac{1}{2,7} = 0,37.$$

På motsvarande sätt får vi

$$p_1 = 0,37$$

$$p_2 = 0,18$$

$$p_3 = 0,06$$

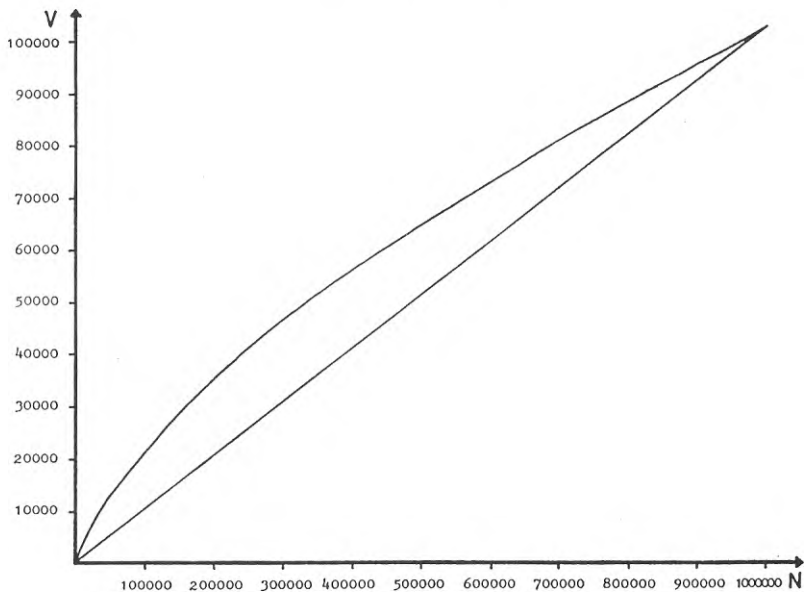
$$p_4 = 0,02$$

Medan den teoretiska frekvensen för det betraktade ordet är 1, skall vi alltså vänta oss att det inte uppträder alls i det aktuella samplet i 37 fall av 100. Vidare skall vi vänta oss att finna frekvensen 1 också i 37 fall av 100, frekvensen 2 i 18 fall av 100, frekvensen 3 i 6 fall av 100 och frekvensen 4 i 2 fall av 100. Det framgår klart att samplingfluktuationen är avsevärd och att det är nödvändigt att gardera sig mot den. Lägg märke till att exemplet är valt så, att textmassan 25 000 löpande ord ungefär motsvarar summan av Sturlas och Ólafs texter och att samplet 50 000 löpande ord ungefär motsvarar Ólafs text.

En annan viktig fråga gäller textlängdernas roll. När författarinnan exempelvis dividerar ett ords frekvens i jämförelsematerialet med 8 för att få fram det väntade frekvensvärdet i Laxdæla, som är 8 gånger kortare, bygger hon på hypotesen att ett ords frekvens är direkt proportionell mot textlängden. Denna hypotes,

som senast har kritiserats av Federico Albano Leoni i uppsatsen *Sagas islandaises et statistique linguistique*. Quelques observations i föregående årgång av denna tidskrift, utgör en alltför grov approximation. Detta framgår av resultat som man kan nå inom kvantitativ vokabulärforskning.

I den undersökning som bedrivs i Göteborg har vi låtit datamaskinen successivt kumulera texter av olika längd och undersöka hur många olika ord varje kumulering ger. På dessa värden har jag byggt en kurva som visar sambandet mellan textlängd eller numerär  $N$  och antalet olika ord eller vokabulär  $V$ . Kumuleringskurvan återges i figur 2. Enheten är i det här fallet det grafiska ordet, men kurvan blir i princip likadan för lemman. Antalet bidragande författare är inemot 1400. Kurvan blir normalt något lägre men i princip likadan vid ett mindre antal bidragsgivare.



FIGUR 2. Kumuleringskurva.

Kurvan visar att  $V$  är ungefär 100 000 när  $N$  är 1 000 000. Räknar man nu proportionsvis, följer värdena den räta linjen

under kurvan. Vid textlängden 200 000 får man då exempelvis i runt tal värdet 20 000 för  $V$ . Det faktiska värdet enligt kurvan är emellertid i runt tal 35 000, alltså inte så långt ifrån dubbelt så stort. Detta innebär att den kvantitativa strukturen hos vokabulären i den kortare texten är en helt annan än strukturen i den längre texten. Vad man gör vid proportionell reducering är att etablera en miniatyr av den längre texten med dennas kvantitativa struktur. Värdena visar att detta är otillfredsställande. Överfört till Laxdælafrågan innebär det bland annat att listorna över plus- och minusord i Laxdæla också i detta avseende rymmer en betydande osäkerhetsfaktor.

Ett teoretiskt resonemang kan belysa problemet ytterligare. Låt oss anta att vi har ett textmaterial som omfattar  $N$  löpande ord och innehåller  $V$  olika ord. Medelfrekvensen blir då

$$\bar{F} = \frac{N}{V}.$$

På motsvarande sätt blir medelfrekvensen i en annan kortare text med  $n$  löpande och  $v$  olika ord

$$\bar{f} = \frac{n}{v}.$$

Förhållandet mellan de båda medelfrekvenserna kan alltså tecknas

$$\frac{\bar{f}}{\bar{F}} = \frac{\frac{n}{v}}{\frac{N}{V}},$$

vilket ger

$$\bar{f} = \frac{nV}{vN} \cdot \bar{F}.$$

Medelfrekvensen i den kortare texten erhålles alltså med hjälp av en faktor som innehåller värdena på såväl vokabulär som numerär i de båda texterna.

Två ting är nu att observera. För det första skall medelfrekvensen inte betraktas som ett direkt uttryck för den kvantitativa strukturen i ett material utan som en indikation på denna struktur. Den är tillräckligt upplysande för vårt resonemang även i den egenskapen. För det andra får den deducerade relationsfaktorn

$$\frac{nV}{\sqrt{N}}$$

inte läggas till grund för beräkning i det enskilda fallet.

Förhållandena är mer komplicerade än så. Bland annat vet vi att direkt proportionalitet faktiskt råder i fråga om de allra vanligaste orden, förutsatt att textlängderna överskrider ett visst kritiskt värde. Se härom min uppsats *Vocabulary data processing 1970*.

Sammanfattningsvis gäller det alltså att på en gång få grepp på samplingfluktuationen, neutralisera textlängdens inverkan och basera sina slutsatser på en känd statistisk modell. Som ett sätt att angripa problemen ser jag följande, som man skulle kunna kalla kumuleringsmetoden. Tanken är att man bör kumulera alla de texter som dras in i undersökningen till ett enda stort material och därefter undersöka ordens fördelning på de ingående deltexterna (samplen). Grafiskt kan vi åskådliggöra materialet såsom i figur 3. Av beteckningarna i figuren står X för den anonyma texten, K för en kandidats text och T för en jämförelsetext.

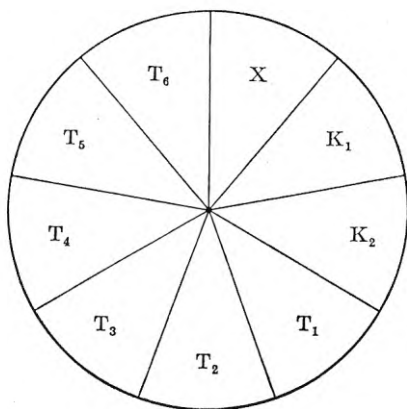
Nollhypotesen blir att varje undersökt ord är jämnt fördelat över deltexterna inom ramen för samplingfluktuationen. Vi fastställer en signifikansnivå, t.ex. 5 %-nivån, och undersöker fördelningen. Detta kan vi göra genom att utnyttja den tidigare nämnda Poissonfördelningen. Finner vi en signifikant avvikelse uppåt i frekvens, kan vi anse oss stå inför ett plusord. Finner vi en signifikant avvikelse neråt, betraktar vi enheten som ett minusord. De flesta enheterna kan väntas falla inom det insignifikanta intervallet. Dessa ord skulle vi kunna kalla för nollord.

I en Poissonfördelning är standardavvikelsen  $s$  lika med roten ur det teoretiska värdet. Använder vi samma symboler som tidigare, kan vi alltså teckna

I ungefär 95 procent av fallen ligger  $f$  inom intervallet  $\pm 2s$ . Vi kan då teckna villkoren för att en enhet med den observerade frekvensen  $f$  skall vara ett plusord, ett nollord respektive ett minusord på följande sätt:

$$\begin{array}{ll} \text{plusord} & f > t + 2\sqrt{t} \\ \text{nollord} & t - 2\sqrt{t} < f < t + 2\sqrt{t} \\ \text{minusord} & f < t - 2\sqrt{t}. \end{array}$$

Såvitt jag vet har en sådan här metodik aldrig använts vid textattribution. En delvis likartad tankegång kan man emellertid finna hos Charles Muller (1968, s. 66, not 1) och nu även hos Federico Albano Leoni (1970, s. 148 ff.).



FIGUR 3. Kumulerad textmassa I.

Vi är så framme vid en prövning av avhandlingens data enligt kumuleringsmetoden. För att kunna beräkna de teoretiska värdena i Laxdæla, hos Sturla och hos Ólafr måste vi först ta reda på numerären  $N$  hos hela materialet (de tre nämnda texterna plus jämförelsetexterna). Vi får  $N = 760\,000$ . Sättes den totala observerade frekvensen för ett ord i materialet till  $F$ , fås följande förväntade beläggsantal i de tre aktuella texterna:

$$\begin{array}{l} \text{Laxdæla} \quad \frac{57}{760} \cdot F = \frac{F}{13,3} \\ \text{Sturla} \quad \frac{201}{760} \cdot F = \frac{F}{3,8} \\ \text{Ólafr} \quad \frac{48}{760} \cdot F = \frac{F}{15,8} \end{array}$$

Jag har låtit bearbeta författarinnans plusordslista och minusordslista i datamaskin enligt de riktlinjer som jag har skisserat. Resultatet redovisas i tabell 2 (plusordslistan) och tabell 3 (minusordslistan) efter uppsatsens text. Tabellerna är uppställda så att de efter vederbörande ordkropp ger i tur och ordning frekvensen i jämförelsematerialet, i Laxdæla, hos Sturla, hos Ólafr och i hela materialet. Därefter följer tre grupper av data som är parallellt uppställda och som gäller Laxdæla, Sturla respektive Ólafr. Inom varje grupp ges först det teoretiska värdet, därefter nollordsintervallet och därefter eventuell signifikant avvikelse. En nolla i den sistnämnda kolumnen innebär att den observerade frekvensen faller inom det väntade intervallet. Det är på grund av denna nolla som jag betecknar vederbörande ord som nollord. Positiva värden i samma kolumn signalerar plusord och negativa värden minusord.

Tabell 1 ger en sammanfattning av beräkningarnas resultat. Tabellen visar hur många av plusordslistans respektive minusordslistans ord som faller ut som plusord, nollord respektive minusord i de tre aktuella texterna. Också relativa värden, angivna i procent, redovisas.

Den verkligt viktiga frågan är naturligtvis hur Laxdæla betar sig. Vi ser att av de 230 enheterna i plusordslistan bara 189 har signifikanta plusordsvärden på 5 %-nivån. Detta innebär ett bortfall av 18 %. Redan det är stort nog att allvarligt rubba värdet hos författarinnans undersökning.

Med tanke på att minusordens frekvenser principiellt är lägre kan man vänta sig att de i ändå större utsträckning stannar inom nollordsintervallet. Detta visar sig också med allt eftertryck vara fallet. Endast 27 av de 186 orden i listan framstår som minusord. Bortfallet är 85,5 %.

TABELL 1. *Plusord, nollord och minusord i avhandlingens plusordslista och minusordslista.*

	Plusord		Nollord		Minusord		S:a	
	Abs	Rel	Abs	Rel	Abs	Rel	Abs	Rel
<i>Plusordlistan</i>								
Laxdæla	189	82,2	41	17,8	0	0,0	230	100,0
Sturla	21	9,1	183	79,6	26	11,3	230	100,0
Ólafr	20	8,7	206	89,6	4	1,7	230	100,0
<i>Minusordlistan</i>								
Laxdæla	0	0,0	159	85,5	27	14,5	186	100,0
Sturla	21	11,3	123	66,1	42	22,6	186	100,0
Ólafr	33	17,7	148	79,6	5	2,7	186	100,0

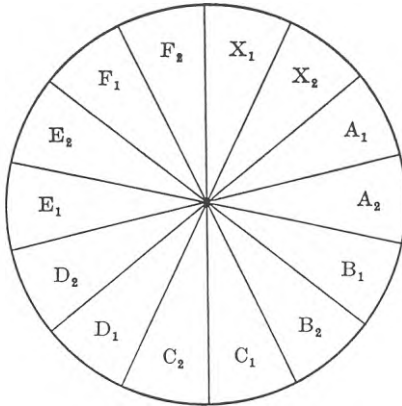
Eftersom sålunda plusordslistan endast delvis är en plusordslista och minusordslistan på det hela taget inte är någon minusordslista, är det inte fruktbart att genomföra några test med avseende på Sturla och Ólafr. Skall man över huvud taget kommentera siffrorna beträffande dessa, kan man påpeka att av de ord som faller utanför nollordsintervallet Sturla i båda listorna har övervikt på minusordssidan, Ólafr på plusordssidan.

För fullständighetens skull har jag också tagit ett par stickprov ur de särskilda egendomligheter som behandlas i slutet av boken. Mina beräkningar gäller de båda första uttrycken. Beräkningarna kan göras inom ramen för den kumulerade textmassan Laxdæla plus Sturla plus Ólafr. Belägg i övriga texter meddelas nämligen inte. Resultatet blir att frekvenserna för de båda uttrycken i deltexterna inte faller utanför de väntade intervallen någonstans, vare sig i Laxdæla, hos Sturla eller hos Ólafr.

Hittills har jag använt vad jag kallar kumuleringsmetoden för att påvisa, att den behandlade avhandlingens slutsatser inte är hållbara. Emellertid bör det ha sitt intresse att tänka på hur metoden skulle kunna utnyttjas positivt i frågor rörande textattribuering.

Tillvägagångssättet skulle kunna bli ungefär följande. Textsituationen ordnas i princip enligt figur 4. Mot den anonyma texten X står alltså ett antal mer eller mindre sannolika kandidater: A, B osv. Markeringarna  $X_1$ ,  $X_2$  osv. anger att både den ano-

nyme skribenten och kandidaterna bör vara representerade av minst två samspel. Man bör hålla epoken och genren konstanta, t.ex. detektivromaner från 1950-talet. Härigenom nedbringas störningen från epok- och genrespecifika egenheter. Genom att varje upphovsman vidare representeras av mer än ett samspel och dessa matchas mot varandra bör också de ämnesspecifika egenheterna väsentligen elimineras. Optimalt bör de olika samplen från en och samma upphovsman hämtas från olika verk, varigenom även de verkspecifika egenheterna (beroende på val av stilläge) hålls nere. Kvar bör då stå författarspecifika egenheter.



FIGUR 4. Kumulerad textmassa II.

Låt oss i anslutning till diskussionen ovan anta att vi arbetar med vokabulären. Vår första uppgift blir då att upprätta en frekvenslista för varje samspel och en frekvenslista för hela materialet. På basis av Poissonfördelningen räknar vi så fram en plusordslista och en minusordslista för varje samspel. För att komma åt de plus- respektive minusord som är karakteristiska för varje upphovsman tar vi därefter fram snittmängderna  $X_1 \cap X_2$  (dvs. antalet för  $X_1$  och  $X_2$  gemensamma plus- respektive minusord),  $A_1 \cap A_2$  osv. Vi kallar dessa snittmängder  $X$ ,  $A$  osv. Vi är nu i stånd att matcha den anonyme upphovsmannen mot var och en



en av kandidaterna och tar därför fram snittmängderna  $X \cap A$ ,  $X \cap B$  osv. De värden som vi får fram på så sätt underkastar vi slutligen en signifikansprövning. I första hand blir det då fråga om att utesluta omöjliga kandidater på basis av signifikant säkerställda skillnader mellan vederbörande snittmängder. För att ge en positiv attribuering med god sannolikhet bör antalet plus- och minusord i  $X$  och vederbörande kandidats text vara av samma storleksordning. Vidare bör snittmängden mellan  $X$  och vederbörande text vara stor jämfört med minimum av plus- och minusord i  $X$  eller vederbörande text.

Man är naturligtvis inte bunden till att arbeta med de signifikanta ordens antal. Man kan också utnyttja deras frekvens. Vidare kan man uppenbarligen öka testets potential genom att undersöka andra slag av språkliga enheter. Som sådana kunde man, för att bara ta ett par exempel, välja förekomsten av genitiv singularis av substantiv eller förekomsten av bisatser. Det krävs att dessa nya faktorer är obundna av tidigare utnyttjade faktorer i den meningen att de inte är förutsägbara med utgångspunkt från dem. Modellen medger att man testar på varje kriterium för sig och på alla kriterierna tillsammans.

Med denna metodologiska skiss slutar jag. Det har framgått att den angivna metodiken förutsätter statistisk analys och språklig databehandling för sin tillämpning. Man kan förutspå att dessa båda vetenskapsgrenar kommer att utöva ett växande inflytande på området. Inte bara textattributionen som sådan kommer att dra nytta härav utan också stilistiken.

TABELL 2. Plusorðslistan.

Jfm	Lax	St	Ól	S: a	Tlax	Intervall	Sign	TSt	Intervall	Sign	TÓl	Intervall	Sign
17	4	17	4	42	3	-.0-6.7	.0	11	4.4-17.8	.0	3	-.0-5.9	.0
0	5	1	0	6	0	-.0-1.8	3.2	2	-.0-4.1	.0	0	-.0-1.6	.0
11	5	3	1	20	2	-.0-3.9	1.1	5	.7-9.9	.0	1	-.0-3.5	.0
1	3	1	0	5	0	-.0-1.6	1.4	1	-.0-3.6	.0	0	-.0-1.4	.0
32	8	32	4	76	6	-.9-10.5	.0	20	11.1-29.1	2.9	5	-.4-9.2	.0
14	6	0	1	21	2	-.0-4.1	1.9	6	.8-10.3	-.8	1	-.0-3.6	.0
29	9	16	6	60	5	-.3-8.7	.3	16	7.9-23.8	.0	4	-.0-7.7	.0
1	3	1	3	8	1	-.0-2.1	.9	2	-.0-5.0	.0	1	-.0-1.9	1.1
5	7	0	0	12	1	-.0-2.8	4.2	3	-.0-6.7	.0	1	-.0-2.5	.0
14	9	3	4	30	2	-.0-5.3	3.8	8	2.3-13.6	.0	2	-.0-4.6	.0
4	4	4	4	16	1	-.0-3.4	.6	4	.1-8.3	.0	1	-.0-3.0	1.0
1	3	0	0	4	0	-.0-1.4	1.6	1	-.0-3.1	.0	0	-.0-1.3	.0
2	4	1	1	8	1	-.0-2.1	1.9	2	-.0-5.0	.0	1	-.0-1.9	.0
1	4	4	0	9	1	-.0-2.3	1.7	2	-.0-5.5	.0	1	-.0-2.1	.0
9	7	1	0	17	1	-.0-3.5	3.5	4	-.3-8.7	.0	1	-.0-3.1	.0
8	8	9	4	29	2	-.0-5.1	2.9	8	2.1-13.2	.0	2	-.0-4.5	.0
2	3	1	0	6	0	-.0-1.8	1.2	2	-.0-4.1	.0	0	-.0-1.6	.0
20	7	27	1	55	4	-.1-8.2	.0	15	6.9-22.2	4.8	3	-.0-7.2	.0
6	5	2	0	13	1	-.0-2.9	2.1	3	-.0-7.1	.0	1	-.0-2.6	.0
21	12	9	0	42	3	-.0-6.7	5.3	11	4.4-17.8	.0	3	-.0-5.9	.0
13	4	6	2	25	2	-.0-4.6	.0	7	1.5-11.8	.0	2	-.0-4.1	.0
12	4	6	0	22	2	-.0-4.2	.0	6	1.0-10.6	.0	1	-.0-3.7	.0
3	7	5	0	15	1	-.0-3.2	3.8	4	-.0-8.0	.0	1	-.0-2.9	.0
8	5	1	0	14	1	-.0-3.1	1.9	4	-.0-7.6	.0	1	-.0-2.8	.0
4	4	1	0	9	1	-.0-2.3	1.7	2	-.0-5.5	.0	1	-.0-2.1	.0
2	3	0	0	5	0	-.0-1.6	1.4	1	-.0-3.6	.0	0	-.0-1.4	.0
27	12	3	0	42	3	-.0-6.7	5.3	11	4.4-17.8	-1.4	3	-.0-5.9	.0
7	5	13	2	27	2	-.0-4.9	.1	7	1.8-12.5	.5	2	-.0-4.3	.0
8	5	5	1	19	1	-.0-3.8	1.2	5	-.5-9.5	.0	1	-.0-3.4	.0



	Jfm	Lax	St	Ól	S:a	TLax	Intervall	Sign.	TSt	Intervall	Sign.	TÓl	Intervall	Sign.
fyrirnuna vb.	0	3	1	0	4	0	-0-1.4	1.6	1	-0-3.1	.0	0	-0-1.3	.0
fyrirsát	6	4	0	0	10	1	-0-2.5	1.5	3	-0-5.9	.0	1	-0-2.2	.0
fýsa(sk) vb.	44	18	25	6	93	7	1.7-12.3	5.7	25	14.7-34.5	.0	6	1.0-10.7	.0
föðurléifð	8	4	15	0	27	2	-0-4.9	.0	7	1.8-12.5	2.5	2	-0-4.3	.0
fórúnautr	80	26	18	2	126	9	3.3-15.6	10.4	33	21.8-44.9	-3.8	8	2.3-13.6	-3
gaman	15	4	8	0	27	2	-0-4.9	.0	7	1.8-12.5	.0	2	-0-4.3	.0
garpr	4	6	0	0	10	1	-0-2.5	3.5	3	-0-5.9	.0	1	-0-2.2	.0
gíptask vb.	17	9	2	0	28	1	-0-5.0	4.0	7	2.0-12.8	.0	2	-0-4.4	.0
gista vb.	13	5	20	0	38	3	-0-6.2	.0	10	3.7-16.4	3.6	2	-0-5.5	.0
gisting	14	5	8	1	28	2	-0-5.0	.0	7	2.0-12.8	.0	2	-0-4.4	.0
glaförð	6	3	2	0	11	1	-0-2.6	.4	3	-0-6.3	.0	1	-0-2.4	.0
gleði	7	4	1	0	12	1	-0-2.8	1.2	3	-0-6.7	.0	1	-0-2.5	.0
gripur	49	20	15	4	88	7	1.5-11.7	8.3	23	13.6-32.9	.0	6	.8-10.3	.0
gullrekinn	5	6	2	0	13	1	-0-2.9	3.1	3	-0-7.1	.0	1	-0-2.6	.0
gullróðinn	6	3	0	1	10	1	-0-2.5	.5	3	-0-5.9	.0	1	-0-2.2	.0
gryla, gerla	19	6	11	0	36	3	-0-6.0	.0	10	3.3-15.7	.0	2	-0-5.3	.0
gersemi, gersemi	18	8	4	3	33	2	-0-5.6	2.4	9	2.8-14.6	.0	2	-0-5.0	.0
gerðvilligr	15	6	0	0	21	2	-0-4.1	1.9	6	.8-10.3	-8	1	-0-3.6	.0
(ó)happ	12	6	0	2	20	2	-0-3.9	2.1	5	.7-9.9	-7	1	-0-3.5	.0
hársliur	0	3	0	0	3	0	-0-1.2	1.8	1	-0-2.6	.0	0	-0-1.1	.0
hefnd	24	16	8	2	50	4	-0-7.6	8.4	13	6.0-20.5	.0	3	-0-6.7	.0
heiman	57	24	15	1	97	7	1.9-12.7	11.3	26	15.5-35.8	-5.6	6	1.2-11.1	-2
heimboð	20	8	3	3	34	3	-0-5.7	2.3	9	3.0-15.0	.0	2	-0-5.1	.0
heimta vb.	54	15	14	6	89	7	1.5-11.8	3.2	24	13.8-33.2	.0	6	.9-10.4	.0
hentur	17	5	3	3	28	2	-0-5.0	.0	7	2.0-12.8	.0	2	-0-4.4	.0
hlíta vb.	8	6	4	2	20	2	-0-3.9	2.1	5	.7-9.9	.0	1	-0-3.5	.0
horfa vb.	25	18	19	2	54	3	-0-8.1	.0	14	6.7-21.8	.0	3	-0-7.1	.0
hraustr	29	10	2	4	45	3	-0-7.0	3.0	12	5.0-18.8	-3.0	3	-0-6.2	.0
hrís	4	3	0	0	7	1	-0-2.0	1.0	2	-0-4.6	.0	0	-0-1.8	.0
hross	51	30	19	0	100	8	2.0-13.0	17.0	26	16.2-36.7	.0	6	1.3-11.3	-1.3
hróf	1	3	0	0	4	0	-0-1.4	1.6	1	-0-3.1	.0	0	-0-1.3	.0
hversdagliga	8	3	16	0	27	2	-0-4.9	.0	7	1.8-12.5	3.5	2	-0-4.3	.0

196	50	61	21	328	25	14.7-34.5	15.5	87	68.1-105.4	-7.1	21	11.6-29.8	.0	
	2	3	0	0	5	.0-1.6	1.4	1	.0-3.6	.0	0	.0-1.4	.0	
höfuðisar														
högr	16	8	8	5	37	0-6.1	1.9	10	3.5-16.0	.0	2	.0-5.4	.0	
jafnaðarmaðr	3	3	3	0	9	.0-2.3	.7	2	.0-5.5	.0	1	.0-2.1	.0	
jafnan	175	38	96	30	339	15.3-35.5	2.5	90	70.7-108.6	.0	21	12.2-30.7	.0	
jafnskjótt	7	5	5	2	19	.0-3.8	1.2	5	.5-9.5	.0	1	.0-3.4	.0	
jaðarmen	2	3	0	0	5	.0-1.6	1.4	1	.0-3.6	.0	0	.0-1.4	.0	
kirkja	94	20	231	36	381	17.9-39.3	.0	101	80.7-120.8	110.2	24	14.3-33.9	2.1	
kirkjuviðr	1	5	0	0	6	.0-1.8	3.2	2	.0-4.1	.0	0	.0-1.6	.0	
knálgir	1	6	3	0	10	.0-2.5	3.5	3	.0-5.9	.0	1	.0-2.2	.0	
konungsnautr	1	6	0	0	7	.0-2.0	4.0	2	.0-4.6	.0	0	.0-1.8	.0	
konungssystir	1	3	2	0	6	.0-1.8	1.2	2	.0-4.1	.0	0	.0-1.6	.0	
kunnigr	46	16	22	12	96	1.8-12.6	3.4	25	15.3-35.5	.0	6	1.1-11.0	1.0	
kurteisi	3	3	0	0	6	.0-1.8	1.2	2	.0-4.1	.0	0	.0-1.6	.0	
kurteiss	7	3	5	1	16	.0-3.4	.0	4	.1-8.3	.0	1	.0-3.0	.0	
kvikifé	7	3	0	0	10	.0-2.5	.5	3	.0-5.9	.0	1	.0-2.2	.0	
kynferð	5	3	0	1	9	.0-2.3	.7	2	.0-5.5	.0	1	.0-2.1	.0	
kynlgir	9	3	1	0	13	.0-2.9	.1	3	.0-7.1	.0	1	.0-2.6	.0	
kyr(r)it adv.	53	28	15	5	101	2.1-13.1	14.9	27	16.4-37.0	-1.4	6	1.3-11.4	.0	
landkaup	0	7	0	0	7	.0-2.0	5.0	2	.0-4.6	.0	0	.0-1.8	.0	
langafasta	4	3	7	0	14	.0-3.1	.0	4	.0-7.6	.0	1	.0-2.8	.0	
letja vb.	38	11	16	0	65	.5-9.3	1.7	17	8.9-25.5	.0	4	.1-8.2	-1	
lößsmurr	21	6	4	0	31	.0-5.4	.6	8	2.5-13.9	.0	2	.0-4.8	.0	
líflát	7	5	5	9	26	2	.0-4.7	.3	7	1.6-12.1	.0	2	.0-4.2	4.8
lítt	69	22	52	10	153	4.7-18.2	3.8	40	27.7-53.2	.0	10	3.4-15.9	.0	
lokrekka	6	3	6	0	15	.0-3.2	.0	4	.0-8.0	.0	1	.0-2.9	.0	
lœkr	8	5	0	0	13	.0-2.9	2.1	3	.0-7.1	.0	1	.0-2.6	.0	
mannaferrð	0	3	4	0	7	.0-2.0	1.0	2	.0-4.6	.0	0	.0-1.8	.0	
mannaforráð	10	3	14	0	27	.0-4.9	.0	7	1.8-12.5	1.5	2	.0-4.3	.0	
mannraun	3	4	4	1	12	.0-2.8	1.2	3	.0-6.7	.0	1	.0-2.5	.0	
máldagi	6	3	0	0	9	.0-2.3	.7	2	.0-5.5	.0	1	.0-2.1	.0	
mein	42	14	16	5	77	1.0-10.6	3.4	20	11.3-29.4	.0	5	.5-9.3	.0	
mennt	8	3	6	3	20	.0-3.9	.0	5	.7-9.9	.0	1	.0-3.5	.0	
merkiligr	7	6	3	0	16	.0-3.4	2.6	4	.1-8.3	.0	1	.0-3.0	.0	

	Jfm	Lax	St	Ól	S: a	T:Lax	Interval	Sign	T:T	Interval	Sign	Tól	Interval	Sign			
metorb	10	3	2	0	15	1	.0-	3.2	.0	4	.0-	8.0	.0	1	.0-	2.9	.0
mikilmenni	1	3	0	0	4	0	.0-	1.4	1.6	1	.0-	3.1	.0	0	.0-	1.3	.0
misseri	27	8	10	0	45	3	.0-	7.0	1.0	12	5.0-	18.8	.0	3	.0-	6.2	.0
motr	0	27	0	0	27	2	.0-	4.9	22.1	7	1.8-	12.5	-1.8	2	.0-	4.3	.0
móurfaðir	5	5	1	0	11	1	.0-	2.6	2.4	3	.0-	6.3	.0	1	.0-	2.4	.0
móðurkyn	2	5	0	0	7	1	.0-	2.0	3.0	2	.0-	4.6	.0	0	.0-	1.8	.0
mægð	7	5	2	1	15	1	.0-	3.2	1.8	4	.0-	8.0	.0	1	.0-	2.9	.0
mæðgin	1	3	8	0	12	1	.0-	2.8	.2	3	.0-	6.7	1.3	1	.0-	2.5	.0
nánd	15	7	8	0	30	2	.0-	5.3	1.8	8	2.3-	13.6	.0	2	.0-	4.6	.0
ofláti	4	3	0	0	7	1	.0-	2.0	1.0	2	.0-	4.6	.0	0	.0-	1.8	.0
ofrefli	8	4	2	2	16	1	.0-	3.4	.6	4	.1-	8.3	.0	1	.0-	3.0	.0
ofsi	10	3	5	2	20	2	.0-	3.9	.0	5	.7-	9.9	.0	1	.0-	3.5	.0
orðtak	5	4	2	1	12	1	.0-	2.8	1.2	3	.0-	6.7	.0	1	.0-	2.5	.0
orlof	18	6	23	1	48	4	.0-	7.4	.0	13	5.6-	19.8	3.2	3	.0-	6.5	.0
óðr	26	9	3	9	47	4	.0-	7.3	1.7	12	5.4-	19.5	-2.4	3	.0-	6.4	2.6
(at) óvörum	2	4	3	0	9	1	.1-	8.2	.0	15	6.9-	22.2	15.8	3	.0-	7.2	.0
páskar, -ir	9	6	38	2	55	4	.0-	4.9	12.1	7	1.8-	12.5	.0	2	.0-	4.3	.0
ráðahag	8	17	2	0	27	2	.0-	2.0	4.0	2	.0-	4.6	.0	0	.0-	1.8	.0
ráðakostr	1	6	0	0	7	1	.0-	2.0	4.0	2	.0-	4.6	.0	0	.0-	1.8	.0
ráðligr	20	6	10	3	39	3	.0-	6.3	.0	10	3.9-	16.7	.0	2	.0-	5.6	.0
reiðfær	2	4	1	0	7	1	.0-	2.0	2.0	2	.0-	4.6	.0	0	.0-	1.8	.0
reiðuligr	9	4	1	3	17	1	.0-	3.5	.5	4	.3-	8.7	.0	1	.0-	3.1	.0
risuligr	1	7	1	0	9	1	.0-	2.3	4.7	2	.0-	5.5	.0	1	.0-	2.1	.0
ritat	23	9	44	9	85	6	1.3-	11.4	.0	22	13.0-	32.0	12.0	5	.7-	10.0	.0
samfór	4	7	1	0	12	1	.0-	2.8	4.2	3	.0-	6.7	.0	1	.0-	2.5	.0
samlendr	5	3	0	0	8	1	.0-	2.1	.9	2	.0-	5.0	.0	1	.0-	1.9	.0
sannligr	7	3	15	5	30	2	.0-	5.3	.0	8	2.3-	13.6	1.4	2	.0-	4.6	.4
selfr	1	3	0	0	4	0	.0-	1.4	1.6	1	.0-	3.1	.0	0	.0-	1.3	.0
silfrhringr	14	4	0	0	4	0	.0-	1.4	1.6	1	.0-	3.1	.0	0	.0-	1.3	.0
sjóðr	2	3	0	0	18	1	.0-	3.7	.3	5	.4-	9.1	-4.4	1	.0-	3.3	.0
skapligr	2	3	0	0	5	0	.0-	1.6	1.4	1	.0-	3.6	.0	0	.0-	1.4	.0
skepstórr	11	4	1	1	17	1	.0-	3.5	.5	4	.3-	8.7	.0	1	.0-	3.1	.0

## skarlatlæklaði, -kyrtill

2	2	8	0	2	12	1	0	-2.8	5.2	3	0	-6.7	0	1	0	-2.5	0
2	4	2	0	2	8	1	0	-2.1	1.9	2	0	-5.0	0	1	0	-1.9	0
3	3	0	0	6	0	0	-0	-1.8	1.2	2	0	-4.1	0	0	0	-1.6	0
13	11	6	1	31	2	2	0	-5.4	5.6	8	2.5	-13.9	0	2	0	-4.8	0
3	3	0	0	6	0	0	-0	-1.8	1.2	2	0	-4.1	0	0	0	-1.6	0
24	7	4	1	36	3	0	-0	-6.0	1.0	10	3.3	-15.7	0	2	0	-5.3	0
28	8	10	0	46	3	0	-0	-7.2	.8	12	5.2	-19.1	0	3	0	-6.3	0
3	4	2	0	9	1	0	-0	-2.3	1.7	2	0	-5.5	0	1	0	-2.1	0
2	3	1	0	6	0	0	-0	-1.8	1.2	2	0	-4.1	0	0	0	-1.6	0
10	6	4	0	20	2	0	-0	-3.9	2.1	5	7	-9.9	0	1	0	-3.5	0
19	16	10	0	45	3	0	-0	-7.0	9.0	12	5.0	-18.8	0	3	0	-6.2	0
21	10	0	2	33	2	0	-0	-5.6	4.4	9	2.8	-14.6	-2.8	2	0	-5.0	0
17	8	3	1	29	2	0	-0	-5.1	2.9	8	2.1	-13.2	0	2	0	-4.5	0
10	3	1	1	15	1	0	-0	-3.2	0	4	0	-8.0	0	1	0	-2.9	0
6	6	7	0	19	1	0	-0	-3.8	2.2	5	5	-9.5	0	1	0	-3.4	0
9	3	9	0	21	2	0	-0	-4.1	0	6	8	-10.3	0	1	0	-3.6	0
6	7	3	0	16	1	0	-0	-3.4	3.6	4	1	-8.3	0	1	0	-3.0	0
7	6	0	0	13	1	0	-0	-2.9	3.1	3	0	-7.1	0	1	0	-2.6	0
3	5	0	7	15	1	0	-0	-3.2	1.8	4	0	-8.0	0	1	0	-2.9	4.1
2	5	1	1	9	1	0	-0	-2.3	2.7	2	0	-5.5	0	1	0	-2.1	0
8	3	2	1	14	1	0	-0	-3.1	0	4	0	-7.6	0	1	0	-2.8	0
5	10	0	2	17	1	0	-0	-3.5	6.5	4	3	-8.7	-3	1	0	-3.1	0
13	6	5	0	24	2	0	-0	-4.5	1.5	6	1.3	-11.4	0	2	0	-4.0	0
4	4	4	5	0	13	1	0	-2.9	1.1	3	0	-7.1	0	1	0	-2.6	0
21	8	4	3	36	3	0	-0	-6.0	2.0	10	3.3	-15.7	0	2	0	-5.3	0
2	4	1	1	8	1	0	-0	-2.1	1.9	2	0	-5.0	0	1	0	-1.9	0
24	9	53	17	103	8	2	2	-13.3	0	27	16.8	-37.7	15.3	7	1.4	-11.6	5.4
147	71	85	18	321	24	14	3	-33.9	37.1	85	66.5	-103.3	0	20	11.3	-29.3	0
11	4	5	1	21	2	0	-0	-4.1	0	6	8	-10.3	0	1	0	-3.6	0
23	7	0	1	31	2	0	-0	-5.4	1.6	8	2.5	-13.9	-2.5	2	0	-4.8	0
4	4	4	0	8	1	0	-0	-2.1	1.9	2	0	-5.0	0	1	0	-1.9	0
8	3	8	2	21	2	0	-0	-4.1	0	6	8	-10.3	0	1	0	-3.6	0
2	7	1	0	10	1	0	-0	-2.5	4.5	3	0	-5.9	0	1	0	-2.2	0
78	20	80	32	210	16	7	8	-23.7	0	56	40.6	-70.4	9.6	13	6.0	-20.5	11.5
12	6	7	5	30	2	0	-0	-5.3	.8	8	2.3	-13.6	0	2	0	-4.6	.4

Jfm	Lax	St	Ól	S: a	TLax Intervall	Sign	TSt Intervall	Sign	TÓl	Intervall	Sign
torsóttligr	1	3	2	0	6	0	-0-1.8	1.2	2	-0-4.1	0
umræða	11	7	0	1	19	1	-0-3.8	3.2	5	-5-9.5	-5
umsjá	4	3	2	0	9	1	-0-2.3	.7	2	-0-5.5	0
umsjá	3	5	0	0	8	1	-0-2.1	2.9	2	-0-5.0	0
uppi	105	31	58	7	201	15	7.3-22.8	8.2	53	38.6-67.7	0
útvíst	6	3	2	0	11	1	-0-2.6	.4	3	-0-6.3	0
útkvæma	0	5	2	0	7	1	-0-2.0	3.0	2	-0-4.6	0
vandlaga, vendliliga	29	9	3	4	45	3	-0-7.0	2.0	12	5.0-18.8	-2.0
vandreiði	45	10	15	2	72	2	-8-10.0	0	19	10.3-27.8	0
varla	20	5	26	8	59	4	-2-8.6	0	16	7.7-23.5	2.5
vaskligr	7	4	5	0	16	1	-0-3.4	.6	4	-1-8.3	0
varkunn	16	5	1	1	23	2	-0-4.4	.6	6	1.2-11.0	-2
veðráttá	2	3	7	0	12	1	-0-2.8	.2	3	-0-6.7	.3
vetrgamall	4	3	1	1	9	1	-0-2.3	.7	2	-0-5.5	0
vetrnatr	6	3	3	0	12	1	-0-2.8	.2	3	-0-6.7	0
vibreign	10	3	0	0	13	1	-0-2.9	.1	3	-0-7.1	0
vingan	8	4	5	0	17	1	-0-3.5	.5	4	-3-8.7	0
vinsell	46	12	11	18	87	7	1.4-11.6	.4	23	13.4-32.6	-2.4
virðing(amaðr)	58	19	8	1	86	6	1.4-11.5	7.5	23	13.2-32.3	-5.2
virðuligr	8	14	6	11	39	3	-0-6.3	7.7	10	3.9-16.7	0
vitsmunur	5	5	0	0	10	1	-0-2.5	2.5	3	-0-5.9	0
vígigr	3	3	2	1	9	1	-0-2.3	.7	2	-0-5.5	0
vænigr	9	6	0	3	18	1	-0-3.7	2.3	5	-4-9.1	-4
vænn	39	23	12	8	82	6	1.2-11.1	11.9	22	12.4-31.0	-4
yfirhit	6	3	1	0	10	1	-0-2.5	.5	3	-0-5.9	0
yfirmaðr	5	3	1	0	9	1	-0-2.3	.7	2	-0-5.5	0
ymissst	8	4	0	0	12	1	-0-2.8	1.2	3	-0-6.7	0
þannig	7	3	28	5	43	3	-0-6.8	0	11	4.6-18.1	9.9
þroski	6	4	3	0	13	1	-0-2.9	1.1	3	-0-7.1	0
þungr	42	10	19	2	73	5	-8-10.2	0	19	10.5-28.1	0
þorf	10	4	1	1	16	1	-0-3.4	.6	4	-1-8.3	0
ætlan	26	19	15	5	65	5	-5-9.3	9.7	17	8.9-25.5	0
cerinn	37	11	4	8	60	5	-3-8.7	2.3	16	7.9-23.8	-3.9



TABELL 3. Minusordlistan.

	Jfm	Lax	St	Ól	S:	a	TLaX Intervall	Sign	TSt Intervall	Sign	TÓl	Intervall	Sign	
almjök	24	1	4	1	30	2	.0-5.3	.0	8	2.3-13.6	.0	2	.0-4.6	.0
alþýða	71	2	13	21	107	8	2.4-13.7	-4	28	17.7-38.9	-4.7	7	1.6-12.0	9.0
andlát	39	2	30	18	89	7	1.5-11.8	.0	24	13.8-33.2	.0	6	9-10.4	7.6
auför	11	0	7	0	18	1	.0-3.7	.0	5	.4-9.1	.0	1	.0-3.3	.0
augurr	50	1	0	0	51	4	.0-7.7	.0	13	6.1-20.8	-6.1	3	.0-6.8	.0
atsókn	21	0	18	7	46	3	.0-7.2	.0	12	5.2-19.1	.0	3	.0-6.3	.7
austmaðr	35	2	26	0	63	5	.4-9.1	.0	17	8.5-24.8	1.2	4	.0-8.0	.0
áglarn	9	0	1	2	12	1	.0-2.8	.0	3	.0-6.7	.0	1	.0-2.5	.0
ákafliga	32	2	18	2	54	4	.0-8.1	.0	14	6.7-21.8	.0	3	.0-7.1	.0
árnaðr	23	0	0	2	25	2	.0-4.6	.0	7	1.5-11.8	-1.5	2	.0-4.1	.0
ártúnaðr	17	0	0	0	17	1	.0-3.5	.0	4	.3-8.7	-3	1	.0-3.1	.0
banahögg	20	0	7	0	27	2	.0-4.9	.0	7	1.8-12.5	.0	2	.0-4.3	.0
bardegi	119	1	49	32	201	15	7.3-22.8	-6.3	53	38.6-67.7	.0	13	5.6-19.8	12.2
bensekr	36	1	0	0	37	3	.0-6.1	.0	10	3.5-16.0	-3.5	2	.0-5.4	.0
bjartr	11	0	3	3	17	1	.0-3.5	.0	4	.3-8.7	.0	1	.0-3.1	.0
blöð	78	4	9	6	97	7	1.9-12.7	.0	26	15.5-35.8	-6.5	6	1.2-11.1	.0
brjóst	37	1	9	1	48	4	.0-7.4	.0	13	5.6-19.8	.0	3	.0-6.5	.0
buandi	20	2	0	0	22	2	.0-4.2	.0	6	1.0-10.6	-1.0	1	.0-3.7	.0
byskup	83	3	708	67	861	65	48.5-80.6	-45.5	228	197.5-257.9	450.1	54	39.6-69.1	.0
dráp	42	3	2	3	50	4	.0-7.6	.0	13	6.0-20.5	-4.0	3	.0-6.7	.0
dröttning	140	1	66	20	227	17	8.8-25.3	-7.8	60	44.5-75.5	.0	14	6.8-21.9	.0
dýrgrípr	17	1	1	1	20	2	.0-3.9	.0	5	.7-9.9	.0	1	.0-3.5	.0
eignask vb.	73	3	1	3	80	6	1.1-10.9	.0	21	12.0-30.4	-11.0	5	6-9.5	.0
einkamal	27	0	14	1	42	3	.0-6.7	.0	11	4.4-17.8	.0	3	.0-5.9	.0
einnæli	15	0	3	0	18	1	.0-3.7	.0	5	.4-9.1	.0	1	.0-3.3	.0
erfbiskup	28	1	235	57	321	24	14.3-33.9	-13.3	85	66.5-103.3	131.7	20	11.3-29.3	27.7
faðerni	9	0	3	0	12	1	.0-2.8	.0	3	.0-6.7	.0	1	.0-2.5	.0
féggjald	11	0	13	4	28	2	.0-5.0	.0	7	2.0-12.8	.2	2	.0-4.4	.0
féggjarn	13	0	0	1	14	1	.0-3.1	.0	4	.0-7.6	.0	1	.0-2.8	.0
finger	10	0	5	0	15	1	.0-3.2	.0	4	.0-8.0	.0	1	.0-2.9	.0

	Jfm	Lax	St	Ól	S:	a	TLax	Intervall	Sign.	TSt	Intervall	Sign.	TÓl	Intervall	Sign.	
fjándmaðr	11	0	0	1	12	1	.0-2.8	.0	3	0-	6.7	0	1	0-	2.5	0
fýja vb.	223	3	47	51	324	24	14.4-34.2	-11.4	86	67.2-104.2	-20.2	20	20	11.4-29.5	21.5	0
forbeini	9	0	1	0	10	1	.0-2.5	.0	3	0-	5.9	0	1	0-	2.2	0
forkunnar	13	0	4	0	17	1	.0-3.5	.0	4	3-	8.7	0	1	0-	3.1	0
forn	78	0	36	10	124	9	3.2-15.4	-3.2	33	21.3-44.2	.0	8	8	2.2-13.4	0	0
forráð	45	1	1	10	57	4	.1-8.4	.0	15	7.3-22.8	-6.3	4	4	0-	7.4	2.6
forstjóri	11	0	1	0	12	1	.0-2.8	.0	3	0-	6.7	0	1	0-	2.5	0
fólk	127	1	52	42	222	17	8.5-24.8	-7.5	59	43.4-74.0	.0	14	14	6.5-21.5	20.5	0
fóstronar	12	0	0	0	12	1	.0-2.8	.0	3	0-	6.7	0	1	0-	2.5	0
fóthvatr	9	0	1	0	10	1	.0-2.5	.0	3	0-	5.9	0	1	0-	2.2	0
freista vb.	29	2	4	1	36	3	0-6.0	.0	10	3.3-15.7	.0	2	2	0-	5.3	0
fullr	89	2	14	2	107	8	2.4-13.7	-4	28	17.7-38.9	-3.7	7	7	1.6-12.0	0	0
fullting	15	0	7	5	27	2	.0-4.9	.0	7	1.8-12.5	.0	2	2	0-	4.3	.7
gagn	30	1	4	0	35	3	.0-5.9	.0	9	3.2-15.3	.0	2	2	0-	5.2	0
galeið	9	0	0	0	9	1	.0-2.3	.0	2	0-	5.5	0	1	0-	2.1	0
gálgi	14	0	0	1	15	1	.0-3.2	.0	4	0-	8.0	0	1	0-	2.9	0
gegnum (f~)	64	4	10	6	84	6	1.3-11.3	.0	22	12.8-31.6	-2.8	5	5	7-	9.9	0
glotta vb.	11	0	0	0	11	1	.0-2.6	.0	3	0-	6.3	0	1	0-	2.4	0
gnógr	25	1	7	0	33	2	.0-5.6	.0	9	2.8-14.6	.0	2	2	0-	5.0	0
greiða vb.	53	1	14	6	74	6	.8-10.3	.0	20	10.7-28.4	.0	5	5	3-	9.0	0
gríð	144	5	202	9	360	27	16.6-37.4	-11.6	95	75.7-114.7	87.3	23	23	13.2-32.3	-4.2	0
grimmr	17	1	2	8	28	2	.0-5.0	.0	7	2.0-12.8	.0	2	2	0-	4.4	3.6
guð	121	3	146	44	314	24	13.8-33.3	-10.8	83	64.8-101.3	44.7	20	20	10.9-28.7	15.3	0
hallari	14	0	0	1	15	1	.0-3.2	.0	4	0-	8.0	0	1	0-	2.9	0
harmingja	22	1	1	1	25	2	.0-4.6	.0	7	1.5-11.8	-5	2	2	0-	4.1	0
hamla vb.	10	0	0	2	12	1	.0-2.8	.0	3	0-	6.7	0	1	0-	2.5	0
handtaaka vb.	27	2	5	3	37	3	.0-6.1	.0	10	3.5-16.0	.0	2	2	0-	5.4	0
hauss	12	0	4	0	16	1	.0-3.4	.0	4	1-8.3	.0	1	1	0-	3.0	0
háseti	48	0	9	1	58	4	2-8.5	-2	15	7.5-23.2	.0	4	4	0-	7.5	0
heilræði	13	0	2	3	18	1	.0-3.7	.0	5	4-9.1	.0	1	1	0-	3.3	0
heimamaðr	49	3	43	0	95	7	1.8-12.5	.0	25	15.1-35.1	7.9	6	6	1.1-10.9	-1.1	0
herbergi	44	3	25	11	83	6	1.2-11.2	.0	22	12.6-31.3	.0	5	5	7-	9.8	1.2

herfang	33	0	6	3	42	3	0	-6.7	0	11	4.4	17.8	0	3	0	-5.9	0	
herfja vb.	183	2	26	35	246	18	65	9.9-27.0	-7.9	65	48.9	81.2	-22.9	16	7.7	-23.4	11.6	
hermaðr	45	1	1	11	58	4	2	-8.5	0	15	7.5	-23.2	-6.5	4	0	-7.5	3.5	
hermaðr	85	1	7	21	114	9	30	2.7-14.4	-1.7	30	19.2	-41.1	-12.2	7	1.8	-12.6	8.4	
heróp	18	1	5	2	26	2	7	1.6	-12.1	0	7	1.6	-12.1	0	2	0	-4.2	0
hersir	60	3	0	0	63	5	4	9.1	0	17	8.5	-24.8	-8.5	4	0	-8.0	0	
herskip	34	1	0	4	39	3	3	0	-6.3	0	10	3.9	-16.7	-3.9	2	0	-5.6	0
herör	17	0	6	0	23	2	6	1.2	-11.0	0	6	1.2	-11.0	0	1	0	-3.9	0
hérblástr	12	0	1	0	13	1	3	0	-7.1	0	3	0	-7.1	0	1	0	-2.6	0
hininn	25	0	2	0	27	2	7	1.8	-12.5	0	7	1.8	-12.5	0	2	0	-4.3	0
hlátr	10	0	0	0	10	1	3	0	-5.9	0	3	0	-5.9	0	1	0	-2.2	0
hlutskipti	10	0	0	1	11	1	3	0	-6.3	0	3	0	-6.3	0	1	0	-2.4	0
horn	40	1	7	1	49	4	13	5.8	-20.2	0	13	5.8	-20.2	0	3	0	-6.0	0
hólmanga	20	0	0	0	20	2	5	7	-9.9	-7	5	7	-9.9	-7	1	0	-3.5	0
hrakning	9	0	2	0	11	1	3	0	-6.3	0	3	0	-6.3	0	1	0	-2.4	0
hugsjúkr	10	0	1	0	11	1	3	0	-2.6	0	3	0	-2.6	0	1	0	-2.4	0
hús	188	9	114	2	313	23	83	64.6	-101.0	13.0	83	64.6	-101.0	20	10.9	-28.7	-8.9	
húsping	18	0	1	1	20	2	5	7	-9.9	0	5	7	-9.9	0	1	0	-3.5	0
hvernig, -ug	71	3	7	1	82	6	22	12.4	-31.0	-5.4	22	12.4	-31.0	5	6	-9.7	0	
höfðingjaleaus	9	0	1	1	11	1	3	0	-2.6	0	3	0	-2.6	0	1	0	-2.4	0
jarldómr	19	0	0	4	23	2	6	1.2	-11.0	-1.2	6	1.2	-11.0	1	0	-3.9	1	
jartegn	80	4	7	6	97	7	26	15.5	-35.8	-8.5	26	15.5	-35.8	6	1.2	-11.1	0	
járn	41	1	13	9	64	5	17	8.7	-25.2	0	17	8.7	-25.2	0	4	0	-8.1	9
kaf	29	0	6	1	36	3	10	3.3	-15.7	0	10	3.3	-15.7	0	2	0	-5.3	0
karl	162	3	25	0	190	14	50	36.1	-64.4	-11.1	50	36.1	-64.4	12	5.1	-18.9	-5.1	
kaupmaðr	49	3	21	4	77	6	20	11.3	-29.4	0	20	11.3	-29.4	0	5	-5	-9.3	0
kaupskip	44	2	0	1	47	4	12	5.4	-19.5	-5.4	12	5.4	-19.5	3	0	-6.4	0	
kennimaðr	21	0	13	8	42	3	11	4.4	-17.8	0	11	4.4	-17.8	0	3	0	-5.9	2.1
kesja	29	0	2	16	47	4	12	5.4	-19.5	-3.4	12	5.4	-19.5	3	0	-6.4	9.6	
konungdóm	68	1	33	12	114	9	30	19.2	-41.1	0	30	19.2	-41.1	0	7	1.8	-12.6	0
konungsmaðr	31	1	28	6	66	5	17	9.1	-25.8	2.2	17	9.1	-25.8	2.2	4	1	-8.3	0
konungsskip	18	1	28	1	45	3	12	5.0	-18.8	6.2	12	5.0	-18.8	6.2	3	0	-6.2	0
krefja	35	1	5	9	50	4	13	6.0	-20.5	-1.0	13	6.0	-20.5	3	0	-6.7	2.3	
kross	30	1	11	0	42	3	11	4.4	-17.8	0	11	4.4	-17.8	0	3	0	-5.9	0

	Jfm	Lax	St	Ól S: a	T Lax Intervall	Sign	T St Intervall	Sign	T Ól Intervall	Sign				
landráð	18	0	6	0	24	2	.0-4.5	.0	6	1.3-11.4	.0	2	.0-4.0	.0
landvegr	9	0	1	0	10	1	.0-2.5	.0	3	.0-5.9	.0	1	.0-2.2	.0
landsfólk	42	0	6	2	50	4	.0-7.6	.0	13	6.0-20.5	.0	3	.0-6.7	.0
landsgæzla	9	0	3	2	14	1	.0-3.1	.0	4	.0-7.6	.0	1	.0-2.8	.0
langfögar	11	0	1	2	14	1	.0-3.1	.0	4	.0-7.6	.0	1	.0-2.8	.0
leynuliga	16	1	14	1	32	2	.0-5.5	.0	8	2.6-14.3	.0	2	.0-4.9	.0
líðsafaðr, -samnaðr	17	0	1	3	21	2	.0-4.1	.0	6	.8-10.3	.0	1	.0-3.6	.0
liðsemð	12	0	0	0	12	1	.0-2.8	.0	3	.0-6.7	.0	1	.0-2.5	.0
liðveizla	71	4	21	4	100	8	2.0-13.0	.0	26	16.2-36.7	.0	6	1.3-11.3	.0
líkarni, -mr	13	0	4	6	23	2	.0-4.4	.0	6	1.2-11.0	.0	1	.0-3.9	2.1
lögmaðr	37	1	55	0	93	7	1.7-12.3	-7	25	14.7-34.5	20.5	6	1.0-10.7	-1.0
læknir	11	0	4	0	15	1	.0-3.2	.0	4	.0-8.0	.0	1	.0-2.9	.0
mannfall	27	0	3	13	43	3	.0-6.8	.0	11	4.6-18.1	-1.6	3	.0-6.0	7.0
metnaðr	15	0	0	0	15	1	.0-3.2	.0	4	.0-8.0	.0	1	.0-2.9	.0
miskunn	20	0	37	8	65	5	.5-9.3	-5	17	8.9-25.5	11.5	4	1-8.2	.0
móttull	11	0	0	1	12	1	.0-2.8	.0	3	.0-6.7	.0	1	.0-2.5	.0
naúðsyn	48	2	14	12	76	6	.9-10.5	.0	20	11.1-29.1	.0	5	4-9.2	2.8
náttverð	9	0	2	0	11	1	.0-2.6	.0	3	.0-6.3	.0	1	.0-2.4	.0
njósnarnaðr	20	0	20	3	43	3	.0-6.8	.0	11	4.6-18.1	1.9	3	.0-6.0	.0
óbalborinn	10	0	1	0	11	1	.0-2.6	.0	3	.0-6.3	.0	1	.0-2.4	.0
ófríðr	75	2	108	31	216	16	8.2-24.2	-6.2	57	42.0-72.2	35.8	14	6.3-21.0	10.0
ósukurð	23	0	22	1	46	3	.0-7.2	.0	12	5.2-19.1	2.9	3	.0-6.3	.0
ósáttir	17	0	4	1	22	2	.0-4.2	.0	6	1.0-10.6	.0	1	.0-3.7	.0
óvigr	17	1	3	2	23	2	.0-4.4	.0	6	1.2-11.0	.0	1	.0-3.9	.0
prestr	84	2	232	26	344	26	15.6-36.0	-13.6	91	71.9-110.1	121.9	22	12.4-31.0	.0
ráðarnaðr	10	0	3	2	15	1	.0-3.2	.0	4	.0-8.0	.0	1	.0-2.9	.0
rangr	16	0	3	2	21	2	.0-4.1	.0	6	.8-10.3	.0	1	.0-3.6	.0
rás	12	0	2	0	14	1	.0-3.1	.0	4	.0-7.6	.0	1	.0-2.8	.0
reising	18	1	0	8	27	2	.0-4.9	.0	7	1.8-12.5	-1.8	2	.0-4.3	3.7
ríkismaðr	48	1	0	10	59	4	.2-8.6	.0	16	7.7-23.5	-7.7	4	.0-7.6	2.4
róg	13	0	1	0	14	1	.0-3.1	.0	4	.0-7.6	.0	1	.0-2.8	.0
saklauss	10	0	9	4	23	2	.0-4.4	.0	6	1.2-11.0	.0	1	.0-3.9	.1

sarminceðr	10	0	4	2	16	1	0	-3.4	0	4	1	-8.3	0	1	0	-3.0	0
sannindi	22	1	7	0	30	2	0	-5.3	0	8	2.3	-13.6	0	2	0	-4.6	0
sáttmál	19	1	1	0	21	2	0	-4.1	0	6	8	-10.3	0	1	0	-3.6	0
sendiför	9	0	1	1	11	1	0	-2.6	0	3	0	-6.3	0	1	0	-2.4	0
sendimaðr	155	0	55	10	220	17	8.4	-24.6	-8.4	58	42.9	-73.4	0	14	6.4	-21.3	0
siðvenja	22	0	8	0	30	2	0	-5.3	0	8	2.3	-13.6	0	2	0	-4.6	0
sigr	164	1	9	32	206	15	7.6	-23.3	-6.6	54	39.7	-69.2	-30.7	13	5.8	-20.2	11.8
sigrsaell	16	0	0	2	18	1	0	-3.7	0	5	4	9.1	-4	1	0	-3.3	0
sjaölderni	16	0	22	1	39	3	0	-6.3	0	10	3.9	-16.7	5.3	2	0	-5.6	0
sjúkr	15	0	14	1	30	2	0	-5.3	0	8	2.3	-13.6	4	2	0	-4.6	0
skáld	100	1	13	5	119	9	3.0	-14.9	-2.0	31	20.3	-42.7	-7.3	8	2.0	-13.0	0
skildagi	11	0	6	0	17	1	0	-3.5	0	4	3	-8.7	0	1	0	-3.1	0
skip(a)kostr	9	0	0	1	10	1	0	-2.5	0	3	0	-5.9	0	1	0	-2.2	0
skjaldborg	16	0	3	1	20	2	0	-3.9	0	5	7	-9.9	0	1	0	-3.5	0
slátr	11	0	2	0	13	1	0	-2.9	0	3	0	-7.1	0	1	0	-2.6	0
snarpr	27	1	4	0	32	2	0	-5.5	0	8	2.6	-14.3	0	2	0	-4.9	0
sóttmaðr	23	0	0	4	27	2	0	-4.9	0	7	1.8	-12.5	-1.8	2	0	-4.3	0
spakr	35	2	3	8	48	4	0	-7.4	0	13	5.6	-19.8	-2.6	3	0	-6.5	1.5
stallari	36	1	7	0	44	3	0	-6.9	0	12	4.8	-18.5	0	3	0	-6.1	0
stefnulaag	25	0	3	2	30	2	0	-5.3	0	8	2.3	-13.6	0	2	0	-4.6	0
stilltr	15	0	6	2	23	2	0	-4.4	0	6	1.2	-11.0	0	1	0	-3.9	0
stræti	18	1	15	0	34	3	0	-5.7	0	9	3.0	-15.0	0	2	0	-5.1	0
svarðagi	19	0	0	5	24	2	0	-4.5	0	6	1.3	-11.4	-1.3	2	0	-4.0	1.0
svikræði	9	0	1	3	13	1	0	-2.9	0	3	0	-7.1	0	1	0	-2.6	4
synja	26	1	4	0	31	2	0	-5.4	0	8	2.5	-13.9	0	2	0	-4.8	0
sýslumaðr	15	0	58	1	74	6	8	-10.3	-8	20	10.7	-28.4	29.6	5	3	-9.0	0
tilkall	21	0	2	0	23	2	0	-4.4	0	6	1.2	-11.0	0	1	0	-3.9	0
titt	45	3	15	0	63	5	4	-9.1	0	17	8.5	-24.8	0	4	0	-8.0	0
tryggð	14	1	1	0	16	1	0	-3.4	0	4	1	-8.3	0	1	0	-3.0	0
umbúnaðr	15	0	3	1	19	1	0	-3.8	0	5	5	-9.5	0	1	0	-3.4	0
umhverfis	20	0	10	3	33	2	0	-5.6	0	9	2.8	-14.6	0	2	0	-5.0	0
underligr	48	3	18	4	73	5	8	-10.2	0	19	10.5	-28.1	0	5	3	-8.9	0
upphaf	69	2	4	4	79	6	1.1	-10.8	0	21	11.8	-30.0	-7.8	5	5	-9.5	0
uppreist	29	1	3	2	35	3	0	-5.9	0	9	3.2	-15.3	-2	2	0	-5.2	0

	Jfm	Lax	St	Ól	S: a	TLex	Intervall	Sign	TSt	Intervall	Sign	TÓl	Intervall	Sign
útböð	14	0	7	5	26	2	0-4.7	.0	7	1.6-12.1	.0	2	0-4.2	.8
útlagi, -gr	34	0	4	0	38	3	0-6.2	.0	10	3.7-16.4	.0	2	0-5.5	.0
vápnðjarfr	11	0	0	3	14	1	0-3.1	.0	4	0-7.6	.0	1	0-2.8	.2
vápnlaus	12	0	2	3	17	1	0-3.5	.0	4	3-8.7	.0	1	0-3.1	.0
verkmaðr	14	0	1	0	15	1	0-3.2	.0	4	0-8.0	.0	1	0-2.9	.0
vetrvist	11	0	2	0	13	1	0-2.9	.0	3	0-7.1	.0	1	0-2.6	.0
vinátta	158	6	54	19	237	18	9.3-26.2	-3.3	63	46.8-78.5	.0	15	7.2-22.7	.0
vináttumál	9	0	0	2	11	1	0-2.6	.0	3	0-6.3	.0	1	0-2.4	.0
vingjöf	20	0	0	3	23	2	0-4.4	.0	6	1.2-11.0	-1.2	1	0-3.9	.0
vinstri	18	0	9	0	27	2	0-4.9	.0	7	1.8-12.5	.0	2	0-4.3	.0
vígsnál	9	0	5	0	14	1	0-3.1	.0	4	0-7.6	.0	1	0-2.8	.0
vörn	51	2	43	7	103	8	2.2-13.3	-2	27	16.8-37.7	5.3	7	1.4-11.6	.0
yfirhöfn	9	0	0	0	9	1	0-2.3	.0	2	0-5.5	.0	1	0-2.1	.0
yfirsókn	21	0	0	1	22	2	0-4.2	.0	6	1.0-10.6	-1.0	1	0-3.7	.0
þakksamligr	9	0	0	2	11	1	0-2.6	.0	3	0-6.3	.0	1	0-2.4	.0
þilja	11	0	0	0	11	1	0-2.6	.0	3	0-6.3	.0	1	0-2.4	.0
þingboð	9	0	2	0	11	1	0-2.6	.0	3	0-6.3	.0	1	0-2.4	.0
þingmaðr	20	0	22	2	44	3	0-6.9	.0	12	4.8-18.5	3.5	3	0-6.1	.0
þjóð	43	0	1	0	44	3	0-6.9	.0	12	4.8-18.5	-3.8	3	0-6.1	.0
þjónusta	34	2	11	6	53	4	0-8.0	.0	14	6.5-21.5	.0	3	0-7.0	.0
þola	42	2	12	6	62	5	3-9.0	.0	16	8.3-24.5	.0	4	0-7.9	.0
þreyta	19	1	1	0	21	2	0-4.1	.0	6	8-10.3	.0	1	0-3.6	.0
örendislok	11	0	1	0	12	1	0-2.8	.0	3	0-6.7	.0	1	0-2.5	.0
örpa	29	1	2	4	36	3	0-6.0	.0	10	3.3-15.7	-1.3	2	0-5.3	.0

## Bibliografi

- Albano Leoni, F., Sagas islandaises et statistique linguistique. Quelques observations. (Arkiv för nordisk filologi 85 (1970), s. 138–162.)
- Allén, S., Nusvensk frekvensordbok baserad på tidningstext. 1. 1970.
- Allén, S., Vocabulary data processing. (The Nordic languages and modern linguistics. Ed. by H. Benediktsson. Reykjavík 1970, s. 235–261.)
- Ellegård, A., A statistical method for determining authorship. 1962.
- Francis, I. S., An exposition of a statistical approach to the Federalist dispute. (The computer & literary style. Ed. by J. Leed. Kent, Ohio 1966, s. 38–78.)
- Hallberg, P., Íslendinga saga och Egla, Laxdæla, Eyrbyggja, Njála, Grettla. *Ett språktest*. (Mål og Minne 1965, s. 89–105.)
- Hallberg, P., Ólafr Þórðarson hvítaskáld, Knýtlinga saga och Laxdæla saga. Ett försök till språklig författarbestämning. Reykjavík 1963.
- Hallberg, P., Snorri Sturluson och Egils saga Skallagrímssonar. Ett försök till språklig författarbestämning. Reykjavík 1962.
- Hallberg, P., Stilsignalement och författarskap i norrön sagalitteratur. 1968.
- Haugen, E., Norwegian word studies. 2. The vocabularies of the Old Norse sagas and of Henrik Wergeland. Madison, Wisc. 1942. [Dupl.]
- Heller, R., Der Verfasser der Laxdæla saga und sein Verhältnis zur Sturlubók. (Afmælisrit Jóns Helgasonar 30. júní 1969. Reykjavík 1969, s. 80–91.)
- Heller, R., Studien zu Aufbau und Stil der Laxdæla saga. (Arkiv för nordisk filologi 75 (1960), s. 113–167.)
- Hesselman, B., Giöta Kiämpa-wisa. En språk-estetisk undersökning. (Språk och stil 7 (1907, tr. 1908), s. 201–236.)
- Johannisson, T., Det dolda stilmönstret. (Dagens Nyheter 29/7 1961.)

