

Judy Ribeck

Steg för steg

Data linguistica 28

<<http://svenska.gu.se/publikationer/data-linguistica>>

Editor: Lars Borin

Språkbanken
Department of Swedish
University of Gothenburg

28 • 2015

Judy Ribeck

Steg för steg

Naturvetenskapligt ämnesspråk som räknas

Göteborg 2015

Data linguistica 28
ISBN 978-91-87850-59-2
ISSN 0347-948X
E-publication <http://hdl.handle.net/2077/40506>

Printed in Sweden by
Ineko AB Göteborg 2015

Typeset in $\text{\LaTeX} 2_{\epsilon}$ by the author

Cover design by Kjell Edgren, Informat.se, and Sven Lindström

Front cover illustration by Anna Ehrlemark ©

Author photo on back cover by My Nyström

ABSTRACT

In this work, I present a linguistic investigation of the language of Swedish textbooks in the natural sciences, i.e., biology, physics and chemistry. The textbooks, which are used in secondary and upper secondary school, are examined with respect to traditional readability measures, e.g., LIX, OVIX and nominal ratio. I also extract typical linguistic features of the texts, typicality being determined using a proposed quantitative method, labelled *the index principle*. This empirical, corpus-based method relies on automatic linguistic annotations produced by language technology tools to calculate what I call *index lists*, rank-ordered lists of characteristic linguistic features of specific text corpora as compared to reference texts.

I produce index lists for typical vocabulary, noun phrase structures and syntactic structures, extracted from a 5.2 million word textbook corpus, compiled as a part of the work presented. As well as being frequent and well dispersed, the linguistic variables selected for the index lists are also characteristic of the text type in question, as is evident when they are compared to a reference corpus, comprising textbooks in the social sciences and mathematics, as well as narrative and academic (university-level) texts.

The results show that textbooks in natural science contain a lot of content-specific, technical vocabulary. This characteristic not only distinguishes natural scientific language from everyday language, but also from social scientific language, which on the lexical level has more in common with narrative texts. On the other hand, the textbook language as a whole is structurally distinguishable from narrative texts, as clearly seen, e.g., in its noun phrase complexity.

In the transition between secondary and upper secondary school, the scores of almost every readability measure go up, indicating an increase in linguistic demands on the readers. In the upper secondary textbooks the words are longer, the vocabulary more varied, the noun phrases longer and more elaborate, and the most typical syntactic structures more complex. Notably, the linguistic development between the form levels is more marked in the natural-science textbooks, compared to social sciences and mathematics. Nevertheless, the textbook language overall shows a relatively low complexity in comparison to academic language.

TILLKÄNNAGIVANDEN

– *Kan du beskriva hur det känns?*

– *Åh, jag tackar Gud!*

Så där ja. Då var det avhandlat.

Svårare är att jag, efter dessa fem år, också är tack skyldig alldeles för många människor alldeles för mycket för att lyckas uttrycka det till fullo, här, i denna form. Men försöka duger, och där bortom himlen finns ju, som bekant, en evighet.

Låt mig så börja med att rikta ett viktigt tack till Elisabet Engdahl, som uppmuntrade mig att söka till forskarutbildningen och som sedan har stöttat mig på vägen. Mina allra innerligaste tack går därefter oavkortat till mina handledare [här krävs en elgitarr]: Lars Borin och Emma Sköldberg. Tack vare att ni är olika på de flesta tänkbara vis har ni lyckats erbjuda vägledning om allt mellan högt och lågt. Lars, tack för detaljerna och sakligheten, och för en svindlande bredd och metaforik. Tack också för högt ställda förväntningar och krav. Emma, tack för helheterna och känsligheten, och för alla noggranna och stilsäkra textgranskningar. Tack också för support och vänskap, genom allt.

Denna avhandling har skrivits inom ramen för forskarskolan *Ämnesspråk i matematiska och naturvetenskapliga praktiker*. Jag vill varmt tacka alla som har deltagit på de återkommande forskarskoleträffarna under dessa år. Speciellt Åsa af Geijerstam, Caroline Liberg och Jenny Wiksten Folkeryd, för löpande respons och stöd. Speciellt också mina värdefulla doktorandkollegor i forskarskolan: Ida Bergvall, Anneli Dyrvold, Tomas Persson och Marie Ståhl. Jag kan inte föreställa mig de här åren utan vår samlade omtanke om varandra. För att inte tala om alla skratt vi delat. Hade vi inte bestämt att också ses framöver hade jag aldrig velat disputeras.

Hjärtligt tack också ni som har speciellläst mitt avhandlingsmanus under arbetets gång. Caroline Liberg (igen!), som kom med värdefulla råd på ett tidigt stadium. Björn Melander, som genomförde en rigorös granskning och bjöd upp till en för avhandlingen avgörande diskussion. Jussi Karlgren, som med sin smittande entusiasm gav arbetet en rejäl skjuts framåt. Johan Prytz,

som bidrog med nya perspektiv och mycket uppmuntran. Och Kristina Nilsson Björkenstam, min slutseminarieopponent, som genomförde en mycket kompetent läsning av texten och förmedlade flera alldeles nödvändiga synpunkter.

Jag vill också tacka mina kontakter på läroboksförlagen för ett gott samarbete: Åsa Stenholt Vernerson, Siv Strömquist och Ulf Wagner på Gleerups, Marie Carlsson, Kenneth Olsson och Sture Sahlström på Liber och Katarina Hellström och Anu Seensalu på Natur och Kultur. Ni har i högsta grad underlättat mitt arbete.

Utan tvekan kommer jag att minnas min tid på forskarutbildningen som god. För detta vill jag, framför allt, tacka mina medarbetare på institutionen för svenska språket vid Göteborgs universitet. Särskilt alla i källaren på våning två, för många stimulerande samtal med högt i tak. Men mest världens bästa rumskompis, Lisa Loenheim, för allt vi har kunnat dela.

Några andra medarbetare som har hjälpt mig lite extra är, i bokstavsordning: Gerlof Bouma, Carina Carlund, Karin Cavallin, Dana Danélls, Inga-Lill Grahn, Martin Hammarstedt, Katarina Heimann Mühlenbock, Per Holmberg, Rickard Johansson, Benjamin Lyngfelt, Andreas Nord, Henrik Rosenkvist, Johan Roxendal, Rudolf Rydstedt och Sofia Tingsell. Tack ska ni ha!

I detta sammanhang vill jag även särskilt tacka två doktorandkollegor, som varit med från början: Kristian Blensenius, vars oefterhärmliga stil förgyllt många lunchraster, och Håkan Jansson, vars djupa engagemang livat upp diverse diskussioner. Stort tack också till resten av mina goda doktorandvänner för en mycket fin gemenskap, som har givit mig en känsla av tillhörighet. Och så mycket skoj vi har haft! Tiden med er har verkligen varit värdefull.

Denna förmånliga yta vore hopplöst skralt utnyttjad om jag inte avslutningsvis tar tillfället i akt och tackar mina grundpelare. Till att börja med, för alla svaren, tack Gagga (jag ser allt hur du ler stolt, där, bakom allt). För gränslösheten och det höga siktet, tack Häggis. För språket och de ovillkorliga vingarna, tack mamma. Till sist, för att du vill stå vid min sida och känna *vinden*, tack My. Tillsammans är ni allt jag är och hoppas kunna bli. Den här boken är till er.

Så där ja. Då var också det avhandlat. Åtminstone efter bästa förmåga. Så, om det nu inte blev som det ska, får det faktiskt vara som det blev.

Det finns en ny början. Nu ökar vi!

Judy Ribeck
Göteborg 21 oktober 2015

INNEHÅLL

Abstract	i
Tillkännagivanden	iii
1 Inledning	1
1.1 Problemformulering	2
1.1.1 Språkets betydelse för lärande	2
1.1.2 Ämnesspråkliga stadiövergångar	3
1.1.3 Naturvetenskapliga läroböcker	4
1.2 Syfte	6
1.2.1 Frågeställningar	7
1.2.2 Avgränsningar	8
1.2.3 Om undersökningens relevans	8
1.3 Avhandlingens disposition	9
2 Teoretisk bakgrund och tidigare forskning	11
2.1 Om skolans språk	11
2.1.1 Vardagsspråk och skolspråk	11
2.1.2 Ämnesspråkliga komponenter	14
2.1.3 A- och B-texter	17
2.2 Akademiskt språk	20
2.2.1 Engelskspråkig forskning	20
2.2.2 Svenskspråkig forskning	21
2.2.3 Engelskspråkiga akademiska ordlistor	23
2.2.4 En svensk akademisk ordlista	26
2.3 Om att undersöka språk i text	29
2.3.1 Kvantitativ stilistik som analysmetod	30
2.3.2 Datamaskinell textanalys	31
2.4 Läroboksforskning	36
2.4.1 Läroböckers roll	36
2.4.2 Studier av lärobokstexter	37
2.4.3 Språket i naturvetenskapliga läroböcker	42
2.4.4 Ord i läromedel – OrdiL	43

3	Material: insamling	45
3.1	Lärobokstexter	46
3.2	Akademiska texter	49
3.3	Berättande texter	53
3.4	Översikt över materialunderlaget	57
4	Material: bearbetning	59
4.1	Lingvistisk annotering	59
4.2	Rensning av texter	62
4.2.1	Lärobokstexter	64
4.2.2	Akademiska texter	68
4.2.3	Berättande texter	71
4.3	Överlappande läroböcker	72
5	Metod	73
5.1	Ämnesspråklig avgränsning	73
5.2	Avhandlingens delstudier	75
5.3	Indexprincipen	78
5.3.1	Selektionskriteriet hög frekvens med jämn spridning	78
5.3.2	Selektionskriteriet texttypiskhet	79
5.4	Delstudiernas övergripande upplägg	80
6	Delstudie: traditionella textanalysmått	85
6.1	Ytliga textanalysmått: presentation	85
6.2	Djupare textanalysmått: presentation	87
6.3	Ytliga textanalysmått: resultat i förhållande till tidigare forskning	88
6.3.1	Meningslängd	88
6.3.2	Ordlängd	91
6.3.3	Läsbarhetsindex, LIX	94
6.3.4	Ordvariationsindex, OVIX	96
6.4	Djupare textanalysmått: resultat i förhållande till tidigare forskning	98
6.4.1	Ordklassfördelning	98
6.4.2	Nominalkvot, NQ	99
6.4.3	Lemmavariationsindex, LVIX	101
6.5	Traditionella textanalysmått: sammanfattning och diskussion	103
7	Delstudie: typiskt ordförråd	107
7.1	Typiskt ordförråd: studiens upplägg	107
7.2	Typiskt ordförråd: resultat	110
7.2.1	Typiska ord	110
7.2.2	Fokusstudier	129

7.3	Typiskt ordförråd i naturvetenskap: sammanfattning och diskussion	136
7.3.1	Förhållandet till ordförråd i matematik och samhälls- vetenskap	136
7.3.2	Stadieövergången mellan högstadiet och gymnasiet . . .	137
7.3.3	Förhållandet till ordförråd i akademiska texter	137
7.3.4	Förhållandet till ordförråd i berättande texter	138
8	Delstudie: typiska nominalfraser	139
8.1	Nominalfrasers funktion och uppbyggnad	140
8.2	Nominalfraser av olika komplexitet	142
8.3	Typiska nominalfraser: studiens upplägg	143
8.4	Typiska nominalfraser: resultat	145
8.4.1	Nominalfrastäthet	145
8.4.2	Nominalfraslängd	147
8.4.3	Nominalfrasuppbyggnad	153
8.4.4	Fokusstudier	170
8.5	Typiska nominalfraser i naturvetenskap: sammanfattning och dis- kussion	177
8.5.1	Förhållandet till nominalfraser i matematik och sam- hällsvetenskap	177
8.5.2	Stadieövergången mellan högstadiet och gymnasiet . . .	177
8.5.3	Förhållandet till nominalfraser i akademiska texter	178
8.5.4	Förhållandet till nominalfraser i berättande texter	179
9	Delstudie: typiska meningar	181
9.1	Typiska meningar: studiens upplägg	181
9.2	Typiska meningar: resultat	183
9.2.1	Parsningsdjup	183
9.2.2	Meningsuppbyggnad	189
9.2.3	Fokusstudier	223
9.3	Typiska meningar i naturvetenskap: sammanfattning och diskussion	226
9.3.1	Förhållandet till meningar i matematik och samhälls- vetenskap	226
9.3.2	Stadieövergången mellan högstadiet och gymnasiet . . .	227
9.3.3	Förhållandet till meningar i akademiska texter	228
9.3.4	Förhållandet till meningar i berättande texter	229
10	Sammanfattning och diskussion	231
10.1	Språkliga drag i naturvetenskapliga ämnesspråk	232
10.1.1	Stadieövergången mellan högstadiet och gymnasiet . . .	232
10.1.2	Förhållandet till ämnesspråk i matematik	233

10.1.3	Förhållandet till ämnesspråk i samhällsvetenskapliga ämnen	234
10.1.4	Förhållandet till språket i akademiska texter	235
10.1.5	Förhållandet till språket i berättande texter	237
10.1.6	Slutsatser	238
10.2	Metodologiska reflektioner	240
10.3	Framtiden	243
	Litteraturförteckning	245
	Appendix	269
A	Förteckning över terminologi	269
A.1	Återkommande förkortningar	269
A.2	Lingvistisk notation	270
B	Materialförteckning	273
B.1	Undersökta läroböcker	273
B.2	Tidskrifter i den akademiska korpusen	277
B.3	Romaner i korpusen med berättande texter	278
C	Rensningsregler	281
D	Pseudokod för indexprincipen	283
E	Samlade resultat	289
E.1	Delstudie: traditionella textanalysmått	289
E.2	Delstudie: typiska ord	303
E.3	Delstudie: typiska nominalfraser	331
E.4	Delstudie: typiska meningar	333

1

INLEDNING

Morgonen efter riksdagsvalet 2014 antyds i Aftonbladets ledare att valet avgjordes redan den 3 december året tidigare. Det var dagen för den s.k. ”Pisachocken”, som, till skillnad från Afro-Dites ”nakenchock” under repetitionerna till Eurovisionsschlagerfestivalen i Tallinn 2002, inte var det minsta sexig. I stället stod det otäckt klart att kunskapsresultaten för svenska skolor alljämt rasade inom samtliga områden som testas: läsförståelse, matematik och naturvetenskap (Skolverket 2014a). Pisachocken kom efter en längre period av diskussioner om problem i den svenska skolan och bedömdes vara det yttersta beviset för regeringen Reinfeldts misslyckande. Dagens Nyheter (2013) rapporterar nyheten med följande ord:

I världens mest prestigefyllda jämförande kunskapsstudie har svenska 15-åringar ännu en gång försämrat sina resultat. Av alla 34 deltagande OECD-länder uppvisar inget annat en så negativ utveckling. Sverige sticker ut oavsett om man jämför med den förra PISA-undersökningen för tre år sedan eller med den första, PISA 2000.

PISA:s senaste läsförståelsetest visar att hela 23 % av de svenska eleverna är ”svaga läsare”, vilket betyder att de bara klarar att ta till sig och tolka de enklaste texter och t.ex. har svårt att läsa läroböcker på gymnasiet (Skolverket 2013; Svenska Dagbladet 2013).

Pisachocken är en av de bidragande orsakerna till att skolfrågan under en tid har placerats högt upp på den politiska dagordningen. Varför det går sämre för Sverige är naturligtvis en komplicerad sak att utreda. Det är emellertid uppenbart att elevers läsförmåga går parallellt med resultaten för proven i naturvetenskap. Flera nationella undersökningar visar dessutom att en likartad resultatnedgång just inom områdena för läsförståelse och naturvetenskap påbörjas redan före de omfattande skolreformerna på 1990-talet (Holmlund m.fl. 2014: 125). Sammantaget väcks därför funderingar kring om det kan finnas ett samband mellan prestationer inom dessa båda kunskapsfält.

2 Inledning

I denna avhandling presenterar jag resultaten av en undersökning som har designats för att indirekt pröva antagandet att sämre läsförståelse gör det svårare att ta till sig naturvetenskapliga ämneskunskaper i skolan. Jag kommer nämligen att undersöka naturvetenskapliga ämnestexters språkliga karaktäristik, i förhållande till språket i andra skolämnen och i förhållande till språket i berättande och akademiska texter. Utgår man från att texters språkliga utformning påverkar läsförståelsen, borde resultat från en sådan jämförande studie kunna tala för eller emot förklaringsförsöket. Några indicier som skulle kunna ge stöd för hypotesen är ifall språket i naturvetenskapliga texter visar sig vara mycket likt språket i akademiska texter, som studenter upplever som svårt (se avsnitt 2.2.2), eller mycket avvikande från språket i berättande texter, vilket elever sägs behärska (se avsnitt 2.1.3).

1.1 Problemformulering

1.1.1 Språkets betydelse för lärande

Skolutbildning är i allt väsentligt av språklig natur. Samtidigt som undervisningen syftar till att utveckla språkliga färdigheter, som att lyssna, tala, läsa och skriva, är det företrädesvis genom sådana språkliga aktiviteter som kunskapen kan förmedlas, tas emot och återges (Anward 1984; Skolverket 2014b). Elevers påvisbara ämneskunskaper är beroende av olika ämnens språkliga krav och det är ofta svårt, om inte omöjligt, att separera språk och kunskap.

I läroplanen för gymnasieskolan fastslås att eleverna på det naturvetenskapliga programmet ska lära sig att producera ”texter som innehåller ett naturvetenskapligt språk” (Skolverket 2011a: 249). Eleverna i den svenska skolan utgör dock en heterogen grupp, och när de börjar skolan är de i mycket varierande utsträckning bekanta med det språkbruk som skolans ämnesundervisning präglas av. Särskilt det naturvetenskapliga språket skiljer sig i många avseenden från både vardagligt språk och det språk som används i andra skolämnen (t.ex. samhällsvetenskapliga ämnen), vilket sannolikt har konsekvenser för elevers framgång i naturvetenskap.

För att hjälpa eleverna satsar Skolverket starkt på att få lärare från alla ämnen att arbeta med språkutvecklande undervisning. På senare tid har också språk som används i skolundervisningen, *ämnesspråk*, fått ökad uppmärksamhet i den akademiska debatten. I en aktuell avhandling konstaterar Hipkiss (2014) att behärskning av olika ämnesspråk är avgörande för elevers skolframgång, inte minst som förberedelse för mötet med det akademiska språk som används på universitet och högskolor.

Engelskt ämnesspråk har varit föremål för forskning under några årtionden. Svenskt ämnesspråk är emellertid fortfarande relativt outforskat. Några betydelsefulla bidrag kommer ändå från projektet *Elevens möte med skolans textvärldar*. Detta projekt pågick 1999–2006 och syftade till att undersöka hur skriftspråk i skolan ser ut och varierar mellan olika ämnen (Liberg m.fl. 1999). I de resulterande avhandlingarna härifrån konstateras bland annat att skolan inte till fullo lyckas hjälpa elever att uppnå de språkliga mål som krävs för att lyckas i matematik och naturvetenskapliga ämnen (jfr Edling 2006; af Geijerstam 2006). Nygård Larssons avhandling, *Biologiämnets texter* (2011), som utreder multimodal kohesion i biologitexter, är också ett viktigt bidrag till forskningen om svenskt ämnesspråk.

1.1.2 Ämnesspråkliga stadiövergångar

Stadieövergångarna i skolgången, t.ex. den mellan högstadiet och gymnasiet, utgör markerade brytpunkter, såväl ämnesmässigt som språkligt och socialt (Ask 2005, 2007; Einarsson 2005). För vissa individer är sådana övergångar mycket besvärliga. En förklaring till svårigheterna är att elever i stadiövergångarna möter hastigt ökade ämnesspråkliga krav, inte minst på läsförmåga, i olika läroböcker.

Speciellt stadiövergången till universitetsstudier innebär en ”diskurschock” för många studenter.¹ Eftersom det ”att uttrycka sig i enlighet med accepterade normer är inträdesbiljetten till en textkultur” (Ask 2005: 89), måste nyblivna studenter tillägna sig de nya stadiernas sätt att uttrycka mening, d.v.s. hur de förväntas skriva och tala om ämnet i fråga. Det akademiska språket representerar ett nytt, mer abstrakt, distanserat och kritiskt tankesätt, där förhållandet till verkligheten bygger på språkliga abstraktioner formulerade i teorier och begrepp (Trondman 1994). Mycket är underförstått i den akademiska textkulturen; den präglas av tyst kunskap och osynliga krav, som synliggörs först när någon bryter mot konventionerna (Blåsjö 2004; Ehn och Löfgren 2004; Ask 2005).

Teleman (1988) menar att all språkanvändning i huvudsak är av den kunskapsart han kallar *procedurkunskap*, vilket motsvarar det att **kunna** läsa, skriva eller tala. Det är emellertid svårare att beskriva hur språket används i olika situationer, eftersom den *propositionella* kunskapen kring processerna ofta saknas; språkbrukare **vet** inte att de gör så eller så. Samtidigt är det väl känt att inlärare har nytta av ”propositionell kunskap om den process som aktiviteten utgör och om kraven på den produkt som aktiviteten ska avkasta” (Teleman

¹Jfr också Bratt 1985: 131 om chocken i övergångarna från låg- till mellanstadiet och från mellan- till högstadiet.

4 Inledning

1988: 7). Trots detta verkar man ibland i skolan tro att procedurkunskap uppstår av sig själv, när det som behövs är vägledning vid olika stadier i elevernas utveckling (Teleman 1988).

Vissa menar att studenterna själva har ansvar för att lära sig ämnesspråken. Dysthe, Hertzberg och Løkesgaard Hoel (2011) förklarar t.ex. att varje kulturell gemenskap delar en historia och vissa underförstådda idéer som inte uttryckligen behöver refereras till. Det är därför inte konstigt att studenter känner sig utanför när de t.ex. börjar läsa ett nytt ämne eller börjar en ny utbildning. De träder emellertid in i ämneskulturen genom att lära sig mer av ämneskunskapen och genom att läsa material och lyssna på människor som tillhör denna kultur.

Åsikten om inlärares eget ansvar i samband med studieövergångar delas dock inte av alla. Ett av de huvudsakliga problem som Ask (2005) pekar ut, i samband med studieövergången till universitetet, är just det att studenterna ofta förväntas tillägna sig kunskap om textnormerna på eftergymnasiala utbildningar på egen hand. En ytterligare komplikation sägs vara att studenterna ofta saknar de skriftspråkliga färdigheter som systemet förutsätter att de har.

I det svenska utbildningssystemet finns traditionellt sett en progressionsaspekt som bygger på att varje institution förser deltagarna med kunskaper och färdigheter som är nödvändiga för att gå vidare till nästa nivå. Skolans ansvar är att ge eleverna en generell skrivkompetens, som de har användning av i privatlivet, i den fortsatta utbildningen och i arbetslivet. Gymnasieskolan (åtminstone de studieförberedande programmen) har ansvar för att förbereda studenterna för skrivande på högre utbildningar, något som explicit finns inskrivet i läroplanen för gymnasieskolan (Skolverket 2011a). Däremot har det aldrig varit skolans uppgift att synliggöra normerna för det akademiska skrivandet; fackskrivandet i den högre utbildningen ska respektive institution ta sig an (Ask 2007; Løkesgaard Hoel 2010).

Studenter deltar i flera ämnen och därmed i flera ämneskulturer, vilka var och en har sitt sätt att beskriva världen, sitt sätt att bygga upp kunskap och sitt eget språk för att utveckla och beskriva denna kunskap. Både nordiska och internationella studier visar att sådana skillnader sällan explicit görs för studenterna (Chandler 1995; Dysthe m.fl. 2006).

1.1.3 Naturvetenskapliga läroböcker

Det råder ingen tvekan om att skolan i mångt och mycket är en textburen verksamhet, i vilken lärobokstexter förekommer som ”en central och integrerad del” (jfr Karlsson och Strand 2012: 114). Enligt Säljö (2014: 221) är läroboken i sig en av de viktigaste orsakerna till höjningen av de kognitiva och kommuni-

kativa kraven på elever i skolan: ”Svårigheter att tillgodogöra sig information via text märks tydligare när läroboken får en så central ställning, när lärandet blir språkligt till sin natur och när en så stor andel av befolkningen förväntas lära sig via text.”

Enligt Ekvall (2011) har naturvetenskapliga lärobokstexter gång på gång beskrivits som svårtillgängliga för elever. Men trots att elever många gånger behöver hjälp med att läsa naturvetenskapliga läroböcker, läggs alltför litet tid på textläsning i undervisningen. Hipkiss undersökning (2014) pekar särskilt på att elever endast kommer i kontakt med ämnesspråken via läroböcker, men inte i klassrumsundervisningen.

I figur 1 återges ett kort exempel på ett avsnitt ur en biologilärobok för högstadiet. Det räcker med en snabb titt på utdraget för att känna igen det som hämtat från en lärobok. En faktastinn rubrik formulerad som en allmän sanning följs av ett ganska kort textstycke i resonerande form. Man behöver heller inte läsa brödtexten för att sluta sig till att det är en naturvetenskaplig text, snarare än t.ex. samhällsvetenskaplig. Skriftspråket varvas med bilder från naturen och bildtexten är informativt formulerad, om än med en högst diffus koppling till bilden (jfr Wallin Victorin 2011).



För att en isbjörn ska bli fullvuxen måste den äta många säl. I slutändan är det mängden växtplankton som styr hur stora populationerna kan bli.

Mängden producenter sätter gränsen

En isbjörnshanne väger cirka 500 kg. Vad har han ätit för att bli så stor? Jo, vanligen säl. För enkelhets skull räknar vi med att det bara är 10 % av sälens vikt som blir till isbjörnskött. Då har isbjörnen behövt äta ungefär fem ton säl för att nå sin vikt. Säl i sin tur äter gärna torsk. För att få fem ton säl behövs då cirka 50 ton torsk. Och för att få 50 ton torsk behövs ungefär 500 ton sill. Sillarna behöver i sin tur stoppa i sig cirka 5 000 ton djurplankton – som i sin tur behöver äta ofattbara 50 000 ton växtplankton!

Med hjälp av den här förenklade räkneövningen förstår du att det inte kan finnas fler isbjörnar i ett ekosystem än vad mängden växtplankton, och deras fotosyntes, sätter gränsen för. Ju fler trofinivåer, desto färre toppkonsumenter. Detsamma gäller för rovdjuret i de svenska skogarna, örnarna längs våra kuster och lejonerna på Afrikas savanner.

Figur 1: Utdrag från en lärobok i biologi för högstadiet (Fabricius, Holm och Nystrand 2013: 164). Återges här med tillstånd från Liber

Moderna läroböcker är i hög grad multimodala (Nygård Larsson 2009) och innehåller också många olika textslag, som t.ex. rubrik-, bild-, bröd- och uppgiftstext (Ekvall 2010). Naturvetenskapliga texter kräver dessutom att mottagaren har både en god allmän läsförståelse och förståelse för formler, tabeller och diagram. I texterna används dessa semiotiska system kompletterande, ef-

6 Inledning

tersom det är omöjligt att koda alla naturvetenskapliga idéer med endast språkliga medel (Lemke 2000). För att komma åt alla betydelser i texterna räcker det därför inte att bara analysera språket; man måste också se till det semiotiska bidrag som t.ex. bilder och typografi står för (Björkvall 2010). I takt med den tekniska utvecklingen har också multimodala analyser av texter blivit ett allt vanligare inslag i textforskningen. Parallellt med sådana arbeten lever dock den mer renodlat språkvetenskapliga traditionen av att undersöka text som språklig struktur och att försöka karakterisera det språkliga systemet i en text eller i en grupp av texter kvar (jfr Svensson och Karlsson 2012 och Karlsson och Strand 2012).

1.2 Syfte

För att teoretiskt kunna underbygga framtidens läroplaner och lärarledda undervisning behövs kunskap om specifika språkliga krav för att lyckas i skolan. Sådana frågor har inte utretts tillräckligt väl. I synnerhet anses ytterligare forskning som utgår från empiriska ämnesspråkliga data behövas (jfr Lindberg 2009). För att kunna lära elever behärska de språkliga verktyg som är nödvändiga, krävs en ökad förståelse för hur olika ämnesspråk är konstruerade – på högstadiet, på gymnasiet och på eftergymnasiala utbildningar. Detta gäller inte minst för naturvetenskapliga ämnen. Härtill fogar Strömquist (1995: 12) en uppmaning om att särskilt uppmärksamma språket i läroböcker. Hon efterlyser uttryckligen mer systematiska och heltäckande studier av hur svenskt läroboksspråk ser ut, både inom olika ämnen och på olika stadier i skolan.

I denna avhandling undersöker jag hur ämnesspråken i biologi, kemi och fysik byggs upp och utvecklas under skoltiden. Jag vill också redogöra för kännetecknen för de naturvetenskapliga ämnesspråken i förhållande till språket i akademiska och berättande texter. Ett ytterligare mål med avhandlingsarbetet är att utveckla och demonstrera en metod för att karakterisera språk i text med språkteknologiska hjälpmedel.

1.2.1 Frågeställningar

I avhandlingen tar jag mig an följande frågeställningar:

1. Vilka språkliga drag kännetecknar ämnesspråk i naturvetenskapliga ämnen
 - (a) på högstadiet,
 - (b) på gymnasiet?
2. Hur förhåller sig ämnesspråk i naturvetenskapliga ämnen, på högstadie- och gymnasienivå, till
 - (a) varandra,
 - (b) ämnesspråk i matematik på motsvarande stadier,
 - (c) ämnesspråk i samhällsvetenskapliga ämnen på motsvarande stadier,
 - (d) akademiskt språk,
 - (e) språk i berättande texter?

Den första frågeställningen går metodologiskt ut på att karakterisera de naturvetenskapliga ämnesspråken på de olika stadierna med fokus på språket som system och på läsbarhet (se vidare avsnitt 5.2 där samtliga variabler presenteras). Den andra frågeställningen angrips genom att resultaten av karakteriseringarna jämförs dels med varandra, dels med motsvarande resultat för, å ena sidan lärobokstexter i matematik och samhällsvetenskap, å andra sidan akademiska och berättande texter. Resultaten presenteras i form av beskrivningar – i det första steget av typiska ämnesspråkliga drag, och i nästa steg av vad stadieövergångarna innebär ur språklig synpunkt.

De stadier som undersöks inom ramen för detta arbete är i huvudsak högstadiet och gymnasiet, men övergången till universitetet behandlas också till viss del, eftersom ett av jämförelsematerialen innehåller akademiskt språk. Då karakteriseringsmetoderna även appliceras på akademiska texter, finns en förhoppning om att också, i viss utsträckning, kunna beskriva hur det svenska akademiska språket typiskt är konstruerat.

8 Inledning

1.2.2 Avgränsningar

I avhandlingen används *ämnesspråk* i betydelsen 'det språk som elever behöver tillägna sig för att beskriva och forma ämneskunskaper i skolan' (jfr Blåsjö 2009). Ämnesspråken undersöks vidare primärt genom att studera ämnestexter, närmare bestämt svenska lärobokstexter, alltså skriftspråkliga källor. Elevproducerade texter eller muntlig kommunikation inom skolundervisningen ingår inte i studien.

Eftersom jag förstår *språk* i enlighet med hur termen används inom traditionell språkvetenskap, omfattar ämnesspråksdefinitionen inga andra semiotiska system. Bilder, grafer och matematiska symboler omtalas i vissa sammanhang också som olika former av "språk", i uttryck som "bildspråk" eller "symbolspråk". Sådana element förekommer också i ämnestexterna och behöver behärskas av eleverna. När jag talar om "ämnesspråk" i ämnestexter avser jag dock enbart den språkliga komponenten.

Min förståelse för ämnesspråkens olika delar har vidare påverkats av hur *literacy* beskrivs i den australiska läroplanen (se vidare 2.1.2). Enligt detta synsätt ligger fokus i undersökningarna av ämnesspråk i denna avhandling på kunskap om ord och grammatik (men inte på kunskap om texters uppbyggnad).

Såväl material- som metodval diskuteras och motiveras utförligare i kapitel 3 och 5.

1.2.3 Om undersökningens relevans

I metodologiskt avseende vilar mitt arbete såväl på traditionell, språkvetenskaplig textforskning som på datalingvistisk textforskning. Det är just i syntesen mellan dessa forskningsfält som jag i första hand önskar bredda kunskapen. Jag vill helt enkelt, för språkvetare, påvisa hur språkteknologiska hjälpmedel kan göra storskaliga undersökningar av lingvistisk data möjliga, och, för andra datalingvister, visa hur våra verktyg kan användas för att besvara viktiga frågor om språk. Relevanta sådana exempel behövs, inte minst på svenska och på svenskt material.

Avhandlingsarbetet gränsar också till ämnesdidaktisk och pedagogisk forskning. I korsningen mellan undervisningsmetodik och språkvetenskap har undersökningen också relevans. I och med denna avhandling vill jag lämna ett empiriskt välgrundat teoretiskt bidrag till kunskapen om hur de naturvetenskapliga ämnesspråken är konstruerade, med en förhoppning om att mina resultat ska ge skolforskare, språkhandledare och andra pedagoger, som aktivt

arbetar med att stödja elevers utveckling av ämnesspråklig kompetens, underlag för vidare diskussion om hur en effektiv språkutvecklande undervisning ska utformas.

1.3 Avhandlingens disposition

Avhandlingen omfattar 10 kapitel.

I kapitel 2 behandlas teori och tidigare forskning, som på olika sätt är relevant för avhandlingsarbetet, med syfte att förankra undersökningen i ett större sammanhang. Den teoretiska kunskapen kommer från flera olika fält, bland annat från datamaskinell textanalys, skolspråksforskning, språkinlärning och textlingvistik.

I kapitel 3 beskrivs och motiveras de textmaterial som studien baseras på. Närmare bestämt rör det sig om korpusar med läroböcker för högstadiet och gymnasiet, samt akademiska och berättande texter.

I kapitel 4 förklaras hur textmaterialen har bearbetats inför undersökningarna. Här beskrivs den språkteknologiska annoteringen av texterna och hur utomspråklig data i texterna, t.ex. i form av matematiska formler, har tagits om hand.

I kapitel 5 redogörs för metodologiska val. Här beskrivs utformningen av den princip som huvudsakligen används för att karakterisera textmaterialen. Dessutom presenteras hur de olika delstudier som ingår i undersökningen är uppbyggda.

I kapitel 6–9 redovisas och diskuteras resultaten av avhandlingens fyra olika delstudier. Den första delstudien (kapitel 6) rör traditionella textanalysmått, som meningslängd, ordlängd, LIX och OVIX. Den andra delstudien (kapitel 7) gäller ordförråd. Den tredje delstudien (kapitel 8) handlar om nominalfraser och den fjärde delstudien (kapitel 9) rör meningar.

I kapitel 10 summeras de huvudsakliga resultaten. I kapitlet diskuteras också metodologiska aspekter av undersökningarna.

2

TEORETISK BAKGRUND OCH TIDIGARE FORSKNING

I detta kapitel sammanfattas forskning som på olika sätt är relevant för avhandlingsarbetet. Bakgrundsteckningen omfattar teoretisk kunskap från flera olika fält. Till att börja med behandlas språkinriktad skolforskning (avsnitt 2.1) och akademiskt språk (avsnitt 2.2). Därefter följer en genomgång av olika angreppssätt för att undersöka språk i text (avsnitt 2.3). Slutligen beskrivs tidigare forskning om läroboksspråk (avsnitt 2.4).

2.1 Om skolans språk

2.1.1 Vardagsspråk och skolspråk

Macken-Horarik (1996) gör en poäng av att skilja mellan tre domäner i vilka undervisning förekommer: den *vardagliga*, den *specialiserade* och den *reflexiva*. Den vardagliga domänen utgörs av hemmets värld och den miljö där barnet växer upp och i första hand socialiseras. Genom interaktion med viktiga förebilder, där språklig kommunikation utgör en naturlig del av samspelet, lär sig barnet familjens och samhällets till största del uttalade etik och moral.

Inom den specialiserade domänen studeras vetenskap genom formell kunskapsförmedling – en lärarledd (hegemonisk) skolundervisning, som förbereder elever för framtida yrken. Lärande i ett utbildningssammanhang innebär att språket används till att utveckla, fördjupa och redovisa kunskaper. I den nya verklighet som skolan utgör måste eleverna urskilja generaliseringar, som relaterar den tidigare vardagskunskapen till en ny systematiserad kunskap. Detta förväntas till stor del ske genom läsning av tekniska och ofta kraftigt metaforiska, skrivna texter. Att påstå att språket är en nyckel till att lyckas i skolan är därför ingen överdrift (jfr Lindberg 2009).

I den reflexiva domänen ses kunskap som en social konstruktion, öppen för granskning, utmaning och förändring. Här tillåts man ifrågasätta den specialiserade kunskapens grundvalar och utmana verkligheten genom kritiska kunskaper. Förståelse utvecklas för att samhället rymmer olika skolor som försöker

vinna anhängare, att olika discipliner står mot varandra, och att det i slutändan kan vara svårt att skilja sig själv och sina värdegrunder från det man vet och kan. Målet med skolgången anses vara att eleverna, i den reflexiva domänen, själva både kan och tillåts reflektera över sitt lärande. I den mån svenska elever/studenter kommer i kontakt med denna slutgiltiga undervisningsform, sker det på universitet och högskolor.

Uppfattningen att det existerar ett slags informellt språkbruk som skiljer sig från det som används inom ramen för undervisning återkommer inom många forskningsgrenar. Exempelvis använder andraspråksforskaren Cummins (1981) termen *BICS* (Basic Interpersonal Communication Skills) för att tala om det språkbruk som används i hemtama miljöer, i direkt kommunikation mellan människor ”här och nu”, och *CALP* (Cognitive Academic Language Proficiency), som är relaterat till läs- och skrivfärdigheter och i synnerhet används i undervisningssituationer. På liknande sätt skiljer Hulstijn (2011) mellan *BLC* (Basic Language Cognition), som är den gemensamma språkkunskap som modersmålstalare har, och *HLC* (Higher Language Cognition), som är färdigheter som modersmålstalare har i varierande utsträckning och som också inkluderar skriftspråkkunskap.

Inom svensk andraspråksforskning har Viberg (1993) myntat begreppet *bas*, som motsvarar den språkbehärskning som barn med svenska som modersmål har vid skolstarten. Hit hör t.ex. ljudsystemet, det morfologiska och syntaktiska systemet (med undantag för vissa drag bundna till vissa stilarter), basordförrådet och grundläggande diskursregler för ”allmänt behärskande av genrer såsom enkla berättelser” (Viberg 1993: 63). Inom engelskspråkig andraspråksforskning talas ofta om fenomenet *core-vocabulary* som närmast är att jämföra med Vibergs basordförråd (se t.ex. Nation 2001).

Föreställningen att barnen kommer till skolan med ett uniformt språkbruk är emellertid problematisk, inte minst för att den individuella variationen mellan elever är stor. Därtill behärskar barn flera olika sociolekter, dialekter och gruppspråk (jfr Anward 1984). Sociolingvistiska studier har också visat att sociala och regionala skillnader i barns uppväxtförhållanden har betydelse för utvecklingen och behärskningen av det språkbruk som uppmuntras i skolmiljön (jfr Bernstein 1971; Romaine 1984). Slutligen har andelen elever i den svenska skolan som inte har svenska som förstaspråk ökat dramatiskt under de senaste decennierna. För dessa elever utgör ordkunskap en särskild utmaning, eftersom deras vokabulär ofta är relativt begränsad, vilket i sin tur medför sämre läsförståelse (Iversen Kulbrandstad 1996; Golden 2005). Vad gäller flerspråkiga elevers utveckling av skolspråkliga drag på svensk mark konstaterar både Magnusson (2010) och Prentice (2010) att startåldern har en avgörande betydelse.

Trots uppenbara problem med att beskriva elevers språkkunskaper vid skolstarten som uniforma, gör också Skolverket (2014b) en tydlig distinktion mellan det språk som elever ”behärskar när de kommer till skolan” och det språk som används i skolundervisningen för att ”beskriva och forma ämneskunskaper” (Blåsjö 2009). Det förra kallas av Skolverket för *vardagsspråk* och det senare för *ämnesspråk* eller *skolspråk* (jfr också Chamot och O’Malley 1994; Bailey m.fl. 2002; Lindberg 2007; Magnusson 2008). I förhållande till vardagsspråket är ämnesspråken mer formella, mer abstrakta och oftare skriftspråkliga, även om variationen mellan skolans olika ämnen är stor.

Långt in på 1990-talet fanns ingen uttalad vetenskaplig disciplin som studerade språket som undervisningsmedel (i stil med psykolingvistik eller sociolingvistik). Termen *språkpedagogik* existerade visserligen, men syftade endast på undervisning i språk. På senare år har dock *ämnesspråk* kommit fram som forskningsområde, i ett försök att bättre fånga den speciella formen av språklig verksamhet som förekommer inom ramen för skolans undervisning. Termen definieras av Blåsjö (2009: 148) som det språk som eleverna behöver tillägna sig för att ”beskriva och forma ämneskunskaper i skolan”.

Skolundervisningen kräver att eleverna ska visa sina kunskaper med hjälp av språkliga uttrycksformer av olika slag. Ett välutvecklat ämnesspråk gör att eleverna kan förstå hur grundläggande principer i ämnet är organiserade och därmed gå in i ämnesspecifika diskussioner och föra relevanta resonemang, samt läsa och skriva de texttyper som dominerar i det aktuella ämnet. Ämnesspråk handlar alltså till stor del om att uttrycka, tolka, förstå och använda begrepp, fakta och centrala tankegångar i tal och skrift, men det inbegriper även de känslor och åsikter som ämnet väcker (Skolverket 2015).

Skolverket (2015) skriver att ämnesspråk är nyckeln till att utveckla kunskaper i alla skolans ämnen. Lärare behöver vara ämnesspråkligt medvetna och inte bara undervisa om ämnets innehåll, utan också om hur argumentationer byggs upp och förs inom ramen för ämnet. I riktlinjerna för skolans arbete med språkmedveten undervisning menar Skolverket (2014b) att lärare kontinuerligt måste analysera det aktuella ämnets texter, för att öka sina egna kunskaper om ämnesspråket och därigenom kunna identifiera sådant som rent språkligt kan utgöra hinder för elevernas kunskapsutveckling. Genom ett sådant förfarande tros gapet mellan elevens vardagsspråk och ämnesspråk kunna minska.

Trots att lärare uppmuntras att aktivt arbeta med ämnesspråk i undervisningen konstaterar Hipkiss (2014), i sin undersökning av hur lärare och elever pratar kring ämnena kemi, biologi och hem- och konsumentkunskap på högstadiet, att ämnesspråken inte inkluderas i diskussionerna i klassrummet, utan att den muntliga kommunikationen mellan lärare och elever sker på vardagsspråk. Eleverna kommer endast i kontakt med ämnesspråken via läroböcker,

men lärobokstexterna inkorporeras inte i undervisningen. Hipkiss tror därför att eleverna möjligen lär sig att förstå ämnesspråken, men inte att själva använda dem. Detta är inte bara problematiskt då elever behöver behärska ämnesspråken för att nå framgångsrika resultat i skolan. Skolans ämnesspråk är också ett sätt att förbereda sig på mötet med det fackspråk som kännetecknar kurslitteratur och undervisning på universitet och högskolor. Här uppstår alltså något som kan liknas vid en kunskapslucka.

2.1.2 Ämnesspråkliga komponenter

Under de senare åren har, enligt Nyström Höög (2010), skola och forskning kommit till två viktiga insikter om skrivpedagogisk utveckling. Dels har det s.k. *diskursiva skrivandet* kommit allt högre upp på agendan, eftersom man är överens om att berättande skrivande inte räcker som förberedelse för livet efter skolan. Dels har man insett att skolundervisningen inte tillräckligt explicit förklarar hur olika texttyper är uppbyggda. Den svenska diskussionen präglas i allt högre grad av tankar från den australiska genrepedagogiken, där just explicitgörandet av texters struktur är central.

Genrepedagogiken kan beskrivas som en pedagogisk tillämpning av den systemisk-funktionella lingvistikens (SFL), och har utvecklats av australiska forskare med särskilt intresse för hur elever kan tillägna sig det naturvetenskapliga sättet att skriva. Situationen i den australiska skolan liknar på många sätt den i dagens svenska skola, med elever med många olika etniska och språkliga bakgrunder. Undervisningssättet är också designat för att ge studieovana elever och elever med språkliga problem möjligheter att lyckas i skolan.

Synen på läs- och skrivkunnsighet i den australiska skolan är präglad av en socialsemiotisk språklig världsbild. Man menar att språket används för att konstruera betydelse i olika sociala och kulturella sammanhang och intresserar sig därför för språklig variation inom och mellan olika situationskontexter (Government of South Australia 2015). Strömningen härstammar från tankar i arbeten av bl.a. Vygotskij (1962) och Halliday och Hasan (1985), vilka beskriver ett ömsesidigt beroendeförhållande mellan social kontext, betydelse och språk. Genom att se språket som socialt förankrat kan man bättre förklara skillnader mellan varieteter som liknar tal- eller skriftspråk, samt den ökande språkliga komplexitet som elever möter under skolgången. Dessa förutsättningar är viktiga när man talar om ämnesspråkligt lärande och lärande av ämnesspråk.

Framför allt är det genrepedagogikens redskap för att koppla ihop språk- och kunskapsutveckling som man har tagit till sig inom den skolrelaterade forskningen i Sverige (Holmberg, Karlsson och Nord 2011). I den australiska läroplanen görs en uppdelning av olika förmågor (eng. *capabilities*) som elever

anses behöva tillägna sig under skolgången. I ett första steg talas om sju generella förmågor. Bland dessa utgör *literacy* ett separat kunskapsfält, om vilket följande står att läsa:

[A]s students engage with subject-based content, they must learn to access and use language and visual elements in the particular and specific ways that are the distinctive and valued modes of communication in each learning area. They need to learn how diverse texts build knowledge in different curriculum areas, and how language and visual information work together in distinctive ways to present this knowledge. [...] As subject-based learning proceeds, particularly in the middle and later school years, the texts that students need to understand and produce take on increasingly formal and academic features employing technical, abstract and specialised ‘written-like’ language forms, in order to communicate complexities of meaning. (Government of South Australia 2015: 3f.)

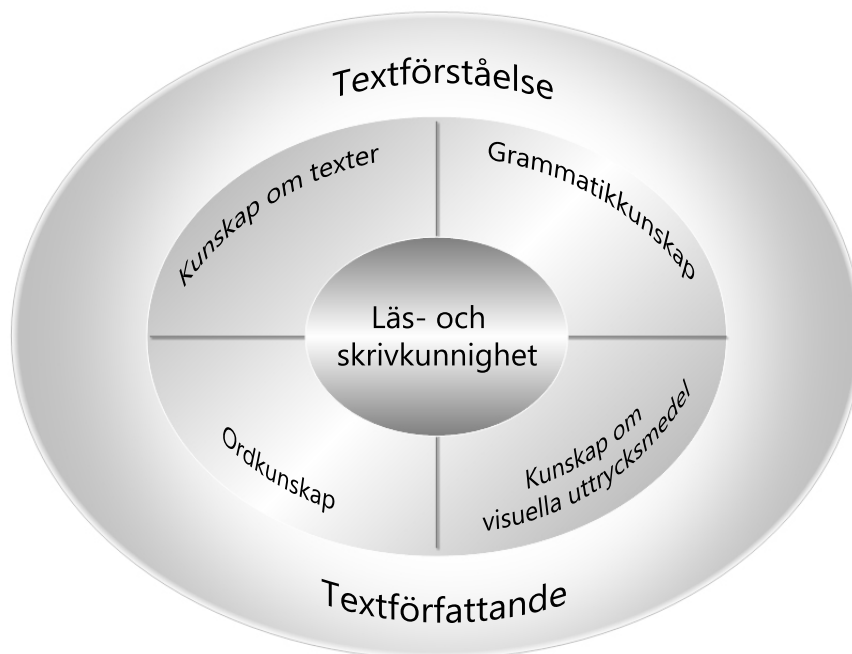
Ett par andra generella förmågor är räkneförmåga och etisk förståelse. De generella förmågorna ses som ett lager som läggs ut på ämneskunskap, som lärs ut inom de olika ämnena. De generella förmågorna är alltså aktuella inom samtliga skolämnen, om än i olika omfattning. Exempelvis är matematikkunskaper användbara i alla skolans ämnen, även om utveckling av räkneförmåga är tätt knutet till matematikundervisningen. På motsvarande vis har etikfrågor en tydlig koppling till samhällsvetenskapliga ämnen.

Literacy, som jag fortsättningsvis, något förenklat, översätter med *läs- och skrivkunnighet*,² delas in i olika komponenter enligt figur 2. De mest övergripande fälten illustrerar hur denna generella förmåga inkluderar att kunna förstå, såväl som författa, texter som förekommer inom olika ämnen.³ Samtidigt omfattar båda dessa övergripande mål fyra kunskapsområden: ordkunskap, kunskap om visuella uttrycksmedel, grammatikkunskap och kunskap om texter.

Kortfattat innefattar ordkunskap att förstå alltmer specialiserade ämnes-specifika ord och kunna avkoda och stava sådana ord. I kunskapsområdet om visuella uttrycksmedel ingår att kunna tolka stilla och rörliga bilder, grafer, tabeller, kartor och andra grafiska representationer, samt att förstå hur bilder och språk samverkar i olika ämnen för att uttrycka betydelse. Grammatikkunskap

²Läsforskare, språkvetare och kulturhistoriker är inte överens om hur den engelska termen ”literacy” bäst översätts till svenska. Se vidare Josephson 2006 om denna debatt.

³I sammanhanget används ”text” i utvidgad bemärkelse för att beteckna alla former av kommunikationsmedel. Texter kan således vara såväl skrivna som talade, och dessutom visuella, digitala och multimodala.



Figur 2: Illustration av komponenter som ingår i läs- och skrivkunnighet (egen översättning efter Government of South Australia 2015: 7)

handlar om att förstå hur olika grammatiska drag påverkar texters betydelse, t.ex. hur olika meningsstrukturer används för att utveckla tankar och hur olika slags ord och fraser kan bära olika slags information. Till denna kunskapsfär räknas också sätt att uttrycka åsikter och värderingar i texter. Slutligen ingår i textkunskap att förstå hur texter struktureras för att uppnå olika syften, som t.ex. att presentera information, förklara processer eller argumentera för något. Textkunskap omfattar också att förstå texters struktur och hur grammatiska, kohesiva, drag fungerar sammanhållande.

Samtliga kunskapsområden i den australiska läroplanen följs upp med specifika inlärningsmål för varje årskurs. Exempelvis anges inom området för grammatik att eleverna efter ett år ska känna till att texter består av ord och fraser som har olika betydelser. De ska också kunna matcha talade och skrivna ord när de läser. Efter två år ska de kunna känna igen och använda substantiv som är knutna till ämnet och också kunna precisera substantiv med hjälp av artiklar och adjektiv. Efter fyra år ska de känna igen och kunna använda adverb och prepositionsfraser för att specificera yttranden. Listan fortsätter i samma

anda ända upp till slutet av år tio, då eleverna ska kunna konstruera texter som innehåller språkligt packande och generaliserande drag, som nominaliseringar och tekniska och abstrakta ord.

Tänkarna om språklig utveckling inom olika ämnesområden i den svenska skolan är, som sagt, starkt påverkade av genrepedagogiska teorier om hur man stöttar elevernas språkutveckling, vilket exempelvis kan ses på Skolverkets hemsida under rubriken ”Vad innebär en språkutvecklande undervisning?” (se Gibbons 2014). Än så länge saknar emellertid den svenska läroplanen en tydlig beskrivning av den förväntade ämnesspråkliga progressionen inom olika ämnen och på olika stadier.

2.1.3 A- och B-texter

När det gäller den språkutveckling som sker inom ramen för skolundervisningen genomgår elevers ordförråd den mest dramatiska utvecklingen (Lindberg 2007). Främst utvecklas ett specialordförråd för skilda fackområden knutna till skolämnen. Vad gäller grammatiken utvecklas förmågan att kombinera redan inlärd syntaktiska strukturer på ett mer varierat sätt för att uppnå olika syften. Elever lär sig även stilistiskt markerade drag, som passivkonstruktioner och nominaliseringar, som är förbundna med skriftspråket (Viberg 1993).

Progressionen handlar emellertid inte bara om att bygga ut ordförråd och grammatik, utan också om att tillägna sig olika repertoarer av språkliga verksamheter – olika *genrer* (Anward 1984; Liberg 2009). Under skoltiden lär man sig t.ex. att behärska resonerande och argumenterade texttyper som hålls samman av logiska relationer (orsak/verkan, argument/motargument) utöver berättelsens tidsbundna då/sen.

Som tidigare har nämnts innebär kommunikation som utspelar sig här och nu ett första steg i den individuella språkutvecklingen. I ett sådant språkligt samspel utgör alltid det ämne som är uppe för behandling någon aspekt av den *omedelbara* situationen. Likaså bygger förståeligheten av yttrandena på den gemensamma upplevelsen av situationen; utanför denna är yttrandena obegripliga och meningslösa. Anward (1984) kallar yttranden som är situerade här och nu för *ingripanden* eller *kommentarer*.

Efter hand lär sig barn att tala om situationer i vilka den omedelbara verksamheten inte ingår. Sådana situationer kallas av Anward (1984) för *medelbara* situationer. I redogörelser för medelbara situationer måste de aspekter av ämnets bakgrund som andra deltagare inte har upplevt eller känner till förklaras med språkliga medel. Med andra ord måste yttrandet sättas in i en *text* som gör att mottagaren kan föreställa sig den situation man talar om. Enligt An-

ward (1984) innebär övergången från att tala om omedelbara situationer till medelbara situationer, i formella termer, att fras- och satsproduktion ersätts av textproduktion.

Elevers genreutveckling som en avgörande uppgift för skolundervisningen påtalas både i forskning och i skolans styrdokument. I kursplanen för svenska i grundskolan står t.ex. att undervisningen ska bidra till att eleverna utvecklar kunskaper om olika sorters skönlitteratur och sakprosa (Skolverket 2011b), och i beskrivningen av gymnasieskolans arbete anges ”genreskrivande” som ett område inom vilket elever på de högskoleförberedande programmen behöver utveckla fördjupade kunskaper (Skolverket 2011a).

Hur utvecklingen av en genrerepertoar går till har beskrivits i olika termer av olika forskare. Samtliga gör dock en grundläggande uppdelning mellan personliga berättelser och mer generella texttyper, som ofta förknippas med sakprosa⁴ och facktexter. De senare texttyperna har dock gått under många olika samlingsnamn, som t.ex. *framställningar* (Anward 1984), *B-texter* (Teleman 1988), *deskriptiva* (Teleman 1989) eller *diskursiva* (Hultman 1989; Nyström 2000).

Eftersom merparten av dessa beteckningar antyder något om texternas funktion, samt att denna i hög grad varierar, väljer jag att i fortsättningen följa den terminologi som Teleman (1988) föreslår. Han särskiljer två basala texttyper som han kallar *A-* och *B-texter*. *A-texter* är talnära och förekommer även i muntlig kommunikation. De utgörs framför allt av berättelser, som kännetecknas av att de återger ett tidsförlopp, har en eller flera klara huvudpersoner och redogör för sinnesdata (sådant som hörs eller syns). *A-texter* är den enklaste texttypen. Den mer avancerade texttypen kallas *B-texter*. Sådana texter är mer generella; de handlar om hur världen ser ut eller bör se ut. I denna texttyp spelar sinnesdata mindre roll. I stället syftar texten till att analysera, resonera och förutsäga aktiviteter som med fördel uttrycks med hjälp av skrivet språk.

Anward (1984) redogör för ytterligare två avgörande punkter på vilka *A-* och *B-texter* skiljer sig. För det första går *A-texter* ut på att beskriva och värdera en medelbar situation på ett sätt som gör att mottagaren kan identifiera sig med huvudpersonerna. För att lyckas med detta måste även talaren/skribenten kunna identifiera sig med karaktärerna och skapa en medryckande och inlevelsefull text. Poängen med *B-texter* är i stället att på ett objektivt sätt beskriva en verksamhet och dess deltagare, så som den uppfattas utifrån. *B-texter* skrivs alltså snarare från ett utomstående perspektiv än från ett deltagarperspektiv.

⁴Termerna *sakprosa*, *normalprosa* och *bruksprosa* används med i stort sett samma betydelse (jfr Thelander 1970; Westman 1974; Englund, Ledin och Svensson 2003).

Den andra skillnaden mellan de två texttyperna är att A-texter bygger på att mottagarna använder sin bakgrundskunskap om sådana situationer som berättelsen handlar om. En sådan aktivering av relevanta erfarenheter av liknande självupplevda situationer krävs för att kunna ta till sig den kunskap som berättelsen vill förmedla. I berättandet är tolkningen halva arbetet, och A-texter är ofta mångtydiga. I B-texter däremot ska läsaren spela en så liten roll som möjligt i tolkningen, eftersom den kunskap som förmedlas ofta rör generaliseringar, samband och slutsatser, vilka i idealfallet enbart ska bedömas utifrån de förutsättningar som beskrivs i texten (Anward 1984; Hultman 1988; Lundberg 2002). B-texter kräver alltså att läsaren håller tillbaka sina förkunskaper om ämnet, ett förhållningssätt man måste lära sig.

Trots att läroplanen från 2011 inte är den första som fastställer att man i undervisningen ska arbeta med olika texttyper, har skolspråksforskning under en lång tid pekat på att främst skönlitterärt läsande och skrivande behandlas inom språkundervisningen i svenska (Hultman 1988; af Geijerstam och Wiksten Folkeryd 2006). Redan i slutet av 1980-talet ser Hultman (1988) ett problem i att många svensklärare i gymnasieskolan ”har sitt hjärta i skönlitteraturen” och håller det berättande skrivandet som målet för eleverna. Han menar att många elever i årskurs 9 är ”durkdrivna” berättare i teknisk mening; de bemästrar alltså texttypen A. Men när det kommer till B-texter går det inte alls lika bra. Detta visar sig när elever, som får i uppgift att skriva en resonerande insändare, väljer att utforma den som en novell, eftersom den berättande formen ligger dem närmare till hands.

Nästan två decennier senare konstaterar Nyström Höög (2006) att berättelser är den enda texttyp som elever på gymnasiet inte har några problem att själva producera. Min personliga erfarenhet av att vara lärare i skrivande inom olika genrer på universitet är att också de flesta vuxna studenter är betydligt säkrare på hur man utformar en fiktiv berättelse än på hur man t.ex. lägger upp en argumenterande text eller kåserar fritt.

Forskare är överens om att läroböcker innehåller mycket B-texter, med en annan språklig repertoar än traditionella berättelser. Detta gör utformningen av läroböcker särskilt betydelsefull för den språkutveckling som sker under skoltiden (se t.ex. Selander 1988; Strömquist 1995; Melander 2003). När elever möter sådana texter kan dock den språkliga framställningen verka främmande (jfr också Teleman 1988; Nyström 2000; Nyström Höög 2006). Därför behöver lärare kunna förklara likheter och skillnader i hur berättande och fackliga texter byggs upp språkligt.

2.2 Akademiskt språk

2.2.1 Engelskspråkig forskning

I engelskspråkig litteratur inom områdena EAP (English for academic purposes) och SLA (Second language acquisition) möter man ofta termerna *academic language* och *academic vocabulary*. Termerna används för att referera till det språk som påträffas i lärandemiljöer för att argumentera, försvara ståndpunkter och framställa och sammanställa information.⁵

Värt att understryka är att *academic language* på engelska inte är direkt översättbart till "akademiskt språk" på svenska, vilket typiskt syftar på språk som hör till universitet och högskolor. Per definition inkluderar den engelska termen nämligen också det som på svenska betecknas som "skolspråk" eller "ämnesspråk". Däremot behandlar de engelska studierna inom området nästan uteslutande material från universitetet, d.v.s. ytterligheten av *academic language*, eller det mest extrema skolspråket, om man så vill.⁶

Klart är att akademiskt språk positionerar sig i ena änden av ett kontinuum som definieras av grader av formalitet, innehållsmässig komplexitet och opersonlig inställning/attityd. I andra änden befinner sig den informella och flyktiga konversationen. Detta gör att det finns vissa gemensamma kännetecken som särskiljer det högt akademiska från det mindre akademiska eller det vanliga samtalsspråket. Snow (2010: 450) beskriver det akademiska språkets särprägel på följande vis:

A major challenge to students learning science is the academic language in which science is written. Academic language is designed to be concise, precise, and authoritative. To achieve these goals, it uses sophisticated words and complex grammatical constructions that can disrupt reading comprehension and block learning.

Komplexiteten i det akademiska språket förklaras närmare genom att det är koncist. Författarna försäkrar sig om uttrycks precision genom och luta sig mot grammatiska processer för att packa samman komplexa idéer med få, informationstunga ord. Verb används för att uttrycka relationer snarare än handlingar,

⁵Dessutom förekommer en rad termer med snarlik användning: *academic English*, *discourse-organising vocabulary*, *language of education*, *language of schooling*, *non-technical terms*, *scientific language*, *semi-technical vocabulary*, *specialised non-technical terms* och *sub-technical vocabulary* (se Paquot 2010: 17–21 för en termutredning).

⁶Coxhead, Liesje och Tinkle (2010) förklarar att forskare har trott att texterna från grundskolan innehåller vanligare ord och därmed är mer lättlästa än universitetsböcker. De säger också att mellan- och högstadietexter eventuellt är mycket svårare än man tror, och därför behöver beforskas mycket mer med korpusstudier.

och långa meningar innehåller oftare inbäddade satser snarare än uppräkningsar. Man kan även observera en auktoritativ röst i akademiska framställningar, vilken härstammar från författarens medlemskap i en gemenskap som intar en delad epistemologisk ställning. Denna inställning markeras genom att undvika personliga pronomen, en förkärlek för generella utsagor snarare än specifika och epistemologiskt grundade evalueringar, som ”rigorous study” eller ”questionable analysis” (i stället för t.ex. ”great study” eller ”funky analysis”). Upprätthållandet av den opersonliga och auktoritativa framställningen skapar en distanserad ton som ofta känns främmande för ungdomar, och är mycket svår att efterlikna (Snow 2010).

För engelskans del är det akademiska ordförrådet tämligen väl utforskat, och med början på tidigt 1970-tal, har en rad olika engelska s.k. *akademiska ordlistor* framställts för att möta behoven hos studenter på avancerad nivå (se t.ex. Champion och Elley 1971; Praninskas 1972; Ghadessy 1979; Xue och Nation 1984 och vidare i avsnitt 2.2.3 om engelska akademiska ordlistor). Det finns också initiativ tagna till att identifiera olika akademiska flerordsuttryck på engelska, t.ex. kollokationer (Gledhill 2000; Marco 2000) och lexikaliserade fraser (Oakey 2002; Simpson-Vlach och Ellis 2010). Däremot är syntaktiska strukturer mindre väl beskrivna eller kända. Korpuslingvisten Biber har visserligen arbetat med att undersöka akademiskt språk på flera olika nivåer (se vidare avsnitt 2.3.2.1 om hans metod), men fältet måste ändå sägas vara långt ifrån uttömt och dokumenterat (se vidare Flowerdew 2002 för en översikt över korpusbaserade undersökningar av akademisk engelska).

Betydelsen av att tillägna sig ett adekvat språk för akademisk verksamhet har sedan en tid fått ökad uppmärksamhet. Till en början gällde intresset i huvudsak akademisk engelska, men nu växer insikten om behoven även för andra språk. Akademiska ordlistor är under utveckling, t.ex. för portugisiska (Baptista m.fl. 2010) och franska (Cobb och Horst 2004). Dessutom arbetar nordiska forskare med att utveckla akademiska ordlistor för de nordiska språken (Johansson Kokkinakis m.fl. 2012). Den svenska akademiska ordlistan presenteras närmare i avsnitt 2.2.4 i detta kapitel.

2.2.2 Svenskspråkig forskning

Ask (2006, 2007) konstaterar att nyblivna studenter ofta upplever kraven på akademiskt skrivande som problematiska, d.v.s. att det finns ett glapp mellan den skriftspråkskompetens som eleverna har efter sina 12 år i skolan och den de förväntas ha när de börjar på universitetet. För många studenter blir problemet akut. De känner sig inte delaktiga i den akademiska gemenskapen, och följer man inte reglerna i denna blir man inte godkänd.

På universitet och högskolor förekommer texter från flera olika genrer, t.ex. avhandlingar, fackartiklar, läroboksartiklar och personliga eller vetenskapliga essäer/ uppsatser. Den vetenskapliga artikeln är i dag huvudgenren för universitetsanställda forskare. Det är en krävande genre med strikta begränsningar. Enligt Dysthe m.fl. (2006) ställs dock samma krav på t.ex. studenters masteruppsatser. Även Ask (2007) konstaterar att de formella kraven på studenternas texter, som främst författas i kunskapskontrollerande syfte, på grundnivå, och även högre upp, ofta är desamma som för kunskapsbyggande skrivande inom den akademiska debatten.

Studenterna i Asks undersökningar påtalar också universitetets oförmåga att synliggöra de krav på skriftlig framställning som faktiskt ställs på universitetsnivå. På många sätt liknar denna kritik de röster som under de senare decennierna har höjts mot den processorienterade skrivpedagogiken som länge har haft ett starkt fäste i skolan (jfr Nyström 2000). Ask (2006) menar att skolan ofta lär eleverna handla inom givna ramar, men utan att på ett intellektuellt plan förstå de redskap som måste användas.

Appropriering av de nya kommunikativa redskap som det akademiska skrivandet förutsätter, kräver att diskurskraven görs tydliga för studenten. Det är alltså föga troligt att studenten på egen hand ska kunna erövra ett akademiskt språk. Ett sådant erövrande kräver [...] en lärandemiljö som öppnar för en medveten utveckling av det akademiska skrivandet. (Ask 2006: 4)

Också Blåsjö (2004) rapporterar att lärarna på universitetet sällan eller aldrig explicitgör vad tänkesätten inom olika ämnen går ut på och hur de manifesterar sig språkligt. Lärarna är i viss mån medvetna om att de tänker på ett speciellt sätt, men har svårt att uttrycka det. ”Just sättet att tänka är troligen det mest omedvetna för en etablerad deltagare i en praktikgemenskap” (Blåsjö 2004: 291).

Både Blåsjö (2004) och Rongen Breivega (2003) undersöker akademiskt skrivande i relation till de normer för texter och språk som gäller i olika ämnesmiljöer. Resultaten av dessa studier ligger i linje med den hävdvunna uppfattningen att det går en tydlig skiljelinje mellan naturvetenskapliga och humanistiska ämnen. Detta förhållande menar Tønnesson (2008) upprätthålls genom att de båda inriktningarna utgör textkulturer med helt skilda normer för hur texter ska skrivas.

Inom den naturvetenskapliga gemenskapen vill man hitta de mest entydiga och korrekta lösningarna på olika problem med hjälp av olika modeller. Alternativa synsätt, genom citat eller referat, presenteras sällan eller aldrig. Över huvud taget saknas oftast ett kritiskt förhållningssätt. Inom den humanistiska

gemenskapen resonerar man genom att ställa frågor ur olika perspektiv utan att nödvändigtvis besvara dem. Texterna innehåller många röster i form av citat och referat, och dessa olika röster diskuteras källkritiskt (Blåsjö 2004).

Rongen Breivega (2003) visar att naturvetenskapliga artiklar disponeras enligt ett standardiserat textformat, medan humanistiska artiklar uppvisar större variation. Personlig stil betraktas som en styrka för en humanist, men uppmuntras inte hos naturvetenskapliga artikelförfattare. Medan naturvetenskapliga artiklar präglas av en objektifierande stil där författaren försöker ställa sig utanför texten, är författarna till humanistiska artiklar genomgående mer synliga i sina egna forskningstexter. Dessutom visar det sig att författarna inom vissa ämnesområden, till exempel historia, inte är lika bundna till en specifik fackterminologi, utan ”förmedlar stoffet genom ett ordval som även lekmän på en viss utbildningsnivå förstår och använder” (Rongen Breivega 2003: 101).

2.2.3 Engelskspråkiga akademiska ordlistor

Läsarens ordförråd är den viktigaste faktorn för både L1- och L2-talare när det gäller textförståelse (Read 2000; Cobb och Horst 2001), eftersom man behöver känna till 95 % av orden i en text för att förstå den (Laufer 1992, 1996). Vid denna gräns anses läsaren kunna börja lära av kontexten och därigenom sluta sig till okända ords betydelse (Nation 1997). Enligt Lindberg och Johansson Kokkinakis (2007) hänger läsarens tolerans för okända ord också samman med andra faktorer, som t.ex. texttyp och förtrogenhet med innehållet i texten, men en bred vokabulär är likväl en viktig tillgång i undervisningssituationer. För den som studerar på sitt andraspråk bedöms ordkunskap till och med vara den enskilt viktigaste faktorn för skolframgång (Saville-Troike 1984; Laufer 1996).

För att uppnå en tillräcklig ordförståelse för högre utbildningar menar Nation (1997, 2001) att man måste behärska tre ”ordlistor”: för det första de 2 000 vanligaste orden eller *basorden*, för det andra *akademiska ord* och för det tredje *tekniska termer*. Basorden, som på engelska ofta kallas *core/basic/nuclear vocabular*, är ord som har hög frekvens i de flesta språkbruk. De består av de mest använda funktionsorden, som *about*, *by* och *some* och innehållsord som *bag*, *put* och *suggest* (Stubbs 1986). Den mest kända listan över basordförråd är *the General Service List* (GSL, West 1953), som innehåller 2 000 s.k.

ordfamiljer⁷, som har valts ut från en korpus med skriven engelska på 5 miljoner ord.⁸ GSL har visats täcka upp till 92 % av skönlitterära texter (Hirsh och Nation 1992) och 76 % av akademiska texter (Coxhead 2000).

Tekniska termer, eller domänspecifika termer, är ord som kräver vetenskaplig kunskap. De karakteriseras av att vara betydelsemässigt specialiserade, motståndskraftiga mot semantisk förändring och sakna exakta synonymer. Tekniska termer uppträder med hög frekvens inom speciella ämnestexter, men förekommer i stort sett inte alls utanför sitt ämnesområde (Nation och Hwang 1995). Mängden teknisk vokabulär antas vara omkring 1 000 ord per ämne och stå för ungefär 5 % av ordkunskapen i en akademisk ämnestext (Coxhead och Nation 2001).

Även om man kan separera vokabulären i basord, tekniska termer och akademiska ord, finns det naturligtvis inga vattentäta skott mellan dessa indelningar (jfr Nation och Hwang 1995). När det gäller att definiera akademiskt språk existerar således ingen exakt gräns, utan de akademiska kännetecknen varierar som en funktion av disciplin, ämne och modalitet (skrift/tal). Så skriver t.ex. Hyland och Tse (2007) att alla discipliner formar orden för sitt eget bruk, vilket visar sig både genom skillnader i kollokationsmönster och varierande betydelser. De menar att om man tar hänsyn till kontext, kotex och användning, existerar inget som kan kallas akademiskt ordförråd (jfr även Chung och Nation 2003). Andra anser emellertid att även om ord kan användas på olika sätt, kan man genom att studera autentiskt språkbruk både identifiera kärnbetydelser hos akademiska ord och se hur de skiljer sig åt lexikogrammatiskt, betydelsemässigt och fraseologiskt mellan olika ämnen (jfr t.ex. Wang och Nation 2004; Eldridge 2008; Granger och Paquot 2009; Paquot 2010).

Forskare som arbetar med att försöka sammanställa akademiska ordlistor har definierat och operationaliserat *akademiska ord* på olika vis. Gemensamt kan dock sägas att termen brukar beteckna en typ av ord som, till skillnad från tekniska termer, utgör en betydande del av löpande akademisk text, oavsett ämnesområde.

Den akademiska ordlista som har haft störst genomslag är *the Academic Word List* (AWL, Coxhead 2000, 2002), som är baserad på en korpus på 3,5 miljoner ord från akademiska artiklar och kurslitteratur. Korpusen är indelad i fyra discipliner, vilka, var och en, innehåller sju ämnesområden. AWL omfattar sammanlagt 570 ordfamiljer. Dessa är fördelade över 10 frekvensbaserade dellistor, vilka är avsedda att vara till hjälp för att kunna sätta upp kortsiktiga

⁷ *Ordfamiljer* används ofta för att gruppera ord för engelskstuderande och avser, enligt Bauer och Nation (1993), en ordstam och alla dess vanliga böjningar och avledningar, t.ex. *react*, *reacting*, *reaction*, *reactionary*, *reactive*, *unreactive*, *reactivate*, *reactor*.

⁸ Urvalet av ordfamiljer till GSL baseras på korpusförekomster av ord. Däremot måste inte alla medlemmar i ordfamiljerna ha påträffats i korpusen.

inlärningsmål. Förutom de ovan nämnda gemensamma kriterierna för akademiska ord, gäller för ordfamiljerna i AWL att deras medlemmar inte får tillhöra de 2 000 vanligaste orden i språket. Detta uppnås genom en jämförelse med ordfamiljerna i GSL (West 1953).

Under det senaste decenniet har AWL använts flitigt i språkundervisning, t.ex. i ordkunskapstest (se t.ex. Schmitt, Schmitt och Clapham 2001) och i läroböcker (se t.ex. Schmitt och Schmitt 2005; Huntley 2006), samt som forskningsunderlag (Coxhead 2011). Listan har dock inte undgått kritik.

En invändning är att användningen av ordfamiljer snedvrider ordurvalet. Ord som inte av egen kraft uppfyller kriterierna ”räddas” genom avledning- ar med annan betydelse. Vidare tas ingen hänsyn till homografi och polysemi; gemensamt etymologiskt ursprung behöver inte innebära att ordstammen har samma betydelse i besläktade ord (Wang och Nation 2004). Ytterligare en kritik mot ordfamiljer, som framförs av Paquot (2010), är avsaknaden av ordklassangivelser i AWL. Man får t.ex. reda på att *issues* är den mest frekvent använda medlemmen i ordfamiljen *issue*, men inte om de olika orden oftast används som verb eller substantiv.

Vidare anser Hyland och Tse (2007) att indelningen och urvalet av korpus- texter är godtycklig och att uppsättningen av ord därför inte är representativ för akademiska texter oavsett ämnesinriktning. Föresatsen att exkludera vardagliga ord har också ifrågasatts, eftersom vissa ord ur basordförrådet anses ha speciella akademiska funktioner (jfr Paquot 2010; Gardner och Davies 2014). Dessutom har man visat att AWL innehåller många mycket vanliga ord, som *adult*, *drama*, *sex* och *tape* (Paquot 2007, 2010).

Olika forskare är alltså inte överens om hur man bäst beskriver och identifierar akademiskt ordförråd. Detta har, på senare år, resulterat i åtminstone två seriösa utmanare till AWL: *the Academic Keyword List* (AKL, Paquot 2010) och *the Academic Vocabulary List* (AVL, Gardner och Davies 2014). Båda dessa ordlistor räknar med lemman i stället för med ordfamiljer. De kräver också att orden är utmärkande för just akademiska texter, s.k. *nyckelord* (se vidare avsnitt 5.3.2 om nyckelord). Detta krav uppfylls genom jämförelser med referenskorpusar. Slutligen måste nyckelorden vara relativt jämnt fördelade över texterna i korpusen.

AKL innehåller 930 akademiska nyckelord, som extraherats ur en korpus på 3 miljoner ord, fördelade över akademisk prosa och studentuppsatser. Texterna kommer från fem olika discipliner. I materialet ingår även en skönlitterär referenskorpus.

AVL består av 3000 ord, framtagna ur ett korpusmaterial på över 120 miljoner ord. Texterna utgörs av tidskriftsartiklar från nio olika discipliner. Som jämförelsematerial används referenskorpusar med nyhetstexter och skönlitte-

ratur. Förutom det stora empiriska underlaget och listans omfattning, är AVL även unik i det att orden inte får förekomma ”oväntat mycket” i någon eller några få discipliner.

2.2.4 En svensk akademisk ordlista

Tillsammans med forskare inom områdena lexikologi/lexikografi och svenska som andraspråk har jag utvecklat en svensk *akademisk ordlista* (AO, Ribeck, Jansson och Sköldberg 2014). Denna är främst avsedd att användas vid produktion av akademisk text på svenska, men också tänkt att kunna tjäna som underlag för utveckling av framtida ordtester och undervisningsmaterial, samt bidra till att dokumentera det svenska akademiska ordförrådet.

Metoden som används för att ta fram kandidater till AO är tydligt inspirerad av tidigare försök att extrahera akademiska ord ur en korpus med akademiska texter. Det första steget består således i att, på bästa sätt, sätta samman en korpus som är representativ för svenskt akademiskt skriftspråk. Nästa steg är att i denna korpus identifiera akademiska ord.⁹

Den svenska akademiska ordlistan baseras på vanligt förekommande ord i *Svensk akademisk korpus* (SveAk). Denna innehåller drygt 26 miljoner token¹⁰ från avhandlingar och publicerade artiklar från humaniora och samhällsvetenskap, och är även den korpus som jag använder mig av i undersökningarna av det akademiska språket i denna avhandling (se vidare avsnitt 3.2 om SveAk:s sammansättning).

I arbetet med ordlistan används lemman istället för ordfamiljer, vilket har tydliga konsekvenser för innehållet i listan. Formellt besläktade ord som *omfatta*, *omfattande* och *omfattning* bildar egna uppslagsord. På så vis synliggörs alla uppslagsord mer och beskrivs utifrån sina egna premisser. Ställningstagandet kan också kopplas till tanken att ordlistan i första hand ska användas vid produktion. Coxhead (2000) menar att bruket av ordfamiljer är befogat, med tanke på att psykolingvistiska studier visat att morfologiska relationer mellan ord troligen finns lagrade i det mentala lexikonet. Paquot (2010) konstaterar att Coxheads resonemang håller för att presentera ordfamiljer för receptiva syften. Däremot är presentationssättet föga meningsfullt vid produktion, då inte alla medlemmar i ordfamiljerna är lika användbara (jfr även Gardner och Davies 2014 som förespråkar användning av lemman i pedagogiska lexikala resurser).

⁹Den som vill ta ytterligare del av våra metodologiska överväganden hänvisas till Sköldberg och Johansson Kokkinakis 2012 och Jansson m.fl. 2012 om insamling av akademiska texter, och till Carlund m.fl. 2012 om ordlistans användning och ordextraktionen.

¹⁰Termen *token* omfattar de huvudsakliga enheter en teckensträng kan delas upp i, d.v.s. ord, skiljetecken och övriga symboler. Se vidare avsnitt 4.1 om lingvistisk textannotering.

Den svenska ordlistan består av 655 huvudord. När dessa ordnas enligt fallande *akademiskt index*¹¹ hamnar följande ord i topp: *dock, studie, beskriva, social, enligt, innebära, samt, form, betydelse* och *fall*.

Både AO och SveAk är fritt tillgängliga genom Språkbankens infrastruktur. Listans användargränssnitt framgår av figur 3.

¹¹Se avsnitt 5.3 för mer information om den s.k. *indexprincipen* som används för att ranka orden.

En svensk akademisk ordlista		Böjning		Betydelse		Språkprov		Engelska	
Ord	Böjning	Betydelse	Språkprov	Engelska	Alfabetiskt ordnad	Akademiskt ordnad	Om listan	Engelska	Engelska
11	begrepp substantiv	begreppet begrepp begreppen	föreställning, uppfattning	I sin bok från 1994 genomför Derrida en analys av <i>begreppet</i> "demokrati" korpus				concept, conception, idea, notion	
12	relation substantiv	relationen relationer relationerna	1. förhållande 2. känslomässigt (ofta sexuellt) förhållande; <även> (formell) förbindelse	I nyare forskning betonar man författarens <i>relation</i> till modernismen korpus				relation, relationship	
13	möjlighet substantiv	möjligheten möjligheter möjligheterna	möjlig utväg, tillfälle	Ett flertal kommittéer har tillsatts för att utreda <i>möjligheterna</i> att minska utsläppen korpus				possibility	
14	bild substantiv	bilden bilder bilderna	foto, teckning, målning etc.; <även bildligt> skildring; <även> liknelse	Mediermas uppgift är att förmedla en rättvisande <i>bild</i> av verkligheten korpus				image, picture	

Figur 3: Användargränssnitt för *Svensk akademisk ordlista*, version 2.0, 2015, <<http://spraakbanken.gu.se/ao/>>

2.3 Om att undersöka språk i text

När det gäller att beskriva språk i texter kan analysmetoderna variera. I stora drag kan dock två huvudlinjer urskiljas: den *kvantitativt stilistiska* ansatsen respektive den *textlingvistiska* ansatsen (jfr Melander 1995 och Ledin och Selander 2003: 101).