- Johannisson, T., Ett språkligt signalement. (Dagens Nyheter 1/8 1961.)
- Johannisson, T., Utredning angående författarskapet till de anonyma breven i Helandermålet. 1960. [Dupl.]
- McKinnon, A., och Webster, R., A method of 'author' identification. (Computer studies in the humanities and verbal behavior 2 (1969), s. 19–23.)
- Meier, M., se Mundt, M.
- Moerk, E.L., Quantitative analysis of writing styles. (Journal of linguistics 6 (1970), s. 223–230.)
- Mosteller, F., och Wallace, D.L., Inference and disputed authorship: *The Federalist*. Reading, Mass. ... 1964.
- Muller, Ch., Initiation à la statistique linguistique. Paris 1968.
- Mundt, M., Sturla Þórdarson und die Laxdæla saga. Bergen ... 1969.
- Mundt (f. Meier), M., Til spørsmålet om Laxdælas forfatter. (Mål og Minne 1965, s. 55–62.)
- Noreen, E., Från Birgitta till Piraten. 1942.
- Noreen, E., Författarfrågor i Lejonkulans dramatik. 1938.
- Radday, Y.T., Isaiah and the computer: A preliminary report. (Computers and the humanities 5 (1970), s. 65–73.)
- Wickmann, D., Eine mathematisch–statistische Methode zur Untersuchung der Verfasserfrage literarischer Texte. Köln und Opladen 1969.



## Vocabulary Data Processing

The Nordic Languages and Modern Linguistics. 1970.

An assumption underlying the project briefly described in this report is that linguists should be concerned with the study of language material as well as with the formulation of linguistic models. There are thus four main divisions of work: the establishment of a corpus, the designing of models of description, the specification of the technical treatment required to attain the descriptive goals, and the description itself.

The project was outlined in 1964 (Allén 1966:4). I set up, as a primary aim, the analysis of the vocabulary system of present-day written Standard Swedish, represented by a corpus of prose from prominent morning newspapers. The project was based from the very beginning on automatic data processing.

In 1966 a research group, *Forskningsgruppen för modern svenska* (Research Group for Modern Swedish), was established at the University of Göteborg in order to carry out the project. Since its establishment, its work has been sponsored by The Bank of Sweden Tercentenary Foundation. At present the group has seven members: the research director, four research assistants (two full-time, two half-time), a programmer, and a secretary/punch-operator. In addition, some special studies are being done outside the group.

A concentrated flow chart of the current shape of the project is given in Figure 1. It will be seen that the central column includes the publication of two frequency dictionaries (E, H) and a monograph (J). The wing columns contribute some other lexical work and a number of research reports on grammatical questions.

### Corpus

The material was received in the form of punched Teletype tapes from the following five newspapers: *Svenska Dagbladet* (Stock-

holm, conservative), *Dagens Nyheter* (Stockholm, liberal), *Stockholms-Tidningen* (Stockholm, social-democratic), *Göteborgs Handels- och Sjöfarts-Tidning* (Göteborg, liberal), and *Sydsvenska Dagbladet Snällposten* (Malmö, conservative). These papers represent the most significant geographical regions and the principal political parties in Sweden. The collecting of the material took place in 1965. The whole year was covered according to the time-table in Table 1 (a comment: GHT doesn't appear on Sundays).

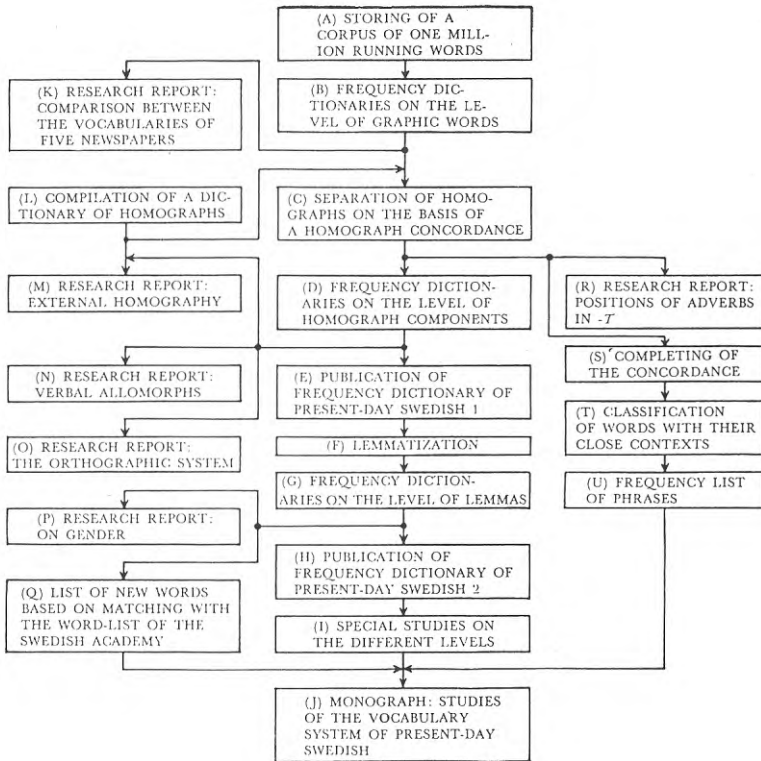


FIGURE 1. Flow Chart of the Project.

TABLE 1. *Time-table of the Collecting of the Material.*

GHT	SVD	ST	DN	SDS
1/1-17/1	18/1-31/1	1/2-14/2	15/2-28/2	1/3-14/3
15/3-28/3	29/3-11/4	12/4-25/4	26/4-9/5	10/5-23/5
24/5-6/6	7/6-20/6	21/6-4/7	5/7-18/7	19/7-1/8
2/8-15/8	16/8-29/8	30/8-12/9	13/9-26/9	27/9-10/10
11/10-24/10	25/10-7/11	8/11-21/11	22/11-5/12	6/12-19/12
20/12-31/12				

The punched tapes were fed into the computer, and the text was stored on magnetic tape. The resulting corpus of one million running words consists of 1,387 articles by 569 authors. The articles belong to three genres: general news reports (43 %), articles on cultural subjects (appearing on fixed pages; 46 %), and reports from foreign correspondents (11 %). The articles of the corpus have also been divided into six topic areas.

The size of the corpus deserves a comment. It should be stressed that there is no upper limit to a vocabulary system; among other things, such infinite sub-classes as compounds and numerals illustrate this point. An enlargement of the corpus thus tends to add to the number of different words, although the list will never be complete. This being the case, the strategy should be as follows. In vocabulary research two types of complementary investigations ought to be undertaken: description on a less advanced level of exceptionally large bodies of material, and description on more advanced levels of large, but by comparison restricted, bodies of material. The present project is of the second type. A project of the first type could be applied to a multi-million sample of a whole year's printing (apart from job-printing). In the case of Swedish, I estimate the total volume at very approximately two thousand million running words.

### **Data Processing**

Time will not permit me to enter upon the technicalities connected with the data processing (some hints are given in Allén 1968;

see also Allén 1969; general introductions are Borko 1967 and Hays 1967). The computer used is a Datasab (D 21). A special system of programmes, called *Språkdata* (Language Data), has been worked out for the purpose of the project. So far the system comprises about twenty different complexes of programmes. Projects in Lund and Helsingfors are also making use of it. The main programming language is Algol-Genius.

### Operative Units

A fundamental question in a quantitative investigation is what units should be counted. The descriptive levels distinguished in the present project are shown in Figure 2. For purposes of practical procedure the input is a string of graphemes (including spaces). A graphic word is defined as a sequence of graphemes delimited by two consecutive spaces. Thus there is a syntagmatic relation (marked -S-) between graphemes and graphic words. There are four kinds of graphic words: alphabetic words (e.g., *enheter* 'units'), hybrid words (e.g., *svensk-isländsk* 'Swedish-Icelandic', *1900-talet* 'the twentieth century', *10.30*), numerical words (e.g., *1900*), and abbreviations or logograms (e.g., %). Let me stress here that, in my opinion, every instance of a grapheme or a graphic word should be counted (cf. Ellegård 1960:222). The omission of, say, proper names or numerical words would distort calculations relating both to the text and to the vocabulary system.

Graphic words in their turn form phrases, phraseological units such as *i varje fall* 'in any case', *i första hand* 'in the first place', *kort sagt* 'in short', and *glömma bort* 'forget' (level 3). Such recurrent combinations are also regarded as relevant to the study of the vocabulary system.

The procedure called here separation of homographs enables us to describe the vocabulary on other interesting levels. A crucial question is how homography should be defined. As far as I can see, it is not feasible to draw a strict line between homography and polysemy on semantic grounds. To take just one

example, how many units should be distinguished in an ordinary word like *blad*, which includes among its meanings 'leaf (on a tree)', 'petal', 'blade (of a knife)', 'blade (of an oar)', 'bowl (of a spoon)', 'sheet (of paper)', 'leaf (in a book)', 'plate', and 'news-paper'? My suggestion is that formal criteria be chosen instead.

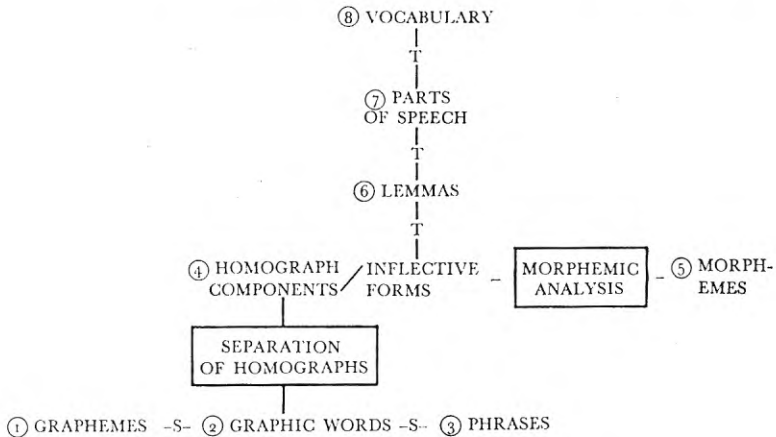


FIGURE 2. *Levels of Description.*

As a first step in the separation of homographs, components belonging to different parts of speech can be distinguished: *filar*, noun (plur. of *fil* 'file'), and *filar*, verb (pres. of *fila* '(to) file'), are separated, and so are *muttrar* 'screw-nuts' and *muttrar* '(he) mutters'. This presupposes a definition of the parts of speech. In the present case a flow-chart grammar that distributes the units among the parts of speech has been worked out. It is based on selected and ordered criteria, in the first place from syntax, in the second place from morphology and graphemic and phonemic form.

A second step is the isolation of components belonging to different lemmas. A lemma is defined as a group of forms within a part of speech which are assignable either to one and the same series of inflexion (in the case of indeclinable words comprising only one form) or to several series of inflexion that converge in

speech and/or writing, the divergences of which show purely facultative variation (free variation). A series of inflexion is of the type *fil, fils, filen, filens, filar, filars, filarna, filarnas* 'file, of a file, the file, etc.', where grammatical constants of a very generic character are added. A flow chart determining the lemmatic status of series of inflexion is given in Figure 3. Homograph components distinguished at this stage are, e.g., *fil* (plur. *filar* 'files') and *fil* (plur. *filer* 'rows'), *nit* (plur. *nitar* 'rivets') and *nit* (plur. *nitar/niter* 'blanks'). On the other hand, a case like *svan* (plur. *svanar/svanor* 'swans') is found to be one lemma. It should be pointed out that supplementary inflexion is not accepted. No one can tell whether, for instance, *värre* in the expression *en värre situation* 'a worse situation' is the comparative of *svår* or *besvärlig* or *vansklig* or *krånglig*, etc., whereas probably any Swedish speaker can tell that it is not the comparative of *ond*, which is the traditional analysis. The series in question is therefore: –, *värre*, *värst*. The kind of homography discussed so far I call external homography, i.e., homography between members of different lemmas.

There is also a kind of homography that can be called internal or sub-lemmatic, viz., homography between members of the same lemma, and a third step that can be taken is the separation of homographs of this type. Then *filade* (pret.) is distinguished from *filade* (past part.), and so is *blad* (sing.) from *blad* (plur.). This separation is based on patterns in the morphological system such as *skrev* 'wrote', *skrivna* 'written' and *fil* 'file', *filar* 'files', respectively.

Provided that a full-scale separation has been carried out, the units on level 4 (in Figure 2) are of four kinds:

THE FORM BELONGS TO A	DECLINABLE LEMMA	INDECLINABLE LEMMA
HOMOGRAPH	<i>muttrar</i> <sub>vb</sub> 'mutters'	<i>bara</i> <sub>adv</sub> 'only'
HETEROGRAPH	<i>skriver</i> 'writes'	<i>inte</i> 'not'

However, the most characteristic feature of the level is that homographs are separated. For this reason it can be called the level of

homograph components. Since the units are also inflective forms (in the wide sense hinted at above), it can also be called the level of inflective forms.

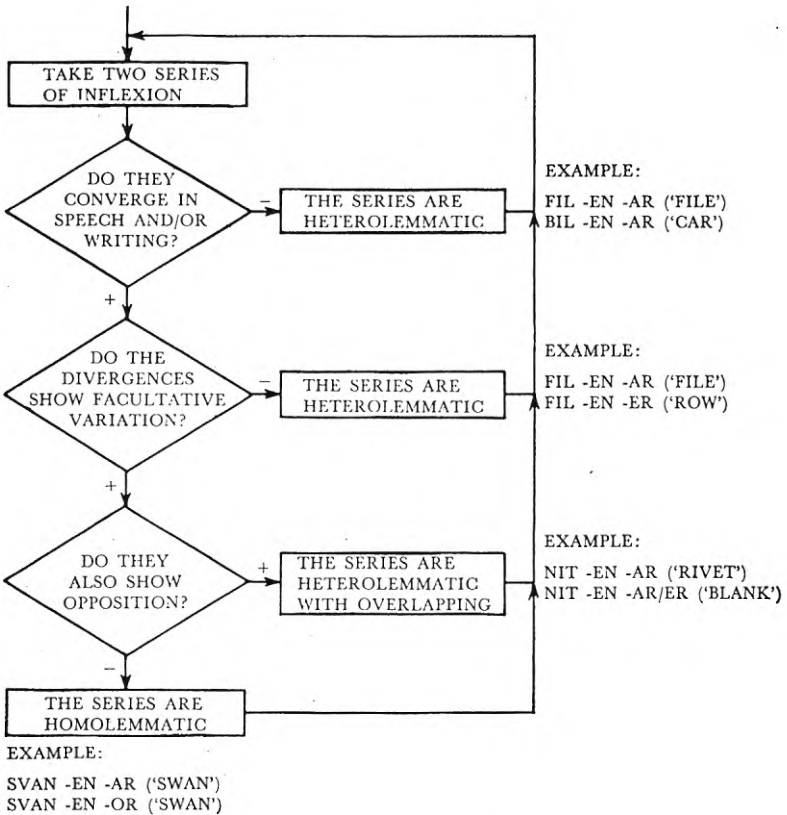


FIGURE 3. *Determination of the Lemmatic Status of Series of Inflexion.*

Between the units on the level just mentioned and the units on levels 6, 7, and 8 there is a so-called taxonomic relation (marked -T-); that is, each and every unit on a certain level can be classified or sub-divided in only one way. This is visualized in Figure 4.

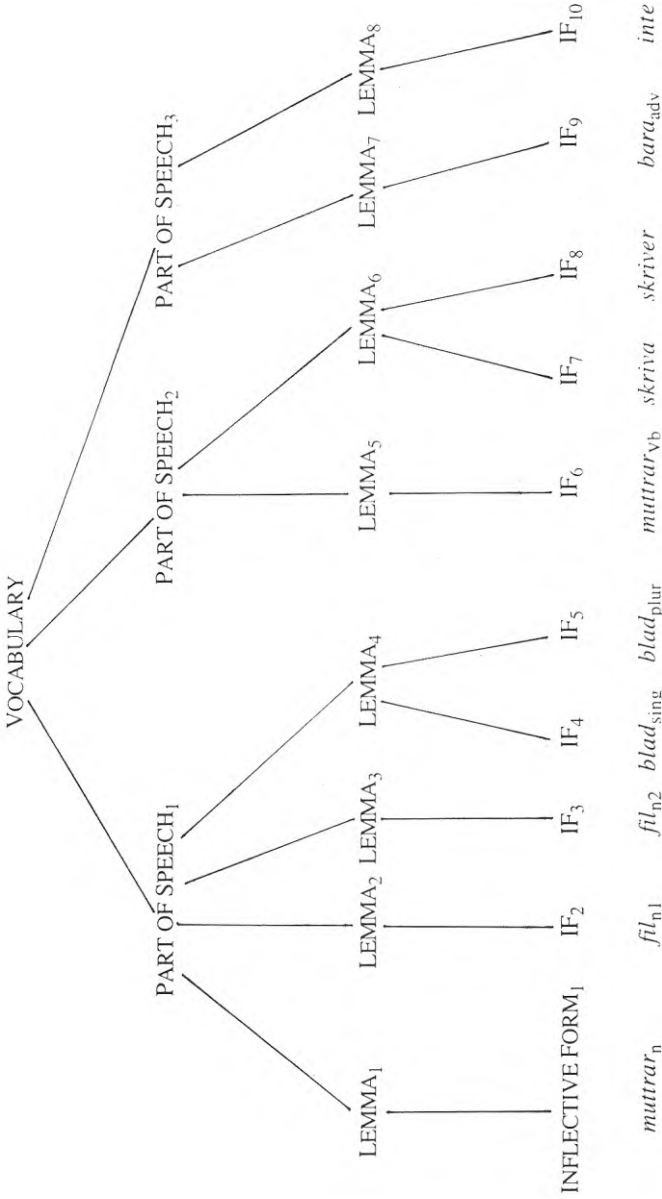


FIGURE 4. Taxonomic Relations between Units on Certain Levels.



The last level indicated in Figure 2 is the level of morphemes, which can be reached by the procedure designated as morphemic analysis. There are of course other levels of interest. One such is formed by categories of word formation, e.g., *bära* 'carry', *bår* 'hand-barrow', *börd* 'birth', *bärrem* 'porter's strap', etc. It goes without saying that such semantic features as can be discerned are of interest, too. I might emphasize that there is nothing in the present analysis that would impede semantic work. On the contrary, it would be very much to the point to do such work within the formal framework established.

The fact that some processes in the analysis presuppose an understanding of the text does not mean that the analysis is based on non-formal criteria. One way to demonstrate this is by a comparison with traditional lexicography, where semantics is one of the fundamental factors. One finds that *grund* 'shallow; (a) shoal' is treated as two units, whereas *djup* 'deep; depth' is treated as one; *ask*, *-ar* 'ash-tree; box' is treated as two units, whereas *konst*, *-er* 'art; pump of a mine' is treated as one; *ned* 'down' and *ner* 'down' are treated as one unit, whereas *icke* 'not' and *inte* 'not' are treated as two, etc. In the present analysis, *djup* as well as *grund* is treated as two units (adjective, noun), *ask* as well as *konst* is treated as one (polysemous) unit (each of the two forms belongs to one single series of inflexion), *ned* and *ner* are treated as two units as well as *icke* and *inte* (neither pair showing any convergence).

### Lexical Treatments

The first lexical result was achieved in 1967, when a frequency dictionary in three parts was produced. This dictionary is entitled 'Frekvenser i svenskan 1965' (Frequencies in Swedish 1965). It records the words on the level of graphic words, part 1 in rank order, part 2 in alphabetical order, and part 3 in reverse alphabetical order. The dictionary has not been published. The research group has had six copies of each part bound. These

copies have been printed by the lineprinter. The dictionary is also available on magnetic tape.

Each of the three parts contains over 1,700 pages in folio size. Under each word one finds the rank number, the total frequency, the frequency in each of the five newspapers, in each of the three genres, and in each of the six topic areas, the number of articles and authors involved, and the number of running words in the relevant article(s). The number of different graphic words is 103,799. The four types of graphic words are kept apart. The distribution of the units over these four types is shown in Table 2.

TABLE 2. *The Distribution of Graphic Words by Word-types.*

	Running Words		Different Words		Average Value of Frequencies	Mean Value of Frequencies
	Abs	Rel	Abs	Rel		
Alphabetic Words	983,220	98.3	98,052	94.5	10.0	1
Hybrid Words	9,538	1.0	4,821	4.6	2.0	1
Numerical Words	7,855	0.8	920	0.9	8.5	2
Logograms	75	0.0	6	0.0	12.5	5.5
Total	1,000,688	100.1	103,799	100.0	9.6	1

In 1968 a second lexical result was achieved. This dictionary is called 'Basvokabulär på grafordsnivå' (Basic Vocabulary on the Level of Graphic Words). It is based on what I call modified frequency (FMOD). This is the product of observed frequency and an index of dispersion. The formula for the latter is

$$D = 1 - \frac{s}{m\sqrt{n-1}}$$

where  $s$  is the standard deviation,  $n$  the number of sub-texts, and  $m$  the average frequency. This formula is the generalized form of a formula used by Juilland and Chang-Rodriguez (1964:LIII). The five newspapers have been chosen as subtexts. The value of

*D* varies from 1 (absolutely even distribution) to 0 (occurrences in only one sub-text). In order to qualify for the basic vocabulary, a graphic word is required to exhibit a minimum FMOD-value of 4.125. This means that a unit with an observed frequency of 5 (which is the lowest frequency taken into account) is admitted to the basic vocabulary only on condition that the five occurrences are evenly distributed, with one in each of the five newspapers.

Out of the 16,946 words that have an observed frequency of 5 or more, 10,605 qualify for the basic vocabulary. Having established a corpus of the present kind and defined the concepts of graphic word and basic vocabulary in the manner stated, the following approximate series is obtained:

Number of Running Graphic Words	1,000,000
Number of Different Graphic Words	100,000
Number of Graphic Words in the Basic Vocabulary	10,000

It is also interesting to find that the words which make up the basic vocabulary account for 83 % of the corpus.

The figures given in the preceding paragraph show that 37 % of the potential members of the basic vocabulary fail to qualify. The list of disqualified alphabetic words is headed by words like *Krupp*, *Powys*, (*i fjor* 'last year' (South Swedish variant of *ifjol*), *Culbertson*, *Påskön* 'Easter Island', *Stadshuset* 'the Town Hall', *fasaner* 'pheasants', *Aronsson*, and *Missionsförbundet* 'the Missionary Society', with values of observed frequency between 59 and 20. Proper names predominate at the top of the list.

Just like 'Frequencies in Swedish 1965', the 'Basic Vocabulary' is unpublished but available in the form of line-printed copies and magnetic tape. Figure 5 is a specimen page from the alphabetical version. DT means dispersion over the five newspapers (only decimals are given). ÄMN indicates occurrence in the six topic areas. FÖR means number of authors. F is the observed frequency and FDIF the difference between F and FMOD. TIDN indicates occurrence in the newspapers. DÄ means dispersion over the topic areas. Finally, X2 is the value of chi-square ( $\chi^2$ ), calculated on the basis of the newspapers.

RNR		FMOD	DT	ÄMN	FÖR	F	FDIF	TIDN	DÄ	X2
6945	RESULTATET	115.9	852	1-6	89	136	20	1-5	850	14.0
6946	RESULTERA	5.176	518	1235	8	10	5	235	524	
6947	RESULTERADE	11.89	743	1-6	15	16	4	1235	635	
6948	RESULTERAT	7.645	478	1-6	15	16	8	125	758	
6949	RESURSER	84.55	813	1-6	72	104	19	1-5	786	12.4
6950	RESURSERNA	13.47	792	2346	15	17	4	1-5	521	
6951	RETA	6.182	773	23456	8	8	2	1-5	715	
6952	RETAR	6.182	773	1256	8	8	2	1-5	656	
6953	RETRÄTT	8.882	683	2356	13	13	4	1-5	643	
6954	REUTERSWÄRD	10.01	834	25	7	12	2	1-5	258	
6955	REVOLT	9.750	696	236	12	14	4	1-5	437	
6956	REVOLTEN	5.245	525	236	8	10	5	2345	500	
6957	REVOLUTION	25.43	637	1-6	26	37	12	1-5	823	
6958	REVOLUTIONEN	33.40	682	12356	28	49	16	1-5	756	24.4
6959	REVOLUTIONENS	6.988	333	12356	8	21	14	2345	519	
6960	REVOLUTIONER	4.395	488	236	7	9	5	2345	398	
6961	REVOLUTIONERANDE	10.01	834	1-6	12	12	2	1-5	776	
6962	REVOLUTIONÄRA	12.76	672	2346	13	19	6	1-5	593	
6963	REVV	9.468	676	1256	8	14	5	1-5	580	
6964	RHODESIA	60.97	762	345	10	80	19	1-5	111	15.4
6965	RHODESIAS	14.15	566	3	5	25	11	2345	000	
6966	RHODESISKA	5.237	374	3	4	14	9	245	000	
6967	RICHARD	27.27	802	236	28	34	7	1-5	507	
6968	RIDER	6.890	689	2356	9	10	3	1245	587	
6969	RIDÄN	7.069	785	2356	8	9	2	1-5	430	
6970	RIK	23.38	866	1-6	23	27	4	1-5	807	
6971	RIKA	53.98	885	1-6	50	61	7	1-5	868	2.68
6972	RIKARE	13.62	801	1-6	17	17	3	1-5	707	
6973	RIKE	12.25	875	12356	13	14	2	1-5	675	
6974	RIKEDOM	42.51	736	12356	17	17	4	1245	509	
6975	RIKET	8.458	651	2356	11	13	5	1-5	385	
6976	RIKETS	10.63	665	2345	13	16	5	1235	553	
6977	RIKLIIG	9.347	719	1235	9	13	4	1-5	278	
6978	RIKLIIGA	4.474	746	1456	6	6	2	2345	587	
6979	RIKLIIGT	8.890	684	12456	11	13	4	1245	641	
6980	RIKSBANKEN	6.282	698	346	5	9	3	1235	332	
6981	RIKSDAG	11.63	775	1234	11	15	3	1-5	606	
6982	RIKSDAGEN	35.66	743	1-6	24	48	12	1-5	550	15.0
6983	RIKSDAGENS	12.34	686	2345	12	18	6	1235	514	
6984	RIKSDAGSMAN	19.41	844	1-6	19	23	4	1-5	843	
6985	RIKSDAGSMÄN	4.776	531	23	6	9	4	235	160	
6986	RIKSFÖRBUND	11.94	746	2345	14	16	4	1-5	577	
6987	RIKSTEATERN	6.837	360	26	4	19	12	234	105	
6988	RIKT	43.55	726	1-6	49	60	16	1-5	774	18.9
6989	RIKTA	11.07	851	12346	13	13	2	1-5	712	
6990	RIKTAD	6.759	751	12346	9	9	2	1-5	747	
6991	RIKTADE	13.13	597	1-6	20	22	9	1-5	703	
6992	RIKTAR	12.02	668	1-6	18	18	6	2345	849	
6993	RIKTAS	9.055	697	1236	13	13	4	1245	548	
6994	RIKTIG	42.19	827	1-6	42	51	9	1-5	794	3.68
6995	RIKTIGA	42.89	858	1-6	43	50	7	1-5	756	4.91
6996	RIKTIGT	214.7	941	1-6	142	228	13	1-5	775	3.19
6997	RIKTLINJER	5.981	598	235	10	10	4	1235	485	
6998	RIKTLINJERNA	5.878	840	23	7	7	1	1-5	184	
6999	RIKTNING	60.21	848	1-6	49	71	11	1-5	736	4.98
7000	RIKTNINGAR	14.09	783	1-6	16	18	4	1-5	772	
7001	RIKTNINGEN	15.89	722	23456	21	22	6	1-5	525	
7002	RIMLIIG	10.74	671	1-6	14	16	5	1245	763	
7003	RIMLIIGA	12.16	579	1-6	20	21	9	2345	676	
7004	RIMLIIGEN	9.822	655	12346	14	15	5	1234	679	
7005	RIMLIIGT	21.66	802	1-6	24	27	5	1-5	871	
7006	RING	13.07	871	1-6	10	15	2	1-5	723	

FIGURE 5. A Page from the Alphabetical Version of the Basic Vocabulary.

Another interesting version of the 'Basic Vocabulary' is obtained by sorting the material according to falling values of DT. One would expect to find one of the most common words at the top of the list. What is, however, the most evenly distributed word is *betydligt* 'considerably' (DT= 0.993, F = 188). On the other hand, the distribution of the most common words is even enough. The 30 words heading the frequency list are among the 500 words heading the dispersion list.

This year (1969) a third dictionary has been produced. It is called 'Nusvensk frekvensordbok 1' (Frequency Dictionary of Present-Day Swedish 1). The level of description is that of homograph components (level 4). The three parts of this dictionary, too, are printed on line-printer and stored on magnetic tape. However, the reverse order version is being prepared for publication (by Almqvist & Wiksell). The plan for the publication of a 'Nusvensk frekvensordbok 2' is that the level of description should be that of lemmas (level 6) and the version the alphabetical one. The basic vocabulary on each of the two levels mentioned will of course also be calculated. These lists (and a number of other lists) will in fact be included in the relevant parts of the dictionary. It may be added that the volumes will be produced by means of computerized printing.

In order to get at the homographs of the material we needed a dictionary of Swedish homographs. Since no such dictionary existed, we had to compile one. The result was a dictionary of 42,000 homographs. It was found that 31,000 of these units occurred in our material and that the number of occurrences was as large as 620,000. This means that more than every second occurrence in the corpus is an instance of a homograph. Internal homography contributes substantially, and so does the group of units at the top of the frequency scale. Notice that the average frequency of the homographs is 20. It thus follows that the average frequency of the heterographs is 5. These figures reflect the increasing specialization that accompanies decreasing frequency.

The separation of homographs has been based on a homograph concordance in a so-called KWIC lay-out; see Figure 6.

The result of the inspection of this concordance has been punched and fed into the computer, which has executed the separation. The manual work required has been reduced by one-third in the following way. There are 26 words of a frequency exceeding 4,000. Of these, 24 are homographs (the two exceptions are *och* 'and' and *inte* 'not'). By sampling these units

NSKA DRAGONER AV TYSKARNAS	PANSAR	OCH ARTILLERI BAKÅT PÅ DEN STÅ	479:105
LITEN TRÄPPFYTA, SNEDSTÄLLT	PANSAR	SOM MEDFÖR ATT TRÄFFAR 'GLIDE	1083:15
STRÄLNING. + STRIDSVAGNENS	PANSAR	SKYDDAR VÅL, BÅDE MOT STÖTVERK	1083:87
AS BESÄTTNINGEN AV GOLVETS	PANSAR	SAMT AV VAGNENS SNABBHET. + VI	1083:99
DET. + ÄVEN DET TJOCKASTE	PANSAR	GENOMSLÅS AV PVROBOTARNAS RIKT	1083:114
SOM PÅMÄNDE OM VADMAL, MED	FRANSAR	LIKT TRASOR HÄR OCH VAR. + KL	1321:14
ERHUVUDENAS ÖVERVÄLDIGANDE	KRANSAR	MED NAMNCHIFFER OCH KRONOR BL	341:23
PÅ LÅKTARNAS BARRIÄRER HAR	KRANSAR	HÄNGTS I EN FÄRGSTARK RAD. +	341:108
RGER SJÖBERG ('KRISER OCH	KRANSAR	')). + DEN SOM OFTAST NÄMNS ÅR	626:48
VANLIGA SIFFERINSLAGNINGEN	STANSAR	HÄLREMSOR, VARS INFORMATION I	191:136
HENNES KNÅ OCH MAC, FADERN	NOSAR	PÅ HENNE SOM EN HUND. + NEDTONI	1416:145
GRA TECKNINGAR AV RINNANDE	FORSAR	HAN EN GÅNG TYCKT HAN LYCKATS	1209:96
RTIER. + I DE STÄMNINGARNA	KORSAR	DELEGATIONERNA VARANDRA NU. +	221:40
A, MÖJLIGHETERNA, TANKARNA	KORSAR	VARANDRA STÄNDIGT. + VÄRDIGHET	327:171
E TROENDE SOM ANDAKTSFULLT	KORSAR	SIG NÄR DE ILAR FÖRBI TILL SIN	1097:87
< DEN STORA BOULEVARD SOM	KORSAR	ETT AV STADENS AKTIVASTE AFFÄR	1360:13
LEN DÄR JÄRNVÄGEN GÅR FRAM	KORSAR	NÄMLIGEN KÖRVÄGEN MOT GRUVOMRÅ	1753:177
ÄR RÄDD OM GOLVEN, OCH MAN	TASSAR	RUNT I LÄNADE TOFFLOR. +	1816:80
LL ORDENTLIGT. + SNÖTÄCKET	PRESSAR	NER ISEN OCH DET FINNS ÖVERAL	127:117
STÅLPRESSTYCKEN I VÄLDIGA	PRESSAR.	+ DE ÄR PÅKOSTADE I DETALJER	250:171
RTS. + DE STIGANDE LÖNERNA	PRESSAR	IHOP VINSTMARGINALERNA OCH FÖ	361:165
VITA BÖNDER SOM FINNS KVAR	PRESSAR	LONDON-REGERINGEN OM ATT BLI	597:124
ING OM TRE PROCENT ÄRLIGEN	PRESSAR	MARGINALERNA SAMTIDIGT SOM JO	597:138
SKINEN MAL SÖNDER GLAS OCH	PRESSAR	SAMMAN PLÅT OCH SLÄPPER SEDAN	1135:182
M TEKNIKEN MED SNABBGÅENDE	PRESSAR	OCH MYCKET ANNAT OCH FÖRÄNDR	1328:16
SKADAN, SE BARA. + OCH HAN	HISSAR	UPP BYXBENET OCH VISAR EN STOR	364:112
ORN, INNESLUTANDE TRAPPOR,	HISSAR,	LEDNINGSSCHAKT. + DESSA TORN	1038:59
ELSERNA I FORM AV TRAPPOR,	HISSAR	OCH LEDNINGSSCHAKT TILL SLUTNA	1038:172
7.28 NÄR STRÖMMEN BRÖTS. +	HISSAR	STANNANDE MELLAN VÄNINGARNA, TU	1629:27
ASTNAT I TUNNELBANETÅG OCH	HISSAR.	+ MYNDIGHETERNA GAV NÄMLIGEN	1629:78
ÅTT SOM MÅNGA ANDRA SCENER	MISSAR	POÅNGEN I ÖVERGÅNGEN TILL ANDR	177:98
DEN ÄR AVSEDD SOM SATIR SÅ	MISSAR	DEN SITT MÅL. +	368:11
REDAN I SITT ORDVAL. + OCH	MISSAR	DE INTE ETT TILLFÄLLE ATT LÄRA	1036:159
EXISTERAR LÄNGRE. + NUMERA	MISSAR	JAG ALDRIG EN ÅHÖRARDAG I SKOL	1364:98
RKAR OCH NILS-OLOF FRANZÉN	MISSAR	SÄLLAN TILLFÄLLEN TILL DEN SOR	1765:60
ODUKTIONEN LIGGER PÅ OLIKA	KROSSAR,	DÄR MAN ÄVEN HAR LICENSTILLV	724:77
OTADE, DET SOM STORDRIFTEN	KROSSAR>	+ ELLER ÄR DET ENDAST GRYMT	908:124
SLÅR UPP STORT, MEN TEXTEN	KROSSAR	DE FÖRHOPNINGAR SOM RUBRIKER	1772:5
L LAST- OCH LEVERANSBILAR,	BUSSAR,	SPECIALFORDON OCH MOTORCYKLAR	324:27
RETAGEN I PARIS (MÉTRO OCH	BUSSAR)	ERHÖLL 1964 STATSSUBVENTIONER	385:154
GSFABRIK FÖR LASTBILAR OCH	BUSSAR	I CASABLANCA ÅGS OCH LEDS AV E	401:96
RNEBU I BESLAGTAGNA NORSKA	BUSSAR	KOM I DEN FÖRVIRRING SOM RÄDDE	541:215
HETER ATT GE LASTBILAR OCH	BUSSAR	SAMMA SERVICE. + < I DAG ÄR DÄ	977:89
N EN AV STATENS JÄRNVÄGARS	BUSSAR	OCH EN FRÅN STOCKHOLMS SPÄRVÄG	1237:89
T SJ HAR 1 800	BUSSAR	SOM KÖR LANDSVÄG. + DE HAR INT	1237:92
SPÄRVÄGAR HAR LÄNAT UT TVÅ	BUSSAR	ÅT 'JAG!', DEN ENA EN GAMMAL	1340:116
TJÄNSTGÖRING PÅ MALMÖ STAD	BUSSAR)	FÖR ATT SÖKA SIG ANDRA YRKEN	1343:12
STEN INTE MAMOR PÅ FYLDA	BUSSAR	OCH VARUHUS, SOM SJÄLVA STÅNKA	1743:157
SK LANDSORTSSTAD, DÄR INGA	BUSSAR	GÅR EFTER NIO PÅ KVÄLLEN < DÄ	1784:117
E INFORMATION FRÅN STABER,	MÄSSAR	OCH TILL OCH MED FRÅN FINLAND.	614:109
E DAGEN SEDAN 1868 OCH	HETSAR:	JÄ, HVIS DE NO VILL BEVISTE E	331:35
ISM OCH EXPRESSIONISM, HAN	HETSAR	SIG SJÄLV TILL SNABBARE TEMPO	449:32
ISKAS LÖJLIGA YTTRE. + HAN	SPETSAR	PÅ ETT STÄLLE SIN FRAMSTÄLLNI	1002:167
MATTOR, PLÅDAR, KNYPLÅDE	SPETSAR	M M VISAS JUST NU I EN SEPARA	1223:2
STE FÖR RINGHANKAR, LÖCKET	SPETSAR	UPPTILL OCH AVSLUTAR MED EN K	1737:63
T MED FÖRBRYTTARBYRÅKRATINS	SPETSAR.	+ (DEN BÄSTA BILDEN AV DENNA	1767:382

FIGURE 6. A Page from the Homograph Concordance.

and calculating frequencies of components and confidence intervals, the running through of 280,000 concordance lines has been replaced by the inspection of less than 50,000.

### Quantitative Data

A small selection of further quantitative data on the various levels follows.

*Graphic words.* The 30 most common words are shown in Table 3. They account for one-third of the corpus. The proportion of a text made up by the most common words has been a matter of wide interest. It has been argued, for instance, that 50 % of a text is covered by the 100 most common words. The percentage of the present investigation is 44. A calculation based on five counts of 100,000 running words each, reported by Carita Hassler-Göransson (1966), gives the figures 56, 48, 50, 37, and 49 %, respectively. Fant reports (1967:27–28) that in a material of 63,000 running words from telephone conversations, 30 words account for 49 %. What factors influence the figures?

One important factor is clearly the medium of expression: speech or writing. Another is the composition of a corpus (cf. the Hassler-Göransson figures). A third significant factor is evidently the length of the text. This appears from the following experiment. A series of accumulations was processed. The text length was systematically increased from 1,000 to 100,000 running words. The principal results are presented in Figure 7. One finds that the proportion of a text covered by a particular number of the most common words decreases as the text length increases – up to a certain point. The critical value for the present material is about 20,000 running words. To state this finding in another way, the number of different words required to account for a given proportion of a text increases as the text length increases, up to a certain point. Then a state of satiation is reached. This interesting phenomenon should be interpreted in terms of the consolidation of the structure word group, after the theme of the text has been

widened to such a degree that the thematic words can no longer assert themselves in the top layer.

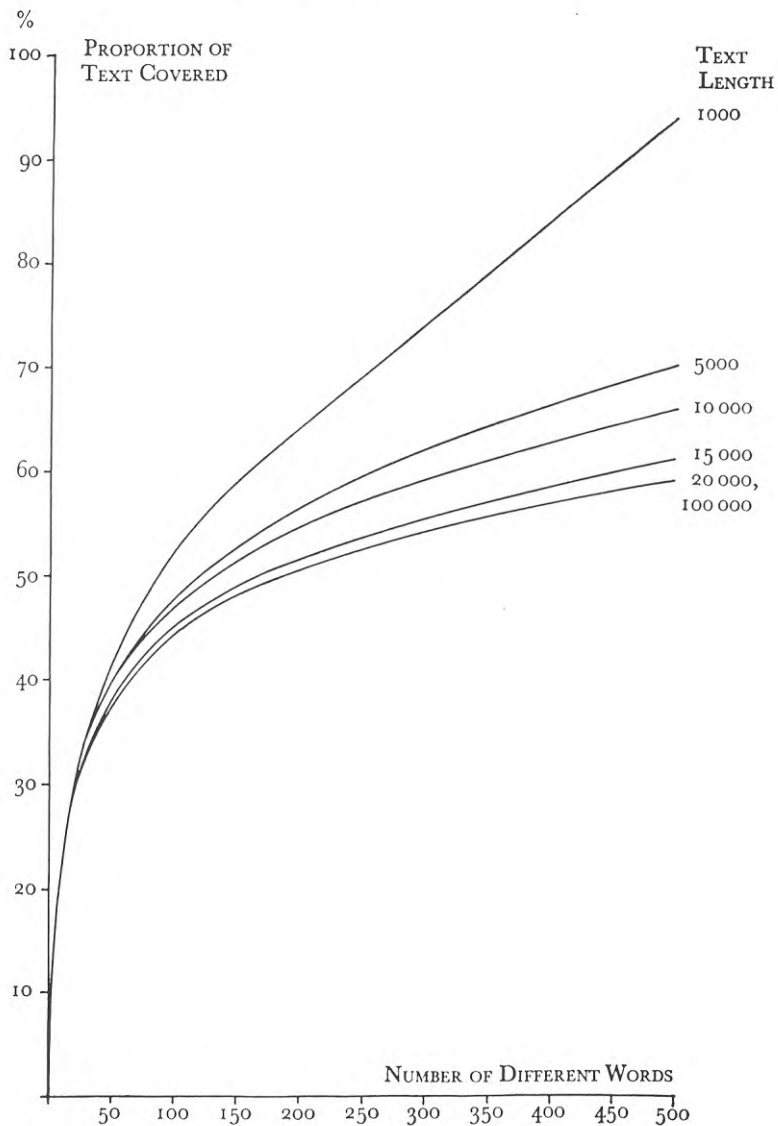


FIGURE 7. *The Proportion of Texts of Different Length Covered by the Most Common Words.*



TABLE 3. *The 30 Most Common Graphic Words.*

Word	Frequency	Word	Frequency
och	30,856	inte	8,901
i	29,551	ett	8,327
att	23,766	om	7,858
en	20,391	man	7,415
som	19,610	han	6,446
det	18,202	men	5,689
är	15,506	sig	5,559
av	15,466	kan	4,479
den	14,668	så	4,233
på	14,161	var	4,141
för	12,399	från	4,042
med	12,051	eller	2,910
de	10,953	sin	2,872
till	10,353	också	2,856
har	9,176	vi	2,739

Formalizing the finding, let  $N$  be the total number of running words (i.e., the text length),  $l$  the critical text length for the pertinent type of material, and  $F_s$  the frequency of the  $s$  most common words. Then,

for  $N \geq l$

$$\frac{F_s}{N} = k$$

This law, if I may call it so, can also be formalized in the following way:

for  $N \geq l$

$$F_s = k \cdot N$$

On the given condition,  $F_s$  is thus directly proportional to  $N$ . The rule has been tested for  $10 \leq s \leq 500$ . The accompanying relative values for  $k$  were found to rise from 21 % to 59 %. Approximately, we get:

for  $N \geq 20,000$

$$F_{10} \approx \frac{1}{5} \cdot N$$

$$F_{30} \approx \frac{1}{3} \cdot N$$

$$F_{200} \approx \frac{1}{2} \cdot N$$

$$F_{500} \approx \frac{3}{8} \cdot N$$

In our case, 50 % of the text is thus covered by the 200 most common words.

*Graphemes.* An extensive investigation of the graphotactics of the basic vocabulary has been carried out. The most general features in the structure of Swedish graphic words (containing one or more medial positions) can be summed up in the following way:

$$\left[ \begin{array}{c} \emptyset \\ sS \end{array} \right] \left\{ A(A) \right\} \cdots \left\{ r(S) \right\} \cdots \left\{ A(A) \right\} \left[ \begin{array}{c} \emptyset \\ n(S) \\ r(S) \end{array} \right]$$

where  $S$  indicates syngrapheme(s) and  $A$  autographeme(s) (for these two concepts, see Allén 1965:48–49, 91–93). Thus, first of all, we can expect zero or an  $s$ -sequence in initial position,  $r$  or an  $r$ -sequence in medial positions, and zero,  $n$ , an  $n$ -sequence,  $r$ , or an  $r$ -sequence in final position. Typical words are *andra* 'other', *svaren* 'the answers', *skildring* 'description', etc.

The average lexical word length in the basic vocabulary is 7 graphemes. The corresponding textual value is 4. These figures illustrate the close connexion between shortness and high frequency in a text. The maximum lexical word length in the whole corpus, as far as alphabetic words are concerned, is 33 graphemes.

The word in question is *strukturrationaliseringsprocessen* 'the process of structural rationalization'. There are considerably longer hybrid words.

*Homograph components.* Table 4 shows the commonest components of the most common homographs. PP is short for preposition, AL for article, VB for verb, PN for pronoun, IE for the

sign of the infinitive, KN for conjunction, AB for adverb, PRT for preterite, and NL for numeral. Asterisks indicate estimated frequency (based on sampling). The list of abbreviations just given indicates the dominance of structure words. An interesting fact not shown in the table is that the noun *man* 'man', the pronoun *var* 'each', the verb *för* 'convey (pres. and imper.)', the verb *så* 'sow', and the noun *men* 'disadvantage' are among the low-frequency components of the homographs listed. Items like these would get lost, if the most frequent words were simply to be cut off from the separation of homographs.

TABLE 4. *Frequency of the Commonest Components of the Most Common Homographs.*

Homograph Component	Frequency	Homograph Component	Frequency
i PP	29,461*	han PN	6,446*
en AL	18,184*	som KN	6,079*
är VB	15,506*	de AL	5,695*
av PP	15,242*	men KN	5,677*
det PN	13,991*	sig PN	5,559*
på PP	13,932*	den PN	5,378*
som PN	13,522*	om PN	5,199*
att IE	12,534*	de PN	4,997*
för PP	11,807*	kan VB	4,479*
med PP	11,700*	det AL	4,203*
att KN	11,232*	från PP	4,042*
till PP	9,865*	så AB	3,979*
den AL	9,290*	var VB PRT	3,839*
har VB	9,176*	om KN	2,283*
ett AL	7,702*	en NL	2,101*
man PN	7,103*		

*Lemmas.* Count no. 4 of Hassler-Göransson concerns the vocabulary of newspapers in 1932/33. This makes possible, in certain respects, a comparison with the present investigation. Let us consider the subsystem of ecclesiastical words. First we notice that words of this kind are clearly overrepresented in the conservative press; see Table 5 and cf. Table 6. It is a positive interest that makes itself felt. Then it appears that the word group is clearly over-represented in the older material; see Table 7. The

difference is strongly significant (chi-square is 68.5). The decrease in frequency seems to me to indicate that so many other fields crave attention in the society of the sixties.

TABLE 5. *Distribution of Ecclesiastical Words among the Newspapers 1965.*

Newspaper	Political Opinion	Difference between Observed and Theoretical Frequency in % of Theoretical Frequency
Svenska Dagbladet	Conservative	+77
Sydsvenska Dagbladet	Conservative	
Snällposten	(later: Liberal)	-16
Dagens Nyheter	Liberal	-16
Göteborgs Handels- och Sjöfarts-Tidning	Liberal	-43
Stockholms-Tidningen	Social-democratic	-59

TABLE 6. *Space Devoted to Ecclesiastical Material in the Newspapers of 1932/33.*

Newspaper	Political Opinion	Number of Column Centimeters per Page
Svenska Dagbladet	Conservative	30
Skånska Dagbladet	Conservative	
	(Agrarian party)	24
Ny Tid	Social-democratic	2
Norrskensflamman	Communist	0

*Parts of speech.* If the 26 most common graphic words (the sampling layer) are separated with regard to their homograph components, lemmatized, and sorted according to parts of speech, it will be found that prepositions dominate the list with a relative frequency of 30 %. An intermediate group is formed by pronouns, conjunctions, articles, and verbs (between 18.5 and 12 %). The bottom group is headed by adverbs and the sign of the infinitive (4 % each). The modest percentage of adverbs

seems to me to be notable, particularly since the lion's share is contributed by only two units, *inte* 'not' and *så* 'so'.

TABLE 7. *Frequency of Ecclesiastical Words.*

Lemma	1932/33	1965 (10 %)
bibel	6	3
biskop	25	7
Gud/gud	49	10
gudstjänst	7	1
Jesus	14	3
komminister	7	0
kristen	6	11
Kristus	13	2
kyrklig	18	4
kyrkoherde	12	1
pastor	10	5
predikan	4	0
präst	11	9
psalm	4	1
Total	186	57

*Morphemes.* An investigation of the suffixes forming adjectives and some pronouns in the basic vocabulary has revealed, among other things, what follows. A total of 26 suffixes were isolated. The forms of the words derived occurred 27,400 times. The five suffixes *-lig* (in *vanlig* 'ordinary', etc.), *-isk* (in *nordisk* 'Nordic', etc.), *-sk* (in *utländsk* 'foreign', etc.), *-ig* (in *nyttig* 'useful', etc.), and *-en* (in *vaken* 'awake', etc.) accounted for 75 % of the occurrences. At the top of the remaining 25 % were found three Romance suffixes, *-ell* (in *kulturell* 'cultural', etc.), *-al* (in *central* 'central', etc.), and *-iv* (in *produktiv* 'productive', etc.). These three suffixes raise the proportion to 88 %.

*Phrases.* The total frequency of the noun *fall* is 642. This means that it is one of the most common nouns. However, consider the following data: the phrase *i varje fall* 'in any case' accounts for 141 of these occurrences, *i alla fall* 'in any case' for 86, and *i så fall* 'in that case' for 79. This means that every second occurrence relates to one of these three phrases. See Table 8 for further information. *Fall* is by no means a unique

case. *Hand*, for instance, very seldom means 'hand'. As a rule it is part of phrases such as *i första hand* 'in the first place', *ta hand om* 'take care of', *efter hand* 'little by little', *på egen hand* 'on one's own account', etc. Theoretical linguistics should perhaps pay more attention to the important part played by ready-made phrases in the formation of sentences.

TABLE 8. *Phraseological Frequencies for the Noun fall.*

Phrase	Frequency	
	Abs	Rel
i varje fall	141	22
i alla fall	86	13
i så fall	79	12
i detta fall	26	4
i vissa fall	24	4
i de (allra) flesta fall	18	3
i (förvånansvärt) många fall	18	3
i bästa fall	14	2
i vanliga fall	9	1
i annat fall	7	1
i en del fall	7	1
i värsta fall	6	1
i (ett) sådant fall	5	1
i så(dana) fall	5	1
i flera fall	4	1
i några fall	4	1
Total	453	71

To a certain extent, phrases can be grouped according to structural and/or semantic features. The word *betydligt* 'considerably', which I mentioned before, may illustrate this point. If the frequency of single phrases is observed, the result is: *betydligt mer(a)* 'considerably more (cf. *much*)' 27 occurrences, *b. större* 'c. bigger' 26, *b. bättre* 'c. better', *b. fler(a)* 'c. more (cf. *many*)', *b. mindre* 'c. less', *b. svårare* 'c. more difficult' 9 each, *öka(s) b.* 'increase c.' 7, *b. högre* 'c. higher' 5, and so on. It is easy to see that forms in the comparative play an important part in these combinations. No less than 163 cases are in fact of this kind. There remain 25 cases. In 23 of these *betydligt* modifies a

gradual verb like *öka* 'increase' or *sjunka* 'sink', which fits in well with the comparative pattern. Only 2 cases are now left. Here *betydligt* modifies a word in the positive degree: *betydligt jordisk* 'considerably earthly', *betydligt ambivalent* 'considerably ambivalent'. These expressions strike me as being somewhat peculiar. They are found in informal articles: *Till detta kom en betydligt jordisk yngling med lockigt hår och avsevärda gåvor i orgelspelning, som gick runt ibland oss och såg oändlig ut och samlade upp tior; Annars kunde nog samt hans inställning till den platta tyska ölsaligheten visa sig betydligt ambivalent, pendlande mellan komiskt road trivsel och den starkaste vämjelse.* However this may be, *betydligt* is strongly linked to adjectives and adverbs in the comparative degree and to the cognate gradual type of verbs.

## Conclusion

I mentioned earlier two types of vocabulary investigations: extensive studies of exceptionally large bodies of material, and intensive studies of large, but by comparison restricted, bodies of material. It would be desirable to have such investigations undertaken at regular intervals. However, a continuous registration of the current vocabulary is also to be desired. All these kinds of work presuppose data processing. They could well be combined with the setting up of text archives on magnetic tape. Already large numbers of texts are being punched in the printing-works. Methods are furthered all the time by new technical aids. At present, terminals, display units, and optical readers attract particular interest.

I think it is fair to say that lexicography (as well as a number of applied disciplines) will profit handsomely from vocabulary data processing. But its potential is greater than this, owing to the structural prerequisite for the definition of the descriptive levels and for the integration of the processing routines.

## Bibliography

- Allén, S. Grafematisk analys som grundval för textedering. (English Summary: Graphemic Analysis as a Basis for Text Editing.) Göteborg 1965.
- Allén, S. Datamaskinell undersökning av tidningsprosa. Språkvård 1966.3.3–8. 1966.
- Allén, S. Om verksamheten vid Forskningsgruppen för modern svenska i Göteborg. Förhandlingar vid sammankomst för att dryfta frågor rörande svenskans beskrivning 4.67–76. Uppsala 1968. [Stenciled.]
- Allén, S. Språklig databehandling. Språkvård 1969.3.3–7. 1969.
- Borko, H. Automated Language Processing. New York, London, Sydney 1967.
- Ellegård, A. Estimating Vocabulary Size. Word 16.219–244. 1960.
- Fant G. Kompendium i talöverföring. 1. Stockholm 1967. [Stenciled.]
- Hassler-Göransson, C. Ordfrekvenser i nusvenskt skriftspråk. Lund 1966.
- Hays, D.G. Introduction to Computational Linguistics. New York 1967.
- Juilland, A., and E. Chang-Rodriguez. Frequency Dictionary of Spanish Words. The Hague 1964.

## Discussion

*Einar Haugen:* For one who did 'vocabulary processing' thirty years ago it is hard not to feel a deep envy of those who today have access to the electronic computer. My 'computer' was the handwork of some 35 persons whom the U. S. government paid to work for me during the great depression. With their help we were able to put on slips some 3,500,000 running words with their page and line numbers, including the entire belles lettres of the Norwegian authors Henrik Wergeland, Bjørnstjerne Bjørnson, Ivar Aasen, and Sigrid Undset, plus the Icelandic



Njála, Egla, and Heimskringla. Part of this material was published in stenciled form (100 copies) as 'Norwegian Word Studies' (2 volumes, University of Wisconsin Press, 1942). Most of it is still unpublished. We secured counts for the most frequent words (what Allén calls 'graphic words') which correspond closely with his results. Our work was inspired by George Zipf's Dynamic Linguistics.

Apropos Allén's discussion of the percentage of running text constituted by the 100 most frequent words, which he found was 44 % and remained fairly stable beyond 20,000 running words: I found that Sigrid Undset's novels averaged 58 % (1,231,364 words), with almost exactly the same percentage in her modern as in her medieval novels; Aasen's writings showed 55.5 % (124,507 words), while the Icelandic sagas showed 53.2 % (400,608 words). These figures clearly reflect the lower incidence of low-frequency (rare, technical, etc.) words in novels than in newspaper materials.

A point that should be made in this connection is that the percentage figures derived in this way for a running text do not actually reflect the proportion of the text which the most frequent words occupy. As is well-known, the frequent words are also shorter and therefore occupy proportionally less space than the less frequent words. This is reflected in Allén's grapheme count, and would be even more striking if he could count the phonemes in actual speech, where the form words are still further reduced. It would be interesting to see these figures combined, an operation that should be easy with a computer.

*Sture Allén:* (1) It is true, of course, that the arrival of the computer has brought about a radical change in the working conditions of the linguist. What is more, it has made possible a variety of investigations that were unattainable before. Let us keep in mind, however, that computational linguistics requires manual work of different kinds, too, and, in my opinion, to a larger extent than most non-computationalists would think.

(2) As I pointed out in my report the percentage constituted by the 100 most common words is largely determined by the composition of the text. Both of us would expect, I am sure, that

the 569 writers of my corpus should make a wider use, totally, of low-frequency words than a single writer like Sigrid Undset.

(3) The proportion of a text occupied by the most common words can in fact be calculated in various ways. One unit would be some linear measure, say centimetres, another the graphic word, a third the grapheme. As long as the unit is specified – in my case, the graphic word – the calculation is in my opinion wholly justified. And from a linguistic point of view the graphic word seems, in this connection, to be no less interesting than the grapheme. If the latter unit is chosen, the proportion will be considerably lower.

Leaving the text proper, the linguist can calculate the pertinent proportion in the spoken counterpart of the corpus (a line of action that would be required to make possible the comparison of figures). One way of doing this would be in terms of the phoneme. It is clear that this would be an interesting operation as would also the correlation with the grapheme figures.

The conversion from writing to speech would have to be undertaken by the computer. This is by no means an easy procedure, owing to the complicated and only partially known graphophonemic rules.

The following members of the Conference also contributed to the discussion: William G. Moulton, Winfred P. Lehmann, Werner Winter, Pardee Lowe.

## Ordklasser

Studier över nusvenskans vokabulärsystem. Språkdata 1969.

Ordklassbestämning är en grammatisk uppgift. För olika ändamål kan olika typer av grammatik bli aktuella. En ordklassbestämmande grammatik som skall kunna läggas till grund för en kvantitativ undersökning av ett omfattande autentiskt material bör enligt min mening motsvara vissa speciella krav. Ett är att varje belägg i materialet med så stor reliabilitet som möjligt skall kunna hänföras till en enda ordklass. Ett annat är att ett litet antal kriterier skall åberopas. Proceduren skall ju vara praktiskt användbar och inte förutsätta utförliga analyser på punkt efter punkt. Ett tredje är att proceduren skall vara formaliserad. Detta är betydelsefullt av flera skäl. Det ger den bästa säkerheten för att konsekvens skall kunna uppnås. Det är enligt min erfarenhet också ett lämpligt sätt att explicit ange hur proceduren fungerar. Vikten av att göra detta har jag tidigare kraftigt understrukt.

Härutöver menar jag att klassifikationen främst skall ta fasta på enheternas uppträdande i ytstrukturen men att en djupare liggande struktur skall kunna åberopas t.ex. vid satsförkortningar. Slutligen bör kriterier liksom vid lemmabestämningen kunna hämtas från såväl tal som skrift. Proceduren tar i detta avseende alltså sikte på sådana mer avancerade språk som har båda uttrycksformerna.

Som ovan nämnts kan en grammatik av Natanael Beckmans snitt inte utan vidare läggas till grund. Detta hänger också samman med att den är både ofullständig och oformaliserad. Nackdelarna blir tydligt märkbara när den konfronteras med ett stort ordmaterial som exempelvis det på många andra sätt förtjänstfulla lexikonet *Illustrerad svensk ordbok*. Man behöver inte bläddra länge i detta för att finna inkonsekvenser och vagheter i ordklassangivelserna. Det torde därför knappast vara nödvändigt att elaborera synpunkten. För att ändå ge ett exempel kan jag peka på *2föga*, oböjligt. Det analyseras som (1) adjektiv: *av föga*

*värde* (2) adverb: *en föga givande bok* (3) substantiv: *härom är föga att säga*. Adjektiv? Substantiv?

En procedur av typen numerisk taxonomi (se Jan Svartvik, *On voice in the English verb* (1966)) bygger på att alltför många kriterier (principiellt hundratals) skall utnyttjas för att den skall kunna läggas till grund för behandlingen av en hel vokabulär. Vid en mera detaljerad analys av delar av vokabulären kan den ifrågakomma.

En generativ modell kan inte läggas till grund av flera skäl. Vi har ingen sådan grammatik för något språk. Om vi hade någon, skulle det möta svårigheter att bestämma en viss beskrivningsnivå där kvantifieringen skulle sätta in. Om en nivå kunde anges, skulle det krävas en synnerligen tidsödande analys av den löpan- de texten för att anpassa den till nivån.

Mitt ställningstagande innebär att jag utgår ifrån den traditionella grammatikens kategorier och preciserar deras innehåll genom att sätta in dem i en formaliserad procedur med ordnade kriterier, ett flödeschema. Den flödesgrammatik som jag sålunda har utarbetat är avsedd att fungera i enlighet med de krav som ovan uppställdes. Det valda förfarandet synes mig vidare ha den fördelen, att en person som önskar ta del av resultatet av den undersökta textmassans kvantifiering och som inte känner till procedurens utformning ändå kan få en approximativ bild av förhållandena. Principen att arbeta med ordnade kriterier ligger i linje med en uppfattning om den metodologiska inriktningen på området som David Crystal gör sig till talesman för i sitt bidrag till volymen *Word classes* (Amsterdam 1967, s. 55): "Throughout this paper, I have tried to underline two main points: that much of the terminology used in the discussion of word classes in English has been badly defined and used uncritically; and that the emphasis in word class analysis and definition should lie on the selection and ordering of criteria – which in turn means a great deal more detailed descriptive work than has yet been done."

Som det viktigaste ordklasskriteriet betraktar jag funktionen, det syntaktiska uppträdandet. Ett bekant exempel på en härpå fotad indelning är Charles C. Fries' *The structure of English*

(1952). En skiss till vidareutveckling ger Alvar Ellegård i uppsatsen *Design for a mechanical distribution analysis of English word classes* (i *Structures and quanta* (1963), s. 5–21). Det visar sig emellertid som man väntar, att indelningsgrunden inte ger helt tillfredsställande resultat. Ord som normalt uppträder i en viss ram (funktion) dyker upp också i andra ramar, och utvidgas textmaterialet i tillräcklig grad kommer detta att gälla hela klasser av ord (jfr Ellegårds uppsats, s. 19 f.), vilket utgör en stor belastning för klassifikationssystemet. Subsidiära kriterier får dras in. Liksom när det gäller lemmabestämningen menar jag att dessa kan hämtas från uttryckslärans och morfologins områden.