Sedan början av 1900-talet har det svenska skriftspråket diversifierats genom att det som tidigare betecknades normal- och lågprosa har lösts upp, och i stället har olika genrer utvecklat specifika drag (Josephson 2004: 22f.). Läroböcker har förekommit i den svenska skolan sedan 1880-talet, men inte förrän i början av 1970-talet började språket i genren beforskas. Vid den tiden lades grunden för det som i Sverige kom att kallas den *kvantitativa stilistiken*, eller *Westmantraditionen*, efter Margareta Westman, som tillsammans med Tor G. Hultman arbetade i projektet *Skrivsyntax*. Projektet, vars huvudsakliga uppgift var att undersöka skrivträning i skolan, startade 1970 i Lund och resulterade i en av världens dittills mest omfattande språkundersökningar (Hultman och Westman 1977).

Under de senare decennierna har intresset hos många forskare mer inriktats mot den textlingvistiska analysmetoden (se t.ex. Nyström 2000; Blåsjö 2004; Karlsson 2006). Denna går ut på att effekterna av språket studeras; de språkliga medlen sätts i relation till texten som helhet, dess funktion och den språksituation den befinner sig i. Selanders omfattande diakrona analys av historieläroböcker (1988) kan sägas finna sig mitt i brytpunkten mellan dessa metoders storhetstider; han räknar grammatiska variabler, tittar på stoffurval och förhåller sig till förändringar av läroplaner m.m. I och med Melanders avhandling om facktexter (1991) kan förskjutningen sägas ha fullbordats. I denna talas det t.ex. om vilka olika ”kunskapsvärldar” som skildras, och om stilistik i termer av t.ex. ”mindre tät och luftig stil”. Naturligtvis är inte alla studier efter 1990 textlingvistiska, men tendensen måste ändå sägas vara tydlig.

På senare år har analyser av texters meningsskapande blivit ett allt vanligare inslag i den språkvetenskapliga textforskningen (jfr Svensson och Karlsson 2012). Det moderna s.k. (*ut*)*vidgade textbegreppet* innebär just att texter betraktas som multimodala, och att viktiga betydelser anses uttryckas både genom språkliga och visuella kommunikationsformer (Nord 2012). Poängen med ett vidgat textbegrepp är att alla delar som bär betydelse ska tas hänsyn till i analys och tolkning av en text. I multimodal textanalys kan dock olika stor vikt läggas vid det språkliga perspektivet. I kritisk diskursanalys likställs språkets betydelse med andra modaliteters. I mer språkvetenskapligt inriktade studier beskrivs oftare vad språket gör i förhållande till det visuella, eller hur språk och bild samverkar (Svensson och Karlsson 2012).

Men alla textstudier syftar inte till att beskriva texters meningspotential, utan kunskapsintresset kan likaväl ligga i att karakterisera språket i en text eller i en grupp av texter. Texter kan alltså analyseras som instanser av språklig struktur (jfr Svensson och Karlsson 2012). Detta slags textanalys ligger nära den som förekommer i den traditionella nordistiken, i vilken man koncentrerar sig på att beskriva texters språkliga drag och motivera dem kontextuellt (jfr Karlsson och Strand 2012).

Sammanfattningsvis kan alltså utvecklingen inom textanalys sägas ha gått från beskrivande, kvantitativt stilistiska studier mot mer funktionellt inriktade textstudier. Beskrivande och kvantitativa undersökningar av språklig variation i text förekommer i dag främst i språkteknologiska tillämpningar (se vidare avsnitt 2.3.2).

2.3.1 Kvantitativ stilistik som analysmetod

Westmantraditionen utvecklades som en rent empiristisk analysmetod som teoretiskt vilar på en idé om att stilistiska variabler systematiskt varierar mellan olika texttyper. Genom att kvantifiera grammatiskt definierade enheter, som t.ex. mängden substantiv, korta meningar eller långa fundament och räkna dessa, tror man sig kunna beskriva denna systematiska variation. När språkdrag har identifierats relateras de till kontexten för att förklaras. Enligt Karlsson och Strand (2012) räcker det dock att koncentrera sig på vissa valda delar av kontexten, som ålder och social bakgrund hos författare och läsare, syftet med texten eller när den skrevs. I och med det att dragen relateras till kontexten beskrivs korrelationer, men inga förklaringar eller orsaker ges.

Teoribildningen är inte mycket mer komplicerad än så, varför den kvantitativa stilistiken ibland har kritiserats som en teoretiskt ogrundad metod utan helhetsperspektiv (se t.ex. Svensson och Karlsson 2012). Till modellens försvar anger Ledin och Selander (2003: 102) att även om metoden i sig inte säger något om samhället i stort, säger språkliga frekvenser en hel del om texters utformning i sociala kontexter. När språkliga kännetecken för olika sociala situationer väl har lagts fram, kan man tolka dessa, d.v.s. koppla dem till de samhälleliga diskurserna. Dessutom kan kvantitativa resultat tjäna som referenspunkter för stilistiska studier överlag (jfr Stubbs 2005).

Westmans avhandling, *Bruksprosa* (1974), vars uppgift var att kartlägga ”den slitstarka bruksprosan” och som skrevs inom ramen för *Skrivsyntax*, utgör den första kvantitativt stilistiska studien av svenska lärobokstexter. Det typiska med *Bruksprosa* är dels studieobjektet, samtida bruksprosa. Denna typ av ”oglamoröst vardagsspråk” (Melander 2003: 135) hade tidigare inte ägnats mycket uppmärksamhet. Ett skäl till det kvantitativa angreppssättets ökande

popularitet var att en metod med tydligt definierade enheter ansågs uppfriskande objektiv och vetenskaplig i jämförelse med tidigare använda metoder. Dessutom möjliggjorde metoden att olika material kunde jämföras med varandra. Härigenom kunde man se hur språket skilde sig mellan olika brukssituationer och hur olika faktorer påverkade språkanvändningen.

2.3.2 Datamaskinell textanalys

2.3.2.1 Registeranalys

Kvantitativt stilistiska undersökningar, som i Westmans efterföljd blev vanliga i Sverige, har relativt få internationella motsvarigheter. Ett närliggande sätt att gripa sig an texter är dock det som under 1980-talet utvecklades av korpuslingvisten Douglas Biber (se t.ex. Biber 1988, 1995, 2006; Biber och Conrad 2009). Bibers datorbaserade undersökningsmetod går under beteckningen *registeranalys* och vilar i mångt och mycket på samma teoretiska förståelse som den kvantitativa stilistiken.

Man kan notera att korpuslingvister som studerar språklig variation kopplad till vissa situationer eller kommunikativa syften ofta, till synes slumpmässigt, väljer att använda antingen *register* eller *genre* som överordnad term för sitt undersökningsobjekt. Biber (2006: 11) urskiljer dock en viss uppdelning, i det att registerstudier, å ena sidan, brukar behandla lexikogrammatiska särdrag och beskriva hur en viss användning av ord, ordtyper och grammatiska egenskaper systematiskt varierar med avseende på situationstyp. Genrestudier, å andra sidan, har ofta ett fokus på att beskriva hur olika sociala aktiviteter utförs med hjälp av språkliga verktyg, vilket medför att de ofta även behandlar frågor som rör makt och ideologier.

I en registeranalys kopplas texters lexikogrammatiska språkdrag till de förhållanden inom vilka texterna producerats. Språkdragen antas vara funktionellt motiverade utifrån dessa förutsättningar. Biber använder alltså *register* på liknande sätt som *genre* används inom SFL, nämligen för att beteckna vad som särskiljer olika texttyper. I mer exakta ordalag beskriver Biber (2006: 11–12; kursiv i original) att termen *register* används:

[to] refer to situationally-defined varieties described for their characteristic lexico-grammatical features. [...] *Register* is used as a cover term for any language variety defined in terms of particular constellation of situational characteristics. That is, register distinctions are defined in non-linguistic terms, including the speaker's purpose in communication, the topics, the relationship between speaker and hearer, and the produc-

tion circumstances. However, [...] there are usually important linguistic differences across registers that correspond to the differences in situational characteristics.

I Bibers analyser används rådata i form av frekvenser för ett flertal olika lingvistiska drag. Enligt Biber (1988: 13) ger (relativa) frekvensvärden "an exact, quantitative characterization of a text, so that different texts can be compared in very precise terms".

I boken *University language* redogör Biber (2006) för det språk som amerikanska studenter möter på universitetet. Undersökningen är fortfarande den mest uttömmande empiriska registeranalys som gjorts. Materialunderlaget utgörs av korpusen *TOEFL 2000 Spoken and written academic language corpus* (Biber m.fl. 2004), som innehåller 2,7 miljoner ord. Korpusen innehåller både tal- och skriftspråk (ca 1,7 respektive 1 miljon ord) från flera olika genrer, som t.ex. lektioner, laborationer, läroböcker och kursplaner. Dessutom täcks de stora ämnesfacken, humaniora, naturvetenskap och samhällsvetenskap, in, på såväl grundnivå som på avancerad nivå.

Några exempel på variabler som räknas är olika ordtyper (som "ord" räknas grundform plus regelbundna böjningsformer), ordklassfördelning, olika semantiska kategorier av substantiv (t.ex. abstrakta, konkreta, tekniska) och verb (t.ex. aspektuella, kausativa, mentala), verbfrasvariation (aspekt, diates, tempus) och olika bisatstyper (adverbiella, relativa, komplementära).

Resultaten visar framför allt att även om målen med läroböckerna och klassrumsundervisningen är desamma, skiljer sig de språkliga uttryckssätten avsevärt mellan de talade och skrivna registren. I läroböckerna förekommer betydligt fler substantiv och färre verb än i lektionsmaterialet. Läroböckerna innehåller också fler relativsatser och ovanliga ord, men färre predikativ och adverbial. Ämnet påverkar också mätningarna; tydligast är att ordvariationen är som störst inom humaniora och samhällsvetenskap. Lektionerna kännetecknas i stället av ett överflöd av adverb som markerar sannolikhet och säkerhet. Båda dessa verbtyper är ovanliga i läroböckerna. I undervisning används också mycket oftare olika slags modala verb och högfrekventa substantiv.

Förutom att beräkna förekomsten av lingvistiska drag bidrar Bibers registeranalys på ett par punkter med nyskapande metoder. För det första utför han en multidimensionell analys av språkdragen med hjälp av faktoranalys. Genom att jämföra värden för olika texter identifieras grupper, eller *dimensioner*, av språkliga drag som samförekommer i texter. Därefter tolkas grupperingarna i funktionella termer. På så vis prioriteras de lingvistiska dimensionerna framför de funktionella. Arbetssättet grundar sig på antagandet om att starkt samförekommande lingvistiska mönster delar viktiga kommunikativa funktioner:

”Features do not randomly co-occur in texts” (Biber 1988: 13). Funktionerna behövs alltså för att redogöra för de observerade samförekommande lingvistiska dragen, men ges inte på a priori-basis.

Några exempel på registeranalytiska tolkningar är att passiver och nominaliseringar representerar en underliggande abstrakt eller informell situation, medan verb i förfluten tid och tredje personens pronomen kan tyda på en berättande stil (Biber 1988). Dimensionerna används alltså för att beskriva olika registertyper och benämns t.ex. *narrative orientation* och *oral versus literate discourse*.

För det andra undersöker Biber (2006) s.k. *lexikala knippen* (eng. *lexical bundles*), som definieras som sekvenser av tre eller fler ord som frekvent samförekommer. Exempel på knippen är *at the end of* och *as well as the* (Biber och Conrad 1999). Lexikala knippen har tre övergripande användningsområden; de kan fungera attitydmässigt, diskursorganisatoriskt eller referentiellt. Dessutom har varje register en typisk uppsättning knippen med speciella diskursiva funktioner (Biber och Barbieri 2007).

I Bibers undersökning (2006) visar sig lexikala knippen förekomma betydligt mer frekvent i klassrumsundervisning än i läroböcker och akademisk prosa. I de talade registren används särskilt knippen med attitydrelaterad betydelse och funktion.¹² De flesta knippena i läroböckerna har strukturen av nominalfras- eller prepositionsfrasfragment. Få knippen förekommer inom alla ämnen. I stället är de flesta typiska för olika ämnen, som t.ex. *an increase in the* som används i naturvetenskap och teknik, och *by the fact that* som används i samhällsvetenskapliga ämnen.

Den maskindrivna analysmodellen bidrar till att långt större material kan behandlas än om undersökningen görs manuellt. I och med det att studierna kan göras mer omfattande och därmed mer generaliserbara, ökar naturligtvis resultatens reliabilitet. Sammanfattningsvis har Bibers arbeten med stora korpusar tydligt underbyggt uppfattningen om att alla (rimliga) registeruppdelningar delar karakteristiska språkdrag.

2.3.2.2 Korpusstilistik

Litterär analys med hjälp av kvantitativa metoder har förekommit sedan 1960-talet, men är fortfarande långt ifrån ett standardiserat område vad gäller metoder, omfång och teoretiska utgångspunkter (Hoover 2008). Idag utförs de flesta kvantitativa litterära analyserna inom det relativt nya forskningsfältet kor-

¹²För svenska visar emellertid Ribeck (2011) och Ribeck och Borin (2014) att ”läroböcker” som register inte räcker, utan uppdelningen måste göras mer specifik, med hänsyn till både ämne och stadium.

pusstilistik eller stilometri (t.ex. Mahlberg 2007; Hoover 2008; Biber 2011). Några användningsområden inom stilometri är t.ex. författaridentifiering (t.ex. Burrows 2007; Craig 2004; Hoover 2001), undersökningar av olika författarstilar (t.ex. Hori 2004; Fischer-Starcke 2010; Mahlberg 2015) eller av enskilda texters stilart, jämfört med andra texter av samma författare eller från samma tidsperiod (t.ex. McKenna och Antonia 2001; Stubbs 2005; Fischer-Starcke 2009). Mahlberg (2007) menar att man sällan upptäcker nya egenheter hos en text eller bok genom att analysera den med en kvantitativ ansats, i stället för att själv bara läsa den. Är man däremot intresserad av likheter och skillnader mellan stora mängder text, som man som enskild forskare inte kan läsa och analysera manuellt, är fördelen uppenbar.

Hoover (2008) menar att utnyttjandet av korpuslingvistiska, statistiska metoder för att beräkna förekomster av grammatiska och lexikala textuella drag är den enda metodologiska grund som kan sägas vara gemensam för korpusstilistiken. Dock kan några viktiga verktyg, som ofta används i de litterära analyserna, urskiljas, t.ex. studiet av nyckelord, ordfrekvenser, samförekomster (också kallade *lexikala knippen* eller *n-gram*), kollokationer och konkordanser. Dessutom betonas ofta vikten av att diskutera resultatens exklusivitet genom jämförelser med andra korpusar, antingen med en standardkorpus för språket i fråga eller med någon mindre, för ändamålet, ihopsatt korpus (se t.ex. Mahlberg 2007; O'Halloran 2007; Hoover 2008; Biber 2011; Nilsson Björkenstam, Gustafson Capková och Wirén 2014).

2.3.2.3 *Korpusjämförelser*

Korpusjämförelser ingår i många olika fält inom korpuslingvistik. Det språkteknologiska forskningsfält som arbetar med att analysera och utvinna information från texter brukar kallas för *text mining* eller *text analytics*. Algoritmerna inkluderar ofta avancerade statistiska metoder, men det är också vanligt att man använder sig av sofistikerad språkvetenskaplig databehandling, i form av t.ex. ordklasstagning och syntaktisk annotation.

Korpusanalys bygger i mångt och mycket på att identifiera och analysera språklig variation. Kvantitativa data från korpusar kan t.ex. användas till sådant som betydelsedisambiguering (Sutcliffe och Slater 1995), genrejämförelser (Mair 2006), koreferenstilldelning (Li m.fl. 2009), maskinöversättning (Teich 2003) och namnigenkänning (Ritter m.fl. 2011). Uppgifter om språklig variation kan också vara användbara för att beskriva skillnader mellan korpusar innehållande texter av olika typer. Sådana undersökningar syftar ofta till att identifiera och redogöra för lexikal variation mellan korpusar.

Allteftersom behovet av olika korpusjämförelser växer, ökar också behovet av en vedertagen metodologi för grundläggande korpuskompilering och processning för att genomföra jämförelserna. I linje med detta har några dataanalytiska arbeten försökt diskutera hur man bäst använder och tolkar kvantitativa data från olika korpusar (se t.ex. Dunning 1993; Rose och Kilgarriff 1998; Rayson och Garside 2000; Kilgarriff 2001; Cavaglia 2002; Gries 2006). Enligt Zhang, Huang och Quattri (2012) finns dock fortfarande inga vedertagna statistiska kriterier för att på ett tillfredsställande sätt ranka skillnader mellan korpusar.

2.3.2.4 Språkmodeller

Data från korpusar kan också ligga till grund för statistiska språkmodeller, vars uppgift är att beräkna sannolikheten för en viss ordsekvens (Chelba 2013). Många olika typer av moderna språkteknologiska tillämpningar är betjänta av sådan information, och språkmodeller används t.ex. till ordprediktion (Bickel, Haider och Scheffer 2005), maskinöversättning (Brown m.fl. 1990), stavningskontroll (Mays, Damerau och Mercer 1991) och taligenkänning (Rabiner och Juang 1993). Härutöver utnyttjas språkmodeller för en rad olika typer av klassificeringsuppgifter (Sebastiani 2002).

Många klassiska språkteknologiska problem, som parsning och ordklasstaggning, går ut på att tilldela en mängd objekt olika fördefinierade klasser, d.v.s. att klassificera objekten. Ett annat brett användningsområde för klassificeringssystem handlar om att sortera in tidigare osedda texter eller dokument i olika klasser. En sådan textkategoriserare bygger upp sin språkmodell genom statistiska inlärningsmetoder. Det vanligaste är att kategoriseraren först tränas på texter med manuellt tilldelade klassangivelser, s.k. övervakad inläring. Men det finns även oövervakade inlärningsmetoder, som använder något slags klustringsalgoritm för att låta inlärningsmetoden själv hitta likheter mellan dokumenten och därigenom göra klassificeringen (Clark och Lappin 2013).

Den mest kända tillämpningen av textklassificering är troligtvis informationssökning (Moens, Uyttendaele och Dumortier 2000). Men metoden kan också användas till att försöka uppskatta sådant som läsbarheten hos en text (Feng, Elhadad och Huenerfauth 2009).

2.4 Läroboksforskning

2.4.1 Läroböckers roll

Många forskare som har studerat läroböcker har beskrivit den viktiga uppgift dessa har i undervisningen. Så deklarerar t.ex. Selander (1988: 38): ”Det explicita syftet med läroboken är att med dess hjälp förmedla grundkunskaper till det uppväxande släktet”. En mer språkligt inriktad beskrivning av detta mål är att läroboken meddelar nya, för läsaren tidigare obekanta fakta, genom att förklara, definiera, precisera och exemplifiera (Danielsson 1975: 1, 88).

Utöver att på detta vis förmedla encyklopedisk kunskap, vilket hör till lärobokens s.k. *figurativa* funktion, fungerar läroböcker ofta såväl *operativt* som *formativt* (Ekvall 1997). Läroböckers operativa roll uppstår då elever uppmanas att själva söka och värdera kunskap, och i formativt syfte överför böckerna olika attityder och ideal. På så vis är alltså läroböcker utformade för att styra; genom boken visar författaren vilket innehåll som är viktigt och vilka värderingar som ska uppmuntras (Ekvall 2001). Englund (2011) sammanfattar lärobokens roll som kunskapsgaranterande, gemensamhetsskapande, auktoriserande och disciplinerande.

Läroboken är en oerhört grundläggande komponent bland s.k. pedagogiska texter, vilket Selander (1988: 17) definierar som texter som återskapar eller reproducerar befintlig kunskap. En pedagogisk text är producerad för en bestämd institutionaliserad användning, ett utbildningssystem organiserat i klasser, lektioner, stadier o.s.v. Den pedagogiska texten är alltså strukturerad enligt vissa pedagogiska krav, men fungerar även själv strukturerande för själva undervisningen (jfr även Säljö 2014; Carlsson och von Brömssen 2011).

Med hjälp av läroböcker konkretiseras till stor del utbildningens roll, mål och syfte (Carlsson och von Brömssen 2011), och i didaktiska studier av läroböckers användning i klassrummet konstateras att lärare ofta planerar sin undervisning utifrån läroböckers form och innehåll (se t.ex. Ekvall 2010; Svensson 2011). Särskilt styrande är läroböcker för mer strukturerade ämnesområden som uppfattas som fakta, t.ex. NO-ämnena, matematik och språk, medan läroböcker används på ett mindre styrande sätt för undervisning i SO-ämnena (Ekvall 2001). Att man utgår från föreställningen om att läroboken som sådan stipulerar kursen blir, enligt Ekvall (1997: 10), extra tydligt när det kommer till matematikböcker, vilka ofta produceras för olika nivåer och program. Böckerna behandlar då samma avdelningar, men vissa meningar kan skilja sig åt mellan utgåvorna.

I slutet av 1980-talet skrev Selander att skolgången innebär att vara samsatt med systematisk inhämtning av textbunden kunskap, och att det främst sker via läsning och bearbetning av lärobokens texter (Selander 1988). I dagens

samhälle, som blir allt mer textbundet (t.ex. Säljö 2014; Josephson 2004), är denna beskrivning än mer aktuell. För att lyckas som elev, måste man kunna inhämta ämnesinnehållet ur läroböckerna och återge det vid förhör och prov. I hur stor utsträckning elever lyckas reproducera lärobokstexten blir ofta ett bevis på lärande (Säljö 2014).

Naturligtvis har skolan intresse av att språket i läroböcker ska vara enkelt och lättfattligt, för att eleverna ska kunna förstå vad de läser. Men en annan aspekt av lärobokstexter är att de läses under den period i livet då man ska erövra ett mer vuxet språk. De ska därför även fungera som stilistiskt föredöme för elevernas eget språk, vad gäller stavning, ordböjning, ordval och meningsbyggnad, samt för hur olika texttyper utformas (Selander 1988; Strömquist 1995; Melander 2003). Genom att läroböcker erbjuder ”till synes rationellt och objektivt organiserade” ämnesspråkliga kategorier och strukturer ska dessutom läroboken hjälpa eleverna att utveckla språkliga tanke- och kunskapsstrukturer, d.v.s. organisera elevernas tänkande om omvärlden (Selander 1988: 17).

2.4.2 Studier av lärobokstexter

Studier av läroböcker har förekommit i Sverige sedan början av 1960-talet. Lärobokstexter kan betraktas från många olika infallsvinklar, t.ex. ideologiska, innehållsliga, pedagogiska eller språkliga (Ekvall 1997: 14). En omfattande systematisk forskningsbibliografi över verk producerade i denna tradition står Reichenberg (2010) för. Jag vill även hänvisa till Melanders (1995; 2003) genomgångar av svensk språkforskning om läroböcker.

Ur språkvetenskapligt perspektiv har språket i läroböcker överhuvudtaget rönt stor uppmärksamhet (se t.ex. Strömquist 1995). Böcker från låg- och mellanstadiet har dock inte studerats i lika stor omfattning som böcker från högstadiet och gymnasiet. På samma vis är naturvetenskap ett ämnesområde som har missgynnats, till skillnad från till exempel historia, som har utforskats mer ingående (Melander 1995, 2003).

Melander (1995, 2003) delar upp läroboksforskningen i olika inriktningar utifrån olika infallsvinklar och frågeställningar. I *deskriptiva* eller *beskrivande studier* besvaras frågan om hur språket ser ut genom att man undersöker syntax, ordförråd och stil. I *funktionellt värderande studier* bemöts frågan huruvida språket är bra eller inte, genom att man t.ex. undersöker om eleverna förstår vad de läser och om texterna fyller sitt syfte. Slutligen förekommer *studier med historiskt perspektiv* i vilka man undersöker språklig förändring över tid.

Denna uppdelning representerar även i viss mån den kronologiska utvecklingen av läroboksstudier; de första studierna från 1960-talet var främst beskrivande, därefter kom det funktionella perspektivet in i bilden och på senaste

tid har historiska studier tillkommit. Värt att påpeka är dock att de olika typerna av läroboksstudier inte är inbördes uteslutande; beskrivningar av språket föranleder t.ex. ofta synpunkter om funktion, svårighetsgrad eller historisk utveckling.

Fortsättningsvis i detta kapitel kommer jag att koncentrera mig på läroboksforskning med särskild relevans för avhandlingsområdet, d.v.s. främst grammatiskt beskrivande studier.

2.4.2.1 Beskrivande studier

Traditionen med beskrivande läroboksstudier inkluderar såväl kvantitativa stilbeskrivningar som textlingvistiska studier. I detta stycke går jag igenom de mest omfattande och inflytelserika kvantitativt stilistiska läroboksstudierna. För beskrivande textlingvistiska studier hänvisas t.ex. till Melin 1992, 1995; Ekvall 1995; Reichenberg 2000; Edling 2006; Nygård Larsson 2011.

I Westmans studie av bruksprosa (1974) beskrivs ett 20-tal texter från gymnasieläroböcker i biologi, samhällskunskap, historia, religionskunskap, psykologi och socialkunskap med avseende på bl.a. ordklassfördelning, ordvariation, meningslängd, bisatstyper och meningsbyggnad. Westman undersöker även en del icke-grammatiska företeelser, som styckeindelning och användning av bildtexter, fotnoter och rubriker. Siffrorna för lärobokstexterna sätts sedan i relation till tre andra bruksprosagenrer: informativa broschyrer, tidningsartiklar och debattartiklar.

En annan viktig del i Skrivsyntaxprojektet är Hultmans och Westmans *Gymnasistsvenska* (1977), där man undersöker om gymnasister behärskar bruksprosan som genre. I den studien fungerar bruksprosaundersökningen som jämförelsematerial och utgör därmed en uttalad stilistisk förebild och norm för det språk studenterna ska erövra i skolans undervisning.

Danielssons avhandling (1975) bidrar också till den kvantitativt stilistiska kartläggningen av svenskt läroboksspråk på 1970-talet. I denna analyseras kvantitativa aspekter av ordförråd och syntax, bl.a. satsantal och meningslängd, i högstadie- och gymnasieböcker i fysik, samhällskunskap, och svenska. Danielsson studerar även ensatsmeningar mer utförligt, samt substantivanvändning ur olika aspekter. Hon ser "läroboksspråk" som löpande text som behandlar ämnets fackstoff och utesluter således t.ex. övningsuppgifter från analysen.

Den bild som Westmans och Danielssons undersökningar ger av läroböcker är samstämmig. Böckerna har ett koncentrerat och faktamättat språk. Ordklassmässigt kännetecknas de av en opersonlig nominal stil, med många substantiv och adjektiv och färre verb och adverb. Vad gäller ordförrådet finns många

ovanliga ord, som tyder på ett precist och innehållsrikt stoff. På det grammatiska planet finns däremot färre utmärkande drag, men syntaxen bedöms vara relativt enkel, med få bisatser. Orden, snarare än syntaxen, bär alltså upp informationen. Trots detta är meningarna ofta relativt långa, beroende på att mycket innehåll packas in i långa nominalfraser, i vilka substantiven modifieras med adjektiv. Westman (1974: 223) använder följande exempel ur en biologilärobok för gymnasiet för att visa texten bakom siffrorna.

Kommunikationen mellan de olika individerna inom en art resulterar ofta i gruppbildningar av olika storlek och styrka. Beteckningar som hjord, flock, stäm och koloni är uttryck för sådana temporära eller mer permanenta samhällsbildningar hos olika djur. Hos insekterna möter vi exempel på en högt organiserad samhällsbildning med stark differentiering mellan individerna (del 1, s. 87). I etologisk mening räknar man emellertid också fortplantningsbeteende, vård av avkomma m.m. till socialt beteende, även när det gäller arter som i övrigt lever solitärt, d.v.s. inte bildar samhällen. En av förutsättningarna för ett socialt beteende är att individerna kan känna igen varandra och att en kommunikation kan åstadkommas mellan dem.

I exemplet lägger man enkelt märke till att substantiven är betydligt fler än verben, vilket är typiskt för den nominala stil som kännetecknar läroböckerna i förhållande till övriga bruksprosaer i undersökningen. Dessutom utmärker sig tydligt många innehållsrika, ovanliga och långa ord, som *samhällsbildningar*, *permanent*, *etologisk* och *fortplantningsbeteende*. Syntaktiskt sett dominerar huvudsatserna och bara några få meningar innehåller bisatser. På grund av långa nominalfraser med prepositionsfraser som kopplar samman olika begrepp är meningarna trots allt relativt långa. Lärobokstexterna i Westmans studie innehåller faktiskt de längsta nominalfraserna i förhållande till övriga texttyper.

Eftersom den språkliga analysen till stor del måste göras för hand på 1970-talet, grundas resultaten från dessa studier på s.k. stickprov ur läroböcker. Westmans sammanlagda läroboksmaterial uppgick t.ex. till 23 000 ord (vilket motsvarar ungefär en fjärdedels lärobok). Hultman och Westman (1977: 6) beskriver analysarbetet i Skrivsyntax-projektet som mycket arbetskrävande: "Varje ord i varje text försågs med utförliga grammatiska och analytiska kodbeteckningar, som sedan stansades på hålkort och lästes in på magnetband." En annan typ av begränsning som manuella textanalysmetoder har framtvingat är s.k. fallstudier. I en diakron studie av historieläroböcker undersöker t.ex.

Selander (1988) avsnitt som behandlar samma innehållsmässiga stoff (storkraftstiden) ur läroböcker från olika tider, för att upptäcka eventuella stilistiska förändringar.

Sedan 1970-talet har textforskningen utvecklats i flera olika riktningar. Inom det mer moderna, tvärvetenskapliga projektet *Svensk sakprosa 1996–2000* (Englund och Ledin 2003), har språkvetare, pedagoger och idéhistoriker arbetat med att teoretisera sakprosforskningen, samt med att undersöka svensk sakprosa ur ett historiskt perspektiv, för att följa utvecklingen från det att svenskarna blev ett läsande folk. Inom ramarna för projektet konstaterar Ekvall (2001) att läroböcker har en viktig roll som mönsterbildande för sakprosan i stort, eftersom de tillhör de mest lästa och spridda sakprosatexterna (jfr också Melander 2003).

Även om sakprosaprojektet inte grundar sig på någon gemensam teoriförståelse är det tydligt att studierna har ett ökat fokus på ”text i kontext” jämfört med tidigare textstudier. Några av de metoder som beskrivs i projektarbetena är uttalat eklektiska, i det att analysenheterna hämtas från olika håll, särskilt ofta från Hallidays funktionella grammatik, Bachtin-traditionen och Faircloughs kritiska diskursanalys. Dock har ingen tagit sig an det kvantitativa angreppssättet, där språket som system utforskas, trots att dagens teknik och ökade kunskap gör det betydligt lättare att räkna språkliga variabler i texter.

Läroböcker från olika ämnen utgör ofta ett samlat underlag för registeranalyser på engelska (Biber och Conrad 2009). För svenskt vidkommande är dock såväl Danielsson (1975) som Selander (1988) tydliga med att *läroboksspråk* inte bör ses som ett sammanhängande system (vilket t.ex. Westman 1974 gör). Danielssons resultat sägs peka mot att ”det synes föreligga skillnader mellan ämnen och mellan stadier, som snarast ger skäl att tala om *olika* läroboksspråk” (Danielsson 1975: 3; *emfas i original*). Selander menar att en viss lärobok är producerad för en speciell kontext och utgör därigenom ett register, en generalisering som relaterar texters formella kännetecken till deras funktioner i bestämda situationskontexter, i detta fall ämnesundervisningskontexten.

Efter Danielssons undersökning har emellertid ingen tagit sig an uppgiften att göra en omfattande registeranalys av moderna, svenska läroböcker från både olika ämnen och olika stadier, vilket gör att de referensvärden för lärobokstexter som anges i stilistiska handböcker (t.ex. i Melin och Lange 2000) kommer från de begränsade undersökningar som gjordes på 1970-talet.

2.4.2.2 Funktionellt värderande studier

Man kan angripa frågan huruvida en text är förståelig på olika sätt. Å ena sidan kan man undersöka inherent egenskaper hos texten som kan göra den mer eller mindre lättillgänglig. På så vis kan till exempel komplicerad meningsbyggnad och långa ord antas ha en försvårande inverkan. Texten kan sedan undersökas med avseende på sådana parametrar, jämföras med andra texter och därigenom placeras i ett större sammanhang.

Å andra sidan kan man se på texten som något som visserligen på ytan har vissa egenskaper som kan påverka begripligheten, men som inte nödvändigtvis säger något om i vilken grad en läsare har förstått texten. För att nå sådana slutsatser måste man studera texten på ett djupare och mer innehållsligt plan. Man ser begriplighet som något som uppstår i mötet mellan texten och läsaren. För att undersöka detta krävs att man analyserar textens innehåll och sätter detta i relation till det läsmål en viss läsare har i en given lässituation, samt hur läsaren läser texten. Härigenom kan man bilda en uppfattning om hur väl anpassad texten är till olika läsare och lässituationer.

Det första angreppssättet kallar Melander (2003: 141) för *läsbarhet* och det andra för *begriplighet*. Läsbarkhetsforskningen har ofta varit inriktad mot ytliga, kvantitativa mått, s.k. *läsbarhetsformler*, som syftar till att snabbt kunna mäta texters svårighetsgrad. För svensk del är den mest kända läsbarkhetsformeln *läsbarhetsindex* (LIX, Björnsson 1968).

LIX-värden har tagits fram för flera olika typer av texter, däribland lärobokstexter för olika stadier. Dessa värden kan användas som jämförelsetal vid textanalys. Resultaten pekar mot att läroboksspråket är relativt avancerat, men inte anmärkningsvärt komplicerat i förhållande till andra texttyper.

Begriplighetsundersökningar föranleder ofta kritiska reflektioner och sammanfattningsvis kan dessa undersökningar sägas peka i samma riktning: läroböckerna framstår som innehållsligt och grafiskt splittrade. Det sägs därför vara svårt att relatera olika delar till varandra (Melin 1992; Ekvall 1995; Melin 1995). Enligt Melander (2003) kan detta tänkas göra läroböckerna svåra att förstå, samt göra dem till dåliga förebilder för studenternas egen språkutveckling och skrivande.

2.4.2.3 Studier med historiskt perspektiv

I tidigare läroboksstudier med historiskt perspektiv (se t.ex. Sandqvist 1995; Ekvall 1997) har man sett att läroböcker verkar följa allmänspråkliga förändringar relativt väl. Under hela 1900-talet har skriftspråket successivt förenklats och närmat sig talspråket vad gäller ordval och syntax. I Sandqvists studie

av historieläroböcker från 1950- och 1980-talet konstateras att de tendenser av förenkling av skriftspråket som karakteriserar svenskan över huvud taget främst märks i läroböckerna genom förkortad text- och styckelängd. Skillnaden märks inte lika tydligt i kortare meningslängd eller färre underordnade satser med tiden (Sandqvist 1995).

På senare tid har man märkt en ökande trend av direkt tilltal och skönlitterär stil (Melander 1995; Ledin och Selander 2003; Lindberg och Johansson Kokkinakis 2007) i läroböcker. Enligt Nygård Larsson (2011) har dessutom användningen av olika läsbarhetsformler, exempelvis LIX, i textproduktionen bidragit till att läroböcker numera ofta verkar domineras av huvudsatser (jfr även Melander 2003 om att läsbarhetsstudier påverkar utformningen av läroböcker).

2.4.3 Språket i naturvetenskapliga läroböcker

Naturvetenskapliga lärobokstexter beskrivs ofta som svårtillgängliga för elever med ordval och grammatik som ligger långt ifrån vardagligt språk (Ekvall 2011). Inom ramen för SFL, som internationellt har belyst de språkliga krav naturvetenskapliga läroböcker ställer, beskrivs språkbruket som särskilt abstrakt, tekniskt, specifikt, informationstätt, distanserat, formellt och objektivt (se t.ex. Halliday och Martin 1993; Martin 1993; Halliday 1998; Martin och Veel 1998; Lemke 2000; Schleppegrell 2004). Samtliga dessa effekter bidrar till ett förhållandevis komplicerat språk.

Komplexiteten förklaras närmare genom att flera och långa ord ofta packas samman i utbyggda nominalfraser, vilket ger naturvetenskapliga texter en mycket hög lexikal densitet (Halliday 1993). På lexikal nivå utmärks texterna också av tydliga tekniska taxonomier, som skiljer sig från vardagstaxonomier, vilka organiseras enligt andra kriterier. Härutöver förekommer många abstrakta substantiv och termer som kan ha både fackspecifik och allmän betydelse.

I en svensk studie visar Edling (2006) att naturvetenskapliga texter innehåller en betydligt högre andel abstrakta och generaliserande substantiv jämfört med texter i samhällsvetenskap och svenska. Graden av abstraktion och generalisering ökar dessutom i alla ämnestexter från mellanstadiet till gymnasiet.

På grammatisk nivå påträffas framför allt många s.k. *ideationella grammatiska metaforer*, speciellt nominaliseringar, vilka bidrar till innehållslig förtätning (jfr Siljan 2011). Dessutom finns speciellt två stilgrepp som gör texterna distanserade: dels många opersonliga konstruktioner med verb i passiv, dels en informationsstruktur i vilken man ofta tematiserar resultatet av en handling, snarare än den som utför densamma.

Naturvetenskapliga texter innehåller också många olika slags texttyper eller textaktiviteter som är kännetecknande för B-texter. Inom SFL-traditionen benämns några av de vanligast förekommande texttyperna i naturvetenskaplig text som "procedur", "rapport", "förklaring", "exposition" och "diskussion" (Veel 1997; Unsworth 2001; Schleppegrell 2004; Martin och Rose 2008). A-texter, eller berättande inslag, är dock som regel sällsynta (Ekvall 2011).

Slutligen är naturvetenskapliga lärobokstexter i hög grad multimodala. De kräver god förmåga att läsa och förstå formler, tabeller och diagram (Kress m.fl. 2001). I svensk kontext uppmärksammar Nygård Larsson (2011) den relativt avancerade multimodala kohesion som kan finnas mellan olika slags visuella textelement i biologitexter.

2.4.4 Ord i läromedel – OrdiL

Projektet *Ord i Läromedel* (OrdiL, Lindberg och Johansson Kokkinakis 2007) pågick 2003–2008 med syfte att kartlägga ordförrådet i läromedel för högstadiet. Genom att först samla in och sammanställa en lärobokskorpus framställdes sedan ordlistor med frekvens- och spridningsuppgifter för orden i korpusen. I dessa listor redovisas vanligt förekommande ämnesoberoende ord (s.k. "allmänt skolrelevanta" ord), typiska fackord inom olika ämnen, samt fackspecifika användningar och vardagliga och mer frekventa ord (Holmegaard m.fl. 2005; Lindberg och Johansson Kokkinakis 2007, 2008).

Tanken är att en fördjupad kunskap om de lexikala krav som läromedel i olika ämnen ställer ska kunna vara till hjälp, dels för andraspråkslevers utveckling av skolrelaterat språk, dels för andraspråkslevers utveckling av skolrelaterat språk, dels för forskare och läromedelsförfattare. Informationen är också tänkt att ligga till grund för ett systematiskt ämnesövergripande pedagogiskt arbete med utvecklingen av ett skolbaserat ordförråd, och då särskilt som underlag för utveckling av olika typer av diagnostiska bedömningsinstrument av elevers ordförråd (Lindberg och Johansson Kokkinakis 2007, 2011).

3

MATERIAL: INSAMLING

I avsnitt 2.4.1 i den teoretiska bakgrunden beskrivs lärobokstexter som de mest typiska pedagogiska texterna inom ämnesundervisning. Det konstateras dessutom att läroböcker spelar en viktig roll för utvecklingen av ämnesspråklig kompetens. Visserligen förekommer även andra pedagogiska texter i undervisningen, både talade och skriftliga, som kan fungera som ämnesspråkliga förebilder, framför allt kanske lärarnas genomgångar, men även kompendier, stenciler och provuppgifter. Av såväl praktiska som metodologiska skäl har mitt material för att undersöka skolspråket emellertid begränsats till ursprungligen skriftliga källor, och mer specifikt till lärobokstexter.¹³ Jag kommer analysera alltså endast den del av ämnesspråken som lärobokstexterna utgör.

Lärobokstexter är beständiga på ett sätt som gör dem särskilt betydelsefulla för språkutvecklingen; deras formuleringar kan läsas om och om igen, till skillnad från t.ex. talad kommunikation, som lärarnas mer flyktiga genomgångar och förklaringar av kursinnehållet. Vad gäller stenciler och kompendier kan de vara svårare än läroböcker att hålla reda på och bevara. De har heller inte en lika given roll i undervisningen som läroböckerna, utan står ofta för mer exkursiva inslag. Läroböckerna däremot utgör typiskt fullständiga, fördjupade och samlade informationskällor till kursinnehållet. Dessutom har det visats att elever främst möter ämnesspråken i de skrivna texterna i läroböckerna. Men i den muntliga kommunikationen talar såväl lärare som elever ett vardagsspråk (se 2.1.1).

Ytterligare en anledning till att denna studie koncentreras kring läroboksspråk är att läroböcker innehåller betydligt *mer* språk än andra ämnesspråkliga källor, samtidigt som ett så stort materialunderlag som möjligt är viktigt för det vidare analysarbetet, samt att skillnader i språkbruk i olika skolämnen

¹³Jfr Nygård Larsson (2009) som kallar all text (skrift, bild, tal) som läraren producerar och/eller använder i undervisningen för "lärartext". Därefter särskiljs läroboken från övrig lärartext, t.ex. stenciler, overhead och anteckningar på tavlan, som ofta omges av lärarens tal. Lärobokens särställning grundar sig i att den utgör referensram och utgångspunkt för både elever och lärare.

märks tydligast i skriven text (Håkiss 2014). Läroboksinsamlingen i sig var tillräckligt tidskrävande, och att samla in andra skriftliga pedagogiska texter bedömdes helt enkelt inte vara mödan värt, i relation till det kvantitativa tillskott som det hade inneburit. De hade troligtvis inte tillfört något avgörande till analysen, än mindre kunnat analyseras separat med sådana kvantitativa, språkteknologiska metoder som är aktuella för just detta avhandlingsarbete.

För att kunna tydliggöra särskilda drag i naturvetenskapliga ämnesspråk jämför jag språket i naturvetenskapliga läroböcker med språket i läroböcker i matematik och samhällsvetenskap. För att kunna följa stadigångarna fullt ut, från högstadiet via gymnasiet till högskolan, har jag även valt att undersöka akademiska texter, vilka jag betraktar som representativa för det mest avancerade utbildningsrelaterade språkbruket. Härutöver är även andra texttyper som elever möter relevanta, speciellt berättande texter, vilka har rapporterats vara den enda texttyp som gymnasieelever inte har ”problem med” och som fokuseras mycket inom svenskämnets språkundervisning (se t.ex. Nyström Höög 2006).

3.1 Lärobokstexter

För att undersöka skolans ämnesspråk använder jag mig av den lärobokskorpus som samlades in i projektet *OrdiL* (se avsnitt 2.4.4). Enligt Lindberg och Johansson Kokkinakis (2008) är de ingående texterna i denna samling representativa för det läroboksspråk som elever kommer i kontakt med under år 7–9 i skolan. *OrdiL*-korpuser innehåller omkring en miljon ord och består av två läroböcker i vart och ett av ämnena biologi, fysik, kemi, matematik, geografi, historia, religion och samhällskunskap (Lindberg och Johansson Kokkinakis 2007: 102). Härutöver har jag samlat in, digitaliserat och bearbetat fler och nyare lärobokstexter i dessa ämnen, specifikt med denna undersökning i åtanke, både för högstadiet och för gymnasiet. För gymnasiet saknades tidigare material helt, och för högstadiet behövdes korpuser kompletteras med både fler och nyare läroböcker.

Lärobokskorpusarnas sammansättning syns i tabell 3.1 och samtliga läroböcker som har ingått i undersökningen listas i appendix B.1. Totalt består lärobokskorpusarna av 63 böcker, varav 49 stycken är nyinsamlade och står för mer än fyra femtedelar av den totala textmassan på drygt 5,2 miljoner token.¹⁴ Som kan utläsas av tabell 3.1 har läroböcker i de naturvetenskapliga ämnena prioriterats i materialinsamlingen. I dessa ämnen föreligger minst 4 böcker och 300 000–700 000 token för varje stadium. Allt som allt undersöks

¹⁴Som tidigare nämnts omfattar termen *token* ord och skiljetecken. Se vidare avsnitt 4.1 om lingvistisk textannotering.

27 läroböcker i naturvetenskap. Även matematik är väl representerat, vilket är bra med tanke på detta ämnes rikliga bruk av symboler, eftersom meningar med sådana försvinner i bearbetningen av texterna (se vidare avsnitt 4.2 om rensning av texterna). Sammanlagt ingår 16 samhällsvetenskapliga böcker i studien. Antalet läroböcker i samhällsvetenskapliga ämnen är alltså inte lika många, men i studien ingår åtminstone två böcker från varje ämne och stadium, och i flera fall fler än så.

Ämne	Högstadiet		Gymnasiet		Totalt	
	Böcker	Token	Böcker	Token	Böcker	Token
<i>Biologi</i>	4	440 510	4	394 413	8	834 923
<i>Fysik</i>	4	315 973	4	370 670	8	686 643
<i>Kemi</i>	5	400 202	6	693 047	11	1 093 249
<i>Matematik</i>	9	851 420	11	721 952	20	1 573 372
<i>Geografi</i>	3	163 561	2	210 222	5	373 783
<i>Historia</i>	3	157 890	2	144 329	5	302 219
<i>Samhällsk.</i>	3	136 636	3	220 953	6	357 589
Totalt	31	2 466 192	32	2 755 586	63	5 221 778

Tabell 3.1: Lärobokskorpusarnas uppbyggnad och storlek

Under materialinsamlingen visade sig läroböcker i religion vara svårast att få tag på i digitalt format. För gymnasiet erhöles endast en bok, vilket gjorde att storleken på religionskorpusen blev betydligt mindre än korpusarna i de övriga samhällsvetenskapliga ämnena. Förutom ett relativt litet underlag hade författarna till denna enda bok också haft ensamrätt till det gymnasiespråk i religion som skulle ha undersökts. Eftersom det inom ramen för detta projekt inte har varit möjligt att få tillgång till någon ytterligare lärobok i religion för gymnasiet, har jag i stället valt att utesluta ämnet ur undersökningarna. Detta gör att ämnesfacket samhällsvetenskap, som vanligtvis omfattar fyra ämnen, i detta arbete representeras av de tre återstående ämnena: geografi, historia och samhällskunskap.

Läroböckerna från OrdIL är från åren 2001 till 2003 och de läroböcker som jag har samlat in är från 2010 till 2013. Inom detta tidsspann har en omformulering av läroplanerna ägt rum och vissa läroböcker från de senare åren följer explicit den nya läroplanen. Min undersökning har ambitionen att vara synkron och alltså inte göra någon skillnad över tid gällande läroboksspråket. Det är därför särskilt viktigt att jag försäkras mig om att de variabler jag studerar inte är tidsbundna på ett sätt som innebär en avgörande skillnad i mätresultaten över denna tid. Denna fråga utreds närmare i avsnitt 6.5.1.

För att ge en uppfattning om hur lärobokstexterna ser ut återges i figur 4, 5 och 6 några korta exempel på avsnitt ur läroböcker för högstadiet i de naturvetenskapliga ämnena. Den första bilden, från biologi, känns igen från inledningskapitlet.



För att en isbjörn ska bli fullvuxen måste den äta många sälar. I slutändan är det mängden växtplankton som styr hur stora populationerna kan bli.

Mängden producenter sätter gränsen

En isbjörnshanne väger cirka 500 kg. Vad har han ätit för att bli så stor? Jo, vanligen sälar. För enkelhets skull räknar vi med att det bara är 10 % av sälens vikt som blir till isbjörnskött. Då har isbjörnen behövt äta ungefär fem ton säl för att nå sin vikt. Sälén i sin tur äter gärna torsk. För att få fem ton säl behövs då cirka 50 ton torsk. Och för att få 50 ton torsk behövs ungefär 500 ton sill. Sillarna behöver i sin tur stoppa i sig cirka 5 000 ton djurplankton – som i sin tur behöver äta ofattbara 50 000 ton växtplankton!

Med hjälp av den här förenklade räkneövningen förstår du att det inte kan finnas fler isbjörnar i ett ekosystem än vad mängden växtplankton, och deras fotosyntes, sätter gränsen för. Ju fler trofinivåer, desto färre toppkonsumenter. Detsamma gäller för rovdjuren i de svenska skogarna, örnarna längs våra kuster och lejonerna på Afrikas savanner.

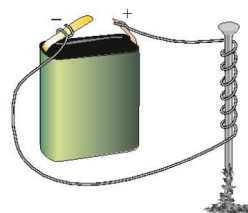
Figur 4: Utdrag från en lärobok i biologi för högstadiet (Fabricius, Holm och Nystrand 2013: 164). Återges här med tillstånd från Liber

Så skapas en elektromagnet

Tänk dig en sytråd som är lindad många varv runt en trådrulle. Om vi på liknande sätt lindar en lång elektrisk ledare får vi en *spole*. När det går ström genom spolen bildas ett *magnetfält* inne i och omkring spolen. Det liknar magnetfältet runt en vanlig stavmagnet.

Med den kunskapen kan vi bygga en enkel *elektromagnet*. Allt vi behöver är en järnspik, en sladd och ett batteri. Vi lindar sladden runt spiken och kopplar sedan sladden till batteriet. När det går ström genom sladden, blir spiken magnetisk och kan till exempel dra till sig järnspån. Om vi ökar strömmen i sladden, eller lindar den flera varv runt spiken, blir elektromagneten starkare. Men det går ingen ström genom spiken, bara i sladden.

Fördelen med en elektromagnet jämfört med en vanlig magnet är att elektromagneten kan stängas av.



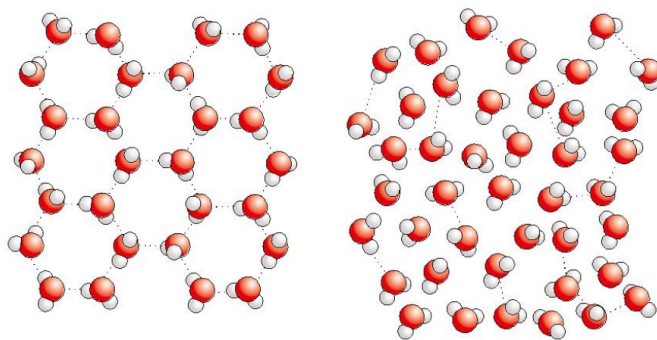
Det här är en enkel elektromagnet. Det är strömmen i den spiralformade ledaren som gör spiken magnetisk. Men genom spiken går ingen ström.

Figur 5: Utdrag från en lärobok i fysik för högstadiet (Undvall och Karlsson 2013: 231). Återges här med tillstånd från Liber

Varför flyter is?

Hur kommer det sig att is flyter ovanpå vatten? Jo, det beror på att is är lättare än vatten.

Det som väger någonting i ett ämne är alltid atomerna och molekylerna. I is ligger vattenmolekylerna ordnade i ett glest mönster. Därför får det inte plats lika många molekyler i en liter is som i en liter vatten. Resultatet blir att en liter is väger mindre än en liter vatten. Det är därför isen flyter ovanpå vattnet.



I is (till vänster) ligger vattenmolekylerna i ett glest stelt mönster. Därför får det inte plats lika många molekyler i en liter som i flytande vatten (till höger). De prickade linjerna visar dragningskraften mellan de positiva och negativa sidorna i molekylerna.

Figur 6: Utdrag från en lärobok i kemi för högstadiet (Nettelblad och Nettelblad 2013: 88). Återges här med tillstånd från Liber

3.2 Akademiska texter

Som en del av projektet *Akademiska ordlistor* har vi satt ihop korpusen *SveAk* (se 2.2.4) på 26 milj. token med publicerade akademiska texter (Ribeck, Jansson och Sköldberg 2014). Denna korpus företräder det akademiska språket i min undersökning. Korpusen innehåller avhandlingar och tidskriftsartiklar från åren 1997 till 2012 och finns tillgänglig via korpusinfrastrukturen *Korp* (Borin, Forsberg och Roxendal 2012) i Språkbanken.

Vid insamlingen av avhandlingar har vi utgått från den nationella databasen *SwePub*.¹⁵ Denna listar alla publikationer från svenska universitet och högskolor enligt Högskoleverkets standard för svensk indelning av forskningsämnen (2011), som, i sin tur, följer OECD:s internationella klassifikation över forskningsämnen.¹⁶ Forskningsämnena klassificeras i första hand i sex olika ämnesområden: naturvetenskap, teknik, medicin och hälsovetenskap, lantbruksveten-

¹⁵<http://swepub.kb.se/>

¹⁶Field of Science and Technology (FOS) Classification in the Frascati Manual

skap, samhällsvetenskap och humaniora. Under vart och ett av dessa finns, i sin tur, mellan fem och elva ämnesgrupper; psykologi och utbildningsvetenskap tillhör exempelvis det samhällsvetenskapliga ämnesområdet.

Tabell 3.2 visar SveAk:s sammansättning. Korpusen består av texter från ämnesområdena humaniora och samhällsvetenskap. Ur dessa har vi valt ut ämnesgrupper där förhållandevis många avhandlingar har funnits tillgängliga som fulltext-pdf:er genom SwePub. De humanistiska ämnesgrupperna som ingår i SveAk är: etnologi, filosofi, historia, konst, litteraturvetenskap, religion och språkvetenskap. De samhällsvetenskapliga ämnesgrupperna är: ekonomi och näringsliv, juridik, medie- och kommunikationsvetenskap, psykologi, social och ekonomisk geografi, sociologi, statsvetenskap och utbildningsvetenskap.

Ämne	Token		Totalt
	Avhandlingar	Tidskrifter	
	Humaniora		
Etnologi	1 396 095	103 306	1 499 401
Filosofi	637 088	328 538	956 626
Historia	2 482 954	1 343 167	3 826 121
Konst	1 649 537	73 723	1 723 260
Litteraturvetenskap	2 083 628	227 355	2 310 983
Religion	2 572 471	–	2 572 471
Språkvetenskap	1 832 567	–	1 832 567
Totalt	12 656 340	2 076 089	14 732 429
	Samhällsvetenskap		
Ekonomi/Näringsliv	1 886 324	–	1 886 324
Juridik	682 560	–	682 560
Medie-/Kommunikationsv.	1 130 813	–	1 130 813
Psykologi	339 848	43 072	382 920
Social/Ekonomisk geografi	1 621 081	–	1 621 081
Sociologi	1 838 189	–	1 838 189
Statsvetenskap	1 556 733	41 983	1 619 861
Utbildningsvetenskap	1 826 970	355 621	2 182 591
Totalt	10 882 518	440 675	11 323 193
SveAk totalt	23 538 858	2 516 764	26 055 622

Tabell 3.2: Sammansättning av den svenska akademiska korpusen SveAk

Tidskrifterna, varifrån artiklarna i SveAk kommer, har valts ut med hjälp av den norska publikationsklassificeringsmodellen som NSD ansvarar för.¹⁷ För att komma på fråga för inkludering i SveAk behövde tidskriften lägst befinna sig på nivå ett enligt detta system, vilket motsvarar ”scholarly” (akademisk, vetenskaplig).

Sammanlagt innehåller SveAk drygt 26 miljoner token. Korpusen består av 500 texter skrivna av fler än 450 olika författarkonstellationer. Som synes i tabell 3.2 utgörs den största textmassan av avhandlingar, medan tidskriftsartiklarna står för ett betydligt mindre bidrag.¹⁸

För att summera de tankar som legat till grund för SveAk:s sammansättning definierar vi det svenska akademiska skriftspråket som bestående av texter skrivna av och för akademiker, på svenska. Vi har således låtit det representeras av texter från forskningsämnesområden där den svenska akademiska produktionen är tillräckligt hög. Inom naturvetenskap (som i SwePub inkluderar matematik) skrivs det nästan uteslutande på engelska på svenska universitet och högskolor (jfr Salö 2010). Efter att ha genomfört ett test att sätta samman delkorpusar med svenska sammanfattningar från naturvetenskapliga avhandlingar konstaterades att den sammanlagda textmassan blev för liten i proportion till övrigt material i korpusen för att anses lönsam och försvarbar.

Jag har också undersökt möjligheten att samla in naturvetenskapliga läroböcker för universitetet till den akademiska korpusen. Det visar sig att det finns ett fåtal svenska läroböcker i biologi som används på universitetsutbildningar. Kurslitteraturen i fysik och kemi är dock uteslutande på engelska redan på grundnivå. Jag har bedömt att vinsten med att inkludera någon lärobok i biologi i korpusen som för liten i förhållande till den arbetsinsats som krävs för att få tillgång till materialet elektroniskt.

Inom det medicinska området produceras förvisso akademiska texter på svenska; i Språkbanken finns t.ex. artiklar från *Läkartidningen* mellan åren 1996 och 2001. Eftersom medicin inte förekommer som skolämne, har jag emellertid valt att inte inkludera dessa artiklar i mitt akademiska textmaterial. För teknik och lantbruksvetenskap gäller dels att det svenska underlaget är mycket tunt, dels att ämnena inte ingår bland dem som undersöks på skolnivå i denna studie.

Exempel på meningar ur de olika ämnesgrupperna i de akademiska texterna visas i (1) till (15).

¹⁷<http://dbh.nsd.uib.no/kanaler/>

¹⁸Dessutom kunde inte vetenskapliga svenska tidskrifter inom samtliga ämnesområden hittas. De ämnesområden och tidskrifter som finns representerade i SveAk listas i tabell B.1 i appendix B.2.

52 *Material: insamling*

- (1) *Att poängtera ett mindre revolutionärt agerande kan därmed, hur paradoxalt det än kan tyckas, tolkas som ett sätt att framhäva östtyskarnas agens. (etnologi)*
- (2) *Världens helhet blir därmed inte en totalitet av ting utan en överskridande ordning och historia som delas. (filosofi)*
- (3) *Kulturarbetarna hade nödvändiga kunskaper om hur man producerar och genomför en föreställning, men utan den estetiska berättelsens identitetsskapande kraft hade förmodligen framgången uteblivit. (historia)*
- (4) *Pågående tonala durationer transformeras därefter till tidsliga enheter simultant hos deltagarna och ju fler saker som pågår samtidigt, desto mer krav ställs på deras simultanitet. (konst)*
- (5) *Jag kommer att visa att dessa dimensioner i texten är betydelsefulla för huvudpersonens inre utveckling liksom för romanens nationella tematik. (litteraturvetenskap)*
- (6) *Vi måste hitta ett sätt att ta tillvara både uppfattningen att det finns en medvetendeoberoende verklighet som är som den är oavsett vad vi tror och tänker om den och uppfattningen att verkligheten alltid är begreppsliggjord. (religion)*
- (7) *Den hypotetiska situationen tycks bli ett sätt att närma sig ett ämne som annars riskerar att bli socialt ansiktshotande. (språkvetenskap)*
- (8) *Min undersökning visar att omsorg, när det gäller lågavlönade ensamstående mödrar också kan handla om att förvärva extra inkomster för att motverka barnens marginalisering. (ekonomi/näringsliv)*
- (9) *Den stora merparten av avverkningar sker dock på skogsmark där det är tillräckligt att anmäla avverkningen till Skogsstyrelsen. (juridik)*
- (10) *Under undersökningsperioden ökade antalet program och inslag per år, i absoluta tal, från ett till drygt etthundra, men i relativa tal har sändningsandelen i förhållande till totala sändningstiden förhållit sig relativt likartad sedan 1956. (medie-/kommunikationsvetenskap)*
- (11) *Lyhördhet innebär att terapeuten har en förmåga att visa sig känslig för patientens behov och känslomässiga tillstånd och att ge gensvar på ett adekvat och konstruktivt sätt. (psykologi)*

- (12) *Ett hållbart lantbruk skulle, om en stor del av kriterierna från det mogna ekosystemet är uppfyllt, präglas av en hög resiliens, d.v.s. systemet är stabilt och har förmåga att klara av störningar.* (social/ekonomisk geografi)
- (13) *Gender School syftade till att sprida kunskap om könssegregeringens komplexitet och att få kvinnor och män medvetna om den könsdiskriminering som förekommer i arbetslivet.* (sociologi)
- (14) *Den aristokratiska och maskulina förmågan att med våld kunna försvara hedern hade avgörande betydelse för den tyska nationens framväxt.* (statsvetenskap)
- (15) *Men i det andra fallet betraktas den gemensamma grundskolan, en skola för alla, nästintill som förtryckande likformig, där individens intresse och begåvning inte tas tillvara, utan alla elever stöps i samma form.* (utbildningsvetenskap)

Exempelmeningarna, som anges i befintligt skick, kan ge en viss inblick i hur det akademiska textmaterialet ser ut. Som för de flesta texttyper förekommer meningar av olika språklig kvalitet och komplexitet.

3.3 Berättande texter

För att kunna undersöka berättande texters språkliga struktur har jag satt ihop en romankorpus. Denna har sammanställts efter noggrant övervägande om vilka slags texter som bäst representerar sådana berättande texter som elever kommer i kontakt med före och under skolan. Lämpligheten för enskilda böcker har bedömts dels efter materialets vardagsspråkighet, dels efter vilken åldersgrupp det riktar sig till (helst från omkring 7 år och uppåt). Jag har försökt uppfylla dessa två kriterier genom en kombination av olika romanmaterial som har varit tillgängliga via Språkbanken.

Korpusen med berättande texter består av sammanlagt 66 romaner och drygt 4,4 miljoner token. Vuxenromanerna är klart överrepresenterade; de är inte bara flest utan utgör också en klar majoritet av den totala textmassan, vilket beror på att de är förhållandevis långa. Samtliga romaner som ingår i korpusen listas i appendix B.3 och en översikt av korpusens sammansättning presenteras i tabell 3.3.

Före skolstarten läser barn få texter på egen hand. Dock kan antas att barnböcker läses **för** barnen, varför sådana texter är lämpliga som materialunderlag för studier av vardagligt berättande språk som riktar sig till barn. Korpusen

	Åldersgrupp	Böcker	Token
Barnböcker	6–9	12	106 756
Barnböcker	9–12	9	291 562
Ungdomsromaner	12–15	5	223 272
Vuxenromaner	15–	40	3 821 430
Totalt		66	4 443 020

Tabell 3.3: Korpus med berättande texter

LäSBarT (Mühlenbock 2009) innehåller bl.a. barn- och ungdomsböcker och finns (delvis) tillgänglig via Språkbanken. Barnböckerna riktar sig till barn i åldern 6–9 år, respektive 9–12 år, alltså på låg- och mellanstadiet, och borde i hög grad motsvara ett slags berättande text som barn är bekanta med när de börjar högstadiet och först möter de naturvetenskapliga lärobokstexterna av mer facklig karaktär. Ungdomsromanerna riktar sig till människor i åldern 12–15 år, alltså i högstadieåldern. Romaner för vuxna har hämtats ur korpusarna *Norstedtsromaner* från 1999 och *StorSUC-romaner* från första hälften av 1990-talet.

I samtliga material har ett noggrant urval gjorts, till förmån för böcker som innehåller **vardaglig** dialog och som utspelar sig i **vardagliga** miljöer och behandlar **vardagliga** situationer. I urvalsprocessen sorterades ungefär hälften av titlarna från samtliga delmaterial ut. För att fånga den berättande genren bland vuxenromanerna togs exempelvis inte memoarer, biografier, kåserier, satirer, vetenskaplig prosa och verk med hög litterär prägel med. I syfte att välja romaner som utspelar sig under vardagliga förhållanden valdes uttalade kriminalromaner, samt en roman som utspelar sig i krigets Bosnien, bort. Bland barnböckerna valdes dessutom *Lätt att läsa*-böcker (t.ex. Helena Bross böcker) bort av anledningen att de är skrivna under speciella förutsättningar, som inte är fokus i min studie.

Korpusen innehåller 12 böcker för åldergruppen 6–9 år och 9 böcker för åldersgruppen 9–12 år. Några slumpvis utvalda exempel på meningar ur den förra kategorin kan ses i (16) och (17) och ur den senare kategorin i (18) och (19). Av flertalet exempelmeningar framgår tydligt att berättelserna utspelar sig i en vardaglig miljö. Hälften av meningarna utgörs av repliker där den tal-språkliga karaktären märks i både ordval och syntax, trots att böckerna återger *fingerat* och inte *autentiskt* talspråk. Även de många personliga pronomina är ett typiskt talspråkligt stildrag (jfr Liljestränd 1993: 151).

- (16) *Berätta om när jag var liten, sa jag och kröp ner under täcket.* (ur *En dag i prinsessan Victorias liv*)
- (17) *Hon äter lite godis och dricker några klunkar läsk.* (ur *Inez långa väg till mormor*)
- (18) *Dessutom har jag ju ingen annan kompis än dig.* (ur *Drömponnyn*)

Mening (19) skiljer sig dock från de andra i det att den är en generell utsaga, som inte berör närvarande deltagare. Den är mer fackartad och skulle lika väl kunna vara hämtad ur en NO-lärobok för mellanstadiet. Trots att barnböckerna tydligt domineras av den berättande texttypen kan genren uppenbarligen även inrymma drag från andra texttyper.

- (19) *En människa kan inte vara ensam med stjärnorna och planeterna.* (ur *Den döde talar*)

Som framgår av tabell 3.3 innehåller korpusen fem ungdomsromaner. Exempel (20) och (21) är hämtade ur ungdomslitteraturen och båda meningarna uppvisar den typiskt berättande strukturen med tydliga karaktärer. Mening (20) är skriven ur ett tredjepersonperspektiv där berättaren återger en romanfigurs inre tankevärld i form av en retorisk fråga. Även om meningen är relativt lång och innehåller flera abstrakta ord, som *uppdrag*, *existera* och *innebära*, är den grammatiskt förhållandevis enkel i sin utslutande samordnande karaktär.

- (20) *Någon måste väl ha känt till både Matildas uppdrag och Ladons uppgift att skydda henne, och att halsbandet existerade och vad det innebar.* (ur *De andra*)

Exempel (21) är en replik ur en ordväxling skildrad ur ett objektvt berättarperspektiv. Repliken är vardagsspråklig och personlig, vilket syns syntaktiskt i att-strykningen och i ”så ADV”-konstruktionen, och lexikalt i bruket av de personliga pronomina och svordomen.

- (21) – *Du tror du är så jävla smart, sa Tuula.* (ur *Vargön*)

Två godtyckliga exempel ur romaner från Norstedts återges i (22) och (23). Även i dessa berättelser riktade till vuxna är vardagsmiljön uppenbar och personliga pronomina vanligt förekommande. De berättande inslagen – centeringen kring olika karaktärer och ett tydligt återgivet händelseförlopp (jfr Virtanen 2010: 56f.) – är också påtagliga, liksom tydligt talspråkliga. Det första exemplet utspelar sig i hemmiljö och beskriver den vardagliga sysslan ’att duka fram frukost’.

- (22) *När hon kommer ut ur badrummet har han börjat duka fram frukost i köket. (ur Det heliga namnet)*

Också mening (23) är starkt informell till sin utformning, vilket märks på stavningsvarianterna *dom* och *mej*, samt på det slangspråkliga ordvalet *snubbe*.

- (23) *Ytterligare två snubbar kommer in och det första dom gör är att spotta på mej. (ur Simple storys [sic!])*

Ungefär hälften av Norstedtsromanerna är översatta till svenska från ett annat språk, vilket oftast är engelska. Tidigare forskning har observerat ett visst mått av överföring av källspråkets struktur till målspråket just i fallet romaner som översatts från engelska till svenska. Enligt Gellerstam (1989, 1996) finns flera exempel på s.k. *översättningssvenska* på såväl ord- som frasnivå. Men också syntaktisk påverkan sägs förekomma till viss del. Hittills har dock inga storskaliga studier på syntax i översatt svenska gjorts, varför Ingo (2007: 29) fortfarande efterlyser en objektiv utforskning av lingvistiska fenomen som aktualiseras vid översättning. Eftersom översättningsindustrin av skönlitteratur till svenska är mycket omfattande,¹⁹ och en stor del av svenska berättande texter därför torde bära spår av översättningssvenska, har jag valt att inkludera de översatta romanerna i materialet. Det är trots allt autentiska, berättande texter som elever kommer i kontakt med som jag vill undersöka, och inte några förebilder ur språkvårdsperspektiv.

StorSUC-romanerna, å andra sidan, är samtliga svenska originalverk och exemplifieras i (24) och (25). Dessa meningar är klart mer skriftspråkliga till karaktären än tidigare exempel ur berättandekorpusen. Här förekommer nominaliseringar/verbalsubstantiv, som *bebyggelse*, participer, som *vitmantlade*, och ovanliga sammansättningar, som *sjöfarare* (jfr skriftspråkliga drag i Liljestränd 1993: 150–153). Att det fortfarande rör sig om berättande texter framgår dock tydligt även av dessa meningar; handlingen kretsar kring mänskliga figurer och framställs i berättande stil.

- (24) *Tidigt på morgonen gick han ut i staden, där det fanns en del äldre bebyggelse men framförallt en stor hamn. (ur Oäktingarna)*
- (25) *Välkomna, kära sjöfarare! utbrast den främste av de vitmantlade med hes stämma. (ur Färdemän)*

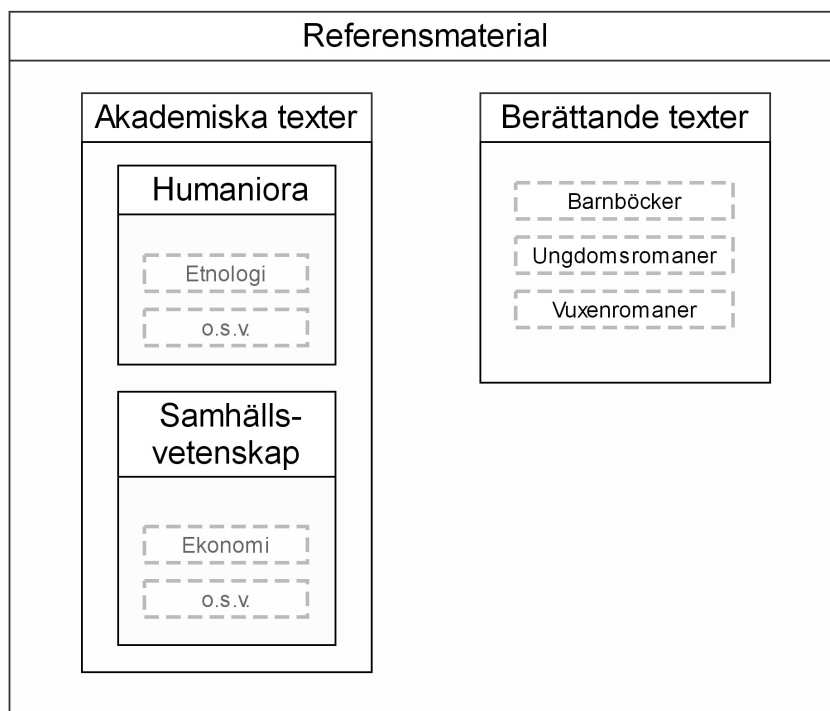
¹⁹2012 var 69 % av alla skönlitterära böcker som gavs ut på svenska översättningar från andra språk, varav engelska var originalspråk i 68 % av fallen (Kungliga biblioteket 2012).

3.4 Översikt över materialunderlaget

Översiktsbilder av läroboksmaterialets respektive referensmaterialets sammansättning visas i figur 7 och 8.



Figur 7: Översiktsbild av läroboksunderlaget



Figur 8: Översiktsbild av referensmaterialet

4

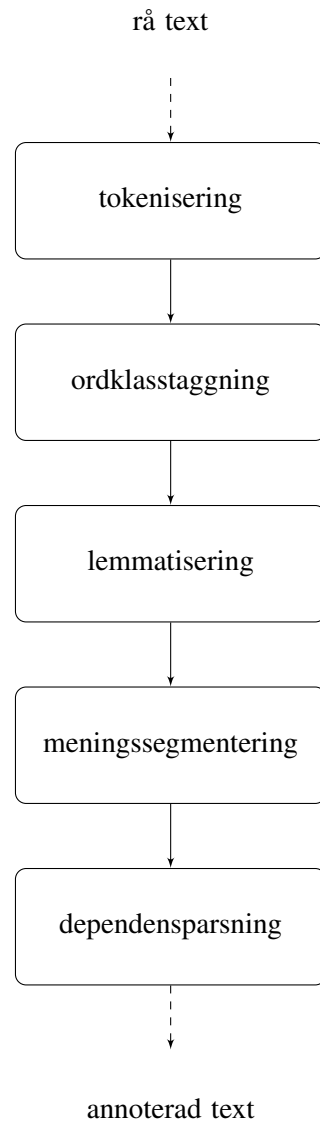
MATERIAL: BEARBETNING

I detta kapitel förklaras hur textmaterialen har bearbetats inför undersökningarna. Först beskrivs den språkteknologiska annoteringen av texterna (avsnitt 4.1). Därefter redogörs för hur texterna har rensats på olika vis, i första hand hur utomspråklig data, t.ex. i form matematiska formler, har tagits om hand (avsnitt 4.2), och därefter hur läroböcker som förekommer i olika versioner för olika gymnasieprogram har behandlats (avsnitt 4.3).

4.1 Lingvistisk annotering

Samtliga textmaterial har annoterats lingvistiskt med hjälp av det paket som i nuläget är standard för analys av svenska och som används av Språkbanken (se vidare Borin, Forsberg och Roxendal 2012). Annoteringsprocessen, som innefattar tokenisering, ordklasstaggning, lemmatisering, meningssegmentering och syntaktisk dependensparsning, illustreras i figur 9.

I det första annoteringssteget, tokeniseringen, delas teckensträngen upp i s.k. *token*, som typiskt omfattar ord och skiljetecken. Hur detta ska göras definieras av en uppsättning reguljära uttryck, som t.ex. beskriver hur ord med bindestreck och förkortningar ska behandlas. När väl token har identifierats kan de *taggas* med ytterligare information. Den första lexikala information som alla ord förses med är en ordklassangivelse. I detta skede skiljs t.ex. substantivet *springa* från det homografa verbet. Härnäst följer lemmatiseringen, genom vilken alla ord tillskrivs ett böjningsparadigm. Verbet *springa* hör t.ex. till mönstret *springa*/VB.



Figur 9: Den språkteknologiska annoteringsprocessen

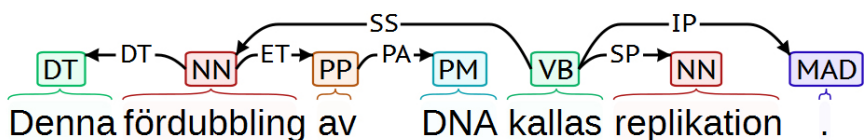
Liksom den lexikala analysen, inleds den syntaktiska *parsningen* med ett slags segmentering i grundläggande enheter. Den här gången handlar det om *meningar*. När gränserna för sådana har bestämts, analyseras varje mening enligt en formell grammatisk beskrivning och förses med taggar som anger de mest sannolika syntaktiska relationerna mellan orden. Den automatiska syntaxanalysen är det mest sofistikerade steget i den analysapparat som jag tillämpar på mina texter, eftersom den bygger på andra, underliggande analyser, som morfologisk analys och ordklasstagning.

Kort kan nämnas att tokeniseringen utförs med *PunktWordTokenizer*, som ingår i *Natural Language Toolkit*, Pythons standardpaket för språkvetenskaplig databehandling. Ordklasstagningen görs med *HunPos* (Halácsy, Kornai och Oravecz 2007; Megyesi 2009),²⁰ som implementerar ett slags *HMM* (Hidden Markov Model), med ordklasstaggar från *SUC2* (Ejerhed m.fl. 1992). Lemmatiseringsalgoritmen utgår från det svenska, elektroniska associationslexikonet *SALDO* (Borin, Forsberg och Lönnngren 2013). Meningarna dependensanalyseras med *MaltParser* (Nivre, Hall och Nilsson 2006), vars statistiska regler genereras från Talbanken (Nivre, Nilsson och Hall 2006). I denna process förses varje ord med en syntaktisk tagg från uppsättningen *MAMBA* (Teleman 1974).

Idén med korpusannoteringen är att informationen som lagrats kan extraheeras och användas på olika sätt. För att t.ex. hitta exempel på ett visst ord i bruk är det smidigt att automatiskt kunna skilja mellan homografa ord från olika ordklasser. För många lexikala sökningar är det också betydligt mer meningsfullt att kunna identifiera böjda ordformer som förekommer av samma ord, än att bara registrera förekomster av *textord* (den ortografiska motsvarigheten till ord) eller *graford* (homografa textord). Den syntaktiska informationen kan t.ex. användas till att analysera olika frastyper, grupper av ord som bildar syntaktiska enheter, eller olika meningsstrukturer.

Ett exempel på en parsad mening ur en kemilärobok för gymnasiet visas i figur 10. Exemplet läses med fördel nedifrån och upp. Längst ner står själva teckensekvensen: *Denna fördubbling av DNA kallas replikation*. Strax ovanför sträcker sig ett antal krullparenteser över de enheter som tolkats som token. Varje token har tilldelats en ordklasstag, som inramas av en rektangel. Exempelvis har ”denna” taggats som ett bestämningsord (DT) och både ”fördubbling” och ”replikation” som substantiv (NN). De pilförsedda bågarna illustrerar slutligen syntaktiska dependensförhållanden mellan olika token. Till exempel har ”DNA” taggats som prepositionskomplement eller rektion (PA) till prepositionen ”av”.

²⁰Denna kommer förmodligen i framtiden att ersättas med ordklasstagaren *Stagger* (Östling 2013).



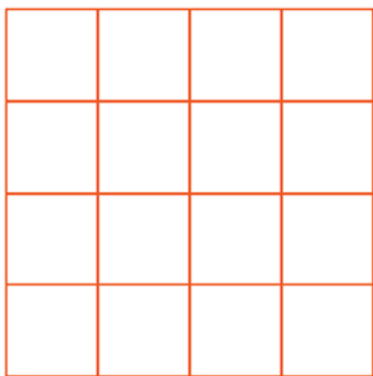
Figur 10: Lingvistisk analys av en mening ur en kemilärobok från gymnasiet

4.2 Rensning av texter

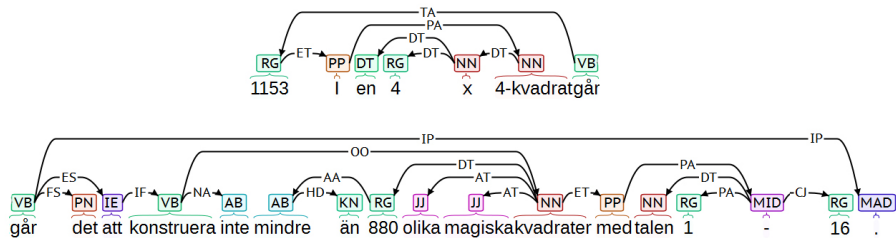
Ett exempel på en sida i en matematikbok för högstadiet visas i figur 11. Det första segment, från just denna sida, som annoteringsverktyget har identifierat som en mening är "1153 I en 4x4-kvadrat går det att konstruera inte mindre än 880 olika magiska kvadrater med talen 1–16". Den fullständiga lingvistiska analysen av denna mening kan ses i figur 12. Som synes har det inledande uppgiftsnumret, "1153", analyserats som huvudord i ett meningsinitialt tidsadverbial (TA), likt årtal kan vara. Vidare har "x" och "4-kvadrat" taggats som substantiv (NN).

1153

I en 4 x 4-kvadrat går det att konstruera inte mindre än 880 olika magiska kvadrater med talen 1–16. Prova att göra minst tre av dessa magiska kvadrater med talen 1–16. Tänk först efter hur stor den magiska summan blir.