Flödesgrammatiken har bl.a. som förutsättning en allmän modell av (svenska) satsers byggnad. Den förutsatta grundstrukturen hos satsen återges i figur 1.

Strukturen beskrivs med hjälp av ett träd-diagram. Grund-schemat har tre delar – subjektsdel, predikatsdel och satsadverbial – eftersom satsadverbialen ju bör placeras direkt under satsnoden. Parentes om beteckningen för en konstituent anger att denna är fakultativ. Som naturligt är avser figuren sålunda normalfallet att satsen är fullständig, har både subjekt och predikat. Ur schemat kan andra fall härledas, t.ex. elliptiska satser som *Sven* (såsom svar på frågan *vem är det som kommer?*) och *hemma* (som svar på frågan *var har du varit?*) eller uppmanings-satser som *kom*. Fall som *hej, fy* är att betrakta som ostrukturerade satsekvivalenter. Diagrammet anger vidare att en sats med eller utan hjälp av en partikel kan kombineras med en annan sats (som i sin tur kan kombineras med en annan sats osv.) till en större helhet.

Enligt modellen utvecklas kopula jämte subjektivt predikativ ur predikatet, jfr satser som *träden grönskar* och *träden är gröna*. Affiniteten mellan subjektivt predikativ och predikativt attribut framgår också, jfr satserna *träden är gröna* och *träden står gröna*. På motsvarande sätt markeras affiniteten mellan predikatsdelens adverbial och agenten och mellan det indirekta objektet och det objektiva predikativet.

De konstituenterna som har ansatts i figuren kan vidareutvecklas på mångahanda sätt. Ur varje adverbial liksom ur det indirekta

objektet kan exempelvis härledas ett uttryck bestående av partikel och rektion i enlighet med vad som anges för satsadverbial, jfr t.ex. *han bor hemma* och *han bor i hemmet, jag gav Sven boken* och *jag gav boken åt Sven*. Ur subjektet kan härledas samordnande uttryck som *han och jag* osv. Ur det direkta objektet kan utvecklas infinitivsyntagmer som i *han älskar att läsa* och nya satser som i *han frågade om Sven kom* osv. Det är denna flexibilitet som gör beskrivningen användbar i praktiskt analytiskt arbete, inte minst genom att antalet funktioner hålls nere.

Satsen byggs upp av understrukturer som kan bestå av ett eller flera ord. En sådan internt strukturerad mikrosyntagm benämner jag fras och dess huvudord fraskärna. Som exempel kan ges *de intensivt gröna träden, Svens bok* och *Sven* (i satsen *Sven kommer*), där *träden, bok* respektive *Sven* är frasernas kärnor. Partiklarna i fraser som *i hemmet, att läsa* anses inte utgöra kärnor. En (mer eller mindre utbyggd) fras av den typ som står under noden subjektsdel och som enligt ovan också förekommer i andra funktioner benämns nominalfras.

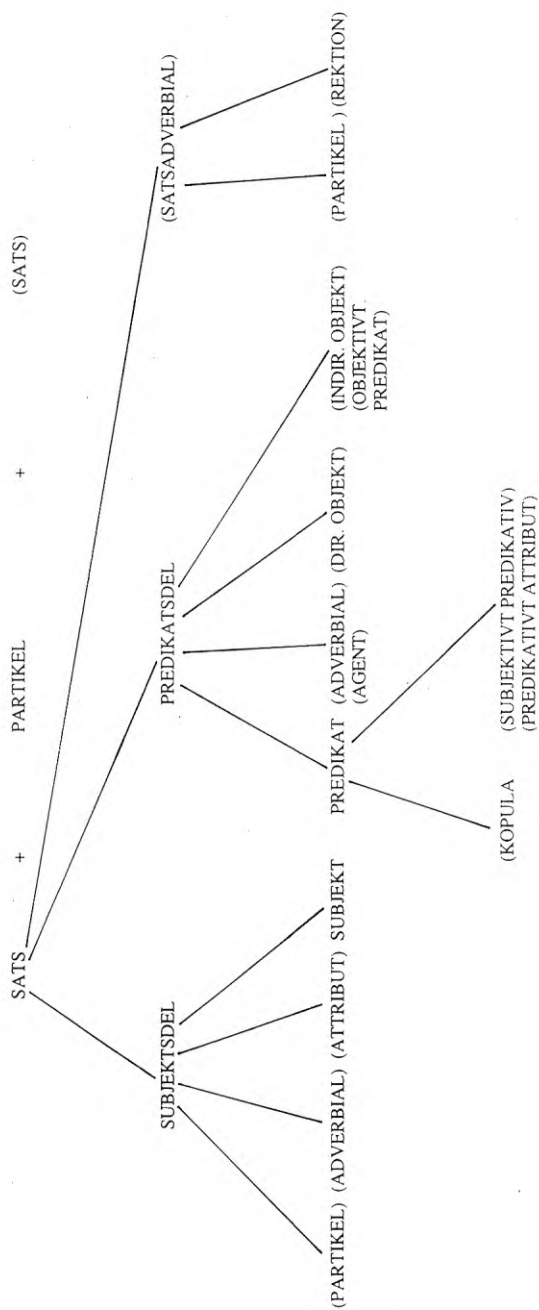
Satsledens funktioner utvisas alltså av trädidiagrammet. Vid analysen appliceras satserna på detta. Forskaren förutsätts därvid kunna supplera de belagda satserna med avseende på satsförkortningar och ellipser.

Vid tillämpningen av modellen finner man i vissa fall att ordmaterial som uppträder i olika funktioner är av samma karaktär. Företeelsen kan belysas med belägg i funktionerna subjekt och direkt objekt.

många barn	har det besvärligt
gamla	
avskedade	

han känner många barn
gamla
avskedade

Detta är ett tecken på att orden i de aktuella funktionerna kan slås samman till en gemensam klass.



FIGUR 1. Satsens grundstruktur.

Som ovan antytts rymmer å andra sidan de funktionsklasser som erhålles i stor utsträckning ett från andra synpunkter heterogent ordmaterial. Detta kan lätt exemplifieras med några ord i funktionsklasserna subjekt och attribut.

många barn	har det besvärligt
gamla	
avskedade	

det gäller magra	hundar
barns	
dresserade	

Normalt uppträder *barn* i den första av funktionerna och *gammal* och *mager* i den andra, medan *avskeda* och *dressera* förekommer i en tredje. Denna fördelning motsvarar skillnader i flexion.

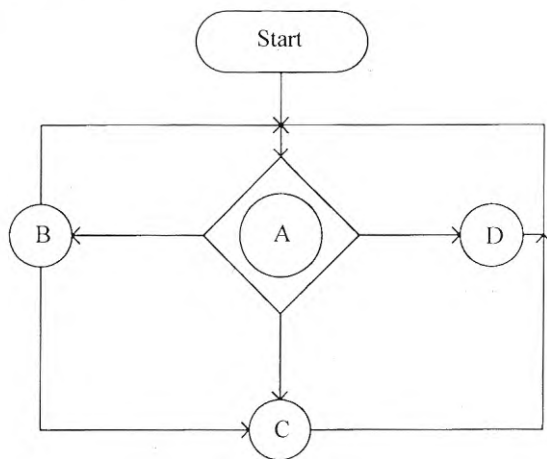
Den fortsatta klassificeringen av det ordmaterial som erhålles i funktionsklasserna sker med hjälp av det grammatiska flödes-schemat. Till de grundläggande syntaktiska kriterierna fogas här som jag tidigare nämnt nyckelkriterier från uttryckslära och morfologi.

Proceduren har fyra delar. Detta har visat sig vara ändamåls-enligt med hänsyn till flera faktorer. En är den interferens från främmande vokabulärsystem som yttrar sig i belägg på icke-svenska ord. Moment A är bl.a. till för att filtrera materialet i detta avseende. Två andra faktorer är olikheterna i karaktär mellan fraskärnor och övriga element och mellan å ena sidan funktionsklasserna subjektivt predikativ och rektion och å andra sidan övriga funktionsklasser. I moment B behandlas fraskärnor i andra konstituenten än subjektivt predikativ och rektion; de enheter som inte faller ut i någon klass under B går vidare till moment C, vari också de konstituenten som inte är fraskärnor behandlas; i moment D slutligen behandlas fraskärnor i subjektivt predikativ och rektion.

För överskådlighetens skull visas i figur 2 sambandet mellan de fyra delarna. Från startpunkten kommer man alltid in i moment A. Där sker efter den ovannämnda filtreringen en växling på så sätt att man hänvisas till ett av momenten B, C eller D beroende på ordets karaktär. Från B kan man återvända till start



antingen direkt eller via C. Från C liksom från D kan man bara återvända direkt till start. Själva flödesgrammatiken återges i figurerna 3–6.



FIGUR 2. *Flödesgrammatikens system.*

Förkortningarna avser följande ordklasser (motsvarande):

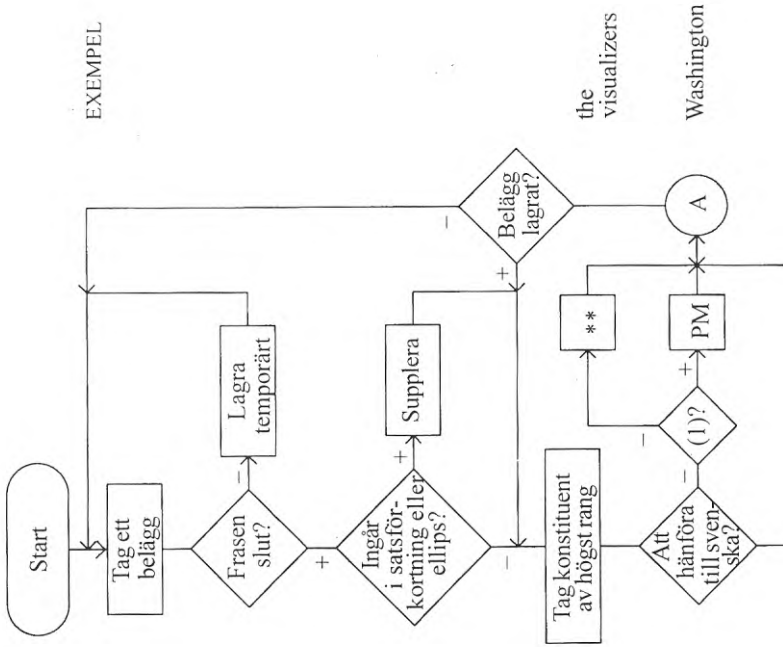
AB	adverb	NL	räkneord (numeral)
AL	artikel	NN	substantiv (nomen substantivum)
AN	förkortning (abbreviation)	PM	egennamn (proprium)
AV	adjektiv	PN	pronomen
IE	infinitivmärke	PP	preposition
IN	interjektion	VB	verb
KN	konjunktion		

Dessa förkortningar har valts för att i görligaste mån ge också utländska läsare av de resulterande lexikonerna hållpunkter för minnet. Beteckningen \*\* avser utländskt språkgods.

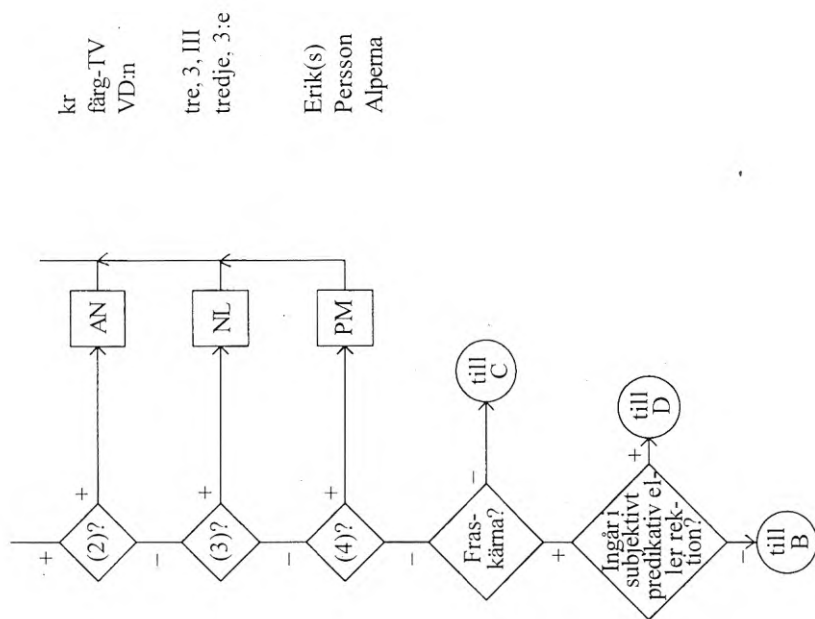
Antalet klasser är alltså tretton. Av dessa är emellertid abbreviationerna ingen ordklass i vedertagen mening utan innehåller förkortade versioner av ord ur andra klasser. Vid bearbetning av omfattande textmaterial är det emellertid lämpligt att särbehandla den stora klassen av abbreviationer. Detta är från språklig synpunkt också av intresse, bl.a. med hänsyn till typo-

UTTRYCKSLÄRA

EXEMPEL



(1) Identifierbart som  
som ort- eller personnamn



- (2) Abbreviaturgram i  
sista semantiska morfem  
(derivationsmorfem följer ej)
- (3) (Alternativt) numeriskt  
graford (ev. följt av flexions-  
morfem)
- (4) Initial versal (ej i variation  
med gemen)

FIGUR 3. Moment A i flödesgrammatiken.

logi och kvantitetsförhållanden. Av de tolv egentliga ordklasserna intar två en särställning genom att de tillsammans endast har en handfull medlemmar: artiklar och infinitivmärke. Redan nu skall dessutom påpekas att klassen pronomen i viss mån också intar en särställning genom att här definieras som en restklass.

För att följa proceduren kan vi välja en typmening som *Bara unga författare debuterade i fjol*. Satsen består av tre fraser: (1) *Bara unga författare* (2) *debuterade* (3) *i fjol*. Vi startar i moment A och tar enligt anvisningen det första belägget som är *Bara*. Frasen är inte slut. Vi håller därför belägget i minnet och tar nästa. Förfarandet upprepas tills frasen är slut, varefter vi går vidare. Det är inte fråga om någon satsförkortning eller ellips, varför vi utan suppling tar den konstituent i frasen som har högst rang, dvs. *författare*. Fråga 1 besvaras jakande, frågorna 2, och 3 nekande. Eftersom belägget är en fraskärna och inte ingår i subjektivt predikativ eller rektion, hänvisas vi till moment B. Här uppfylls inte villkoret under 5 men väl villkoren under 6, varför belägget faller ut som substantiv.

Från moment B återförs vi till punkten A i moment A. På frågan om det finns något belägg lagrat är svaret ja. Vi tar då konstituenten *unga* och hänvisas efter hand med den till moment C. Här uppfylls omedelbart villkoren under 7. Belägget faller ut som adjektiv.

Vi förs så åter till punkten A, tar belägget *Bara* och får hänvisning till C igen. Här besvaras fråga 8 positivt, och belägget faller ut som adverb. Vi hänvisas därefter till punkten A igen. Inget mer belägg är lagrat, och följaktligen kommer vi tillbaka till start.

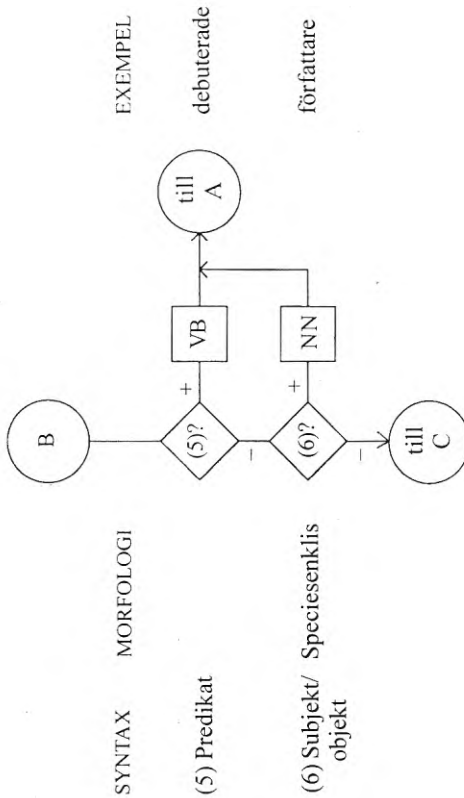
Nästa fras innehåller endast belägget *debuterade*. Moment A hänvisar oss till B. Belägget faller ut som verb vid fråga 5. Därefter åter via punkten A till start.

Den sista frasen, *i fjol*, tas in. Fraskärnan *fjol* ingår i rektion, varför vi får gå till D. Här uppfylls villkoret 23 – *fjol* förekommer enbart i rektion – och belägget utfaller alltså som substantiv.

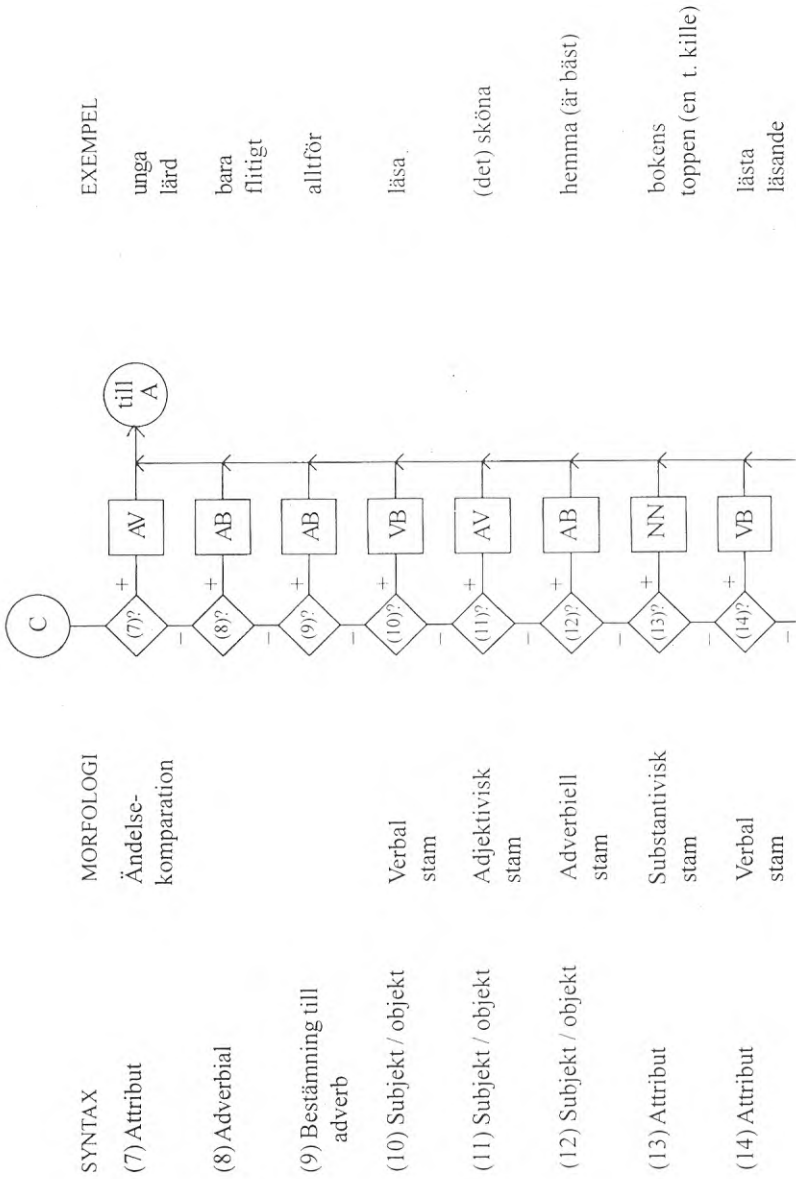
Via punkten A får vi slutligen belägget *i* med oss till moment C. Här uppfylls villkor 18, varigenom belägget klassificeras som

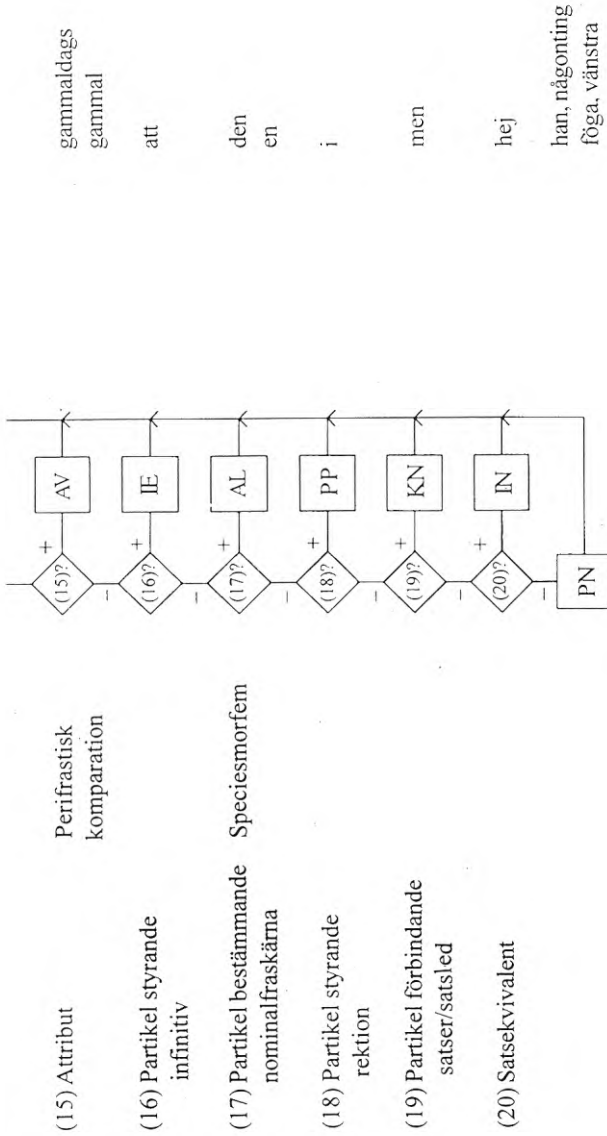
preposition. Via punkten A återförs vi till start och är beredda att behandla en ny sats.

Beträffande villkorens formulering skall två påpekanden göras. Markeringen subjekt/objekt är en av utrymmesskäl motiverad förkortning av subjekt/direkt objekt/indirekt objekt/objektivt predikativ/apposition. På motsvarande sätt markerar attribut också predikativt attribut.



FIGUR 4. Moment B i flödesgrammatiken.





FIGUR 5. Moment C i flödesgrammatiken.

Villkoren skall så belysas något närmare. Om den betraktade enheten är att hänföra till svenskan (se moment A) är i vissa fall en avvägningsfråga. Syftet med filtret är att skilja ut vad som är klart icke-svenskt: *the* osv. Normalt ger morfologin (t.ex. *alperna* med svensk böjning) och/eller uttrycksformen (t.ex. *Ktesibios*, *visualizers* med främmande skepnad) utslag. En särskild roll spelar proprierna, som ju i stor utsträckning är ett slags internationellt commune bonum. De får därför möjlighet att inlemmas i kategorin proprier, även om de har bedömts som utländskt gods (villkor 1). För det praktiska arbetets skull har villkoret begränsats till att gälla (utan svårighet) identifierbara ortnamn och personnamn. Dessa får ju anses vara de två viktigaste kategorierna.

Närmast följer tre villkor som hänför sig till uttrycksläran och som därför är oberoende av syntaktiska funktionsklasser. Om lämpligheten i att utsortera abbreviationerna har jag tidigare talat. I anslutning till villkorets formulering kan nämnas att ett belägg som *kr* 'kronor' grafematiskt analyseras <k÷r> där ÷ betecknar abbreviaturgrafemet (se Grafematisk analys, s. 122). Preciseringsringen i villkor 2 innebär att ordsluten blir avgörande: ord som *kr*, *färg-TV*, *VD:n* 'verkställande direktören', *A* 'Anders' osv. utfaller. Ord av typen *TV-tittare* har alltså däremot ansetts böra gå samma väg som *tittare*.

Nästa villkor sorterar ut räkneorden, de enheter som i skriften kan växla mellan alfabetiska och numeriska grafem. Här utfaller t.ex. *tre*, *3*, *III*, *tredje*, *3:e* osv. Placeras kriteriet på detta sätt, blir också *miljon* m.fl. enheter räkneord (i överensstämmelse med betydelseinnehållet; jag har i min uppsats *Indelningen av nussvenskans pronomen* (1958) betecknat tallinjen som ordklassens ryggrad).

Det sista villkoret på uttryckslärans område gäller initial versal som ej uppträder i fakultativ variation med gemen. Här utfaller proprierna: *Erik*, *Eriks*, *Persson*, *Perssöner* osv. Villkorets formulering innebär att exempelvis *Slottet* och *Kungen* som varierar med *slottet* respektive *kungen* inte faller ut. Uttryck som *Kulla Gunnarstorp* innehåller två proprier, uttryck som *af Geijerstam* ett proprium, *Geijerstam*. Jämför härmed att denna enhet i mycket stor utsträckning ensam får representera kombinationen



vid omnämmande av namnbäraren. På fraseologisk nivå bildar givetvis *af Geijerstam* en enda enhet, ett proprium. Vidare skall här nämnas att vad jag betecknar som junkturell majuskularitet (Grafematisk analys, s. 119), dvs. obligatorisk versal i början av mening, negligeras i proceduren.

Att fraskärnorna behandlas med prioritet beror på att deras bestämningar måste kategoriseras i relation till dem. Att subjektivt predikativ särbehandlas beror på att funktionsklassen har en speciellt vid variationsbredd med avseende på beläggens karaktär och att dessa belägg bör kategoriseras med hänsyn till vederbörande enheters övriga uppträdande.

författarna är flitiga  
uppmärksammade  
sådana  
autodidakter  
hemma

Rektion särbehandlas av samma skäl.

Om villkoren i moment B kan jag fatta mig kort. De syntaktiska kriterierna tar sin utgångspunkt i predikatsfunktionen. Här utfaller sålunda verbets finita former. Funktionskriterierna subjekt/objekt osv. kombineras med det morfologiska nyckelkriteriet slutartikel. Kan den betraktade enheten ta detta karaktäristiska speciesmorfem, som kan betecknas som funktionsklassens speciella flexionsart, föreligger substantiv. Självfallet får man observera att speciesmorfemet i vissa fall kan vara representerat av nollallomorfen: *början, verkan, universum* o.d.

I moment C urskiljs först funktionen som attribut (inklusive predikativt attribut). Som morfologiskt nyckelkriterium framstår i denna funktionsklass ändelsekomparation. Utfallande ord betecknas som adjektiv. Frågan om huruvida ändelsekomparation är möjlig avgörs i föreliggande fall principiellt genom konsultation av Illustrerad svensk ordbok. Om materialet såsom i fråga om *upphöjd* visar att komparation förekommer där ordboken inte har någon sådan uppgift, ges materialet vitsord. I de fall där belagda ord saknas i ordboken får den analyserandes kännedom om språket fälla utslaget. Ett särskilt omnämmande förtjänar i

detta sammanhang ord som *lärdd* och *driven* (liksom det ovan nämnda *upphöjd*). Belägg av typen (*en*) *lärdd* (*person*) faller ut som adjektiv (jfr *en lärddare person*), medan belägg av typen (*han har blivit*) *lärdd* (*att göra så*) går vidare. Båda slagen är belagda i ett uttryck som *vår som så frigjord ansedda tid*.

Villkoret 8 samlar i klassen adverb de enskilda ord som enligt schemat över satsens grundstruktur fungerar som adverbial. Här utfaller ord som *bara* och *inte* men också t.ex. *t*-avledningar av adjektiv (*flitigt* osv.) Det följande villkoret för också bestämningar till de så klassificerade adverbena till ordklassen i fråga.

Kriterierna 10–12 appliceras på sådana enheter i funktionerna subjekt/objekt osv. som inte faller ut under 6 i moment B. Här utnyttjas det förhållandet att den fortskridande proceduren genom villkoren 5, 7 och 8–9 definierar ett stort antal enheter som verb, adjektiv respektive adverb och därmed deras stammar som verbala, adjektiviska respektive adverbiala. Under 10 utfaller i enlighet härmed exempelvis *läsa* i *han älskar att läsa* som verb, under 11 *sköna* i *det sköna är evigt* som adjektiv och under 12 *hemma* i *hemma är bäst* som adverb.

På motsvarande sätt behandlas i punkterna 13–15 återstående enheter i den attributiva funktionsklassen. Punkt 15 samlar de icke utfallna enheter i funktionsklassen som har perifrastisk komparation. I vissa fall kan man råka i tvivelsmål om huruvida perifrastisk komparation är möjlig (uppgifter i lexikon kan härvidlag inte läggas till grund). Förvånansvärt ofta kan man emellertid peka på kontexter som tillåter komparation även av enheter som anses motsträviga. Den som är både hjärtdöd och hjärndöd kan exempelvis sägas vara *en mera död person* än den som enbart är hjärndöd. Det finns dock ingen anledning att pressa för hårt på den här punkten. Faller sig en perifrastisk komparation inte naturlig, får orden gå vidare.

Under punkt 13 utfaller exempelvis *bokens* i uttrycket *bokens dag* och *toppen* i *en toppen kille* som substantiv, under 14 *lästa* i *den lästa boken* och *läsande* i *den läsande allmänheten* som verb och under 15 i *en gammaldags roman* som adjektiv. Också *gammal* utfaller under 15 (och inte under 7) beroende på att s. k.

suppletiv böjning inte accepteras i den här förevarande analysen. Se närmare härom i III:4.

Punkterna 16–19 fördelar de i grundstrukturen urskilda partiklarna på infinitivmärke, artiklar, prepositioner och konjunktioner. När det gäller punkt 17 kan man ibland vara osäker om huruvida partikeln har ren speciesfunktion. Detta sammanhänger delvis med att skriften inte noterar prosodi. Den besvärligaste av punkterna är emellertid nummer 18. Svårigheterna knyter sig till analysen av förbindelser där både verb och nominalfraser är inbegripna, typen *gå till någon/något*. Här föreligger ett problemkomplex som skulle behöva en egen monografi. Med hänsyn till det stora antal belägg som materialet tillhandahåller har efter försök med olika modeller ett grovt och enkelt men för kvantitativa behov av allt att döma tillräckligt effektivt analyschema fastställts. Om den betraktade enheten vid uppläsning måste vara betonad, bedöms den som adverbial, dvs. villkoret 8 är uppfyllt. Ett exempel är *det går till tio kronor*. Om enheten kan vara obetonad, bedöms den som rektionsstyrande partikel och faller ut som preposition. Ett exempel är *han går till båten*. Det kan tilläggas att enheten när den står i absolut ställning analyseras som adverbial. Exempel: *sillen går till*.

Vissa problem knyter sig också till punkt 19. De rör förhållandet mellan konjunktioner och adverb. Som exempel väljes analys av några typiska fall av *så*. I typen *Så röstade ledamöterna om förslaget* med omvänd ordföljd anses adverb föreligga, i *Så ledamöterna röstade om förslaget* med rak ordföljd anses konjunktion föreligga. I typen *han ordnade så ledamöterna (inte) röstade om förslaget* med bisatsordföljd anses konjunktion föreligga och i *han ordnade så att ledamöterna röstade* räknas *så* som adverb (hänförligt till *ordnade*) och *att* som konjunktion.

Under punkt 20 faller interjektionerna ut. Här har verklig satsekivalens krävts. Fall som *eld!* betraktas som elliptiska och suppleras sålunda i början av moment A. Analysen av *eld* sker följaktligen på basis av (*giv*) *eld!*

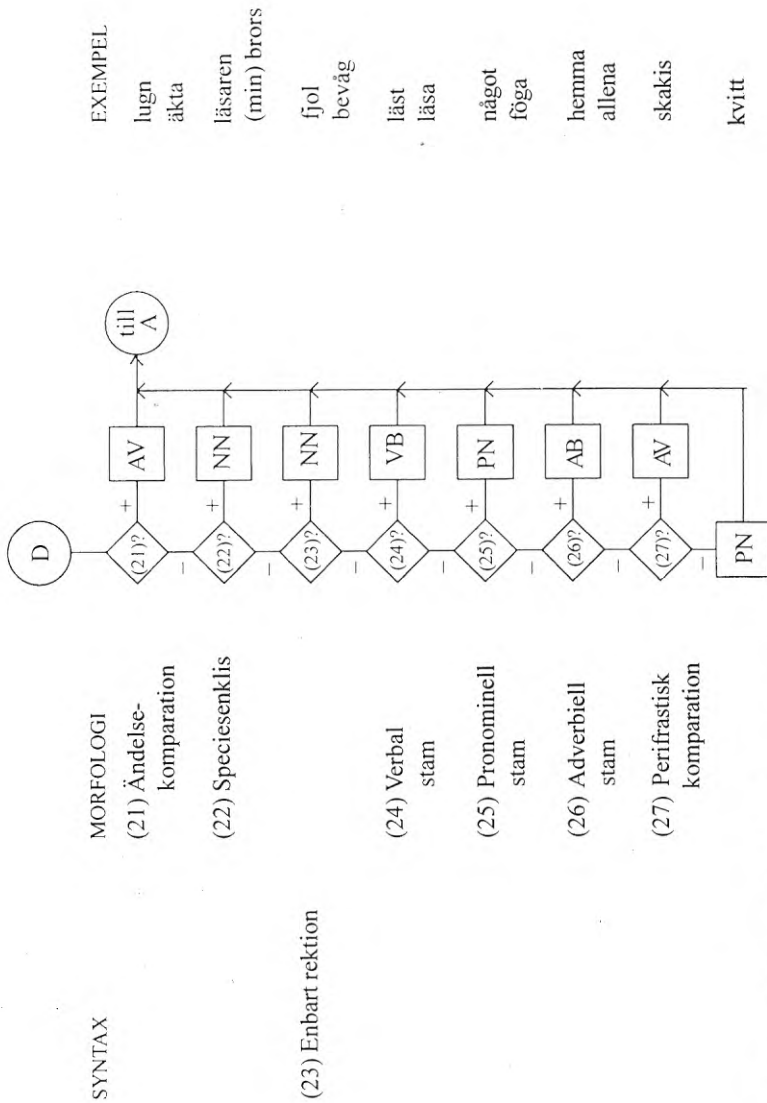
Som sista kategori i moment C (och D) står en restklass, vars medlemmar kännetecknas av att sakna alla de kriterier som appliceras. På de nämnda punkterna utfaller ord som *han, detta,*

*sin, alla, ingen, någonting, vem* m.fl. men också t.ex. *vänstra* (i *den vänstra kolumnen*), *föga* (i *av föga värde*) och *kvitt* (i *vara kvitt*). Å andra sidan utfaller inga ord som kan kompareras. Den traditionellt vagt bestämda och heterogena ordklassen pronomen har alltså dels fått släppa ifrån sig medlemmar, dels tillförts nya. Som namn på klassen kunde *i* och *för* sig något annat än pronomen väljas. Eftersom kärntruppen emellertid utgörs av mycket frekventa ord med den traditionella beteckningen pronomen, har jag bibehållit namnet. Det förtjänar att påpekas att klassens rent negativa karakterisering inte utgör något hinder från klassifikatorisk synpunkt. Jämför exempelvis att fonemet /i/ respektive grafemet <u> enbart kännetecknas av frånvaro av de särdrag som urskiljs i de analyser som återges på s. 78 f. i min uppsats *Förhållandet mellan skrift och tal* (1967).

Moment D slutligen behandlar fraskärnor i subjektivt predikativ och rektion. Villkoren 21 och 22 sorterar ut adjektiv och substantiv enligt tidigare applicerade kriterier. Villkoret 23 kategoriserar vidare sådana enheter som enbart uppträder i rektion som substantiv: (*i*) *ffjol*, (*i*) *går*, (*på eget*) *bevåg*, (*på någons*) *vagnar*, (*till*) *reds*, (*till*) *salu*. Sedan de verbala stammarna utfallit under 24, kommer vi till punkt 25, där de i moment C erhållna pronomina stammarna utnyttjas som kriterium. Under 26 utfaller *hemma* i *han är hemma* osv. som adverb. Slutligen bestämmer kriteriet perifrastisk komparation liksom i moment C ytterligare vissa enheter som adjektiv, t.ex. *skakis*. Restklassen har tidigare omnämnts.

Självfallet är det långt ifrån alla enskildheter i ordklassbestämning som har kunnat tas upp här. Det stora material som proceduren tillämpats på har resulterat i att en omfattande kasuistik har tillförts forskningsgruppens samlingar. Vid genomgången av modellen har jag hämtat exempel därur. Utöver de enskilda fallen har som nämnts emellertid också vissa större grupper av enheter krävt speciella ställningstaganden vid tillämpningen. Ytterligare ett par av dem skall beröras.

Den stora gruppen sammansättningar av typen (*en*) *semesterfirande* (*turist*) har behandlats på så sätt, att verbal stam har ansetts föreligga om Illustrerad svensk ordbok upptar en mot



FIGUR 6. Moment D i flödesgrammatiken.

svarande infinitiv (i exempelfallet *\*semesterfira*). Om så inte är fallet, har enheterna generellt kategoriserats som adjektiv, oaktat komparation kan vara långsökt (som är fallet med *mera semesterfirande*). På motsvarande sätt har typen *höftplacerad* behandlats.

En generell tillämpningsföreskrift har också fått gälla de stora grupperna av sammansättningar på *-isk* och *-ig*, vilka generellt räknats som adjektiv. På samma sätt har den stora gruppen sammansättningar med komparerbar slutled, typen *knäkort*, behandlats.

Ovan nämndes att tidningsartiklar av metaspråklig karaktär uteslöts ur korpusen. Fenomen av det nämnda slaget uppträder naturligtvis likafullt i texten. De bör bli föremål för särbehandling. Uttryck som *med många om och men* bryts sålunda ut ur proceduren (*om* och *men* räknas som konjunktioner). På motsvarande sätt bryts titlar på böcker, skådespel, filmer och liknande ut och underkastas intern analys i enlighet med reglerna för analysen i övrigt.

Vid genomgången av flödesgrammatiken har jag resonerat under den förutsättningen att proceduren antas ha varit i verksamhet i den utsträckningen att de utfallande kategorierna är välfyllda. Detta har betydelse i sådana fall där kriterier definieras av proceduren. Som exempel kan nämnas att ordet *läsa* i ett som subjekt fungerande *att läsa* inte faller ut under 10 i moment C så länge verbstammen *läs-* inte har etablerats i moment B. Den hamnar alltså temporärt i restklassen pronomen. Allt eftersom beläggen strömmar in i proceduren, får därför enheterna i restklassen omprövas, varvid till en början uppflyttning i andra klasser ofta sker. Denna successiva revision av restklassen har inte särskilt markerats i flödesschemat. Det kunde emellertid ske genom att en order om revision efter exempelvis var tusende sats lades in omedelbart efter start i moment A. Det hela är en teknisk fråga.

En helt annan relevans har den kontinuerliga revision av klassen pronomen som motiveras av språksystemets dynamik. Ett ord som *vänstra* i uttrycket *den vänstra kolumnen* hamnar inom ramen för det nuvarande standardspråket i kategorin pro-

nomen (komparation förekommer ej). Om emellertid ett språkbruk av den stundom iakttagna typen *den vänstraste kolumnen* slog igenom, skulle enheten omkategoriseras som adjektiv. Strukturförändringar av detta och liknande slag ger och skall ge utslag.

I viss mån reflekterar proceduren den graduella karaktär som kännetecknar språkliga kategorier. De ändelsekomparerade och de perifrastiskt komparerade adjektiven kan sägas bilda punkter på en skala (som längre ner också har en punkt för vissa av enheterna i klassen pronomen). Motsvarande gäller de substantiv som kategoriseras på basis av speciesenkliks och de substantiv som utfaller på villkoret att de endast står i rektion.

Den ursprungligen uppställda modellen har successivt modifierats på basis av erfarenheterna från konfrontationen med materialet. Med hänsyn till denna ständiga återkoppling från beläggen kan arbetssättet betecknas som pragmatiskt. Denna karakteristik är också giltig i det avseendet att det vid utformningen av proceduren i väsentlig utsträckning har varit fråga om att underbygga, vidareutveckla och precisera kategoriseringen i en traditionellt given analys.

## Litteratur

- Allén, S., Grafematisk analys som grundval för textedering. Göteborg 1965.
- Allén, S., Indelningen av nusvenskans pronomen. (Arkiv för nordisk filologi 73 (1958), s. 51–70.)
- Chrystal, D., English. (i: Word classes. 1967. Reprinted from *Lingua* 17, s. 24–56.)
- Ellegård, A., Design for a mechanical distribution analysis of English word classes. (i: Structures and quanta. Interdisciplinary studies from the Scandinavian summer university. 12 (1963), s. 5–21.)
- Fries, Ch.C., The structure of English. New York 1952.
- Illustrerad svensk ordbok. Red. B. Molde. 3. uppl. Stockholm 1964.
- Svartvik, J., On voice in the English verb. The Hague 1966.

## Semantische Strukturen. Bemerkungen zur Arbeit von Inger Rosengren

Studia Neophilologica 40: 2. 1969.

Als semantische Strukturen können eine Reihe verschiedener Erscheinungen bezeichnet werden. Demgemäss weist die Bedeutungsforschung Strukturbeschreibungen sehr wechselnder Art auf. Um meinen folgenden Gesichtspunkten einen Hintergrund zu geben, erinnere ich an einige.

Eine Art begegnet uns in der sogenannten Feldtheorie Triers, derzufolge die Bedeutung der Wörter Felder bilden, die durch bestimmte Beziehungen zwischen den vorhandenen Einheiten gekennzeichnet sind. In diesem Modell wird nicht mit stufenweisen Übergängen gerechnet.

Eine andere Art finden wir bei den Glossematikern. Die Analyse der Inhaltsseite der Sprache ist bei Hjelmslev prinzipiell von derselben Art wie die der Ausdrucksseite. Grössere Strukturen werden also als Strukturen beschrieben, die aus minimalen Einheiten – Figuren – aufgebaut sind. Der Inhalt 'Kuh' ist, um ein gewöhnliches Beispiel zu nehmen, in den Figuren 'Rind' und 'Femininum' strukturiert.

Lyons (*Structural Semantics*, 1963) betrachtet semantische Strukturen als Beziehungen zwischen den Einheiten eines lexikalischen Teilsystems. Er betont insbesondere die Beziehungen Inkompatibilität oder Unvereinbarkeit, z. B. 'rot' – 'grün', 'Major' – 'Oberst', Antonymie oder Gegensatz, z. B. 'verheiratet' – 'unverheiratet', 'männlich' – 'weiblich', und Hyponomie oder Einordnung, z. B. 'Blume' – 'Tulpe'.

Bei Katz und Fodor (*The Structure of a Semantic Theory*, Language 1963), die mit der semantischen Seite der generativen Grammatik arbeiten, wird eine bestimmte Bedeutung durch eine gewisse generative Geschichte gekennzeichnet und durch eine gewisse Strecke in einer graphischen Darstellung (Baumdiagramm) repräsentiert. Die Verzweigungsstellen des Diagrammes nennen sie "grammatical markers", welche die Zugehörigkeit zu



verschiedenen Wortarten angeben, "semantic markers", welche die Bedeutung in das semantische Bezugssystem des Vokabulars einfügen, und "distinguishers", welche das hinzufügen, was für die betreffende Bedeutung charakteristisch ist. Ihr Beispiel ist *bachelor*, was u. a. als "noun (human) (male) [who has never married]" und "noun (human) [who has the first or lowest academic degree]" gekennzeichnet wird.

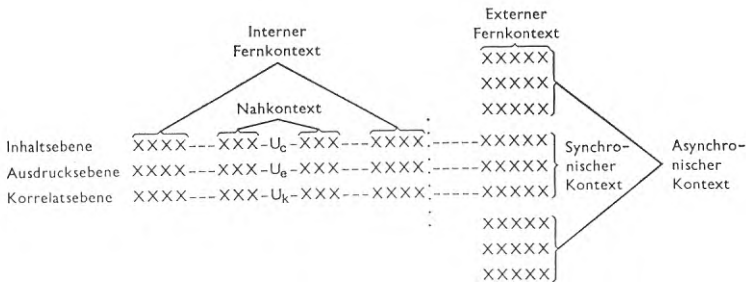
Auch das Auftreten von Wörtern im Kontext ist eine Art semantischer Struktur. Von den Forschern, die ich bisher genannt habe, betont besonders Lyons die Wichtigkeit des Kontextes. Bei verschiedenen Forschern hat der Begriff Kontext eine jeweils recht verschiedene Bedeutung erhalten. Malinowski hat, von ethnographischen Gesichtspunkten aus gesehen, den Ausdruck "context of situation" vorgeschlagen und die Bedeutung der Wörter als ihre Funktion im Rahmen des Satzes in der gegebenen Situation bezeichnet.

Ganz durchgesetzt hat sich der Kontextbegriff in der englischen Schule, in der Firth der wichtigste Name ist. Die Theorie Malinowskis wird hier auf solche Weise weiterentwickelt, dass Bedeutung als ein Komplex von kontextuellen Beziehungen auf allen Ebenen definiert wird: der situationalen, der grammatischen und auch der phonologischen. Der Kontextbegriff wird nicht nur auf die syntagmatische, sondern auch auf die paradigmatische Dimension ausgedehnt.

Die Ansicht, dass das Auftreten von Wörtern in Kontexten ein Ausdruck für ihre Bedeutung sei, kann mit Harris (*Distributional Structure*, Word 1954) auch so formuliert werden, dass Bedeutungsunterschiede mit Distributionsunterschieden in einer Korrelation stehen. Der Grad der Bedeutungsverschiedenheit entspricht also nach Harris dem Grade der Verschiedenheit in der Wortumgebung. Weil die Umgebung eines Wortes in gewisser Masse formalisiert werden kann, ermöglicht diese Betrachtungsweise eine statistische Behandlung von semantischen Strukturen.

Es ist in diesem Zusammenhang interessant, den Kontextbegriff vom Gesichtspunkt des den Korpus analysierenden Forschers näher zu untersuchen, sei seine Einstellung semantisch,

lexikographisch oder realphilologisch, syntaktisch oder morphologisch, phonologisch oder "graphonomisch", stilistisch oder anderer Art. Meine Meinung über diese Frage wird in Figur 1 veranschaulicht. Der Faktor, der direkt beobachtet werden kann, ist der sprachliche Ausdruck – der geschriebene oder (was spätere Jahrzehnte betrifft) der gesprochene. Die beobachtete Einheit auf der Ausdrucksebene – nicht notwendigerweise ein einzelnes Wort – wird  $U_e$  bezeichnet, übrige Einheiten X. Die nächste Umgebung der beobachteten Einheit, die auf verschiedene Weise abgegrenzt werden kann, wird Nahkontext genannt.



FIGUR 1. *Kontext.*

Die übrigen Teile des Korpus werden interner Fernkontext genannt. Der Korpus (als ein prinzipiell synchronisches Material konzipiert) und damit  $U_e$  steht ausserdem in Kontextbeziehung zu anderem Material, dem externen Fernkontext, der synchronisch oder asynchronisch sein kann. In der Figur markiert die punktierte Linie die Grenze zwischen dem untersuchten Korpus und dem übrigen Material.

Der sprachliche Inhalt erscheint als eine Funktion des Ausdrucks. Man muss aber bemerken, dass die Inhaltsebene nicht den linearen Charakter der Ausdrucksebene hat. Die Umgebung der auf der Inhaltsebene beobachteten Einheit  $U_e$  – des Inhaltskontexts – lässt sich mit denselben Termini beschreiben wie die Ausdrucksseite. Die beiden jetzt genannten Ebenen, die Aus-

drucksebene und die Inhaltsebene, können als die linguistischen Ebenen bezeichnet werden.

Als eine weitere Funktion der Ausdrucksebene erscheint eine extralinguistische Ebene, die Korrelatebene oder Applikationsebene. Es handelt sich um die wirklichen oder gedachten Erscheinungen, Vorgänge usw., auf welche sich der sprachliche Ausdruck bezieht. Bei einer genaueren Errichtung der Korrelatebene spielt das sogenannte *cultural overlap* (Lyons S. 40 ff.) eine wichtige Rolle, wenn es sich um andere Zeitperioden und Sprachen handelt, als diejenigen, denen der Forscher selbst angehört.

Es bleibt dahingestellt, ob der Fernkontext formalisiert werden kann. Der Nahkontext andererseits ist auf jeden Fall in gewissem Masse formalisierbar. Im allgemeinen dürfte gelten, dass die semantische Beschreibung um so gröber wird, je kleiner der Teil des operativen Kontexts ist, der bei einer Untersuchung beachtet wird. Doch auch eine Beschreibung mit einem groben semantischen Massstab, der sich auf ein quantitatives Kontextstudium gründet, ist von grossem Interesse als Beitrag zur Diskussion über Bedeutungsdefinition und Bedeutungsklassifikation.

Eine Untersuchung dieser Art wurde kürzlich von Inger Rosengren als Nr. 38 in der Reihe "Lunder germanistische Forschungen" (herausgegeben von Tage Ahldén) vorgelegt. Der Titel der Dissertation ist *Semantische Strukturen. Eine quantitative Distributionsanalyse einiger mittelhochdeutscher Adjektive* (Lund und Kopenhagen, 1966). Es handelt sich also um die Untersuchung eines mittelhochdeutschen Materials auf distributionaler Grundlage und mit statistischer Technik.

Das untersuchte Material besteht aus einer Gruppe von Adjektiven mit *senfte* und *milte* als Kern. Für das Frühmittelhochdeutsche führt die Verfasserin 44 Wörter mit insgesamt 1679 Belegen an. Im späteren Mittelhochdeutschen belegt sie 12 dieser Wörter nebst 7 Gegenwörtern mit insgesamt 2445 Stellen. Das ganze Material umfasst also etwa 4000 Belege.

Was oben als Ausdruckskontext bezeichnet wurde, entspricht in der Dissertation der Kategorie Funktion, und was als Inhalts-

kontext bezeichnet wurde, der Kategorie Kontext. Die Verfasserin richtet prinzipiell ihre Aufmerksamkeit auf diese beiden Ebenen. Die erstere Kategorie wird in 10 Gruppen eingeteilt und die letztere in 24. Zweck der Untersuchung ist, das Verhältnis der Wörter zu diesen Gruppen festzustellen. Die statistische Technik, die vor allem verwendet wird, ist der  $\chi^2$ -Test. Dieser Test ist eine Methode, mit welcher die Signifikanz der Häufigkeitsunterschiede zwischen Gruppen bestimmt wird.

Mit den angegebenen Methoden behandelt die Verfasserin zuerst das frühmittelhochdeutsche Material. Dies wird im zweiten Abschnitt der Dissertation vorgenommen (S. 28–61), der den umfassendsten Teil bildet. Hier werden die Wörter anfangs in 4 Gruppen, a, b, c und d eingeteilt je nach ihrem Auftreten in den beiden Kontexten *Gott* und *Maria* (S. 35). Danach wird die prozentuale Übereinstimmung im Kontext innerhalb der Wortgruppen und zwischen ihnen untersucht. Dabei erscheint ein signifikanter Unterschied zwischen der Übereinstimmung innerhalb der Gruppen und der zwischen ihnen.

Die kontextuelle Übereinstimmung innerhalb der Gruppen ist aber verschieden gross. Mit Rücksicht auf diesen Faktor und unter Ausschluss solcher Wörter, die weniger als 10 Belege haben, erhält die Verfasserin die auf S. 43 angeführten Gruppen: a<sup>1</sup>, c<sup>1</sup>, a<sup>2</sup> und b<sup>1</sup>+d. Dann werden die in die genannten Gruppen eingehenden Wörter nach einem neuen Prinzip klassifiziert, und zwar nach der Beziehung zwischen der Belegzahl der Wörter innerhalb bzw. ausserhalb der Kontexte 1–5, der sogenannten göttlich–menschlichen Sphäre. Die Umgruppierung ergibt 3 Gruppen, als 1, 2 und 3 bezeichnet (S. 44).

Sowohl die mit Buchstaben als auch die mit Ziffern bezeichneten Wortgruppen sind aus einem Studium der Kontexte hervorgegangen. Eine Analyse ihrer Beziehung zu den Funktionsgruppen zeigt, dass sich die Wortgruppen signifikant voneinander unterscheiden, auch hinsichtlich ihrer Frequenzverteilung auf die Funktionsgruppen. Zusammenfassend konstatiert die Verfasserin, dass es sich bei den Wörtern nicht so sehr um eine Struktur als vielmehr um mehrere, sich teilweise deckende Strukturen handelt.

Im nächsten Abschnitt (S. 62–85) wird das spätere mittelhochdeutsche Material untersucht. Hier wird teils die Frage gestellt, ob die 12 behandelten Wörter während der beiden Perioden eine verschiedene Verteilung aufweisen, teils die Frage, wie sich die angewandte Methode bewährt.

Zuerst wird nachgewiesen, dass die Korrelation zwischen den Rangordnungen der Kontexte während der beiden Perioden signifikant ist, was die Verfasserin so deutet, dass die Wörter ungefähr denselben Verwendungsbereich während der betreffenden Perioden haben. Danach werden die Wörter nach denselben Prinzipien wie vorher eingeteilt, d. h. teils nach Vorkommen bzw. Nicht-Vorkommen in den Kontexten *Gott* und *Maria*, teils nach der Beziehung zwischen den Belegzahlen innerhalb und ausserhalb der göttlich–menschlichen Sphäre. Es zeigt sich, dass das Ergebnis in beiderlei Hinsicht im grossen und ganzen für die beiden Perioden dasselbe ist: die Gruppen A und B (S. 68) bzw. I–IV (S. 69).

Unter Ausschluss gewisser Kontexte nimmt die Verfasserin eine Klassifikation in 7 Kontextgruppen (S. 71) vor, und danach einen Vergleich zwischen den Wortgruppen A (*milte, senfte, stille, süeze*) und B (*blæde, bræde, kranc, linde, lîse, senftmüetec, swach, weich*) im Frühmittelhochdeutschen und späteren Mittelhochdeutschen hinsichtlich ihrer Frequenzverteilung auf die Kontextgruppen. Die beiden Gruppen A weisen einen signifikanten Unterschied auf. Dies veranlasst eine Untersuchung der 4 Wörter der Gruppe hinsichtlich der Rangordnung zwischen den einzelnen Kontexten während der beiden Perioden. Die Korrelation erweist sich durchgehend als signifikant. Wo es um den Unterschied zwischen den Gruppen A geht, den der  $\chi^2$ -Test zeigte, denkt sich die Verfasserin (und streift hier die Korrelatsebene), dass er davon abhängen kann, dass gewisse Themen und damit gewisse Kontexte in den betreffenden Perioden begünstigt wurden. In Ermangelung der Frequenzuntersuchungen des Wortschatzes während der beiden Perioden können wir, wie die Verfasserin sagt, diese Frage nicht entscheiden.

Am Ende des besprochenen Abschnittes wird die Frequenzverteilung der Wörter auf Funktionsgruppen während der beiden

Perioden untersucht. Sie erscheint im grossen und ganzen als unverändert. Zusammenfassend wird konstatiert, dass die behandelten Wörter semantische und syntaktische Stabilität zeigen. Diese Stabilität wird als eine indirekte Kontrolle der angewandten Methoden beurteilt.

Der vierte Abschnitt (S. 86–100) behandelt einige Probleme der Gegenwörter. Dabei werden 7 Wörter untersucht (*bitter, gesunt, hart, scharpf, starc, strenge, sÛr*), von denen angenommen werden kann, dass sie in einem semantischen Gegensatzverhältnis zu den 12 Wörtern des späteren Mittelhochdeutschen stehen.

Im Hinblick auf ihre Frequenz innerhalb und ausserhalb der göttlich-menschlichen Sphäre werden die Wörter in 3 Gruppen eingeteilt (S. 86 f.): *gesunt; bitter, scharpf, sÛr; hart, starc, strenge*. Bei einem Vergleich zwischen diesen und gewissen früher behandelten Wörtern findet die Verfasserin vor allem kontextuelle Unterschiede zwischen *gesunt* und *kranc*, zwischen der Gruppe *hart, starc, strenge* und den einzelnen Wörtern *milte, senfte* und *sÛeze*. Diese Unterschiede scheinen signifikanten Unterschieden in der Funktion zu entsprechen.

Im Abschnitt 5 (S. 101–104) wird schliesslich als Kontrolle ein Vergleich angestellt zwischen den Ergebnissen, die in bezug auf die semantischen Strukturen im Frühmittelhochdeutschen ermittelt wurden, und den Wortbedeutungen, die in Lexers Mhd. Wörterbuch zu finden sind. Vier Punkte werden hervorgehoben.

Die religiösen Termini *barmherze, diemüete, gedultec, gehörsam* und *genædec* werden in der Gruppe a<sup>1</sup> gefunden und gehören auch zur Gruppe 1.

Die Wörter, für die Lexer die Bedeutung 'freundlich, gütig' angibt, findet man erstens in der Gruppe a<sup>2</sup> (diese Wörter gehören, *sÛeze* ausgenommen, auch zur Gruppe 2), zweitens in der Gruppe a<sup>1</sup> (nämlich die Wörter der Gruppe, die auch zur Gruppe 2 gehören) und drittens – im Falle *holt* – in den Gruppen c<sup>1</sup> bzw. 1. Ausserhalb dieser bei Lexer angeführten Gruppe, aber innerhalb der entsprechenden Gruppen in der Einteilung der Verfasserin, erscheinen die Wörter *stille, undertân* und *siech*.

Die Wörter, für die Lexer die Bedeutung 'schwach, angenehm, weich' angibt, entsprechen der Gruppe 3 in der Einteilung der Verfasserin (gehören ausserdem mit einer Ausnahme zu den Gruppen b<sup>1</sup> und c<sup>1</sup>). Das Wortpaar *bræde-blæde* ist ein Sonderfall.

Schliesslich entsteht eine Restgruppe von 5 wenig belegten Wörtern: *erbarmec, laz, lise, miteware* und *træge*.

Die Übereinstimmung ist im ganzen als gut zu betrachten. Die Verfasserin sieht in ihrem Ergebnis einen Ausgangspunkt für eine weitere Distributionsanalyse, die u. a. nähere Auskunft in Fällen wie *siech* und *stille* geben könnte.

In einigen Punkten veranlasst die Dissertation Kommentare in verschiedener Hinsicht. Ich werde einige Fragen über das Material, die Einteilung in Funktions- und Kontextgruppen; die Gruppierung der Wörter während der beiden Perioden, die statistische Behandlung und die Methode im allgemeinen erörtern.

Was die Repräsentanz des Materials betrifft, soll man hinsichtlich des Frühmittelhochdeutschen beachten, dass das Bairische stark überwiegt (S. 28), bezüglich des späteren Mittelhochdeutschen, dass die grosse "Fachliteratur" nur schwach vertreten ist (S. 22). Die Verfasserin soll aber nicht deswegen getadelt werden. Ihre Stichprobe beruht letzten Endes darauf, was an mittelhochdeutscher Literatur überliefert und herausgegeben ist.

Der Ausgangspunkt für die Aufstellung der untersuchten Wortgruppen ist selbstverständlich ein beliebiger. Dagegen wäre die Darstellung eleganter gewesen, wäre die Verfasserin von einem Worte ausgegangen statt von zweien, die ohnedies schon eine Gruppe bilden. Die Abgrenzung der Gruppe ist auch beliebig: die Verfasserin hat solche Wörter gesammelt, von denen angenommen werden konnte, dass sie *senfte* und *milte* semantisch nahe ständen (s. S. 21). Es ist möglich, dass es einen Zusammenhang gibt zwischen der Abgrenzungsweise und der oben genannten Restgruppe, die bei dem Vergleich mit Lexers Mhd. Wörterbuch entsteht.

Der Umfang des Kontextes wird von der Verfasserin auf das betreffende Adjektiv und das von diesem bestimmte Substantiv beschränkt. Dies scheint mir, wie die Untersuchung angelegt ist, eine weise Begrenzung zu sein. Es bedeutet jedoch, dass man sich einer Menge Information begibt. Den gewählten Kontext beleuchtet die Figur 1 oben aus einer gewissen Perspektive. Das beobachtete Adjektiv entspricht  $U_e$ , sein Hauptwort einem nebenstehenden  $X$ , der Inhalt des Hauptwortes dem darüber stehenden  $X$  (da die Inhaltsebene hier als eine direkte Entsprechung der Ausdrucksebene angesehen werden kann). Die Notwendigkeit eines vermehrten Kontextes hat die Verfasserin besonders bei der Behandlung der Gegenwörter bemerkt, was nicht überrascht (s. S. 104).

Die Einteilung in Funktions- und Kontextgruppen erfordert einige Kommentare. Was die Funktionsgruppen (S. 24) betrifft, soll bemerkt werden, dass die angegebene Einteilung gewöhnlich nicht durchgeführt wird. Die Gruppen V–IX werden in der Praxis zu einer Gruppe r (Restgruppe) zusammengeschlagen, während die Gruppe X, Rest genannt, nicht berücksichtigt wird. In der Tabelle 113 (S. 96) werden auf eine etwas verwirrende Weise die früher verwendeten Bezeichnungen für die fünf resultierenden Funktionsgruppen aufgegeben. Die Ziffern 1–5 über den einzelnen Reihen bezeichnen anderswo Kontextgruppen und sollten gegen a, k, v, s und r ausgetauscht werden.

Gegen die Kontexteinteilung (S. 25 ff.) habe ich gewisse Einwände. Der wichtigste ist, dass die Einteilung nicht konsequent ist. Die Verfasserin unterscheidet eine göttlich–menschliche Sphäre (Gruppe 1–5), teilt aber einschlägige Wörter teils dem Kontext 8 (*Herz*), teils dem Kontext 19 (*Land*) zu. Es handelt sich hier um eine Verwendung im übertragenen Sinn: *Herz* als Bezeichnung für den ganzen Menschen, *Land* als Bezeichnung für die Einwohner eines Landes. Den genannten Gruppen können die Wörter nur aus formalen Gründen zugeordnet werden, was die Systematik beeinträchtigt und, wie zu erwarten, im folgenden zu gewissen Konsequenzen führt.

Der auf S. 43 verwendete Einteilungsgrund (die Beziehung zwischen der Belegzahl innerhalb und ausserhalb der göttlich–



menschlichen Sphäre) gibt demzufolge einen Ausschlag für *diemüete*, den die Verfasserin ablehnt, mit dem Hinweis dass er durch Belege im Kontext *Herz* verursacht wurde. Der Vergleich mit Lexer bei *undertan* veranlasst die Verfasserin, ebenso die Behandlung des Kontexts *Land* als "nicht glücklich" (S. 102) zu bezeichnen. Die Lösung des Problems scheint mir zu sein, dass man mit zwei Kontexten *Herz* und zwei Kontexten *Land* rechnet: *Herz*<sub>1</sub> und *Land*<sub>1</sub> innerhalb der göttlich-menschlichen Sphäre, *Herz*<sub>2</sub> und *Land*<sub>2</sub> ausserhalb dieser Sphäre. In diesem Zusammenhang soll auch genannt werden, dass die Kontextgruppen 21–23 (neutrale Formen) nur aus formalen Gründen angesetzt worden sind.

Wir kommen dann zu der Gruppierung der Wörter im frühmittelhochdeutschen Material. Auf S. 43 f. wird eine Umgruppierung der Wörter auf Grund der oben erwähnten Beziehung zwischen der Belegzahl innerhalb und ausserhalb der göttlich-menschlichen Sphäre vorgenommen. (Übrigens möchte ich einige falsche Zahlangaben auf S. 43 korrigieren: *senfte* hat 78, nicht 98 Belege innerhalb der Sphäre (vgl. die Tabelle 63, S. 70); *guotlich*, *lieplich*, *undertân* haben zwischen  $\frac{1}{4}$  und  $\frac{1}{8}$ , nicht zwischen  $\frac{1}{3}$  und  $\frac{1}{7}$  ihrer Belege ausserhalb der Sphäre, *diemüete* hat auf entsprechende Weise  $\frac{1}{4}$ , nicht  $\frac{1}{3}$ .) Den Gruppen 1, 2 und 3 auf S. 44 werden Wörter zugeordnet, die innerhalb der genannten Sphäre viel häufiger, häufiger bzw. seltener sind als ausserhalb der Sphäre. Hier spielt nun der Kontext *Herz* eine gewisse Rolle. Das oben genannte *diemüete* wäre der Gruppe 2, nicht der Gruppe 1 zugeteilt worden, wenn die *Herz*-Belege nicht zu der göttlich-menschlichen Sphäre gebracht worden wären. Auch andere Wörter haben aber einige ihrer Belege im Kontext *Herz*, u. a. haben dort *bræde* 10 und *mitewære* 1 (s. Tabelle III, S. 110). Die Konsequenz erfordert, dass diese Wörter wie *diemüete* behandelt werden. Ihre Relationszahlen werden da von  $\frac{16}{24}$  bzw.  $\frac{13}{2}$  zu  $\frac{26}{14}$  bzw.  $\frac{14}{1}$  geändert. Daraus ergibt sich, dass *bræde* von der Gruppe 3 in die Gruppe 2 übergeht, *mitewære* von der Gruppe 2 in die Gruppe 1.