Figur 11: Layoutexempel från en matematikbok för högstadiet (Mårtensson, Sjöström och Svensson 2012:23). Återges här med tillstånd från Gleerups



Figur 12: Lingvistisk analys av en mening ur en matematiklärobok för högstadiet

Denna mening är på intet sätt utvald som ett skräckexempel; åtminstone matematiktexterna utgörs till stor del av symboler, siffror, formler och tabeller. De annoterade texterna innehåller därför oerhört många dylika och ännu ”värre” exempel på hur taggningen störs av utomspråkliga data i materialet. Verktøyen som används är utvecklande för naturligt språk, och saknar därför kvalificerade regler för hur ickespråkliga element ska taggas. Det faktum att t.ex. multiplikationsoperatoren, ”x”, räknas som ett substantiv, inverkar naturligtvis menligt på kvantitativa beräkningar av sådant som ordklassfördelning, nominalfraslängd och nominalkvot, och gör resultaten svårtolkade.

För att de språkliga variabler som studeras faktiskt ska begränsas till att omfatta just språkliga data, har jag valt att rensa lärobokstexterna på icke-språkliga data i största möjliga mån.²¹ För detta syfte har jag utvecklat automatiska rensningsregler som opererar på de redan annoterade texterna.²²

När rensningsreglerna formulerades utgick jag från läroboksmaterialet. Inledningsvis prövades en relativt enkel lösning, som grundades på antal tecken som ordklassstaggats som MID (små skiljetecken som + / =) eller som RG (räkneord, grundtal). Efter att manUEllt ha utvärderat hur denna rensning fallit ut, genom att, från varje ämne, kontrollera 100 slumpvist utvalda meningar som rensats bort och 100 som behållits, konstaterades att reglerna behövde vara mer sofistikerade. Stickproven studerades då mer i detalj. Därefter skrevs rensningsprogrammet om, utvärderades på samma vis igen, och skrevs om igen, tills det att ett tillfredsställande resultat uppnåddes.

²¹Jfr hur Danielsson (1975: 12) manUEllt utesluter meningar med ekvationer och formler från fysiktexterna i sin undersökning, eftersom ”meningar med ekvationer etc. vållar problem såväl vid ordräkning som vid sats- och meningsräkning”.

²²För andra typer av textanalyser, t.ex. för diskursanalyser med multimodala ansatser, hade det förmodligen varit naturligare att försöka utveckla verktygen för att kunna ta hand om texterna mer i sin helhet.

Det viktiga för en registeranalys är inte att ha kompletta texter, utan att ha representativ text som kan analyseras med de metoder som används. I utvärderingarna har jag därför valt att prioritera kvaliteten på materialet som går vidare till analys, precisionen, framför storleken eller täckningen (eng. *recall*). Förutom att isolera språket i texterna och sortera bort meningar som innehåller matematiska formler, sorterar reglerna därför också bort en del felsegmenterade meningar. Samtidigt vill jag naturligtvis inte rensa bort alltför mycket gott material i onödan, eftersom jag vill kunna analysera så stora material som möjligt. Den slutliga korrektheten i att skilja önskvärda, välformade meningar, från icke-önskvärda, uppskattas (genom stickprovskontroller enligt ovan) vara 97,5 %.

De slutgiltiga rensningsreglerna återfinns i sin helhet i appendix C. I stora drag innebär kontrollerna att en mening inte får (1) inledas eller avslutas med RG, (2) innehålla en siffra eller bokstav följt av en punkt eller högerparentes, (3) innehålla sammanlagt fler än fyra RG, (4) innehålla ett likhetstecken eller ett plustecken, (5) innehålla fler än tre förekomster av MID, PAD eller RG och minst två olika typer av dessa tecken, (6) innehålla fler än 53 ord eller (7) sakna VB.

I förhållande till läroböcker och akademiska texter innehåller de berättande texterna mycket litet symboler, tabeller och förteckningar. Trots att de fenomen jag avser rensa bort är marginella i berättande texter har jag ändå valt att rensa dessa, för att de kommande undersökningsresultaten ska grunda sig på texter som har bearbetats jämlikt. Samtliga textmaterial som ingår i undersökningarna genomgår således rensningsprocessen.

4.2.1 Lärobokstexter

Storleken på läroboksmaterialet efter rensningen framgår av tabell 4.1. Som synes har det återstående materialet decimerats i olika utsträckning beroende på ämne och benägenheten att använda symboler i ämnestexterna. I biologi och de samhällsvetenskapliga ämnena finns mer än 80 % av den ursprungliga textmassan kvar, medan knappt 70 % återstår av fysik- och kemitexterna. Matematiktexterna har reducerats klart mest; endast 28 % av textmassan har passerat rensningen.

Värdena i tabellen visar också att andelen symboler i de naturvetenskapliga ämnena, främst i fysik och kemi, är högre på gymnasiet än på högstadiet. I matematik ligger andelen på en relativt konstant, mycket hög nivå. I de samhällsvetenskapliga ämnena märks ingen skillnad i andelen symboler på de olika stadierna.

Ämne	Högstadiet		Gymnasiet		Totalt	
	Token	Andel	Token	Andel	Token	Andel
<i>Biologi</i>	382 471	87 %	331 741	84 %	714 212	86 %
<i>Fysik</i>	255 146	81 %	217 994	59 %	473 140	69 %
<i>Kemi</i>	329 934	82 %	414 525	60 %	744 459	68 %
<i>Matematik</i>	249 036	29 %	191 582	27 %	440 618	28 %
<i>Geografi</i>	139 995	86 %	163 684	78 %	303 679	81 %
<i>Historia</i>	137 095	87 %	130 736	91 %	267 831	89 %
<i>Samhällsk.</i>	115 980	85 %	192 928	87 %	308 908	86 %
Totalt	1 609 657	65 %	1 643 190	60 %	3 252 847	62 %

Tabell 4.1: Lärobokskorpusarnas storlek efter rensning

4.2.1.1 Kvarvarande meningar

I utvärderingen av den slutgiltiga rensningen uppskattas precisionen ligga på 98 % och täckningen på 97 % i det att identifiera naturliga meningar. Exempel (26) – (32) visar slumpmässigt utvalda prov på meningar som passerat rensningen ur läroböcker för gymnasiet för samtliga ämnen som ingår i undersökningen.

- (26) *För det tredje leder en mutation i en viktig gen oftast bara till att cellen dör, och lämnar mer plats för cellen bredvid att dela sig och ersätta den döda cellen.* (biologi)
- (27) *I den punkt där strålen träffar spegeln tänker vi oss en linje vinkelrät mot spegeln, en normal till spegelytan.* (fysik)
- (28) *Då blir det kvar en positivt laddad partikel – väteatomen bildar en vätejon H^+ .* (kemi)
- (29) *När man drar höjden i en triangel måste man tänka på att den ska gå från ett hörn, vinkelrätt mot basen.* (matematik)
- (30) *Nötningen blir störst i berggrundens mjuka och porösa partier och där gräver floderna först ut typiska V-formade floddalar.* (geografi)
- (31) *Genom industrialiseringen förändrades den militära tekniken och nu krävdes ett modernare försvar med fler soldater.* (historia)

- (32) *Under de yrkesverksamma åren betalar vi tillbaka till systemet genom skatterna.* (samhällskunskap)

Stickprovskontrollerna, av 100 meningar från varje ämne, tyder på att den absoluta huvudparten av meningarna i det kvarvarande materialet inte innehåller symboler i någon större utsträckning, och dessutom är välformade i klassisk grammatisk bemärkelse. Några exempel på mindre bra meningar förekommer dock i materialet, beroende på olika faktorer. Exempel (33) torde exempelvis utgöras av en rubrik.

- (33) *Svaren på minns du?* (kemi)

Huruvida denna mening är ”felaktig” eller ”korrekt” kan diskuteras. Det är en grafisk mening, som börjar med stor bokstav och avslutas med stort skiljetecken, och även en grammatisk mening (eller makrosyntagm), d.v.s. ett syntaktiskt självständigt yttrande (jfr Westman 1974: 42).²³ Meningen har alltså taggats på bästa möjliga sätt, under förutsättning att ”meningar” betraktas som en grundläggande enhet som all text kan delas upp i, vilket är den princip meningssegmenteraren följer. Att meningen inte utgörs av en fullständig sats är en annan sak.

Ett liknande resonemang kan föras kring exempelmening (34), som är ett exempel på en svarskonstruktion som torde vara kännetecknande för just lärobokstexter. Denna konstruktion återfinns i facit till övningsuppgifter och består av ett svar följt av en motivering, d.v.s. en grafisk mening och två makrosyntagmer.

- (34) *Australien därför att alla talar engelska och bor i samma land.*
(geografi)

Värt att påpeka är att stickprovskontrollerna pekar mot att alla ”felaktiga” meningar som behållits antingen beror på att den lingvistiska annotationen är tvivelaktig, eller på hur texterna är beskaffade, d.v.s. att de inte genomgående består av ”korrekta” meningar. Eftersom målet med rensningen inte är att förbättra själva parsningen, måste dessa felkällor anses oundvikliga. Under rådande förutsättningar fungerar därför rensningen optimalt.

²³En makrosyntagm ”utgör ett syntaktiskt sammanhängande led av maximalt omfång” (Loman och Jörgensen 1971: 10). Makrosyntagmer betecknar alltså syntaktiska enheter som fungerar likvärdigt i tal- och skriftspråk, vilket, förutom meningar, även täcker meningsfragment, interjektioner och tilltal.

4.2.1.2 Bortrensade meningar

I det att identifiera meningar som innehåller för mycket utomspråkligt innehåll eller är felparsade, felsegmenterade eller icke-fullständiga, uppskattas precisionen ligga på 97 % och täckningen på 98 %. Meningarna (35), (36) och (37) är riktiga skolboksexempel (!) på att matematiska formler har orsakat en bortrensning, vilken för mina syften är av godo.

- (35) *Den totala impulsen varje sekund blir då som lägst $6,70 \cdot 1016 \cdot 1,25 \cdot 10^{-27} = 8,34 \cdot 10^{-11}$ Ns och som högst $2 \cdot 6,70 \cdot 1016 \cdot 1,25 \cdot 10^{-27} = 1,67 \cdot 10^{-10}$ Ns. (fysik)*
- (36) *Då blir $n(\text{Ti}) = m(\text{Ti}) / M(\text{Ti}) = 475 \text{ g} / 47,9 \text{ g} \cdot \text{mol}^{-1} = 9,916 \text{ mol}$. (kemi)*
- (37) *1182 Vilket eller vilka av talen nedan är a) ett naturligt tal 1183 a) (32 – 18) / 2 14 1184 a) · 25 1185 a) (–6) · 3 1186 a) 0,52 b) ett rationellt tal 3,5 b) 0,05 31 2 b) ++ 523 b) 6 – (–3) b) (–6)2 A: 3 7 B: –5 C: p D: 52 E: 0,37 1187 Skriv talen i grundpotensform. (matematik)*

Exempel (38) har också på goda grunder försvunnit i rensningen. Denna ”mening” uppfyller faktiskt flera olika kriterier för bortrensning, t.ex. innehåller den för många grundtal och för många ord. Som synes kommer den från en innehållsförteckning och består av en rad kapitelrubriker som var och en föregås av ett sidnummer. Att ett kapitel råkar vara namngivet med en hel frågesats som avslutas med ett stort skiljetecken är det som tolkas som meningsslut. Denna teckensekvens innehåller mängder av makrosyntaxmer och skulle därför kunna anses vara felsegmenterad; att den undanhålls vidare analys måste anses fördelaktigt.

- (38) *124 Ekosystemen till vår tjänst 126 Mat till nio miljarder 128 Hållbar utveckling 130 Ekonomi - på naturens villkor 131 Miljöarbete i Sverige 132 Miljömål för nästa generation 133 Begränsad klimatpåverkan 134 Friskluft 137 Bara naturlig försurning 138 Levande sjöar och vattendrag 140 Grundvatten av god kvalitet 140 Myllrande våtmarker 141 Ingen övergödning 142 Hav i balans samt levande kust och skärgård 143 Ett rikt växt-och djurliv 144 Levande skogar 147 Ett rikt odlingslandskap 148 Säker strålmiljö 148 Storslagen fjällmiljö 149 Skyddande ozonskikt 149 Giftfri miljö 150 God bebyggd miljö 152 Hela jorden är vårt ekosystem 154 Sammanfattning och frågor 156 Uppgifter 157 Människan 160 Att vara människa 162 Att må bra och att må dåligt 163 Stenåldersmänniskan i dataåldern 164 Kroppens organ samarbetar 166 Vad är ett funktionshinder? (biologi)*

En mening som har rensats bort på grund av avsaknad av verb är (39). Denna meningstyp härrör förmodligen från bildtexter, och är inte lika tydligt önskad för min undersökning. Analysen i detta arbete torde emellertid, på det stora hela, gynnas av att meningsnotationen i materialet, så väl som möjligt, stämmer med den syntaktiska meningen (som innehåller ett predikat/finit verb).

- (39) *Churchill, Roosevelt och Stalin på Jaltakonferensen i februari 1945.*
(historia)

Även om en övervägande majoritet av de bortgallrade meningarna är just sådana som jag avser avlägsna med rensningsreglerna, förvinner några meningar som gärna hade fått passera till vidare analys. Ett exempel på en sådan naturlig mening som rensats bort är (40), som har gallrats ut på grund av att den innehåller naturligt språk som representeras av tecken som till formen liknar matematiska symboler.

- (40) *Det nazistiska Tyskland (1933 – 45) ansåg att ”den ariska rasen”, som tyskarna påstods tillhöra, var överlägsen judar, zigenare och slaver (östeuropéer).* (samhällskunskap)

I detta sammanhang kan det också vara värt att förklara att meningar ibland kan rensas bort på grund av felaktigt tilldelad lingvistisk annotation. Inledande imperativer taggas t.ex. genomgående som substantiv, vilket ofta resulterar i bortrensning, då meningen tycks sakna verb. Så är t.ex. fallet i mening (41).

- (41) *Sök i uppslagsverk efter riksdagsval.* (geografi)

4.2.2 Akademiska texter

Efter rensning av de akademiska texterna återstår 14 740 203 token, vilket motsvarar 57 % av originaltexterna. Det är något mindre än lärobokstexternas 62 %. En manuell granskning av (100 + 100) stickprov efter rensningen tyder först och främst på att de allra flesta meningar som har bevarats, omkring 95 %, är lämpliga för vidare analys. Totalt uppskattar jag alltså att max 5 % onaturliga meningar har passerat rensningen, något som oftast beror på missar i meningssegmenterarens förarbete (se t.ex. mening 42).

- (42) *Keywords: family values, gender roles, society, picturebooks, motherhood 1960- och 1970-talen var en omvälvande tid för svensk barn- och ungdomslitteratur.*

Stickprovskontrollerna av de meningar som rensats bort visar att dessa grovt sett kan kategoriseras i fyra grupper. Omkring 20 % av meningarna som rensats bort är sådana som uppenbart kommer från litteraturlistor, se (43), innehållsförteckningar, se (44), tabeller, se (45), eller dylikt, och i allt väsentligt inte lämpar sig för syntaktisk analys.

- (43) *George E. Haggerty (New York & London, 2000), "Sexology" i The Encyclopedia of Lesbian and Gay Histories and Cultures II.*
- (44) *61 Personliga lärdomar i förundran 61 Desillusionerade dekonstruktioner 64 Försvar av mischmasch och apolitiska liv 71 Diskussion 75 Existentiell sårbarhet 75 Stärker och skaver 76 Nationen 76 Historiernas kamp 78 Finhistoriens besvärjelse och börda 80 En beköad Historia 81 Dilemman och medlanden 82 Kapitel 3.*
- (45) *144 FMU 3494, 3657, 3658, 3660, 4283 (1472 03 08, 1473 datum saknas, 1476 11 06, 1490 02 21).*

Ungefär en lika stor andel av de bortrensade meningarna visar sig vara utländska, som (46). I sådana kan parsern inte identifiera något verb.

- (46) *This is an Open Access article distributed under the terms of the Creative Commons Attribution-Noncommercial 3.0 Unported License (<http://creativecommons.org/licenses/3.0/>), permitting all non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.*

Cirka 30 % av de bortrensade meningarna kommer från brödtexten och innehåller citat, som (47), och referenshänvisningar, som (48). Dessa meningar kan försvinna eftersom årtal och sidnummer ser ut som matematiska formler, och eftersom citattecken, bindestreck, parenteser och kolon m.m. räknas som skiljetecken.

- (47) *Sociologen Rosmari Eliasson som kritiskt granskade den sektoriserade psykiatrin, ställde frågan om det nya i den nya psykiatrin var "nytt i betydelsen en ny förpackning åt gamla funktioner eller nytt i betydelsen väsensskilt från psykiatrins gamla funktioner för den enskilde patienten och för samhället?" (Eliasson 1979:14).*
- (48) *De handlar dels om naturens stora betydelse i romanen och för Malin i synnerhet, dels den polyfona strukturen (Edström 2000, 82 – 85), och slutligen romanens existentiella djup (Edström 2007, 124).*

Slutligen utgörs resterande 30 % av riktigt långa ”meningar”, som är svåra att beskriva generellt. Se (49) och (50) för ett par exempel på hur dessa kan se ut.

- (49) *ena foten ... ett knä ... 150 Pregnoscape andra benet är strax bredvid ... där ser ni profilen ... panna ... nästipp ... och så kommer det en hand ... den rör sig ... svår att fånga ... armen ... handen är där vid knät ... man ser en sida i taget ... andra armen ligger där ... handen är där bakom huvudet ... ligger och [ohörbart] precis som du gör [till pappan som sitter bakåtlutad med händerna bakom huvudet] skratt ... det svarta är vatten ... det ser fint ut ... [ohörbart] ... det gråa är ... [ohörbart] ... [någon knackar på, barnmorskan har fått ett telefonsamtal] säg att jag ringer upp ... nu ska jag mäta då ... huvudet och magen.*
- (50) *I boken Nattvarden i evangeliskt gudstjänstliv nämner Yngve Brilioth de nytillkomna nattvardspsalterna, bl a Frostensons 'ljusa och innerliga': 'Du öppnar, o himmelske Fader, i Kristus din famn ...' samt bearbetningen av enhetsbönen ur Didake: psalmen vars inledande ord börjar: 'Som spridda sädeskornen'.⁵³⁶ I inledningskapitlet ger Brilioth en summering av olika arbeten vilka alla visar på svårigheten, att med utgångspunkt från bibelberättelserna, fastställa huruvida Jesu sista måltid hade samband med en rituellt etablerad sabbatsmåltid i anslutning till större högtider, eller om den ingick som en bland övriga judiska måltider, i vilken husfadern enligt sed välsignade brödet och vinbägaren.*

Några få välformulerade långa meningar har emellertid försvunnit enbart på grund av sin längd eller längden och den höga enkla nominalkvoten, som t.ex. (51).

- (51) *Den andra följdfrågan handlar om huruvida det inom professionen går att urskilja vilka grupper som huvudsakligen är mer positiva till publikorientering, i termer av att betona publikstudiernas vikt för företaget och dess journalister, samt vilka grupper som ställer sig mer kritiska till publikorientering.*

Trots att många meningar med hänvisningar och referat försvinner i rensningen, avviker ändå de flesta av de bortrensade meningarna i så hög grad från naturligt (svenskt) språk att rensningen bedöms vara välgörande.

4.2.3 Berättande texter

Efter rensning av de berättande texterna återstår 3 941 886 token, d.v.s. 89 % av texterna. Stickprovskontroller av 100 + 100 meningar pekar mot att samtliga av de meningar som återstår är välformulerade och önskvärda. Värt att notera är ändå att vissa anföringsuttryck, som (52), förekommer bland de meningar som har klarat kontrollen. Att de skilts från sin föregående replik beror förmodligen på att denna avslutats med stort skiljetecken. Fristående anföringsuttryck är dock typiskt för berättande texter (Liljestrand 1993: 156f.) och betraktas därmed som oproblematiskt.

(52) *suckade Anne.*

Bland de meningar som sållats bort tycks de flesta vara av typen makrosyntagmer som utgörs av ofullständiga sätser, som (53) och (54).

(53) *En fantastisk personlighet.*

(54) *Men varför?*

Ett fåtal meningar som försvunnit är sådana som är mycket långa, som (55). Dylika meningar hade förstås gärna fått vara kvar. Men denna rensningsregel gör överlag mer gott än ont, då de flesta teckensekvenser som taggats som mycket långa meningar i själva verket inte motsvarar grafiska meningar. Och, som sagt, alla delmaterial behandlats åtminstone på samma vis, vilket gör att riktigt långa meningar försvinner i alla texter.

(55) *Banken lånade ut pengar till dom bortkollrade seminariedeltagarna och lotsade dom till Gillrays för att köpa och sälja konst (och fick provision av Gillrays i bägge riktningarna), och när man ändå höll på arrangerade man VIP-lyxiga konstronder i New York, varpå man såg till att få överta alla deras bankaffärer, personliga och på företagsplanet.*

4.3 Överlappande läroböcker

När det gäller läroböcker i matematik för gymnasiet har jag utfört en ytterligare rensning av texterna. Natur och Kulturs läroböcker, *Matematik 5000*, finns i olika versioner för olika program: teknik-/naturvetenskapsprogrammet, de serviceinriktade yrkesprogrammen, ekonomi-/estetiskt/humanistiskt eller samhällsvetenskapsprogrammet, samt vuxenutbildningar. Gleerups matematikböcker, *Exponent*, finns också i olika versioner, där a-versionerna riktar sig till yrkesprogrammen och b- och c-versionerna till de studieförberedande programmen (jfr avsnitt 2.4.1).

Dessa matematikserier innehåller ofta avsnitt med samma rubriker, men med olika detaljerat innehåll. För att undvika att snedvrída frekvenserna av de språkliga parametrar som mäts, har jag rensat dessa delvis överlappande böcker på de mest uppenbara dublettformuleringarna. Denna rensning utfördes genom att jag skrev ett program som först kände igen identiska meningar som förekom exakt en gång i (minst två) olika böcker inom samma serie, och sedan endast behöll en av förekomsterna.

De olika upplagorna av naturvetenskapliga läroböcker i PULS-serien har genomgått samma rensningsprocess med avseende på meningsdubletter.

5

METOD

Avhandlingsarbetet anknyter till den inriktning inom textforskning inom vilken man intresserar sig för text som språklig struktur och som syftar till att karakterisera språket i en text eller i en grupp av texter. Mer specifikt ansluter jag mig till den kvantitativt stilistiska analystraditionen, som var som mest populär i Sverige på 1970-talet (se avsnitt 2.3).

I viss mån tar jag alltså ett steg tillbaka i tiden – och beskriver på nytt lärobokstexter objektivt utifrån grammatiska egenskaper. På samma gång vill jag föra textforskningen några steg framåt – genom att undersöka texterna med moderna, språkteknologiska verktyg. De variabler som jag utgår från kommer ifrån området för språkvetenskaplig textanalys och läsbarhetsforskning, och har operationaliserats för att kunna identifieras och studeras datamaskinellt. I stora drag kan metoden beskrivas som: *deskriptiv, grammatisk, korpusbaserad, kvantitativ, språkteknologisk, stilistisk* och *synkron*.

5.1 Ämnesspråklig avgränsning

Utformningen av de undersökningar jag redovisar i avhandlingen har influerats av teorier om olika undervisningsdomäner/stadier och om skolspråk/ämnesspråk. De tre undervisningsdomänerna, den *vardagliga*, den *specialiserade* och den *reflexiva* domänen (se avsnitt 2.1.1), kan kopplas till de språkbruk som jag undersöker, nämligen *vardagsspråk*, *ämnesspråk* och *akademiskt språk*. De textmaterial som jag har valt som representanter för dessa språkbruk är berättande texter, lärobokstexter och akademiska texter (materialvalen motiveras närmare i kapitel 3). Olika ämnesspråk undersöks genom läroböcker i olika skolämnen, och stadieövergångarna inom skolan undersöks genom att läroboksmaterialet delas upp i högstadie- och gymnasieböcker.

För att precisera de delar av ämnesspråk som jag studerar använder jag det tankeschema över läs- och skrivfärdighet från den australiska skolan som presenteras i avsnitt 2.1.2 i den teoretiska bakgrunden. Enligt modellen delas läs-

och skrivkunnighet upp i fyra grundläggande kunskapsområden: ordkunskap, kunskap om visuella uttrycksmedel, grammatikkunskap och kunskap om texter.

I figur 13 har jag framhävt de kunskapsområden som jag berör inom ramar för avhandlingen, och samtidigt tonat ner de fält som inte omfattas. De delstudier som ingår i avhandlingen behandlar alltså ord- och grammatikkunskap, viktiga komponenter i ämnestexter enligt modellen. Kunskap om visuella uttrycksmedel ingår inte i min definition av ämnesspråklig kompetens (jfr 1.2.2) och studeras därför inte. Men inte heller kunskap om texter omfattas av delstudierna, vilket givetvis begränsar vad jag i slutändan kan säga om ämnesspråken. Valet av delstudier och språkliga variabler att studera beskrivs närmare i avsnitt 5.2 i detta kapitel.



Figur 13: Illustration av de delar av ämnesspråk som undersöks (efter Government of South Australia 2015: 7)

5.2 Avhandlingens delstudier

Avhandlingen innehåller fyra olika delstudier, vilka, i tur och ordning, benämns: *traditionella textanalysmått*, *typiskt ordförråd*, *typiska nominalfraser* och *typiska meningar*. Inom varje delstudie analyseras vidare en eller flera olika mätbara egenskaper. De fyra delstudierna och samtliga variabler som undersöks listas i tabell 5.1.

Delstudie	Variabel
Traditionella textanalysmått	<i>lemmavariation</i> <i>läsbarhetsindex</i> <i>meningslängd</i> <i>nominalkvot</i> <i>ordklassfördelning</i> <i>ordlängd</i> <i>ordvariation</i>
Typiskt ordförråd	<i>ord</i>
Typiska nominalfraser	<i>nominalfraslängd</i> <i>nominalfrastäthet</i> <i>nominalfrasuppbyggnad</i>
Typiska meningar	<i>meningsuppbyggnad</i> <i>parsningsdjup</i>

Tabell 5.1: De variabler som undersöks i avhandlingens fyra olika delstudier

Valet av delstudier och variabler har påverkats av flera olika faktorer. En utgångspunkt har varit att undersöka ämnesspråkliga drag i ett så brett perspektiv som möjligt och att påvisa fördelar med språkteknologisk textanalys på flera olika språkliga nivåer. Samtidigt har jag naturligtvis velat studera sådana variabler som är meningsfulla att beskriva utifrån syftet med undersökningen. Den hypotes som framförs i inledningskapitlet, att texters utformning påverkar för-

ståelsen för dem, föranleder delstudier som på något sätt indikerar olika grader av språklig svårighet eller komplexitet. En annan grundläggande förutsättning har varit att språkdragen ska kunna kvantifieras och analyseras automatiskt.

Än så länge är språkteknologiska analyser i stort sett begränsade till konstruktioner som bygger upp meningar. Språkliga medel för att skapa sammanhang i text (koherens och kohesion), som har en central roll inom textlingvistik, kan t.ex. knappast mätas automatiskt. Jag har också bedömt att materialunderlaget är alltför litet för att försöka identifiera olika typer av flerordsuttryck.²⁴

Olika slags ytliga textanalysmått (jfr Heimann Mühlenbock 2013), som ordlängd och meningslängd, har sedan 1960-talet använts inom den kvantitativa stilistiken och läsbarhetsforskningen (jfr Svensson och Karlsson 2012) för att komma åt lingvistiska drag och språklig komplexitet. Flera av studierna inom denna tradition har pekat mot intressanta resultat. Denna avhandlings första delstudie kan ses som en replik på denna tidigare textforskning, med start på 1960-talet. Den främsta anledningen till att inkludera dessa traditionella kvantitativa textanalysmått är att möjliggöra enkla jämförelser mellan mina resultat och andras. Valet av variabler i denna delstudie är också influerat av de parametrar som Heimann Mühlenbock (2013) finner värdefulla för bedömning av texters svårighetsgrad.

Med de efterföljande delstudierna går jag bortom tidigare kvantitativa studier, i vilka indikatorer på språkliga drag kartläggs, genom att i stället mäta olika språkdrag direkt. Dessa delstudier har konstruerats för att undersöka några viktiga aspekter på språklig uppbyggnad i texterna – ordförråd, informationspackning och strukturell komplexitet. Gemensamt för dessa delstudier är att resultatredovisningarna bygger på s.k. *indexlistor* (se vidare avsnitt 5.3 i detta kapitel).

Delstudie nummer två berör lärobokstexternas ordförråd. Denna studie är till viss del en vidareutveckling av tidigare forskning om framtagning av olika slags ordlistor, som frekvenslistorna i OrdiL-projektet och de akademiska ordlistor som utvecklats för engelska och svenska (se avsnitt 2.2.3 och 2.2.4). Att skapa förteckningar över typiska ord i textsamlingar är ett sätt att innehållsmässigt beskriva texttyper. Appliceras metoden på lärobokstexter från olika ämnen och stadier kan såväl ämnestypiska som stadietypiska ord extraheras. I den andra delstudien redovisas indexlistor över typisk ämnesspråklig vokabulär, som tagits fram enligt metoder som utnyttjar både lingvistisk och statistisk information.

²⁴Under arbetets gång har jag dock testat att automatiskt identifiera flerordsuttryck i delar av mitt läroboksmaterial (se Ribeck 2011; Ribeck och Borin 2014).

Delstudierna tre och fyra syftar till att mäta olika aspekter av språklig komplexitet. För att processa en mening måste man kognitivt greppa relationerna mellan meningens olika ord. Hur lätt eller svårt det är att upprätta denna mentala modell är svårt att bedöma med sådana textanalysmått som dominerade den tidiga läsbarhetsforskningen. Långa meningar kan t.ex. orsakas av syntaktiska konstruktioner med mycket olika svårighetsgrad.

På senare tid har emellertid mer sofistikerade textanalysmått utvecklats för att mäta meningars komplexitet direkt och därmed mer korrekt (jfr Lin 1996). De vanligaste sådana läsbarhetsmått är: genomsnittligt antal nominal- eller verbfraser per mening, genomsnittligt antal bisatser per mening och genomsnittligt parsningsdjup (Schwarm och Ostendorf 2005; Collins 2014). För att beräkna dylika värden automatiskt, med språkteknologiska metoder, behöver orden i en mening stå i beroendeförhållanden till varandra, vilket de t.ex. gör i en dependensgrammatisk analys (se Mel'čuk 1988).

Förutom att "svårt" språk ofta är strukturellt komplext och påfrestar korttidsminnet genom att anstränga vår parsingskapacitet, är det också ofta innehållspackat och kräver mentalt processande av stora propositionsmängder på kort tid (jfr Melin 2007). Dessa två aspekter på svårt språk, innehållspackning och strukturell komplexitet, undersöks i de två avslutande delstudierna.

I tidigare studier har man funnit att långa meningar i läroböcker främst orsakas av långa nominalfraser (jfr avsnitt 2.4.2.1). I delstudie nummer tre studeras just nominalfraser, i syfte att mäta informationspackning på bättre sätt än med t.ex. nominalkvot. Dessutom borde indexlistor över typisk nominalfrasuppbyggnad kunna ge en tydligare bild av komplexiteten i nominalfraser.

Den avslutande delstudien, nummer fyra, handlar om typiska meningar. I delstudien mäts strukturell komplexitet, genom att undersöka meningars maximala parsningsdjup, en strategi som baseras på antagandet att kortare dependenskedjor är lättare att processa än längre (jfr Mel'čuk 1988). Dessutom beskrivs komplexiteten med hjälp av indexlistor över typisk meningsuppbyggnad.

Dessa avslutande delstudier, av nominalfraser och meningar, är något mer explorativa till sin natur än de inledande delstudierna. Möjligheterna att studera syntaktiska konstruktioner i större material har tidigare varit begränsade. Det finns därför inte riktigt någon tradition att ansluta sig till. Av den anledningen var det inte alldeles självklart att det skulle fungera att applicera indexprincipen på flerordiga variabler. Metoden utvecklades nämligen ursprungligen för att identifiera texttypiska ord, men har i och med detta avhandlingsarbete justerats på några punkter efter arbetets teoretiska förståelse och materialunderlag (se vidare avsnitt 5.3 i detta kapitel).

5.3 Indexprincipen

I flera delstudier beskrivs ämnesspråken genom att s.k. *indexlistor* för olika språkliga variabler framställs. Detta sker genom den s.k. *indexprincipen*, som är en metod som jag har utvecklat för att automatiskt identifiera och ranka variabler i en korpus efter hur texttypiska de är. Pseudokod för algoritmen finns i appendix D.

I metodutvecklingen har jag framför allt inspirerats av korpusstilistisk forskning och av hur forskare tidigare har undersökt läroboksspråkskorpusar och akademiskt ordförråd. Från början användes principen i syfte att kunna identifiera akademiska ord, i samband med framtagningen av en svensk akademisk ordlista (se avsnitt 2.2.4). I och med detta avhandlingsarbete har jag emellertid både utökat och prövat metoden ytterligare. En viktig justering har varit att tillåta algoritmen att appliceras på flerordiga variabler, som nominalfraser och meningar.

Måttet tar hänsyn till variablernas frekvens, spridning, och nyckelvärde (vilket beräknas genom jämförelse med ett annat register). Om undersökningsregistret består av texter från flera ämnesområden garanterar även selektionsmetoden att variablerna är utbredda över samtliga dessa.

I alla steg av indexprincipen är datorns kraft oundgänglig; algoritmbaserad identifikation av ord med särskild lingvistisk och statistisk karakteristik i större textmassor är helt enkelt alltför tidskrävande för att utföras manuellt.

5.3.1 Selektionskriteriet hög frekvens med jämn spridning

Det första steget i indexprincipen går ut på att försöka identifiera variabler som är ”vanliga” i en textsamling. Men, som Savický och Hlaváčová (2002) påpekar, finns det ingen formell definition på den intuitiva betydelsen av ”vanlighet” när det kommer till att ranka sådant som ord i språket. Oftast används dock ordens absoluta eller relativa frekvens som måttstock, trots att ingen av dessa mått tar någon hänsyn till hur frekvensen är fördelad över de ingående texterna i korpusen. En enskild text som handlar om ett visst ämne får då lätt för stort genomslag i korpusen som helhet.

För att mer objektivt skatta hur vanlig en viss variabel är i en korpus, finns några olika s.k. *korrigerade* frekvensmått, som, vid sidan av ren frekvens, även tar hänsyn till spridning (eng. *dispersion*).²⁵ Det korrigerade frekvensmått som har valts till indexprincipen är den *reducerade* frekvens (RF), som beskrivs av

²⁵Ett spridningsmått kan användas för att sortera variabler på så vis att snedfördelning kommer längre ner i listan. Snedfördelade variabler ses alltså som mindre representativa för texten som helhet.

Hlaváčová (2000), och beräknas på följande vis:

Låt $f(x)$ vara frekvensen av ordet x i en korpus bestående av N ord. Dela sedan upp hela korpusen i $f(x)$ intervaller $\langle i, j \rangle$. För $n = 1, 2, \dots, f(x)$, är det n :te intervallet $\langle \frac{(n-1)N}{f(x)} + 1, \frac{nN}{f(x)} \rangle$.

Låt Fx vara den partiella frekvensen av x enligt:

$Fx(n) = 1$, om x förekommer inom intervall n ,

$Fx(n) = 0$, om x inte förekommer inom intervall n .

$RF(x)$ beräknas då enligt (56).

$$RF(x) = \sum_{n=1}^{n=f(x)} Fx(n) \quad (56)$$

RF är alltså summan av alla partiella frekvenser för x , då korpusen delats upp i lika många delar som den absoluta frekvensen för variabeln i fråga. Med hjälp av RF säkerställs att frekvensen sprids i korpusen, utan att korpusen behöver delas upp i delkorpusar utifrån t.ex. texttyper eller genrer. Detta schematiska tillvägagångssätt är därför både mer lätthanterligt och effektivt än många andra spridningsmått (Savický och Hlaváčová 2002: 216f.).

Den reducerade frekvensen för variabler med en jämn distribution över korpusen ligger närmare värdet för den absoluta frekvensen. För variabler med mer ojämnt fördelade frekvenser blir värdet för RF lägre än den absoluta frekvensen. Genom hela processen räknas dock strikt med relativa frekvenser, för att kompensera för korpusarnas varierande storlek.

När ämnesfack studeras säkerställs variabelernas utbredning (eng. *range*) över de olika ämnena genom att den reducerade frekvensen måste uppgå till ett bestämt minimumvärde inom alla ämnesområden (jfr t.ex. angreppssättet i Stubbs 2005). Dessa värden skiljer sig åt mellan delstudierna, och redovisas därför i anslutning till respektive delstudies upplägg.

5.3.2 Selektionskriteriet texttyptypiskhet

Det andra och avslutande steget i indexprincipen syftar till att urskilja texttyptypiska variabler, d.v.s. variabler som är vanligare, eller mer utbredda, i en textsamling än i en annan. Identifieringen sker genom att variabelernas (reducerade) frekvenser i det register som undersöks, undersökningsregistret, ställs mot motsvarande värden i ett annat register, jämförelseregistret.

I en sådan jämförelse kallas ord med statistiskt signifikant högre frekvens i undersökningsregistret än i jämförelseregistret för *nyckelord* (jfr "key words" i Scott 1997). Sådana kan extraheras för att peka ut vad en viss texttyp "handlar om", d.v.s. dess ämnesinnehåll och centrala innehållsliga element (Baker m.fl. 2008). Inom korpusstilistiken har nyckelordsmetoden använts för att beskriva enskilda författarstilar (jfr avsnitt 2.3.2.2) och inom engelskspråkig forskning har metoden används för att extrahera typiskt akademisk vokabulär (jfr avsnitt 2.2.3).

I indexprincipen har nyckelordskonceptet utvidgats till att omfatta inte bara ord, utan språkliga variabler generellt. Nyckelvariabelförhållandet baserat på reducerade frekvenser benämns *index*; ju högre detta värde är, desto mer texttypisk är variabeln enligt definitionen. Ett index på 2,0 betyder enkelt uttryckt att en variabel är dubbelt så spridd i undersökningsregistret som i jämförelseregistret.²⁶ För att en variabel ska ingå i den slutliga indexlistan måste dess index uppgå till minst 1,1.²⁷

Resultatet av en analys enligt indexprincipen sammanställs i en s.k. *indexlista*, i vilken variablerna listas i fallande ordning, med de mest texttyptypiska överst. Då indexprincipen identifierar variabler som är typiska för det register som undersöks jämfört med ett annat register, beskriver den resulterande indexlistan endast undersökningsregistret. Posterna på listan är alltså inte reflexiva och gällande för den omvända jämförelsen. I resultatredovisningen presenteras därför för samtliga jämförelser mellan två register två stycken indexlistor, vilka kompletterar varandra. Den första visar variabelvärden typiska för undersökningsregistret (i förhållande till jämförelseregistret), den andra variabelvärden typiska för jämförelseregistret (i förhållande till undersökningsregistret).

5.4 Delstudiernas övergripande upplägg

De lärobokstexter som ingår i undersökningarna delas upp utifrån två grundläggande dimensioner – ämne och stadium. Dessutom grupperas läroböcker i ämnena biologi, fysik och kemi under ämnesfacksbeteckningen "naturvetenskap" och läroböcker i geografi, historia och samhällskunskap under ämnes-

²⁶I sällsynta fall inträffar det att en variabel förekommer i den ena korpusen i jämförelsen men inte i den andra. Därför tillämpas ett slags utjämningsteknik genom vilken sådana frekvenser uppskattas (jfr *smoothing* i t.ex. Jurafsky och Martin 2009).

²⁷De tekniska parametrarna i den automatiska extraktionsmodellen har bestämts på heuristisk väg. I arbetet med att ta fram en ordlista över svensk akademisk vokabulär (se 2.2.4) testades kombinationer av alla krav som tidigare använts för att extrahera akademiska ord. De resulterande listorna utvärderades sedan manuellt utifrån vår intuition om ords användning och stilvärde. De slutliga kriterierna och tröskelvärdena är de med vilka vi erhöll bäst precision enligt denna kvalitativa bedömning.

facksbeteckningen ”samhällsvetenskap” (jfr *kunskapsfamiljer* i Selander 2001: 127). Ämnesfackstilldelningen grundar sig på några homogenitetstester som utfördes på en delmängd av läroboksmaterialet i ett tidigt stadium i avhandlingsarbetet. I dessa tester fann jag ingen grund för att gruppera NO-ämnena tillsammans med matematik, vilket man ibland har gjort i tidigare forskning (t.ex. i OrdiL-projektet). Mina matematiktexter platsar inte i någon övergripande ämnesfacksgruppering och behandlas därför som ett separat ämnesfack. Genom att betrakta läroböcker i ett visst ämnesfack eller ett visst ämne som olika register, blir det möjligt att jämföra dels olika ämnesfackliga språk, dels olika ämnesspråk, med varandra.