Die Abgrenzung zwischen den drei Wortgruppen ist subjektiv. Eine Abgrenzung nach völlig objektiven Gründen kann nicht gemacht werden, aber es scheint mir, als wäre eine grössere Trennungsschärfe möglich. Ein einfaches Mass für die Beziehung zwischen den Belegzahlen innerhalb bzw. ausserhalb der Sphäre erhält man dadurch, dass man sie durcheinander dividiert. Man erhält folgende Werte:

die Gruppe 1	9-∞
die Gruppe 2	1-13
die Gruppe 3	< 1.

Die Überschneidung der Gruppen 1 und 2 wird durch die Relationszahl von *lîse* (13/1) verursacht. Das Ergebnis lässt es meiner Meinung nach angebracht erscheinen, dieses Wort von der Gruppe 2 zur Gruppe 1 überzuführen. Die Skala wird da: 9-∞, 1-8, <1.

Es wäre aber auch nicht unangebracht, *guotlich* mit der Relationszahl 49/7 und *undertân* mit der Relationszahl 103/15 (nachdem die 16 Belege im Kontext *Land* zur göttlich-menschlichen Sphäre übergeführt worden sind) zur Gruppe 1 zu bringen. Dies würde die Skala 7-∞, 1-6 (in der Praxis 1-3), <1 ergeben. Diese Berichtigung würde u. a. zur Folge haben, dass das oben erwähnte Problem *undertân* beseitigt worden wäre (S. 102). Die Umgruppierungen, die ich vorgeschlagen habe – *bræde* von der Gruppe 3 zur Gruppe 2, *mitewære*, *lîse*, *guotlich* und *undertân* von der Gruppe 2 zur Gruppe 1 – bedeuten, dass die Wortzahl in den Gruppen 1-3 von 7, 13 bzw. 8 auf 11, 10 bzw. 7 geändert wird. Es handelt sich also um eine recht beträchtliche Änderung.

Die wortauswahl für die spätere mittelhochdeutsche Periode ist, soviel ich sehen kann, nicht nach einem eindeutigen Prinzip vorgenommen worden. Einerseits gibt es unter den zwölf Wörtern kein Wort von der Gruppe 1. Andererseits gibt es dort zwei, *senftmüetec* und *swach*, die während der älteren Periode eine so geringe Frequenz haben, dass sie schon früh aus der Diskussion verschwinden und also zu keiner der Gruppen 1-3 gehören.

Über die Einteilung der zwölf Wörter in frequente und weniger frequente Wörter werden auf Seite 21 und Seite 64 widersprechende Angaben gemacht. Eine Kontrolle der Tabelle 59 auf Seite 65 (die Rangordnung der Kontexte der weniger frequenten Wörter) lässt erkennen, dass die Angabe auf Seite 64 die korrekte ist: die frequente Gruppe besteht aus *kranc*, *stille* und *süeze* (*milte* ist also beim Aufzählen der weniger frequenten Wörter auf Seite 21 weggefallen).

Die statistische Behandlung ist nach meinem Urteil gut. Einige kleinere Versehen möchte ich hier erwähnen. Der  $\chi^2$ -Test ist gewissen Einschränkungen unterworfen. Wenn die Anzahl der Kolumnen grösser ist als 2 kann der Test verwendet werden, wenn weniger als 20% der Zellen einen kleineren theoretischen Wert haben als 5, und wenn keine Zelle einen kleineren theoretischen Wert hat als 1. Werden die Bedingungen nicht erfüllt, muss man Reihen vereinen, um die Werte zu erhöhen oder, wenn dies nicht sinnvoll ist, darauf verzichten, den Test zu verwenden. In der Tabelle 123 (S. 99) beträgt nun der theoretische Wert 0,6 auf Zeile 1 in Reihe 4. In der Tabelle 45 (S. 57) beträgt der theoretische Wert 2 sowohl auf Zeile 1 als auf Zeile 2 in Reihe 3. In diesen Fällen hätte also der  $\chi^2$ -Test nicht verwendet werden sollen.

Auf Seite 85 fasst Inger Rosengren ihre Gesichtspunkte zur Methodik der Dissertation zusammen. Sie erwähnt, dass gewisse Unterschiede zwischen den Sprachperioden festgestellt werden konnten, dass aber die Wörter im grossen und ganzen semantische und syntaktische Stabilität zeigen. Wie oben gesagt, betrachtet sie diese Stabilität als eine indirekte Kontrolle der angewandten Methoden. Meiner Meinung nach ist dieser Gedankengang im Prinzip unhaltbar. Man kann nicht gleichzeitig diachronische Unterschiede nachweisen und die Methoden kontrollieren. Eine Prüfung der Methoden muss an vergleichbarem synchronischem Material vorgenommen werden. Dies kann anhand von älteren Texten geschehen, könnte aber aus mehreren Gründen an modernem Material erfolgen, was grössere Sicherheit gewährleisten würde.

Behandelt man modernes Material, werden Einteilungen, die sich auf das Vorkommen bzw. Nicht-Vorkommen in einem gewissen Kontext gründen, zuverlässiger. Hierbei bietet sich auch, soweit ich sehen kann, die ansprechende Möglichkeit mit numerischer Taxonomie zu arbeiten. Diese Methode umfasst teils eine numerische Schätzung der Ähnlichkeit zwischen beobachteten Einheiten auf Grund einer grossen Anzahl von Kriterien, teils eine Gruppierung der Einheiten nach Ähnlichkeit (s. Svartvik, *On Voice in the English Verb*, 1966). Gerade darauf will ja die semantische Distributionsanalyse hinaus. Durch Verwendung einer so grossen Anzahl von Kriterien dürfte es möglich werden, die Formalisierung des Nahkontextes bis ins Detail durchzuführen, was seinerseits zu einem nuancierteren Bild der semantischen Strukturen führen würde.

Die Verfasserin beschäftigt sich gegenwärtig mit der Erforschung von Material aus der modernen deutschen Zeitungssprache. Die in ihrer Dissertation erzielten Ergebnisse lassen erwarten, dass auch diese neue Arbeit interessante Ergebnisse zeitigt.

## Engelskt och svenskt. En översikt av det språkliga inflytandet

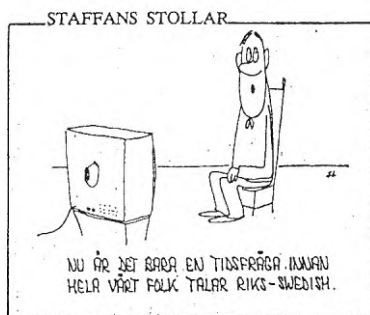
Språkdata 1968.

En samlad behandling av engelskans inflytande på svenskan saknas ännu. Här väntar en viktig uppgift. En monografi kunde bli både omfattande och mångfacetterad: den kunde ta upp influenser inom hela det språkliga systemet, inom olika språkmiljöer (sociala, kulturella, geografiska) och under olika tidsperioder. Den principiella frågan om språklig påverkan kunde också belysas. I det följande skall relationernas karaktär och några drag i deras utveckling behandlas.

De båda nedan återgivna dokumenten belyser förhållandena i runsvenskan och nusvenskan, de båda ändpunkterna i det tidskede som skall överblickas.

sar [suin] x uarþ x tuþr  
 a x iutlati x an skulti  
 fara x til x iklanþs x  
 kuþ x ialbi x an s at uk  
 salu x uk x us muþir x  
 betr x þan an karþi til

Runsten från Husby-Lyhundra,  
 Uppland, c. 1020



SvD (B) 29/4 1966

Texten till vänster är senare delen av en inskrift på en runsten från Husby-Lyhundra i Uppland (återgiven efter Sven B. F. Jansson, Runinskrifter i Sverige 1963). Den handlar om en man som hette Sven. Han dog i Jutland. Han skulle fara till England, står det. Och inskriften slutar: Gud och Guds moder hjälpe hans ande och själ bättre än han förtjänade. De utomspråkliga relationerna mellan Sverige och England på 1000-talet hade två sidor: svenska vikingar drog som inskriften visar i västerled, engelska missionärer för till Sverige. Ett vittnesbörd om de punktuella

språkliga inflytelser som kontakten ledde till har vi på runstenen i formen *salu* (rad 5), en böjningsform av *sal* 'själ'. Läget niohundra år senare sammanfattar Staffan Lindéns TV-tittare för sin del i repliken: Nu är det bara en tidsfråga innan hela vårt folk talar riks-Swedish.

De båda utförligaste studierna av det engelska inflytandet på svenskan är Nils Bergstens uppsats Om engelska lånord i svenskan, tryckt i *Språk och Stil* 1915, och Elof Hellquists kapitel Engelska och engelsk-amerikanska i andra delen av *Det svenska ordförrådets ålder och ursprung* 1930. Båda är som framgår av titlarna av lexikalisk typ. Norskan, tyskan och holländskan är i vissa avseenden bättre tillgodosedda genom Aasta Stenes arbete *English loan-words in modern Norwegian. A study of linguistic borrowing in the process* 1946, Broder Carstensens *Englische Einflüsse auf die deutsche Sprache nach 1945*, som kom 1965, respektive R. W. Zandvoorts *English in the Netherlands. A study in linguistic infiltration* 1964.

Jag nämnde inledningsvis runsvenskan. När det gäller periodindelningen erinrar jag om att den fornsvenska perioden närmast motsvaras av den medelengelska, som brukar anges till tiden mellan 1100 och 1500. Det kan härvid observeras att de medelengelska källorna under 1100-talet flyter mycket sparsamt. Tiden före 1100 är den fornengelska perioden särskilt förr även kallad den anglosaxiska. Det är alltså denna som närmast motsvarar den runsvenska. Tiden efter 1500, den nyengelska perioden, motsvaras av den nysvenska.

Det är uppenbart, att det fornengelska inflytandet på svensk kristen terminologi var starkt under ett tidigt skede i den svenska kyrkans historia, skriver Carl-Eric Thors i sitt arbete *Den kristna terminologien i fornsvenskan* 1957. Man lånade in sådana ord som behövdes i det kristna missionsarbetet. Det finns engelska inslag bland dop- och nattvardstermerna, namnen på de viktigaste högtiderna och de liturgiska orden. Bland de ord som visar på engelskt inflytande i någon form nämner Thors *kirikia*, *skruþer*, *kalker*, *fonter*, *skript* 'skriftermål', *askuopinsdagher*, *fastuganger* 'fastans början vid fastlags söndagen', *mässa*, *ottusanger* och *aptunsanger*. I vissa fall är hithörande ord endast

belagda i fornvästgötskan, så *penkost* 'pingst' och *husl* 'nattvarden' (av fe. *hus(e)l*). Det fornengelska inflytandet tycks emellertid ha upphört förhållandevis tidigt. Ett av tecknen på detta är att just formen *sal* av ordet *själ* endast är belagd på runstenar. Den nuvarande formen av ordet *själ* återgår på kontinentalgermanska former.

Också vissa icke-kyrkliga fornsvenska ord brukar med större eller mindre säkerhet betraktas som fornengelska lån. Sådana är *bläck*, *klut*, *kölna* 'torkhus' och *på-* i *påfågel*. Vidare återgår fornsvenska *hirþ-* i sammansättningarna *hirðhränger*, *hirþgarper*, *hirþman* på fornengelska *hir(e)d* 'familj, broderskap'. Ordet *hird* har f.ö. långt senare upptagits på nytt i svenskan från den fornvästnordiska litteraturen. Jag vill inte heller underlåta att nämna fornvästgötska *lavarþer* 'husbonde', som är lån från fornengelska *hláford* < \**hláfweard*, egentligen 'brödvaktare'. Senare har vi lånat in ordet på nytt, nämligen i den moderna formen *lord*. Det tidigaste belägget är från 1788.

Även på ett annat område gjorde sig ett engelskt inflytande gällande under vikinga- och missionstiden. Jag tänker då på den anglosaxiska skriften som starkt har påverkat den nordiska skriften, särskilt den västnordiska men även den fornvästgötska. Av de särskilda skrivtecknen kan jag nämna þ och ð. Tecknet þ som vann störst spridning av de två i fornsvenska texter, hade upptagits i den anglosaxiska insulärskriften från runskriften, tecknet ð, dvs. d med diakritiskt tecken, hade nybildats.

Engelskans syntaktiska inflytande på svenskan tycks inte ha varit särskilt kraftigt. Det är därför så mycket intressantare att ta del av Carl Ivar Ståhles resonemang på en viktig punkt i arbetet Syntaktiska och stilistiska studier i fornnordiskt lagspråk 1958. Grundstommen i fornsvenskt lagspråk utgörs, säger Ståhle, av bestämmelser som består av enkel frågeformad konditionalsats följt av eftersats inledd med konjunktiv. Ett exempel är *Hængir klocka i kyrkiu falder i hovod mannæ. bøte sopl firi. niv marchum* ÄVgL.

Mot detta grundmönster bryter framför allt en lag, nämligen Äldre Västgöotalagen, som har ett ganska stort antal belägg på inledande konjunktionsbisatser, *æn*-satser: *æn maþer krævær*

*man skuld, þa skal han a grannæ hans kallæ ...* Ståhle ställer frågan varför det är så och erinrar då om att Västgötalagen upp-tecknades under inflytande från Norge. Han ser den som en nedteckning av muntlig lagsaga i samband med nyredigering efter det enda skriftliga mönster som fanns till hands, nämligen den norska lagen, med vilken lagman Eskil bör ha varit för-trogen. Och i västnordiskt lagspråk är konjunktionsbisatser, *ef*-satser, det normala uttrycket för förutsättning. Även om nu den konditionala konjunktionsbisatsen i och för sig är rent nordisk, är det frågan, menar Ståhle vidare, om inte den höga frekvensen i de västnordiska lagarna sammanhänger med främmande inflytande. Han pekar då på att de kungar som under 900- och 1000-talen verkade för kristendomens införande i Norge och därvid i första hand hade engelska missionärer till hjälp också får antas ha gjort insatser i norskt rättsväsen. Skriftsystemet i Norge kom som nämnts att bestämmas av engelska mönster, och det förefaller inte osannolikt, menar han, att de tidiga och viktiga förbindelserna med det kulturellt så överlägsna England med dess rika lagstiftning på folkspråket också kan ha inverkat på själva språkformen i de västnordiska lagarna. Och i fråga om det språkliga uttrycket för förutsättningen föreligger en central likhet mellan fornväst-nordisk och fornengelsk lag: mot de fornvästnordiska *ef*-satserna svarar de fornengelska *gif*-satserna. (Bakom dessa ligger för övrigt i sin tur det latinska lagspråkets *si*-satser.) Ståhle gör alltså troligt att den syntaktiska egendomligheten på en viktig punkt i vår äldsta lagkodex beror på engelskt inflytande, förmedlat av norskan.

Vissa fornsvenska ord är att återföra på medelengelskan snarare än på fornengelskan. Ett sådant är *daggare* 'dolk', biform till *dagge*, som sammanställs med engelska *dagger*. Ett annat är *nobel* 'ett slags mynt, spec. eng. guldmynt'. Erik Holmkvist har i ett par uppsatser i Arkiv för nordisk filologi 1950 och 1951 fäst uppmärksamheten på relationerna mellan medelengelskan och fornsvenskan när det gäller bergverksterminologin. Han har observerat att *grovsmed* heter *swartasmid* i Arboga stads tänkebok och att även *hvitasmid* är belagt i fornsvenskan. Han uppfattar dessa som översättningslån från medelengelskans motsvarigheter



till *blacksmith* respektive *whitesmith* och påminner om att de tyska orden är *Grobschmied* respektive *Kleinschmied*, motsvarande fornsvenska *grofsmider* respektive *klensmedher*. Det är dessa båda senare former som har segrat i svenskan, dvs. det engelska inflytandet har undanträngts av det tyska. Holmkvist tar också upp *puta* med betydelsen 'liten gruva', som han menar återgår på engelska *pit*. Detta ord har i medelengelskan ofta *u*-former: *pute*, *putte* m.m. Han nämner för övrigt också *sutare* 'skomakare' som han vill återföra på medelengelska *soutere* snarare än med Hellquist på fornengelska *sutere*.

Jag har här pekat på en del spår av engelskt inflytande under fornspråklig tid. Innan jag behandlar de följande perioderna, kan det vara på sin plats att i någon mån gå in på principfrågan om kontakt mellan olika språk. Denna fråga har diskuterats en hel del under senare år. Bland internationellt uppmärksammade arbeten på området kan jag nämna Einar Haugens *The Norwegian language in America* 1953 och Uriel Weinreichs *Languages in contact* samma år. Man kan säga att det har gällt att komma åt de allmänna principerna för det språkliga beteendet i den tvåspråkiga situationen. Det har befunnits att kontaktfenomenen kan placeras in på en skala från vad Haugen kallar kodbyte eller kodskifte via interferens till integration. Kodbyte föreligger när språkbrukaren fullständigt går över från det ena språkssystemet till det andra, vilket t.ex. kan inträffa vid en ord- eller meningsgräns. Men mycket ofta är kodskiftet inte fullständigt. Det kan för att ta ett enkelt exempel vara fråga om ljudsubstitution. I sådana fall föreligger interferens mellan de båda språkssystemen. Interferensen behöver emellertid inte förbli interferens. Om företeelsen upprepas ofta kan den bli en del av normen, och då föreligger integration. Gränsen mellan interferens och å ena sidan kodbyte, å andra sidan integration är naturligtvis svårdragen, om ens möjlig att dra. Själva den teoretiska modellen förefaller mig emellertid klagörande.

Går vi med de här begreppen till vokabulärsystemet, finner vi till att börja med att vad som brukar kallas citatord just är ett exempel på kodbyte. Citatorden brukar ju också markeras med citationstecken, kursiv eller någon bestämning av typen *s.k.*

Vidare finner vi att det som ofta kallas främmande ord är ett exempel på interferens mellan två språkssystem. Interferensen kan ta sig uttryck i accenten, den fonotaktiska strukturen, böjningen eller annat. Integration slutligen har vi exempel på bland de fall som Karl-Hampus Dahlstedt i skriften *Främmande ord i nysvenskan 1962* kallar "hemtama" ord. Svårigheten med gränsdragningen mellan kodbyte, interferens och integration motsvaras av svårigheten med gränsdragningen mellan citatord, främmande ord och "hemtama" ord. Denna svårighet ligger i sakens natur. Det är som jag nämnde fråga om en skala, ett kontinuum. Och det försiggår en ständig övergång av citatord till främmande ord och av främmande ord till "hemtama" ord.

Från historisk eller etymologisk synpunkt utgörs de främmande orden av lånord, medan de "hemtama" orden utgörs av dels arvord, dels lånord. Tar vi med också dessa fakta i skalan, får den följande utseende:

kodbyte	interferens	integration
citatord	främmande ord	"hemtama" ord
	lånord	arvord
		lånord

Lånorden kan vara av flera olika slag, dvs. interferensen kan gälla respektive ha gällt olika delar av de berörda språkliga systemen. En ordinär typ kan exemplifieras med ordet *jobb*, som är tämligen neutralt med hänsyn till relationen tal-skrift. I ett fall som *nylon* har skriften dominerat. Somliga skulle kalla det ett synlån. I ett fall som *tejp* har däremot uttalet dominerat. Det blir då ett hörlån med en motsvarande term. I vissa fall lånas en betydelse in och kombineras med en nyskapad inhemsk uttrycksform såsom i fallen *ångare* efter engelska *steamer* och *formgivare* efter engelska *designer*. I sådana fall talar man om översättningslån. Med betydelselån slutligen avser man sådana fall, där en betydelse lånas in och fogas till ett befintligt ord. Man vidgar således ordets polysemi. Ett exempel är betydelsen 'kärleksaffär' som på senare år har lånats in och kombinerats med ordet *romans*. I den nya betydelsen är det tidigast belagt 1956 i en kontext där prinsessan Margaret figurerar. Om vi lämnar vokabu-

lärsystemet och betraktar den språkliga interferensen i stort, kan vi konstatera att vi har en grupp termer som ofta är användbara, nämligen anglicism och dess motsvarigheter.

Totalt sett får engelskans inflytande på svenskan under fornspråklig tid sägas vara ganska ringa. Det tycks emellertid snarast vara ännu mindre under den nysvenska periodens båda första århundraden. En viss kulturell och politisk kontakt fanns naturligtvis. Erik den XIV:s Elisabethanska planer hör dit. När Cromwells ambassadör Whitelocke i mitten på 1600-talet besökte Uppsala uppvaktades han av studenter med verser på engelska, såsom Tönnes Kleberg har påpekat i uppsatsen Engelskspråkig diktning i stormaktstidens Sverige, publicerad 1942. I 1600-talets gratulationspoesi fanns också ett engelskt inslag. Kleberg erinrar om att även Lasse Lucidor har producerat sig på engelska i ett par sammanhang. Johan Ekeblad, som var en mycket språkkunnig och internationellt synnerligen välorienterad person och som dessutom besökte London, stod inte främmande för engelskan. Men några nämnvärda spår i hans språk har den så vitt jag kunnat se inte satt.

Under äldre nysvensk tid upptas över huvud taget bara enstaka engelska ord i den svenska vokabulären. Fastän det endast tycks vara ett drygt dussintal som man med någorlunda god säkerhet kan få fram med hjälp av SAOB:s etymologiska parenteser, så kan man emellertid redan nu urskilja vissa linjer. På sjöfartsområdet finner vi *bäring*, *logg* och *paketbåt*, dvs. ett slags passagerar-, post- och godsartyg. Inom textilbranschen finner vi *kambrik*, ett slags linneväv, och *schagg*, ett slags ylletyg. På det kulinariska området finner vi *pudding* och drycken *rom*. Hit får vi väl dessutom föra *potatis*. Också i det här skedet har vi ett religiöst inflytande som omvittnas av orden *kväkare*, *metodist* och *puritan*. Ordet *bill* 'lagförslag' belagt 1712 tror jag man skall betrakta som ett exempel på kodskifte. Också det antyder emellertid en linje. Ytterligare en linje antyds av det ord som tycks vara tidigast belagt av dem som dyker upp under äldre nysvensk tid, nämligen *dogg* 1638. Ett betydelselån föreligger i fallet *pressa* 'tvångsvis värva' som är belagt i denna betydelse 1632 men som förekommer redan i fornsvenskan. Läger vi härtill *lanter* 'prat,

dumheter; ett slags kortspel', efter engelska *lanterloo* – ordet används f.ö. senare av Bellman – tror jag vi har fått med det mesta.

1600-talet var det århundrade, under vilket det intellektuella Sverige upptäckte England, men det var under frihetstiden som de lärda förbindelserna mellan de bägge länderna nådde full intensitet, skriver Sven Rydberg i sin avhandling Svenska studieresor till England under frihetstiden (1951). Bland englandsresenärer nämner han särskilt diplomater, ekonomer och bergsmän. Under den här tiden seglar ju också England bokstavligt talat upp som Europas främsta handels- och sjömak. Även den engelska litteraturen slår igenom på allvar med namn som Pope, Swift, Defoe, Fielding osv. När det gäller läsvanorna skall man emellertid komma ihåg att den engelska litteraturen under ganska avsevärd tid i stor utsträckning lästes i fransk eller tysk version. En antydan om detta ger, som Rydberg påpekar, bl.a. en förteckning över Lars Salvius böcker. Under alla förhållanden stegras det engelska inflytandet efter hand som den yngre nysvenska perioden fortskrider, och det understöds som bekant under periodens senare del av ett alltmer ökat amerikanskt inflytande.

De synnerligen omfattande influenserna på det svenska vokabulärsystemet under yngre nysvensk tid kan jag här inte annat än antyda. Jag gör det genom att nämna ett antal ämnesområden och ge ett exempel från varje: sjöfart *pir*, kläder *dress*, matlagning *grill*, politik *budget*, kommunikation *tunnel*, teknik *fläns*, tidningsväsendet *reporter*, affärsvärlden *check*, religion *Frälsningsarmén*, djurvärlden *terrier*, dans *foxtrot*, väder *cyklon*, medicin *penicillin*, naturvetenskap *laser*; därtill beteckningar för hela ämnesområden som *film* och *sport*.

Översättningslånen är inte så få i denna import. Jag nämnde *Frälsningsarmén*. Gösta Bergman ger en lista i skriften Främmande ord i nusvenskan. Bland exemplen kan jag nämna *frilans*, *grodman*, det diskutabla *högbrynt* efter engelska *high-brow*, vidare *järnridån* och *ljudvallen*. Även fraseologiska översättningslån förekommer: *rädda ansiktet* efter *save one's face*, *stjåla föreställningen* efter *steal the show* och *har kommit för att*

*stanna efter has come to stay*. Som exempel på betydelselån kan jag nämna *diskretion* i betydelsen 'omdöme': "Använd er diskretion." Också betydelsen 'korsa' hos ordet *krossa* i uttrycket *krossa Atlanten* kan hänföras hit.

Ett par lustiga fall är *five-o'clock tea* som i England heter *afternoon tea* och *smoking* som i England kallas *dinner-jacket*. I fallet *smoking* har man antagit en förväxling på kontinenten mellan engelskans *smoking-jacket* och *dinner-jacket* (Låftman i Arkiv för nordisk filologi 1942). Jag kunde kanske också nämna några rena amerikanismer. Sådana är *indiansommar* efter *Indian summer* och *mormon* efter *Mormon*.

Inom vissa fackspråk är den s.k. svengelskan mycket utbredd. Det är ju fallet i åtskilliga vetenskapliga jargonger, det är fallet i idrottspråket och i många andra sammanhang. Jag skall här nöja mig med att ge ett enda exempel, radio-TV-folkets vokabulär. Beläggen hämtar jag från en lista i personaltidningen Antennen 1964: *beam* 'elektronstråle i ett kamerarör', *break* 'avbrott', *cue* 'startsignal', *fejka* 'göra något på låtsas', *life-sändning* 'direktsändning', *scanner* 'projektor för omvandling av film till TV-bilder', *streaking* 'typ av bildstörning i form av horisontella streck', *switcha* 'koppla över' och *trickmixer* 'apparat för utförande av speciella bildeffekter'.

Det engelska inflytandet är också betydande inom den del av vokabulären som utgörs av egennamn. Särskilt märkbart är det bland förnamnen. I stort sett kan man säga att detta inflytande sätter in efter 1800. Härvid spelar västkusten med Göteborg en aktiv roll. Roland Otterbjörk nämner (i Svenska förnamn 1964) från detta tidiga skede namn som Charles, John, Henry, William, Betty, Jenny, Mary och Nelly. Efter hand får vi rent av en morfologisk mönstertyp på -y, -ie: Billy, Eddie, Harry, Jonny, Teddy, Annie, Emmy, Billy, Lizzie, Nancy, Peggy, Sally. Kontakten med Nordamerika genom emigrationen blev också av betydelse. Före sekelskiftet har vidare namn som Alfred, Allan, Edvin, Frank, Morgan, Ralf, Alice, Edith, Ellen, Jane, Kate, Lilian, Maude, Marion och Mildred etablerat sig i svenskan. Under och efter första världskriget är namn som Allan, Edvard, Harry, Henry och John vanliga bland Lunds studenter enligt vad

Gudrun Österlund har funnit (Lunds studenters tilltalsnamn under femtio år 1966). Under en ny våg i samband med andra världskriget blev namn som Tom, Bill, Dick, Kennet, Roger, Ronny och Tommy populära. På kvinnosidan var inflytandet under båda vågorna betydligt blygsammare. Gudrun Österlund nämner dock namn som Alice, Harriet, Lilian, Lilly och Lizzie.

Om franska lånord i svenska dialekter har det skrivits men så vitt jag vet inte om engelska lånord. Det är nog också ganska klart att engelskans inflytande på de svenska dialekterna har varit mycket begränsat. Västkusten har kanske varit särskilt utsatt för det inflytande som trots allt har förekommit, och jag skall därför ge några exempel från bohuslänskan och från vad Björseth kallar göteborgsdialekten. Ett spritt ord i bohuslänskan är det som uppträder i former som *manevår* och *manoår*. Det betyder 'krigsfartyg' och återgår på engelskans *man-of-war*. Till sjöterminologin hör också *bräckvater* 'vågbrytare' av engelska *breakwater*, *skalejt* 'takfönster på kajuta' motsvarande engelska *skylight* och *flängjepp* 'försegl' motsvarande engelskans *flying jib*. Från göteborgsdialekten uppger Björseth i Göteborgsspråket 1958 verbaluttrycket *all rejta* 'klara av' till engelskans *all right*, *jumpa i* 'bli våt' till engelskans *jump* och *tjomme* som väl bör sammanföras med engelskans *chum*. Självt har jag upptecknat *tejta tell* 'dra åt (om knop)', som jag vill sammanställa med engelskans *tighten*. *Spid* 'fart' har varit och är vanligt. Ett nytt och som man får anta tillfälligt synlån som har avlyssnats på Tjörn är *basón-fläsk* som avser *bacon*.

Vissa ord i bohuslänskan skall säkerligen förklaras med inflytande från amerikansvenskar, särskilt kanske hemvändande emigranter. För *sutt* 'beklädnad' antar Nilén i Ordbok öfver allmogemålet i Sörbygden 1879 sådant inflytande, jfr *suite*. Alldeles uppenbart förefaller det mig vara i ett av Nilén inte anförda fall som *amerkatronk* 'amerikaväska', jfr *trunk*. Om emigrant-svenskan fördes en diskussion i Svenska Dagbladet 1960 mellan Einar Haugen och Vilhelm Moberg. Utgångspunkten var Mobergs behandling av amerikansvenskan i det stora utvandrar-eposet. Einar Haugen gjorde gällande, på grundval av resultaten i sin undersökning *Bilingualism in the Americas* 1956, att

nybyggarspråket i romanserien var ett sammelsurium av orimligheter hopade på varandra utan hänsyn till daglig praxis bland utvandrade svensk-amerikaner. Huvudinvändningarna gällde att Moberg i alltför hög grad hade brutit mot två regler för låntagning från ett språk till ett annat, nämligen att strukturord sällan eller aldrig lånas och att främmande ord inte används när ett inhemskt ord som liknar det fonetiskt står till buds. Alltså inte *how många* utan *hur många*, inte *dä har dröjt så long* utan *dä har dröjt så länge*. Moberg svarade att han i allt väsentligt hade byggt på en dagbok från åren 1854 till 1898 av en viss Andrew Peterson. Han rekommenderade Haugen att studera dagboken, som finns i St. Paul, Minnesota. Einar Haugen tog rådet ad notam och for till St. Paul. I en ny artikel skrev han att svaret på frågan om det i Andrew Petersons dagbok finns någon täckning för Mobergs språkbruk måste bli ett kategoriskt nej. Andrew Peterson visade sig följa samma praxis som Haugen hade funnit för norsk-amerikanskans vidkommande. Vilhelm Moberg blev inte svarslös, men man kan säga att sakk Diskussionen var avslutad i och med detta. Jag vill gärna tillägga att Haugen inte diskuterar det estetiska intrycket av Mobergs amerikansvenska.

Jag har uppehållit mig något vid den här diskussionen därför att jag menar att den ger en god belysning av några viktiga språkvetenskapliga problem som knyter sig till emigrantsvenskan. En mycket positiv händelse på området under senare år är de inspelningsresor som Folke Hedblom har företagit i de amerikanska svenskbygderna. Preliminära rapporter om resultatet har han givit i uppsatsform i Svenska landsmål.

Efter att ha uppehållit mig vid det engelska inflytandet på vokabulären under nysvensk tid övergår jag nu till att ta upp några andra sidor av interferensen mellan de båda språkliga systemen under perioden.

Först går jag då in på uttryckslärans område. Termen *offside* hade på fotbollsplanerna under min uppväxttid uttalet *offsejd*. Man hörde också *ållrejt* eller *hållerejt* för *all right*. Och det heter ju allmänt *streck*. Problemet kan belysas lite närmare med hjälp av det intressanta ordet *guide*. Detta lånades in från franskan alldeles i början av 1800-talet i betydelsen 'vägvisare' och hade

då uppenbarligen det franska uttalet [gid']. På 1870-talet lånades det in på nytt, nu från engelskan och i betydelsen 'styrlist'. Det skrevs i denna användning redan från början *gejd* alternerande med *guid(e)*. SAOB tryckte artikeln *gid* 1929 och uppgav då endast uttalen [gi'd] och [gid']. I dag torde normaluttalet då 'vägvisare' avses vara [gajd], även om Wessén i Våra ord ger både [gaj'd] och [gäj'd]. Det franska lånordet, som hela tiden åtminstone alternativt har bibehållit originalstavningen, har alltså på senare tid associerats med engelskans *guide*. Mot varandra står då här 1870-talets stavning och uppenbarligen uttal med *ej*-diftong och 1960-talets uttal med *aj*-diftong. Mot varandra står på samma sätt 1870-talets fackliga term *strejk* och 1960-talets bowlingterm *strike*.

Underlaget till en förklaring av de här förhållandena bör vara att söka i engelskans fonologiska system. Diftongen [ai] återgår på långt *i* via en centraliserad diftong [ə] enligt Dobsons m.fl. mening. Nu kan man utläsa ur äldre 1900-talsarbeten av Ekwall och Daniel Jones dels att diftongens första komponent hade en allofonisk variation mellan olika *a*-artikulationer och ett mycket öppet *æ*, dels att många, särskilt norrut i England, använde den centraliserade varianten [əi]. Också i ett modernt arbete som A.C. Gimson's *An introduction to the pronunciation of English* 1962 nämns likväl varianterna med [æ] och [ə]. Mot denna bakgrund förefaller det som om man kan tänka sig flera olika möjligheter till förklaring av former som *offsejd*, *gejd* och *strejk*. En första komponent som var tillräckligt *ä*-haltig för att diftongen skulle associeras med svenskt *ej* kan tidigare ha varit vanligare i engelskt uttal. Långgivande befolkningsgrupper kan ha haft uttalen [æi] eller [əi]. Substitution med *ej* kan ha känts naturlig i svenskan, som har denna diftong i vissa högfrekventa inhemska ord: *mej*, *dej*, *sej*, *säja*. I vissa trakter kan en tendens till hyperkorrektion ha spelat in. Dagens former *gajd*, *strajk*, *tajma* osv. får å andra sidan en väsentlig del av sin förklaring genom skolans standardiserande inverkan. Vi skall emellertid också vänta oss vissa tidiga fall av *aj* och vissa sena fall av *ej*. Av det förra slaget är *paj*, belagt 1883, och av det senare slaget exempelvis det uttalsalternativ [gäj'd] 'vägvisare' som Wessén ger i Våra ord.



Också när det gäller den engelska diftongen [ei] kan man göra en del intressanta iakttagelser. Det heter å ena sidan *räd*, *räls* och *jed* eller *jäd* – mineralet *jade* – å andra sidan *tejp* och *spray*. Av dessa ord är *jäd* belagt redan 1788, *räls* 1836 och *räd* 1889, medan ju *tejp* och *spray* är rena nykomlingar. I det här fallet ser vi antagligen en reflex av engelsk ljudutveckling. Diftongen [ei] är sista stadiet i utvecklingen från medelengelskt [a] via [æ] och [ɛ]. Och detta stadium är sent. Man brukar anse att diftongen föreligger omkring 1800 men naturligtvis inte samtidigt och likformigt över hela språkområdet.

Ett par andra exempel på fonologisk interferens kan också för tjäna att nämnas. Det ena är *shop* som har lånats in två gånger och som första gången togs upp i formen *schapp* och andra gången nu i formen *shop*. Det andra exemplet, och där är både vokalismen och konsonantismen inbegripna, är vårt *hurts* som i början av 1900-talet inlånades från engelskans *hutch*. Här har alltså våra apikoalveolarer fått återge den engelska affrikatan.

Innan jag lämnar fonemsystemen vill jag nämna den svårighet som engelskans [æ] för närvarande vållar i svenskt tal. Ulf Teleman har i en uppsats (i Modersmålsläraarnas förenings årskrift 1965) pekat på att vi till och med har minimala par som visar opposition mellan å ena sidan /æ/ och /ä/, å andra sidan /æ/ och /a/: *slacks* – *släcks*, *bagar* – *baggar*. Hittills har vi normalt haft ljudsubstitution med /a/ såsom fallen *back*, *banta*, *handikapp*, *jazz*, *match* och *Harry* visar. *Pentry* står väl snarast som undantaget som bekräftar regeln.

Inflytandet på uttryckslärans andra område, skriften, har jag indirekt delvis berört. Jag skall ta upp ett par punkter till. Vi har en väletablerad tradition att skriva lånord av typen *jobb*, *klubb*, *logg* med dubbelteckning i slutställningen. Denna grundregel i det svenska systemet har på sistone börjat brytas i fall som *jet* och *pop*. Vidare har engelskans *sh* som beteckning för *sj*-ljud under senare decennier vunnit allt större spridning. Detta reflekteras i sista upplagan (1950) av Svenska Akademiens ordlista, som alternativt anger t.ex. *sherry* med engelsk stavning. På denna punkt har alltså vårt skriftsystem tillförts en ny grafotaktisk kombination.

Det mest slående interferensfenomenet på det morfologiska området är behandlingen av den engelska pluraländelsen *-s*. Förvecklingarna kan belysas med hjälp av ordet *potatis*. Det dyker upp under de allra sista åren av den äldre nysvenska perioden och har då den engelska formen *potatoes*. *S*-et biter sig fast, och Dalin ger i sin ordbok från 1800-talets mitt singularformerna *potatoes* och *potatis*, pluralformerna *potäter* och *potatisar*, den senare med inskränkningen ”dock ej i kollektiv mening”. De båda pluralformerna representerar två tendenser som man också fortsättningsvis kan urskilja. I formen *potäter* har den främmande ändelsen utbytt mot en inhemsk, i formen *potatisar* har en svensk ändelse tillagts efter den främmande. I anslutning till pluralformen *potäter* bildas naturligtvis lätt singularformer som de i dialekterna spridda *potäta*, *potät* m.m. Typen *potatis*, s.k. *keps*-ord, har vi speciellt i sådana fall där kollektiv betydelse är vanlig, t.ex. i *shorts* och *koks*. Typen *potatisar* har vi i fall som *jumpersar* och *bebisar*, som numera med som jag tycker ganska god framgång motarbetas. Typen *potäter* har vi i fall som *inter-vjuer* och *ulstrar*. Pluralis på *-ar* av engelska lånord på *-er*, alltså just typen *ulstrar*, förefaller mig vinna spridning. Radiofolket säger t.ex. *scannrar*. En del av de ord som hör hit vållar emellertid svårigheter, t.ex. *partner*, som inte gärna kan heta *partnrar* i pluralis. Här sker då i stället anslutning till typen *musiker*, pluralis *musiker*. Även om sålunda svenskan har goda möjligheter att hävda sin struktur på den här punkten, kan man naturligtvis få se hybrida fall. Jag har t.ex. ett belägg på pluralformen *bambas* av det nu så spridda ordet *bamba* ’barnbespisning’. Kanske har skribenten tvekat om huruvida det skulle heta flera *bamban* eller *bambor* och då gripit till formen *bambas*.

En annan morfologisk företeelse som man ser gripa omkring sig och som i varje fall till stor del torde bero på engelsk-amerikanskt inflytande är den analytiska firma- och varunamnstypen. *Corn Flakes* har vi kunnat äta länge. Nu kan vi också äta *Fras Kuddar*, *Honungs Puffar* eller *Chokla' Puffar*. Allmänt sett vill jag emellertid understryka att det svenska ordbildningssystemet har visat sig mycket motståndskraftigt. *Sporty* blir *sportig*, *indi-*

*cate* blir *indikera*, *electron* (med betoningen på andra stavelsen) blir *elektron* och *-minded* blir *-sinnad* som i *fotosinnad*.

På det syntaktiska området tycks i varje fall det omedelbart skönjbara engelska inflytandet vara av synnerligen liten omfattning också i modern tid. En belysande jämförelse kan göras med Carstensen's monografi över det engelska inflytandet på tyskan, där avsnittet om syntax på samma gång är ytterst kortfattat och innehåller en rad ganska lösa förmodanden. Översättningslån av typen *har kommit för att stanna*, *gå in för* m.fl. kan naturligtvis nämnas här. Likaså förbindelser som *ta över* i stället för *överta* efter *take over* och *följa upp* efter *follow up*. Vidare väl också konstruktioner som *flyga SAS* i stället för *flyga med SAS*. Erik Wellander anför i Riktig svenska ett språkprov som visar att ett engelskt syntaktiskt mönster har kalkerats: Denna fråga tål att tänka på, och så gör skolfrågan. Jag tror emellertid inte att man kan se den konstruktionstypen ofta. Oftare ser man då den lätt löjeväckande typen *namn plus ålder*: Anders Johansson, 32.

I emigrantsituationen är influensen däremot påtaglig. Jag citerar från Hedblom:

Di från Järbo, di trodde att di visste hur te tala och vi visste inte hur; deras var rätt och vårt var inte rätt.

Också i översättarens situation kan det syntaktiska inflytandet bli ganska kraftigt. Autentiska och avskräckande exempel på vad som kan förekomma i sämre översättningar av t.ex. detektivromaner, serier och bruksanvisningar ger Börje Söderlund i sin bok Att översätta 1965. Jag citerar ett stycke av en bruksanvisning för " 'Splosh' vatten pistolen":

Först skjut pistongen [på] baksidan av bösspipan helt hem. Sedan doppa mynningen i vattnet och långsamt drag pistongen tillbaka till dess fulla utsträckning. Pistolen är numera lastad. Att avfira pistolen krama avtryckaren i den vanliga vägen.

Engelskan är som bekant inte det enda främmande språk som har utövat ett väsentligt inflytande på svenskan. Det är framför allt två andra språk som kommer i fråga, tyskan och franskan. Det tyska inflytandet var som störst under yngre fornsvensk tid

genom lågtyskan och under äldre nysvensk tid genom högtyskan. Det franska inflytandet var som störst under senare delen av äldre nysvenskan och förra delen av yngre nysvenskan. Det ligger nära till hands att försöka jämföra omfattningen av inflytandena från de tre språken. Som första kriterium skulle jag därvid vilja uppställa influensen på strukturen, dvs. på tal- och skriftsystemen, ordböjningen, ordbildningen, syntaxen. Som andra kriterium skulle jag vilja uppställa influensen på ordförrådet. Som ett tredje kriterium kan man tänka sig graden av infiltration i dialekterna. Vad nu först tyskan beträffar, så har den starkt påverkat den svenska språkstrukturen. Vi har fått en rad ordbildningselement som i *grevinna*, *studentska*, *trohet* osv. Vi har assimilerat en mängd formord eller grammatiska ord sådana som *bli*, *måste*, *men*, *dock*, *sådan*, *hurudan*, *ganska*, *strax*, *redan*. Blandningen mellan tyskt och svenskt under fornsvenskans senare skede var säkerligen också en väsentlig orsak till det fornsvenska fyrkasussystemets upplösning. Att jämföra är vad som skedde med fornengelskan efter den normandiska erövringen. Också franskan har utövat ett betydande strukturellt inflytande framförallt på ordbildningen: *montera*, *montör*, *bilist*, *realism* osv. I jämförelse härmed framstår engelskans rent strukturella inflytande såsom föga märkbart. När det så gäller vokabulären har det påståtts att de tyska lånorden i svenskan är fler än arvorden. Under alla omständigheter är kvantiteten mycket stor. Också den franskimporterade delen av vår vokabulär är försvarlig, och det kan vi numera säga också om den engelskimporterade. Vad infiltrationen i dialekterna beträffar går det tyska djupast, och därefter kommer det franska. Tillsammans taget synes mig detta innebära att rangordningen mellan de hittillsvarande inflytandena från de tre huvudkonkurrenterna är densamma som tidsföljden: tyskt, franskt, engelskt.

Till detta vill jag allra sist föga att det engelska inflytandet ju ännu står i fullt flor och att vi inte vet vart det kan leda. Å andra sidan har vi en allt aktivare språkvård. Och om denna först och främst inriktas på vården av svenskans struktur i den mening som jag här har angivit – vilket jag ser som språkvårdens huvud-

uppgift – bör vi ha goda utsikter att till nästa generation överbringa en språkform som inte förtjänar att kallas riks-Swedish.

## Litteratur

- Bergsten, Nils, Om engelska lånord i svenskan. 1915.
- Björseth, Bertil, Göteborgsspråket. 1958.
- Carstensen, Broder, Englische Einflüsse auf die deutsche Sprache nach 1945. 1965.
- Dahlstedt, Karl-Hampus, Bergman, Gösta, & Ståhle, Carl Ivar, Främmande ord i nusvenskan. 1962.
- Gimson, A.C., An introduction to the pronunciation of English. 1962.
- Haugen, Einar, Bilingualism in the Americas. 1956.
- Haugen, Einar, The Norwegian language in America. A study in bilingual behavior. 1–2. 1953.
- Hedblom, Folke, Om svenska folkmål i Amerika. 1962.
- Hedblom, Folke, Bandinspelningsexpeditionen till Svensk-Amerika 1964. 1965.
- Hedblom, Folke, Den tredje inspelningsexpeditionen till Svensk-Amerika. En rapport från resan 1966. 1966.
- Hellquist, Elof, Det svenska ordförrådets ålder och ursprung. En översikt. 2. 1930.
- Holmkvist, Erik, Mera om engelska lånord i svenska yrkesspråk. 1951.
- Jansson, Sven B.F., Runinskrifter i Sverige. 1963.
- Kleberg, Tønnes, Engelskspråkig diktning i stormaktstidens Sverige. 1942.
- Nilén, Nils Fr., Ordbok öfver allmogemålet i Sörbygdén. 1879.
- Otterbjörk, Roland, Svenska förnamn. Kortfattat namnlexikon. 1964.
- Rydberg, Sven, Svenska studieresor till England under frihetstiden. 1951.
- Stene, Aasta, English loan-words in modern Norwegian. A study of linguistic borrowing in the process. 1946.

- Ståhle, Carl Ivar, Syntaktiska och stilistiska studier i fornnordiskt lagspråk. 1958.
- Söderlund, Börje, Att översätta. 1965.
- Teleman, Ulf, Språkvårdens argument. 1965.
- Thors, Carl-Eric, Den kristna terminologien i fornsvenskan. 1957.
- Weinreich, Uriel, Languages in contact. Findings and problems. 1953.
- Zandvoort, R.W., English in the Netherlands. A study in linguistic infiltration. 1964.
- Österlund, Gudrun, Lunds studenters tilltalsnamn under femtio år. 1966.

# Om nominalkomponerade verb i nordiska språk

Språkdata 1968.

## 1. Inledning

Under de senaste decennierna har vad man kan kalla nominalkomponerade verb, typen *landsätta*, tilldragit sig ganska stor uppmärksamhet. En orsak härtill är säkerligen att verbtypen har visat en markant ökad frekvens under denna tid. Som en annan orsak framstår den omständigheten att verbtypen är av betydande intresse från ordbildningssynpunkt. I föreliggande studie ges först en forskningsöversikt, därefter diskuteras verbens bildning och slutligen behandlas deras förekomst.

Allmänt bestämmer jag de nominalkomponerade verben såsom verb med minst en nominal (substantivisk eller adjektivisk) förled. De flesta verb som uppfyller kraven kan hänföras till någon av följande fyra morfologiska typer (av förkortningarna står N för nominal led, V för verbal led och P för partikel):

- |     |               |           |
|-----|---------------|-----------|
| (1) | landsätta     | N-V-a     |
| (2) | landskampa    | N-(s)-N-a |
| (3) | luftlandsätta | N-N-V-a   |
| (4) | ilandföra     | P-N-V-a   |

En partikel kan också stå omedelbart före verbmorfemet såsom i *landsförvisa*, *läkarundersöka*, da. *kobberforhude* och da. *tvangsindlægge*. I no. *avtaleforplikte* (*seg*) (Nordiska språkfrågor 1955, s. 31) finner vi partiklar på två ställen. Som exempel på verb med adjektiviska morfem som förleder kan anföras *allmänbilda*, sv. dial. *barstjälja* 'frånstjälja någon allt vad han äger', *bredsida* (fotbollsterm), nisl. *blödlygna* 'bli stilla', *djupfrysa*, no. *frifinne* 'frikänna', *friställa*, *fritidshugga*, nisl. *lútilvirða* 'ringakta', *möjliggöra*, *pånyttföda*, sv. dial. *röleta* 'rödfärga', nisl. *skammstafa* 'förkorta', *simultantolka*, *snabbutbilda*, *torrförvälla*, da. *totalfred*, *tvärvetenskapa* och no. *virkeligjøre*.

Materialet till min undersökning är av storleksordningen 2 500 verb. Den ena huvuddelen av detta är hämtad ur riksspråkliga och dialektala ordböcker och ordlistor från det nordiska området jämte arkivmaterial. Den andra huvuddelen har jag samlat in under 1950- och 1960-talen ur i första hand moderna svenska texter.

## 2. Forskningsöversikt

Johan Rydqvist behandlar i korthet ordtypen i femte bandet 1874 av Svenska Språkets Lagar. Rydqvist utgår från Jacob Grimms uppfattning, enligt vilken ingen sammansättning av substantiv med verb utöver den i infinitiv och particip erkänns för de germanska språkens vidkommande. De svaga finita verbformerna betraktas som avledda av nominalsammansättningar och de starka som oorganiska. Rydqvist hävdar att huvudtanken hos Grimm inte synes böra bestridas, fastän ett till grund för verbet liggande nomen i flertalet fall som han uttrycker det får sökas i teorin och fastän såväl starka som svaga finita former föreligger (s. 63 f.). Bland starka former i fornsvenskan nämner han *hals-hiog* och *laghböþ*. Han anför också att friheten att använda finita former nu är utvidgad och överförd till flera ord. Som exempel på verb med stark böjning anför han *deltaga*, *hemfalla*, *munhuggas*, *köpslå* och *rådslå*, som exempel på verb med svag böjning *stadfästa*, *korsfästa*, *hopföra*, *härbärga* och *nödsakas*.

I sin avhandling Om de sammansatta verben i isländskan 1877 arbetar också Hans Hainer i Grimms anda. Han gör emellertid ett viktigt påpekande, nämligen att en stor del av de nominalkomponerade verben torde ha utgått från participformer, särskilt preteritum particip (s. 23). Som exempel ger han bl.a. *jarðsetja* av *jarðsettr*. Elva år senare publicerade han en kort uppsats med titeln Om de sammansatta verben i nysvenskan. Denna kan snarast betecknas som utdrag ur en ordnad materialsamling, hämtad från Svenska Akademiens ordlista och Dalins Ordbok öfver svenska språket. Hainer ger dock vissa antalsuppgifter och



meddelar några korta iakttagelser om förledens relation till det enkla verbet.

Hos Fredrik Tamm i *Sammansatta ord i nutida svenskan 1900* möter vi för första gången belägg hämtade från tidningar. Av hans fall kan nämnas *tvångsvisitera* (GHT) och *grundlagsbestämma* (UNT). Tamm betecknar dessa som "mera tillfälligt bildade sekundära verb" och hänvisar till substantiven *tvångsvisitation* respektive *grundlagsbestämmelse*.

Samma år som Tamms bok utgavs kom *Dansk-Norskens syntax i historisk fremstilling* av Hjalmar Falk och Alf Torp. Här hävdas att primära komposita, bildade utan förmedling av nomina, i fornvästnordiskan dominerat över avledda komposita (s. 358 ff.). Exempel skulle vara *knésetja*, *krossfesta*, *trúlofa* m.fl.

Olof Östergren skrev 1905 en uppsats i *Nordisk Tidskrift* med titeln *Språklig nyskapelse*. Som exempel på nybildningar ger han bl.a. vad han betecknar som de modernt klingande främlingsorden *generalrepetera*, *röntgenfotografera* och *principdebattera* (s. 588). Han anför också *bildhugga*, *jordbäva*, *liktala*, *krypskjuta* och *träsnida*. Hos Östergren möter vi några av de iakttagelser som kommer att spela en viktig roll i den följande diskussionen: dels uppgiften att verben börjar bli ganska vanliga, dels påpekandet att de å ena sidan ansluter sig till förut befintliga enkla verb och å andra sidan till sammansatta substantiv.

Gustav Cederschiöld har i boken *Fresta duger 1914* ett kapitel som heter *För riksspråket nya verb, som bildats genom avledning* (s. 69 f.). Han ger där några hithörande belägg, uppdelade efter betydelse. Bland dessa kan nämnas *solstråla*, *mordbränna*, *fondmäkla* och *skuggspela*. I *Framtidssvenska* (s. 75 f.) ger han på likartat sätt bl.a. exemplen *blomsterkasta* (Selma Lagerlöf), *dagteckna* (Viktor Rydberg) och *tidsindela*. Om det sista av verben säger han att det väl är bildat efter *tidsindelning*.

Det i principiellt avseende viktigaste av de tidigare bidragen är Erik Wellanders uppsats *Ett par produktiva typer av skriftspråklig nybildning 1915*. Wellander utgår från substantivbildningar sådana som *maskinskrivning*, *postbefordran* och *telegrafsvär*. En bildning av detta slag utgör en som Wellander uttrycker det "sluten enhet, såväl i semologiskt som i morfologiskt avseende,

d.v.s. den uttrycker på ett enhetligt sätt en viss given totalföreställning” (s. 45). När denna totalföreställning på nytt behöver uttryckas, visar det sig ofta lämpligt att bilda ett verb av substantivet. Man erhåller så *telegrafsvara*, *skridskosegla*, *postbefordra* osv.

Wellander tvekar att kalla dessa verbbildningar retrograda. För språkkänslan ingår verbalnomina i verbets böjningsschema, och all avledning sker med verbalstammen som utgångspunkt, menar han. Dock reserverar han sig beträffande verb som *kroppsvisitera*, *vådaskjuta* och *knivhugga*. Dessa fall finner han det mera berättigat att beteckna som retrograda, därför att han inte känner verbalnomina *visitation*, *skott* och *hugg* såsom avledda av respektive verbalstammar. Han fortsätter: ”Här sker dock nybildningen i själva verket på samma sätt, analogiskt, ehuru det ifrågavarande enkla verbet ensamt avger mönstret” (s. 47). Såvitt jag förstår menar han med detta att inte heller *kroppsvisitera* bör karakteriseras som en retrogradering.

I fortsättningen nämner Wellander att inte bara nomina actionis har givit upphov till verb av den aktuella typen. Även nomina agentis och participen har spelat sin roll. Av exemplen kan anföras *kannstöpa* efter *kannstöpare* respektive *skräckslå* efter *skräckslagen*. Han nämner också att åtskilliga av verben är mer eller mindre gamla i svenskan, t.ex. *korsfästa*, *våldtaga*, *hushålla*, men att den stora massan av verb är ny, t.ex. *brandförsäkra*, *taxeringsvärdera*, *blomsterhölja*. Han framhåller vidare förenklingen i verbens konstruktionssätt. Liksom vid substantivens bildning utelämnas den preposition som uppehåller den syntaktiska förbindelsen i motsvarande fras: *läsa korrektur på en avhandling* men *korrekturläsa en avhandling*. Slutligen påpekar han att denna ”denominativa verbbildning” är originell så tillvida som den inte beror på eller stöts av tyskt mönster (s. 50).

Adolf Noreen har i sjunde bandet 1919 av Vårt språk ett mindre kapitel om substantivs sammansättning med verb (s. 469 ff.). Kapitlet är nästan uteslutande en materialsamling, ordnad efter formen hos verbens förleder. Inledningsvis skriver Noreen att nybildningarna på sistone har blivit allt vanligare och att de oftast utgår från redan befintliga sammansättningar. Hän-

visning sker till Wellanders uppsats. Noreen har också en motsvarande paragraf om adjektivs sammansättning med verb (s. 485 ff.).

Övermåttan vanliga i modern danska är ord som *sygemælde*, *ligbrændes*, *statsunderstøttes*, *ovnlakere*, *prisedsætte*, *støvsuge* och *ferieforelske sig*, skriver Kristian Sandfeld i Sprogvidenskaben (andra upplagan 1923, s. 45). Han ser dessa som resultat av att förhållandet mellan stam och ändelse i de bakomliggande sammansättningarna *sygemælding*, *ligbrænding*, *statsunderstøttelse* osv. som han uttrycker det har misstolkats genom analogiskt inflytande från motsvarande enkla substantiv och verb.

Till Wellander anknyter vidare Erik Noreen i uppsatsen Bidrag till fornsvensk lexikografi 1941 (MASO 3). Verben behandlas här under rubriken Fsv. *hjärtstinga* och *rantaka*. När det gäller verbtypens bildning menar han att nomina actionis är den viktigaste utgångspunkten i modern svenska, medan preteritum particip torde vara viktigast för äldre språkperioder: *korsfæsta* av *korsfæster*, *rantaka* av *rantakin* (s. 16). Erik Noreen återkom till verbtypen i en kort uppsats med rubriken Fläckurtag Eder kostym! i boken Svensk stilparodi 1944 (s. 137 ff.). I denna uppsats betonar han att det är en strävan efter koncentration som har gett upphov till verb av typen *fläckurtaga*.

I sin bok Objekt och adverbial 1942 kommer K.G. Ljunggren i ett kort kapitel in på syntagmer av typen *skriva maskin*, *predika prov*, *sitta kväll* (s. 177 ff.). Detta ger honom anledning att något beröra bildningen av verb som *provpredika*. Han finner det sannolikt att både *provpredika* och *predika prov* är bildade till *provpredikan* (som i sin tur återger tyska *probepredigt*). Det kan också nämnas att han i förbigående noterar att verbet *kvällsitta* är belagt i dialekterna.

I andra delen av Elias Wesséns Svensk språkhistoria 1943 och senare år används inte termen retrogradering vid beskrivningen av de aktuella verben. Författaren framhåller emellertid att de sammansatta verben med nominal förled har uppkommit sekundärt i anslutning till sammansatta nominala bildningar. Bland exemplen förekommer såväl verb som bildats till verbalsubstantiv av olika slag som verb som bildats till particip: *maskinskriva*,

*kannstöpa, korsfästa*. Till verb sammansatta med nominal förled hör egentligen inte verb som *hushålla, lovorda, landsmåla*, skriver Wessén. Dessa verb betecknas som avledningar till substantiven *hushåll, lovord* respektive *landsmål*. Från dialekterna anför han ett verb, nämligen *buföra (buffra)* 'flytta med korna till fäbodarna' (4. uppl. 1965, s. 108 not 3). Fjärde upplagan innehåller inga ändringar i sak utan framför allt en utförligare exemplifiering.

Också i Peter Skautrups fyrbandiga verk *Det danske sprogs historie* (1944–1968) omnämns verbtypen. I första bandet upptas några exempel under rubriken *Sammensætninger* (s. 295): *huztrykæ, laghbiuthæ, tacsættæ* 'sätta i borgen' och ett par till. Detsamma är fallet i andra bandet (s. 95): *skrinlægge, torvføre, hådrage, mensværge, klargøre, mangfoldiggøre* m.fl. På s. 241 f. och 386 anføres emellertid att particip av typen *bundfrossen* i vid utsträckning har tjänat som grundlag för "subtraktionsdannelser" som *bundfryse, stenlægge, bønghøre, bagbinde*. I tredje bandet konstateras att typen blir vanligare efter 1700 (s. 350) och i fjärde bandet att nybildning nu sker mycket fritt, i synnerhet i journalistspråk (s. 244, 267 f.). I det senare bandet anges också att de allra flesta bildningarna är "tillbagedannelser". Bland de moderna exemplen är *begrepsbestemme, fjernsynssende, helårsbeskæftige, kortslutte, toneillustrere, værdisikre*.

Att döma av det korta avsnittet om verben i *Elementær dansk Grammatik* (1. uppl. 1946, 2. uppl. 1957) betraktar Paul Diderichsen verben genomgående som avledningar. "Samdannede Verber, afledt af Verbaladjektiver og Verbalsubstantiver" är hans beteckning (2. uppl., s. 239). Att de är "Samdannelser" innebär att de är avledda av mer än en rot (s. 78). Av exemplen kan nämnas *irettesætte, pantsætte, dødsdømme, renskrive* och *koldsvede*. Någon plats för primära nominalkomponerade verbsammansättningar finns det inte i Diderichsens beskrivning, såvida man inte till dem vill räkna den speciella typen *stortude* med adjektiv i stamform som anger grad eller sätt, vilken Diderichsen uppför som en av tre sorters äkta sammansättningar (de båda andra är *adlyde* med partikel och *øsregne* med verb i stamform).

Från sin utgångspunkt kommer också Rudolv Körner in på verbtypen i sin avhandling Studier över syntaktisk nybildning i svenskan 1 1948. Konstruktionstypen *detta värdesättes högt* betraktas av Körner – säkerligen med rätta – som en allvarlig konkurrent till konstruktionstypen *de borde tas vara på*, prepositionell passivbildning (s. 285). Verben kan ju obehindrat användas passivt: *han har måst magpumpas, uppfinningen har patentsökts* osv. Denna iakttagelse har för övrigt gjorts av Erik Wellander redan i den tidigare nämnda uppsatsen från 1915. Han skriver där: ”I satsen *när färgade regementen blomsterkastast af Rysslands allierade* (Sv. D.) torde detta kompositum (dvs. *blomsterkastast*) vara valt för att slippa å ena sidan det tunga och tråkiga *vara föremål för blomsterkastning* och å andra sidan undvika det alltför lediga *kastast blommor på*” (s. 49). Det är som Wellander påpekar alltså inte bara prepositionella passivbildningar utan också konstruktioner av typen substantivsjuka som nominalkomponerade verb verkar återhållande på. Av Körners exempel på hithörande verb kan jag nämna *blodprovas, fartprovas, hjärtundersökas, intelligensprövas, magpumpas, patentsökas, prisregleras, protokollföras, rumphuggas, smakättas, vingklippas*.

I festskriften till Karl Helm 1951, *Erbe der Vergangenheit*, ingår en uppsats av Ture Johannisson med titeln *Über Rückbildung im heutigen Schwedisch* (s. 257 ff.). I anslutning till Walter Henzens handbok *Deutsche Wortbildung* behandlar författaren där tre typer av retrogradering: nomina postverbalia (typen *glid*), substantiv bildade av adjektiv (typen *spänst*) och de här aktuella verben. Vad dessa beträffar sammanfattas forskningsläget och ges ett antal nya exempel, bl.a. *kohandla*.

Den som har skrivit mest om ordtypen är Åke Åkermalm. Hans första bidrag är en uppsats i *Nysvenska studier* 32 (tryckt 1954) med titeln *Om verbet atombomba och liknande bildningar i nutida dagspress*. Uppsatsen föreligger i något omarbetat skick i hans bok *Rubriksvenska* 1965.

När det gäller verbens förekomst skiljer författaren mellan ockasionella och usuella verb, å ena sidan t.ex. *bankettkonversera* och *morgonkaffeväcka*, å andra sidan t.ex. *spurtbesegra*

och *valtala* (1965, s. 61 ff.). Han framhåller också att rubrikerna svarar för mer än 500 av de 1200 verben i hans material. Vidare hänför sig en stor del av beläggen till sportsidorna. Slutligen pekar han på de möjligheter verben ger kåsörerna att vinna individuellt stilistiska effekter: *pigstava*, *bengå*, *ormtjusa*, *sängfösa* (s. 72 ff.).

I fråga om verbens bildning urskiljer Åkermalm fyra typer (s. 76 ff.): denominativa verb, analogibildningar, verkliga komposita och översättningslån. De denominativa verben indelar han i avledningar, t.ex. *sabelhugga* av *sabelhugg*, och retrograderingar, t.ex. *brandförsäkra* av *brandförsäkring*, *fasadklättra* av *fasadklättrare*, *eldhärja* av *eldhärjad*. Bland analogibildningarna urskiljer han typen *rekordskjuta* efter andra verb på *rekord-*, *kramforstala* efter andra verb på *-tala* och *kedjeäta* efter mönster av *kedjeröka*. Med komposita menar han enbart verb som har uppstått genom direkt sammanfogning av ett substantiv och en verbform. Exempel är *minnesturista* och *myrkrypa*.

I sin avhandling Fornnordiska verb med substantivisk förled 1955 utgår Åkermalm från antagandet att studiet av de moderna nybildningarna kan belysa motsvarande verb från äldre språkskeden. Som arbetshypotes uppställer han därför, att den denominativa typen är den normala i fornspråken (s. 38). På ett annat ställe i avhandlingen säger han att vad som brukar kallas Scaligers lag är en lämplig arbetshypotes (s. 12). Denna *regium Scaligeri praeceptum*, underbyggd också av Jacob Grimm, går såsom indirekt framkommit tidigare ut på att skenbart sammansatta verb i själva verket är avledda av nominalkomposita. I ett tredje sammanhang (s. 44) säger han att en lämplig arbetshypotes är att förutsätta ett enhetligt bildningssätt för varje verb.