Resultaten av delstudierna redovisas i var sitt kapitel. Vidare utförs analyserna på olika sätt för mått med rent kvantitativa resultat och för de variabler på vilka indexprincipen appliceras. Också resultatredovisningen skiljer sig för dessa olika slags undersökningar.

För kvantitativa mått som t.ex. LIX och nominalfrastäthet, genomförs analyserna på följande vis: i ett första steg undersöks ämnesfacken (naturvetenskap, matematik och samhällsvetenskap). I samma steg undersöks även samtliga läroböcker under kategorin ”hela läroboksmaterialet”. I nästa steg delas naturvetenskap och samhällsvetenskap upp i sina ämnesmässiga beståndsdelar. Samtliga dessa analyser genomförs separat för högstadie- respektive gymnasietexterna.

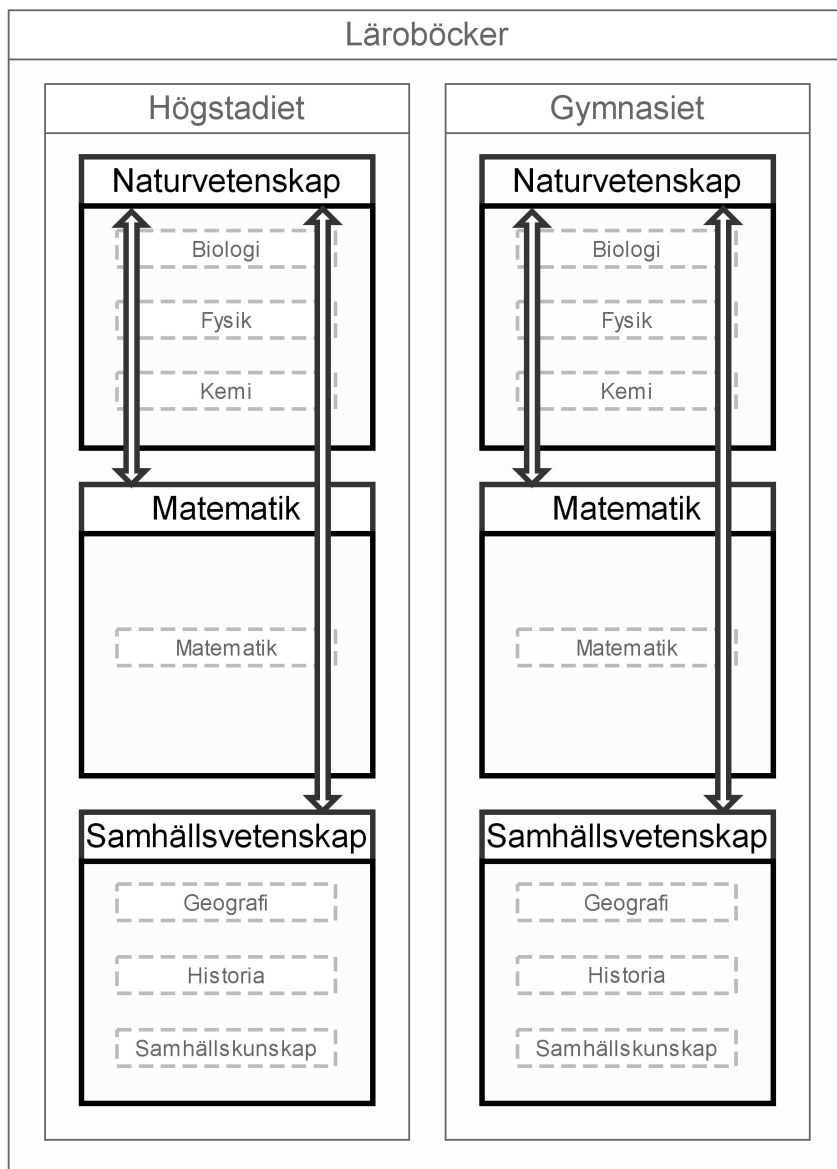
Resultaten för de kvantitativa måtten redovisas i första hand i form av linjediagram över medelvärden. För läroboksmaterialet i dessa diagram finns två diskreta mätpunkter, en för högstadietexter och en för gymnasietexter. Dessa två mätvärden binds sedan samman med linjär interpolation, för att visa på utvecklingen i stadieövergången mellan högstadiet och gymnasiet.

I linjediagrammen ingår även värden för referenskorpusarna med akademiska och berättande texter. Dessa korpusar delas inte upp i sina beståndsdelar och får därför bara ett värde var. Anledningen till att de berättande texterna inte delas upp i barnböcker, ungdomsromaner och vuxenromaner är att storlekskillnaderna mellan korpusarna har bedömts vara för stora. Närmare bestämt är barn- och ungdomskorpusarna för små för att redovisas separat (jfr avsnitt 3.3).

I vissa fall kompletteras linjediagrammen med låddiagram, för att bättre åskådliggöra fördelningen av variabelvärden. Resultaten av de kvantitativa måtten redovisas också genomgående i tabellformat i appendix, där även medianvärden och standaravvikelser anges.

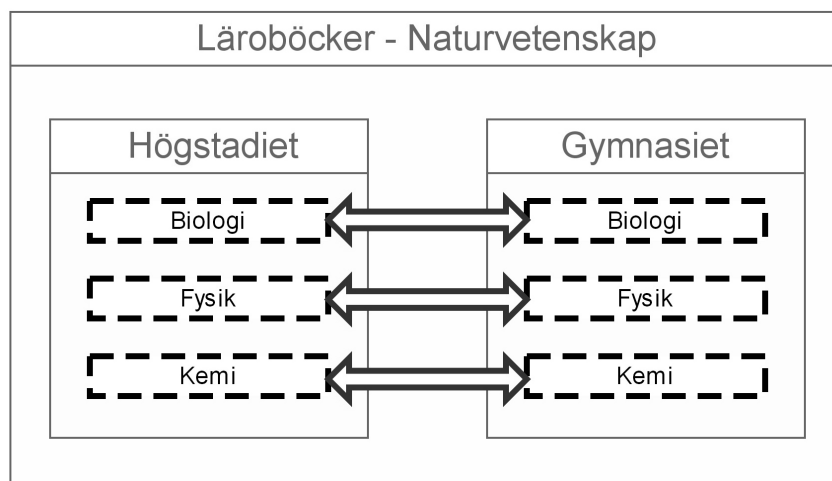
I delstudierna av ord, nominalfraser och meningar används indexprincipen för att extrahera språkliga variabler som är typiska för ämnesspråken. I denna process produceras indexlistor för samtliga för avhandlingen kritiska ämnen/ämnesfack och stadier. Analysgången för de variabler som studeras utifrån

indexprincipen är tredelad och illustreras därför av tre olika figurer, 14–16. I dessa står de dubbla pilarna för de jämförelser som görs i varje steg. Varje sådan jämförelse resulterar i två indexlistor. I det första steget (se figur 14) jämförs naturvetenskap som ämnesfack med de andra ämnesfacken, matematik och samhällsvetenskap, stadium för stadium.



Figur 14: Steg 1 i analysgången för variabler som studeras utifrån indexprincipen

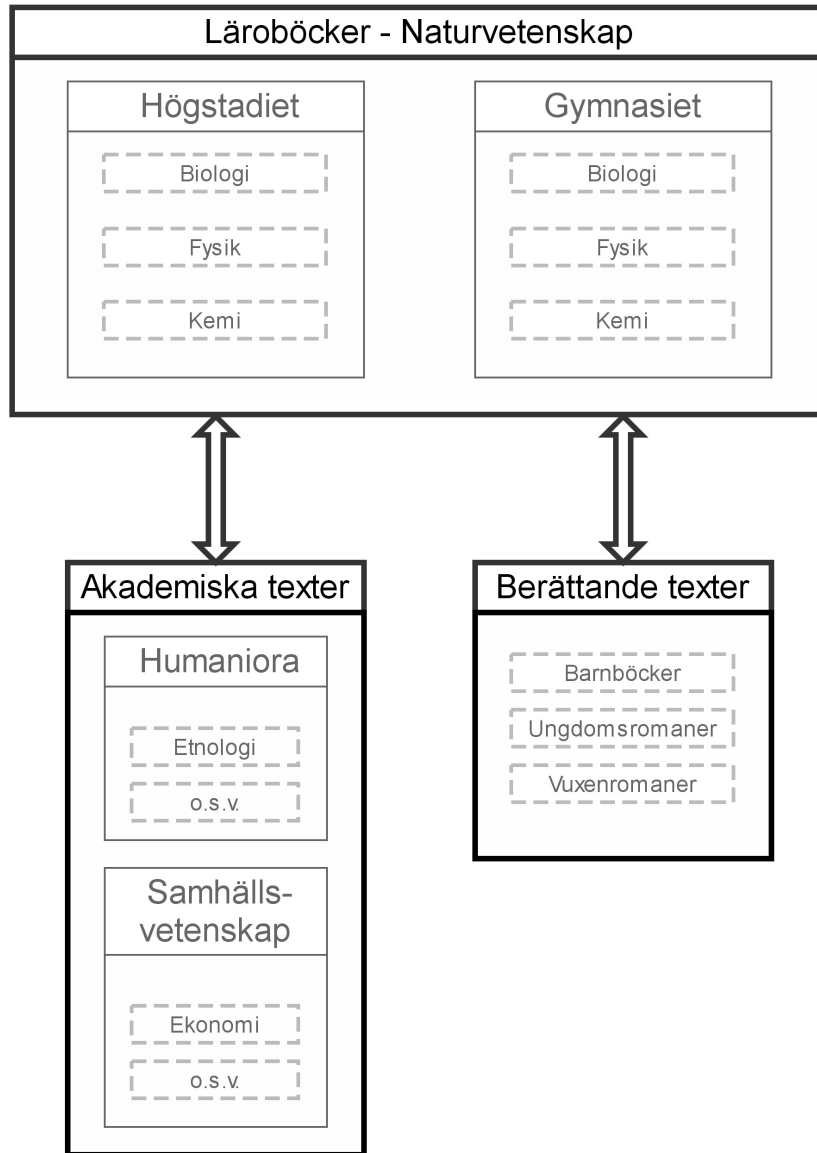
I det andra steget (se figur 15) delas naturvetenskap upp i sina ämnesmässiga beståndsdelar. Här jämförs sedan varje ämne på högstadiet med samma ämne på gymnasiet för att urskilja stadietypiska ämnesspråkliga variabler.



Figur 15: Steg 2 i analysgången för variabler som studeras utifrån indexprincipen

I det tredje och avslutande steget (se figur 16) jämförs naturvetenskap som ämnesfack med referensmaterialen. I dessa jämförelser innehåller naturvetenskapskorpusen såväl högstadie- som gymnasietexterna. Det finns flera anledningar till att jag valt att inte inkludera någon stadiespekt i detta steg. För det första är referensmaterialen, som bekant, inte nivåupplade (motsvarande högstadiet och gymnasiet). Vidare framkom i förberedande tester att mycket små skillnader märks i indexlistorna, om dessa tas fram separat för högstadiet och gymnasiet; på sin höjd byter några konstruktioner plats med varandra i stadiesövergången. Slutligen undersöks skillnaden mellan naturvetenskap på högstadiet och gymnasiet ändå tillräckligt väl i steg ett och två i analysgången.

I delstudierna framkommer ibland resultat som föranleder ytterligare eller mer djupgående undersökningar av någon aspekt. Några sådana utvecklingar har också genomförts och jag kallar dem för *fokusstudier* (se t.ex. 7.2.2). Resultaten av fokusstudierna redovisas alldeles före den sammanfattande diskussionen för delstudien i fråga.



Figur 16: Steg 3 i analysgången för variabler som studeras utifrån indexprincipen

6

DELSTUDIE: TRADITIONELLA TEXTANALYSMÅTT

Delstudien av traditionella textanalysmått består av två avsnitt. Det första behandlar följande vanliga ytliga textanalysmått: meningslängd, ordlängd, läsbarhetsindex och ordvariationsindex. Därefter följer beräkningar av några litet djupare textanalysmått, som i viss utsträckning nyttjar lingvistisk information: ordklassfördelning, nominalkvot och lemmavariationsindex.

För samtliga dessa variabler presenteras grafiska översikter över mätresultaten i det följande. I figurerna har referenslinjer för de berättande respektive akademiska texterna infogats som fasta mätpunkter att förhålla värdena för lärobokstexterna till. Dessutom återfinns alla exakta mätvärden i appendix E.

6.1 Ytliga textanalysmått: presentation

6.1.1 Meningslängd

Enligt Hellspong och Ledin (1997: 75) är en beräkning av texters genomsnittliga *meningslängd* i antal ord ett sätt att ”ta ett allmänt grepp om syntaxen” och ”överblicka meningarna”. Anledningen till att meningslängden är intressant är att långa meningar ofta medför en hög syntaktisk komplexitet, eftersom sådana oftare har en invecklad, hypotaktisk struktur med bisatser av olika underordningsgrad (Platzack 1973: 111). Men denna korrelation är inte entydig. Exempelvis hävdar Hellspong och Ledin (1997) att långa meningar i texter med nominal stil beror på långa nominalfraser och samordningar, medan långa meningar i texter med verbal stil i stället orsakas av långa prepositionsadverbial och många underordnade satser. Också antagandet att korta satser skulle vara mer lättlästa har ifrågasätts, då studier har visat att korta texter och satser ofta utgör hinder för läsförståelsen och är symptomatiska för bristande logisk uppbyggnad av innehållet (jfr Hvenekilde 1993; Reichenberg och Lundberg 2011).

Några exempel på läroboksstudier som innehåller beräkningar av meningslängd är: Westman 1974; Danielsson 1975; Melin 1995; Sandqvist 1995; Ekvall 1997.

6.1.2 Ordlängd

Ordlängd är också ett mått som kan användas för att uppskatta texter svårighetsgrad. Texter med fler längre ord anses vara mer lexikalt komplexa att processa. Detta har att göra med att långa ord, som i svenskan ofta bildas genom sammansättning eller avledning, ofta är ovanliga, svåra och innehållsrika, medan korta ord ofta är vanliga, enkla och mer allmänna (Hultman och Westman 1977).

6.1.3 Läsbarhetsindex, LIX

Läsbarhetsindex, LIX, är ett mått för att snabbt uppskatta hur lättläst en text är. Måttet skapades av Björnsson (1968) och används fortfarande flitigt (Lagerholm 2008), även om det ofta har blivit kritiserat under årens lopp (se t.ex. Lundberg och Reichenberg 2008).²⁸

LIX beräknas enligt (57), där ”långa textord” innehåller fler än sex bokstäver.²⁹ Värdet på denna gräns beror på att frekvensen av ord med sju eller fler bokstäver har visat ett direkt samband med stigande upplevd svårighetsgrad hos texten. För ord med sex bokstäver eller mindre är tendensen den motsatta; ju fler korta textord, desto enklare upplevs texten (Thelander 1970: 31).

$$LIX = \frac{\text{textord}}{\text{meningar}} + 100 \times \frac{\text{långa textord}}{\text{textord}} \quad (57)$$

LIX har ofta använts i undersökningar av lärobokstexter (se t.ex. Danielsson 1975 och Melin 1995). Björnsson (1968) rapporterar att LIX för läroböcker för årskurs två är 22, för årskurs fem 31 och för årskurs åtta 37. Härutöver anges barn- och ungdomsböcker ha ett LIX på 27, skönlitteratur för vuxna 33,

²⁸Några moderna exempel på avhandlingar som metodologiskt baseras på LIX är Partanen 2006 och Kuisma 2013. Härutöver finns en uppsjö av studentuppsatser som använder måttet för att bedöma texters svårighetsgrad (se t.ex. Bergstrand Grenner och Ekblad 2011; Kiland 2012; Fagerström och Karlsson 2014).

²⁹På webbsidan <www.lix.se> kan användare själva enkelt beräkna LIX i sina texter.

saklitteratur 47 och facklitteratur 56. Melander (2003) nämner också att andra läroboksstudier som beräknat LIX har uppmätt värden mellan 41 och 45 för högstadieböcker och mellan 47 och 50 för gymnasieböcker.

6.1.4 Ordvariationsindex, OVIX

Ordvariationsindex, OVIX, är ett mått på hur varierat ordförrådet i texter är.³⁰ Måttet är utvecklat av Hultman och Westman (1977) och beräknas enligt (58):

$$OVIX = \frac{\log(\text{textord})}{\log\left(2 - \frac{\log(\text{textordstyper})}{\log(\text{textord})}\right)} \quad (58)$$

Melin (1995) och Sandqvist (1995) använder OVIX-måttet i sina lärobokundersökningar. Ett högt OVIX-värde har tidigare visat sig korrelera med dels goda elevtexter, dels höga betyg (Hultman och Westman 1977; Johansson Kokkinakis och Magnusson 2009; Nordenfors 2011).

6.2 Djupare textanalysmått: presentation

6.2.1 Ordklassfördelning

Eftersom olika ordklasser har olika informationsvärde är en beräkning av *ordklassfördelning* en del av karakteristiken för texter. Exempelvis har nominala ordklasser, som substantiv, particip och prepositioner, ett högre informationsvärde än verbala ordklasser, som verb, adverb och pronomen (Melin och Lange 2000).

Westman (1974) redovisar ordklassfördelningen i sina läroböcker för gymnasiet.

³⁰Att mäta ordvariation i texter är inte helt enkelt, bl.a. eftersom sannolikheten att nya ord introduceras minskar med textlängden. OVIX har valts som mått på ordvariation eftersom det är relativt oberoende av textlängd jämfört med andra populära mått, som t.ex. *type-token ratio*, TTR, och *vocational density*, VocD (jfr Malvern m.fl. 2004).

6.2.2 Nominalkvot, NQ

Nominalkvot, NQ, är ett etablerat mått på innehållspackning i text (Einarsson 1978) och beräknas enligt (59):

$$NQ = \frac{\text{substantiv} + \text{prepositioner} + \text{particip}}{\text{pronomen} + \text{adverb} + \text{verb}} \quad (59)$$

Sandqvist (1995) beräknar nominalkvot i sin diakrona studie av läroböcker i historia. Melin och Lange (2000: 48) skriver att "normalvärdet" för NQ är 1,0, och att läroböcker för högstadiet befinner sig på denna nivå.

6.2.3 Lemmavariationsindex, LVIX

Lemmavariationsindex, LVIX, beräknas för lemman enligt samma formel som OVIX beräknas för textord. Enligt Heimann Mühlenbock (2013: 36) är LVIX ett bättre mått på lexikal variation än OVIX, då det tar hänsyn till det släktskap som råder mellan regelbundna böjningsformer av ord i språket. I OVIX-beräkningen räknas frekvensen av alla ordformer i texten vilket gör att t.ex. *cirkel* och *cirkeln* betraktas som olika typer. Med hjälp av lemmatisering, d.v.s. att för varje ordform ange vilken grundform denna tillhör, kan man i stället beräkna frekvensen av olika grundformer. På så vis räknas här omtalade ord som instanser av samma ordtyp. Ordförrådets lemmabaserade struktur är en viktig lexikal aspekt att ta hänsyn till när man analyserar autentiskt textmaterial (Svensén 2004: 65f.).

6.3 Ytliga textanalysmått: resultat i förhållande till tidigare forskning

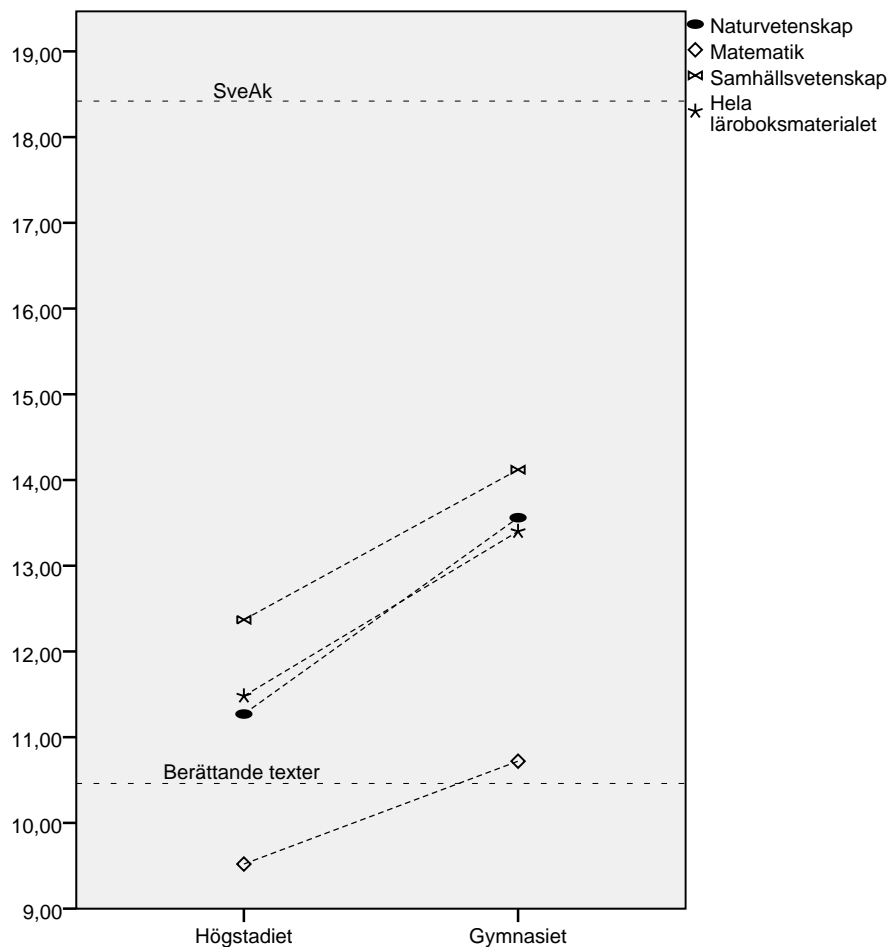
Mätvärden för samtliga ytliga mätningar finns i appendix E.1.1.

6.3.1 Meningslängd

I denna undersökning av meningslängder i hela läroboksmaterialet, räknas antal *ord* i meningarna (vilket exkluderar skiljetecken). De detaljerade mätvärdena för meningslängd återfinns i tabell E.1 i appendix.

Figur 17 visar till att börja med en översiktlig utveckling av den genomsnittliga meningslängden över stadierna i hela läroboksmaterialet och i ämnesfacken. I figuren syns tydligt att meningslängden ökar i samtliga ämnesfack mellan högstadiet och gymnasiet. Ökningen sker kraftigast i naturvetenskap.

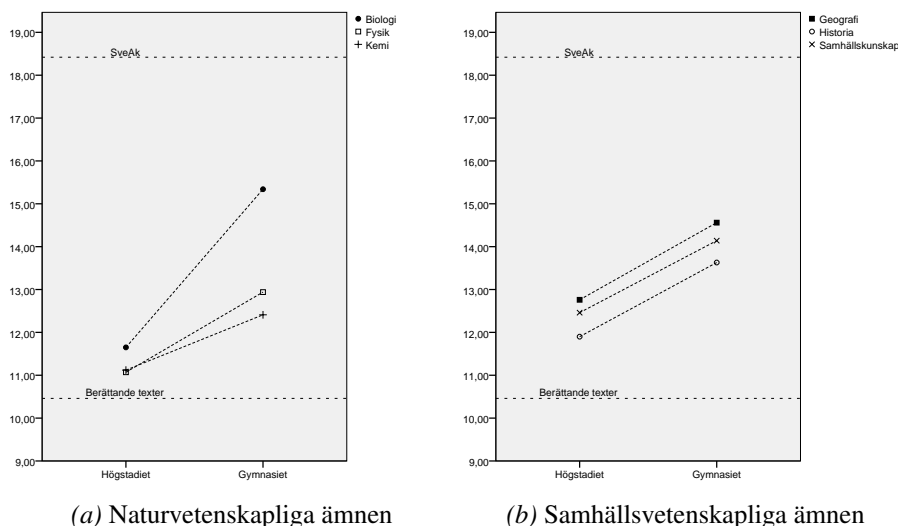
Meningarna i samhällsvetenskapliga texter är generellt sett längre än i övriga ämnesfack, både på högstadiet och på gymnasiet, medan meningarna i de naturvetenskapliga texterna närmare följer utvecklingsgången i hela läroboksmaterialet. Matematiktexterna utmärker sig med betydligt lägre värden över stadierna; på högstadiet är meningslängden i de matematiska texterna till och med kortare än i de berättande texterna.



Figur 17: Översiktlig utveckling av genomsnittlig meningslängd över stadierna

Figur 18 visar vidare den ämnesspecifika utvecklingen av den genomsnittliga meningslängden över stadierna i naturvetenskap och samhällsvetenskap. Betraktar man de ämnesspecifika kurvorna noteras i samtliga fall en ökning av meningslängden i stadiövergången. Det finns dock en tydlig skillnad mellan de naturvetenskapliga och samhällsvetenskapliga ämnena. I naturveten-

skap ligger värdena för högstadiet mycket närmare varandra än värdena för gymnasiet; meningslängderna i fysik och kemi är till och med statistiskt sett likvärdiga på högstadiet ($p = 0,299$).³¹



Figur 18: Ämnesspecifik utveckling av genomsnittlig meningslängd över stadierna

I biologitexterna sker en förhållandevis dramatisk ökning av meningslängden i stadiövergången, vilket medför en mer splittrad bild av ämnesfacket naturvetenskap på gymnasiet. I alla de samhällsvetenskapliga ämnena är ökningen av meningslängd i stort sett lika omfattande, vilket gör att ämnesfacket är ungefär lika homogent på gymnasiet som på högstadiet.

Den genomsnittliga meningslängden för texterna i mitt moderna läroboksmaterial är något lägre än värdena som rapporteras av Westman (1974) och Danielsson (1975). Exempelvis ligger mitt medelvärde för högstadietexter i fysik och samhällskunskap på 11,7 respektive 12,5 ord, medan Danielssons motsvarande värden är 12,5 respektive 12,9 ord, och mina gymnasievärden för dessa ämnen är 12,9 respektive 14,1 ord, och Danielssons 15,2 respektive 15,0 ord. Den genomsnittliga meningslängden för hela mitt gymnasiematerial uppgår till 13,4 ord, vilket kan jämföras med meningslängden i Westmans gymnasietexter, som är 14,5 ord.

³¹Värdet för p , som står för *probability*, beräknas genom ett t-test och anger sannolikheten för att skillnaden mellan två medelvärden har orsakats av slumpen. Jag följer traditionen och använder 0,05 som övre gräns för 'statistisk signifikans', och 0,001 som övre gräns för 'stark statistisk signifikans'.

Mina värden för historietexterna kan jämföras med resultaten för historieläroböcker från 1950- respektive 1980-talet (Sandqvist 1995). För högstadieläroböcker ser utvecklingen ut såhär: 1950 – 12,1; 1980 – 11,6; nutid – 11,9 ord per mening. För gymnasieläroböcker är värdena: 1950 – 15,3; 1980 – 12,7; nutid – 13,6 ord per mening.

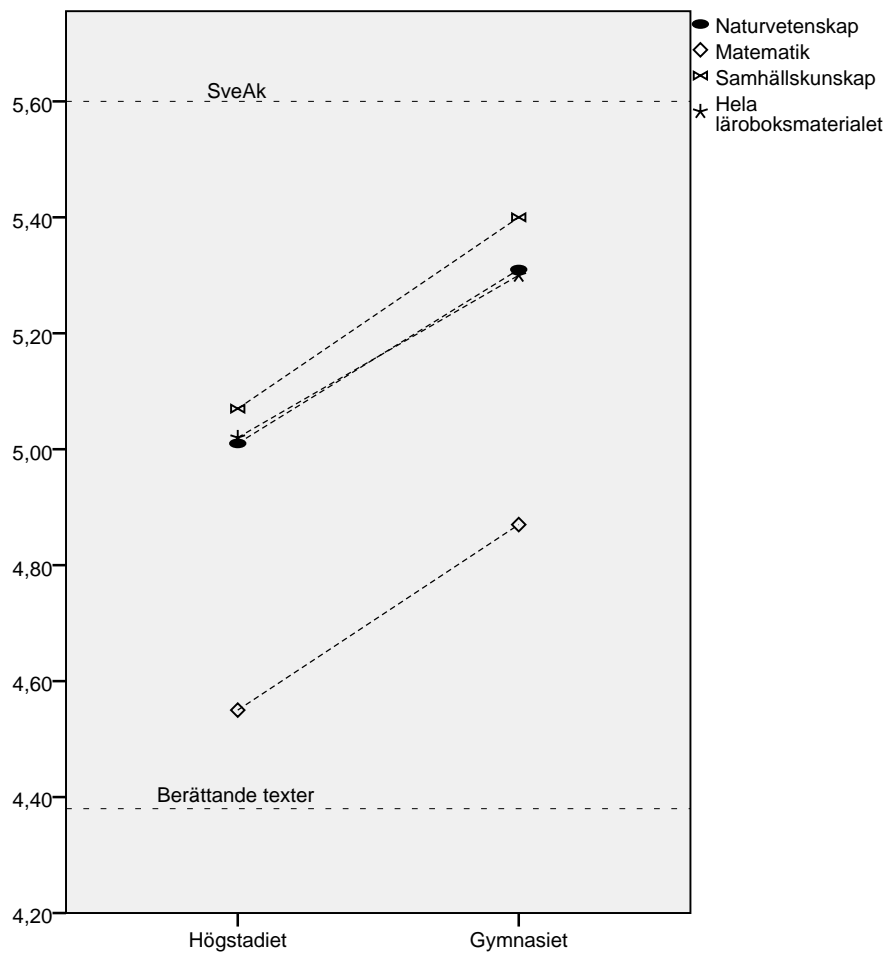
6.3.2 Ordlängd

Samtliga mätvärden för ordlängd listas i tabell E.2 i appendix.

Översikten av den genomsnittliga ordlängdsutvecklingen i hela läroboksmaterialet och i ämnesfacken, som visas i figur 19, påminner mycket om bilden av de olika ämnesfackens meningslängder över stadierna. I samtliga ämnesfack ökar ordlängden i övergången till gymnasiet. Samhällsvetenskap befinner sig hela tiden högst på skalan och naturvetenskap följer utvecklingskurvan för hela läroboksmaterialet. Värdena för matematik är betydligt lägre, men ämnet ligger redan på högstadiet över värdet för de berättande texterna.

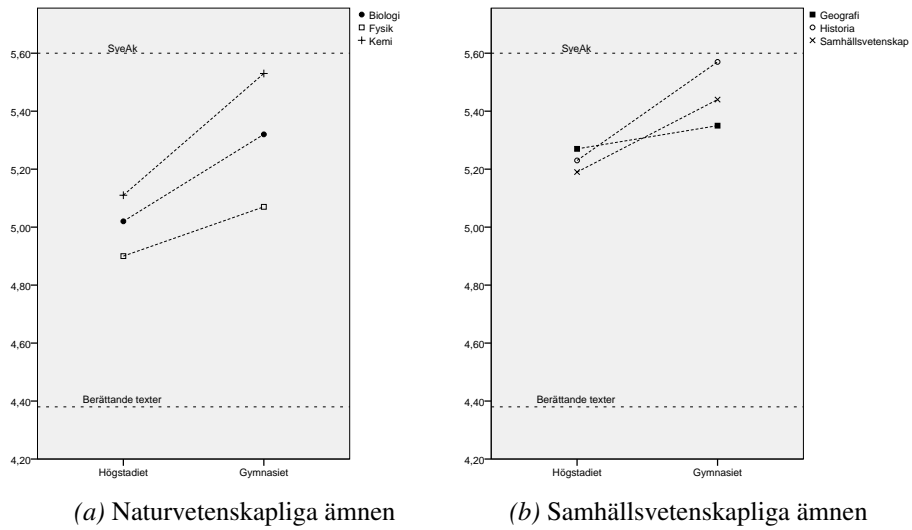
Enligt Hultman och Westman (1977: 77) är den genomsnittliga ordlängden i gymnasieläroböcker 5,59 bokstäver. I mitt material är den genomsnittliga ordlängden i gymnasietexter något kortare – 5,30 bokstäver.

Figur 20 visar den ämnesspecifika utvecklingen av den genomsnittliga ordlängden över stadierna i naturvetenskap och samhällsvetenskap. Som synes ökar ordlängden mellan högstadiet och gymnasiet i samtliga ämnen. I de naturvetenskapliga ämnena sker ökningen i relativt olika omfattning. Mest dramatisk är ökningen i kemi, som på gymnasiet i det närmaste når ordlängden i de akademiska texterna. Även historia når på gymnasiet nästan upp till ordlängdsnivån i de akademiska texterna. Bland de samhällsvetenskapliga ämnena märks även geografi, vars ordlängdsökning är mest modest i stadiövergången.



Figur 19: Översiktlig utveckling av genomsnittlig ordlängd över stadierna

6.3 Ytliga textanalysmått: resultat 93

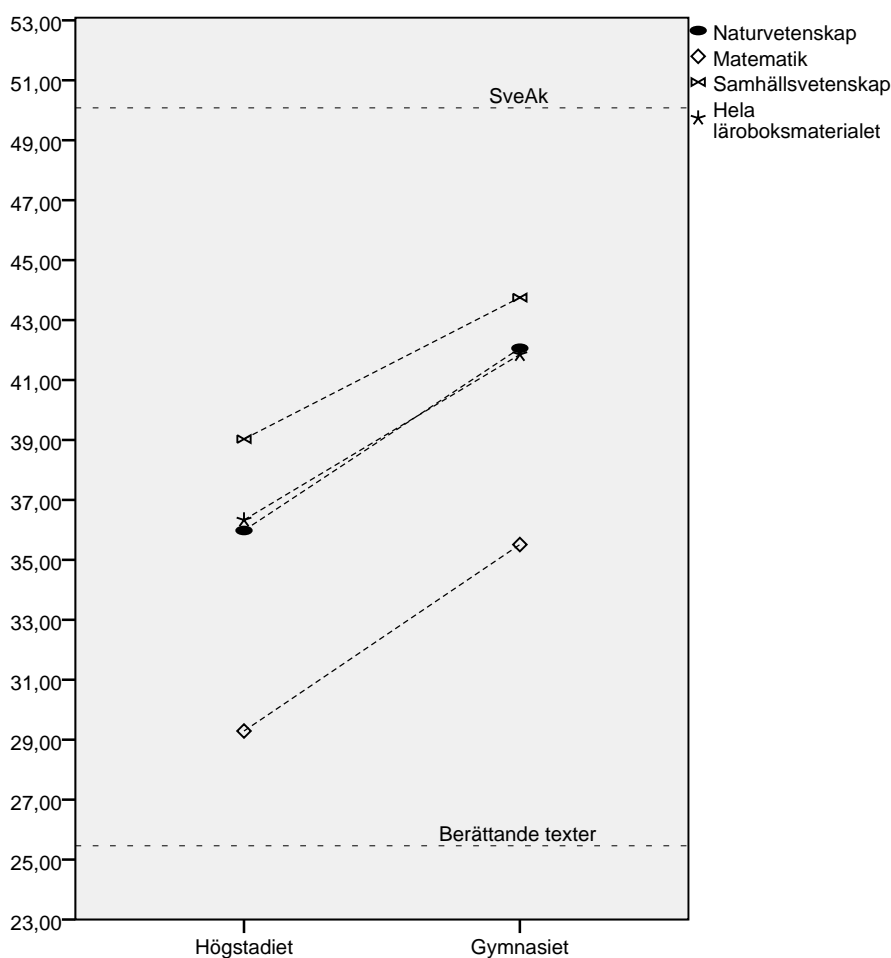


Figur 20: Ämnesspecifik utveckling av genomsnittlig ordlängd över stadierna

6.3.3 Läsbarhetsindex, LIX

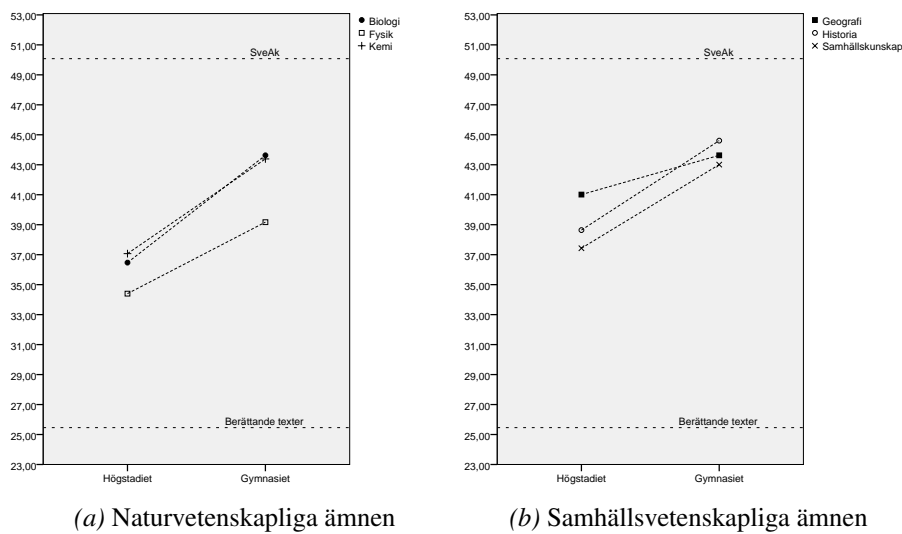
Mätvärdena för LIX återfinns i tabell E.3 i appendix.

Figur 21 illustrerar hur LIX utvecklas över stadierna i hela läroboksmaterialet och i ämnesfacken. Bilden är snarlik både den för meningslängd och den för ordlängd (vilket såklart beror på de parametrar som ingår i beräkningen av LIX). På högstadiet är medelvärdet för hela läroboksmaterialet 36, vilket enligt Melin och Lange (2000) ligger precis mellan värdet för ”lätt” och ”medelsvår” text. På gymnasiet har LIX ökat till 42, vilket visserligen är klart över värdet för ”medelsvår” text, men ändå ligger närmare ”medelsvår” än ”svår” text.



Figur 21: Översiktlig utveckling av LIX över stadierna

Den ämnesspecifika utvecklingen av LIX i naturvetenskapliga och samhällsvetenskapliga ämnen framgår av figur 22. När ämnena delas upp märks ännu tydligare att LIX är beroende av både meningslängd och ordlängd. För de naturvetenskapliga ämnestexterna resulterar detta kombinerade mått på läsbarhet i att biologi och kemi knappt kan särskiljas, varken på högstadiet eller på gymnasiet. Sedan tidigare vet vi dock att den sammanlagda komplexitetsökningen i biologi främst beror på längre meningar på gymnasiet, medan den i kemi framför allt grundar sig i längre ord.