Undersökningens material utgörs i princip av alla i de nordiska fornspråken belagda verb med substantivisk förled. Det äldsta fornsvenska belägget är från 1222 och gäller verbet *muldsköta* (*mulscöta*) 'överlåta jordegendom genom skötning'. Av de drygt 300 verben bedöms drygt 200 som denominativa, ett 80-tal som analogibildningar, 15 som översättningslån o.d. och 5 som verkliga komposita (s. 217). Av de drygt 200 denominativa verben

hänförs drygt 150 till den från preteritum particip utgående typen.

I uppsatsen Om sammansatta verb i svenskan, publicerad i Nysvenska studier 34 (tr. 1955), ställer Ture Johannisson typen *maskinskriva* i relation till typen *nedkomma* liksom till verb med förleder av annan ordklass. Han visar där att det inte råder någon principiell skillnad mellan typerna *maskinskriva* och *nedkomma* med hänsyn till kompositionens ursprung: bakom båda ligger nominalbildningar (s. 174). Den förra typen har samnordiska anor, den senare tillhör i huvudsak den nysvenska perioden. När det gäller de sammansatta verbens bildning understryker författaren vidare framför allt det olämpliga i att som Åkermalm gör beteckna det starka verbet *sabelhugga* som en avledning av substantivet *sabelhugg* (s. 167). Han pekar också på den roll som det enkla verbet kan ha spelat vid bildningen av ord som *atombomba* och framhåller att dylika ords uppkomst medger en dubbel uppfattning, varvid än det ena, än det andra alternativet kan komma i första rummet (s. 167).

Åke Åkermalm tog upp frågorna om verbens bildning och om terminologin i en kort uppsats med titeln Om termernas avledning och retrograd avledning i Arkiv för nordisk filologi 71 (1956). Uppsatsen ingår i omarbetat skick också i hans bok Rubriksvenska 1965. I uppsatsens första version säger sig Åkermalm förstå betänkligheterna mot att kalla *sabelhugga* för avledning men menar att detta trots allt är den bästa lösningen (s. 67 f.). I den andra versionen däremot säger han att verbets starka eller svaga böjning såvitt han kan se inte borde spela någon som helst roll i diskussionen (s. 102). Det starka *sabelhugga* och det svaga *ordkasta* har identiskt bildnings sätt, återgår på *sabelhugg* respektive *ordkast*, menar han. Det finns därför ingen anledning att skilja dem åt genom olika terminologi. Åkermalm gör vidare gällande att retrogradering också är ett slags avledning, och han jämför därvid med den tyska termen *retrograde Ableitung* (1956, s. 67 f.).

I omedelbar anslutning till Åkermalms uppsats i Arkiv för nordisk filologi tryckte K. G. Ljunggren ett eget helt kort yttrande i frågan med titeln Ytterligare om termerna avledning och

retrograd avledning (s. 69 f.). Han vänder sig där mot Åkermalms huvudkriterium, att en retrogradering skall vara kortare eller i varje fall inte längre än sin utgångspunkt. Han menar att vi också måste ta hänsyn till den typ av verbalabstrakter som kallas korta avledningar. Eftersom då både *maskinskrivning* och *sabelhugg* karakteriseras som avledningar bör, menar han, både *maskinskriva* och *sabelhugga* karakteriseras som retrograderingar.

Ragnhild Söderbergh nämner naturligtvis också ordtypen i sitt arbete *Svensk ordbildning* 1967. Hon ger huvudsakligen exempel på retrograderingar – hon använder också själva termen – och anför *livrädda* av *livräddning*, *vinterbada* av *vinterbadare* och *handmåla* av *handmålad* som typexempel (s. 179 f.).

Som exempel på avledningar ger hon *frimärka* (s. 162), *hushålla* (s. 164) och *lovorda* (s. 165). Av vad som för övrigt intresserar oss i detta sammanhang kan jag nämna, att hon i stället för beteckningen kort avledning föreslår benämningen avledning med nollsuffix (s. 79, 97 f.).

För de nordiska språkens vidkommande skall slutligen nämnas att Elias Wessén 1968 återkommer till typen i boken *Vårt svenska språk*. Man kan i hans framställning där urskilja fyra typer (s. 106 ff.): avledningar (*hushålla*, *lovorda*), bildningar sekundära till sammansatta verbalsubstantiv och particip (*sjösätta*, *dagteckna*), bildningar direkt återgående på finita verbaluttryck (inget exempel) och översättningslån (*hemställa*, *rådfråga*).

I de västgermanska språken är de nominalkomponerade verben mera sparsamt belagda, och de har därför förstäligt nog inte tilldragit sig tillnärmelsevis samma intresse inom motsvarande forskning som inom nordistiken. Några inlägg skall emellertid beröras. Från germanistikens område nämnde jag tidigare Walter Henzens behandling i *Deutsche Wortbildung*. Under rubriken *Rückbildung* ger Henzen exempel som *nussknacken* (av *Nussknacker*), *handlangen*, *notlanden*, *wegleiten* och *herabschneiden* (s. 241).

Från anglistikens område kan jag först nämna Otto Jespersens uppsats *A few backformations* 1936. Jespersen behandlar där retrograderingarna som en underavdelning av vad han kallar *metanalysis* (s. 117), dvs. det förhållandet att ord eller ordgrupper



analyseras annorlunda av en yngre generation än av en äldre (jfr Sandfeldt ovan). Det tidigaste av de exempel han ger är *backbite* från 1300. Av de övriga exemplen kan jag nämna *caretake*, *globe-trot*, *handshake*, *sight-see*, *sunburn*, *type-write* och *word-paint*.

Utan att referera till Jespersens uppsats tar Robert A. Hall, Jr., upp verbtypen i uppsatsen *How we noun-incorporate in English* 1956. Hans intryck är att verbens frekvens har ökat mycket starkt under 1940- och 1950-talen (s. 87), och han ger en lista över ord som han har träffat på i skrift och i tal. Av exemplen kan jag nämna *brain-wash*, det treledade *deep-fat-dry*, *folk-dance* och *loan-translate*. Hall konstaterar att verb av denna typ visserligen ständigt nybildas (han anger t.o.m. tempot: ett var tionde dag, se s. 88 not 11) men att många av dem väcker munterhet eller motstånd. Han urskiljer tre bildningstyper: retrogradering av nomina agentis (*kidnapper*), retrogradering av nomina actionis (*muck-raking*) och konvertering av nominalkomposita (*fingerpaint*).

I *The categories and types of present-day English word-formation* 1960 betecknar Hans Marchand de nominalkomponerade verben såsom "pseudo-compound verbs" (s. 58 ff.). I betraktande av verbens 400-åriga historia, skriver han, verkar det inte som om verkliga komposita (icke utgående från nominalkomposita) skulle uppkomma (s. 64). Marchand delar in verben i två grupper: typen *spotlight* som återgår på vad han kallar en nominal sammansättning och typen *stage-manage* som återgår på vad han kallar en syntetisk sammansättning, i detta fall *stage-manager*. Den senare typen betecknar han som "backformation". Den första typen är den vanligaste. Tidigare exempel är få; ett är *safeguard* från 1494. Fr.o.m. 1800-talet blir produktiviteten emellertid stor, särskilt i amerikansk engelska. Exempel är *blacklist* 1884, *daydream* 1884, *week-end* 1901, *streamline* 1913, *foxtrot* 1921, *lipstick* 1926, *crosscheck* 1951. Ett tidigt exempel på den andra typen är *blindfold* före 1225. Senare exempel är *tonguetie* 1557, *henpeck* 1688, *stallfeed* 1763, *sight-see* 1835, *ghostwrite* 1928, *proofread* 1934, *baby-sit* 1947.

Av forskningsöversikten har det framgått, att en rad värdefulla resultat har lagts fram på det aktuella området. Det har emellertid

också framgått, att det fortfarande råder olika meningar om såväl verbens bildningssätt som de begrepp och termer som skall användas vid beskrivningen. Av dessa skäl och av det skälet att vissa aspekter endast flyktigt eller inte alls har berörts i diskussionen kan det vara befogat att ta upp hela problemkomplexet till behandling.

### 3. Bildning

Nybildning av ord sker i princip efter modell, dvs. efter mönster av större eller mindre grupper av inom vokabulärsystemet etablerade enheter. Vi kan kalla detta analogi. Det bör omedelbart tillfogas dels att flera olika slag av analogi kommer i fråga, dels att dessa kan samverka i växlande kombinationer. Vilka typer som är relevanta när det gäller de nominalkomponerade verben skall jag försöka sammanfatta i ett principresonemang, där jag av teoretiska skäl renodlar fallen.

Först är det då viktigt att slå fast att mönstren föreligger i det synkroniska systemet. Nybildning är resultatet i den diakroniska dimensionen av dettas inneboende dynamik.

Vilka mönster och därmed vilka slag av analogi har vi att räkna med? Det mest generella slaget av analogi är väl det, där mönstret helt enkelt utgörs av själva den sammansatta verbtypen, alltså förled plus verbmorfem. Ett sådant mönster föreligger sedan äldsta tid i de partikelkomponerade verben, som här kan exemplifieras med gotiskans *distairan* 'slita sönder', *fairgreipan* 'gripa fast', *ufargaggan* 'gå över'. Låt oss kalla detta ordtyps-analogi. Detta och övriga slag av analogi förtecknas i bildningsschemat i tabell 1, till vilket jag här generellt hänvisar.

Ett annat också generellt slag av analogi är det, där mönstret utgörs av ordpar som *eld* – *elda*, *tvätt* – *tvätta*. Det är här alltså fråga om parvis förknippade nomen och verb. Vi kan kalla denna analogityp för nomen–verb-analogi.

Generellt är vidare det slag av analogi, där mönstret utgörs av grupper som *skriva* – *skriven* – *skrivning* – *skrivare* etc., dvs. ett verb med förknippade nominalbildningar. Detta slags analogi

TABELL 1. *Bildningsschema.*

Analogislag	Synkroniskt mönster	Normalt diakroniskt förlopp	Förlopp i det enskilda fallet	Ordbildningstyp
<i>Generell analogi</i>				
Ordtypsanalogi	sammansatt verb	-	<i>biff + bulta</i> → <i>biffbulta</i>	Primärt kompositum
Nomen-verb-analogi	<i>eld - elda</i> etc.	<i>eld</i> → <i>elda</i> etc.	<i>landskamp</i> → <i>landskampa</i>	Avledning
Verb-nomen-analogi	<i>skriva - skrivning</i> etc.	<i>skriva</i> → <i>skrivning</i> etc.	jäkt ← <i>jäkta</i> <i>elda</i> → <i>eldning</i>	Kort retrogradering Avledning
<i>Speciell analogi</i>				
Simplexanalogi	<i>fall - falla</i>	-	<i>knäfall</i> → <i>knäfalla</i>	Simplexanalogt kompositum
Ledanalogi	<i>rekord- + verbmorfem</i> - nominalmorfem + <i>-tala</i>	-	<i>rekord- + hoppa</i> → <i>rekordhoppa</i> <i>hembygd + -tala</i> → <i>hembygdstala</i>	Förledsanalogt kompositum Efterledsanalogt kompositum
Hapaxanalogi				
	<i>vanetenke</i>	-	<i>rekord- + -tala</i> → <i>rekordtala</i> <i>vanetenke</i> → <i>vaneføle</i>	Flerledsanalogt kompositum Hapaxanalogt kompositum (inomspråkligt)
	<i>schanfilen</i>	-	<i>schanfilen</i> → <i>skamfila</i>	Hapaxanalogt kompositum (tvärspråkligt)

kan vi kalla verb–nomen-analogi. De tre slag av analogi som jag nu har nämnt kan sammanfattas under beteckningen generell analogi. Det behöver knappast tilläggas att effekten av mönstren kraftigt förstärks, så snart de har givit upphov till nominalkomponerade verb.

Utöver den generella analogin har vi att räkna med olika typer av vad man kan kalla speciell analogi. Hit för jag då fall där en direkt specificerbar enhet eller en grupp av sådana enheter har avgett mönstret. Att exempelvis simplex *falla* har spelat in vid bildningen av *knäfalla* framgår av detta ords böjning. Vi kan kalla detta slag av speciell analogi för simplexanalogi.

Ett par typer av speciell analogi förutsätter att nominalkomponerade verb redan har etablerats. I den första av dessa typer står en grupp nominalkomponerade verb med samma för- eller efterled som mönster. Som exempel kan från senare tid anföras de förut nämnda verben med förleden *rekord-* som i *rekordhoppa*, *rekordlöpa* osv. respektive efterleden *-tala* som i *hembygdstala*, *tacktala* osv. Detta slags analogi kan vi kalla ledanalogi.

I vissa fall är det fråga om sådan speciell analogi, där enskilda verbkomposita har varit mönsterbildande. Ett gott exempel synes mig föreligga i norska *vaneføle* efter *vanetanke*: *Man både vantenkte og vanefølte* (Nordiska språkfrågor 1961 och 1962, s. 137). Denna typ av speciell analogi kan kallas hapaxanalogi. Hit hör i grunden också översättningslånen, ehuru mönstret i dessa fall hämtas utifrån. Som exempel kan anföras *skamfila* eller lågtyska *schamfilen*. De båda nämnda typerna kan betecknas som inomspråklig respektive tvärspråklig hapaxanalogi.

När nu de mönster som föreligger i språkssystemet har tagits fram, är det dags att uppmärksamma tidsfaktorn. Den är från ordbildningssynpunkt särskilt relevant i de båda fall av generell analogi som här har betecknats som nomen–verb-analogi och verb–nomen-analogi. Det historiska normalförhållandet mellan enheterna i dessa båda fall är som bekant (högerriktad pil markerar 'föregår')

eld → elda  
 tvätt → tvätta  
 dröm → drömma  
 tejp → tejpa  
 etc.

## respektive

skriva → skrivning ...  
 elda → eldning ...  
 löpa → löpning ...  
 läsa → läsning ...  
 etc.

I båda fallen betecknas produkterna som avledningar.

Går man nu till materialet rörande de nominalkomponerade verben, finner man å ena sidan

landskamp → landskampa  
 landsmål → landsmåla  
 vildkatt → vildkatta (om människa)  
 väderkvarn → väderkvarna (om propeller)  
 precedensincident (om föregiven olyckshändelse) → precedens-  
 incidenta (TV-pjäs 25/11 1968)

men å andra sidan t.ex. (vänsterriktad pil markerar 'föregås av')

landsätta ← landsättning  
 korrekturläsa ← korrekturläsning  
 skraddarsy ← skraddarsydd  
 deltaga ← deltagande (pres. part.)

Det finns alltså från historisk synpunkt anledning att kalla typen *landskampa* för avledning men inte typen *landsätta*. I klarhetens intresse bör den senare få en beteckning som inte alls erinrar om avledning. Retrogradering är därför bättre än retrograderad avledning. Retrogradering kan alltså bestämmas som avvikelser från det normala historiska resultatet av generell analogi. Jag understryker att termen retrogradering uteslutande hör hemma i det diakroniska studiet. Den saknar mening i ett synkroniskt resonemang, eftersom ju t.ex. *korrekturläsa* och *korrekturläsning* står på jämställd fot med *läsa* och *läsning* i ett givet vokabulär-system. Det är på den här punkten som Erik Wellanders resonemang så vitt jag kan se brister.

Jag påpekar i detta sammanhang att definitionen också ger klart utslag i fall som exempelvis

jäkt ← jäkta  
 hugg ← hugga  
 spring ← springa  
 glid ← glida

Vi konstaterar att det föreligger avvikelse från det normala historiska förloppet (*tvätt* → *tvätta*). Orden i vänstra ledet är sålunda retrograderingar. I latinska termer kan de som jag nämnt kallas nomina postverbalia. En något mera svensk beteckning vore kort retrogradering, en term som karakteriserar dem i förhållande till föregående retrograda typ.

Under alla omständigheter är det olämpligt att kalla typen *jäkt* kort avledning, eftersom elementet "avledning" i detta sammanhang är vilseledande. K. G. Ljunggrens resonemang "är det riktigt att beteckna ord av typen *hugg* som ett slags avledningar, så är det också riktigt att beteckna *sabelhugga* som en retrogradering" bygger sålunda på en olämplig terminologi snarare än på den språkliga verkligheten, och jag kan därför inte ansluta mig till det. Jag vill emellertid förutskicka att mitt ställningstagande inte innebär att jag betraktar *sabelhugga* som en avledning.

Ragnhild Söderberghs ovannämnda förslag att beteckna typen *jäkt* som avledning med nollsuffix är från diakronisk synpunkt likaså vilseledande, och från synkronisk synpunkt måste det betecknas som överflödigt. Från den sistnämnda synpunkten står ju *jäkt* och *jäkta* på jämställd fot med *tvätt* och *tvätta* (jfr mitt tidigare påpekande om *korrekturläsa*). När det gäller terminologin i andra språk kan man konstatera att engelskans *back-formation* och tyskans *Rückbildung* är bra men inte tyskans *retrograde Ableitung*.

Vi kommer så till frågan om de nominalkomponerade verbens böjning. Intresset knyter sig här i första hand till stark kontra svag böjning. Till en början kan vi konstatera att de retrograda verben inte vållar något problem. Här är ett slags simplexanalogi så att säga inbyggd, och sammansättningarna får därför genomgående samma böjning som simplex:

deltaga – deltog  
 korrekturläsa – korrekturläste  
 fondmäkla – fondmäklade

Annorlunda ställer sig saken när det gäller de verb som ansluter sig till nomen-verb-mönstret. För simplex gäller här i äldre som i nyare tid att verben genomgående får svag böjning: *eldade*, *dömde*, *tvättade*, *tejpade* osv. Det är inte att vänta att förhållandet skulle vara ett annat, när utgångspunkten är ett kompositum. Kontrollerbara fall som *landskampade*, *landsmålade*, *skärmbildade* bekräftar enhetligheten i fråga om böjningen.

Det finns sålunda all anledning att betrakta å ena sidan svag böjning som ett kriterium på avledning (dock inte ett tillräckligt kriterium, jämför nedan), å andra sidan stark böjning som ett kriterium på icke-avledning (ett tillräckligt kriterium).

Frågan kan belysas närmare med de i diskussionen anförda verben *knäfalla* och *hushålla*, sedda i relation till *knäfall* respektive *hushåll* (jfr Åkermalm 1955, s. 77, respektive Johannisson 1955, s. 167 not 13). Verbet *knäfalla* kunde i fornvästnordiskan böjas svagt – ”Ólafr konungr knéfallaði fyrir hánum með ǫllu lífúllæti” – vilket visar att ordet i det språkskedet alternativt framstår som en avledning. Vår starka böjning *knäföll* å andra sidan visar att verbet i denna språkmiljö inte är att betrakta som en avledning. Eftersom substantivet *knäfall* emellertid också här får antas i tiden föregå verbet *knäfalla*, förtjänar verbet enligt de här uppställda principerna inte heller beteckningen retrogradering. Kvar står att beteckna det som ett simplexanalogt kompositum efter den faktor som uppenbarligen har varit verksam vid ordbildningen. I fråga om *hushålla* ligger det till så att verbet i svenskan har böjts starkt från 1500-talet och till mitten av 1700-talet, därefter svagt. Substantivet *hushåll* är belagt tidigare än både *hushöll* och *hushållade*. Vid böjningen *hushöll* har vi sålunda att göra med ett simplexanalogt kompositum, vid böjningen *hushållade* med en avledning. Böjningsförändringen kan vara betingad av att *hushåll* efter hand kom att uppfattas som en i förhållande till simplex *håll* – *hålla* allt mer självständig enhet. På motsvarande sätt bör verbet *sabelhugga* sett i relation till *sabelhugg* beskrivas som en simplexanalog bildning. Vi har alltså på

det nu berörda området att göra med tre ordbildningstyper som kräver tre termer:

landsätta	retrogradering
hushålla (svagt)	avledning
hushålla (starkt)	simplexanalogt kompositum

Frågan kan härutöver belysas med motsvarigheterna på verb-sidan till substantivet *öknamn* (att vi i detta fall har en verbal förled är irrelevant från den anlagda synpunkten). Något verb är inte belagt i fornsvenskan, men i F. A. Dahlgrens Glossarium finner man ett belägg från äldre nysvenskan: *han hafver öknamnat honom* (Linköpings Capitels Dombok). Formen *öknamnat* visar att vi har att göra med en avledning. Annorlunda ställer det sig med den form A. F. Dalin ger i sin ordbok från 1800-talets mitt: *öknämna*. Som ett exempel på denna forms användning kan anföras följande belägg: *Invånarna i s[ödra] delen av socknen öknämnas Bönepungarna* (Ortnamnen i Göteborgs och Bohus län X (1942), s. 1). Till skillnad från det tidigare *öknamna* får *öknämna* betraktas som ett simplexanalogt kompositum – avledning med beledsagande omljud är ingen levande ordbildningstyp. Detta fall visar att man inte behöver gå utanför de svaga verbens område för att påvisa den aktuella skillnaden i ordbildning.

Vi har så att behandla typerna *telegrafsvara* och *atombomba*. Vid deras sida står den aktionella sammansättningen *telegrafsvar* respektive den icke-aktionella sammansättningen *atombomb*. Där står vidare i båda fallen osammansatta substantiv och verb. Eftersom nu båda de nominalkomponerade verben är svaga och böjda efter första konjugationen, kan vi inte avgöra huruvida det i fall som dessa är fråga om simplexanalogi eller avledning – eller samverkan av bådadera. Jämför härmed att i ett fall som Lars Gyllenstens *skenlever* (Diarium spirituale 1968, s. 130) böjningen (liksom vokalväxlingen) visar, att vi har att göra med en simplexanalog bildning och inte med en avledning av *skenliv*.

Jag nämnde tidigare att jag renodlade frågeställningarna, dvs. jag betraktade varje bildningssätt för sig. Som jag just antydde har det emellertid i praktiken ofta varit fråga om samverkande



krafter, med andra ord multipel bildning. Jag betraktar sålunda den av Åkermalms ovannämnda arbetshypoteser som går ut på att förutsätta ett enhetligt bildningssätt för varje verb såsom föga välgrundad. Jag tror att man i varje fall delvis kan återföra hypotesen på det faktum att Åkermalm tar de nusvenska förhållandena till utgångspunkt och underlåter att studera tidsförhållandet mellan nominalbildningarna och verben. En mera fruktbar metod hade varit att utgå från de bildningsförlopp som kan studeras under den nysvenska perioden som helhet och därvid utnyttja SAOB:s stora informationsmängd. Om så hade skett, hade arbetshypotesen måst revideras. I tabell 2 ger jag några exempel som visar detta.

TABELL 2. *Tidsordnade belägg I.*

BELÄGG	ÅR	BELÄGG	ÅR
deltagande (part.)	1727	grundmurning	1712
deltagare	1743	grundmur	1734
<i>deltaga</i>	1746	grundmurad	1751
deltagerska	1746	<i>grundmura</i>	1850
deltagande (subst.)	1747		
deltagelse	1818		
deltagarinna	1835		
hushållning	1541	kannstöpare (eg. bet.)	1686
hushållare	1544	kannstöpare (fig. bet.)	(1729) 1755
hushåll	1559	kannstöperi	1766
hushållerska	1572	kannstöpare (part.)	1839
<i>hushålla</i> (starkt)	1573	<i>kannstöpa</i>	1843
hushålleri	1638	kannstöperska	1913
hushållande (subst.)	1726		
<i>hushålla</i> (svagt)	1750		
korrekturläsning	1821	maskinskriven	1898
korrekturläsare	1831	maskinskrivare	1904
<i>korrekturläsa</i>	1888	maskinskriverska	1904
korrekturläserska	1929	maskinskrift	1906
		maskinskrivning	1908
		<i>maskinskriva</i>	1931
rekordlöpning	1890	sabelhugg	1771
rekordlopp	1906	sabelhuggare	1865
<i>rekordlöpa</i>	1924	sabelhuggning	1873
		sabelhuggen	1950
		<i>sabelhugga</i>	1954

Vad de enskilda beläggen i tabellen beträffar skall det nämnas att SAOB:s data har kompletterats med hjälp av Svenska Akademiens ordboksredaktions språkprovssamlingar (SAOS). Årtalet för *sabelhugga* återgår på Åke Åkermalms uppsats Om verbet *atombomba* etc. Det beläggen säger om verbens bildning har sammanfattats i tabell 3.

Av de valda exemplen visar sig alltså fyra vara uppkomna genom retrogradering, nämligen *deltaga* och *kannstöpa* med tre belagda utgångspunkter, *korrekturläsa* med två och *maskinskriva* med fem, tre genom retrogradering i kombination med simplexanalog bildning, nämligen *hushålla* (starkt) med fyra belagda utgångspunkter, *rekordlöpa* med två och *sabelhugga* med fyra, ett genom retrogradering i kombination med simplexanalog bildning alternativt avledning, nämligen *grundmura* med tre belagda utgångspunkter, och slutligen ett genom avledning, nämligen *hushålla* (svagt) som alltså har en enda utgångspunkt. Multipel bildning är sålunda en utbredd företeelse under den nysvenska perioden. Det förefaller mig sannolikt att förhållandet har varit jämförbart under fornspråklig tid.

Ett vackert exempel på vad jag kallar ledanalogiska bildningar finner vi i Stig Dagermans radiopjäsa Vår lilla sommar (1951), där det på ett ställe heter (manus s. 8; för citatkontroll tackar jag docent Gunnar Hallingberg): "Jag behöver bara påminna de närvarande om vartill den ansvarsmedvetna medborgaren använder sin sommar: Han vårtalar, han hembygdstalar, han jungfrutalar ...". Den produktiva efterleden *-tala* förekommer också i t.ex. *hälsningstala*, *invigningstala*, *jubileumstala*, *kamrattala* 'tala till medlemmar i de nordiska militära kamratföreningarna', *radio-tala*, *tacktala*, *välkomsttala* och *öppningstala*. Andra frekventa mönster för efterledsanalogi ger *-debutera* som i *filmdebutera*, *köpingsdebutera*, *radiodebutera*, *seriedebutera*, *skivdebutera*, *stumfilmsdebutera*, *Sverigedebutera* och *säsongsdebutera*, *försäkra* som i *bilförsäkra*, *brandförsäkra*, *livförsäkra*, *rån-försäkra*, *olycksfallsförsäkra*, *sjukförsäkra*, *sjöförsäkra*, *skadeförsäkra*, *stormskadeförsäkra* och *stöldförsäkra*, *-lägga* som i *beslaglägga*, *blötlägga*, *bordlägga*, *dimlägga*, *färglägga*, *grav-*

TABELL 3. *Multipel bildning.*

Verb	Retrogradering av nominalkompositum på <i>-ning -eri -l -are -erska -ad</i> etc. <i>-ande</i> (part.)	Avledning	Simplexanalogt kompositum alternativt avledning	Simplexanalogt kompositum
deltaga	+	+		
grundmura	+	+	+	
hushålla (starkt)	+	+		+
hushålla (svagt)			+	
kannstöpa	+	+		
korrekturläsa	+	+		
maskinskriva	+	+	+	
rekordlöpa	+			+
sabelhugga	+	+		+

*lägga, handlägga, kartlägga, korslägga, planlägga, stenlägga och taklägga, -spela som i avslutningsspela, bandyspela, försvarsspela, gästspela, kombinationsspela, kortspela, kvalspela, slutspela, triangelspela, uppvisningsspela, vänskapsspela och växelspela, -sätta som i betygsätta, färgsätta, gravsätta, landsätta, sjösätta, skuldsätta, smaksätta, stapelsätta, stensätta, strandsätta, syrsätta, tonsätta och värdesätta, -vinna som i debutvinna, finalvinna, lottvinna, millimetervinna, poängvinna, sensationsvinna, stafettvinna och tennisvinna.*

Förledsanalog verkan har likaså en rad element, bland dem *gäst-* som i *gästdirigera, gästföreläsa, gästpredika, gästregistrera, gästspela* och *gästsuppträda, kontroll-* som i *kontrollläsa, kontrollmäta, kontrollräkna, kontrollstämpla* och *kontrollväga, premiär-* som i *premiärboxa, premiärrulla, premiärträna, premiärtvätta, premiärveva* och *premiärvisa, pris-* som i *prisförbilliga, priskontrollera, prisnotera, prisreglera* och *prisöka, prov-* som i *provföreläsa, provgrädda, provhoppa, provodla, provprenumerera, provraka (sig)* och *provtvätta, sommar-* som i *sommarekipera, sommarföda, sommarförvara, sommarplugga, sommarplöja, sommarproviantera* och *sommarstarta, vinter-* som i *vinterbikta, vinterbona, vinterföda, vinterlagra, vintermata, vintersemestra, vintersåga* och *vinterutbilda.*

Efterhand som grupperna av för- respektive efterledsanaloga mönster vidgas, ökas möjligheterna till flerledsanaloga (dubbelt ledanaloga) bildningar. Förutom det ovannämnda *rekordtala* kan som exempel nämnas *premiärstrejka, premiärvinna, prissäta, provspela* och *vårtala.*

Att det även när det gäller de ledanaloga bildningarna i stor utsträckning är fråga om multipel bildning kan belysas med verbet *radiotala*. Efterledens produktivitet framgår av att redan Adolf Noreen i den ovannämnda materialsamlingen i sjunde bandet av *Vårt språk 1919* anför *baktala, buktala* och *liktala*. Substantiven *radiotalare* och *radiotal* är enligt SAOB tidigast belagda 1923 resp. 1934. Motsvarande belägg på verbet *radiotala* är från 1933. Till sin uppkomst är sålunda *radiotala* å ena sidan ett efterledsanalogt kompositum, å andra sidan en retrogradering till *radiotalare*. Med hänsyn till den härfina tidsskillnaden

mellan *radiotala* och *radiotal* kan också simplexanalogi alternativt avledning ifrågakomma.

Det skall tilläggas att semantiskt samhöriga element också kan spela in vid bildningen. En efterledsanalog grupp som *finalvinna*, *poängvinna*, *stafettvinna* m.fl. stöds förutom av substantiven *finalvinst* osv. också av de betydligt vanligare substantiven *finalseger* osv. Mera utpräglat framträder detta förhållande vid verb som *vänskapsspela* och *avslutningsspela* som otvivelaktigt stöds av substantiven *vänskapsmatch* respektive *avslutningsmatch*.

För fullständighetens skull nämner jag i detta sammanhang den semantiska urvattning som man kan iaktta hos vissa mycket produktiva förleder. Som exempel kan jag nämna *hem* i *hembjuda*, *hembära*, *hemdraga* osv. och *storm-* i *stormfundera*, *stormgräla*, *stormläsa* osv.

I tabell 1 gav jag som exempel på hapaxanaloga komposita det inomspråkliga *vaneföle* efter *vanetenke* och det tvärspråkliga *skamfila* efter *schamfilen*. Ytterligare exempel på den inomspråkliga typen är *djävulsförklara* efter *helgonförklara* och *teatersucka* efter *teaterviska* (Åkermalm 1965, s. 81), på den tvärspråkliga typen *dagtinga*, fsv. *daghþinga*, efter lågtyska *dagedingen* med varianter och *rådfråga*, fsv. *radhfragha*, efter lågtyska *ratvragen* (Åkermalm 1955, s. 181, 184). Också *korsförhöra* torde vara att betrakta som ett fall av tvärspråklig hapaxanalogi. Av språkprovssamlingarna vid Svenska Akademiens ordboksredaktion kan man utröna dels att verbet är belagt 1871 medan substantivet *korsförhör* är belagt först 1884, dels att de tre äldsta beläggen på verbet hänför sig till översättningar av Dickens (1871) och Twain (1873, 1874). Förebilden synes alltså vara engelskans *cross-examine*.

De ordbildningstyper som jag nu gått närmare in på – avledningar, retrograderingar, simplexanaloga, ledanaloga och hapaxanaloga bildningar – har det gemensamt att de är sekundära i förhållande till nominala och/eller verbala enheter. De förtjänar alltså beteckningen sekundära komposita. Som primära komposita framstår de fall där inget av de nu behandlade bildningssätten kan påvisas eller göras sannolikt, utan bildningen får antas ha skett efter mönster av ordtypen som helhet. Sannolikheten för

primära komposita bör öka i samma mån som antalet nominalkomponerade verb och deras frekvens ökar. I nusvenskan är alltså förutsättningarna rätt goda. Exemplet *biffbulta* i tabell 1 är hämtat från en artikel i Sydsvenska Dagbladet (2/8 1965): "Sven-Biff, kallad så i forskarkretsar för förnamnsinitialernas skull, vart så glad över fyndet, att han ryggdunkande biffbultade landsantikvarie som journalister." Andra fall är *ateljéstopp*, *spurta*, *firmaro*, *fotstegsåka*, *pennpeta* 'peta med penna på', *radoroa* (*sig*), *refrängjazza*, *räddningspaddla*, *silverskjuta* 'skjuta godkänt för den militära skyttemedaljen i silver' och *single-klå* 'besegra i singelmatch'. Av belägg från skönlitterära författare kan jag nämna Frödings *kysslängta*, Eyvind Johnsons *karamellsuga*, Olle Svenssons *interiörpremiere* och *kvalitetsplacera*, Bernt Eriksons *drunkningsgunga*, *dödsspika* och *luftcirkla*, Vilgot Sjömans *fyllglo* och Sandro Key-Åbergs *isdö*.

Av de fyra morfologiska huvudtyper som jag inledningsvis uppställde har jag hittills uppehållit mig vid de två första: *landsätta* och *landskampa*. Den tredje typen, *luftlandsätta*, är endast en utvidgning av den första typen med ytterligare ett nominalt led. Man kan bland hithörande verb urskilja två undergrupper, allt eftersom den viktigaste konstituentgränsen går mellan det första och det andra eller det andra och det tredje semantiska morfemet. Av det förra slaget är just *luftlandsätta* 'landsätta från luften', av det senare slaget exempelvis *landslagsturnera* 'turnera med landslaget'. Mina belägg tyder på att den första undergruppen är den minsta. Till den hör bl.a. *jubileumsrekordflyga*, *lagerbokföra*, *planhushålla* och *yrkesvägleda*. Att hänföra till den andra undergruppen är bl.a. *arbetsmarknadsanpassa*, *fotfelsförklara*, *fotstegsåka*, *kolosförgifta*, *kylrumsförvara*, *kyrkobokföra*, *skjutjärnsintervjua*, *skyddskoppsympa*, *skärmbildsfotografiera*, *syrgasbehandla* och *världsrekordsimma*. Ett fyrledat kompositum föreligger i det danska belägget *langtidsplanlægge* (Nordiska språkfrågor 1961 och 1962, s. 109).

Återstår så den intressanta typen *ilandföra*, där vi har en partikel före det nominala morfemet. Såvitt jag har kunnat finna har påfallande litet sagts om bildningen av denna verbtyp. Åkermalm är emellertid inne på den i sin avhandling med anledning av

fornsvenska *astadhkoma*. Han säger där (s. 210): "Den väsentliga skillnaden mellan typerna *bakbinda* och *astadhkoma* är den, att i förra fallet på grundval av *binda a bak* bildades p.p. *bakbundin*, varvid prepositionen 'föll', medan till *a stadh koma* ingen nominalsammansättning (vare sig particip eller substantiv) veterligen har skapats i fsv. eller nsv. tid. Därför finns prepositionen kvar i detta och liknande fall." Enligt min mening har Åkermalm knappast träffat det väsentliga i förloppet. Det finns sålunda en rad fall där partikelförsedda nominalkomposita är belagda före motsvarande verb. Exempelen i tabell 4 har jag huvudsakligen hämtat ur SAOS, eftersom SAOB är alltför knapphändig när det gäller relationen mellan å ena sidan participen och å andra sidan de finita formerna och infinitiven.

TABELL 4. *Tidsordnade belägg II.*

BELÄGG	ÅR	BELÄGG	ÅR
iakttagelse	1749	iståndsättiande (s.)	1734
iakttagare	1756	iståndsatt (part.)	c. 1740
iakttagning	1764	<i>iståndsätta</i>	1792
<i>iakttaga</i>	1802		
ifrågakomne	1822	ombordkomst	1854
<i>ifrågakomma</i>	1880	<i>ombordkomma</i>	1903
ifrågasatt (part.)	1833	pånyttfödd	1810
<i>ifrågasätta</i>	1859	pånyttfödelse	1820
		pånyttfödning	1833
		<i>pånyttföda</i>	1841
		pånyttfödare	1843
iscensättning	1882	åsidosatte (part.)	c. 1740
<i>iscensätta</i>	1886	<i>åsidosätta</i>	c. 1750
iscensättare	1892	åsidosättande (s.)	1786

För min del vill jag gärna tänka mig den förklaringen, att leden i prepositionsfrasen har kommit att kännas så intimt samhöriga i de aktuella fallen, att de har smält ihop till en (adverbiell) enhet före sammansättningarnas bildande. Parallellen *ihjäl* inställer sig osökt. Av förbindelsen verb + adverb har sedan såväl sammansatta verb som nominalsammansättningar direkt kunnat bildas. Typen *bakbinda* representerar normalfallen, där tyngdpunkten

har legat på participet och/eller andra nominalbildningar, medan typen *åstadkomma* representerar specialfallen, där prepositionsfrasen relativt sett har haft större vikt. Vissa ambivalenta fall har man att vänta. Ett sådant är *landsätta/ilandsätta*. Jämför å ena sidan uteslutande *landstiga*, å andra sidan uteslutande *ilandföra*.

I vissa fall får man också när det gäller dessa partikelförsedda verb räkna med tvärspråklig hapaxanalogi. Som exempel härpå kan jag ge *åvägabringa* efter danska *tilvejebringe*, lågtyska *to wege bringen*, anført av Esaias Tegnér i uppsatsen *Tyska inflytelser på svenskan* (Arkiv för nordisk filologi 5 (1889), s. 336). Till den morfologiska huvudtyp jag är inne på ansluter sig numera tämligen många fall. Utöver de ovan anförda kan jag nämna *avdagataga*, *igångsätta*, *ihågkomma*, *iordningställa*, *omhänderha(va)*, *tillfångata(ga)*, *tillgodose*, *tillmötesgå*, *tillryggalägga*, *tillrättavisa* och *ådagalägga*. En särskild variant representerar *förhandsgranska* 'granska på förhand', där partikeln förekommer redan i prepositionsfrasens substantiv.

I detta avsnitt skall jag avslutningsvis nämna några mera speciella verbtyper. En bildas av sådana verb som innehåller en förkortning. Hit hör t.ex. *korpskjuta* och *kvalspela*, vilkas förleder återgår på *korporation* respektive *kvalificering*. Flerledat är *teolodtors-avhandla* (*pietismen*) (DN 5/1 1968). Längre drivna förkortningar uppvisar *mc-dödas* och *TV-filma* för att inte tala om *NYWF-flanera* 'flanera på New York World Fair' (SDS 20/5 1965). Idrottsspråket har en viss flora av typen *DM-tävla*, *EM-boxas*, *VM-starta* (där det är fråga om distriktsmästerskap etc.) och *OS-träna* (där det gäller olympiska spel). Dit hör också *k.o.-besegra* 'besegra på knockout'.

Engelskt inflytande manifesterar sig i verb av typen *time-chartra* 'tidsbefrakta' (SAOL), där ändelser ur svensk morfologi har fogats till främmande ordkroppar. Till denna typ hör *freelanca* 'arbeta som frilans' (SDS 1/8 1965). Ett annat belägg är *cocktail-partya*: "Man skall helst hänga med i svängen och cocktail-partya också – alltid får man veta något." (SDS 29/7 1965). Ytterligare ett är *play-backa*: "Tidigare har Lasse Bagges bakgrundspiano *play-backats*." (DN 19/10 1968).



Beträffande verb som *annonsbraska*, *kemtvätta* m.fl. ber jag få hänvisa till uppsatsen Denominala verbalfraser.

#### 4. Förekomst

De nominalkomponerade verbens förekomst i fornspråkens litteratur har Åke Åkermalm redogjort för i sin ovan refererade avhandling. Sammanlagt omfattar hans material 317 verb med substantivisk förled, nämligen 144 fornsvenska, 127 fornvästnordiska, 44 forndanska och 2 forngutniska (s. 216). Som Åkermalm påpekar är siffrorna avhängiga av den bevarade litteraturens omfång inom de olika språkområdena och kan därför inte läggas till grund för några slutsatser om olikheter i produktiviteten hos ordbildningstypen.

Antalet verb ökar under de närmaste århundradena ganska långsamt, och det är först i sen tid som fördämningarna sprängs. Utvecklingen låter sig väl följa i de successiva delarna av Peter Skautrups danska språkhistoria (se ovan avsnitt 2). Åkermalm sätter med skäl verbens definitiva genombrott i samband med dagspressens snabba utveckling under senare tid (Rubriksvenska, s. 59). Som hållpunkter för de svenska verbens vidkommande kan jag nämna att Hainer 1888 angav antalet till omkring 120 på basis av Dalins ordbok och SAOL (Om de sammansatta verben i nysvenskan, s. 3), att Adolf Noreen 1919 anförde c. 250 verb med substantivisk förled (Vårt språk 7, s. 469 ff.) och att Åkermalm åren 1942–1952 samlade in ungefär 1200 verb, huvudsakligen ur dagspressen (Rubriksvenska, s. 61). Under ett halvår 1951 samlade jag själv in ungefär 1000 verb ur i första hand tidningar.

För att belysa verben från frekvenssynpunkt kan jag utnyttja mitt tidningsmaterial från 1965. Bland de (drygt 10 000) ord som bildar detta materials basvokabulär – dvs. de ord som uppfyller vissa bestämda krav på frekvens och på jämnhet i spridningen – ingår följande nominalkomponerade verb: *åstadkomma*, *delta*, *tillgodose*, *frigöra*, *fastställa*, *tjänstgöra*, *ifrågasätta*, *redogöra*, *klarlägga*, *kartlägga*, *fullfölja*, *sysselsätta* och *tillfredsställa*.

Innehållsligt har verbgruppen som synes en ganska generell karaktär, motsvarande ett vidsträckt användningsområde.

Hur är läget i danskan? En genomgång av Nudansk ordbog (c. 1000 tvåspaltiga sidor) ger ett resultat av ungefär 225 verb. Antalet är sålunda inte överväldigande. Som avvikelser från svenskan framstår bl.a. *blidgøre* 'blidka', *irettesætte* 'tillrätta-visa', *personliggøre* 'personifiera', *skamskænde* 'misshandla' och *varetage* 'tillvarata'. Av de tretton nyssnämnda orden i den svenska basvokabulären har nio direkta motsvarigheter: *afstedkomme*, *deltage*, *frigøre*, *kortlægge*, *klarlægge*, *redegøre*, *sysselsette*, *tilfredsstille* och *tjenstgøre*. Verben *fastställa*, *fullgöra*, *ifrågasätta* och *tillgodose* saknar formella motsvarigheter.

Vi kommer så till det västnordiska området. Genomgången med avseende på norskan gäller Leiv Heggstads Fornorskinsordliste (c. 100 trespaltiga sidor). Antalet påträffade fall är omkring 120. Att detta är lägre än det föregående skall ses mot bakgrunden av att bokens omfång (trots den avsevärt högre tätheten) är betydligt mindre än Nudansk ordbogs. Bland avvikelser från svenskan kan nämnas *bakvaske* 'baktala', *huglegge* 'förälska sig i', *imøtekomme* 'tillmötesgå', *ivareta* 'tillvarata' och *synkverve* 'förvränga synen på'. Fem direkta motsvarigheter till verben i basvokabulären förekommer i ordlistan: *delta*, *redegjøre*, *kartlegge*, *sysselsette* och *tilfredsstille*.

Att jag som föremål för genomgången valt just Fornorskinsordliste hänger samman med att den ger vissa inblickar i förhållandet mellan bokmål och nynorska på den aktuella punkten. I en rad fall återges ett bokmålsverb med ett annat nominalkomponerat verb i nynorskan, bland dem följande:

<i>bokmål</i>	<i>nynorska</i>
bekjentgjøre	kunngjera
brulegge	steinleggja
bønnefalle	naudbeda
festligholde	høgtida
landsforvise	landlysa
lemleste	skamslå
livnære	livberga
offentliggjøre	kunngjera
rådslå	rådleggja
veilede	rettleia

Inte mindre intressanta är de fall där som motsvarighet ges ett enkelt verb, ensamt eller i en lös förbindelse. Hit hör:

<i>bokmål</i>	<i>nynorska</i>
blidgjøre	blidka
delta	ta del
frita	ta fri
hemmeligholde	halda løynd
iakttta	ansa etter
imøtegå	leggja imot
imøtese	venta
irrettesette	gje(va) ei skrape
istandssette	setja i stand
ledsage	følgja
lukksaliggjøre	gjera sæl
løslate	lata laus
løsrive	riva laus
radbrekke	bryta sund
redegjøre	gjera greie for
tilbakelegge	leggja attom seg
tilveiebringe	skaffa
vareta	ta(ka) vare
virkeligjøre	setja i verk

Eftersom det rör sig om ungefär en sjättedel av fallen, bör det kunna betecknas som en tendens.

Isländskan uppvisar en stor mängd nominalkomponerade verb. I andra upplagan av Isländsk-svensk ordbok av Leijström-Magnússon-Jansson (c. 400 tvåspaltiga sidor) förekommer omkring 500. Inget av dessa motsvarar något verb i den svenska basvokabulären. Över huvud taget är det en rätt liten del av dem – ett trettiotal – som i ordboken återges med en formell svensk motsvarighet. Jag kan nämna *baktala*, *bókfæra*, *botnfrjósa*, *búsetja*, *eiðfesta*, *fríðlýsa*, *fullgera*, *hálshöggva*, *handhafa*, *kaldhamra*, *lofsyngja*, *lögsækja*, *prestvígja*, *ráðfæra*, *skrinleggja*, *stálsetja*, *steinleggja*, *trúlofa* och *þrælbinda*.

Från morfologisk synpunkt kan man lägga märke till att spetsställd partikel som i *afvegaleiða* 'förleda' är sällsynt. Vidare kan man iakttta enstaka fall av treledade verb: *handleggsbrotna* 'bryta armen', *votheysfergja* 'silopressa' (till *vothey* 'ensilage').

De nominalkomponerade verben är ett av vapnen i den isländska purismens arsenal. Detta kan belysas med bl.a. följande fall:

*bólusetja* 'vaccinera', *dáleiða* 'hypnotisera', *gerilsneyða* 'pastörisera', *hraðrita* 'stenografera', *kvikmynda* 'filma', *ljósmynda* 'fotografera', *rafvirkja* 'elektrifiera', *símrita* 'telegrafera', *sótt-hreinsa* 'desinficera', *steinprenta* 'litografera' och *þjóðnýta* 'socialisera'.

När det gäller de nordiska riksspråken kan vi sammanfattningsvis konstatera att verben är rikt företrädde. Det vill vidare synas som om isländskan vid sidan av svenskan skulle ha den största benägenheten för ordtypen.

Om de nominalkomponerade verbens förekomst i dialekterna ges praktiskt taget inga uppgifter i litteraturen. Det är belysande att Elias Wessén som jag nämnt endast anför ett verb, *buföra* 'flytta med korna till fåbodarna', i andra delen av Svensk språkhistoria. I stället betonar han starkt hur skriftspråklig företeelsen är (s. 108 not 4). I själva verket är verbtypen emellertid rikligt företrädd i våra dialekter.

En genomgång av Rietz' dialektlexikon ger över 150 hit hörande verb: *barstjåla*, *eggvålla*, *liabränna*, *vöttusätt*, *ögonblåkas* osv. För att få en uppfattning om ordförrådet på denna punkt i en mera utförlig lexikografisk framställning har jag gått igenom det hittills färdigställda avsnittet A–G i den opublicerade Bohuslänsk ordbok (föreligger på kort i samlingarna vid Institutet för ortnamns- och dialektforskning i Göteborg). Genomgången gav i runt tal 100 verb, vilket är ungefär fem gånger så många som motsvarande parti hos Rietz uppvisar. Med reservation för ojämnheter i de båda ordmaterialen skulle man alltså kunna räkna med att en fullbordad version av den bohuslänska ordboken skulle komma att innehålla nominalkomponerade verb till ett antal av snarare över än under 500. Det är att märka att endast tre verb är gemensamma för Bohuslänsk ordbok och Rietz i det jämförda partiet, nämligen *eggvålla*, *eggvända* (*eggvånna*) och *gallskrika*. En liknande tendens gör sig gällande när andra källor från det nordiska området dras in (se nedan). Det sammanlagda antalet olika verb i våra dialektordböcker och dialektsamlingar torde därför vara avsevärt högre än den ovan angivna siffran.

Genom sin mängd bör verben spela en ganska betydelsefull roll i den dialektala vokabulären. Detta intryck förstärks när man grupperar dem efter ämnesområden och finner, att de i största utsträckning hänför sig till det dagliga livet: arbetet, relationerna till medmänniskorna, företeelserna i omgivningen. I den följande exemplifieringen står R för Rietz och B för Bohuslänsk ordbok.

Till arbetet i jord och mark hänför sig *bottenhacka* 'hacka djupt' (B), *djupgrava* (B), *flåhacka* 'hacka av grässvål' (B), *grunddika* 'täckdika' (B), *kasföra* 'sammanföra ris m.m. i hopar' (R), *körttröska* 'tröska med ojämna slag' (R), *landfära* 'uppdraga tvärfåror vid ändarna av en åker' (R), *liabränna* 'avhugga gräset så nära att gräsroten skadas av solhetta eller frost' (R), *spanntröska* 'tröska (säd) och i betalning erhålla en viss mängd därav' (R), *trävsätta* 'sätta sädeskärvorna uti en rad' (R).

Från boskapskötselns område kan förutom det ovan anförda *boföra* (*buföra*) nämnas *bogskälva* 'skälva i bogen av överanstängning (om hästar)' (B), *bogsticka* 'sticka ett djur i bogen vid slakt' (B), *glugg-gäta* 'valla kor, men endast se till dem ibland' (B), *gånghälla* 'binda samman ett framben och ett bakben på får' (B), *görbräkka* 'stänga (kreatur) så att gorret går ut' (R), *köttmäta* 'skilja lårstycket av kött från benet att det kan torkas' (R), *stakaränna sej* 'vid hoppning över gärdesgård rädda en stör i livet på sig (om djur)' (R), *vållränna* 'skena' (R).

Till skogliga sysslor hör *billhugga* 'skräda grovt' (B), *bomflytta* 'vid sågning för hand flytta den främre bommen varpå stocken vilade' (B), *bomskära* 'såga en stock intill den främre bommen som stocken vilade på' (B), *finskräda* (B), *grovskräda* (B).

Sysslor av hantverkskaraktär åsyftar bl.a. *bordfylla* 'sätta bord i en båt' (B), *brattröst* 'bygga hög takresning' (R), *brädfodra* (B), *eggvälla* 'ståla eggen (av ett skärande eller huggande, förut stål-satt redskap)' (B, R; Rietz upptar ordet också under *ä* och med en något avvikande betydelse), *finhyvla* (B), *kinnbain* 'bila (en timmerstock) på två sidor där den skall intimras i husknut' (R), *lerslå* 'överstryka väggar med lerbruk' (R) *röleta* 'rödfärga' (R), *tjärbrå* 'överstryka med tjära' (R).

Till fiske och segling hör *blodsalta* 'salta sill utan att först gäla den' (B), *bottenreva* (B), *fjölhala* 'sträcka seglet för hårt vid bidevindssegling' (B), *grundstöta* (B), *isnööt* 'draga isnot' (R), *limtåål* 'övergjuta avkok av fiskfjäll och skinn över färgade fiskredskap att färgen ej må så snart upplösas' (R), *ljusföra* 'ljustra' (R).

Till hushållsarbete hänför sig en lång rad verb, bland dem *blåskumma* 'skumma ytterst noga (om mjölk)' (B), *bomvinda* 'bobina (garn)' (B), *eggvända* (*eggavänna*) 'sidvända ett lakan' (B, R), *finbråta* 'finbråka (lin)' (B), *finkarda* (B), *gräddfalla* 'bilda grädde (om nymjölk)' (B), *hóseförföta* 'förföta (strumpor)' (R), *kósthåll* 'bespisa' (R), *matreda* 'göra i ordning mat' (R), *nástgjera* 'utrusta med reskost' (R), *skirtvätta* 'tvätta ren' (R), *såpetvätta* (R), *vöttusätt* 'lägga i vatten' (R).

Till arbete i allmänhet refererar *blodslita* 'arbeta hårt' (B), *blåhånga* 'slöa' (B), *hannsóla* 'skaffa sig biförtjänst' (R), *skålverka* 'skaffa sig biförtjänst' (R), *strängsätta* 'hårt anstränga' (R).

Till kroppsliga ting hänför sig *aumkyl* 'förorsaka att man blir ömkyld' (R), *blodrinna* 'dö av blodförlust' (B), *blåfrysa* (B), *lemmstita* 'stöta en lem så att värk uppkommer' (R), *årela* 'åderlåta' (R), *åreslå* 'åderlåta' (R).

En del verb står i relation till lag och rätt, t.ex. *barstjåla* 'frånstjåla någon allt vad han äger' (R), *bysätta* (B), *farkänna* 'uppgeva såsom barnfader (om ogift kvinna)' (R), *fläckkföra sig* 'lämna ifrån sig all sin förmögenhet mot underhåll till dödsdagar' (R), *tjyvstjåla* 'stjåla allt löst som finnes på ett ställe' (R).

Verb som anger mänskligt beteende i kontaktsituationer av olika slag bildar den största gruppen. Till denna hör bl.a. *blodpina* 'pina svårt' (B), *fillrista* 'skaka (en person) grundligt' (B), *flatskratta* (B), *flepläpa* 'göra spektakel åt någon' (B), *fulltacka* 'tacka tillräckligt' (B), *fulltala* 'tala mycket för att vilja övertyga' (B), *fångbrytas* 'brottas' (B), *gallskrika* (B, R), *godtala* 'tala väl med någon' (B), *handgangäs* 'slåss' (R), *handpala* 'handplagga' (R), *huveröska* 'lugga' (R), *kallflina* 'hånligt le' (R), *korannta* 'ystert springa omkring såsom kalvar' (R), *långledes* 'ha ledsamt' (R), *mulbännas* 'munhuggas' (R), *näsbränna* 'skarpt tillrättvisa någon' (R), *olhuggas* 'träta' (R), *olkastas*

'munhuggas' (R), *påsarista* 'hårt skaka' (R), *rómprökkja* 'ständigt jaga efter' (R), *småvura* 'förklena' (R), *snömylas* 'slåss med snöbollar' (R), *spórrökkä* 'förfölja någons spår' (R), *sänglaga* 'hava samlag' (R), *trynhuggas* 'kyssas' (R), *trälbärja* 'slå någon brun och blå' (R), *tupplivas* 'slåss' (R), *vattlekramma* 'klämma och trycka sin motpart mycket hårt' (R), *ögonblåkas* 'se kärligt på varandra' (R).

Till sist några verb som rör väderleken:

*bottenfrysa* (B), *daggfalla* 'bli daggigt' (B), *da(g)mena* (om dagsmejan) (R), *grundfrussta* 'bottenfrysa' (R), *skulregna* 'starkt slagregna' (R), *thordöna* 'åska' (R).

Beläggstätheten är som jag nämnde hög i det bohuslänska materialet. För att ge en bild av hur samlingar från andra håll ställer sig i relation dels till den bohuslänska ordboken, dels till Rietz skall jag redovisa några resultat från en genomgång av partiet A–G i tre andra ordböcker, nämligen den gotländska ordboken, Feilbergs ordbok och Aasens ordbok.

Gotländsk ordbok ger ett femtontal verb, vilket ligger något under antalet hos Rietz. Siffran är den lägsta av de erhållna. Även den aviserar emellertid en icke obetydlig ordsfatt i lexikonet som helhet. Gemensamma med Rietz är verben *barstjåla*, *eggvända* 'stadvända', *fläckföra* (*sig*) 'taga undantag' och *grundfrost*. Mot *bannsätta* svarar vidare *bannsatt(er)* hos Rietz. Exempel på verb som saknas hos Rietz (och i Bohuslänsk ordbok) är *arsjämka* 'jämka sig fram', *beckvalda* 'björklaka', *döds skjuta*, *eldbåda* 'värma en yxa', *fiskföda* (*sig*) och *gatulägga* 'uppriva gärdesgården kring en åker och lägga den ihop med gatan'.

Antalet verb i det aktuella partiet av H. F. Feilbergs Bidrag till en ordbog over jyske almuesmål överensstämmer med det man finner hos Rietz. Det är alltså fråga om ett tjugotal. Motsvarigheter hos Rietz har *endevende* (*ännavänna*) och *flækføre*, i den bohuslänska ordboken *bovstikke*, *dagsætte* och *fuldtakke*. Bland verb som saknas hos Rietz är *bagspringe* ('när tyren fejler koen under parringen'), *baldehule* (*æn træsku*) 'hule plads ud for fodens balde', *blindgribe* (i samband med leken blindbock) och *brudegive* 'bortgive som brud'.

I Norsk Ordbog av Ivar Aasen är antalet verb långt högre än i Rietz' ordbok. Siffran är ungefär sextio (orden på *aa-* har utelutits) och ligger alltså mitt emellan talen för Rietz och den bohuslänska ordboken. Överensstämmelserna med Rietz är påfallande få: *buføra* och *fletføra* (Rietz *fläckføra*). Bland de många övriga fallen är *baktyggja* 'baktala', *bannstøyta* 'förbanna', *blodvekkja* 'såra', *bordsyda* 'bekläda med bräder', *botnarsetja* 'sätta fisknät på botten så att de inte når till vattenytan', *braadvenda*, *dagverda* 'äta frukost', *endesnu*, *flatrydja* 'hugga timmerstockar så att sidorna blir flata', *godmæla* 'värdesätta', *gorstinga* 'sticka i inälvorna' och *grønsalta*.

Också i löpande dialekttext kan verbtypen naturligt nog beläggas. Exempel hämtar jag från Djurklou och Bondeson. I Svenska sagor och sägner talas det om "ho mang' krötter de kónn' vinterfö' te mans" (6 (1953), s. 141). I Historiegubbar på Dal finner man: "Men där var ändå ett par karlar, som gick åt skogen och ville se, hur han skogkörde timret, varföre de klev upp i var sitt träd" (s. 109); "Och de går på och slår, och fan fick stickebryna [ 'vässta lien med brynsticka' ] titt och brått" (s. 125).

Vi kan sålunda konstatera att verbtypen är väl spridd inte bara i de nordiska riksspråken utan också i dialekterna. I och för sig borde detta också vara fallet med tanke på utbredningen i fornspråken. Den konkreta eller om man så vill livsnära prägel som det dialektala ordförrådet har ger det emellertid en särställning i förhållande till det fornspråkliga (liksom för övrigt det riksspråkliga). Detta skall säkerligen sättas i samband med att den bevarade litteraturen på fornspråken långt ifrån är representativ för hela det språkliga fältet. De nominalkomponerade verben torde ha spelat en påtaglig roll i de talade fornspråken liksom de nu gör det i dialekterna.



## Litteratur

### *Otryckt*

Bohuslänsk ordbok (Institutet för ortnamns- och dialektforskning).

### *Tryckt*

Aasen, I., Norsk Ordbog. 4 Udg. Kristiania 1918.

Cederschiöld, G., Framtidssvenska. 1917.

Cederschiöld, G., Fresta duger. 1914.

Diderichsen, P., Elementær dansk Grammatik. København 1946. (2. Udg. 1957.)

Feilberg, H. F., Bidrag till en ordbog over jyske almuesmål. 1. Kjøbenhavn 1886–93.

Gustavson, H., Gotländsk ordbok 1. 1918–40.

Hainer, H., Om de sammansatta verben i isländskan. 1877.

Hainer, H., Om de sammansatta verben i nysvenskan. (Redogörelse för de allm. läroverken i Blekinge 1887–88. 1888, s. 3ff.)

Hall, R. A., Jr, How we noun-incorporate in English. (American Speech 1956, s. 83 ff.)

Heggstad, L., Fornorskingsordliste. 11. oppl. Oslo 1968.

Henzen, W., Deutsche Wortbildung. 1947. (3. Aufl. Tübingen 1965.)

Jespersen, O., A few back-formations. (Englische Studien 70 (1936), s. 117 ff.)

Johannisson, T., Om sammansatta verb i svenskan. (Nysvenska studier 34 (tr. 1955), s. 162 ff.)

Johannisson, T., Über Rückbildung im heutigen Schwedisch. (Erbe der Vergangenheit. Festgabe für Karl Helm. Tübingen 1951, s. 257 ff.)

Körner, R., Studier över syntaktisk nybildning i svenskan. 1. 1948.

Leijström, G., Magnússon, J., & Jansson, S. B. F., Isländsk-svensk ordbok. 2 uppl. 1955.

Ljunggren, K. G., Objekt och adverbial. 1942.

- Ljunggren, K. G., Ytterligare om termerna avledning och retrograd avledning. (Arkiv för nordisk filologi 71 (1956), s. 69 f.)
- Marchand, H., The categories and types of present-day English word-formation. Alabama 1960.
- Nordiska språkfrågor 1955. (Nämnden för svensk språkvård 28. 1963.)
- Noreen, A., Vårt språk. 7. 1919.
- Noreen, E., Bidrag till fornsvensk lexikografi. (Meijerbergs arkiv för svensk ordforskning 3 (1941), s.15 ff.)
- Noreen, E., Svensk stilparodi. 1944.
- Nudansk ordbog. 1-2. 5. udg. København 1967.
- Rietz, J. E., Svenskt dialektlexikon. Nytryck Lund 1962.
- Rydqvist, J., Svenska Språkets Lagar. 5. 1874.
- Sandfeld, K., Sprogvidenskaben. 2. uppl. Kjøbenhavn 1923.
- Skautrup, P., Det danske sprogs historie. 1-4. København 1944-68.
- Söderbergh, R., Svensk ordbildning. 1967.
- Tamm, F., Sammansatta ord i nutida svenskan. 1900.
- Tegnér, E., Tyska inflytelser på svenskan. (Arkiv för nordisk filologi 5 (1889).)
- Wellander, E., Ett par produktiva typer av skriftspråklig nybildning. (Språk och stil 15 (1915), s. 44 ff.)
- Wessén, E., Svensk språkhistoria. 2. 1943. (4. uppl. 1965.)
- Wessén, E., Vårt svenska språk. 1968.
- Åkermalm, Å., Fornnordiska verb med substantivisk förled. 1955.
- Åkermalm, Å., Rubriksvenska. 1965.
- Östergren, O., Språklig nyskapelse. (Nordisk tidskrift 1905, s. 588 ff.)

## Verbal allomorfi i modern svenska

Språkdata 1967.

Det är välbekant att en rad verb i modern svenska har alternativ tempusböjning. Elias Wessén ger åtskilliga exempel i första delen av Svensk språkhistoria (§ 198 ff.): *tvinga – tvingade/tvang – tvingat/tvungit*; *växa – växte – växt/vuxit*; *betala – betaladel/betalte – betalat/betalt* osv. Det är också välbekant att en rad verb har alternativformer av annat slag: *karakterisera – karaktärisera*, *poängtera/pointera*, *tal/ta*, *göra/göres* osv. Ofta får endera varianten språkvårdens rekommendation; sålunda ges t.ex. *aktivera* och *effektivera* förord framför *aktivisera* och *effektivisera* (Gösta Bergman, Rätt och fel i språket, s. 63). Syftet med föreliggande uppsats är att studera denna verbens allomorfi från kombinerat kvalitativ-kvantitativ synpunkt.

Den moderna svenskans verbsystem har på senare tid kortfattat behandlats av Lars-Gunnar Hallander 1955 i uppsatsen Morfofonemiska förhållanden vid svensk verbböjning och av J. Lotz 1964 i uppsatsen Morphophonemic alternations in the Swedish strong verb (Ablaut). Kvantitativa uppgifter om verben i äldre stockholmskt talspråk ges av Carl Ivar Ståhle 1966 i Verbböjningstyper och verbfrekvens i stockholmskt berättarspråk.

Den av uppsatserna som har störst intresse från den här anlagda synpunkten är den första. I Hallanders beskrivning av verbböjningstyperna får imperativen utgöra verbens stam. Härigenom kan de tre svaga konjugationerna slås samman till en enda; valet av ändelseallomorf betingas av stammorfemets fonematiska form (s. 145 f.). Preteritummorfemet uppträder sålunda som /te/ efter tonlösa fonem (*köpte*), som /dde/ efter vokalfonem med huvudaccent (*bodde*) och som /de/ efter övriga tonande fonem (*kallade*, *dömde*). Motsatsen mellan svag och stark böjning låter sig däremot inte beskrivas som fonematiskt betingad. För varje stammorfen måste därför anges om det har svag eller stark böjning. Analysen utmynnar i att tre böjningstyper uppställs: A (*kalla*, *bo*), B (*komma*, *bita*), AB (*gå*).