Figur 22: Ämnesspecifik utveckling av LIX över stadierna

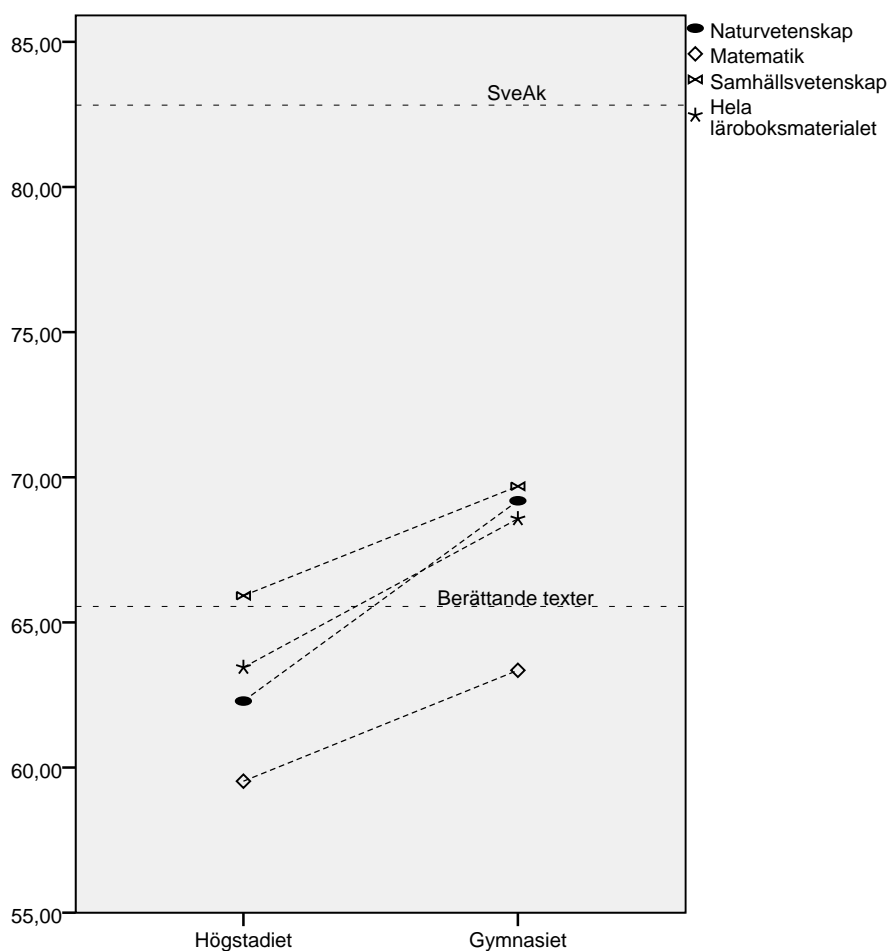
LIX för de enskilda ämnena ligger förvånansvärt tätt intill varandra på gymnasiet – mellan 43 och 44 i samtliga ämnen utom i fysik och matematik, som har betydligt lägre värden.

På 1970-talet beräknade Danielsson (1975) LIX för lärobokstexter i fysik och samhällskunskap. I förhållande till hennes uppmätta LIX-värden är mina genomgående lägre. LIX för fysiktexterna för högstadiet och gymnasiet i mitt material är 34,4 respektive 39,2, och Danielssons 39,3 respektive 45,0. Värdena för mina samhällskunskapstexter är 37,4 respektive 43,0 och Danielssons 39,5 respektive 50,5.

6.3.4 Ordvariationsindex, OVIX

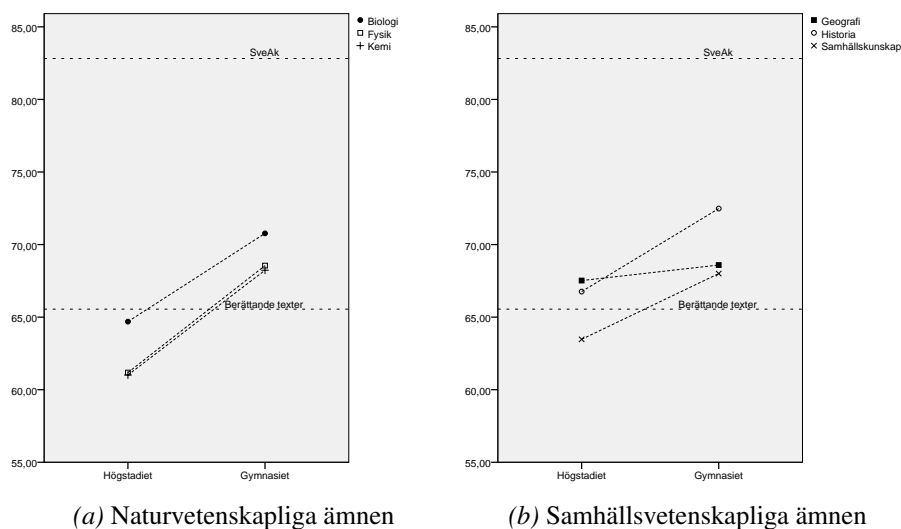
Exakta mätvärden för OVIX återfinns i tabell E.4 i appendix.

I figur 23, som översiktligt visar hur OVIX utvecklas över stadierna i hela läroboksmaterialet och i ämnesfacken, upprepas än en gång mönstren från de tidigare ytliga undersökningarna. I alla ämnesfack ökar OVIX mellan högstadiet och gymnasiet, och ökningen är störst i naturvetenskap. Samhällsvetenskap ligger högst och matematik lägst.



Figur 23: Översiktlig utveckling av OVIX över stadierna

Den ämnesspecifika utvecklingen av OVIX över stadierna i naturvetenskap och samhällsvetenskap tecknas i figur 24. Bland de naturvetenskapliga ämnena följer fysik och kemi varandra mycket tätt genom stadiövergången, där värdet för de berättande texterna passeras. Biologi befinner sig redan på högstadiet högre upp på skalan, och behåller ett visst övertag även på gymnasiet.



Figur 24: Ämnesspecifik utveckling av OVIX över stadierna

De samhällsvetenskapliga ämnena betar sig mer olika sinsemellan. I geografi ökar OVIX endast marginellt i stadiövergången, medan ökningen är betydligt mer markant i samhällskunskap och, framför allt, i historia. Historia är det ämne som når högst OVIX-värde på gymnasiet – 72,5.

För Westmans gymnasietexter anges ett medianvärde för OVIX på 71,0 (Hultman och Westman 1977: 56) och ett medelvärde på 78,6 (Melin och Lange 2000: 167). Det genomsnittliga värdet på OVIX i mitt material för gymnasiet ligger på 68,6, vilket alltså är lågt i jämförelse med 1970-talets lärobokstexter.

Man kan också jämföra OVIX-värden för historieläroböcker från 1950- respektive 1980-talet (Sandqvist 1995) och OVIX-värden för mina historietexter. Följande värden har uppmätts för högstadieböcker: 1950 – 78,6; 1980 – 73,3; nutid – 66,8. För gymnasieböcker ser utvecklingen ut såhär: 1950 – 85,5; 1980 – 76,1; nutid – 72,5. Tydligen har ordvariationen i historieböcker minskat under perioden.

6.4 Djupare textanalysmått: resultat i förhållande till tidigare forskning

Mätvärden för de djupare mätningarna presenteras i appendix E.1.2.

6.4.1 Ordklassfördelning

Ordklassfördelningen i hela läroboksmaterialet samt i referensmaterialen redovisas i appendix E.1.2, där först värdena för högstadietexterna presenteras, därefter värdena för gymnasietexterna, och slutligen värdena för referentexterna.

Sett till ordklassfördelningen utmärker sig de naturvetenskapliga texterna, både på högstadiet och på gymnasiet, med en något större andel subjunktioner än i texterna från övriga ämnesfack. På högstadiet ligger även naturvetenskap i topp när det gäller andel substantiv och verb. Mest säregen karakteristik visar dock matematiktexterna över stadierna. Både på högstadiet och på gymnasiet förekommer ordklasserna bestämningsord och grundtal i större utsträckning i matematik. Matematiktexterna innehåller också hela tiden betydligt fler stora skiljetecken, vilket kan förbindas med den relativt korta meningslängden i ämnet. Över stadierna märks även tydligt att bruket av konjunktioner och particip inte är lika omfattande i matematik som i övriga ämnesfack. Samhällsvetenskapliga texter innehåller genomgående en större andel adjektiv.

Bland de naturvetenskapliga ämnena framkommer att fysik, liksom matematik, mer utnyttjar bestämningsord och grundtal, och mindre konjunktioner och particip. Fysik innehåller högst andel pronomen av de naturvetenskapliga ämnena, vilket märks tydligast på högstadiet. Biologi utmärker sig med störst andel prepositioner, såväl på högstadiet som på gymnasiet. På gymnasiet förekommer relativt många adverb och verb i biologitexterna. I kemitexterna märks tydligast att relativt många substantiv används på gymnasiet.

I samhällsvetenskap framkommer att geografi är det ämne där adjektiv och substantiv används flitigast. På högstadiet förekommer dessutom betydligt fler prepositioner i geografi än i övriga samhällsvetenskapliga ämnen. I samhällskunskap används i stället relativt få prepositioner över stadierna. Däremot innehåller samhällskunskapstexterna flest pronomen, subjunktioner och verb, över hela skolgången.

I de båda referensmaterialen, berättande och akademiska texter, distribueras flera viktiga ordklasser mycket olika. De berättande texterna utgörs av fler adverb, pronomen och verb, medan de akademiska texterna innehåller betydligt fler adjektiv och substantiv.

I Westman 1974 redovisas några olika ordklassers relativa bidrag till det totala antalet ord i lärobokstexterna (för gymnasiet). I tabell 6.1 visas ordklassfördelningen i hela mitt läroboksmaterial jämte motsvarande värden för

Westmans lärobokstexter.³² Jämför man resultaten slås man av att de nästan är identiska för ordklasserna konjunktioner, prepositioner och substantiv. Men även värdena för adjektiv, particip och subjunktioner är relativt lika. De största skillnaderna märks i andel adverb och verb. Gymnasieläroböckerna från 1970-talet innehåller en högre andel adverb, medan dagens texter består av en högre andel verb.

	Hela läroboksmaterialet	Westman 1974
adjektiv	6,44	7,05
adverb	5,39	7,12
konjunktion	4,14	4,18
particip	1,23	1,01
preposition	10,89	10,81
subjunktion	1,49	2,22
substantiv	25,16	25,14
verb	15,44	13,41

Tabell 6.1: Ordklassfördelning (i procent av totalt antal token) i gymnasieläroböcker i denna undersökning och i Westman 1974

6.4.2 Nominalkvot, NQ

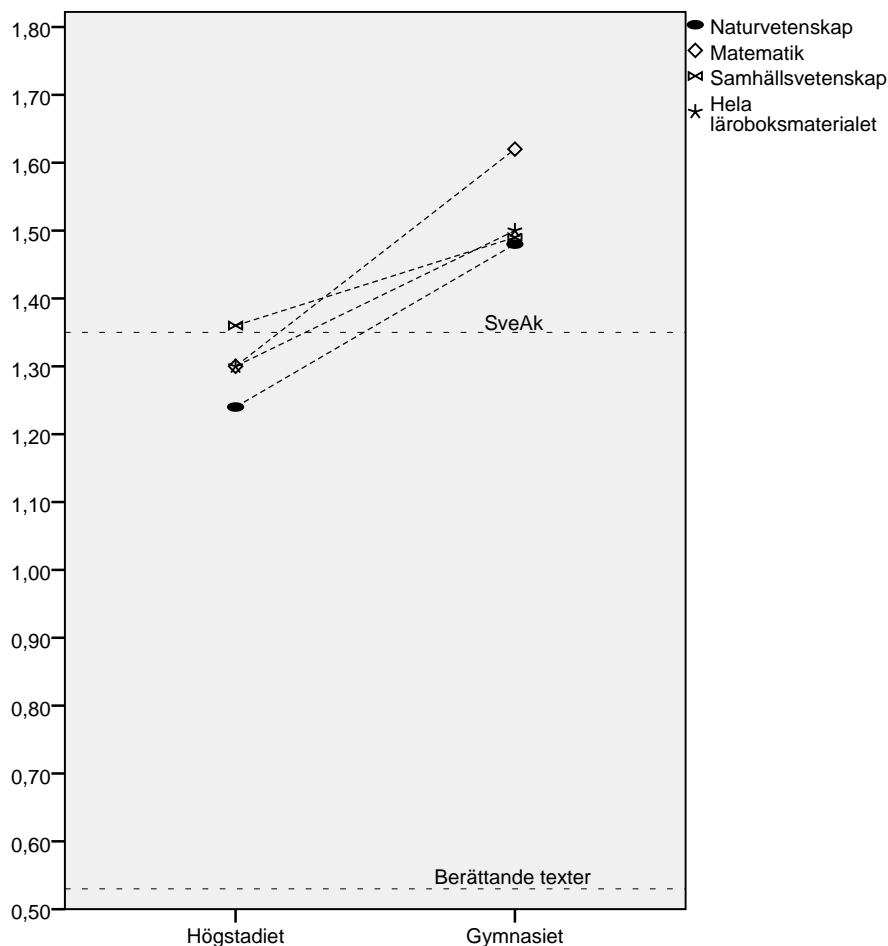
Samtliga mätvärden för NQ kan ses i tabell E.12 i appendix.

Lärobokstexterna för högstadiet har i genomsnitt en NQ på 1,30 och texterna för gymnasiet en NQ på 1,50. Båda dessa resultat är mycket högre än de värden för läroböcker som anges av Melin och Lange (2000), nämligen 1,0 för högstadiet och 1,18 för gymnasiet.

Utvecklingen av NQ över stadierna i hela läroboksmaterialet och i ämnesfacken åskådliggörs i figur 25. I denna känns inte riktigt mönstren igen från tidigare undersökningar. Detta beror främst på att matematiktexterna uppvisar den kraftigaste ökningen av NQ i stadiövergången, vilket resulterar i att matematik ligger i topp bland ämnesfacken på gymnasiet. Förhållandet mellan samhällsvetenskap och naturvetenskap är detsamma som i de ytliga mätningarna, men eftersom matematik ligger på en mycket högre nivå hamnar naturvetenskap som ämnesfack längst ner på skalan, såväl på högstadiet som på gymna-

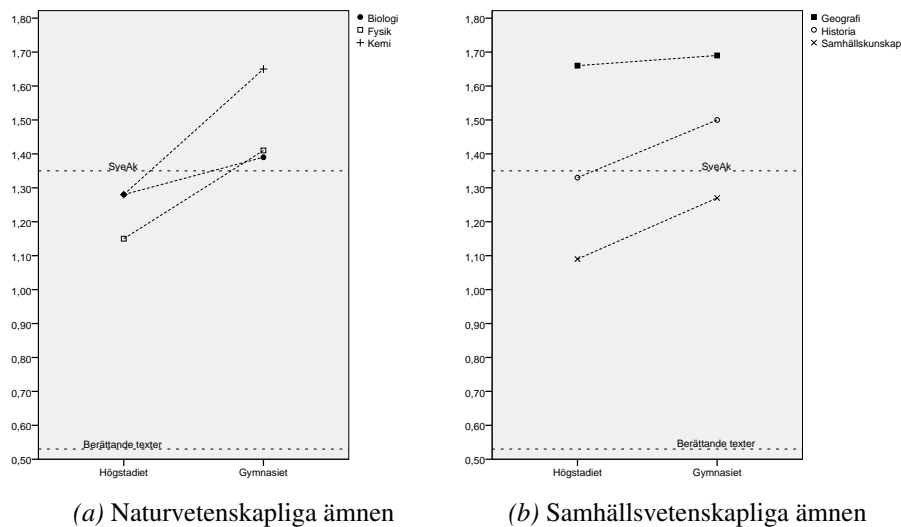
³²Tabellen innehåller endast de ordklasser för vilka Westman anger värden och våra ordklassdefinitioner stämmer överens. För jämförbarhets skull har Westmans värden räknats om till procentandel av antal token.

siet. Ytterligare en väsentlig skillnad mot de ytliga analyserna är att samtliga ämnesfack på högstadiet ligger nära värdet för de akademiska texterna, och faktiskt passerar detta i stadiövergången till gymnasiet.



Figur 25: Översiktlig utveckling av NQ över stadierna

Granskar man vidare utvecklingen av NQ i de olika naturvetenskapliga och samhällsvetenskapliga ämnena (se figur 26), kan man notera att kemi särskiljer sig från övriga naturvetenskapliga ämnen genom en relativt drastisk ökning av NQ. På gymnasiet är NQ i kemi betydligt mycket högre än i de akademiska texterna. Biologi, som på högstadiet ligger i nivå med kemi, har den lindrigast stigande utvecklingskurvan och hamnar, för en gångs skull, längst ner på gymnasiet.



Figur 26: Ämnesspecifik utveckling av NQ över stadierna

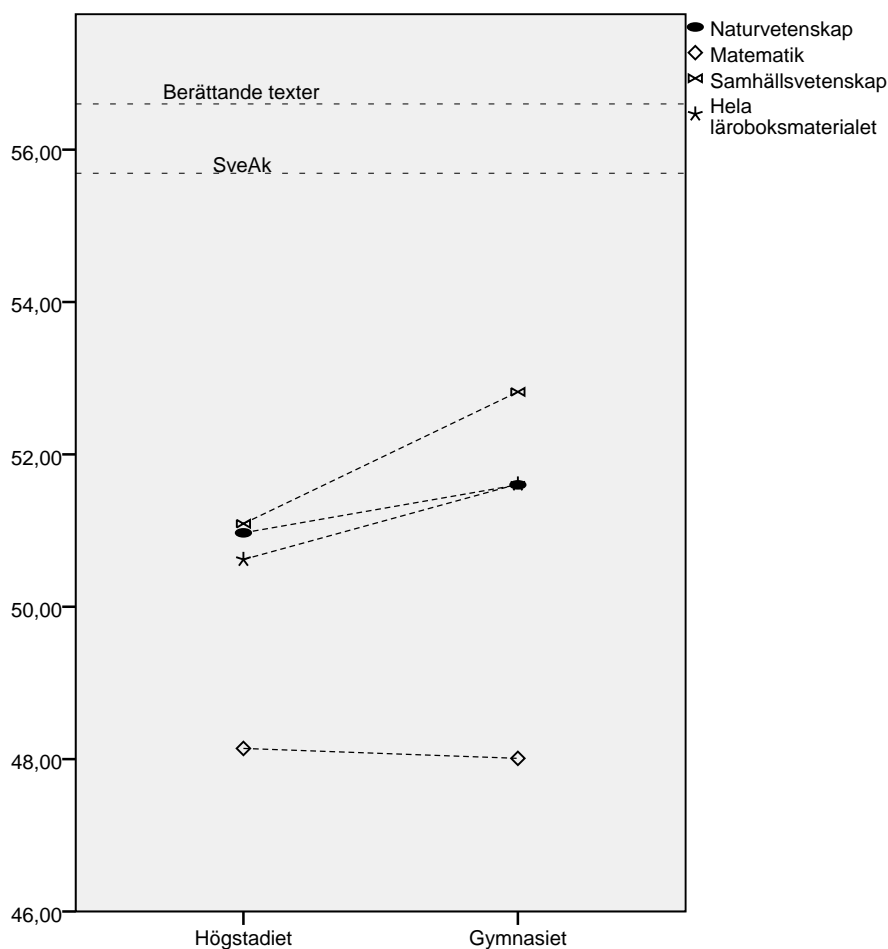
Utvecklingen av NQ i de olika samhällsvetenskapliga ämnena är totalt disparat; genom hela skolgången har de olika ämnena individuella NQ-värden. Högst är dessa i geografi, som befinner sig långt över de akademiska texterna redan på högstadiet.

Mina NQ-värden för historietexter kan jämföras med NQ-värden för historieläroböcker från 1950- respektive 1980-talet (Sandqvist 1995). För högsta-dietexter har följande värden noterats: 1950 – 1,20; 1980 – 1,22; nutid – 1,33. För gymnasietexter är värdena: 1950 – 1,59; 1980 – 1,36; nutid – 1,50.

6.4.3 Lemmavariationsindex, LVIX

Detaljerade mätvärden för LVIX presenteras i tabell E.13 i appendix.

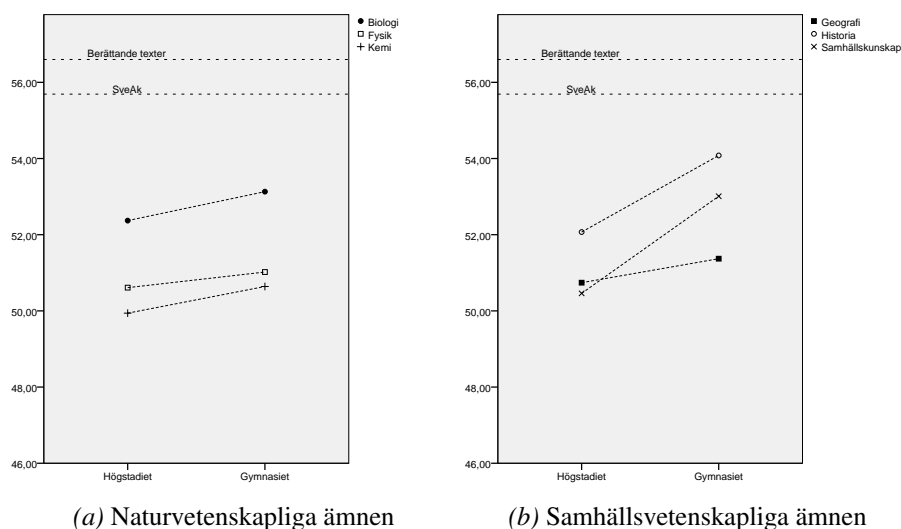
Figur 27 visar LVIX över stadierna i hela läroboksmaterialet och i ämnes-facken. Av denna figur framgår att LVIX-värdet för berättande texter är betydligt högre än värdena för lärobokstexterna, men även något högre än värdet för de akademiska texterna. Detta resultat kan tyckas oväntat, men kan förklaras med att LVIX i de olika delmaterialen i korpusen med berättande skiljer sig kraftigt åt mellan, å ena sidan, skönlitteraturen för barn- och ungdomar, med värden på 54,0 respektive 54,6, och, å andra sidan, romanerna för vuxna, med ett värde på 61,2. Mina värden kan jämföras med Heimann Mühlenbocks värden för barnbokstexter på 55,5 och vuxenromaner på 62,3 (Heimann Mühlenbock 2013).



Figur 27: Översiktlig utveckling av LVIX över stadierna

Ett annat oväntat resultat är att LVIX-utvecklingen i matematik är sjunkande över stadierna.

Den ämnesspecifika utvecklingen av LVIX över stadierna i naturvetenskap och samhällsvetenskap framgår av figur 28. De flesta ämnen beter sig ungefär på samma vis i LVIX- och OVIX-mätningarna. Undantagen är fysik och geografi. I förhållande till OVIX-kurvorna märks att LVIX-kurvan för fysik har dalat på liknande vis som den i matematik, även om trenden i fysik alltså är uppåtående. Vad gäller de samhällsvetenskapliga ämnena ligger geografi-kurvan för LVIX på en betydligt högre nivå i förhållande till övriga ämnen än motsvarande kurva för OVIX. Lutningen är dock i stort sett likvärdig.



Figur 28: Ämnesspecifik utveckling av LVIX över stadierna

6.5 Traditionella textanalysmått: sammanfattning och diskussion

6.5.1 Materialkontroll

Beräkningarna av de mått som redovisas i detta kapitel föregicks av separata undersökningar av högstadieläroböckerna från början av 2000-talet och de från början av 2010-talet, för att kontrollera att materialet verkligen hålls samman över perioden, särskilt då andra läroplaner var rådande i början av 2000-talet. För ordlängd, OVIX, NQ och LVIX är mätvärdena i stort sett identiska i de äldre och nyare böckerna, och för meningslängden är skillnaden mycket liten. Den språkliga utformningen har följaktligen bedömts tillräckligt konstant över tid för att läroböckerna ska kunna betraktas som ett synkront register.

Vidare har jag taggat en delmängd av Westmans material med mina metoder, för att säkerställa att mina resultat, som har beräknats maskinellt, är jämförbara med de siffror som rapporteras i 1970-talsstudierna, som beräknades

manuellt.³³ De uppmätta resultaten (se tabell 6.2) visar att de språkteknologiska metoderna inte påverkar resultatet av mätningarna, och att det därför är fullt möjligt att jämföra värdena från 1970-talet med värdena i min forskning.³⁴

	Min metod	Westmans metod
Ordlängd	5,6	5,6
Meningslängd	14,9	14,5
OVIX	74,9	78,6

Tabell 6.2: Värden för några läsbarhetsmått på Westmans texter, dels med mina språkteknologiska metoder på en delmängd av hela materialet, dels med Westmans manuella metoder på hela materialet

6.5.2 Resultat

De ytliga mätningarna visar tämligen lika mönster i läroboksmaterialet. Samhällsvetenskap tycks genomgående ha ett mer komplicerat språk än naturvetenskap, och matematik har ett förhållandevis enkelt språk. Men också ordklassfördelningen i de olika ämnesfacken kan tolkas på samma vis. Matematik innehåller t.ex. få konjunktioner och många stora skiljetecken, båda tydliga signaler på enklare syntax. Naturvetenskap innehåller relativt många subjunktioner, vilket antyder ett mer komplicerat språk, med en förhållandevis hypotaktisk utformning. Samhällsvetenskapliga och akademiska texter innehåller fler adjektiv, vilket, på grund av adjektivens syntaktiska funktion i utbyggda nominalfraser, också pekar mot ett mer komplicerat språk.

Samtliga ytliga parametrar som mätts pekar på en språkutveckling som sker i stadiövergången mellan högstadiet och gymnasiet; meningarna blir längre och likaså orden. Ordförrådet blir dessutom mer varierat. I vissa av undersökningarna verkar det som att språkutvecklingen i naturvetenskap sker i något snabbare takt än i övriga ämnesfack. I detta tycks både biologi och kemi spela

³³Danielssons material är inte tillgängligt, och av Westmans material finns endast en två lärobokstexter tillgängliga, genom Talbanksprojektet. Närmare bestämt uppgår materialet till 5 484 ord, d.v.s. ungefär 25 % av Westmans totala lärobokstexter.

³⁴Visserligen skiljer sig värdet för meningslängd och OVIX något. Detta kan ha flera orsaker, som t.ex. urvalet av de Westmanska texterna som ingår i testet. En annan trolig orsak till att jag får något högre genomsnittlig meningslängd är att fristående makrosyntagmer, i t.ex. rubriker eller bildtext, med mina metoder ofta meningssegmenteras tillsammans med nästföljande mening. Westman räknar inte med dessa alls. Vad gäller OVIX-värdet tror jag att hela förklaringen till skillnaden är att testet bara utförts på en delmängd av Westmans texter. Som förklaras i avsnitt 6.1 är måttet inte helt okänt för texter av olika längd.

viktiga roller. Biologi utmärker sig bland de naturvetenskapliga ämnena genom en kraftigt ökad meningslängd på gymnasiet. Biologitexterna verkar alltså bli syntaktiskt mer komplexa, vilket förmodligen också speglas i ämnets förhållandevis många adverb och verb på gymnasiet. Kemitexterna verkar i stället bli mer lexikalt komplexa på gymnasiet, vilket framför allt den dramatiska utvecklingen av ordlängd visar, men som också märks på ämnets höga NQ-värde. På gymnasiet förmedlas mycket information i kemi genom orden, eftersom informationstunga ordklasser (framför allt substantiv) överväger.

Undersökningarna av NQ och LVIX bidrar emellertid till en mer komplex bild av utvecklingen i materialet. NQ är förvånansvärt hög för matematik och likaledes låg för de akademiska texterna. Av allt att döma har andelen substantiv mest betydelse i dessa resultat. Matematik ligger faktiskt i topp i andel substantiv på gymnasiet och de akademiska texterna består av en väsentligt mindre andel substantiv än lärobokstexterna.

I undersökningarna av LVIX och OVIX framkommer tydligt att de båda måtten mäter språket på olika nivåer. Tendenserna mellan mätresultaten för de olika måtten skiljer sig nämligen på ett par avgörande punkter. För det första visar LVIX-utvecklingen i matematik en negativ trend, från att i OVIX-mätningen ha uppmätt en lika kraftig ökning som samhällsvetenskap. Ytterligare en stor skillnad från OVIX-mätningen (och samtliga andra tidigare mätningar) är att LVIX för de berättande texterna överstiger LVIX för de akademiska texterna. De berättande texterna hamnar i LVIX-mätningen i den absoluta toppen, medan de akademiska texterna står i ungefär samma förhållande till lärobokstexterna som i OVIX-mätningen. Det faktum att historia är det skolämne som får högst LVIX-värde skulle kunna hänga samman med att texter i detta ämne bär många likheter med berättande texter.

Vidare sker knappt någon ökning av LVIX i matematik och fysik i stadiövergången. Samtidigt kan noteras att just matematik och fysik hela tiden har de lägsta värdena för ordlängd av alla ämnen. Att en kortare genomsnittlig ordlängd bör korrelera med en låg ordvariation har föreslagits i litteraturen (jfr Hultman och Westman 1977). Att endast det djupare ordvariationsmåttet, LVIX, förmår spegla detta skulle kunna tyda på att det ytliga ordvariationsmåttet, OVIX, tecknar en alltför förenklad bild av den språkliga verkligheten. I så fall är alltså ordvariationsmåttet LVIX att föredra framför OVIX (jfr Heilmann Mühlenbock 2013 som också konstaterar att LVIX är en god indikator för ordvariation).

I Thelanders undersökning av kvantitativa variabler i sakprosatexter under perioden 1908–1968 presenteras några mätningar som kan vara värda att nämna i sammanhanget. I en jämförelse mellan sakprosa och fiktionsprosa finner han t.ex. att sakprosatexterna (under de senare åren) kännetecknas av en signifikant längre meningslängd och ordlängd, samt ett högre LIX. Dessutom hittas

färre verb och fler adjektiv i sakprosatexterna (Thelander 1970). Alla dessa stilistiska förhållanden föreligger även i jämförelserna mellan hela mitt läroboksmaterial och de berättande texterna, vilket visar att dessa lexikogrammatiska variabler alltså är stilistiskt markerade i sammanhanget.

I jämförelse med tidigare kvantitativa analyser av läroboksmaterial, gjorda på 1970-talet, är värdena för meningslängd, ordlängd, LIX och OVIX genomgående lägre i mitt material. Även om mina uppmätta värden är lägre än både Danielssons och Westmans, kan det vara värt att påminna om att fem av sex ämnen i Westmans material är samhällsvetenskapliga (jfr 2.4.2.1). Samtidigt får mina samhällsvetenskapliga texter genomgående högre värden än texterna i både naturvetenskap och matematik. För att inte dra några förhastade slutsatser om läroboksspråkets historiska utveckling, kan det därför vara värt att jämföra Westmans siffror med mina resultat för de samhällsvetenskapliga texterna, och därigenom minimera effekten av att våra läroböcker faktiskt kommer från olika ämnesområden. Det visar sig dock att också värdena för meningslängd, ordlängd och OVIX för mina samhällsvetenskapliga texter överlag är lägre än Westmans. För meningslängd är värdena 14,1 mot 14,5, för ordlängd 5,44 mot 5,59 och för OVIX 69,7 mot 78,6.

7

DELSTUDIE: TYPISKT ORDFÖRRÅD

I denna delstudie undersöks det typiska ordförrådet i lärobokstexterna.

7.1 Typiskt ordförråd: studiens upplägg

Troligtvis är de flesta elever i den svenska skolan bekanta med svenska böjningsregler. De borde därför relativt lätt känna igen olika ordformer av det som i *Svenska Akademiens grammatik* (SAG, Telemann, Hellberg och Andersson 1999, del 2: 3) kallas för *lexikonord*. Dessa är grammatikens byggstenar och består av ensamma rotmorfem (*blind*) eller flera morfem som fogas ihop i sammansättningar (*färgblind*, *blindstyre*) eller avledningar (*blindhet*). Enligt SAG (del 2: 5) är funktionen och betydelsen av böjda ord förutsägbara utifrån suffixen. Ord som utgörs av flera morfem kan också vara semantiskt förutsägbara utifrån kompositionen, men helhetsbetydelsen skevar oftast på något sätt, och ringar endast in någon möjlig betydelseaspekt.

Vissa typer av konstruktioner kan dock vara besvärliga att entydigt kategorisera som lexikonord eller syntaktiskt bildade fraser. Till sådana svårkategoriserade ordförbindelser hör exempelvis fasta uttryck som *idag/i dag*, partikelverb som *ge upp*, flerordsprepositioner som *på grund av*, flerordssubjunktioner som *så att*, flerordskonjunktioner som *både... och*, flerordsadverb som *så pass* och förkortningar som *o.s.v.*

När språk ska analyseras elektroniskt måste begreppet ”ord” operationaliseras på ett sätt som möjliggör bästa möjliga automatiska identifikation och analys. I denna delstudie av texters ordförråd har jag valt att tala om ”ord” i termer av *lexikonenheter*. I denna beteckning ingår SAG:s lexikonordbegrepp, som jag förstår som den oböjda grundformen av ett ord och alla böjningsformer, t.ex. adjektivet *blind* tillsammans med böjningsformerna *blint*, *blinda* (jfr Hultman 2003). Härutöver inkluderar min teoretiska utgångspunkt de flesta fasta uttryck och flerordningar, även om många sådana av praktiska skäl inte kan identifieras.

I delstudien implementeras lexikonenheter som en kombination av grundform och ordklass, och utgör tillsammans de element som räknas i de annoterade lärobokstexterna. Exempelvis resulterar en mening som *Vi hör och ser sämre* (något förenklat) i följande lingvistiska analys: *vi/PN höra/VB och/KN se/VB dålig/JJ och/KN dålig/JJ*. Viktigt att förklara är också att lexikonenheterna känns igen på formella kriterier, men inte på betydelsemässiga. Ett polysemt ord, som *cell*, som t.ex. kan betyda ”litet isolerat rum” eller ”minsta levande enhet i en levande organism” (Svensk ordbok 2009), förs således ihop i formenheten *cell/NN*. För långt mer än hälften av alla ord i lexikonet behövs dock inte någon semantisk analys, då orden bara har en betydelse. Detta är ofta fallet med mindre vanliga ord, som t.ex. ämnesspecifika termer. Rik polysemi gäller främst de allra vanligaste orden i språket (Borin 2010).

Det typiska ordförrådet i lärobokstexterna identifieras enligt indexprincipen, som förklaras i avsnitt 5.3. De resulterande indexlistorna (jfr appendix E.2) bildar ett slags ordlistor över texttyptypiska ord och erhålls med en tröskelnivå på minst 15 förekomster av lexikonenheterna per miljon ord, vilket är samma nivå som vi använder för att extrahera akademiska ord till den svenska akademiska ordlistan (se avsnitt 2.2.4).

Extraktionsalgoritmen räknar alfabetiska ord och hybrida ord, som *1800-tal*, men inte numeriska ord eller logogram (jfr ordtyperna i Allén 1970). Skiljetecken, som kommatecken, parenteser och citationstecken, ingår inte heller i beräkningarna i denna studie. Vidare har jag valt att utesluta ord som taggats som egennamn, ordningstal och utländska ord, i den händelse de skulle dyka upp på listorna. Slutligen har listorna manuellt rensats på flerordsförkortningar, som t.ex., och teckensekvenser som hamnat på listorna på felaktiga grunder (t.ex. beroende på felaktig tokenisering).

De ordklasstaggar som används i ordlistorna presenteras i tabell 7.1. Ordklasserna följer standarden för SUC (Källgren 2006) – som brukar räknas som referenskorpus för svenska (Nilsson Björkenstam, Gustafson Capková och Wirén 2014).

Den morfologiska analysen (grundformstilldelningen) baseras på associationslexikonet SALDO (Borin, Forsberg och Lönnngren 2013).³⁵ Grundformer av orden fås av de s.k. *lemgrammen* i SALDO, i vilka bl.a. grundform, ordklass och böjningsmönster specificeras.

I enlighet med SAG (del 2) behandlas particip som avledning av verb, och skiljs alltså från verbformer i ordlistorna. I tagsettet för lemgrammen förekommer dock inte particip som en egen ordklass, utan participformer tillhör böj-

³⁵ Annoteringen av texterna beskrivs närmare i avsnitt 4.1.

Tagg	Betydelse
AB	adverb
DT	bestämningsord
HA	frågande/relativt adverb
HD	frågande/relativt bestämning
HP	frågande/relativt pronomen
HS	frågande/relativt possessivuttryck
IN	interjektion
JJ	adjektiv
KN	konjunktion
NN	substantiv
PC	particip
PL	partikel
PN	pronomen
PP	preposition
PS	possessivuttryck
SN	subjunktion
VB	verb

Tabell 7.1: Uppsättning av ordklasstaggar och deras betydelse

ningsparadigm för verb och/eller adjektiv m.m. Participformen *levande* ingår t.ex. som en böjningsform i både verbet *leva* och adjektivet *levande*. Eftersom particip finns i SUC-tagsettet kan ordklassen ändå identifieras.³⁶

³⁶Se vidare avsnitt 7.2.1 om participbehandlingen.

7.2 Typiskt ordförråd: resultat

7.2.1 Typiska ord

I resultatredovisningen visas till att börja med stadiesbundna ordlistor över ämnesfacken. I dessa markeras stadiespecifika ord, d.v.s. unika ord för högstadie- eller gymnasielistan, med indrag. Därefter följer separata förteckningar för högstadiet och gymnasiet över typiskt ämnesspråkliga ord i de olika naturvetenskapliga ämnena. Dessa listor utgör dels i sig själva lexikala beskrivningar av ämnesspråken, dels kan de jämföras med varandra för att undersöka stadieövergångarna ur lexikaliskt perspektiv. Till sist jämförs ordförrådet i naturvetenskapliga texter (för både högstadiet och gymnasiet tillsammans) med akademiska respektive berättande texter.

Som redan påpekats, redovisas listorna i par som är inbördes avhängiga av varandra. Den första kolumnen visar ord som är typiska för den första texttypen jämfört med den andra kolumnens texter. Den andra kolumnen visar de ord som är mest typiska för denna andra texttyp jämfört med den första.

Antal ord i listorna varierar mellan 813 och 2482, men i detta kapitel inkluderas endast de 20 överst placerade orden ur varje jämförelse, vilket alltså är de ord som är mest typiska i jämförelsen mellan de båda texttyperna. Något längre listutdrag, topp 100, är samlade i appendix E.2. Indexlistorna innehåller, i tur och ordning, information om ordens grundform, ordklassstillhörighet och framräknat index. För varje listpar presenteras även en sammanfattande tabell med uppgifter om totalt antal ord i listorna, andel långa ord (med ≥ 6 bokstäver) och ordklassfördelningarna över hela listorna.

Behandlingen av particip behöver eventuellt förklaras lite mer noggrant innan indexlistorna presenteras. I beräkningarna har nämligen presens och perfekt particip av praktiska skäl slagits ihop till en enhet som räknas tillsammans.³⁷ Det huvudord som anges för en particippost, presens eller perfekt particip, utgår sedan från vilket ord som är vanligast i texterna. För alla particip på ordlistor som redovisas i avhandlingen finns information om fördelningen mellan olika ordformer som räknats till detta huvudord i tabell E.23. Av denna tabell framgår att participberäkningen i mycket liten utsträckning påverkas av hopslagningen, eftersom det konkurrerande ordet till huvudordet sällan ens utgör en procentandel av förekomsterna.

³⁷Anledningen till detta är att beräkningarna utgår från de enkla ordklassangivelser som kommer från SUC, vilka är tillräckliga för övriga ordklasser.

7.2.1.1 Naturvetenskapliga texter jämförda med matematiktexter

De första 20 posterna på indexlistan över typiska ord i naturvetenskapliga texter och matematiktexter för högstadiet jämförda mot varandra visas i tabell 7.2. I listan markeras alltså stadiespecifika ord för högstadie- eller gymnasiet med indrag under kolumnen märkt ”unik”.

	Naturvetenskap		Matematik	
	UNIKT	Index	UNIKT	Index
1	<i>ämne</i> NN	23,04	<i>area</i> NN	33,71
2	<i>växt</i> NN	17,50	<i>tal</i> NN	33,12
3	<i>energi</i> NN	13,44	<i>avrunda</i> VB	30,57
4	<i>atom</i> NN	12,22	<i>triangel</i> NN	24,19
5	<i>kemisk</i> JJ	11,91	<i>uttryck</i> NN	18,59
6	<i>syre</i> NN	11,26	<i>kosta</i> VB	18,31
7	<i>cell</i> NN	8,92	<i>värde</i> NN	15,74
8	<i>koldioxid</i> NN	8,48	<i>omkrets</i> NN	15,54
9	<i>molekyl</i> NN	8,42	<i>vinkel</i> NN	14,53
10	<i>metall</i> NN	8,16	<i>ekvation</i> NN	14,51
11	<i>luft</i> NN	7,82	<i>pris</i> NN	14,49
12	<i>egenskap</i> NN	7,52	<i>kvadrat</i> NN	14,46
13	<i>stark</i> JJ	7,36	<i>betala</i> VB	14,38
14	<i>vätska</i> NN	7,28	<i>rektangel</i> NN	14,28
15	<i>djur</i> NN	7,13	<i>bråk</i> NN	13,81
16	<i>gas</i> NN	7,01	<i>procent</i> NN	13,11
17	<i>ström</i> NN	6,54	<i>svara</i> VB	13,07
18	<i>natur</i> NN	6,53	<i>teckna</i> VB	12,63
19	<i>blod</i> NN	6,51	<i>räkna ut</i> VB	11,90
20	<i>sprida</i> VB	6,51	<i>figur</i> NN	11,38

Tabell 7.2: Indexlista över ord i läroböcker i naturvetenskap/matematik för högstadiet, topp 20

För det första är indexen genomgående högre för matematikorden, vilket tyder på att den lexikala variationen är mindre i matematik, åtminstone bland de allra mest utmärkande ämnesorden.