Att välja imperativen som utgångspunkt för beskrivningen har betydande fördelar. Det är exempelvis också den som Sven E. G. Öhman utgår ifrån i uppsatsen *Generativa regler för det svenska verbets fonologi och prosodi* (1966, s. 76 ff.). En konsekvens är att verb med alternativ böjning får anses ha alternativa stammar: *betalade* och *simmade* återgår på *betala-* respektive *simma-*, *betalte* och *sam* på *betal-* respektive *simm-*. Detta tycks också ha varit Hallanders mening (jfr s. 146, not 2).

En invändning mot analysen nämner Hallander själv (s. 146, not 2): ett antal A-verb med stam på /n/ och /l/ har preteritum-allomorfen /te/, t.ex. *brynte*, *betalte*. Jag kan inte ansluta mig till författarens (visserligen försiktigt formulerade) förslag, att /n/ och /l/ i dessa fall skulle vara att betrakta som från fonematisk synpunkt tonande men från morfematisk synpunkt tonlösa. I stället torde vederbörande stammar få omhändertas av en särskild regel, som fördelar ändelseallomorferna efter stammarnas identitet (*betal-te*, *ställ-de* osv.).

Det nämnda förhållandet är ett av skälen till att den analys jag lägger till grund för den följande behandlingen i viss mån avviker från Hallanders. Ett annat är att eventuell vokalväxling inom stammorfemet enligt min mening genomgående bör beaktas vid typiseringen (jfr Hallander s. 147).

TABELL 1. *Verbböjningstyper*.

Typ	(1) Stamslut på vokal		(2) Vokal- växling	(3) Preteritum på				(4) Supinum på		
	obeton.	beton.		de	te	dde	Ø	t	tt	it
1	+			+				+		
2a			+	+				+		
2b			+	+				+		
2c			+		+			+		
2d			+		+			+		
3		+				+			+	
4a			+				+			+
4b			+	+			+			+

I tabell 1 redovisas de åtta grundläggande verbböjningstyper som jag urskiljer. Indelningen bygger på fyra kriterier: stamslutande fonem, vokalväxling i stammen, preteritumsuffixets form och supinumsuffixets form. Typbeteckningarnas siffror ansluter sig till Beckmans konjugationsbeteckningar (§ 141). Som exempel på verb tillhörande de olika klasserna kan följande anföras:

1	kalla	kallade	kallat
2a	böja	böjde	böjt
2b	välja	valde	valt
2c	köpa	köpte	köpt
2d	bringa	bragte	bragt
3	tro	trodde	trott
4a	komma	kom	kommit
4b	fara	for	farit

I ett fall som *slå* (*slog – slagit*) kan stammen betraktas som slutande på konsonant. Verbet är då att hänföra till klass 4b. Ett antal verb uppvisar blandad böjning, t.ex. följande:

2d+1	veta	visste	vetat
3+4b	le	log	lett
4b+1	ligga	låg	legat

Om preteritum av hjälpverbet *skola* uppfattas som *skulle* < *skulde*, kan verbet klassificeras som 2b+1.

Från kvalitativ synpunkt skiljer jag sålunda mellan tre slag av böjning, det tredje fördelat på två kategorier:

- |  |       |
|--|-------|
| 1) ordinär böjning                                   |       |
| ex. <i>kalla – kallade – kallat</i>                  | 1     |
| 2) blandad böjning                                   |       |
| ex. <i>le – log – lett</i>                           | 3+4b  |
| 3) alternativ böjning                                |       |
| a) oinskränkt  |       |
| ex. <i>tvinga – tvingade/tvang – tvingat/tvungit</i> | 1/4b  |
| b) inskränkt   |       |
| ex. <i>växa – växte – växt/vuxit</i>                 | 2c/4b |

Teoretiskt sett är det av vikt att relationen mellan begreppen morfem och lemma (se mitt arbete Studier över nusvenskans

vokabulärsystem, s. 46 ff., särskilt s. 60) klargörs. Fallen under 3 ovan ger exempel på allomorfi inom ett och samma lemma, homolemmatisk allomorfi:

tvinga-de	väx-t
tvang-Ø	vux-it

Stamformerna *tvinga-* och *väx-* har allomorferna *tvang-* respektive *vux-*, böjningsändelserna *-de* och *-t* har allomorferna *-Ø* respektive *-it*.

Å andra sidan föreligger heterolemmatisk allomorfi mellan t.ex. *väx-* i verbet *växa* och *väx-* i substantivet *växt*, *-de* i preteritum *tvingade* och *-te* i preteritum *växte*. Enheterna morfem och lemma opererar alltså oberoende av varandra. Detta sammanhänger med två fundamentala olikheter dem emellan: dels tillhör de skilda stukturella strata, dels är morfemet till skillnad från lemmat betydelsebaserat.

Undersökningens grundmaterial omfattar cirka 250 verblemman (exklusive fallen med växling *-es/s* i presens, se tabell 17 nedan), sammanställda ur ordböcker, grammatikor och uppsatser. De torde på ett fylligt sätt representera allomorfifenomenet. I drygt 100 av dessa fall rör det sig om alternativa böjningsformer, i återstoden av fallen om alternativa grundformer.

I kvantitativt avseende gäller undersökningen vilka av alternativformerna som är kuranta i modern svenska, med andra ord hur det i källorna angivna formsystemet utnyttjas. De kvantitativa uppgifterna hämtas ur lexikonserien *Frekvenser i svenskan 1965*, som torde ge en representativ bild av läget i modernt svenskt standardspråk. I vissa fall är de studerade formerna behäftade med homonymi, extern homonymi som i fråga om *rykte* (verb, substantiv) och *svalt* (verb, adjektiv), intern homonymi som i fråga om *tvingade* (preteritum, particip), interlingval homonymi som i fråga om *la* (svenskt verb, utländsk artikel). I regel går formerna emellertid fria, och inte sällan är homonymin irrelevant med avseende på den aktuella frågeställningen.

Av de nämnda cirka 250 verblemmarna (alternativa heterolemmatiska enheter har alltså räknats som olika enheter) är ett

hundratal så lågfrekventa, att de antingen inte belagts alls eller också belagts endast i någon här icke studerad form i Frekvenser i svenskan 1965. De har med andra ord en relativ frekvens som i allmänhet understiger 1/1 000 000 i den språkform frekvensundersökningen avser. I en mindre del av dessa fall är det fråga om alternativ tempusböjning. Hit hör i huvudsak följande:

1/2c	drösa, förlika, förtenna, knysta, nästa, röta, stäcka, vässa
1/2c/4b	fisa
1/4b	skrinna
2a/2b	förtälja, kvälja, tämja
2a/2c	mäla
2a/4b	kvida, kväda
2c/4b	genomlöpa, gitta, nypa, nysa, rysa, skrympa, utlöpa
3/4b	två

De största grupperna är alltså de som alternerar mellan klasserna 1 och 2c respektive 2c och 4b.

I huvuddelen av de icke belagda fallen rör det sig om alternativa grundformer, homolemmatiska eller heterolemmatiska. Av hithörande enheter kan följande anföras:

aluminera/aluminisera	avvälta/avvältra
barlasta/ballasta	bortschackra/bortsjackra
chosa/sjåsa	cleara/klira
entresolera/entresollera	förkoxa/förkoxa
gipa/gippa	grensla/gränsla
harska/harskla	höta/hytta
klichera/klischer	licensera/licensiera
loma/lomma	mopeda/mopedera
morna/morgna	mortifiera/mortificera
mussera/moussera	petrificera/petrificera
ragla/raggla	rekognoscera/rekognosera
resumera/resymera	sjappa/schappa
splitsa/splissa	sulagera/soulagera
såvera/sauvera	taschera/tascha
tüllregla/tüllrigla	velurisera/velourisera
volma/vålma	vyssa/vyssja

Exemplen visar variation av flera olika slag, avseende såväl bildning som uttal och stavning.

Mängden av lågfrekventa lemman är intressant i och för sig som ett uttryck för utnyttjandegraden hos vederbörande enheter.

Huvudintresset knyter sig emellertid naturligtvis till de frekventare enheter, där valet av alternativformer kan studeras. I det följande behandlas först böjning efter alternativa böjningsklasser och sedan övriga alternativformers uppträdande. I det förra avseendet ställs först svag böjning mot stark och sedan olika slag av svag böjning mot varandra.

Jag skiljer från kvantitativ synpunkt mellan två huvudtyper av böjning: icke saxad böjning och saxad böjning. Med saxning menar jag växling mellan böjningsklasser som i fallet *dyka* – *dök* – *dykt* (där alternativet *dykte* sålunda ej utnyttjats); jämför den kvalitativa beskrivningens blandade böjning. Under huvudtypen icke saxad böjning skiljer jag mellan konsekvent böjning efter viss böjningstyp (inga belägg på alternativa former), dominant böjning efter viss böjningstyp (vissa belägg på alternativa former) och odeciderad böjning (likartad fördelning på alternativa former). Under huvudtypen saxad böjning skiljer jag på motsvarande sätt mellan konsekvent och dominant böjning. De olika slagen av böjning blir alltså följande:

- 1) icke saxad böjning
  - a) konsekvent  
ex. *sitta* – *satt* (144 belägg) – *suttit* (23)/*setat* (0)
  - b) dominant  
ex. *tvinga* – *tvingade* (46)/*tvang* (8) – *tvingat* (37)/*tvungit* (1)
  - c) odeciderad  
ex. *koka* – *kokade* (2)/*kokte* (1) – *kokat* (3)/*kokt* (9)
  
- 2) saxad böjning
  - a) konsekvent  
ex. *dyka* – *dök* (20)/*dykte* (0) – *dykt* (10)
  - b) dominant  
ex. *sprida* – *spridde* (1)/*spred* (14) – *spritt* (11)/*spridit* (0)

De åtta verbböjningstyperna är i växlande omfattning inbegripna i fall med alternativ böjning. Antalet i mitt frekvensmaterial belagda lemman med potentiellt alternativ böjning samt arten av denna böjning registreras i tabell 2. Det framgår där att av de 28 principiellt möjliga kombinationerna 10 är potentiellt förekommande.



TABELL 2. *Antal belagda verblemmen med potentiellt alternativ böjning.*

<i>1</i>	<i>2a</i>	<i>2b</i>	<i>2c</i>	<i>2d</i>	<i>3</i>	<i>4a</i>	<i>4b</i>
<i>1</i>	<i>1</i>	<i>1</i>	<i>23</i>	<i>1</i>			<i>11</i>
	<i>2a</i>	<i>2</i>	<i>2</i>				<i>9</i>
		<i>2b</i>					
			<i>2c</i>				<i>16</i>
				<i>2d</i>			
					<i>3</i>		<i>3</i>
						<i>4a</i>	
							<i>4b</i>

De vanligaste mönstren är 1/2c (*koka – kokadel/kokte – kokat/kokt*), 2c/4b (*växa – växte – växt/vuxit*), 1/4b (*besluta – besluta-del/beslöt – beslutat/beslutit*) och 2a/4b (*sprida – spriddelspred – spritt*) i nu nämnd ordning. Antalet olika potentiella kombinationer som var och en av de åtta verbböjningstyperna ingår i framgår av tabell 3. Vanligast i detta avseende är alltså klass 1. Den lilla klassen 4a (*komma, sova, vara*), och endast den, uppvisar ingen kombination. De kvantitativa uppgifterna i de följande tabellerna avser formerna i presens (aktivum) – när böjningstyp 1 är inbegripen – preteritum (aktivum + *s*-form) och supinum (aktivum + *s*-form), vilket torde ge en god bild av förhållandena. De med \* markerade sifbertalen är behäftade med viss osäkerhet på grund av homonymi.

TABELL 3. *Antal potentiella kombinationer för de olika verbböjningstyperna.*

Typ	Antal kombinationer
<i>1</i>	<i>5</i>
<i>2a, 4b</i>	<i>4</i>
<i>2c</i>	<i>3</i>
<i>2b</i>	<i>2</i>
<i>2d, 3</i>	<i>1</i>
<i>4a</i>	<i>0</i>

I tabell 4 förtecknas verb med böjningsalternativen 1 och 4b. Homonymi med substantiv föreligger i fallen *klang* och *tiger* (fall

som *teg* torde kunna negligeras), homonymi med proprium i fallet *sam*. Den interna homonymin mellan preteritum och particip av typen *avslutade* har inte särskilt markerats. Anmärkas bör, att de sammanlagt 28 belägg på *slök(s)* (alla förekommande i en artikel) som upptas på s. 1 249 i andra delen av Frekvenser i svenskan 1965 hänför sig till propriet *Slök* (egentligen *Sløk*, danskt släktnamn).

TABELL 4. *Böjningsklass 1 alternativt 4b.*

Verb	Presens		Preteritum		Supinum		S:a	
	1	4b	1	4b	1	4b	1	4b
avsluta	10		23		13		46	0
besluta	15		12	49	33		60	49
förtiga		5				3	0	8
klinga	6		3	*20			9	*20
simma	2		2	*10	1		5	*10
sitta		178		144		23	0	345
sluka	5		4		1		10	0
sluta	47		71	8	14	4	132	12
stupa	4		6		2		12	0
tiga		*5		2		1	0	*8
tvinga	19		46	8	37	1	102	9

Av de 11 verben i tabellen visar 3 – *avsluta*, *sluka* och *stupa* – konsekvent svag böjning och likaså 3 – *förtiga*, *sitta* och *tiga* – konsekvent stark böjning. Verben *sluta* och *tvinga* har dominant svag böjning, medan de övriga – *besluta* och med viss reservation *klinga* och *simma* – har odeciderad böjning. Serien *avsluta* – *sluta* – *besluta* går alltså från konsekvent svag via dominant svag till odeciderad böjning (det skall dock noteras att *sluta* bildar två lemman: *sluta* – *slöt* – *slutit* 'stänga m.m.'; *sluta* – *slutade*/*slöt* – *slutat* 'fullborda m.m.').

Det är intressant att konstatera, att alla de 6 verb som visar konsekvent svag eller stark böjning från kvalitativ synpunkt sett har inskränkt alternativ böjning, medan de övriga 5 har oinskränkt alternativ böjning. Vidare är det intressant att de 3 verben med konsekvent svag böjning har sitt böjningsalternativ i preteritum (*avslutade*/*avslöt*, *slukade*/*slök*, *stupade*/*stöp*), medan de 3 verben med konsekvent stark böjning har sitt alternativ i

supinum (*förtogat/förtigit, setat/suttit, tegattigit*). Ett karakteristikon för gruppen som helhet är övervikten (i fråga om antalet olika lemman) för svag presensböjning.

Tabell 5 upptar de verb som har böjningsalternativen 2a och 4b. Av de båda fylligast belagda verben har *lyda* odeciderad böjning och *sprida* dominant saxad böjning. I kvalitativt avseende har båda enligt Illustrerad svensk ordbok (ISO) inskränkt alternativ böjning, i det att supinum enbart bildas efter klass 2a. Odeciderad böjning finner vi vidare hos *begrava* och *bestrida*. För gruppen som helhet kan vi notera en klar övervikt för å ena sidan starkt preteritum och å andra sidan svagt supinum. Frånräknas de båda homonymibelastade verben, fås sammanlagt 35 belägg på starkt preteritum mot 12 på svagt preteritum och 16 belägg på svagt supinum mot 2 på starkt supinum. Detta sammanhänger till en del med att 5 av verben har inskränkt alternativ böjning (hit hör enligt frekvensmaterialets vittnesbörd inte *bestrida*, för vilket verb ISO endast uppger supinum *bestritt*). Att *smälla* egentligen bildar två verblemman, varav endast det intransitiva *smälla* – *smälldelsmall* – *smällt* här är av intresse, inverkar på grund av vederbörande beläggstals låga värden inte på den statistiska bedömningen av gruppen.

TABELL 5. Böjningsklass 2a alternativt 4b.

Verb	Preteritum		Supinum		S:a	
	2a	4b	2a	4b	2a	4b
begrava	1	1	2		3	1
bestrida		3	1	1	1	4
duga		3			0	3
häva	3	*7	1		4	*7
lyda	8	10	2		10	10
skälva	1	*1			1	*1
smälla	2	2			2	2
sprida	1	14	11		12	14
strida		2		1	0	3

Böjningsalternativet 2c/4b redovisas i tabell 6. Av de aktuella verben visar *hjälpa* konsekvent svag böjning och *dyka* konsekvent saxad böjning. Av dessa två har *dyka* inskränkt alterna-

tiv böjning. Vidare torde man kunna räkna med odeciderad böjning för *växa*. Åtskilliga av formerna i tabellen är behäftade med homonymi, och fyra av verben (*frysa*, *skvätta*, *smälta*, *svälta*) bildar två lemman. Likafullt kan en tydlig tendens beträffande gruppen som helhet iakttas: den markanta övervikten för svagt supinum. Tendensen beror endast till en del på inskränkt alternativ böjning hos verben. Iakttagelsen överensstämmer med den som gällde föregående grupps supinum.

TABELL 6. *Böjningsklass 2c alternativt 4b.*

Verb	Preteritum		Supinum		S:a	
	2c	4b	2c	4b	2c	4b
belöpa	1				1	0
dräpa			1		1	0
dyka		20	10		10	20
fnysa		1	1		1	1
frysa		6	1	2	1	8
förlöpa	1	*18	1		2	*18
hjälpa	25		18		43	0
löpa	4	*36	1		5	*36
ryka	*34	*9	1		*35	*9
skvätta	2	*1			2	*1
smälta	2	*2	4		6	*2
strypa		1			0	1
svälta		*11	*8		*8	*11
vika		*5	*40		*40	*5
vräka	1		1		2	0
växa	41		*20	40	*61	40

I en fjärde grupp växlar böjning efter klass 3 med böjning efter klass 4b. Här tar jag upp 3 verb – *bel/bedja*, *förse* och *gel/giva* – med uteslutande av prefigurerade enheter som *ange*, *avge*, *bege* m.fl. Beläggen anförs i tabell 7. Det framgår att *bel/bedja* och *förse* har konsekvent saxad böjning, medan *gel/giva* har odeciderad böjning. Kvalitativt sett har alla tre verben inskränkt alternativ böjning.

TABELL 7. Böjningsklass 3 alternativt 4b.

Verb	Preteritum		Supinum		S:a	
	3	4b	3	4b	3	4b
be/bedja		*39	*10		*10	*39
förse		7	18		18	7
ge/giva		284	83	87	83	371

När det gäller de nu behandlade verbgrupperna, där alltså svag böjning står mot stark, kan vi summera våra kvantitativt grundade iakttagelser på följande sätt:

konsekvent svag böjning: *avsluta* (typ 1), *sluka* (1), *stupa* (1), *hjälpa* (2c)  
 dominant svag böjning: *sluta* (1), *tvinga* (1)  
 konsekvent stark böjning: *förtiga* (4b), *sitta* (4b), *tiga* (4b)  
 konsekvent saxad böjning: *dyka* (4b/2c), *be/bedja* (4b/3), *förse* (4b/3)  
 dominant saxad böjning: *sprida* (4b/2a)  
 odeciderad böjning: *besluta* (1/4b), *klinga* (1/4b), *simma* (1/4b), *begrava* (2a/4b), *bestrida* (2a/4b), *lyda* (2a/4b), *växa* (2c/4b), *ge/giva* (3/4b)  
 allmänt karakteristikon: 1/4b svag presensböjning  
 2a/4b stark preteritumböjning  
 2a/4b, 2c/4b svag supinumböjning

Närmast betraktar vi så de verb, där båda alternativen gäller svag böjning. I tabell 8 har sammanförts tre verb, vilkas böjningsalternativ följer å ena sidan klass 1, å andra sidan klass 2a, 2b respektive 2d. Av dessa verb visar *leva* dominant böjning efter klass 2a, de båda övriga odeciderad böjning. Belägget på supinum av *töras/tordas* har formen *tordats*.

TABELL 8. Böjningsklass 1 alternativt 2a, 2b respektive 2d.

Verb	Presens				Preteritum				Supinum				S:a			
	1	2a	2b	2d	1	2a	2b	2d	1	2a	2b	2d	1	2a	2b	2d
leva		*131			52				9	14			9	*197		
töras/tordas			5						1				1		5	
bringa	2					3					5		2			8

Den största av alla alternativböjningsgrupperna har växling mellan klasserna 1 och 2c, se tabell 9. Frekvenstalen är också i majoriteten av fallen höga. Av de 23 verben i tabellen bildar 9 en

framträdande undergrupp med konsekvent böjning efter typ 1: *ana*, *låna*, *mena*, *ropa*, *skapa*, *spela* (2 lemman), *tala*, *tjäna* och *visa*. Härtill ansluter sig *förtjäna* med dominant böjning efter typ 1. Å andra sidan har verben *mista* och *fästa* konsekvent respektive dominant böjning efter typ 2c. Odeciderad böjning visar *betala*, *förmena* (dock 2 lemman) och *koka*. Som allmänt karaktistikon framstår en tydlig benägenhet för böjning efter klass 1.

TABELL 9. *Böjningsklass 1 alternativt 2c.*

Verb	Presens		Preteritum		Supinum		S:a	
	1	2c	1	2c	1	2c	1	2c
ana	16		12		5		33	0
betala	33		21		9	28	63	28
byka		1					0	1
frälsa				1		2	0	3
fästa		24	1	*14		6	1	*44
förmena	1			1	1	3	2	4
förmåla		2					0	2
försona	1		1				2	0
förtjäna	29		1		1	1	31	1
koka	7		2	1	3	9	12	10
krysta			2		1		3	0
låna	12		11		19		42	0
mena	209		60		17		286	0
mista		6		*12		9	0	*27
ropa	10		17		2		29	0
skapa	63		39		44		146	0
skona			2		1		3	0
spela	165		72		59		296	0
sprätta	1						1	0
tala	246		95		49		390	0
tjusa			1			*4	1	*4
tjäna	33		9		8		50	0
visa	386		188		161		735	0

De återstående verben förtecknas i tabell 10. Det gäller två verb med alternativen 2a/2b och två verb med alternativen 2a/2c. Den förra gruppen visar konsekvent 2b-böjning, den senare konsekvent 2a-böjning. I fråga om den sistnämnda gruppen har beläggen i supinum inte angetts i tabellen, eftersom formerna efter klasserna 2a och 2c är identiska på denna punkt.

TABELL 10. Böjningsklass 2a alternativt 2b respektive 2c.

Verb	Preteritum			Supinum			S:a		
	2a	2b	2c	2a	2b	2c	2a	2b	2c
dväljas		1					0	1	
svälja		1			*11		0	*12	
påminna	9		0						
tåla	3		0						

I fråga om verben med svag alternativböjning kan vi sammanfattningsvis notera följande kvantitativa böjningsegenskaper:

konsekvent typ 1:	ana, låna, mena, ropa, skapa, spela, tala, tjäna, visa
dominant typ 1:	förtjäna
konsekvent typ 2a:	påminna
dominant typ 2a:	leva
konsekvent typ 2b:	svälja
konsekvent typ 2c:	mista
dominant typ 2c:	fästa
odeciderade:	töras/tordas (1/2b), betala (1/2c), förmena (1/2c), koka (1/2c), bringa (1/2d)
allmänt karakteristikon:	1/2c böjning efter typ 1

Exempel på saxad böjning saknas.

Totalt sett har vi fått en ganska klar uppfattning om de reella böjningsförhållandena hos ett fyrtiotal enskilda verb med potentiellt alternativ tempusböjning. Det har visat sig att drygt hälften av dessa i praktiken konsekvent böjs efter viss böjningstyp. Från historisk synpunkt innebär detta en stabilisering av vederbörande verbs flexion. Med växlande säkerhetsgrad, beroende på frekvenstalens storlek, kan det i vissa fall ses som slutfasen i en övergång från en böjningstyp till en annan; så t.ex. i fallen *hjälpa* och *svälja* (stark till svag), *tiga* (svag till stark), *låna* och *visa* (svag till annan svag böjning). I andra fall, t.ex. *mista* och *tala*, är det att betrakta som en återgång till konsekvent ursprunglig böjning. I åter andra fall, såsom *dyka*, tycks blandad böjning etableras. Den andra hälften av verben omfattar en mindre grupp med dominant böjning efter viss typ, t.ex. *tvinga* och *fästa*, och en större grupp med odeciderad böjning, t.ex. *besluta*, *koka*, *lyda* och *växa*. Dessa grupper kan från den anlagda synpunkten ses

som stegvis mindre utpräglade motstycken till de konsekvent böjda verben.

De hittills behandlade fallen av verbal allomorfi har hänfört sig till växling mellan olika verbböjningstyper. I fortsättningen tas övriga slag av alternativformer upp. Nästan genomgående blir det fråga om växling inom givna verbböjningstyper. Först behandlas stamvariation (sådan som inte sammanhänger med tempusflexionen), därefter böjningsvariation. Vid studiet av alternativa stamformer anges inte i varje särskilt fall huruvida alternativen innehåller ett morfem i likhet med *säg/säj-* eller flera i likhet med *karaktärisera/karaktärisera-*.

TABELL 11. *Variation i stammars uttrycksform.*

Växling	Alternativ 1	Frekvens	Alternativ 2	Frekvens
e/ä	karaktärisera-	27	karaktärisera-	23
	modernisera-	10	modärnisera-	0
i/y	hiva-	1	hyva-	0
	rist-	1	ryst-	0
i/ö	flirta-	1	flörta-	0
ou/u	desavouera-	1	desavuera-	0
	goutera-	1	gutera-	0
	outrera-	2	uttrera-	1
u/y	nunna-	0	nynna-	1
	refusera-	2	refysera-	0
u/å	bugna-	0	bågna-	1
in/äng	pointera-	0	poängtera-	13
c/k	handicapa-	0	handikappa-	54
	indicera-	1	indikera-	0
	camouflera-	1	kamouflera-	3
cc/ks	vaccinera-	3	vaksinera-	0
d/g	fradga-	*3	fragga-	0
dj/j	rödj-	0	röj-	20
g/j	säg-	1333	säj-	9
n/nd	hunsa-	2	hundsä-	0
r/rj	svär-	1	svärj-	0
s/ss	hisna-	11	hissna-	2
ss/ts	låssa-	1	låtsä-	20



Den första aspekten på stammarna gäller växling i den grafematiska formen, se tabell 11. I de behandlade fallen är denna växling i allmänhet inte förknippad med växling i den fonematiska formen. Flertalet hithörande stammar är lågfrekventa. I en tredjedel av fallen erhålles dock ett klart utslag. Sålunda kan man konstatera ett konsekvent bruk av stammarna *modernisera-*, *poängtera-*, *handikappa-* och *röj-* och ett dominant bruk av stammarna *säg-*, *hisna-* och *låtsa-*, medan bruket i fråga om *karaktiseralkaraktärisera-* är odeciderat. I huvuddelen av de lågfrekventa fallen står från långgivande språk hämtade stavningar mot försvenskade stavningar: *i/ö*, *ou/u*, *u/y*, *c/k*, *cc/ks*. Tendensen bland de åtta ifrågakommande stammarna går i den förra variantens favör: 12 belägg mot 4.

TABELL 12. *Frekvenser för vissa alternativa grundformer.*

Alternativ 1	Frekvens	Alternativ 2	Frekvens
be	21	bedja	5
klä	15	kläda	2
rå	*3	råda	24
spä	0	späda	*6
trä	*38	träda	*18
dra	122	draga	8
ta	646	taga	13
bli	968	bliva	2
ge	510	giva	*2
ha	1484	hava	4

Vissa enmorfemiga stammar slutande på *-d*, *-g* och *-v* har allomorfer utan dessa fonem/grafem. För vederbörande verbs vidkommande innebär detta en växling mellan en- och tvåstaviga grundformer: *bel/bedja*, *dral/draga*, *blil/bliva* osv. Frekvenserna för tio hithörande verbs grundformsalternativ ges i tabell 12. För de fem stammarna med *g-* och *v-*alternativ är det kvantitativa utslaget mycket tydligt, i det att bruket av de kortare varianterna är starkt dominant. Gruppen med *d-*alternativ har dels betydligt lägre frekvens, dels heterogen frekvensfördelning. I de enskilda fallen föredras *be*, *klä* respektive *råda* (förhållandena vid *späl/späda* och *trä/träda* undandrar sig bedömning på grund av

homonymi). Totalt innebär detta att kortformen dominerar i 7 av 8 fall.

Kanske inte helt väntat (jfr Bergman, Rätt och fel i språket, s. 39) ger grundformerna av sammansättningarna med de tio nämnda verben i huvudsak samma bild, se tabell 13. Inom gruppen med stammar på *-g* och *-v* är kortformerna klart dominanta. Frånräknas siffrorna för homonymerna *andra* (jämte motsvarigheten *andraga*) och *ådra*, blir beläggstalen för *-draldraga* 40 mot 3, *-taltaga* 171 mot 41, *-bli/bliva* 15 mot 1, *-gelgiva* 101 mot 3 och *-halhava* 6 mot 0. Sammanlagt står alltså 333 belägg mot 48. Huvudparten av beläggen på de längre formerna kommer på *-taga*, som härigenom visar en egen profil, dock utan att bryta den allmänna tendensen. Verben med *d*-alternativ är ytterst sporadiskt belagda. Belägget på *-råda* ansluter sig till den dominerande av motsvarande simplexformer, medan de tre beläggen på *-kläda* ansluter sig till den mindre kuranta simplexvarianten.

TABELL 13. *Frekvenser för vissa sammansättnings alternativa grundformer.*

Alternativ 1	Frekvens	Alternativ 2	Frekvens
-be	0	-bedja	0
-klä	0	bekläda	2
		av-	1
-rå	0	tillråda	1
-spä	0	-späda	0
-trä	0	-träda	0
föredra	7	bidraga	3
bi-	29	an-	1
till-	2		
an-	*1882		
undan-	1		
upp-	1		
å-	*2		
tillvarata	3	vidtaga	6
vid-	9	god-	1
god-	10	före-	2
före-	12	del-	9
fri-	2	till-	1

del-	*34	an-	7
an-	23	in-	4
in-	15	upp-	1
från-	1	över-	2
upp-	5	för-	1
återupp-	3	iakt-	1
omhänder-	2	mot-	4
åter-	4	av-	1
över-	23	på-	1
för-	5		
iakt-	6		
mot-	7		
av-	1		
å-	6		
utebli	3	förbliva	1
för-	12		
tillkännage	2	avgiva	3
med-	11		
be-	4		
fri-	1		
del-	2		
om-	4		
an-	21		
in-	10		
upp-	7		
åter-	18		
över-	11		
ut-	5		
av-	5		
handha	3	-hava	0
inne-	3		

Kortare och längre stamformsvarianter uppträder också vid vissa verb på *-era*. I de fall som anförs i tabell 14 gäller det dels suffixvariation *-eralisera*, *-eraliera* – dels variation inom bas-morfemet – *verifi/verific-*, *pulvr/pulver-*. Med undantag av *effektivisera* väljs i frekvensmaterialet genomgående den kortare formen. För gruppen totalt är siffrorna 18 mot 6.

TABELL 14. *Frekvenser för stammar på -era.*

Alternativ 1	Frekvens	Alternativ 2	Frekvens
aktivera-	3	aktivera-	0
effektivera-	0	effektivera-	6
onanera-	1	onanisera	0
injicera-	3	injicera-	0
projicera-	2	projicera-	0
verifiera-	7	verificera-	0
pulvriserar-	2	pulvriserar	0

Inom presensböjningen förekommer också varianter av olika längd i en rad fall. Jag tar först upp de tio verben med *d-*, *g-* och *v-*alternativ, se tabell 15. Det mest framträdande draget i tabellen är att de fem *g-* och *v-*verben uppvisar konsekvent kortformsböjning i presens aktivum; sammanlagt föreligger inemot 12 000 belägg. I passivum (där infinitivformerna av formella skäl medtagits) har kortformerna inte ett lika förkrossande övertag – dock är de klart dominanta. Vad *d-*verben beträffar kan vi konstatera att formerna *ber* och *råder* tydligt föredras. Totalt gäller sålunda att kortformen dominerar hos 6 av 7 verb.

TABELL 15. *Frekvenser för vissa alternativa presensformer. I.*

Presens aktivum				Presens (+infinitiv) passivum			
Alternativ 1	F	Alternativ 2	F	Alternativ 1	F	Alternativ 2	F
ber	25	beder	0				
klär	14	kläder	*57				
rår	*2	råder	89				
trår	0	träder	*22	spås	0	spåds	1
drår	105	drager	0	dras	55	drages	6
						dragas	3
tar	386	tager	0	tas	108	tages	*3
						tagas	5
blir	1394	bliver	0				
ger	652	giver	0	ges	98	gives	2
						givas	0
har	9176	haver	0				

I tabell 16 förtecknar jag dels ett antal verb med *j*-variant i presens, dels ett par verb med speciella kortformer. För den förra gruppens vidkommande framgår det att *j*-lös form dominerar i aktivum, medan bruket i fråga om *s*-formerna tycks vara mera odeciderat. När det gäller *spar/sparar* favoriseras i någon mån den längre formen, kanske på grund av tryck från former som *varar*, *svarar*, *parar*. Av allomorferna *ska* och *skall* slutligen dominerar den senare eftertryckligt. Det kan tilläggas att ungefär hälften (155) av beläggen på *ska* hänför sig till en av de fem tidningar som är representerade i frekvensmaterialet, nämligen Stockholms-Tidningen.

TABELL 16. *Frekvenser för vissa alternativa presensformer. II.*

R-former				S-former			
Alternativ 1	F	Alternativ 2	F	Alternativ 1	F	Alternativ 2	F
gläder	5	glädjer	1	gläds	1	glädjes	0
				skils	1	skiljs	5
						skiljes	1
stöder	19	stödjer	2	stöds	*3	stödjdes	0
svär	3	svärjer	0				
spar	2	sparar	7				
ska	338	skall	2005				

Ett särskilt uppmärksammat slag av allomorfi i presensböjningen är växlingen *-es/s*. Denna har från olika utgångspunkter berörts av Rudolv Körner 1954 (Nutida tendenser i svensk språkutveckling. I, s. 99 ff.), Ture Johannisson 1961 (Det dolda stilmönstret, DN (A) 29/7 1961, s. 4), Gösta Bergman 1962 (Rätt och fel i språket, s. 35 ff.) och Sven Engdahl 1962 (Studier i nusvensk sakprosa, särskilt s. 133 f.).

I mitt material föreligger i runt tal 4 000 belägg på formerna i fråga. Deras fördelning på allomorfer och verbböjningstyper framgår av tabell 17. Vissa former har uteslutits ur tabellen på grund av homonymi, i de flesta fall beroende på den ena variantens sammanfall med preteritumformen. Fallen är *sänd(e)s*, *insänd(e)s*, *utsänd(e)s*, *tänd(e)s*, *vänd(e)s*, *använd(e)s*, *lyft(e)s*, *fäst(e)s*, *befäst(e)s*, *skvätt(e)s*, *åstadkom(me)s*, *möt(e)s* och

*läng(e)s*. Det sammanlagda antalet belägg på dessa former är så lågt – 73 belägg på *es*-former och 174 belägg på *s*-former – att deras uteslutande inte inverkar på totalbedömningen av företeelsen.

TABELL 17. *Fördelning på verbböjningstyper av presensformer på -es alternativt -s.*

Verb- böjnings- typ	Former på <i>-es</i>			Former på <i>-s</i>		
	Antal belägg	Antal lemman	Antal belägg per lemma	Antal belägg	Antal lemman	Antal belägg per lemma
2a	109	61	1,8	868	132	6,6
2b	54	19	2,8	137	16	8,6
2c	131	14	9,4	592	50	11,8
2d	8	6	1,3	71	15	4,7
4b	115	46	2,5	1833	66	27,8
1/2b				5	1	5,0
1/4b				4	2	2,0
2a/2c	1	1	1,0	2	2	1,0
2a/4b	4	2	2,0	13	5	2,6
<i>Totalt</i>	422	149	2,8	3525	289	12,2

Av de sammanlagt 3 947 förekomsterna svarar *finns* ensamt för 40 % (1 573 belägg; antalet belägg på *finnes* är 11). I övrigt är det endast tre former som har ett beläggsantal överstigande 100, nämligen *tycks* 367, *behövs* 107 och *synes* likaså 107 (motsvarande siffror är för *tyckes* 0, *behöves* 0 och *syns* 27). Formen *synes* torde för övrigt normalt ha en från *syns* avvikande betydelse, vilket kvalitativt sett innebär att de båda böjningsändelserna är i färd med att utveckla morfematisk distinktivitet. Även om beläggen på *finn(e)s* frånräknas, dominerar formerna på *-s* mycket markant: 1 952 belägg mot 411. Detta yttrar sig också i det genomsnittliga antalet belägg per lemma, som under samma förutsättning är 6,8 mot 2,8. Den kraftiga dominansen för formerna på *-s* innebär från stilistisk synpunkt att standardspråkets skriftliga och muntliga former nu står nära varandra i det aktuella avseendet.

En möjlig utveckling synes vara att den längre böjningsformen i icke högtidlig framställning väsentligen kommer att kvar-

stå i fall där en betydelseskiljande funktion konsolideras; jämför att *synes* i det studerade frekvensmaterialet står för drygt 25 % av beläggen på *es*-former. Härutöver kräver givetvis verbstammar på *-s* av formella skäl ändelsen *-es*: *upplöses* osv. Dessa stammar är emellertid så lågfrekventa att inget belägg förekommer i Frekvenser i svenskan 1965. Vad beträffar Körners påpekande (med instämmande citerat av Bergman) att former på *-s* särskilt ofta möter som predikatsverb till plurala subjekt torde mot bakgrunden av frekvenstalen i tabell 17 kunna anföras, att företeelsen är att betrakta som ett övergångsfenomen.

Historiskt sett är det vidare av intresse, att formerna på *-s* dominerar inom alla ifrågakommande verbböjningstyper. Schematiskt förefaller det svenska skriftspråkets utveckling i det aktuella avseendet i och med detta kunna tecknas på följande sätt (jfr Elias Wessén, Svensk språkhistoria. 1, § 211). I utgångsläget föreligger en tämligen strikt fördelning av ändelserna: starka verb och ja-verb tar *-s*, ia-verb och e-verb tar *-es*. Med reformationstiden inträder ett promiskuitetsskede, under vilket sambandet med ursprungskategorierna *upplöses*. Efterhand stabiliseras bruket; ett stadium uppnås, där *-es* är normalformen (utom i vissa icke-passiva funktioner). I innevarande skede utträngs den så av *-s* under inverkan från talspråket, där denna form tidigare segrat.

Inom preteritumböjningen tar jag upp de alternativformer som förtecknas i tabell 18. Av de där anförda enheterna är *la* och *las* belastade med interlingval homonymi (artikelformer). Vidare är *sas* homonymt med det skandinaviska flygbolagets namn. Vi kan emellertid konstatera att de längre formerna *lade(s)* och *sade(s)* håller sig väl. När det gäller formerna *skil(j)de(s)* dominerar de kortare, *j*-lösa varianterna. Tendensen är densamma i supinum, där det föreligger 3 belägg på *skilt(s)* mot 0 på *skiljt(s)*.

Det mest markanta kvantitativa draget hos de typer av allomorfi som har behandlats i uppsatsens andra hälft är den massiva övervikten för de kortare alternativen. Det är endast på enstaka punkter som de längre formerna väger tyngst: *råda*, *-kläda*, *effektivisera*, *sparar*, *skall*, *synes*, *lade*, *sade*. Resultatet avspeglar

den utveckling mot en ledigare och mera effektiv prosa som ägt rum under senare år.

TABELL 18. *Frekvenser för vissa alternativa preteritumformer.*

Preteritum aktivum				Preteritum passivum/deponens			
Alternativ 1	F	Alternativ 2	F	Alternativ 1	F	Alternativ 2	F
la	*104	lade	50	las	*22	lades	17
sa	72	sade	297	sas	*34	sades	15
skilde	9	skiljde	1	skildes	5	skiljdes	0

Undersökningens konsekvenser för beskrivningen av modern svenska kan enklast demonstreras med hjälp av några lexikografiska exempel. Gängse ordböcker har ju i själva verket en kvalitativ-kvantitativ karaktär i och med att de ger dels grammatiska (morfologiska) uppgifter, dels bruklighetsuppgifter. I lexikonerna givna data har hittills med nödvändighet väsentligen varit impressionistiska. En fastare grund kan läggas genom på frekvensundersökningar baserade studier.

Under (1) nedan ges kortare utdrag ur fyra artiklar i vår mest aktuella ordbok, Illustrerad svensk ordbok (tredje, reviderade upplagan 1964). Under (2) ges motsvarande uppgifter i en version som närmare motsvarar det faktiska språkbruket.

(1)

**beslut|a** pres.: *-ar* el. *-er*;  
imperf.: *-ade* el. *beslöt*;  
sup.: *-at* el. *-it*.

**draga** el. **dra**, pres.: *drager*  
el. *drar*; imperf.: *drog*; sup.:  
*dragit*; pres. pass. *drages* el.  
*dras*. De längre formerna  
tillhör skr.

**giva** el. **ge** *gav*, *givit* el.  
*gett*, pres. sing. *giver* el.  
*ger*; längre formerna skr.

**poängter|a** äv. **pointer|a** *-ade*

(2)

**beslut|a** pres.: *-ar* säll. *-er*;  
pret.: *beslöt* el. *-ade*;  
sup.: *-at* säll. *-it*.

**dra** äv. **draga**, pres.: *drar*  
säll. *drager*; pret.: *drog*;  
sup.: *dragit*; pres. pass. *dras*  
äv. *drages*.

**ge** äv. **giva** *gav*, *gett* el.  
*givit*, pres. *ger* säll. *giver*.

**poängtera** säll. **pointera** *-de*



I detta sammanhang aktualiseras också den teoretiskt intressanta frågan om relationen mellan språkbruket och den normativa lexikografin. En närmare diskussion av denna fråga faller emellertid utanför uppsatsens ram.

## Litteratur

- Allén, S., Frekvenser i svenskan 1965. 1-3. Göteborg 1967. [Duplik.]
- Allén, S., Studier över nusvenskans vokabulärsystem. Göteborg 1967. [Duplik.]
- Beckman, N., Svensk språklära för den högre elementarundervisningen. 9. uppl. Stockholm 1945.
- Bergman, G., Rätt och fel i språket. Stockholm 1962.
- Engdahl, S., Studier i nusvensk sakprosa. Några utvecklingslinjer. Uppsala 1962.
- Hallander, L.-G., Morfofonemiska förhållanden vid svensk verbböjning. (I: Nordisk sommeruniversitet 1954. Moderne viden-skab. Orientering og debat. 4, s. 144-147.)
- Illustrerad svensk ordbok. Red. B. Molde. 3. rev. uppl. Stockholm 1964.
- Johannisson, T., Det dolda stilmönstret. (I: Dagens Nyheter (A) 29/7 1961, s. 4.)
- Körner, R., Nutida tendenser i svensk språkutveckling. I. (I: Modersmållärarynas Förening. Årsskrift 1954, s. 98-113.)
- Lotz, J., Morphophonemic alternations in the Swedish strong verb (Ablaut). (I: Förhandlingar vid sammankomst för att dryfta frågor rörande svenskans beskrivning. 1, Stockholm 1963. Utg. av C.-C. Elert. Stockholm 1964, s. 55-61.) [Duplik.]
- Stähle, C.I., Verbböjningstyper och verbfrekvens i stockholmskt berättarspråk. (I: Förhandlingar vid sammankomst för att dryfta frågor rörande svenskans beskrivning. 3, Göteborg 1965. Utg. av S. Allén. Göteborg 1966, s. 5-27.) [Duplik.]
- Wessén, E., Svensk språkhistoria. 1. Ljudlära och ordböjningslära. 7. uppl. Stockholm, Göteborg & Uppsala 1965.

Öhman, S.E.G., Generativa regler för det svenska verbets fonologi och prosodi. (I: Förhandlingar vid sammankomst för att dryfta frågor rörande svenskans beskrivning. 3, Göteborg 1965. Utg. av S. Allén. Göteborg 1966, s. 71–87.) [Duplik.]

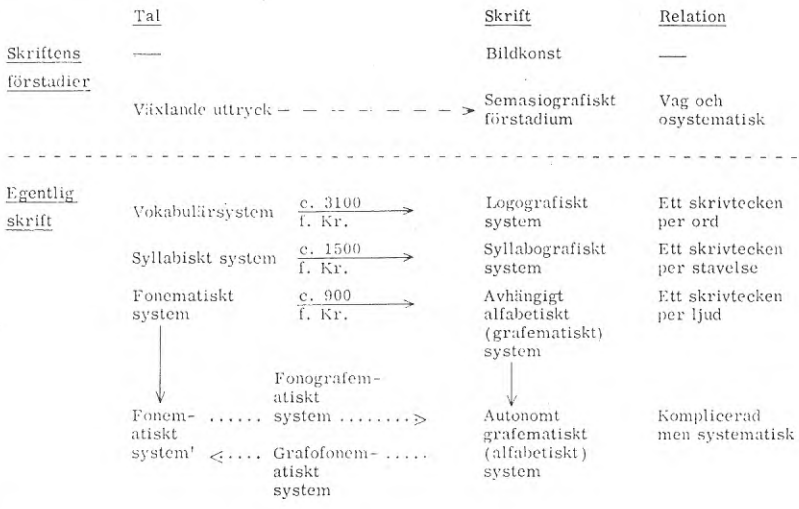
## Förhållandet mellan skrift och tal

Språk, språkvård och kommunikation. 1967.

Min utgångspunkt är den att språket har två viktiga manifestationsformer – skrift och tal. Jag ser som min uppgift att betrakta relationen mellan dessa huvudsakligen från teoretisk synpunkt och att dra fram vissa av betraktelsesättets konsekvenser för språkvården.

Låt mig i klarhetens intresse genast framhålla att jag med beteckningar som tal, talat språk, talspråk genomgående avser ljudmanifestationer och med beteckningar som skrift, skrivet språk, skriftspråk analogt grafiska manifestationer. En företeelse i skrivet språk som av någon anledning leder tankarna till talspråk – en form som *la 'lade'* eller en fras som *hör du du* – ryms alltså inte under beteckningen talspråk i denna skarpt avgränsade mening. Att inflytande från talspråk ligger bakom är en annan sak.

För att så börja från början: det är sannolikt att talet, som de flesta menar, är äldre än den egentliga skriften. Bevarade dokument tyder på en utvecklingshistoria för skriften, som schematiskt kan återges i enlighet med fig. 1. Det framgår där att skrift i egentlig mening är ungefär 5 000 år gammal. I sin äldsta form var den ett logografiskt system som reflekterade talets vokabulär-system. Kinesisk skrift är ett exempel. Syllabografiska system, stavelseskrift, framträder från omkring 1500 f.Kr. Japansk skrift är i väsentliga delar syllabografisk. Omkring 900 f.Kr. skapar så grekerna på grundval av nordsemitisk skrift ett alfabetiskt system, som avspeglar språkljuden. Här upphör i stort sett skrifthistorikernas intresse i principiellt avseende. Utvecklingen går emellertid vidare och leder i en rad fall till uppkomsten av autonoma grafematiska system. Nusvensk skrift är ett gott exempel. Reglerna för konvertering från skrift till tal och vice versa, det grafofonematiska respektive det fonografematiska systemet, är på detta stadium förhållandevis komplicerade.



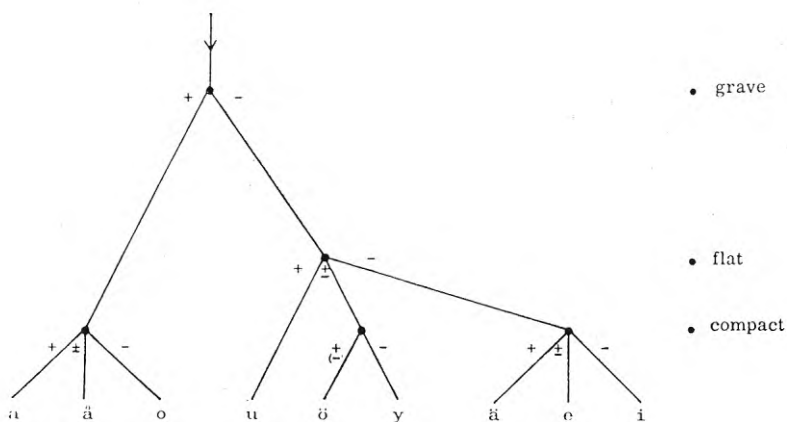
FIGUR 1. *Skriftens utvecklingshistoria.*

För säkerhets skull understryker jag att illustrationen är schematisk. Särskilt vill jag nämna den roll som blandformer mellan logografiska och syllabografiska system (s.k. logosyllabiska system) spelar. Den omtalade kretensiska skriftformen Linear B är exempelvis logosyllabisk. Någon närmare diskussion av den egentliga skriftens tre första utvecklingsstadier går jag av naturliga skäl inte in på utan hänvisar för deras vidkommande till gängse standardverk (Gelb 1963, Jensen 1958, Diringer 1962). I stället vill jag peka på några av de egenskaper som konstituerar skriften som ett autonomt system.

Av internt skriftspråkliga utvecklingar nämner jag då först framväxten av ett system av minuskler (i tryck: gemena typer) vid sidan av majuskulerna (i tryck: versaler). Härigenom kan man i skrift skilja mellan *proprier* och *icke-proprier*, mellan *Olle* och *olle*, *Lisa* och *lisa*, *Ekstam* och *ekstam*, *Sträng* och *sträng*, *Kina* och *kina*, *Person* och *person*, *Torson* och *torson*. I de båda sista paren motsvaras olikheten i skrift av en accent- respektive kvantitetsskillnad i tal. Redan i ytstrukturen skiljer skriften men inte

talet mellan fall som *den stolen är gjord av Alm* och *den stolen är gjord av alm*.

Vidare gäller att minuskelsystemet är uppbyggt av ett begränsat antal distinktiva faktorer, typologiska särdrag, på ett med det fonematiske systemet jämförbart sätt men analytiskt oberoende av detta. Som exempel visar jag i fig. 2 Gunnar Fants analys av nusvenskans vokalfonem (se Fant 1961, fig. 1–9) – modifierad så till vida att uppdelningen i korta och långa har utelämnats – och i fig. 3 en av mig utförd analys av nusvenskans autografem (i skrivmaskinsstil). Jag påpekar här att grafemens indelning i autografem och syngrafem inte grundas på deras motsvarigheter i tal utan på deras kombinatoriska egenskaper. Autografemen utgör, enkelt uttryckt, den grupp av grafem som är representerad i varje grafiskt ord (se närmare Allén 1965, s. 91 ff.). I fig. 3 motsvarar i fråga om diakritiska tecken + två stycken, ± ett och – inget, i fråga om ögla + stor ögla, ± liten ögla och – ingen ögla och i fråga om öglans höjdläge + övre och – nedre läge. Det framgår av figurerna att fonemen grupperar sig i tre bakre vokaler, tre rundade främre vokaler och tre orundade främre vokaler.



FIGUR 2. Nusvenskans vokalfonem.



grafemen i den där analyserade stilen inte har mer än en allograf var. Förhållandet belyses däremot av exempel i det följande, jämför bl.a. fig. 4.

I linje med att skriften uppvisar ett självständigt system av särdrag ligger att också allografsystemet är oberoende av allofonsystemet. I modern handstil förekommer t.ex. allografer som de som återges på de två första raderna i fig. 4, fakultativa varianter av grafemen <r>, <z> respektive <ä>. Dessa har inga unika motsvarigheter i tal. I stället har fonemet /r/ allofoner som främre och bakre *r* och /ä/ allofoner som de i *väder* och *värde*. I äldre handstilar förekom i fall som *sol*, *sko* och *nos* tre olika allografer av <s>, se den nedersta raden i fig. 4. Dessa hade alla samma motsvarighet i tal. Komplementärt distribuerade enheter som dessa tre s-allografer (av vilka i princip den tredje står finalt, den andra framför k och den första i övriga fall) kan klassifikatoriskt föras samman genom att de alla växlar med samma majuskulära enhet – S eller motsvarande. Denna skriftens egenskap kan sägas metodologiskt motsvara kombinationen distinktiva faktorer och komplementär distribution vid fonemanalys.

r	vär	z	zenit	ä	län
r	vär	z	zonit	ä	län
f	fol	g	gdo	s	nos

FIGUR 4. Allografer.

Systemen av grafem och fonem i nusvenskan avviker också markant från varandra. Enligt Gunnar Fants analys har vi, bortsett från skillnaden mellan korta och långa vokaler, de 31 segmentala fonem som i hans notation ges nedan, 9 vokalfonem och 22 konsonantfonem (se Fant 1961, s. 15).

/a e i o u y å ä ö/

/b d f g h j k l m n ng p r rd rl rn rt s sj t tj v/

Upplöses /rd rj rn rt/ i sina komponenter, fås 27 fonem. Upplöses också /ng sj tj/, fås 24 fonem (jämför Öhman 1966 a, s. 75 f.). Såvitt jag hittills har kunnat finna har vi 28 alfabetiska grafem, 9 autografem och 19 syngrafem:

<a e i o u y å ä ö>

<b c d f g h j k l m n p r s t v w x z>

Några strategiskt valda minimala par:

celt 'slags yxa'	ceder	sak	lys	vatt(-gröt)	son
kelt	seder plur.	sax	lyx	watt	zon

Den lågfrekventa typen q torde vara att betrakta som allograf av <k> med hänsyn till fall som dessa:

kvinna	kvantum	Dahlkvist
quinna	quantum (satis)	Dahlqvist

Av syngrafemen har <c w x z> ingen egen, enkel huvudmotsvarighet i fonemsystemet. Sett från andra hållet gäller motsvarande för de alternativt upplösbara fonemen.

I kombinatoriskt avseende är skillnaderna likaså avsevärda. Grafotaktiska sekvenser som exempelvis initialt <gj hj kj lj sj skj stj> saknar fonotaktiska motsvarigheter (jämför Sigurd 1965 b, s. 42). Unika för skriften är olikheter som *och-ock*, *Olsson-Ohlsson-Olson-Ohlson-Olzon-Ohlzon*. Här kan också nämnas skillnader i oppositionsbärande enheters position som t.ex. i följande fall:

son	skjutit	skott
zon	skrutit	skota

I det första paret upp bärs oppositionen i tal av vokalfonemen /å/ och /o/ i andra position, i skrift av syngrafemen <s> och <z> i första position. I de båda övriga fallen är det i tal inte fråga om minimala par.

I all korthet nämner jag ytterligare några olikheter. I skrift utvecklas tidigt olika typer av abbreviaturer, förkortningskonven-



tioner. Dessa kan för nusvenskans del exemplifieras med fall som TV, *oms* och tecknet % (ekvivalent med den alfabetiska sekvensen *procent*). I fråga om språkets gränssignaler, junkturerna, gäller att de för skriftens vidkommande är segmentala enheter – skiljetecknen hör dit – medan de för talets vidkommande är suprasegmentala, prosodiska, enheter. Fler olikheter kunde nämnas.

Relationerna på grafem–fonem-nivån (där överensstämmelsen likväl är störst) är alltså komplicerade. Själva existensen av de relationssystem som jag har pekat på visar, att vi har att göra med två självständiga manifestationssystem. Historiskt sett beror detta dels som jag nämnt på internt skriftspråkliga utvecklingar, dels på skriftens konservativa karaktär gentemot talets förhållandevis progressiva. Som exempel på en fonografematisk relation kan jag nämna en serie om ett fyrtiotal grafematiska motsvarigheter till fonemet /ʃ/:

< sju sked skjuta stjärna ryssja addition division diskussion schack  
 chef logi bagage suggestion reflexion motion apache shoppa  
 fashionabel crescendo religiös beige jalusi damejeanne mars for-  
 cera herrskap karls försjunka försköna förskjuta harts nerz väst-  
 göte värdshus gärdesgård Nordenskiöld Cederschiöld Swartz  
 Växjö (Bergslagen) >

Som exempel på en grafofonematisk relation kan tjäna en rad fonematiska motsvarigheter till grafemet <g>:

/gå:            jö:˘ra        ʃeni:˘        vaɲn        bokse:˘ra/

I fall som <gjuta ghanan> är <g> det ena ledet i en tvåledad grafemsekvens, vars motsvarighet är ett fonem. När det gäller konvertering från tal till skrift och omvänt se vidare Allén 1964 och 1965 (s. 38 ff.) samt Engström.

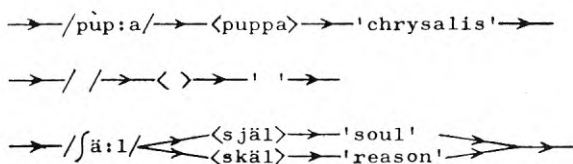
Studiet av det språkliga uttryckets substans bör i anslutning härtill bedrivas efter tre linjer: studium av tal, av skrift och av relationerna mellan dem. På talsidan får vi dels en fonetik, dels en fonem(at)ik, dels en fonotaktik (fonotax). På skriftsidan får vi parallellt härmed en grafetik (innefattande paleografi, epigrafi m.m.), en grafem(at)ik och en grafotaktik (grafotax). Vill man

använda termer som inte implicerar en strikt hierarkisk struktur, kan man välja fonologi respektive grafonomi. Den senare termen har föreslagits av Charles F. Hockett (Hockett 1951, s. 445) – termen grafologi är ju upptagen. Relationsstudiet omfattar dels en fonografem(at)ik, dels en grafofonem(at)ik. För den normativa aspekten på de båda relationerna används beteckningarna ortografi respektive ortofoni. En term saknas, en term vars non-existens ofta snedvrider lingvistiska resonemang. Jag syftar på en gemensam beteckning för fonologi och grafonomi. Att fylla denna lucka med en internationellt gångbar term vore en betydelsefull språkvårdsinsats.

På det grammatiska området, särskilt inom syntaxen, har man under senare år gjort avsevärda teoretiska landvinningar genom utvecklingen av den generativa grammatiken. När det gäller det språkliga uttrycket är diskussionen emellertid i traditionell anda ensidigt inriktad på talet (se t.ex. Chomsky 1964 och 1965). Det förefaller likväl uppenbart, att en grammatik som genererar inte bara fonologiskt utan också grafonomiskt välformade satser är mer adekvat än en som genererar enbart den förra typen. Ett par påpekanden som jag nyss gjorde bör därvid vara ett memento: dels har de distinktiva faktorerna, som i generativ fonologi spelar en viktig roll, inte samma strukturella uppgift inom grafonomi som inom fonologin, dels är det på fonem–grafem-nivån som tal och skrift kommer närmast varandra. Det bör tilläggas att restlös konvertering mellan uttryckssystemen på basis av enbart uttrycksformen inte kan ske (fall som substantiven *gjord–hjord–jord* med samma uttal, *loge* med två olika uttal). Som jag ser det leder detta till att man får anta en för uttryckssystemen gemensam underliggande struktur, en det språkliga uttryckets djupstruktur.

Skriftens autonomi i förhållande till talet leder till den konsekvensen för det språkliga tecknet, att detta är att betrakta som en treledad konstellation, uppbyggd av en fonematisk, en grafematisk och en semantisk storhet. Jag har funnit det vara smidigt att åskådliggöra det språkliga tecknet som ett flöde i enlighet med det översta exemplet i fig. 5. Den generella formen, endast

notationsbeteckningarna, ges på andra raden i figuren. Relationer mellan språkliga tecken kan enkelt demonstreras så som exemplet nederst i figuren visar. När schemat där genomlöpes två gånger efter alternativa vägar, produceras två homofona tecken.



FIGUR 5. Språkliga tecken.

Fonematisk och grafematisk form utgör viktiga led i ett mer omfattande system av centralt intresse, det som reglerar förhållandet mellan språkets ord. Detta system kan beskrivas på olika sätt. För egen del är jag böjd för att lägga stor vikt vid formella kriterier, särskilt med tanke på behovet av att kunna avgöra vad som är "samma ord" och vad som är "olika ord". Den grundläggande betydelsen av detta för kvantitativa undersökningar men också för grammatiken och lexikografin är uppenbar.

De ytterligare kriterier som man enligt min mening främst bör beakta är tillhörighet till ordklass och tillhörighet till böjningsklass. Eftersom begreppet ordklass är överordnat begreppet böjningsklass bör kriterierna ordnas på motsvarande sätt. De nu nämnda kriterierna är numrerade från 1 till 4 i fig. 6. Likhets-tecken i en kolumn innebär givetvis att jämförda enheter är lika med avseende på vederbörande kolumns faktor, olikhetstecken ( $\neq$ ) att de är olika. I fallen under A, de homomorfa, är både fonematisk och grafematisk form lika. I fallen under B, de semi-homomorfa eller partiellt homomorfa, är endera formen lika. Vi får dels fonohomomorfa och dels grafohomomorfa fall. I de återstående fallen, de heteromorfa, är både fonematisk och grafematisk form olika. Uppställningen kräver egentligen rätt utförlig motivering och diskussion, men i det här sammanhanget finns det bara utrymme för några korta kommentarer.

Termer		Fonematisk form 1	Grafematisk form 2
A. Homomorfi	Identitet	1 =	=
	Flexivisk variation	2 =	=
	Homonymi i in-skränkt mening	3 =	=
		4 =	=
B. Semihomomorfi	Grafematisk variation	5 =	≠
	Homofoni i in-skränkt mening	6 =	≠
7 =		≠	
8 =		≠	
	Fonematisk variation	9 ≠	=
	Homografi i in-skränkt mening	10 ≠	=
		11 ≠	=
		12 ≠	=
C. Heteromorfi	Fullständig uttrycksvariation	13 ≠	≠
	Heteronymi	14 ≠	≠
15 ≠		≠	
16 ≠		≠	

FIGUR 6. *Relationer mellan ord.*

Det är inte utan vidare klart att olikhet i fråga om uttrycksform eller böjningsklass är att betrakta som ordskiljande. Det kan vara fråga om variation – ett välkänt lingvistiskt fenomen – inom ramen för ett ”ord”, en lemmatisk klass. I den mån dylik växling förekommer är den att vänta i sådana fall inom grupperna A, B respektive C, där antalet föreliggande olikheter är minst, dvs. där

Ord- klass 3	Böjnings- klass 4	Lemmatisk klass 5	Exempel
=	=		<i>puppa; konst</i>
=	≠	=	<i>svan -ar -or</i>
=	≠	≠	<i>fil -ar</i> <i>fil -er</i>
≠	≠		<i>fara</i> subst. <i>fara</i> verb
=	=	=	<schal sjal>
=	=	≠	<set>
=	≠		<sätt>
≠	≠		>själ> -ar <skäl> -ø
			<hjärna> subst. <gärna> adv.
=	=	=	/ki: ʔo çi: ʔo/
=	=	≠	/fö: rslag/ /försla: ɡ/
=	≠		/lo:v/ -ar /lå:v/ -ø
≠	≠		/dälta/ subst. /dè:lta/ verb
=	=	=	<i>bror broder</i>
=	=	≠	<i>lepra; lepra</i>
=	≠		<i>spetälsk; föda</i>
≠	≠		<i>angrepp; lepra</i>
			<i>attack; skåp</i>
			<i>lepra</i>
			<i>läsa</i>

vi i de båda första grupperna har en olikhet, i den sista två olikheter. De fall det blir aktuellt att undersöka är alltså paren 2–3, 5–6, 9–10 och 13–14.

På medlemmar av en lemmatisk klass ställer vi nu två krav. Det första är att det i flexionsserien skall föreligga minst en formell konvergens i tal och/eller skrift. Detta krav tillgodoses i serier som följande:

svanar		fil	
svan		fil	
svanor		filer	
/ʃal/	/ʃa:lar/	/sät/	/sät/
<kilo>	<kilo>	<förslag>	<förslag>
bror			
	bröder		
broder			

Det tillgodoses däremot inte i ett fall som följande:

lepra		lepran
spetälska		spetälskan

På de enheter som uppfyller det första kravet ställer vi så krav nummer två, som innebär att i tal och/eller skrift divergerande former vid kommutationsprov skall visa fakultativ variation, dvs. kunna utbytas mot varandra utan att (den denotativa) betydelsen ändras. Kravet uppfylls i följande exempelfall:

svanar	<schal>	/ki:ˈlo/
svanor	<sjal>	/çi:ˈlo/

men inte i följande:

fil	<set>	/fö:ˈrslag/
filer	<sätt>	/fö:rsˈla:g/

Vi har sålunda funnit att det i fallen 2, 5, 9 och 13 rör sig om homolemmatiska former, i fallen 3, 6, 10 och 14 om heterolemmatiska former. De fyra slag av variation som har visat sig föreligga kan vi beteckna som flexivisk variation (*svan*), grafematisk variation (*schal*), fonematisk variation (*kilo*) respektive fullständig uttrycksvariation (*bror*).

Det inses vidare lätt att de båda uppställda kraven tillgodoses av fall nummer 1 men inte av fallen 4, 7–8, 11–12 och 15–16. Av dessa senare ingår 4, 7–8 och 11–12 tillsammans med de nyss påvisade heterolemmatiska fallen 3, 6 respektive 10 i vad som kan betecknas som homonymi, homofoni respektive homografi i inskränkt mening. Homonymi i vid mening omfattar även homo-

foni och homografi i inskränkt mening. Homofoni och homografi i vid mening omfattar även homonymi i inskränkt mening. Jämför på den här punkten Hesselman 1919 och se Dahlstedt 1964 b och 1966 samt Allén 1964 och 1965.

När det gäller oböjliga ord utgår givetvis kolumn 4, och den drar med sig fallen 2–3, 7, 11 och 15 samt det första kriteriet på homolemmatiska former (formell konvergens i flexionsserie). Med detta kriteriums försvinnande utgår också möjligheten att föra heteromorfa enheter (t.ex. *bara* och *endast*) till samma lemmatiska klass, dvs. alla heteromorfa oböjliga enheter är heterolemmatiska. Eftersom vidare olikhet i kolumn 3 när det gäller böjliga ord alltid medför olikhet i kolumn 4, blir olikhet i kolumn 3 genomgående ordskiljande. Denna konsekvens bör givetvis olikhet i kolumn 3 få också när det gäller oböjliga ord. I sådana fall slutligen där det föreligger likhet i kolumn 3 men olikhet i kolumn 1 eller 2 (t.ex. <och> – /åk/, /å/) blir kommutationsprovet, det kvarstående lemmatiseringskriteriet, avgörande.

Den referens till betydelseförhållandena som det andra kriteriet på en lemmatisk klass innebär är av formell karaktär så tillvida som den enbart gäller utbytbart i kontexter, frågan om likhet eller olikhet. Det är för övrigt samma kriterium som utnyttjas vid fonematisk och grafematisk analys. Tar man ordens semantiska valör över huvud i betraktande, finner man inom olika lemman en mycket växlande spännvidd: från det lämpligen som monosemt betecknade *puppa* till det polysema *konst* 'artistisk förmåga; gruvpump', från det förhållandevis monosema *schal/sjal* till det polysema *reflexion/reflektion* 'återkastning; eftertanke'. På motsvarande sätt förhåller sig olika lemman till varandra: från de homosema (eller synonyma – jämför Dahlstedt 1966, s. 54) *lepra* och *spetälska* till de lämpligen som heterosema betecknade *lepra* och *föda*.

Polysemi och homonymi är från den synpunkt jag här har anlagt artschilda företeelser och således klart avgränsade från varandra. Polysemi avser semantisk skiftning inom ett lemma, homonymi vissa typer av formell relation mellan olika lemman. Jag inskjuter i detta sammanhang, att nusvensk lexikografi i fråga om lemmatisering och därmed också i frågor som rör polysemi

och homonymi har en arbiträr prägel, i det att etymologiska överväganden, praktiska hänsyn (lättlästhet, utrymmesekonomi) och en säkerligen i inte oväsentlig grad på lexikografisk tradition grundad "språkkänsla" spelar in. En jämförelse mellan några exempel ur Illustrerad svensk ordbok (ISO), som i huvudsak följer Svenska Akademiens ordlista (SAOL), och fig. 6 ovan belyser skillnaderna:

		ISO (SAOL)	FIG. 6
<i>hög</i>	subst.; adj.	homonymi	homonymi
<i>djup</i>	subst.; adj.	polysemi	homonymi
<i>för</i>	(subst.; adj.; verb;) adv.; prep. (o. adv.); konj.	homonymi	homonymi
<i>innan</i>	adv.; prep.; konj.	polysemi	homonymi
<i>konst -er</i>	'artistisk förmåga; gruvpump'	polysemi	polysemi
<i>ask -ar</i>	'slags träd; mindre låda'	homonymi	polysemi

Under århundradenas lopp har det skett ett ömsesidigt givande och tagande, i växlande utsträckning, mellan svenskt tal- och skriftspråk. För utformningen av det talade riksspråket har skriftspråket spelat en avgörande roll. Som Elias Wessén framhåller har härvid uppläsning av texter (religiösa, officiella, litterära) betytt mycket (Wessén 1965, s. 152). Det talade högreståndsspråk som under 1600- och 1700-talet utvecklades i Mälardalen anpassade sig efterhand till skriftspråket. På bred front gör sig influenserna gällande från och med senare delen av 1800-talet, då analfabetismen efterhand försvinner som en följd av sociala förändringar, främst införandet av allmän skolplikt (1842). Vid den aktuella tiden uppvisar skriftspråket en hög grad av stabilitet och är följaktligen väl rustat att på allvar fungera som riksläkare. Den moderna pressen och den växande bokproduktionen förstärker skriftspråkets inflytande. Under de senaste decennierna har rikstalspråket i sin tur fått viktiga kraftcentra genom först radion och sedan televisionen. Indirekt, genom uppläsningsspråket, går dock etermedierna i viss mån också skriftspråkets ärenden.



Påverkningarna gör sig gällande på många olika strukturella nivåer. Gamla uttalsformer som t.ex. /çö`rka/ och /*(ut)å:* / har inte kunnat motstå trycket från skriftens <kyrka> respektive <*(ut)av*>. Skriftens *bh* blir i tal /be:˘hå/ och återkommer i skrift som *behå*. Skriftens *kem. tvätt* blir i tal efterhand /çe:˘mtvåt/ och återkommer i skrift som *kemtvätt*. Talspråksformer som /bli:/, /ha:/, /ta:/ tränger in i skriftspråket.

Det finita verbets enhetsformer (typen *vi går*) som länge tillhört talspråket har under vårt århundrade slagit igenom i skriftspråket. Å andra sidan sätter skriftspråkliga böjningsformer som <taket>, <gator>, <räknade> i allt högre grad sin prägel på det talade språket. Utbytet inom vokabulären måste också nämnas.

Syntaxen i spontant talspråk är genom sin brutna och impulsiva karaktär ofta rätt invecklad.

Det påminde om en ... ingen gurka men du vet dom där som vi åt var det inte i Athen.

Skriftspråkets syntax är också den ofta komplex, men av andra skäl.

Lärare inom närbesläktade vetenskaper, oavsett om de tillhöra samma universitet, så ock annan som meddelar undervisning vid universitet må efter särskilt av Kungl. Maj:t på förslag av universitetskanslersämbetet meddelat beslut bilda en ämnesgrupp med uppgift att handlägga sådana till dessa vetenskaper anknyttande ärenden, vilka enligt gällande föreskrifter eller särskilda beslut hänvisas till gruppen.

Tendensen på riksplanen tycks vara – och det är vad man skulle vänta – att talspråket går mot ökad stringens, skriftspråket mot större ledighet (jfr Åkermalm 1966, s. 42 f.).

Vilken riktning bör så språkvårdsarbetet ta med hänsyn till de synpunkter jag här har framfört? Jag vill markera några riktpunkter som förefaller mig väsentliga. De gäller dagens språkliga system i förhållande till dels gårdagens och dels morgondagens.

Förbindelselänken bakåt är skriftspråket. Jag betraktar kontinuiteten, tillgängligheten till äldre generationers dokument, som ett värde. Riktningen i denna dimension bör enligt min mening

därför vara att bibehålla skriftbildens. Dess funktionsduglighet är ju väl prövad. I synnerhet bör man undvika att manipulera med frekventa företeelser. Teoretiskt är varsamhet med den grafoniska formen motiverad av att denna är en integrerande del av det språkliga tecknet. Vi har som jag ser det inte anledning att beklaga att den junggrammatiskt inspirerade anstormningen mot skriftspråket under decennierna kring sekelskiftet har bedarrat och att den inte gav fler resultat än den gjorde (ändring av typerna *qvarn, jagt, jern, godt, hvem, af, lefva*). Ett par språkprov ur tidskriften *Nystavaren* kan förtjäna att anföras. Ruben G:son Berg skriver (5: 1, 1899, s. 17):

I rättstavningsfrågan som i andra frågor går ju utvecklingen jenom vågbärj och vågdalar, ock den dal, vari denna legat under några år, tycks nu vara slut, ock det tycks som om vi hölle på att singla oss uppför ett vågbärj ijänn. Även ett nytt avslag å petitionsjonen kämmer väl inte att skrämna någon, utan endast mana fram nya åtjäder, avsedda att åstadkåmma ett nytt nådigt sirkulär, som jenomför några större ändringar.

Hugo Pipping (4, 1892–97, s. 141):

Til sist anholder jag at fo ytra nogra ord til ursekt för at jag vogat framtrede med reformförslag, som sanolikt enligt mongas tanke ej kuna anat en öka den bestoende oredan. Min fasta övertygelse er, at intet av de i Sverje framsteldta nystavningsförslagen kan, eler ens bör alment omfatas i Finland, ty ala polega de os et altför drygt arbete.

Utan att gå in på någon närmare diskussion påpekar jag att utformningen av en ”judenlig” stavning som framgår av exemplen långt ifrån är självklar och att detta slags stavning i princip måste följa med när uttalet ändras.

I överensstämmelse med det tidigare sagda bör en riktpunkt framåt vara att på nivåer under den syntaktiska stödja talspråkets tendens att närma sig skriftspråket och att på den syntaktiska nivån stödja skriftspråkets tendens mot enkelhet och talspråkets tendens mot stringens. Längs samma syftlinje ligger uppgiften att verka för att nyheter i språket anpassas efter rådande system. Principiellt bör därvid fyra konformitetsvillkor obligatoriskt uppfyllas av ett nytt uttryck: konformitet med gängse bildnings-

och böjningsmönster och med såväl det grafonomiska som det fonologiska systemet. Därutöver bör det givetvis tillföra språket ett visst plusvärde – begreppsmässigt, internationellt, funktionellt, socialt – för att förtjäna rekommendation (se närmare Allén 1966). För att bara ta två exempel: formerna *mediatek* och *spray* bör förkastas på grund av att de inte är konforma med hänsyn till bildning respektive stavning. Formerna *mediotek* och *sprej* uppfyller alla konformitetsvillkoren, och eftersom de dessutom representerar begreppsmässiga plusvärden bör de rekommenderas.

En teoretiskt grundad språkvård som på detta vis ser samspelet mellan skrift och tal i tidsperspektiv fullgör enligt min mening på bästa sätt sin uppgift.

## Litteratur

- Allén, S., *Grafem och fonografem*, i: Förhandlingar vid sammankomst för att dryfta frågor rörande svenskans beskrivning. 2, Lund 1964. Utg. av B. Loman & B. Sigurd. Lund 1966 (Duplik.).
- Allén, S., *Grafematisk analys som grundval för textedering med särskild hänsyn till Johan Ekeblads brev till brodern Claes Ekeblad 1639–1655*. 1965.
- Allén, S., *Språkvård i teori och praktik*, i: Antennen. Personal-tidning för Sveriges radio, juni 1966.
- Berg, Ruben G:son, i: *Nystavaren* 5: 1, 1899.
- Chomsky, N., *Current Issues in Linguistic Theory*. 1964.
- Chomsky, N., *Aspects of the Theory of Syntax*. 1965.
- Dahlstedt, Karl-Hampus, *Om homonymi i nusvenskan*, i: Förhandlingar = Allén 1964 (Dahlstedt 1964 b).
- Dahlstedt, Karl-Hampus, *Homonymi i nusvenskan*, i: Nysvenska studier 45 (1965, tr. 1966).
- Diringer, D., *Skrift*. De olika skriftsystemen och deras utveckling. 1962.
- Engström, S., *Överföring från ortografisk till fonematisk svenska med hjälp av datamaskin*. Stockholm (KVAL) u.å. (Duplik.)

- Fant, Gunnar, *Structural Classification of Swedish Phonemes*, i: STL-QPSR 1/1961.
- Gelb, I. J., *A Study of Writing*. Rev. ed. Chicago & London 1963.
- Hesselman, B., *Ortografiska reformer i språkhistorisk belysning*, i: Språk och Stil 19 (1919).
- Hockett, Ch. F., (Rec. av) John De Francis, *Nationalism and Language Reform in China*, i: Language 27 (1951).
- Illustrerad svensk ordbok*. Red. B. Molde. 3. rev. uppl. 1964.
- Jensen, H., *Die Schrift in Vergangenheit und Gegenwart*. 2. Aufl. 1958.
- Loman, B., *Rökrunorna som grafematiskt system*, i: ANF 80 (1965).
- Nystavaren*. Tidskrift för rättstavningsfrågor. 4–5: 1, Uppsala 1892–99.
- Paine, R. M., *Preparation for Optical Character Recognition*, i: The Computer Journal 9: 3 (1966).
- Pipping, Hugo, i: *Nystavaren* 4, 1892–97.
- SAOL = *Svenska Akademiens ordlista över svenska språket*. 9 uppl. 1950.
- Wessén, Elias, *Svensk språkhistoria*. 1. Ljudlära och ordböjningslära. 7. uppl. 1965.
- Åkermalm, Å. *Modern svenska*. Språk- och stilfrågor. 1966.
- Öhman, Sven, *Generativa regler för det svenska verbets fonologi och prosodi*, i: Förhandlingar vid sammankomst för att dryfta frågor rörande svenskans beskrivning. 3, Göteborg 1965, utg. av S. Allén. Göteborg 1966. (Duplik.) (Öhman 1966 a.)

## Ordleksepisoden i Gutasagan och textdokumentens stemma

Språkdata 1967.

Det berättas i Gutasagan att befolkningen på Gotland ökade så mycket att landet inte kunde föda alla. Man lottade då bort var tredje invånare. De utstötta vägrade att lämna ön men drevs därifrån. Deras sökande efter nytt land förde dem efterhand långt från hemön. Sagan förtäljer:<sup>1</sup>

so fierri foru þair at þair / quamu til griclanz[s], þar baddus þa / ir byggias firir af grica kunungi, vm. / ny. oc niþar kunungr þann lufafi þa- / im Oc hugði at ai miþ ann manarþ / wari. siþan gangnum manafi wildi / hann þaim bort wisa En þair aniz / suararv þa at ny oc niþar / wari .e. oc. e. Oc quapu so sir wa // ra lufat. þissun þaira wipratta quam / firir drytningina vm siþir þa segði / han. Minn herra kunungr þu lufafi / þaim byggia vm ny. oc niþar. þa ir / þet .e. oc .e. þa matt þu ai af þaim / taka. So bygpus þair þar firir. / oc enn byggia. oc enn hafa þair suint / af waru mali

Gutarna kom alltså ända till Grekland. De bad kungen om att få slå sig ner i landet *vm ny oc niþar* 'under ny och nedan'. Kungen fattade detta som ett uttryck för 'en månad' och gav dem sitt tillstånd. Som det sedan visade sig inlade emellertid gutarna däri betydelsen *e oc e* 'alltid'. Genom drottningens ingripande avgjordes tvisten till gutarnas förmån, och de bosatte sig i landet.

Till sitt allmänna innehåll är denna episod fullt klar: gutarna framställer sin begäran i form av en ordlek, och listen lyckas. Däremot har ingen allmänt omfattad explikation givits av det centrala avsnitt, vari kungens uppfattning av innebörden i uttrycket *vm ny oc niþar* anges:

kunungr þann lufafi þaim Oc hugði  
at ai miþ ann manarþ wari.

<sup>1</sup> Se Pipping 1905–07, 63: 8–19. Se beträffande första hälften av textstycket även figur 1 nedan, rad 9 ff.

Man antar genomgående att stället är korrumpert.

Gutasagan föreligger i en enda forngutnisk handskrift, Cod. Holm. B 64 från c. 1350, förvarad i Kungl. biblioteket, Stockholm. Det är sålunda dess läsart som efter Hugo Pippings edition har givits ovan. I växlande språkformer föreligger emellertid större eller mindre partier av sagan i fem dokument från de närmast följande tre hundra åren.<sup>1</sup> Tidsmässigt grupperar sig dokumenten i tre äldre, från tiden 1401 till c. 1500, och två yngre, från 1600-talet. Med hänsyn till språkformen grupperar de sig i ett på fornsvenska, tre på danska och ett på tyska. Några data i översiktsform om de fem dokumenten:

Cod. Holm. B 65	ty. 1401	Kungl. biblioteket, Stockholm
Cod. Holm. D 2	fsv. c. 1450	Kungl. biblioteket, Stockholm
N:o 408 8:0	da. c. 1500	Ny Kongelig Samling, Köpenhamn
N:o 2414 4:0	da. c. 1625	Gammel Kongelig Samling, Köpenhamn
Peder Syv, Betenkninger om det Cimbriske Sprog	da. c. 1650 (tr. 1663)	

Som uppställningen antyder är de fyra första dokumenten handskrivna, det femte (vari endast ett fragment av Gutasagan ingår) tryckt.

Av de fem texterna innehåller fyra – alla utom D 2 – det här aktuella avsnittet. Formuleringarna är följande.<sup>2</sup>

B 65	das irloubete en der könyng vnde meynte das is eyn mande were.
408	Ok han loffw[adhe] them thz: ok meýntith icka ware mera en monadh
2414	deth thillstede Kongenn thennom och mentte att thedt haffde weritt Enn manede tiidt
Syv	hvilket og Kongen bevilgede, tænkedes, at det varede ikkun en maanets tiid;

<sup>1</sup> Se Ljunggren 1959, s. 9 ff.

<sup>2</sup> Schlyter 1852, s. 163, resp. Ljunggren 1959, s. 20, 31, 90.

Alla avviker som synes väsentligt från formuleringen i B 64, och ingen av dem ger något osökt uppslag till klargörande av denna den äldsta bevarade versionen.

De som har kommenterat textstället i B 64 har som nämnts betraktat det som fördärvat och har följaktligen föreslagit restitutioner av olika slag. C. J. Schlyters konjektur lyder: *at ai maira þen ann manapr wari*.<sup>1</sup> Denna står strukturellt nära Johan Hadorphs översättning från 1687:<sup>2</sup> 'ey längre än en Månads tijd warade'. Den ligger också nära läsarten i 408 – *icka ware mera en monadh* – vilken dock inte var bekant för Schlyter (och f. ö. inte heller för Pipping). På samma linje går Lis Jacobsen, som emellertid under hänvisning till 408 föreslår en liten modifiering: *at ai maira þann manapr vari*.<sup>3</sup> Hon menar, utan att gå in på detaljer, att förvanskningen av ett *maira þann* till *miþ ann* lätt låter sig förklara i en handskrift som likt B 64 innehåller en mängd felaktigheter. Häremot har K. G. Ljunggren invänt, enligt min mening med rätta, att det behövs en verklig förklaring till hur en dylik misstyding gått till.<sup>4</sup> Någon sådan har inte lämnats: Holmbäck-Wesséns försök till rekonstruktion av förloppet – *maira þan > mi þan > miþ ann* – ges ingen motivering och kan inte betecknas som övertygande.<sup>5</sup> Schlyter-Jacobsens konjektur är ju emellertid innehållsmässigt acceptabel, och den läggs till grund för Holmbäck-Wesséns översättning:<sup>6</sup> 'Konungen lovade dem det och trodde, att det ej var mera än en månad.'

Den andra huvudlinjen i konjekturdiskussionen utgår från Hugo Pipping. Han föreslår i uppsatsen *Nya gotländska studier*, att urtexten har haft *at ain niþ ann manapr wari*.<sup>7</sup> Formen *niþ* uppfattar han som femininum snarare än som neutrum pluralis. Som översättning ger han: 'att ett nedan var lika med en månad'. "Men dette var merkelig tænkt af Kongen", invänder Lis Jacob-

<sup>1</sup> Schlyter 1852, s. 95 not 11.

<sup>2</sup> Hadorph 1687, s. 48.

<sup>3</sup> Jacobsen 1911, s. 57.

<sup>4</sup> Ljunggren 1959, s. 92.

<sup>5</sup> Holmbäck-Wessén 1943, s. 303.

<sup>6</sup> Holmbäck-Wessén 1943, s. 292.

<sup>7</sup> Pipping 1904, s. 14 f.

sen.<sup>1</sup> "En 'Nid' er ... en halv Maaned, og sikkert vilde man ikke oprindeligt have skrevet, at en Nid, Næde var en Maaned, hvorimod man ypperligt kunde have skrevet, at 'Ny og Næde var en Maaned'", fortsætter hon. Hon hänvisar också till att ingen av de övriga handskrifterna pekar i riktning mot Pippings konjektur. Holmbäck-Wessén instämmer i att denna konjektur från betydelsens synpunkt är betänklig.<sup>2</sup>

K. G. Ljunggren ansluter sig emellertid för sin del till Pippings förslag och vidareutvecklar det.<sup>3</sup> Läsningen av *ai miþ* som *ain niþ* betecknar han som ett från paleografisk synpunkt elegant tillämpande av lectio-difficilior-principen, klart överlägset Schlyters förslag. De semantiska betänkligheterna söker han avvisa genom att anföra en parallell. En betydelseförskjutning till 'månad' är inte mer påfallande, menar han, än den förskjutning till betydelsen 'måne' som enligt Feilberg är iakttagen i Vendsyssel: *no æ nēj n gon åp. gon nir*.<sup>4</sup> För egen del finner jag inte parallellen avgörande. Jag är snarare böjd för att betrakta den Pipping-Ljunggrenska konjekturen som en nödlösning.

Diskussionen kring textstället avspeglas i Noreens fornsvenska läsebok. I första upplagan 1892–94 går Adolf Noreen på Schlyters linje: *at ai maira þan ann manapr wari*.<sup>5</sup> I andra upplagan 1904 för han in Pippings konjektur: *at ain niþ ann manapr wari*.<sup>6</sup> I tredje upplagan 1921 återgår han till Schlyters förslag men nu med Lis Jacobsens modifikation (i normaliserad form): *at ai maira þan manapr wari*.<sup>7</sup> Denna version står oförändrad kvar i Erik Noreens utgåva 1932.<sup>8</sup> Detsamma är fallet i den av Sven Benson 1954 ombesörjda upplagan, men i kom-

<sup>1</sup> Jacobsen 1911, s. 56.

<sup>2</sup> Holmbäck-Wessén 1943, s. 303.

<sup>3</sup> Ljunggren 1952, s. 59: 1959, s. 91 f.

<sup>4</sup> Feilberg s. v. *næde*.

<sup>5</sup> Noreen 1892–94, s. 38.

<sup>6</sup> Noreen 1904, s. 38.

<sup>7</sup> Noreen 1921, s. 38.

<sup>8</sup> Noreen 1932, s. 153.



mentaren betecknas också det Pipping-Ljunggrenska restitutionförslaget som tänkbart.<sup>1</sup>

Vi återvänder nu till det forngutniska textställets egen form. Vi tar som vår uppgift att närmare granska läsningen av sekvensen *ai miþ ann* och att belysa den med hjälp av i första hand handskriftens övriga text för att därigenom söka komma till rätta med avsnittet.

Om translittereringen av de enskilda graferna torde ingen tvekan kunna råda, se figur 1, rad 13 (figurerna 1 och 2 återgår på vid Kungl. biblioteket utförda fotokopior av handskriften). Däremot finns det anledning att ta mellanrummet mellan *niþ* och *ann* i betraktande. Det framgår omedelbart att det är avsevärt mindre än exempelvis både det före *miþ* och det efter *ann*. Detta kan tyda på att vi har att göra med ett fall i gränzonen mellan spatium och vad jag har föreslagit att kalla sektur, avbrott inuti ett grafiskt ord.<sup>2</sup> Ser man på avståndet mellan *þ* och *a* i vad Pipping uppfattar som grafiska ord, finner man en ganska vid variation. Ett med det här diskuterade fallet jämförbart exempel har vi mitt på översta raden i figur 1: *Siþan*. Under sådana omständigheter är det belysande och tänkvärt att den gutniska textens förste utgivare Johan Hadorph – och efter honom Karl Schildener – har läst *at ai miþann manapr wari*.<sup>3</sup> (De översättningar som ges kan emellertid i båda fallen snarast betecknas som omskrivningar: 'ey längre än en Månadz tijd warade' respektive 'dass es nicht[!] einen Monat lang wahren würde'.) Det synes mig sålunda finnas fog för att alternativt läsa stället *ai miþann*.

I handskriften är det vidare så, att finalt *-nn* inte så få gånger står för väntat *-n*. Pipping anför bl. a. *hann* för *han*, *enn* för *en*, *þann* för *þa en*.<sup>4</sup> Han anför också exempel på medialt *-nn-* för väntat *-n-* som i *synnum*. Samma växling avspeglas i Lis Jacobsens restitutionförslag *ai maira þann* jämfört med Adolf Noreens *ai maira þan*. Om orsaken till växlingen – ”felet” enligt

<sup>1</sup> Noreen-Benson 1954, s. 153, 178.

<sup>2</sup> Allén 1965, s. 61.

<sup>3</sup> Hadorph 1687, s. 48; Schildener 1818, s. 107.

<sup>4</sup> Pipping 1905–07, s. XXXIV med not 2.

44

þann þeir þepau. Si þau foru þair bo  
 m. i. f. i. rōvna. Oc bygþus þaz firu  
 þaz gati. þair ai sic vþi haldir vti  
 foru. rama oþ wipi auitland si haitu  
 dagapi. Oc bygþus þaz firu. Oc giez  
 þu burgama si em. Gims þaz gati  
 þair oc ai sic haldir. Vtan foru vþ at  
 wacu þi si haitu dýna. Oc vþ gmu  
 rýza land so fieri foru þair at þair  
 quamu til griclanz. þaz baddus þa  
 u byggiaz firu af grica kumngi. vni  
 ny. Oc mþaz kumngi þau lufapi þa  
 m. Oc hugþi at ai mþauu manapi  
 wari. si þan sauguu manapi wile  
 þann þann þort wisa. Þu þair anz  
 suaru so fir wana at ny oc mþaz  
 wari. e. oc. e. Oc quapu so fir wa

FIGUR 1. Cod. Holm. B 64, blad 44 recto.

Pippings formulering – nämner Pipping i första hand att skrivtecknet i något skede av traderingen har både utskrivits och försetts med förkortningstecken, i andra hand att ouppmärksamhet har föranlett förväxling av snarlika ord som *hann* och *han*. Fortfarande utan att ta ställning till betydelsen kan vi alltså konstatera, att den aktuella sekvensen i B 64 med hänsyn till skrivvanorna i dokumentet kan återge ett ursprungligt *ai miþan*.

Hur skall så den ursprungliga lydelse som texten i B 64 pekar mot tolkas? Nyckeln till tolkningen – och till de tidiga missförstånden vid traderingen – ligger enligt min mening i betydelsen hos ordet *ai*. Fram till Pippings uppsats Nya gotländska studier 1904 räknade man med att *ai* i forngutniskan endast var negation. I denna uppsats påpekade emellertid författaren, att det finns ett ställe i Gutalagen, där sammanhanget förbjuder att *ai* tolkas som 'ej'.<sup>1</sup> Passusen lyder (se figur 2, rad 3–5):<sup>2</sup>

Biaupi bot manni en / hann orkar sen atmeli ir ut gangit / oc lati at meli ai milli.

Schlyter ger *a milli* men har naturligtvis sett i-et, och han påpekar i en not att *a* kan läsas *ai*.<sup>3</sup> Med hänsyn till B-handskriften av lagen, som har *lati è atmelj at millan*, gör Pipping den bedömningen att det är fråga om ett adverb *ai* 'alltid'. Mellan detta och *e* 'alltid' föreligger då, framhåller han, samma förhållande som mellan fvn. *ei* 'alltid' och *æ* 'alltid'. Jag finner Pippings bedömning helt riktig.

Enligt min mening är det samma *ai* 'alltid' som möter i Gutasagens *ai miþann*. Liksom i det förra fallet *ai* följs av adverbet *milli* '(där)emellan', följs det här av adverbet *miþann* 'under tiden, därunder'. Den diskuterade meningen tolkar jag alltså:

[Där bad de grekernas kung att få bo under ny och nedan.]  
Kungen lovade dem (det) och hade i tankarna, att det därunder alltid var (en) månad.

<sup>1</sup> Pipping 1904, s. 3.

<sup>2</sup> Pipping 1905–07, 17: 10–12.

<sup>3</sup> Schlyter 1852, s. 27 med not 41.

ogurumstan man þa dragi so bandu  
 su gutum. Þu han a boli sielfs fins  
 tor a guilandi. **B**iaupi tor man en  
 han ozkar sen atueh u ut sangit  
 oc lat at mehi ai milli. oc biaupi þey  
 suaz om þey az. Þu hynn sei maðz ofe  
 siemdz þau en han taki at fyrsta sin  
 ni þegar þupit verþ. **Þ**a en han wil  
 ai taka at fyrsta sinni. oc ai at andz  
 þa taki han at þupia. sen þey az ni  
 vt gangin wil han ai þa taka. þa  
 skal biera a þing firri alla lyþi. Taki  
 en þaz en han wil. þa en han ai wil  
 þa rapin allir lyþir firri fe en han sei  
 maðz o sacar. **Þ**a en hynn callaz ai ve  
 ra frest lipin. en han callaz vara lipin  
 su vereldi biaupe. þa hafir þam rittuþ  
 su lipit callaz niþ þinn þarra maná  
 schiellu su niþ hann at garþi wau.

FIGUR 2. Cod. Holm. B 64, blad 12 recto.

Adverbet *mipan* är i B 64 också belagt på ett ställe i Gutalagen.<sup>1</sup>

Kirchju þa syki / hann þau andra mipan firi þy so et ai / magu þair  
beþir til annir at fara

*Mipan* är även belagt som konjunktion.

I ordleksepisoden syftar *mipann* på en explicit angiven tidsperiod, *vm ny oc nipar*. Det får därför en mera bestämd betydelse, 'därunder, under den tiden', än den tämligen obestämda 'under tiden' som det har i belägget från Gutalagen. Den senare nyansen är den vanligare i fornnordiskan,<sup>2</sup> och det kan därför vara befogat att anföra ett par paralleller från fornsvenska texter till användningen i Gutasagan. Den första hämtar jag från Fem Moseböcker på fornsvenska.<sup>3</sup>

Halder haarith sin förra lit / tha ær arret myrkt oc ey meer  
inswnkith æn annath om kring / tha skal han innelykkia i siw  
dagha Wæxa mædhan malata mark / tha skal prester döma han  
oskiran oc fra manna sammæru

I detta exempel syftar alltså *mædhan* på uttrycket *i siw dagha*.

Det förtjänar särskilt att framhållas, att handlingen i den föregående satsen inte åsyftas. Detta framgår om man prövar att ersätta ordet med en omskrivning i anslutning till föregående sats. Man får:

Växer tecknen till spetälska under de sju dagarna ...

Helt analogt är det i Gutasagan:

... att det alltid var (en) månad under ny och nedan.

I följande fall, hämtat från Siælinna thrøst, syftar *Mædhan* på *j xl dagha*:<sup>4</sup>

<sup>1</sup> Pipping 1905–07, 16: 26–27.

<sup>2</sup> Se exemplen hos Söderwall s. v. *māpan*, Fritzner s. v. *medan*.

<sup>3</sup> MB 1 A, s. 362 (se även MB 1 B, s. 281).

<sup>4</sup> ST, s. 7.

Moyses gig oppa berghit til wan herra / oc bleff ther j xl. / dagha  
Mædhan ofwer gaff folkit sin gudh oc herra

Traderingen av textstället utvisar som nämnts att svårigheter med tolkningen förelegat redan under medeltiden. Omformuleringar ligger då nära till hands. De behöver inte vara så omfattande som i vissa av handskrifterna. Även en sådan detalj kan komma i fråga som *-nn* i *mīþann* (*mīþ ann*). Gemineringen kan ha tillkommit under något stadium på grund av attraktion från *manapr* (jfr Schlyters konjektur *ai maira þan ann manapr*).

Förklaringen till att den ursprungliga läsarten så tidigt spårat ur och så konsekvent motstått tolkningsförsöken är som jag ser det följande. Forngutniskans *ai* 'alltid' är i det språkskede vi kan iaktta ett obsolet språkligt tecken, vars uttrycksform *ai* normalt förbinds med innehållet 'ej' och vars innehåll 'alltid' normalt förbinds med uttrycksformen *e*. De kvantitativa förhållandena i det samspel som B 64 kan sägas utgöra belyser klart omständigheterna. Pippings ordförteckning upptar 192 belägg på *ai* 'ej' och 19 belägg på *e* 'alltid' men som antyts bara 1 belägg på *ai* 'alltid'.<sup>1</sup> Att interpretens associationer går från uttrycksformen *ai* till innehållet 'ej' är under rådande förhållanden fullt naturligt. Historiskt ligger det säkerligen till så som Pipping tänkte sig det, att B 64:s *ai* 'alltid' är en rest av ett promiskuitetsstadium, i vilket såväl *ai* som *e* betydde såväl 'ej' som 'alltid' liksom fallet är med fornvästnordiskans *ei* och *æ*.<sup>2</sup>

Med hänsyn till det nu sagda framstår den här givna läsningen *ai mīþann* såsom *lectio difficilior*. Enligt min tanke ger oss textstället en glimt av ett för länge sedan förflutet. Det är helt följderiktigt att den centrala passus det gäller är allittererad.

Ordleksepisoden är intressant också från den synpunkten att den spelar en framträdande roll i diskussionen om textdokumentens stemma. De viktigaste bidragen till denna diskussion har lämnats av Lis Jacobsen och K. G. Ljunggren.

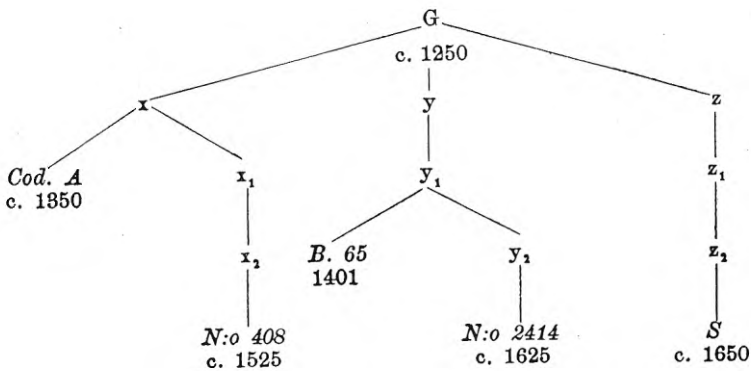
<sup>1</sup> Pipping 1905–07, glossaren s. 3, 17.

<sup>2</sup> Pipping 1904, s. 3.

Lis Jacobsen anser att stemmat har tre grenar, av henne be-tecknade x, y och z (se figur 3).<sup>1</sup> Till x-grenen för hon B 64 och 408, till y-grenen B 65 och 2414 och till z-grenen Syvs version (S). Grenarna x och y är ansatta på mycket goda grunder. För-fattarinnan uttalar själv tvivel i fråga om z-grenen.<sup>2</sup> Hon bygger sitt resonemang om att S representerar en från x och y skild gren på det textställe som i B 64 innehåller ordformen *eluiſt* 'för-häxat':<sup>3</sup>

þa war gut / land so eluiſt at þet daghum sanc Oc / natum war uppi.

Också y-grenen har uppenbarligen haft *eluiſt*, eftersom B 65 har den härav förvanskade formen *eliuch*. Däremot förutsätter enligt hennes mening motsvarande uttryck i S, *lidet til at see*, ett *oliuſt*. Denna form är en konjektur som framförts av Hadorph och upptagits av Schlyter.<sup>4</sup>



FIGUR 3. *Stemmat enligt Lis Jacobsen.*

<sup>1</sup> Jacobsen 1911, s. 66 ff.

<sup>2</sup> Jacobsen 1911, s. 68 not 1.

<sup>3</sup> Pipping 1905–07, 62: 3–4.

<sup>4</sup> Hadorph 1687, registret s. v. *eliuſt*. Schlyter 1852, s. 93 not a.

Antagandet om en särskild z-gren betecknas av Nils Tiberg såsom knappast ens sannolikt.<sup>1</sup> Också K. G. Ljunggren riktar kritik mot tanken. Såsom ett av flera alternativ pekar han på att den ganska fritt arbetande upphovsmannen till S kan ha omskrivit ett oförstått *elui*st som *lidet til at see* på basis av kontexten, av vilken ju framgår att Gotland av olika skäl var osynligt dygnet runt.<sup>2</sup> För min del ansluter jag mig till denna uppfattning med den modifikationen, att omskrivningen (i en form jämförbar med den i S) kan ha uppkommit på ett tidigare stadium i traderingen. På detta sätt erhålles en naturligare förklaring, synes det mig, än genom konstruktionen med ett *oli*ust, vare sig denna form såsom hos Lis Jacobsen antas ha tillhört S:s förlaga eller såsom alternativt hos K. G. Ljunggren antas ha föresvävat upphovsmannen till S (som tolkning av ett obegripligt *elui*st).

Något annat argument än fallet *elui*st stöder sig Lis Jacobsen trots K. G. Ljunggrens uppgift härom inte på när det gäller z-grenen. Själv tar Ljunggren som naturligt är också hänsyn till ordleksepisoden. Att han härvid är hänvisad till att laborera med alternativen *ai maira þan(n)* och *ain niþ ann* som utgångspunkter vållar märkbara svårigheter. Sammanfattningsvis finner han emellertid, att tillvaron av z-grenen inte kan anses styrkt.<sup>3</sup> Slutatsen synes mig nu ha vunnit i säkerhet.

Var skall så S placeras, sedan tanken att den representerar en särskild gren skrinlagts? Utan att väga skälen för x- respektive y-grenen mot varandra resolverar K. G. Ljunggren, att S ”på sakens nuvarande ståndpunkt” kan inplaceras på x-grenen (se figur 4).<sup>4</sup>

Underlaget för en bedömning är begränsat till fallen *ai miþann* och *elui*st. I det förra fallet har S *at det varede ikkun en maanets tiid*. Utan att utnyttja iakttagelsen i sin diskussion har Lis Jacobsen påpekat, att formuleringen står nära motsvarigheterna i B 65 och 2414, *das is eyn mande were* respektive *att thedt haffde weritt Enn manede tiidt*.<sup>5</sup> De ovan flera gånger citerade formule-

<sup>1</sup> Tiberg 1946, s. 30.

<sup>2</sup> Ljunggren 1959, s. 91.

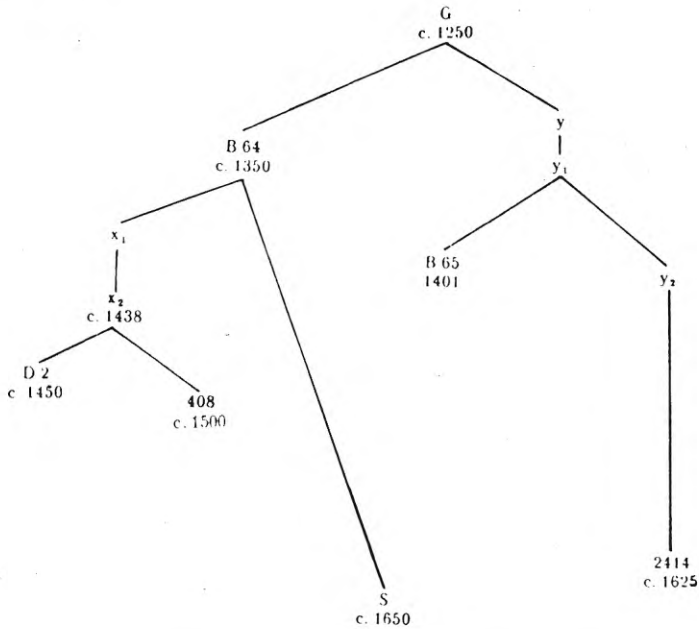
<sup>3</sup> Ljunggren 1959, s. 97.

<sup>4</sup> Ljunggren 1959, s. 91, 97.

<sup>5</sup> Jacobsen 1911, s. 57.



ringarna i B 64 och 408 avviker radikalt härifrån (D 2 saknar stället). Förhållandena på denna punkt talar alltså för att S tillhör y-grenen.



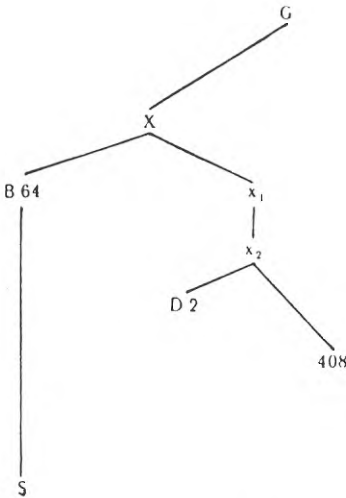
FIGUR 4. Stemmat enligt K. G. Ljunggrens alternativ I.

När det gäller motsvarigheterna till B 64:s *elulist* är läget det, att B 65 har *eliuch* och 2414 *øde*.<sup>1</sup> Det senare ordet är att betrakta som en fri omskrivning, och den står i semantiskt avseende mycket nära formuleringen i S: *lidet til at see*. Som ett positivt argument för S:s tillhörighet till y-grenen kan likheten emellertid inte användas, emedan textstället saknas i x-grenens versioner D 2 och 408. Förhållandena på denna punkt kan alltså beskrivas så, att de inte talar mot att S tillhör y-grenen. Tas då båda punkterna i betraktande, blir slutsatsen att S på grundval av det knappa material som föreligger bör hänföras till y-grenen.

<sup>1</sup> Se Ljunggren 1959, s. 29.

Var på y-grenen S skall placeras är osäkert. Dess formuleringar tyder snarast på att den är sidoordnad i förhållande till B 65 och 2414.

Utöver placeringen av S skiljer sig K. G. Ljunggrens stemma enligt alternativ 1 från Lis Jacobsens främst däri, att B 64 hos Ljunggren intar en överordnad position i förhållande till 408, medan de båda versionerna hos Lis Jacobsen är sidoordnade. Den återstående skillnaden utgörs av att Ljunggren har placerat in den av Lis Jacobsen obeaktade D 2 i stemmat. Placeringen bygger på en rad textställen och förefaller mig invändningsfri.



FIGUR 5. X-grenen enligt K. G. Ljunggrens alternativ 2.

Alternativ 2 hos Ljunggren (se figur 5) är, bortsett från S och D 2, identiskt med Lis Jacobsens stemma. Huvudskälen till den utformning av x-grenen det här gäller är två. Det första avser läsarna i ordleksepisoden. Som tidigare framgått menade Lis Jacobsen att den ursprungliga texten är bäst bevarad i 408 (*icka ... mera en*). B 64 kan då inte utgöra ett led i 408:s traderingshistoria. Slutsatsen är enligt K. G. Ljunggren inte nödvändig. Han formulerar sitt ställningstagande så, att han efter noggrant

övervägande alltjämt måste hålla det för fullt möjligt, att 408:s version kan ha uppkommit som ett försök att klara ett till sin språkliga form dunkelt men rent innehållsmässigt ganska tydligt ställe (nämligen ett *ain niþ ann*).<sup>1</sup> I ordleksepisodens nya ljus upplöses Lis Jacobsens argument, och K. G. Ljunggrens specialförklaring blir överflödig. Lydelsen i B 64 återger den ursprungliga avfattningen och övriga bevarade versioner – den i 408 lika väl som de i B 65, 2414 och S – är att betrakta som omskrivningar.

Det andra huvudskälet återstår likväl, och det är det som K. G. Ljunggren med åberopande av Lis Jacobsen direkt anger som motivering till alternativ 2. Det textställe det här gäller är beskrivningen av Awair Strabain. B 64 har:<sup>2</sup>

þy et hann war senieldir. Oc / fiel kunungur so sum  
saghur af ganga

408 har:<sup>3</sup>

Thÿ Man war sneller ok selff konnigh / som sawngth  
goor aff

Man torde vara överens om att B 64:s *fiel kunungur* reflekterar ett ursprungligt *fielcunnugr* 'mångkunnig', motsvarande fornvästnordiskans *fiþlkunnigr*. Vad Lis Jacobsen säger om relationen mellan 408 och B 64 på den här punkten är väsentligen, att formen *konnigh* i 408 leder en till att tro att en förlaga har haft *kun(n)ung(u)r* liksom B 64 men att detta inte är nödvändigt.<sup>4</sup> Också K. G. Ljunggren konstaterar att förhållandena talar för ett beroende, men han vill "inte helt förneka möjligheten" att överensstämmelsen "i enlighet med Lis Jacobsens uppfattning" härrör från en (för x, och B 64) gemensam källa.<sup>5</sup>

<sup>1</sup> Ljunggren 1959, s. 96, 92.

<sup>2</sup> Pipping 1905–07, 64: 12–13.

<sup>3</sup> Ljunggren 1959, s. 21.

<sup>4</sup> Jacobsen 1911, s. 61.

<sup>5</sup> Ljunggren 1959, s. 96: jfr s. 41.

Till det sist anförda är att säga att Lis Jacobsen enligt min läsning inte menar, att överensställmelsen härrör från en gemensam källa. Hon torde i stället mena, att denna har haft *kunnugr* (denna form ansätter hon som tillhörande B 64:s förlaga). Detta *kunnugr* har så lett till missförstånd på båda håll. För 408:s vidkommande har missförståndet också omfattat *fiol*. Det är då anmärkningsvärt att medan x-grenens båda huvudlinjer oberoende av varandra skulle ha drabbats av detta missförstånd, y-grenen har gått helt fri. Som Lis Jacobsen nämner har B 65 *wol sprechende vorsichtich vnde manches dynges wys* och 2414 *weltalinnde, forsünlig och ij alting wis och klogh*.<sup>1</sup> Med hänsyn tagen till samtliga bevarade versioner synes mig den rationella förklaringen vara, att 408:s *sneller ok selff konnigh* – liksom D 2:s *snielder oc cloker konunger*<sup>2</sup> – återgår på B 64:s *senieldir. Oc fiel kunungr*.

De få ställen som i övrigt är att beakta när det gäller förhållandet mellan  $x_1$  (och därmed 408) och B 64 är av det slaget, att de med K. G. Ljunggren låter sig väl förklara som i  $x_2$  företagna korrigeringar av B 64.<sup>3</sup> När nu enligt ovan det första och som det framkommit i grunden viktigaste av huvudargumenten för alternativ 2 av x-grenens utformning bortfallit och det andra visat sig vara mindre ett argument för än den osannolikare av två tänkbara förklaringar till relationen mellan de inbegripna textställena, finner jag det inte rimligt att längre bibehålla alternativ 2.

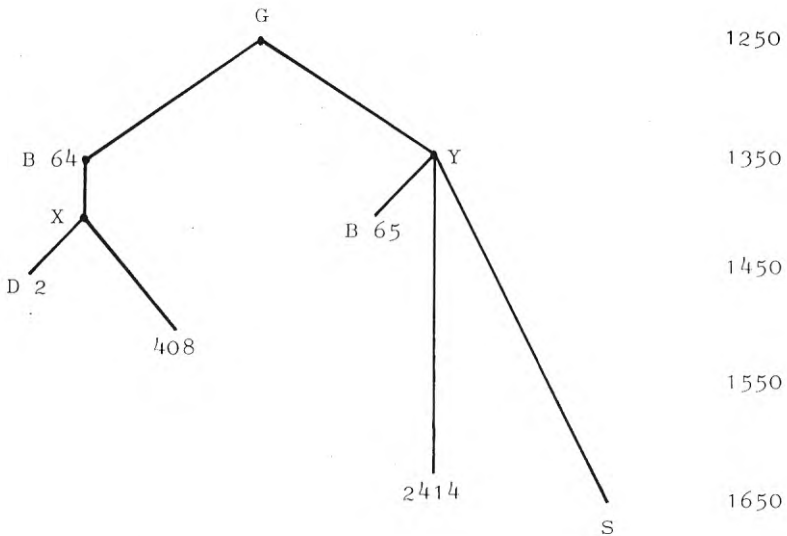
Det stemma jag har kommit fram till återges i figur 6. Det är att betrakta som ett principstemma, där antalet mellanled är ospecificerat. Orsaken till att jag har givit det denna form är inte, att jag har ansett att vägande skäl saknas för av Lis Jacobsen och K. G. Ljunggren ansatta mellanled. Orsaken är i stället att jag har velat ge stemmat en enhetlig abstraktionsnivå, och med hänsyn till svårigheten att med någon säkerhet specificera mellanleden på y-grenen har jag valt en något högre abstraktionsnivå. Vidare följer

<sup>1</sup> Jacobsen 1911, s. 61.

<sup>2</sup> Se Ljunggren 1959, s. 15.

<sup>3</sup> Ljunggren 1959, s. 95 f.

av stemmats karaktär att den angivna tidsskalan i detalj är giltig endast för de bevarade textdokumenten.



FIGUR 6. Stemmat enligt bedömningen i föreliggande uppsats.  
X och Y betecknar ett ospecificerat antal mellanled.

Sammanfattningsvis kan följande punkter anföras. Den livligt diskuterade mening i Gutasagans ordleksepisod som behandlas i undersökningens förra del återges av Hugo Pipping efter B 64 (jfr figur 1, rad 12–14):

kunungr þann lufapi þaim Oc hugpi  
at ai niþ ann manar wari.

Textstället återfinns i fyra senare dokument: B 65, 408, 2414 och S. Alla har väsentligt avvikande formuleringar. Stället har genomgående betraktats som korrumperat. Konjekturen *ai maira þan(n) (ann)* stöds av C. J. Schlyter, Lis Jacobsen och Elias Wessén, konjekturen *ain niþ ann* av Hugo Pipping och K.

G. Ljunggren. Adolf Noreen har i tur och ordning stött den första, den andra och den första igen. Skrivvanorna i B 64 gör det emellertid troligt att stället med Johan Hadorph skall läsas

at ai miþann manapr wari.

Skrivvanorna gör det också troligt att *ai miþann* återger ett ursprungligt *ai miþan*. Jag tolkar här *ai* som 'alltid' och *miþan* som 'under tiden därunder' och hänvisar till parallellen *ai milli* 'alltid emellan' i Gutalagen (se figur 2, rad 5). Uttrycket *ai miþann* syftar på *vm ny oc niþar* i föregående mening. Förklaringen till de problem som stället redan tidigt vållat är att *ai* i B 64 normalt betyder 'ej' (192 belägg), medan de båda fallen av *ai* 'alltid' är en rest av ett äldre promiskuitetsstadium.

I undersökningens senare del behandlas konsekvenserna för textdokumentens stemma av ordleksepisodens explikation. Lis Jacobsen ansätter tre grenar, x, y och z (se figur 3). Mot ansättandet av en z-gren, endast representerad av S, har kritik riktats av Nils Tiberger och K. G. Ljunggren. Den senares slutsats att tillvaron av en z-gren inte kan anses styrkt synes mig nu ytterligare befast. Ljunggren placerar S på x-grenen (se figur 4). Materialet för en bedömning är knappt – fallen *elulist* och *ai miþann* – men talar enligt min mening för y-grenen. Ljunggren ger vidare två alternativa utformningar av x-grenen. Den andra är (bortsett från S och D 2) identisk med Lis Jacobsens (se figur 5). Utformningen bygger i huvudsak på fallen *ai miþann* och *fiel kunungur*. I ordleksepisodens nya ljus framstår det enligt min mening inte som rimligt att längre bibehålla detta alternativ. Ett principstemma ges i figur 6.

## Bibliografi

### Otryckt litteratur

Cod. Holm. B 64, Kungl. biblioteket, Stockholm.

### Tryckt litteratur

- Allén 1965 Allén, S., Grafematisk analys som grundval för textedering. Göteborg 1965.
- Feilberg Feilberg, H. F., Bidrag til en ordbog over jyske almuesmål. 2. Kjøbenhavn 1894–1904.
- Fritzner Fritzner, J., Ordbog over Det gamle norske Sprog. 2. 2. uppl. Kristiania 1891.
- Hadorph 1687 Gothlandz-Laghen På gammal Göthiska/ Med en Historisk Berättelse vid ändan... Utg. av J. Hadorphius. Stockholm 1687.
- Holmbäck-Wessén 1943 Svenska landskapslagar. Tol. o. förkl. för nutidens svenskar av Å. Holmbäck & E. Wessén. 4. Skånelagen och Gutalagen. Stockholm 1943.
- Jacobsen 1911 Jacobsen, L., Gamle danske Oversættelser af Gutasaga. (ANF 27 (1911), s. 50–75.)
- Ljunggren 1952 Ljunggren, K. G., Isl. *nið*, nysv. *nedan* (månfasen) och deras närmaste släktingar. (ANF 67 (1952), s. 54–69.)
- Ljunggren 1959 Ljunggren, K. G., En fornsvensk och några äldre danska översättningar av Gutasagan. Lund 1959. (SFSS h. 224.)

- MB 1 A Svenska medeltidens bibel-arbeten. Utg. av G. E. Klemming. 1. Stockholm 1848. (SFSS 9:1.)
- MB 1 B Fem Moseböcker på fornsvenska enligt Cod. Holm. A 1. Utg. av O. Thorell. Uppsala 1959. (SFSS 60.)
- Noreen 1892–94 Altschwedisches lesebuch mit anmerkungen und glossar. Utg. av A. Noreen. Halle 1892–4.
- Noreen 1904 Altschwedisches lesebuch mit anmerkungen und glossar. Utg. av A. Noreen. 2. uppl. Stockholm & Halle 1904.
- Noreen 1921 Altschwedisches lesebuch mit anmerkungen und glossar. Utg. av A. Noreen. 3. omarb. uppl. Stockholm 1921.
- Noreen 1932 Fornsvensk läsebok. På grundval av A. Noreens Altschwedisches Lesebuch utarb. av E. Noreen. Lund 1932.
- Noreen-Benson 1954 Fornsvensk läsebok. Utg. av E. Noreen. 2. bearb. uppl. utg. av S. Benson. Malmö 1954.
- Pipping 1904 Pipping, H., Nya gotländska studier. (Göteborgs högskolas årsskrift 1904, s. 3–24.)
- Pipping 1905–07 Gutalag och gutasaga jämte ordbok. Utg. av H. Pipping. København 1905–07.
- Schildener 1818 Guta-Lagh das ist: Der Insel Gothland altes Rechtsbuch. Utg. av K. Schildener. Greifswald 1818.
- Schlyter 1852 Codex iuris Gotlandici. Utg. av C. J. Schlyter. Lund 1852. (Corpus iuris Sueo-Gotorum antiqui. VII.)



- ST Siælinna thrøst. Utg. av S. Henning.  
Uppsala 1954. (SFSS 59.)
- Söderwall Söderwall, K. F., Ordbok öfver  
svenska medeltidsspråket. 2. Lund  
1891–1900.
- Tiberg 1946 Tiberg, N., Utvandringsberättelsen i  
Gutasagan. (Gotländskt arkiv 1946, s.  
16–47.)

## Ordforskaren och datamaskinen

Göteborgs Handels- och Sjöfarts-Tidning 31 mars 1964.

Den i en tidigare artikel (21/1) nämnde ordboksredaktören F. de Tollenaere, verksam i Nederländerna, har under de senaste åren med stor grundlighet satt sig in i den lexikologiska automatiseringens problem och metoder. Ett resultat av hans studier föreligger sedan förra året i form av ett synnerligen läsvärt arbete om de nya vägarna inom lexikologien (*Nieuwe wegen in de lexicologie*, North-Holland Publishing Company, Amsterdam). Boken är som synes skriven på nederländska. En utförlig resumé på engelska ingår. Boken innehåller också en värdefull bibliografi och en ganska stor planschbilaga.

Tollenaere intar samma ståndpunkt som den lexikologiska automatiseringens pionjärer, nämligen den, att det språkliga materialets alltmer växande omfattning gör de traditionella arbetsmetoderna otillräckliga. Den nödvändiga precisionen och fullständigheten kan endast vinnas genom automatisering. Och införandet av datamaskinen som ett hjälpmedel inom lexikologien innebär en omvälvning av samma storleksordning som den utnyttjandet av boktryckarkonsten på sin tid åstadkom.

Invändningen att automatiseringen kommer att medföra lexikologiens "avhumanisering" avvisar Tollenaere med det fullt riktiga påpekandet, att det är den rutinmässiga och tråkiga delen av arbetet som maskinen utför, den del som varken kräver intelligens eller fantasi. Planeringen, programmeringen och den slutliga bearbetningen och bedömningen åvilar lexikologen.

Det bör nämnas att författaren i termen lexikologi inlägger en vid betydelse. Till lexikologien bör enligt hans mening räknas semantiken (betydelseläran), morfologien (formläran), ordbildningsläran, etymologien m.m. Termen lexikografi å andra sidan bör reserveras för begreppet ordboksarbete. De nämnda exemplen på arbetsuppgifter visar, att det långt ifrån endast är ordboksredaktörer som kan dra nytta av lexikologisk automatisering. Och de rika möjligheter som maskinerna redan erbjuder

kommer att ökas högst avsevärt, när inom en sannolikt mycket nära framtid en fotoelektrisk textläsare har färdigutvecklats. Man kan då överföra en tryckt text direkt till en kod, som är läsbar för datamaskinen, och slipper det än så länge nödvändiga manuella stansandet av hålremsa. Behovet av ett språkvetenskapligt databehandlingsinstitut vid varje universitet kommer att accentueras i allt högre grad, så mycket mer som lexikologien bara är ett av de språkvetenskapliga områden, där datamaskintekniken äger tillämpning. Ett sådant institut synes böra fungera som ett självständigt organ i intimt samarbete med den akademiska datacentralen, de språk- och litteraturvetenskapliga institutionerna och universitetsbiblioteket.

Det här förda resonemanget om de svenska förhållandena utgick från Tollenaeres diskussion om ett lexikologiskt centrum i lärdomsstaden Leiden. Avslutningvis tar författaren upp en fråga, som han återkommer till flera gånger i sin bok och som även framskymtat ovan. Det gäller de olika hållningar till automatiseringen som man möter. Den första är den avvisande hållningen, grundad på ett postulat om oförenligheten av kulturvetenskaperna (geesteswetenschappen) och automatiseringen. Tekniken utgör enligt denna uppfattning ett hot. En andra hållningstyp är den blinda tron på maskinens allmakt. Man trycker på en knapp och får händerna fulla av de mest lysande resultat. Mellan dessa två ytterlighetsståndpunkter står den tredje, den nyktra och sakliga hållningen. Dennas företrädare ställer frågan: vad har automatiseringen att bjuda? Varken fruktan för maskinen eller blindhet med avseende på dess begränsning får spela in. Det är denna kritiskt positiva inställning som författaren vill bibringa sina läsare. När han framhåller, att människan alltid kommer att vara viktigare än maskinen, är han inte den förste som gör det. Men synpunkten förtjänar att hållas aktuell.

Till sin huvuddel är boken en förträfflig översikt över vad som hittills utträttats på det aktuella området. Huvudintresset riktar sig åt framför allt tre håll: Italien, Frankrike och USA.

I Italien har R. Busa sedan omkring 1950 varit sysselsatt med mekanisering av lexikologiskt arbete, tidigast med utnyttjande av hålkortsmaskiner, senare även datamaskiner. Busas mål är att upprätta en fullständig ordlista och en fullständig konkordans till Thomas ab Aquinos verk. Det är här fråga om en textmassa av 13 miljoner ord. Vid det lexikologiska centrum i Gallarate som Busa skapat har man också behandlat Döda havsrullarna. Till dessa har producerats olika typer av ordlistor och en konkordans, som har haft betydande textkritiskt värde.

I Frankrike är B. Quemada den ledande forskaren på området. Vid det av honom ledda vokabulärinstitutet i Besançon är man bl.a. sysselsatt med ett ordhistoriskt projekt, *Répertoire historique du vocabulaire français*.

På amerikansk botten är det framför allt det arbete som bedrivs under ledning av S.M. Parrish som drar uppmärksamheten till sig. Parrish är huvudansvarig för den serie av konkordanser som produceras vid Cornell University. Hittills är konkordanser över Matthew Arnolds och W.B. Yeats' dikter publicerade, och de kommer att följas av sådana över Ben Jonson, Emily Dickinson m.fl.

I detta sammanhang kan två något vilseledande tryckfel i resuméns USA-avsnitt påpekas. Det sägs där, att amerikanerna Josselson och Ellison redan 1935 utförde lexikologiskt arbete med hålkorts- respektive datamaskinutrustning. Det rätta året är i båda fallen 1953. De första datamaskinerna tillkom på 1940-talet. I bokens nederländska text ges de rätta årtalen.

Tollenaeres bok är emellertid mer än en forskningsöversikt. Den är också ett inlägg i en fråga av stor vikt för språkvetenskapen. Författaren pläderar för upprättandet av ett centrum för automatiserad lexikologi i Leiden. Han förutsäger också att ett eller flera sådana centra kommer att upprättas i varje land.

För Sveriges vidkommande kan man utan överdrift påstå, att behovet av lexikologisk automatisering redan gjort sig gällande. Låt oss se lite närmare på vad som med datamaskinens hjälp kan uträttas på detta område. Som nämnts kan man till en given text upprätta en konkordans jämte ordlistor av olika slag. Viktiga

typer av sådana är t.ex. frekvensordlistor, baklängesordlistor (av stor betydelse för bl.a. ordbildningsstudier), morfemindex (d.v.s. index över ordens minsta betydelsebärande enheter), ordlängdsindex (orden ordnade efter antalet ingående tecken) och rimordlistor. En del av dessa ordlistor har direkt praktisk betydelse för språkvårdsarbetet. I samband med konkordansen kan kontexterade (d.v.s. med vederbörande textställes lydelse försedda) ordlappar produceras och ställas till lexikografernas förfogande. Andra typer av uppgifter är exempelvis etablerandet av samlade förteckningar över ordbeståndet i viktiga grupper av lexika, t.ex. äldre lexika och fackområdeslexika.

## Massmedium, etermedium, mediotek

Svenska Dagbladet 6 oktober 1964.

Nästan dagligen kan man läsa eller höra olika former av ordet *massmedium*. Också andra bildningar vari ordet *medium* i någon form ingår liksom det enkla ordet självt blir allt vanligare. Det är en viktig grupp av ord och begrepp inom vårt samhälles språk som det här är fråga om. I det följande skall jag behandla några av dessa och därvid göra vissa rekommendationer.

Ordet *massmedium* är så nytt i svenskan, att det inte upptas i vare sig Svenska Akademiens ordlista (1950) eller Illustrerad svensk ordbok (1955). I Lilla uppslagsboken (1958) finns det med. De överföringsmedel ordet avser växlar något, men i stort sett kan man säga att det utgör en samlingsbeteckning för de fyra enheterna radio, TV, press och film.

Ordets pluralböjning är vacklande. Den ytterst från latinet inlånade formen *-media* konkurrerar med den efter inhemskt mönster bildade formen *-medier* (jfr *stipendier* m. fl.). De nämnda ordböckerna ger endast pluralformen *medier*, och detsamma gäller Elias Wesséns *Våra ord* (1961). Från språkvårdssynpunkt är denna form att föredra, inte minst därför att den ger möjlighet att göra en formell distinktion mellan obestämd och bestämd form: *medier* – *medierna*. Den stundom hörda invändningen att pluralformen *medier* leder tankarna till förmedlare med andevärlden väger allt lättare i ett samhälle, där massmedierna spelar allt större roll och spiritismen allt mindre.

Som förled i sammansättning uppträder *massmedium* i formerna *massmedie-* och *massmedia-*. Efter mönster av t. ex. *stipendiefond*, *studiecirkel* bör vi stanna för den förstnämnda formen, alltså tala om *massmedieforskning*, *-kultur*, *-utbud*.

På grund av den nära samhörigheten mellan de båda massmedierna radio och TV är det av behovet påkallat att ha en gemensam beteckning för dem. En sådan har tills nyligen saknats. I titeln till ett radioprogram 17/9 1963 förekom emellertid

ett ord med den önskade betydelsen : "Radio, TV och vi – om etermedierna av Ivar Ivre och Håkan Unsgaard".

Det första belägget på ordet *etermedium* är sannolikt följande: "vi har vant oss vid att lita till Sveriges Radios initiativ i alla hänseenden, så snart det gäller våra etermedier" (Gunnar Hallingberg i Folkuniversitetet 1963:3). För ordets tillkomst redogör Hallingberg, som är biträdande sekreterare i radioutredningen, i sin bok Kultur för miljoner (1963), där han genomgående använder ordet. Mitt förslag till författaren grundade sig givetvis på det förhållandet, att radio och TV till skillnad från press och film enligt gängse språkbruk kommunicerar med mottagarna via etern.

Ordet ansågs av Ivar Ivre i artikeln Kulturmedier i etern (Tiden 1964:1) ha en framtid i våra språkvanor, och att det har vunnit spridning skall visas med en del ytterligare exempel: "debatt kring etermedia" (Ingemar Lindblad i DN 27/12 1963), "kontakt med etermedierna från insidan" (L. R. i GHT 3/1 1964), "etermediernas historia" (Leif Furuhammar i SvD 14/1 1964), "båda etermedierna" (Henrik Sandblad i GHT 18/3 1964), "globalt samarbete mellan etermedia" (TT 23/6 1964), "radion är den enskildes etermedium" (Rolf Lundgren i Sveriges Radios Årsbok 1963).

*Etermedium* ansluter sig naturligt till en grupp andra sammansättningar med samma förled. Äldre exempel är *eterhav(et)*, *-rymd*, *-svängning*, *-våg*. Nyare fall är *eterkrig*, *-röst(er)*, *-vana*, *-värld(en)*.

I detta sammanhang kan framhållas, att Gösta Forsströms påpekande (Perspektiv 1964:6), att teorien om den substanslösa etern sedan länge är vederlagd, inte är något vägande motargument. *Etermedium* är att betrakta som ett bildligt uttryck av en typ som språket är fullt av och inte klarar sig utan. Forsströms eget förslag till beteckning för etermedierna, *telemedier*, förefaller ge mig felaktiga associationer, i det förleden *tele-* säkerligen främst förknippas med *telefon* och *telegraf*. När Forsström talar om "vårt svenska teleföretag" och därmed menar Sveriges Radio är man heller inte villig att följa honom.

Andra språk har enligt tillgängliga uppgifter knappast något att ge. Danskan, norskan och isländskan torde sakna ord för begreppet, franskan likaså. I engelskan, främst dess amerikanska variant, reder man sig i varje fall i vissa sammanhang genom att använda termen *broadcasting* i utvidgad betydelse. Åt samma håll tenderar tyskans *Rundfunk* (*Funk*). Att på motsvarande sätt töja innehållet i orden *radio* och *rundradio* är emellertid knappast vad man önskar, så fast knutna till ljudradioverksamheten som de är i gängse språkbruk.

Det skulle kunna hävdas, att vi egentligen borde tala om *etermassmedium*, detta för att utesluta eventuella tankeförbindelser med radiotelefonering. I ord av denna långa typ har språket emellertid en tendens att göra sig kvitt mellanledet (s.k. ellipsis), såsom t.ex. i *kamgarn* för *kamullsgarn*. Utelämnandet av ett ordled på detta sätt får inga konsekvenser för betydelsen. Ännu mindre behöver icke-förekomsten av *-mass-* i *etermedium* spela in. Vi vänjer oss helt enkelt vid att "etermedium" är ett "massmedium" underordnat begrepp.

Möjligheten att anknyta förleden i det fristående ordet *etermedium* till bedövningsmedlet eter finns givetvis. I en löpande framställning är risken för missförstånd emellertid praktiskt taget obefintlig. Och det lilla köttben till rikets lustigkurrar som termen innefattar kan etermediernas företrädare lugnt kosta på sig.

Bakom den påtalade användningen av formen *-media-* i *massmediaforskning* m.fl. ligger förmodligen användningen av pluralformen *massmedia*. Sannolikt är det på motsvarande sätt pluralformen *media* som ligger bakom det av Skolöverstyrelsen använda nya ordet *mediatek* (*Skolvärlden* 1964:25). Dettas betydelse är "samling av en skolas hjälpmedel ('böcker, AV-materiel och programmerat material)". Belägg på ordet föreligger från 1962.

Nyordet i fråga är uppenbarligen bildat efter den numera relativt stora mönstergrupp som omfattar ord som *bibliotek*, *diskotek*, *filmotek*, *fonotek*, *glyptotek*, *kartotek*, *mikrotek*, *pinakotek*. Dessa ord har det etymologiska draget gemensamt, att de betecknar (lokal för) en samling av det som respektive ords första led markerar (*apotek* och *hypotek* är av annan typ). Ordets senare led *-tek* återgår på ett grekiskt ord med betydelsen "förvaringsrum".



Av de nämnda orden är *bibliotek* med månghundraåriga anor det äldsta i svenskan. De båda relativt ovanlig orden *pinakotek* "tavelsamling" och *glyptotek* "skulptursamling" uppträder i vårt språk från 1800-talets första del. De övriga är 1900-talsord med *kartotek* som det tidigaste, belagt från tiden kring första världskriget. Av de återstående orden upptas inget i Svenska Akademiens ordlista (1950), *diskotek* och *filmotek* däremot i Svensk uppslagsbok samma år. *Filmotek* nämner uppslagsboken även 1949 under *cinematek*, ett ord som det att döma av Bibliotekstjänsts förteckning *Bibliotekstermer* (1963) nu har trängt undan. Dock kan detta någon gång påträffas med hänsyftning på utländska nationalarkiv för film, särskilt det franska (*La cinémathèque française*). I förbigående sagt kunde en smidigare beteckning för detta senare begrepp, *kinotek*, erhållas med utgångspunkt antingen från *kino*, danskans och tyskans kortform av *kinematograf*, eller från ett efter klassiskt mönster bildat *kinematotek*. Orden *fonotek* och *mikrotek* är nytillskott från de senaste åren. Det arkiv vid Kungl. Biblioteket som nu heter Nationalfonoteket betecknas i riksbibliotekariens ämbetsberättelse för år 1959 som *ljudarkiv*, i berättelsen för år 1960 som *fonotek*. Termen *mikrotek* begagnas av Sveriges Radio sedan ungefär samma tid som beteckning för dess mikrofilmsarkiv.

Mönstergruppens förleder slutar regelmässigt på *-o* (det sporadiska *cinematek* undantaget). Tendensen är så stark, att vi även i ord som *kartotek* och *filmotek*, vilkas förleder återgår på latinets *charta* respektive engelskans *film*, har *o*-form. Mot denna formella enhetlighet bryter nu *mediatek*. Såvitt jag kan se finns det därför all anledning att istället använda formen *mediotek*. För formen *mediotek* talar också sammansättningar som den språkvetenskapliga termen *mediopassivum*.

Aktivt språkvårdsarbete kan inte inriktas på ett konsekvent tillämpande av en gång för alla fastlagda principer. Bedömning får i stället ske från fall till fall genom avvägning av för- och nackdelar. En princip bör emellertid enligt min uppfattning i det längsta fasthållas: strävan efter att ansluta nybildningar till lämpliga existerande mönstergrupper. Härigenom kan vi i görligaste mån undvika att ytterligare komplicera vårt redan tillräckligt

komplicerade språkliga system. Låt oss sålunda i enlighet med denna grundsats använda formerna *massmedier*, *etermedier* och *mediotek*.

## Indelningen av nusvenskans pronomen

Arkiv för nordisk filologi 73. 1958.

Frågan om vad man för nusvenskans vidkommande skall mena med pronomen och hur man lämpligen skall indela denna ordklass har aktualiserats genom en uppsats om attributiva pronomen av Bengt Loman i ANF 71, 1956, s. 218 ff. I föreliggande uppsats, vartill Lomans arbete givit den närmaste impulsen, diskuteras indelningen av ordklassen i dess helhet på basis av ett vidgat ordmaterial och från delvis andra utgångspunkter.

Ehuru ingen tillfredsställande definition av termen pronomen har givits, råder likväl en viss enighet om vilka tecken som är att föra till ifrågavarande ordklass. Sådana är t. ex. *jag, hans, denna, sådan, vars, någras, andra* och *vem*. Ordklassen kan alltså i stora drag bestämmas genom uppräknig. Av framlagda definitionsförslag skall här endast det i SAOB publicerade granskas. För den tidigare diskussionen i frågan hänvisas framför allt till Adolf Noreen, *Vårt språk V*, 1904, s. 63 ff., och till Viggo Brøndal, *Ordklasserne*, 1928, s. 20 ff.

I SAOB bestäms *pronomen* på följande sätt (s. v. P 2041, arket tryckt 1954): 'benämning på en viss grupp av ord (utgörande en särskild ordklass) med substantivisk l. adjektivisk funktion o. med ett blott i vissa avseenden fixerat, efter ordets användning skiftande betydelseinnehåll, som användas i stället för det ord (i sht substantiv l. adjektiv) l. uttryck som utgör namnet på l. den egentliga beteckningen för ett visst begrepp, med syftning på detta begrepp, ersättningsord'. Denna definition är inte adekvat men kan möjligen betraktas som ett närmevärde. Den framhåller tre egenskaper som utmärkande för pronomenen. 1. De har substantivisk eller adjektivisk funktion. Detta gäller normalt. Dock kan vissa pronomen ha adverbial funktion, t. ex. i uttryck som *en något föråldrad bil*. 2. Deras betydelseinnehåll är blott i vissa avseenden fixerat och skiftar efter ordens användning. Härvidlag är skillnaden mellan språkets ord uppenbarligen en gradskillnad – betydelseinnehållet är normalt inte i alla av-

seenden fixerat hos mer än ett fåtal ord (t. ex. *hundra* i uttryck som *hundra plus fem är hundrafem*). En avgränsning på denna grund mot övriga substantiviska och adjektiviska tecken kan därför inte bli skarp. Å andra sidan torde man emellertid kunna konstatera en påfallande skillnad i pregnans mellan t. ex. *det* och *palatset/skeppet/porträttet/stallet, vem* och *presidenten/sjömannen/målaren/hästen, några/många* och *ett stim/hundrafem/oräkneliga/spolformiga*. De senare alternativens konnotationer är avgjort innehållsrikare. 3. De används i stället för namnet på eller den egentliga beteckningen för ett visst begrepp med syftning på detta. Denna egenskap kan svårligen anses utmärka de attributiva tecknen, jfr uttryck som *den där bilen, en sådan bil, den ena bilen, några bilar, vilken bil?* ”I sin egentliga betydelse: ’ord som står i stället för ett nomen’, passar ordet pronomen blott på de substantiviska pronomina. De adjektiviska ha sitt nomen (substantiv) bredvid sig” (Natanael Beckman, Svensk språklära, 1945, § 54 Anm. 4). Emellertid täcker termen ersättningsord knappast ens de substantiviska pronomenen – vad ersätter t. ex. *somliga, vem, vem som helst?* ”No one doubts that *nobody* and the interrogative *who* are pronouns, but it is not easy to see what nouns they can be said to be substitutes for” (Otto Jespersen, *The Philosophy of Grammar*, 1924, s. 82). – På liknande sätt bestämmer Beckman ordklassen (§ 54).

Man kan alltså ta fasta på att pronomenen normalt har substantivisk eller attributiv funktion och relativt svag pregnans; i semantiskt avseende är de i övrigt ganska olika, jfr t. ex. *jag, som, några, vem*. Man kan tillägga, att de i stor utsträckning har synnerligen hög frekvens. Av en ordförrådsundersökning tillkommen vid mitten av 1950-talet, vars resultat publicerats av Carita Hassler-Göransson i *Skola och samhälle* 1957, s. 141 ff., kan man sålunda utläsa, att av de hundra vanligaste orden (böjningsformer och avledningar räknas som självständiga ord) i en löpande text, vilka utgör ungefär hälften av ordmassan, i runt tal en tredjedel är pronomen. De nedan som personliga betecknade pronomenen (*jag, min, mig* osv.) representeras därvid av fler tecken än de övriga typerna tillsammans. – Här skall emellertid inte något försök göras att formulera en definition.

Intresset skall i det följande huvudsakligen inriktas på ordklassens indelning.

Vad man då först kan konstatera är, att den traditionella indelningen av ordklassen inte är tillfredställande, något som Loman också kraftigt framhåller och ganska utförligt demonstrerar (se särskilt s. 241 ff.). Ej heller bland grammatikerna själva tycks emellertid tillfredsställelse råda. Sålunda faller t. ex. Eskil Källqvist i sin resonerande Svensk språklära, 1954, s. 138, anmärkningen, att den gängse indelningen av pronomenen – vilken han själv i stort sett bibehåller – ”har den svagheten, att den inte följer någon bestämd indelningsgrund”. Jfr också Rudolv Körners kritik av ”pronominalläran i skolgrammatiken” i Moderna Språk XXX, 1936, s. 69 ff.

Ställer man Källquists anmärkning i relation till den traditionella grammatikens standardverk, Beckmans Svensk språklära, finner man, att den inte är obefogad. Pronomenen indelas där till en början ”efter sin betydelse” i tre kategorier (§ 57): definitiva eller bestämda (sammanhanget ger besked om vem eller vad pronomenet åsyftar), interrogativa eller frågande (den talande vill ge den hörande en uppmaning att utfylla, vad han själv ej vet) och indefinita eller obestämda (den talande vet icke och begär ej heller att få veta, hur pronomenets betydelse skall utfyllas). De definitiva pronomenen indelas sedan ”efter olika grammatiska egenskaper” i fyra grupper: personliga, possessiva, demonstrativa och relativa. ”Samtliga pronomina sönderfalla därför i sex grupper.”

Det är naturligtvis otillfredsställande att på detta sätt räkna med sex grupper, av vilka fyra egentligen är underavdelningar av en med de båda övriga jämförbar kategori. Det visar sig vidare, att de ”grammatiska egenskaper”, efter vilka de definitiva pronomenen indelas, omfattar dels funktionella, dels formella och dels semantiska karakteristika i växlande konstellationer (se Beckman §§ 58, 63, 66 och 74; jfr även Loman s. 242), varför indelningen inte är så enkel som formuleringen låter förmoda. Svävningarna i indelningsprincipen tar sig uttryck i oklarhet vid behandlingen av ordklassens grupper. Detta gäller t. ex. de demonstrativa pronomenen och, i hög grad, de indefinita. Vad denna sista grupp skall

anses omfatta, ger Beckman egentligen inte något besked om. Ordklassens gräns mot de s. k. obestämda räkneorden blir härigenom mycket suddig (se vidare nedan).

Liksom Källquist och Körner påtalar Loman "bristen på enhetlig indelningsgrund" (s. 242). Den leder till "sådana diskutabla fall som att *förste* karakteriseras som ordningstal men *siste* som adjektiv, *hundra* som grundtal men *många* som obestämt pronomen/obestämt räkneord, och att *båda*, *bägge*, *hel*, *halv*, *talrika* och *fåtaliga* helt lämnas utanför den pronominella indelningen medan *somliga* betecknas som ett obestämt pronomen".

En schematisk översikt av Beckmans indelning av pronomen och räkneord ter sig på följande sätt:

Pronomen					
definita				interrogativa	indefinita
personliga	possessiva	demonstr.	relat.		
<i>jag, mig</i> <i>du, tåg</i> <i>ni, er</i> sing. <i>han, honom</i> <i>han</i> <i>hon, henne</i> <i>hennes</i> <i>vi, oss</i> <i>I, ni, e(de)r</i> <i>de, dem</i> <i>deras</i> <i>sig</i> reflex. ( <i>varandra</i> recipr.)	<i>min</i> <i>vår</i> <i>din</i> <i>e(de)r</i> sing. <i>e(de)r</i> plur. <i>sin</i> reflex. <i>de mina</i> <i>de dina</i> <i>de sina</i> <i>de våra</i> ( <i>min egen</i> ) ( <i>allas vår</i> ) ( <i>var sin</i> )	<i>den, det</i> <i>dens, dess</i> <i>de, dem</i> <i>deras</i> <i>den här</i> <i>den där</i> <i>denne</i> <i>densamme</i> <i>samme</i> <i>sådan</i> <i>sådan där</i> <i>sådan här</i> ( <i>dylik</i> ) ( <i>slik</i> ) ( <i>likadan</i> ) ( <i>tocken</i> ) ( <i>så</i> )	<i>som</i> <i>vars</i> <i>vilken</i> <i>vad</i> <i>den</i> <i>den där</i> (obest. relat.)	<i>vem</i> <i>hø</i> <i>vars</i> <i>vad</i> <i>vad för en</i> <i>vilken</i> <i>vilkendera</i> <i>hur(u)dan</i>	<i>man, en, ens</i> <i>någon, några</i> <i>någondera</i> <i>någoting</i> <i>somlig, somliga</i> <i>somt</i> <i>vem som helst</i> <i>vad som helst</i> <i>vilken som helst</i> <i>hurudan som helst</i> <i>eho</i> <i>evad</i> <i>vem (... än)</i> <i>vad helst</i> <i>vilkendera (... än)</i> <i>annan, andra.</i> ( <i>var</i> ) ( <i>varje</i> ) ( <i>envar</i> ) ( <i>var och en</i> ) ( <i>alla</i> ) ( <i>mången</i> ) ( <i>flere</i> ) ( <i>få</i> ) ( <i>ingen</i> ) ( <i>den och den</i> )

Räkneord		
grundtal	ordningstal	obestämda
<i>ett, två etc.</i> <i>tu</i> <i>tvenne</i> <i>trenne</i> <i>(tolft)</i> <i>(dussin)</i> <i>(tjog)</i> <i>(skock)</i>	<i>den förste, den andre etc.</i>	<i>alla</i> <i>många</i> <i>få</i> <i>några</i> <i>flere</i>

De vanliga läroböckerna följer i stort sett Beckmans principiella uppläggning. Man räknar alltså i allmänhet med sex grupper. Stundom uppförs dock en sjunde grupp, determinativa pronomen, vilket blott ytterligare poängterar det oenhetliga i framställningen. Från pedagogisk synpunkt måste det också framstå som en mindre lycklig konsekvens, att samma ord i olika grammatikor betecknas på olika sätt: *hans, hennes, dess, deras* än som personliga pronomen, än som possessiva; *alla, åtskilliga, somliga* m. fl. än som obestämda pronomen, än som obestämda räkneord; *båda* och *bägge* än som obestämda pronomen, än som räkneord osv.

Lomans ovannämnda undersökning går i korta drag ut på att studera framförställda attributiva bestämningars ordföljds- och betydelseförhållanden. Författaren finner, att dessa bestämningar kan hänföras till sju positionsklasser (PK). Av dessa visar sig den sjunde innehålla kvalitetsbetecknande attribut, medan de sex föregående föreslås få utgöra de attributiva pronomenen (vilket innebär en utvidgning av ordklassen). De sex grupperna analyseras semantiskt, varvid respektive grupper grundbegrepp fastställs. På grundval av dessa föreslås så en indelning av de attributiva pronomenen enligt följande (s. 244): "Det är ur det nutida svenska riksspråkets synpunkt mera ändamålsenligt att indela attributiva pronomena i totala (PK 1), determinativa (PK 2), possessiva (PK 3), kvantitativa (PK 4), selektiva (PK 5) och komparativa (PK 6)." De av Loman behandlade attributiva pronomens fördelning framgår av översikten på följande sida.

Attributiva pronomen					
totala	determinativa	possessiva	kvantit.	selektiva	komparat.
<i>hela</i>	<i>denna</i>	<i>min</i>	<i>myckna</i>	<i>annan</i>	<i>sådan</i>
<i>halva</i>	<i>detta</i>	<i>mitt</i>	<i>två</i>	<i>annat</i>	<i>sådant</i>
<i>all</i>	<i>dess</i>	<i>mina</i>	<i>tre etc.</i>	<i>enda</i>	<i>sådana</i>
<i>allt</i>	<i>den (här/där)</i>	<i>din</i>	<i>båda</i>	<i>enstaka</i>	<i>likadant</i>
<i>alla</i>	<i>det (här/där)</i>	<i>ditt</i>	<i>bägge</i>	<i>första</i>	<i>likadana</i>
<i>båda</i>	<i>de (här/där)</i>	<i>dina</i>	<i>många</i>	<i>andra etc.</i>	<i>dylikt</i>
<i>bägge</i>	<i>en</i>	<i>sin</i>	<i>talrika</i>	<i>sista</i>	<i>dylika</i>
<i>samtliga</i>	<i>ett</i>	<i>sitt</i>	<i>få</i>	<i>senaste</i>	<i>slikt</i>
	<i>(en viss)</i>	<i>sina</i>	<i>fätaliga</i>	<i>övriga</i>	<i>slika</i>
	<i>(en bestämd)</i>	<i>hans</i>	<i>otaliga</i>	<i>ena</i>	<i>hurdan</i>
	<i>någon</i>	<i>hennes</i>	<i>mycken</i>	<i>senare</i>	<i>sådana här</i>
	<i>något</i>	<i>dess</i>	<i>mycket</i>	<i>tidigare</i>	
	<i>några</i>	<i>(ens)</i>	<i>mången</i>		
	<i>varje</i>	<i>vår</i>	<i>månget</i>		
	<i>ett par</i>	<i>vårt</i>			
	<i>flera</i>	<i>våra</i>			
	<i>ingen</i>	<i>er</i>			
	<i>inget</i>	<i>ert</i>			
	<i>intet</i>	<i>era</i>			
	<i>inga</i>	<i>deras</i>			
	<i>var</i>	<i>allas (vår/er)</i>			
	<i>somliga</i>	<i>allas (vårt/ert)</i>			
	<i>åtskilliga</i>	<i>allas (våra/era)</i>			
	<i>vissa</i>	<i>vilkens interr.</i>			
	<i>vilken inter-</i>	<i>vems</i>			
	<i>rog., relat.</i>				

De i den refererade undersökningen blottade ordföljds- och betydelseförhållandena är oavsett ens inställning till frågan om pronomenens indelning av utomordentligt stort intresse. När det gäller den föreslagna indelningen, kan några överväganden emellertid ge upphov till en viss tvekan beträffande dess ändamålsenlighet. 1. Sannolikt vore ett system som täckte hela ordklassen att föredra. 2. Helst borde en och samma princip motivera såväl gruppering som terminologi. 3. Den valda grupperingsprincipen borde helst konsekvent kunna genomföras. Placeringen av interrogativerna *vilkens*, *vems* och *hurdan* får Loman låta bestämmas av formella och semantiska skäl (s. 232), likaså placeringen av *ena* (s. 228). Komplikationer uppträder



vidare vid den allmänna avgränsningen av den femte positionsklassen. På s. 228 konstateras, att medlemmar av PK 6 stundom även kan stå "efter (emfatiskt betonade) medlemmar av beskrivningsfältet" (dvs. kvalitetsbeteckningar); som exempel ges *Nej, jag vill ha en gul sådan likadan hatt. Gul* torde få uppfattas som stående i femte position, jfr *två gula sådana hattar*. Inget hindrar oss från att uttala ett uttryck som *en gul sådan hatt* helt oemfatiskt eller med *sådan* eller *en* emfatiskt betonade. Därvid kan inte *gul* av fonetiska skäl avföras från PK 5. Man kan peka på andra fall, där en (alternativt) oemfatisk kvalitetsbeteckning står i femte position: *en bekant dylik situation (möter hos Bellman)*, *två tänkbara sådana förslag (har framställts)* osv. Här kan inte *bekant* respektive *tänkbara* under hänvisning till emfas uteslutas ur PK 5. Man torde vid avgränsningen av den femte positionsklassen i ganska stor utsträckning få lita till de semantiska förhållandena.

4. En rationell gruppering borde medge en så naturlig inplacering som möjligt av varje ifrågakommande tecken. Detta önskemål kan bli mindre väl tillgodosett i ett system, som kräver hänsynstagande till potentiella kombinationsmöjligheter, *någon min vän, vilket mitt hus talar du om?* o. d. (s. 230). En del mer eller mindre svårbestämbara pronomen, t. ex. *varenda, vad för en, själva*, tas inte upp. (Författaren betonar, att han inte eftersträvat fullständighet; härvidlag hänvisar han till "den språkstatistiska forskningen" (s. 225).) Om t. ex. *själva* torde gälla, att det liksom da. *selve* kommer i första position (se Paul Diderichsen, *Elementær dansk Grammatik*, 1957, s. 226), varigenom den redan av *halva* något splittrade semantiska enhetligheten i PK 1 ytterligare starkt splittras. 5. Då terminologien byggs på betydelseförhållandena, bör de semantiska relationerna inom varje grupp (de "vertikala") vara starkare än de mellan medlemmar av olika grupper (de "horisontella"). Positionsklass 2 uppfyller knappast detta krav (s. 230). 6. Ett ändamålsenligt system borde äga en viss anpassningsförmåga inför språkliga novationer. Skulle t. ex. talspråkets tendens att undvika konstruktioner som *denna hans åtgärd* slå igenom i skriftspråket, är urskiljandet av positionsklasserna 2 och 3 inte längre motiverat (se Loman s. 231). Den positionella grunden för grupperna determinativa och

possessiva pronomen faller, och den resulterande positionsklassen 2 blir i semantiskt avseende ytterligare heterogen. – Det spill som uppkommer vid indelningen av de attributiva pronomenen på positionell-semantisk grund kan måhända uppfattas så, att en fullt konsekvent genomförbar dylik indelning skulle förutsätta en större exakthet hos språket än detta besitter.

Vill man söka ett alternativ till de båda nu diskuterade indelningarna, förefaller det mig, att man bör pröva möjligheten att låta betydelseförhållandena ensamma utgöra indelningsgrund, att alltså bygga såväl gruppering som terminologi direkt på semantisk analys och att låta denna omfatta såväl i funktionellt avseende attributiva som substantiviska pronomen. I det följande skall ett förslag till indelning på denna grund i grova drag skisseras. Vad man därvid kan inrikta sig på är att söka åstadkomma en så motsägelsefri indelning som språkets karaktär medger. De urskiljbara grupperna kan inte väntas stå i ett oantastbart logiskt förhållande till varandra.

Sedan pronomenen indelats i grupper efter likartade gemensamma betydelseegenskaper, kan givetvis inom varje grupp andra (formella, funktionella, semantiska) kännetecken framhävas för en närmare beskrivning eller överskådlig uppställning av respektive grupps medlemmar. Stundom får ett tecken föras till mer än en grupp. Detta gäller t. ex. *den*, *vilken* och *en del*. Man kan överväga att beteckna detta som fall av homonymi i en vid mening av denna term. Det visar sig lämpligt att vid grupperingen av tecknen lämna kvaliteterna genus (resp. sexus), numerus, kasus och species utan avseende. Detsamma gäller av naturliga skäl variationer i stilvärde.

Den fornärvda indelningen i definitiva och indefinita pronomen synes mig i princip ha ett icke oväsentligt värde därigenom, att den ger en första antydning om den semantiska karaktären hos tecknen och möjliggör en grovsortering av dem. Vad man närmare angivet skall mena med de båda kategorierna har formulerats på olika sätt. Här skall definitionerna ges följande formuleringar: ett definit pronomen hänсыftar på något som avsändaren uppfattar som bekant; ett indefinit pronomen hänсыftar på något som avsändaren uppfattar som obekant. Exempel på den förra

typen är *jag, den där*, på den senare *somliga, vem*. Vid indelningen i dessa båda grundläggande kategorier kan man bortse från specialfall som t. ex. *somliga kommer för sent* (utsagan riktad till en eller flera just inträdande), där det normalt indefinita *somliga* har karaktären av ett definit pronomen. – I förbigående vill jag nämna, att jag föredrar termerna definit och indefinit, vilka dels bidrar till att ge terminologien en enhetlig prägel, dels och framför allt lättare kan ges det önskade betydelseinnehållet.

En vidare indelning av de definita pronomenen kunde ske i deiktiska pronomen, vilka hännyftar på något utom kontexten befintligt, genom talsituationen givet (t. ex. *den där*), anaforiska pronomen, vilka hännyftar sig till korrelat som anförts i det föregående (t. ex. *som*), och determinativa pronomen, vilka hännyftar sig till korrelat som anförts i det följande (t. ex. *dens*). För vissa behov är det också praktiskt att räkna med dessa typer. Två mycket starka skäl talar emellertid mot att hänsyn tas till denna indelningsgrund vid den fortsatta analysen. Det ena skälet är, att detta slags indelning inte är tillämplig för de indefinita pronomenen, det andra, att grupperna ifråga endast i ringa utsträckning motsvaras av egna pronomen. Som ofta sker, bör man i stället tala om pronomen i deiktisk, anaforisk eller determinativ funktion.

Om grammatisk person, dvs. förhållandet till avsändaren i en okomplicerad språksituation, upplyser de allra flesta pronomen. Bland de få undantagen befinner sig t. ex. *varandra, var sin* och *själv*. Första personen betecknas av helt få pronomen (t. ex. *jag, vår*), likaså andra personen (t. ex. *ni, dig*). Det stora flertalet pronomen betecknar som rimligt är normalt tredje personen (t. ex. *hans, den ene, många, vem*).

En stor del av tecknen uttrycker possession (varmed inte nödvändigtvis ett ägandeförhållande avses, jfr t. ex. *deras uppfostran*). Detta gäller *min, mitt, mina, hans, dennes, dennas, dettas, dessas, vars, den enes, den enas, det enas, ingens, ingets, ingas, någons, någots, någras, vems, vilkens, vilkets, vilkas, ens* osv. Denna semantiska egenskap är genom sin omfattande utbredning i och för sig olämplig som gruppstiljande karakteristikum. Dessutom rymmer nästan alla dessa possessiva tecken andra utmärkande betydelseegenskaper, egenskaper som de har

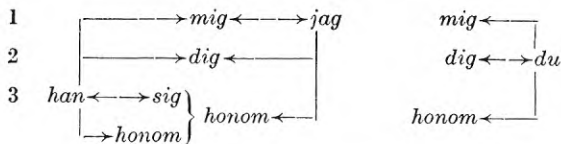
gemensamma med andra tecken. Sålunda har t. ex. *dennes*, *dennas*, *dettas* och *dessas* på samma sätt som *denne*, *denna*, *detta* och *dessas* en utpekande, demonstrativ, innebörd, *någons*, *någots* och *någras* liksom *någon*, *något* och *några* en kvantitativ innebörd, *vems* liksom *vem* en interrogativ innebörd osv. Det är på dessa senare betydelseegenskaper – de skulle kunna kallas tecknens distinktiva semantiska karakteristika – som man enligt min mening med fördel kan ta fasta. I den mån de saknas (så t. ex. hos *jag*, *dig*, *hans*) kan själva detta faktum uppfattas som det för vederbörande tecken karakteristiska. I det följande kommer alltså betydelseegenskapen possession inte att göras grupp-skiljande. Ett tecken som *denne* blir ofta ett praktiskt skrivsätt för *denne*, *dennes*, *denna*, *dennas*, *detta*, *dettas*, *dessas* och *dessas*, liksom t. ex. *min* för *min*, *mitt* och *mina*, *vem* för *vem* och *vems* osv.

Tillvägagångssättet vid indelningen kan med några få exempel illustreras i ett principschema av följande slag:

Tecken	Betydelseegenskaper				Kategori	Grupp
	def.	gramm. pers.	—	—		
<i>jag</i>	def.	gramm. pers.	—	—	def.	personliga
<i>min</i>	»	»	poss.	—	»	»
<i>var sin</i>	»	—	»	distribution	»	distributiva
<i>den ene</i>	»	»	—	selektion	»	selektiva
<i>den enes</i>	»	»	»	»	»	»
<i>några</i>	indef.	»	—	kvantitet	indef.	kvantitativa
<i>någras</i>	»	»	»	»	»	»
<i>man</i>	»	»	—	generalisation	»	generella
<i>ens</i>	»	»	»	»	»	»

Som en första grupp bland de definita pronomenen kan man urskilja dem, som saknar vad som ovan betecknats som distinktivt semantiskt karakteristikum. Den positiva förbindelselänken mellan dessa tecken å andra sidan är uttryckandet av grammatisk person. Endast i denna grupp finns uttryck för alla de tre personerna. Hit hör t. ex. *jag*, *mig*, *du*, *din*, *han*, *honom*, *sig*, *hans*, *sin*, *de*, *dem* och *deras*. De kan med en utvidgning av den vedertagna termens omfattning betecknas som *personliga* pronomen.

Tecken som *mig* och *dig* (första resp. andra personen) innesluter i sig de båda nyanser som i tredje personen uttrycks med *sig* respektive *honom*: *jag ser mig (på fotot)/han ser mig* gentemot *han ser sig/jag ser honom*. Detta har sin naturliga förklaring i att det endast är vid språkliga förlopp inom tredje personen som ett behov av en formell distinktion mellan två möjligheter föreligger.



*Mig* och *dig* kan, som pilarna antyder, alltså reflektera första respektive andra personen, medan *sig* och *honom* alltid respektive aldrig reflekterar tredje personen. Motsvarande gäller för de plurala tecknen *oss* etc. och för gruppens possessiva medlemmar. De icke-reflexiva tecknen *mig*, *dig*, *honom* etc. kan av funktionella skäl benämnas oblika, varjämte de specifika uttrycken för de grammatiska personerna, *jag* etc., kan benämnas rektiva.

Ett schema över de personliga pronomenen får följande utseende:

Person	Rektiv	Possessiv		Reflexiv	Oblik
1 sing.	<i>jag</i>	<i>min</i>		<i>min</i>	<i>mig</i>
2 sing.	<i>du</i>	<i>din</i>		<i>din</i>	<i>dig</i>
	<i>ni</i>	<i>er</i>		<i>er</i>	<i>er</i>
3 sing.	<i>han</i>	<i>hans</i>		<i>sin</i>	<i>sig</i>
	<i>hon</i>	<i>hennes</i>		<i>sin</i>	<i>sig</i>
	<i>den</i>	<i>dess, dens</i>		<i>sin</i>	<i>sig</i>
	<i>det</i>	<i>dess</i>		<i>sin</i>	<i>sig</i>
1 plur.	<i>vi</i>	<i>vår</i>		<i>vår</i>	<i>oss</i>
2 plur.	<i>ni</i>	<i>er</i>		<i>er</i>	<i>er</i>
3 plur.	<i>de</i>	<i>deras</i>		<i>sin</i>	<i>sig</i>

Tecknen *sig* och *sin* kan reflektera även andra korrelat än personliga pronomen, t. ex. *den där, den ene, somliga, vem, vem som helst, ingenjören, familjen, människor*, vilka ju alla normalt betecknar tredje personen. *Vår* och *er* får en något modifierad betydelse genom ett framförställt *allas, bådas* eller *bägges*. En specialbetydelse har *de mina, de dina, de sina, de våra*: 'mina anhöriga' etc.

En särskild grupp bildar *varandra* och *varann*, som uttrycker växelverkan eller ömsesidighet. För dessa bibehålles lämpligen termen *reciproka* pronomen. Från de personliga pronomenen skiljer de sig förutom i ovannämnda avseende även genom att de som tidigare nämnts inte betecknar någon bestämd grammatisk person utan hänför sig till plurala korrelat i allmänhet.

Också *var sin* och *en* etc. *var/varadera* hänför sig till plurala korrelat i allmänhet. För dessa pronomen, som ju uttrycker ett distributivt possessionsförhållande, föreslås termen *distributiva* pronomen.

Obundna till grammatisk person (och även till personens numerus) är vidare *själv* och dess possessiva motsvarighet *egen*, förslagsvis benämnda *intensiva* pronomen. Dessas distinktiva betydelseegenskap synes vara en intensifiering i semantiskt avseende av de tecken vartill de hänför sig.

En annan grupp bildas av de pronomen, som har en utpekande innebörd – de *demonstrativa*. Sådana är *den där, den här, den, denne, han* och *hon*. De något mindre starkt utpekande *ovanstående* (Ragnar Eklund, Svensk språklära, 1948, § 201) och *nedanstående* kunde möjligen också föras hit, liksom kuralstilens särpräglade *bemälde, nämnde, berörda* osv. (Beckman § 70 Anm. 3). Hos vissa, andra grupper tillhöriga tecken möter ett mer eller mindre tydligt demonstrativt betydelseelement, så t. ex. hos *den förre* och *sådan (där/här)*.

Besläktade med de demonstrativa pronomenen är som just antytts tecken som *sådan, dylik, slik* och *likadan*. Dessa får, som Loman klart framhållit, sitt betydelseinnehåll genom "referens till känd modell" (s. 237). Det de hänsyftar på får alltså sägas vara på indirekt väg bekant för avsändaren och pronomenen däri-genom definitiva. Som Beckman observerat, kan dessa pronomen

emellertid sägas vara ”de minst bestämda av de definitiva pronomina” (§ 73 Anm.). Med Lomans goda term kallar jag hithörande tecken komparativa pronomen. Utöver de ovan nämnda för jag till denna grupp *samme* och *densamme* samt de något mer deiktiska *sådan där* och *sådan här*. I uttryck som *en sådan vacker tavla!* är det komparativa betydelsemomentet hos *sådan* ganska urblekt. En jämförelse med tavlor i allmänhet kan dock anses ske. På motsvarande sätt kan typen *vilken vacker tavla!* bedömas, varigenom detta *vilken* också torde kunna föras till de komparativa pronomenen.

En karakteristisk grupp inom ordklassen bildar *som*, *vars*, *vilken* och *vad*. Genom dem ställs i allmänhet ett föregående korrelat i relation till en följande utsaga. För denna grupp kan den traditionella termen *relativa* pronomen användas. Stundom saknas korrelat, så t. ex. i uttryckstypen *han gjorde, vad han blivit tillsagd*.

Om *annan* skriver Beckman bl. a. (§ 95): ”Har jag blott två att välja på så tjänstgör *annan*, i best. form *andre*, *-a*, som definit pronomen. Ex. *Den ene gick, den andre kom*. På samma sätt kan det vara i pluralis. Ex. *Du får stanna kvar, de andra få gå*.” Är valet inte på angivet sätt begränsat, får *den ene*, *den andre* indefinit karaktär, se nedan. Till samma semantiska typ *som den ene* och *den andre* hör *den förre* och *den senare*, *den förstnämnde* och *den sistnämnde*, vilka dock enligt Beckman trots sin betydelse ej bruka upptagas bland pronomina” (§ 70 Anm. 3). Från rent semantisk synpunkt finns det emellertid knappast anledning att ställa dem utanför ordklassen. Eklund inkluderar *den förre* och *den senare* (§ 201). I semantiskt avseende betecknar de alla ett bestämt urval. De är att jämföra med större delen av den grupp, för vilken Loman föreslagit termen *selektiva* pronomen, nämligen – förutom *ene* och *andre* – *ende*, *senaste*, *sista* och *övriga*. Till denna typ kan också räknas de av Loman som determinativa betecknade *en viss* ’en bestämd, angiven’ och *en bestämd* samt vidare *näste*, *föregående* och *följande*. Om de besläktade tecknen *nollte*, *förste* etc. (de s. k. ordningstalen) och *varannan* etc. se nedan.

De pronomen, som på ett eller annat sätt betecknar antal, synes mig hittills inte ha ägnats den uppmärksamhet de förtjänar. Till utgångspunkt för en analys av dem kan man välja ett uttalande av Beckman (§ 96): "Nästan alla här upptagna obestämda pronomen beteckna även på obestämt sätt antal. Därför övergå obestämda pronomen omärkligt i obestämda räkneord". De pronomen det gäller är *några* (§ 91), *somliga* (§ 92), *andra* (§ 95) m. fl. I termen obestämda pronomen är obestämd synonymt med indefinit och anger då enligt Beckmans definition att "den talande icke vet och icke heller begär att få veta, hur pronomens betydelse skall utfyllas" (§ 57:3). I termen obestämda räkneord har obestämd däremot ungefär betydelsen 'som betecknar ett icke närmare angivet, approximativt antal'. Att använda samma ord i dessa båda klart skilda betydelser är olämpligt. Terminologien blir koncigare, om man i stället för obestämda pronomen använder uttrycket indefinita pronomen.

I själva verket betecknar ju termerna obestämda pronomen och obestämda räkneord i stort sett samma ordgrupp, betraktad från olika semantiska synpunkter. Det ena uttrycket framhäver, att orden ifråga är pronomina och närmare bestämt indefinita, det andra, att de är antalsbeteckningar och som sådana approximativa. Erik Wellander, som i sin språkspalt i SvD den 8/4 1957 är inne på denna fråga, skriver: "Lägger man mera tonvikt på den innehållsfattigdom som anses känneteckna pronomen, så blir man böjd för att kalla sådana ord som *några* för pronomen, lägger man åter mera tonvikt på antalet, så lutar man givetvis åt att beteckna dylika ord som räkneord." Gränsen mellan de båda grupperna betecknas t. ex. hos Eklund följdriktigt som "svår att ange" (§ 214). Det rimliga är att föra samtliga ifrågakommande ord till pronomenen (jfr nedan), alltså till samma ordklass som Beckman i varje fall för *någon*, *någondera*, *någoting*, *somlig*, *somt* och *annan*.

Icke alla antalsbetecknande pronomen kan emellertid hänföras till kategorien indefinita. Vad ett tecken som *alla* hänsyftar på är i princip bekant för avsändaren. K. G. Ljunggren anför i Svensk uppslagsbok om bl. a. just *alla*, att vad det anger "strängt taget icke är något obestämt" (s. v. *pronomen*). Jfr även Loman s. 240.



Lomans välvalda term för hithörande tecken är *totala* pronomen. Som sådana räknar jag i överensstämmelse med honom också *samtliga*, tecknen *båda* och *bägge*, vilka sannolikt alltid implicerar ett totalitetsförhållande och därför hänförs enbart hit, samt mängdbeteckningen *hel* – däremot inte *halv*, varom se nedan. Till denna grupp för jag vidare *allesamman(s)*, *allihop(a)*, *alla tre* etc., *allting*, *bådadera*, *båda två*, *bäggedera* och *bägge två* samt de visserligen individualiserande men dock klart totalitetsbetecknande *varje*, *var*, *var och en*, *envar*, *varenda*, *vardera* och det vardagliga *vareviga* (SAOL s. v. *var*). Sådana formellt inkongruenta uttryck som de i radio avlyssnade *Det gäller för var och en att själva pröva* (26/3 1957) och *Två poäng för vardera lagen* (9/4 1957) eller det i Erik Wellanders *Riktig svenska*, 1948, s. 389, citerade *Varje hans ord har upptecknats, och de ha sedan bevarats med största omsorg* låter dessa pronomens betydelseinnehåll framträda.

Också *inga*, *intet* och *ingenting* torde uttrycka ett totalitetsförhållande, ehuru aspekten här är negativ. Må det tillåtas mig att citera en strof av Pär Lagerkvist.

Ingenting får störa vår stund med varandra,  
inga vindar blåsa, inga skyar gå.  
Allt skall vara stilla i världen,  
ingen vandra över markens stillhet mera än vi två.

Att *ingenting får störa* innebär ordagrant, att *allting vägras störa*: *alla vindar att blåsa, alla skyar att gå*, sammanfattat i *allt skall vara stilla*. På motsvarande sätt är *ingen vandra* en utsaga om *var och en*. Jfr vidare t. ex. två med varandra ekvivalenta satser som *alla är närvarande* och *inga är frånvarande*. Man kan upplösa *ingenting* i *allting ... icke-*, *inga* i 'alla ... icke-' och *ingen* i 'var och en ... icke-' eller, i andra fall, i 'varje ... icke-' eller 'all ... icke-'. Den vanliga analysen *inga* 'icke några' torde bygga på *några* i vad som nedan betecknas som generell betydelse, 'några som helst', och som där säges stundom ligga nära total betydelse. Denna uppfattning avviker från den av Loman framförda, enligt vilken *inga* betecknar "'indifferent plural', d. v. s. utan närmare värdering av kvantiteten", (s. 238). Förhållandet

mellan *ingen* och *inte någon* skulle också kunna belysas med hjälp av en typmening som *man har ingen möjlighet (att klara krisen)*.

man har ingen möjlighet  
man har inte någon möjlighet

man saknar varje möjlighet  
man saknar möjlighet

man har inga möjligheter  
man har inte några möjligheter

man saknar alla möjligheter  
man saknar möjligheter

I de med avseende på pronomenet positiva formuleringarna motsvaras så vitt jag kan se *ingen*, *inga* av *varje*, *alla*, dvs. de har total betydelse. Den positiva motsvarigheten till *inte någon*, *inte några* (generell betydelse) uttrycks däremot utan hjälp av pronomen.

Beckmans mening, att *ingen* "betecknar en bestämd siffra, 0" (§ 96), är vilseledande. *Ingen* och *noll* är inte synonyma. I regel ger *noll* liksom t. ex. *hundra* blott en antalsuppgift utan relation till totalitet: *vindstyrkan är noll meter per sekund gentemot inga vindar blåser*. I vissa fall dominerar starkt den negerande betydelsen hos *ingen* på bekostnad av den totala: *du är inget barn längre, men ingen Jenny kom, han hade ingen stor röst* (jfr SAOB s. v. I 473). Samma relation som föreligger mellan *inga* och *alla* föreligger också mellan *ingendera* och *bådadera*, *bäggedera*. *Ingendera* är alltså också ett totalt pronomen.

Därmed är de pronomen som jag vill ge beteckningen definitiva behandlade. Denna kategori har kommit att omfatta nio grupper: personliga, reciproka, distributiva, intensitiva, demonstrativa, komparativa, relativa, selektiva och totala pronomen. Följande sammanställning ger en överblick av kategorien:

Definita pronomen				
personliga	reciproka	distributiva	intensitiva	demonstrativa
jag min, mig du din, dig ni er han hans, sin sig, honom hon hennes, henne den dess, dens det vi vår, oss ni er de deras, sin sig, dem	varandra varann	var sin en etc. var en etc. vardera	själv egen	den där den här den denne han hon ovanstående nedanstående bemålde etc.

komparativa	relativa	selektiva	totala
sådan sådan där sådan här dylik stik likadan samme densamme vilken	som vars vilken vul	den förre den senare den förstnämnde den sistnämnde den ene den andre den ende en viss en bestämd siste senaste näste föregående följande övriga	alla allesamman(s) allihop(a) alla tre etc. allting hel samtliga båda bådadera båda två bägge bäggedera bägge två varje var var och en envar varenda vardera vareviga inga intet ingenting ingendera

Bland antals- och mängdbeteckningarna kan man, förutom de redan behandlade *alla* och *hel*, urskilja typerna *många*, *mycken* och *somliga*, *somt*. Dessa skiljer sig från *alla* och *hel* genom att de är indefinita. En vidare analys ger vid handen, att medan *många* och *mycken* utöver kvantitetsangivelsen inte uppvisar någon karakteristisk betydelseegenskap, uttrycker *somliga* och *somt* ytterligare ett slags begränsning eller inskränkning (limitation), varom närmare nedan, på motsvarande sätt som *alla* och *hel* kännetecknas av sin totalitetsbetydelse. Det anförda kan sammanfattas på följande sätt:

<i>alla</i>	antal	totalitet
<i>hel</i>	mängd	totalitet
<i>många</i>	antal	—
<i>mycken</i>	mängd	—
<i>somliga</i>	antal	limitation
<i>somt</i>	mängd	limitation

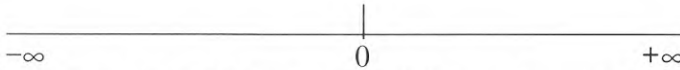
Liksom totalitetsbetydelsen är den karakteristiska och förenande semantiska egenskapen hos *alla* och *hel*, är det den limitativa betydelsen som utmärker och förenar *somliga* och *somt*. De får därför beteckning limitativa pronomen. Hos *många* och *mycken* framstår kvantitetsangivelsen som den distinktiva semantiska egenskapen. En uppdelning av typerna *många* och *mycken* i antals- och mängdbeteckningar (numeriska resp. i strikt mening kvantitativa pronomen) vore inkonsekvent (och dessutom opraktisk).

De k v a n t i t a t i v a pronomenen bildar en av de största pronomina grupperna. Jämte *många* får man föra hit *åtskilliga*, *fler(a)*, *flest*, *några*, *någoting*, *få*, *färre* och *ett par* 'några' samt varianterna *flera stycken*, *några stycken*, *några få* och *ett par stycken*. Pluralformen *ena* i t. ex. *ena stackare* (Beckman § 50 Anm. 3) hör också hit. Detsamma gäller förutom *mycken*, *mycke(t)* även mängdbeteckningarna *föga* och *lite(t)* jämte varianten *något lite(t)*.

Uttryck av typen *ett fåtal* fungerar också som kvantitativa pronomen. På det formella planet bekräftas detta av den av Erik Wellander i den ovannämnda språkspalten diskuterade konstruktionstypen *ett fåtal lärjungar äro kuggade*, där participet

(och hjälpverbet) uppvisar plural kongruensböjning. Uttryck av denna typ är vidare *ett flertal, ett antal, en mängd, en massa, en myckenhet, en (hel) del, en smula*. De är att jämföra t. ex. med engelskans *a lot of, a great deal of, a large number of*. Inkluderar man *ett fåtal, en massa* osv. i pronomenen, kvarstår givetvis *fåtal, massa* osv. som substantiv på samma sätt som t. ex. substantivet *ting* vid sidan av pronomenet *ingenting*, eng. *body* vid sidan av *everybody* osv. Det är osäkert om den positionella metoden medger, att uttrycken inräknas bland pronomenen, eftersom de liksom genitivformer av substantiv och egennamn i sin tur kan bestämmas av attribut: *en hel del, en väldig massa* osv. Märk, att detta funktionella kriterium inte gäller *ett par*, varför detta uttryck av Loman kan upptas bland pronomenen. Också den semantiska typen *gott om* har anknytning till de kvantitativa pronomenen.

Loman föreslår, att också grundtalen (utom *en, ett*) av positionella skäl skall införlivas med de kvantitativa pronomenen (s. 227). Man får då räkna med dels det attributiva pronomenet *hundra* (i t. ex. *hundra frimärken*), dels det substantiviska tecknet *hundra* (i t. ex. *hundra plus fem är hundrafem*). Det bjuder emot att räkna ett så pregnant och till sin betydelse väl definierat tecken som detta substantiviska *hundra* som pronomen; det torde få föras till någon annan ordklass. Också det attributiva *hundra* avviker emellertid semantiskt från *många* osv. Genom att det betecknar exakt antal och alltså kan omskrivas med siffror är det till sin betydelse mer fixerat än de kvantitativa pronomenen. Jfr också Diderichsen s. 28: "Sammenligner man Ordene *begge to* eller *de* og *tyve*, ser man, at de alle mangler beskrivende Indhold: man siger ikke noget om Beskaffenheden af det, der nævnes. Men medens *begge* og *de* altid betegner Genstande (R[elatum]) i Flertal, er *to* og *tyve* rene Talstørrelser, der strengt taget kun betegner Pladser i Talrækken, selv om man i daglig Tale mest bruger dem om et vist Antal af Genstande." Med hänsyn härtill framstår en uppspaltning av *hundra* på två ordklasser knappast som tillräddlig. Det torde från semantisk synpunkt i stället vara befogat att föra samtliga de tecken som har avseende på punkter på den s. k. tallinjen



till en särskild ordklass, räkneord. Denna kommer då inte att omfatta endast grundtalen utan också ordningstalen. Göres på detta sätt tallinjen till ryggrad i ordklassen räkneord, framstår inkluderingen av "de obestämda räkneorden" i ordklassen pronomen närmast som ofrånkomlig.

Till räkneorden bör då föras *noll, en etc., tvenne, trenne, tu, halv, halvannan, kvart(s)* och *tredjedel(s)* etc. Vidare kan man som räkneord uppfatta antalsbeteckningar som *par, dussin, tolf, tjog* och *gross* (jfr *million, milliard* etc.). Till ordningstalen, *nollte, förste* etc., kan med viss fördel läggas *varannan* etc., ehuru för denna ordtyp alternativa placeringar på tallinjen är möjliga. Vill man föra också *siste* hit, får man göra avkall på den logiska bundenheten till tallinjen. Följande schema kan alltså uppställas över ordklassen räkneord:

Räkneord	
grundtal	ordningstal
<i>noll, en etc.</i>	<i>nollte, förste etc.</i>
<i>tvenne, trenne</i>	<i>varannan etc.</i>
<i>tu</i>	
<i>halv</i>	
<i>halvannan</i>	
<i>kvart(s)</i>	
<i>tredjedel(s) etc.</i>	
<i>par</i>	
<i>dussin</i>	
<i>tolft</i>	
<i>tjog</i>	
<i>gross</i>	

Den ovan urskilda gruppen *limitativa* pronomen, typen *somliga, somt*, kan sägas vara en sorts indefinit motsvarighet till de selektiva pronomenen. Hithörande ord betecknar inte som de ett bestämt urval utan ett slags begränsning eller inskränkning, varigenom en viss kvantitet ställs i motsättning till viss annan. "Vid *somliga* tänker man nästan alltid ett *andra icke*" (Beckman

§ 92 Anm.). Som limitativa får man utöver *somliga* och *somt* anse *vissa* 'ej närmare angivna', *en del andra*, *en*, *en annan*, *en och annan*, *den ene* och *den andre*. I *en och annan* inryms motsättningen i själva pronomenet. Som limitativa möter *den ene* och *den andre* i fall som *den ene tycker si*, *den andre så*. Limitativa är vidare *endera* 'en av två' och *någondera* 'någon av två'. Stundom uppträder också *några* i limitativ betydelse.

Ett pronomen som *vem* hänsyftar på något som avsändaren uppfattar som obekant – pronomenet är alltså indefinit – men som han gör till föremål för bestämningsförsök. Egenskapen att ge uttryck häråt får betecknas som pronomenets distinktiva karakteristikum. Den traditionella termen *interrogativa* pronomen är alltså lämplig. Till samma grupp som *vem* hör *vad*, *vad för en*, *vilken*, *vilkendera* och *hurdan*, varjämte frågeuttrycken *hur många*, *hur mycket* etc. också kan inlemmas (jfr Beckman § 55:6).

En sista tydlig grupp bland de *indefinita* pronomenen bildar uttryck av typen *vem som helst*, som utmärks av att beteckna en generalisation – "valet är alldeles likgiltigt" (Beckman § 94). Denna ganska stora grupp kallar jag därför *generella* pronomen. Hit hör jämte *vem som helst* till en början *vad som helst*, *vad för en som helst*, *vilken som helst*, *vilkendera som helst*, *hur många* etc. *som helst* och *hurdan som helst*, vidare *några som helst*, vilket även utan *som helst* kan uppvisa generell betydelse, jfr ovan. Huruvida *som helst* också ger *inga* och *varje* generell betydelse eller endast förstärker den totala betydelsen är mera tveksamt. Typen *vem som helst* kan till betydelsen närma sig typen *alla*: *Du kan fråga vem som helst, det veta alla människor* (Beckman § 94:I). I jämförelse med *alla* innebär då *vem som helst* snarast en skillnad i synpunkten på det sagda: 'den eller den eller den ...' gentemot 'den och den och den ...'.

Till de generella pronomenen hör också *man*, dess possessiva och oblika motsvarigheter *ens* och *en* samt det egenartade *den och den* (jfr Beckman 68 1/2:5). Dessutom ansluter sig osökt typen *evad* till gruppen. De vanliga nusvenska synonymerna till detta ord är *vad ... än* och *vad helst*, där *än* och *helst* icke är nödvändiga – varvid dock *vad* "nästan alltid" är tryckstarkt

(Beckman § 94 H:II) men tjänar till att klarare framhäva pronomenens generella karaktär. Av samma slag är *vem ... än* och *vem helst, vad för en ... än* och *vad helst för en, vilken ... än, vilkendera ... än, hur många* etc. ... *än* och *hurdan ... än*. Man observerar, att i de fyra sista fallen konstruktion med *helst* icke kan förekomma.

Inom kategorien indefinita pronomen har jag sålunda fått anledning särskilja fyra grupper: kvantitativa, limitativa, interrogativa och generella pronomen. En sammanställning av denna kategori får följande utseende:

Indefinita pronomen			
kvantitativa	limitativa	interrogativa	generella
<i>många</i>	<i>somliga</i>	<i>vem</i>	<i>vem som helst</i>
<i>åtskilliga</i>	<i>somt</i>	<i>vad</i>	<i>vad som helst</i>
<i>fler(a), flest</i>	<i>vissa</i>	<i>vad för en</i>	<i>vad för en som helst</i>
<i>flera stycken</i>	<i>en del</i>	<i>vilken</i>	<i>vilken som helst</i>
<i>några</i>	<i>andra</i>	<i>vilkendera</i>	<i>vilkendera som helst</i>
<i>några stycken</i>	<i>en</i>	<i>hur många</i> etc.	<i>hur många</i> etc. <i>som helst</i>
<i>några få</i>	<i>en annan</i>	<i>hurdan</i>	<i>hurdan som helst</i>
<i>någoting</i>	<i>en och annan</i>		<i>några (som helst)</i>
<i>få, färre</i>	<i>den ene</i>		<i>man, ens, en</i>
<i>ett par</i>	<i>den andre</i>		<i>den och den</i>
<i>ett par stycken</i>	<i>endera</i>		<i>evad</i>
<i>ena plur.</i>	<i>någondera</i>		<i>vem (... än)</i>
<i>mycken, -e(t)</i>	<i>några</i>		<i>vem helst</i>
<i>föga</i>			<i>vad (... än)</i>
<i>lite(t)</i>			<i>vad helst</i>
<i>något lite(t)</i>			<i>vad för en (... än)</i>
<i>ett fåtal</i>			<i>vad helst för en</i>
<i>ett flertal</i>			<i>vilken (... än)</i>
<i>ett antal</i>			<i>vilkendera (... än)</i>
<i>en mängd</i>			<i>hur många</i> etc. (... än)
<i>en massa</i>			<i>hurdan (... än)</i>
<i>en myckenhet</i>			
<i>en (hel) del</i>			
<i>en smula</i>			
<i>gott om</i>			



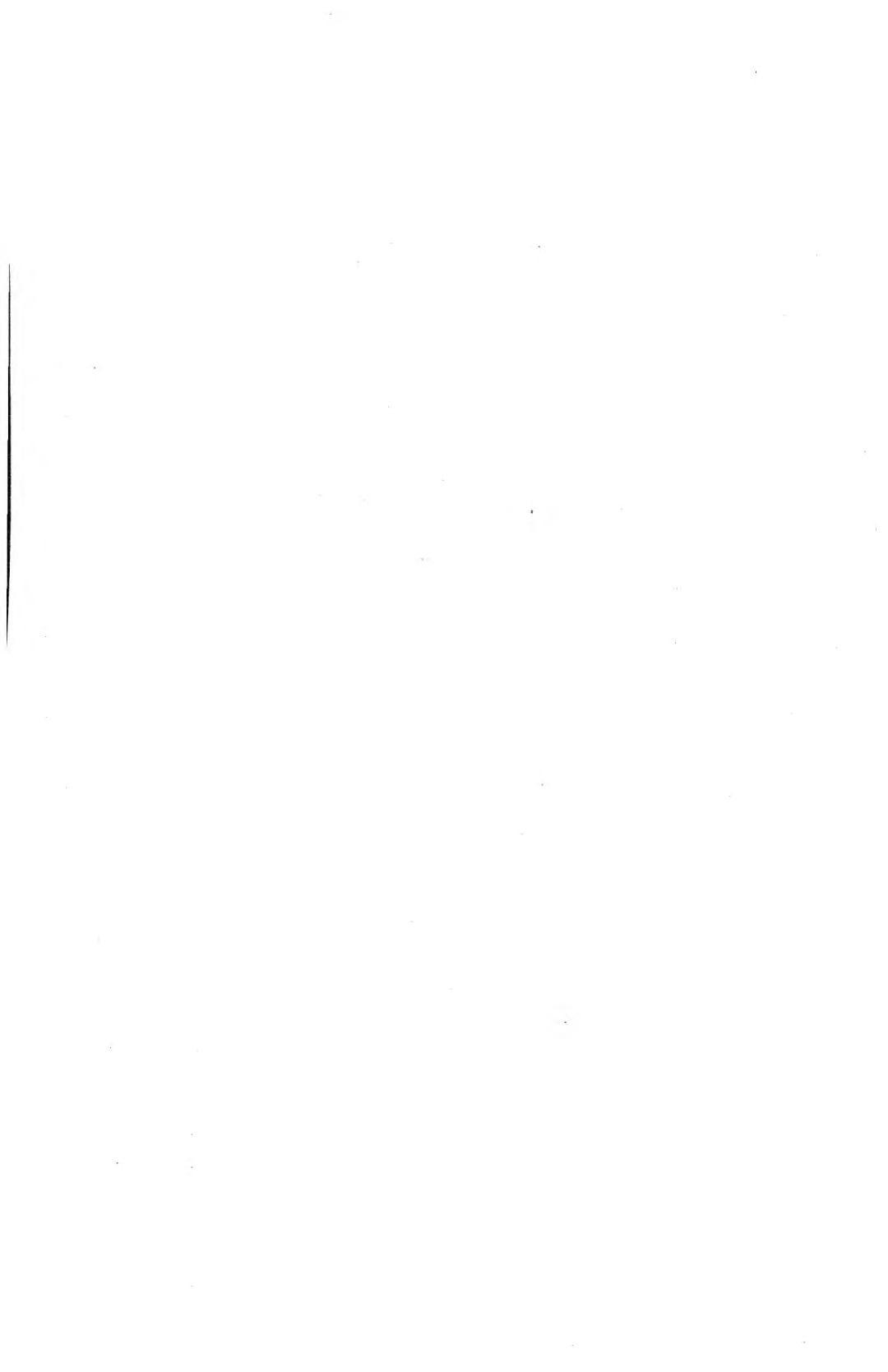
Tecknen *en* och *den* i sina försvagade betydelser faller utanför den pronomina indelningen. Jämte den obestämda artikeln *en* bör man tala om den bestämda fristående artikeln *den* – ej om adjektivets bestämda artikel, en term som med ordklassen pronomens här förutsatta omfattning blir missvisande.

Sammanlagt har alltså materialet fördelats på tretton olika grupper: nio definitiva, nämligen personliga (*jag*), reciproka (*varandra*), distributiva (*var sin*), intensitiva (*själv*), demonstrativa (*den där*), komparativa (*sådan*), relativa (*som*), selektiva (*den förre*) och totala (*alla*), samt fyra indefinita, nämligen kvantitativa (*många*), limitativa (*somliga*), interrogativa (*vem*) och generella (*vem som helst*). Fem av grupperna – personliga, reciproka, demonstrativa, relativa och interrogativa – har betecknats med traditionella termer och har, bortsett från de personliga, givits ett i stort sett traditionellt innehåll (jfr dock särskilt de demonstrativa). Fyra grupper – komparativa, selektiva, totala och kvantitativa – har betecknats med av Loman föreslagna termer. Termernas betydelse har emellertid skärpts, varför dessa gruppers innehåll i större eller mindre grad avviker från det av Loman givna. För de fyra återstående grupperna har de nya termerna distributiva, intensitiva, limitativa och generella pronomen föreslagits.

Från praktisk synpunkt kan antalet grupper och därmed antalet termer förefalla stort. En indelning på semantisk grund är givetvis, liksom varje annan indelning, i viss mån en avvägningsfråga. I valet mellan färre men semantiskt heterogena grupper och fler men semantiskt homogena grupper har det ansetts tjänligt att inom vissa gränser ge det senare alternativet företräde. Måhända kan fördubblingen av antalet grupper och därmed antalet termer i jämförelse med den traditionella indelningen uppvägas av att man vinner en bättre överblick över ordklassen än tidigare och att betydelseförhållandena framstår klarare.

*Korrekturnot:* I en uppsats i Modersmåslärarnas förenings årsskrift 1957, s. 63 ff., vartill jag nu kan hänvisa, behandlar Rudolv Körner *var sin*, "ett enhetligt uttryck med distributiv betydelse" (s. 64).







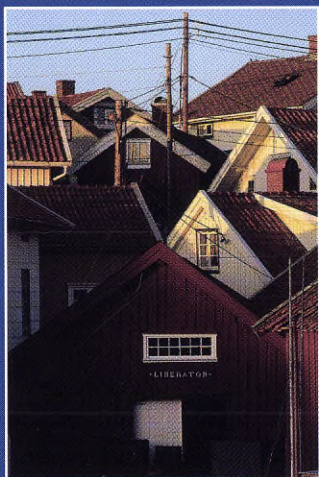


Meijerbergs Arkiv för svensk ordforskning (MASO) utges av styrelsen för Meijerbergs institut för svensk etymologisk forskning vid Göteborgs universitet. Styrelsen består av Filosofiska fakulteternas gemensamma donationsnämnds ordförande, biträdande professor Britt-Mari Näsström, ordförande, professorn i jämförande språkforskning med sanskrit Folke Josephson, professorn i nordiska språk Bo Ralph (självskrivna ledamöter), samt professor i modern svenska Lars-Gunnar Andersson, utsedd av humanistiska fakulteten.

Arkivet utges i tvångsfria delar om minst fem ark.

Prenumeration kan verkställas i bokhandel eller direkt hos Institutet, adress Institutionen för svenska språket, Box 200, 405 30 Göteborg.

Bidrag till Arkivet torde sändas till samma adress.



**STURE ALLÉN** har varit verksam inom svensk språkvetenskap sedan 1950-talet. Under denna tid har han lämnat viktiga bidrag på flera olika områden. Ett antal av hans artiklar har här samlats till en

volym. Urvalet omfattar i första hand sådana arbeten som på olika sätt belyser svenskan, i enlighet med bokens titel: *Modersmålet i fäderneslandet*. Strängt taget alla bidragen har varit publicerade tidigare i någon form men kan nu vara svåråtkomliga.

Parallellt med – eller snarare som en del av – utforskandet av svenskan har Sture Allén byggt upp ett särskilt examensämne vid Göteborgs universitet, språkvetenskaplig databehandling. Till detta har en professur varit knuten sedan 1972. Sture Allén var dess förste innehavare, fram till sin pensionering vid utgången av 1993. Många av bidragen återspeglar ämnets framväxt och utveckling. En del bidrag anknyter också till Sture Alléns arbete inom Svenska Akademien.

*Modersmålet i fäderneslandet* ger prov på ett mångsidigt vetenskapligt författarskap med en tyngdpunkt i svensk vokabulärforskning. Det är naturligt att boken ges ut i serien Meijerbergs arkiv för svensk ordforskning, som nr 25, speciellt som Sture Allén under en lång tid var engagerad inom Meijerbergs institut, senast i dess styrelse. Omslagsbilderna, som erinrar om Sture Alléns västsvenska ankargrund, har Mats Johansson som upphovsman.