Vidare är orden i naturvetenskap till största delen ämnesrelaterade substantiv, som *energi*, *atom*, *cell* och *syre*, medan motsvarande ordförråd i matematik innehåller flera ämnesrelaterade verb som *avrunda*, *teckna*, *svara* och *räkna ut*. De toppnoterade verben i matematik visar att de allra mest ämnesspecifika

orden i matematiktexterna till stor del utgörs av uppmaningar om att utföra matematiska beräkningar, medan texterna i naturvetenskap i första hand fokuserar deklarativ ämneskunskap.

Av de ord som är unika för högstadielistorna syns endast verbet *teckna* i topp 20-tabellen (i matematikkolumnen). Av de 100 första orden på de båda listorna är dock 12 stycken unika för högstadielistan i naturvetenskap, jämfört med matematiklistans 4 unika ord. Det typiska naturvetenskapliga ordförrådet verkar alltså vara mer stadiiebundet än de typiska orden i matematikläroböcker.

I tabell 7.3 visas översiktlig information om indexlistan över ord i läroböcker i naturvetenskap/matematik för högstadiet.³⁸ Matematiklistan är längre, vilket betyder att läroboksspråket i högstadieläroböcker i matematik innehåller fler ämnesspecifika ord än högstadieläroböcker i naturvetenskap. De typiska matematikorden är dessutom oftare långa ord.

	Naturvetenskap	Matematik
Antal ord	1069	1877
Långa ord %	52,4	62,5
Långa ord på topp 100	46	57
<i>Ordklasser %</i>		
Substantiv	30,6	58,2
Adjektiv	11,3	8,0
Pronomen	1,9	1,1
Verb	26,8	15,6
Particip	2,9	4,4
Adverb	14,2	5,3
Prepositioner	5,7	2,3
Konjunktioner	1,0	0,5
Subjunktioner	0,7	1,3
Interjektioner	0,1	0,2

Tabell 7.3: Information om indexlistan över ord i läroböcker i naturvetenskap/matematik för högstadiet

³⁸De ordklasser som ingår i tabellen är de huvudsakliga ordklasser som tas upp i SAG (förutom egennamn, räkneord och infinitivmärke).

Ordklassfördelningen visar att det typiska ordförrådet i naturvetenskap innehåller förhållandevis många verb, adverb och prepositioner jämfört med matematik. Men framför allt består det typiska ordförrådet i matematik till mycket högre grad av substantiv; en nästan dubbelt så stor andel av orden på matematiklistan är substantiv.

Undersöker man vilken typ av substantiv som förekommer på de båda listorna (tabell 7.2) märker man snabbt att substantiven på den naturvetenskapliga listan nästan uteslutande tillhör klassen ämnesrelaterade ord, som t.ex. *energi*, *atom* och *koldioxid*. Dessa ord hör till den naturvetenskapliga terminologin, eller de naturvetenskapliga ämnesspråken om man så vill. Substantiven på matematiklistan kan i stället, i stora drag, delas upp i två klasser: ämnesrelaterade respektive vardagliga ord. Ämnesrelaterade ord är t.ex. *area*, *omkrets* och *ekvation*, medan vardagliga substantiv, som *kompis*, *affär* och *pool*, förmodligen härstammar från uppdiktade räkneuppgifter. Detta senare användningsområde är en viktig del i förklaringen av matematiktexternas ymniga substantivbruk.

Topp 20 på gymnasielistan över det typiska ordförrådet i naturvetenskapliga texter och matematiktexter ställda mot varandra visas i tabell 7.4. Liksom indexlistan för naturvetenskap på högstadiet inleds gymnasielistan med ett flertal ämnesrelaterade substantiv, t.ex. *molekyl*, *protein* och *elektron*. Men i toppen denna gång återfinns något fler verb, som *reagera* och *binda*, vilka tydligt beskriver naturvetenskapliga fenomen. Toppverben i matematik, som *motivera* och *multiplitera*, tycks vara orienterade kring exempel och elevuppgifter.

Naturvetenskap			Matematik	
	UNIKT	Index	UNIKT	Index
1	<i>molekyl</i> NN	19,38	<i>triangel</i> NN	22,07
2	<i>reaktion</i> NN	16,10	<i>ekvation</i> NN	20,27
3	<i>cell</i> NN	12,20	<i>tal</i> NN	20,07
4	<i>elektron</i> NN	11,80	<i>räknare</i> NN	18,37
5	<i>jon</i> NN	10,08	<i>vinkel</i> NN	16,62
6	<i>reagera</i> VB	9,92	<i>motivera</i> VB	11,78
7	<i>protein</i> NN	9,61	<i>graf</i> NN	11,23
8	<i>atom</i> NN	9,48	<i>area</i> NN	10,56
9	<i>kemisk</i> JJ	8,82	<i>sann</i> JJ	10,55
10	<i>binda</i> VB	8,68	<i>multiplitera</i> VB	9,90
11	<i>energi</i> NN	8,62	<i>värde</i> NN	9,65
12	<i>ämne</i> NN	8,08	<i>variabel</i> NN	9,47
13	<i>bilda</i> VB	8,07	<i>funktion</i> NN	9,45
14	<i>luft</i> NN	7,41	<i>pris</i> NN	9,43
15	<i>växt</i> NN	7,38	<i>betala</i> VB	9,29
16	<i>gas</i> NN	7,11	<i>uttryck</i> NN	9,06
17	<i>vatten</i> NN	7,09	<i>kvadrat</i> NN	8,97
18	<i>syre</i> NN	6,89	<i>procent</i> NN	8,79
19	<i>ske</i> VB	6,88	<i>nämnare</i> NN	8,57
20	<i>metall</i> NN	6,73	<i>kosta</i> VB	8,54

Tabell 7.4: Indexlista över ord i läroböcker i naturvetenskap/matematik för gymnasiet, topp 20

Naturvetenskapslistans 100-i-topp innehåller 18 nya ord och matematiklistan 9 stycken. Även gymnasielistorna pekar således mot att läroboksspråket i naturvetenskapliga läroböcker är mer stadibundet, eller varierar mer mellan stadierna, medan matematikspråket är sig mer likt i stadiövergången till gymnasiet. Fortfarande är indexen i matematik alltigenom högre än motsvarande positioner i naturvetenskap, om än inte i lika hög grad som för högstadielistan.

I tabell 7.5, som sammanfattar information om indexlistan över ord i läroböcker i naturvetenskap/matematik för gymnasiet, syns en del intressanta tendenser i de båda läroboksspråken. Båda listorna innehåller en större andel långa ord på gymnasiet, såväl på topp 100 som på listan i sin helhet, men i matematik tillkommer fler långa ord än i naturvetenskap. Dessutom utökas det, redan för högstadiet, relativt stora typiska matematikordförrådet i övergången till gymnasiet.

	Naturvetenskap	Matematik
Antal ord	990	2470
Långa ord %	55,4	68,9
Långa ord på topp 100	53	58
<i>Ordklasser %</i>		
Substantiv	29,8	63,9
Adjektiv	10,4	8,5
Pronomen	1,4	1,1
Verb	28,0	12,3
Particip	3,9	5,2
Adverb	15,2	3,7
Prepositioner	5,2	2,0
Konjunktioner	1,1	0,4
Subjunktioner	0,6	0,1
Interjektioner	0,0	0,1

Tabell 7.5: Information om indexlistan över ord i läroböcker i naturvetenskap/matematik för gymnasiet

Ordklassfördelningen för gymnasielistorna är tämligen lik den för högstadiet. Mest anmärkningsvärt är fortfarande det att substantiv är dubbelt så vanliga på matematiklistan. Liksom för högstadiespråket beror detta på att läroboksspråket i matematik beskriver två domäner: den inomvetenskapliga, ämnesspråkligt matematiska, och den vardagliga.

7.2.1.2 *Naturvetenskapliga texter jämförda med samhällsvetenskapliga texter*

Resultatet av jämförelsen av ordförrådet i naturvetenskapliga och samhällsvetenskapliga läroböcker för högstadiet syns i tabell 7.6. Denna gång är det naturvetenskapliga ordförrådet det som är minst stadiespecifikt; bara 13 ord på högstadielistan för naturvetenskap (topp 100) är unika, medan hela 23 ord i samhällsvetenskap endast förekommer på högstadielistan. Listan i naturvetenskap bär på det hela taget många likheter med högstadielistan som föll ut i jämförelsen med matematik.

	Naturvetenskap		Samhällsvetenskap	
	UNIKT	Index	UNIKT	Index
1	<i>ämne</i> NN	14,85	<i>företag</i> NN	11,16
2	<i>atom</i> NN	13,50	<i>regering</i> NN	8,91
3	<i>cell</i> NN	12,00	<i>riksdag</i> NN	8,56
4	<i>kropp</i> NN	11,14	<i>krig</i> NN	7,78
5	<i>syre</i> NN	10,86	<i>parti</i> NN	7,41
6	<i>bakterie</i> NN	10,33	<i>makt</i> NN	7,31
7	<i>blod</i> NN	9,17	<i>stad</i> NN	7,23
8	<i>molekyl</i> NN	8,42	<i>land</i> NN	7,12
9	<i>kemisk</i> JJ	7,99	<i>vara</i> NN	6,83
10	<i>vätska</i> NN	7,48	<i>bo</i> VB	6,50
11	<i>växt</i> NN	6,40	<i>stat</i> NN	6,47
12	<i>innehålla</i> VB	6,23	<i>Sverige</i> NN	6,45
13	<i>energi</i> NN	6,01	<i>pengar</i> NN	6,27
14	<i>egenskap</i> NN	5,80	<i>karta</i> NN	6,26
15	<i>hjärna</i> NN	5,66	<i>politisk</i> JJ	6,16
16	<i>metall</i> NN	5,63	<i>handel</i> NN	5,62
17	<i>ljus</i> NN	5,62	<i>invånare</i> NN	5,60
18	<i>koldioxid</i> NN	5,43	<i>folk</i> NN	5,60
19	<i>hud</i> NN	5,42	<i>befolkning</i> NN	5,43
20	<i>muskel</i> NN	5,28	<i>betala</i> VB	5,37

Tabell 7.6: Indexlista över ord i läroböcker i naturvetenskap/samhällsvetenskap för högstadiet, topp 20

Listan för samhällsvetenskap toppas av flera för disciplinen karakteristiska substantiv, som *riksdag*, *makt* och *handel*. Granskar man hela topp 100 i samhällsvetenskap märker man dock att många av de specifikt samhällsvetenskapliga orden är gemensamma med vardagsspråkliga ord, t.ex. *bo*/VB,

fattig/JJ, *jobb/NN*, *man/NN* och *jag/PN*. Majoriteten av orden på indexlistan för naturvetenskap, som t.ex. *atom/NN*, *cell/NN* och *syre/NN*, är av mer vetenskaplig natur och kan lättare definieras exakt. De mest all dagliga naturvetenskapliga orden på topp 100 är *vatten/NN*, *djur/NN*, *lätt/AB* och *röd/JJ*.

I översikten över listornas innehåll (se tabell 7.7) syns att de är ganska lika beträffande längd och andel långa ord. Inte heller ordklassfördelningen visar några betydande skillnader. Substantiv och particip förekommer i något större utsträckning på listan för samhällsvetenskap, medan verb är vanligare på listan för naturvetenskap. Visserligen innehåller listan i naturvetenskap också 9 gånger så många subjunktioner som listan i samhällsvetenskap, men de absoluta värdena är så låga att man knappast kan dra några slutsatser av dem (på dessa listor motsvarar en procent knappt 9 förekomster).

	Naturvetenskap	Samhällsvetenskap
Antal ord	866	892
Långa ord %	49,3	54,4
Långa ord på topp 100	48	58
<i>Ordklasser %</i>		
Substantiv	35,3	40,9
Adjektiv	11,2	10,7
Pronomen	2,0	1,6
Verb	27,3	22,1
Particip	3,0	4,9
Adverb	10,6	10,8
Prepositioner	3,8	4,1
Konjunktioner	0,9	0,6
Subjunktioner	0,9	0,1
Interjektioner	0,1	0,1

Tabell 7.7: Information om indexlistan över ord i läroböcker i naturvetenskap/samhällsvetenskap för högstadiet

Indexlistan för gymnasiet mellan naturvetenskap och samhällsvetenskap kan ses i tabell 7.8. I naturvetenskap är flera av orden på topp 20 nya substantiv, som *elektron*, *protein* och *koncentration*. I samhällsvetenskap känns däremot topp 20 igen från högstadielistan, med ord som *stad*, *företag* och *politisk*, undantaget det nya substantivet *kommun*. Denna tendens visar sig fortsätta över hela topp 100; sammanlagt innehåller naturvetenskapslistan hela 24 nya ord, och samhällsvetenskapslistan endast 5 nya. Det typiska naturvetenskapliga ordförrådet utökas följaktligen mycket mer i studieövergången till

gymnasiet än den samhällsvetenskapliga vokabulären. Inräknat att böckerna i samhällsvetenskap redan på högstadiet uppvisar en förhållandevis vardaglig lexikal repertoar, ligger det nära till hands att sluta sig till att de naturvetenskapliga ämnena är betydligt mer lexikalt krävande, såväl på högstadiet som på gymnasiet.

	Naturvetenskap		Samhällsvetenskap	
	UNIKT	Index	UNIKT	Index
1	<i>cell</i> NN	20,69	<i>land</i> NN	11,61
2	<i>molekyl</i> NN	19,38	<i>stad</i> NN	11,51
3	<i>elektron</i> NN	15,50	<i>regering</i> NN	11,22
4	<i>ämne</i> NN	13,93	<i>ekonomisk</i> JJ	9,48
5	<i>atom</i> NN	12,91	<i>politisk</i> JJ	9,43
6	<i>reaktion</i> NN	12,51	<i>riksdag</i> NN	8,71
7	<i>protein</i> NN	11,16	<i>krig</i> NN	8,70
8	<i>jon</i> NN	10,08	<i>stat</i> NN	8,43
9	<i>energi</i> NN	8,32	<i>företag</i> NN	8,13
10	<i>vätska</i> NN	7,14	<i>parti</i> NN	8,08
11	<i>reagera</i> VB	7,03	<i>makt</i> NN	7,93
12	<i>partikel</i> NN	6,93	<i>samhälle</i> NN	7,36
13	<i>laddning</i> NN	6,62	<i>svensk</i> JJ	7,35
14	<i>syre</i> NN	6,61	<i>kommun</i> NN	6,95
15	<i>kropp</i> NN	6,22	<i>folk</i> NN	6,77
16	<i>avge</i> VB	6,18	<i>vara</i> NN	6,67
17	<i>bindning</i> NN	6,12	<i>demokrati</i> NN	6,33
18	<i>koncentration</i> NN	6,10	<i>befolkning</i> NN	6,06
19	<i>kemisk</i> JJ	5,99	<i>beslut</i> NN	5,89
20	<i>binda</i> VB	5,93	<i>pengar</i> NN	5,82

Tabell 7.8: Indexlista över ord i läroböcker i naturvetenskap/samhällsvetenskap för gymnasiet, topp 20

Tabell 7.9 visar information om indexlistan över ord i läroböcker i naturvetenskap/samhällsvetenskap för gymnasiet. Av denna framgår att den samhällsvetenskapliga listan både är något längre och innehåller en något större andel långa ord än den naturvetenskapliga listan. Skillnaden i fördelningen mellan substantiv och verb, som framkom i högstadijämförelsen, är nu nästan utjämnad. Dessutom har naturvetenskap på gymnasiet ett visst övertag vad gäller particip, i motsats till förhållandet på högstadiet.

	Naturvetenskap	Samhällsvetenskap
Antal ord	813	988
Långa ord %	53,4	58,9
Långa ord på topp 100	58	63
<i>Ordklasser %</i>		
Substantiv	36,5	38,9
Adjektiv	11,4	12,4
Pronomen	1,2	1,0
Verb	25,6	24,0
Particip	5,2	3,9
Adverb	10,6	10,0
Prepositioner	4,2	4,4
Konjunktioner	0,4	0,8
Subjunktioner	0,9	0,2
Interjektioner	0,0	0,1

Tabell 7.9: Information om indexlistan över ord i läroböcker i naturvetenskap/samhällsvetenskap för gymnasiet

7.2.1.3 Biologitexter på högstadiet och gymnasiet

Orden på topp 100 av indexlistan över biologitexter på högstadiet och gymnasiet har alltigenom lägre index än orden i de tidigare jämförelserna mellan olika ämnesfack. Dessa relativt låga index indikerar att en del av lexikonet är gemensamt över stadierna i ämnestexterna. Denna del hör sannolikt samman med själva ämnets praktik, och manifesteras genom verb och adverb. De ord som varierar mellan stadierna speglar i stället huvudsakligen det att olika ämnesområden behandlas i böckerna för högstadiet och gymnasiet. Vilka ämnesområden det rör sig om syns tydligast på substantiven.

En titt på orden på topp 20 (se tabell 7.10) bekräftar att de flesta orden är just biologinära ämnesord, t.ex. *druvsocker* och *groddjur* på högstadielistan, och *cellmembran* och *receptor* på gymnasielistan. På högstadielistan förekommer emellertid några personliga och possessiva pronomen, *du*, *jag* och *din*. Dessa borde bidra till ett tydligare läsarinluderande tilltal och tydligare berättarröst i högstadiet böckerna. Gymnasielistans allra översta del utmärks i stället av några tydligt textstrukturerande ord, som är typiskt skriftspråkliga till sin karaktär, t.ex. *dock*, *vilket*/HP och *respektive*/AB.

	Högstadiet	Index	Gymnasiet	Index
1	<i>drog</i> NN	5,59	<i>dock</i> AB	7,32
2	<i>alkohol</i> NN	5,16	<i>organisk</i> JJ	6,31
3	<i>druvsocker</i> NN	4,49	<i>därmed</i> AB	6,12
4	<i>dricka</i> VB	4,12	<i>vilket</i> HP	4,61
5	<i>skriva</i> VB	3,89	<i>denna</i> DT	4,33
6	<i>narkotika</i> NN	3,65	<i>glukos</i> NN	4,20
7	<i>Sverige</i> NN	3,61	<i>binda</i> VB	4,19
8	<i>hjärta</i> NN	3,61	<i>denna</i> PN	4,05
9	<i>betyda</i> VB	3,59	<i>utgöra</i> VB	3,84
10	<i>din</i> PS	3,34	<i>molekyl</i> NN	3,71
11	<i>röka</i> VB	3,30	<i>cellmembran</i> NN	3,57
12	<i>du</i> PN	3,26	<i>evolution</i> NN	3,50
13	<i>andas</i> VB	3,12	<i>receptor</i> NN	3,35
14	<i>slida</i> NN	3,12	<i>membran</i> NN	3,26
15	<i>hjärna</i> NN	3,04	<i>arké</i> NN	3,22
16	<i>jag</i> PN	3,02	<i>eukaryot</i> NN	3,20
17	<i>blodkärl</i> NN	2,96	<i>process</i> NN	3,16
18	<i>känsla</i> NN	2,91	<i>jon</i> NN	3,13
19	<i>groddjur</i> NN	2,87	<i>struktur</i> NN	3,05
20	<i>blod</i> NN	2,86	<i>respektive</i> AB	3,03

Tabell 7.10: Indexlista över ord i biologiläroböcker, topp 20

Av tabell 7.11 kan utläsas att ordlistan för högstadiet innehåller något fler poster än gymnasielistan, samt att orden på gymnasielistan oftare är långa. Vad gäller ordklassfördelningen kan sägas att högstadielistan framför allt innehåller en högre andel substantiv, medan gymnasielistan i stället har ett relativt överskott av adjektiv och particip. Tilläggas kan också att dessa listor, där ämnet är konstant och stadierna varierar, i enlighet med resonemanget ovan, innehåller en mindre andel verb och adverb, och en högre andel substantiv, än de listor som jämför ämnesfack.

	Högstadiet	Gymnasiet
Antal ord	2482	2067
Långa ord %	64,7	75,3
Långa ord på topp 100	62	75
<i>Ordklasser %</i>		
Substantiv	58,4	52,0
Adjektiv	8,6	11,4
Pronomen	1,2	0,8
Verb	17,2	17,0
Particip	4,0	6,9
Adverb	5,4	6,1
Prepositioner	1,5	3,0
Konjunktioner	0,2	0,5
Subjunktioner	0,2	0,3
Interjektioner	0,0	0,1

Tabell 7.11: Information om indexlistan över ord i biologiläroböcker

7.2.1.4 *Fysiktexter på högstadiet och gymnasiet*

Även topporden på listorna för fysik (se tabell 7.12) innehåller genomgående ord med lägre index än listorna för ämnesfacken. Bland orden på högstadielistans topp 100 är visserligen flera typiska fysikord, som t.ex. *transistor*, *elektromagnet* och *ljudvåg*, men många av orden har också en vardagsspråklig karaktär, som *sak*, *farlig* och *se ut*. Gymnasielistan innehåller färre vardagligt klingande ord, samt orienteras tydligare runt matematiska övningar, vilket högt placerade ord som *beräkna*, *lösning*, *värde* och *uppskatta* visar.

	Högstadiet	Index	Gymnasiet	Index
1	<i>lins</i> NN	7,06	<i>konstant</i> JJ	8,83
2	<i>anteckna</i> VB	6,69	<i>fält</i> NN	6,99
3	<i>måne</i> NN	5,59	<i>accelerera</i> VB	6,80
4	<i>konvex</i> JJ	5,42	<i>acceleration</i> NN	6,02
5	<i>kikare</i> NN	5,22	<i>beräkna</i> VB	5,75
6	<i>transistor</i> NN	4,51	<i>figur</i> NN	5,64
7	<i>himmel</i> NN	4,41	<i>lösning</i> NN	5,44
8	<i>diod</i> NN	4,36	<i>enligt</i> PP	5,37
9	<i>ljud</i> NN	4,24	<i>värde</i> NN	5,27
10	<i>konkav</i> JJ	4,10	<i>area</i> NN	5,27
11	<i>ämne</i> NN	3,94	<i>fjäder</i> NN	5,18
12	<i>synas</i> VB	3,63	<i>rörelsemängd</i> NN	5,04
13	<i>hävstång</i> NN	3,51	<i>medium</i> NN	4,30
14	<i>sol</i> NN	3,42	<i>bestämna</i> VB	4,21
15	<i>elektromagnet</i> NN	3,37	<i>graf</i> NN	4,16
16	<i>stjärnbild</i> NN	3,36	<i>hastighet</i> NN	4,07
17	<i>spegel</i> NN	3,32	<i>radie</i> NN	4,04
18	<i>så här</i> AB	3,31	<i>nuklid</i> NN	4,03
19	<i>materia</i> NN	3,27	<i>positiv</i> JJ	3,87
20	<i>magnet</i> NN	3,24	<i>foton</i> NN	3,72

Tabell 7.12: Indexlista över ord i fysikläroböcker, topp 20

Sammantaget kan man av tabell 7.13 utläsa att gymnasielistan i fysik innehåller en högre procentandel långa ord än högstadielistan. Beträffande ordklassfördelningen i de båda listorna är den påtagligt konstant över stadierna. Andelen adjektiv och particip kan dock sägas öka marginellt på bekostnad av adverb.

	Högstadiet	Gymnasiet
Antal ord	2245	1918
Långa ord %	62,4	71,7
Långa ord på topp 100	60	67
<i>Ordklasser %</i>		
Substantiv	52,4	52,2
Adjektiv	8,8	9,4
Pronomen	1,2	1,0
Verb	18,6	18,7
Particip	4,5	5,3
Adverb	8,0	6,3
Prepositioner	2,5	3,6
Konjunktioner	0,4	0,6
Subjunktioner	0,4	0,3
Interjektioner	0,0	0,1

Tabell 7.13: Information om indexlistan över ord i fysikläroböcker

7.2.1.5 *Kemitexter på högstadiet och gymnasiet*

Topp 20 på indexlistan över kemitexter visas i tabell 7.14. Den absoluta toppen av gymnasielistan i kemi innehåller de högsta indexen av de ämnesspecifika listorna. Gymnasieböckerna i kemi är alltså tydligt uppbyggda kring vissa centrala ämnesanknutna områden, kopplade till ord som *koncentration*, *mol* och *jämvikt*.

	Högstadiet	Index	Gymnasiet	Index
1	<i>eld</i> NN	7,98	<i>koncentration</i> NN	12,76
2	<i>sak</i> NN	5,76	<i>mol</i> NN	11,36
3	<i>brinna</i> VB	5,34	<i>figur</i> NN	9,19
4	<i>äta</i> VB	4,95	<i>fas</i> NN	7,36
5	<i>faktiskt</i> AB	4,79	<i>beräkna</i> VB	7,32
6	<i>förvandla</i> VB	4,77	<i>volym</i> NN	7,06
7	<i>mark</i> NN	4,74	<i>jämvikt</i> NN	6,21
8	<i>jord</i> NN	4,59	<i>ange</i> VB	6,17
9	<i>du</i> PN	4,58	<i>följande</i> PC	5,42
10	<i>varför</i> HD	4,52	<i>kovalent</i> JJ	5,17
11	<i>växt</i> NN	4,50	<i>bestämma</i> VB	4,95
12	<i>olja</i> NN	4,17	<i>substansmängd</i> NN	4,81
13	<i>se ut</i> VB	4,12	<i>respektive</i> AB	4,80
14	<i>kropp</i> NN	4,01	<i>oxidera</i> VB	4,63
15	<i>mat</i> NN	3,99	<i>vätebindning</i> NN	4,62
16	<i>trä</i> NN	3,99	<i>prov</i> NN	4,53
17	<i>tillverka</i> VB	3,98	<i>därmed</i> AB	4,39
18	<i>tänka</i> VB	3,98	<i>pH</i> NN	4,31
19	<i>malm</i> NN	3,95	<i>ekvivalenspunkt</i> NN	4,23
20	<i>djur</i> NN	3,88	<i>kolonn</i> NN	4,14

Tabell 7.14: Indexlista över ord i kemiläroböcker, topp 20

Högstadiet fokuserar tydligare kemiämnetts förklarande del, med ord som *faktiskt* och *varför*/HD, som motiverar att ämnet över huvud taget är intressant att ägna sig åt. I gymnasiet tycks studiet av kemi tas mer för givet, och texterna innehåller därför fler räkneppgifter med anknytning till ämnet (jfr ord som *beräkna* och *bestämma*), samt textbindande ord som *därmed* och *enligt* (av vilka det senare syns på topp 100).

Uppgifterna om de båda ordlistorna i kemi (se tabell 7.15) visar att ordklasserna fördelas ganska olika. På högstadiet är de typiska orden oftare verb, och på gymnasiet substantiv och particip. På gymnasielistan är en högre andel ord långa. Däremot innehåller de båda listorna i stort sett lika många ord.

	Högstadiet	Gymnasiet
Antal ord	2196	2223
Långa ord %	63,8	75,3
Långa ord på topp 100	48	68
<i>Ordklasser %</i>		
Substantiv	52,1	56,6
Adjektiv	8,7	10,1
Pronomen	1,5	0,8
Verb	19,8	15,0
Particip	4,1	7,1
Adverb	6,8	5,3
Prepositioner	2,3	2,2
Konjunktioner	0,5	0,5
Subjunktioner	0,3	0,2
Interjektioner	0,1	0,2

Tabell 7.15: Information om indexlistan över ord i kemiläroböcker

7.2.1.6 *Naturvetenskapliga texter jämförda med akademiska texter*

Resultatet av jämförelsen mellan naturvetenskapliga och akademiska texter presenteras i tabell 7.16. En intressant iakttagelse vad gäller listan för naturvetenskap är att den i jämförelsen med akademiska texter innehåller flera högt placerade formord som är vanliga i vardagsspråket, som t.ex. *som*/HP, *den*/DT och *än*/KN. I tidigare jämförelser med lärobokstexter från andra ämnesfack har naturvetenskaplistorna starkt dominerats av ämnesrelaterade substantiv.

	Naturvetenskap		SveAk	
1	<i>som</i> HP	193,19	<i>som</i> KN	91,93
2	<i>den</i> DT	102,18	<i>andra</i> JJ	23,06
3	<i>vatten</i> NN	27,99	<i>jag</i> PN	19,82
4	<i>än</i> KN	22,52	<i>vilken</i> HP	19,09
5	<i>cell</i> NN	16,79	<i>både</i> KN	11,38
6	<i>energi</i> NN	16,56	<i>min</i> PS	8,93
7	<i>bilda</i> VB	14,02	<i>social</i> JJ	8,69
8	<i>atom</i> NN	13,64	<i>relation</i> NN	7,97
9	<i>växt</i> NN	13,46	<i>studie</i> NN	7,39
10	<i>molekyl</i> NN	13,34	<i>perspektiv</i> NN	6,42
11	<i>ämne</i> NN	12,96	<i>utifrån</i> PP	5,87
12	<i>elektron</i> NN	12,37	<i>avhandling</i> NN	5,84
13	<i>varje</i> DT	12,16	<i>diskussion</i> NN	5,43
14	<i>luft</i> NN	11,98	<i>politisk</i> JJ	5,27
15	<i>kemisk</i> JJ	11,18	<i>erfarenhet</i> NN	4,9
16	<i>syre</i> NN	10,83	<i>syfte</i> NN	4,79
17	<i>börja</i> VB	10,74	<i>såväl</i> KN	4,71
18	<i>bakterie</i> NN	10,17	<i>respektive</i> AB	4,62
19	<i>jord</i> NN	9,97	<i>fråga</i> NN	4,62
20	<i>protein</i> NN	9,25	<i>situation</i> NN	4,45

Tabell 7.16: Indexlista över ord i läroböcker i naturvetenskap/akademiska texter, topp 20

Även de mest typiska akademiska orden är i många fall både korta och vanliga ord, som *som*/KN, *andra*/JJ, *jag* och *min*/PS. De relativt sett få akademiskt klingande substantiven på topp 20, exempelvis *relation*, *studie* och *perspektiv*, visar sig vara allmänt spridda, och inte särskilt knutna till något visst ämnesområde. Detta är i och för sig väntat med tanke på extraheringsmetoden som just garanterar spridning över olika ämnesområden, men ändå värt att påpeka.

De båda aktuella listorna omfattar i stort sett lika många ord, vilket framgår av informationen i tabell 7.17. Den akademiska listan innehåller dock en betydligt högre andel långa ord. Den allra största skillnaden mellan de typiska ordförråden hittas i participfördelningen; den akademiska listan innehåller nästan tre gånger så många particip som den naturvetenskapliga listan. Också adjektiv visar sig vara vanligare i typisk akademisk vokabulär, medan de typiska orden i naturvetenskap oftare är substantiv.

	Naturvetenskap	SveAk
Antal ord	2204	2221
Långa ord %	60,4	75,4
Långa ord på topp 100	41	69
<i>Ordklasser %</i>		
Substantiv	52,4	44,2
Adjektiv	8,8	14,2
Pronomen	0,9	0,8
Verb	21,1	18,5
Particip	3,6	9,8
Adverb	6,6	7,6
Prepositioner	3,2	1,4
Konjunktioner	0,2	0,7
Subjunktioner	0,3	0,2
Interjektioner	0,1	0,1

Tabell 7.17: Information om indexlistan över ord i läroböcker i naturvetenskap/akademiska texter

7.2.1.7 Naturvetenskapliga texter jämförda med berättande texter

Indexlistorna över naturvetenskapliga och berättande texter kan ses i tabell 7.18. Resultatet för de naturvetenskapliga texterna i denna jämförelse är mer likt resultatet för de naturvetenskapliga texterna i jämförelserna med läroböcker från andra ämnesfack. En stor del av orden i naturvetenskap är ämnesrelaterade substantiv och verb, varav de flesta av de senare, i typfallet, är statiska till sin karaktär, som t.ex. *innehålla* och *bestå (av)*. Verben i de berättande texterna, som *säga*, *fråga* och *undra*, hör oftare till ett agentivt subjekt.

	Naturvetenskap	Index	Berättande	Index
1	<i>ämne</i> NN	24,57	<i>jag</i> PN	84,69
2	<i>bilda</i> VB	23,19	<i>hon</i> PN	40,40
3	<i>energi</i> NN	16,62	<i>han</i> PN	24,93
4	<i>cell</i> NN	15,33	<i>min</i> PS	14,73
5	<i>olik</i> JJ	14,88	<i>hennes</i> PS	12,42
6	<i>innehålla</i> VB	13,91	<i>hans</i> PS	11,87
7	<i>atom</i> NN	13,79	<i>säga</i> VB	10,53
8	<i>molekyl</i> NN	13,31	<i>ja</i> IN	9,33
9	<i>elektron</i> NN	12,36	<i>fråga</i> VB	8,90
10	<i>växt</i> NN	12,11	<i>ni</i> PN	8,84
11	<i>kemisk</i> JJ	11,17	<i>ansikte</i> NN	6,86
12	<i>syre</i> NN	10,72	<i>blick</i> NN	6,83
13	<i>bestå</i> VB	10,63	<i>dörr</i> NN	6,67
14	<i>bakterie</i> NN	10,06	<i>nej</i> IN	6,26
15	<i>använda</i> VB	9,63	<i>väl</i> AB	6,16
16	<i>exempel</i> NN	9,38	<i>röst</i> NN	5,53
17	<i>protein</i> NN	9,26	<i>nog</i> AB	5,43
18	<i>koldioxid</i> NN	8,75	<i>undra</i> VB	5,33
19	<i>temperatur</i> NN	8,29	<i>ringa</i> VB	5,33
20	<i>gas</i> NN	8,15	<i>aldrig</i> AB	5,31

Tabell 7.18: Indexlista över ord i läroböcker i naturvetenskap/berättande texter, topp 20

Orden i indexlistan över det typiska ordförrådet i berättande texter är påtagligt korta, vardagliga, vanliga och enkla. Listan domineras av toppnoterade personliga och possessiva pronomen, som *hon* och *min*. Även de vanliga interjektionerna *ja* och *nej* är starkt utmärkande för de berättande texterna. Tydligt är också att de mest typiska substantiven används i utsagor om mänskligt agerande; antingen refererar de till människors kroppsdelar, som t.ex. *ansikte* och *huvud* (på plats 25), eller också till mänskliga uttrycksmedel, som *blick* och *röst*.

Av tabell 7.19 kan man utläsa att de typiska orden för berättande texter är avsevärt färre än de typiska orden för naturvetenskapliga texter. Det beror troligtvis på att ordförrådet i de berättande texterna är mer varierat (jfr OVIX- och LVIX-mätningarna i 6.3.4 och 6.4.3). Ytterligare en anledning till att indexlistan för berättande texter innehåller färre poster kan vara att det lexikala överlappet mellan registren utgör en större del av lexikonet i de berättande texterna.

	Naturvetenskap	Berättande
Antal ord	2030	1380
Långa ord %	66,9	46,0
Långa ord på topp 100	53	24
<i>Ordklasser %</i>		
Substantiv	57,5	27,1
Adjektiv	9,9	9,0
Pronomen	0,7	2,5
Verb	16,7	36,4
Particip	5,0	0,9
Adverb	5,1	11,9
Prepositioner	2,6	4,5
Konjunktioner	0,5	0,4
Subjunktioner	0,0	0,9
Interjektioner	0,0	1,2

Tabell 7.19: Information om indexlistan över ord i läroböcker i naturvetenskap/berättande texter

Av tabellen framgår också att orden i listan över naturvetenskapliga texter oftare är långa ord. Den naturvetenskapliga listan innehåller också en mycket högre andel substantiv och particip, medan listan över berättande texter till större del utgörs av pronomen, verb och adverb.

7.2.2 Fokusstudier

7.2.2.1 Typiska ämnesfackliga ords grad av vardagsspråkighet

Orden på listorna över typisk samhällsvetenskaplig vokabulär (jfr 7.2.1.2) ger intryck av att vara förhållandevis vanliga i vardagsspråket. Frekvensuppgifter för de allra högst placerade orden på högstadie- och gymnasielistorna i naturvetenskap, matematik och samhällsvetenskap, i några av de textmaterial som ingår i Språkbanken, visas i tabell 7.20.³⁹

³⁹Sökningen genomfördes den 7 mars 2015.

		GP 2013	Bloggmix 2014	Romaner
Naturv.	<i>cell</i> /NN	13,6	5,3	21,3
	<i>molekyl</i> /NN	1,3	0,5	0,8
	<i>växt</i> /NN	31,4	21,8	35,5
	<i>elektron</i> /NN	0,2	0,1	1,7
	<i>reaktion</i> /NN	52,6	26,3	41,0
Matematik	<i>area</i> /NN	1,5	5,2	0,7
	<i>tal</i> /NN	168,9	145,9	185,3
	<i>avrunda</i> /VB	2,2	6,0	2,1
	<i>triangel</i> /NN	1,3	2,6	2,5
	<i>ekvation</i> /NN	5,7	2,7	2,5
Samhällsv.	<i>land</i> /NN	717,0	339,8	296,5
	<i>stad</i> /NN	138,0	340,4	249,6
	<i>regering</i> /NN	389,6	126,6	78,7
	<i>företag</i> /NN	438,1	198,6	51,9
	<i>riksdag</i> /NN	71,8	42,0	14,6

Tabell 7.20: Antal träffar per miljon ord för toppnoterade indexord i läroböcker i naturvetenskap, matematik och samhällsvetenskap, i olika korpusar i Språkbanken

Sökningen visar att flera av de allra mest typiska orden i naturvetenskap och matematik är extremt eller mycket sällsynta i tidningstext, bloggtext och berättande text. För de sökord som är något mer vanligt förekommande, *reaktion*/NN och *tal*/NN, visar dessutom konkordanserna att det nästan aldrig rör sig om den tekniska betydelsen av ordet. De vanligaste betydelseerna av dessa ord illustreras i exempelmeningarna (60) och (61), som är hämtade ur blogg-korpusen.

(60) *Jag och Andreas kunde inte sluta skratta åt Annicas reaktioner på Marcelos svar!*

(61) *På tal om London så var vi med om en sjukt pinsam olycka i London när vi var på väg hem efter en utekväll.*

Jag har också undersökt täckningen (eng. *coverage*) för samtliga ord på listorna i naturvetenskap och samhällsvetenskap i korpusen med berättande texter. Resultaten presenteras i tabell 7.21. Det visar sig att de samhällsvetenskapliga listorna täcker absolut högst andel av orden i de berättande texterna. De naturvetenskapliga ordlistorna täcker betydligt färre berättande ord.⁴⁰

	Högstadiet (%)	Gymnasiet (%)
<i>Naturvetenskap</i>	23,4	15,5
<i>Matematik</i>	30,7	27,3
<i>Samhällsvetenskap</i>	37,0	39,6

Tabell 7.21: De ämnesfackliga ordlistornas täckning i korpusen med berättande texter

Denna kontroll stärker slutsatsen att de samhällsvetenskapliga lärobokstexterna till relativt stor del liknar A-texter (jfr avsnitt 2.1), i alla fall vad gäller ordförrådet. Särskilt intressant är att notera att de naturvetenskapliga läroböckerna på gymnasiet innehåller ett ännu mer specialiserat ordförråd, i förhållande till A-texter, än högstadieläroböckerna. I samhällsvetenskap och matematik ökar inte den lexikala specifikationen tillnärmelsevis lika kraftigt i stadiövergången.

7.2.2.2 Samhällsvetenskapliga texter jämförda med akademiska texter

I jämförelsen mellan naturvetenskapliga och akademiska texter skiljer sig texterna åt både ämnesmässigt och stadiemässigt (se avsnitt 7.2.1.6). Stadieskillnaden mellan texterna gör att vissa ord med mer vardaglig prägel hamnar på naturvetenskapslistan. Då de akademiska texterna kommer från humanistiska och samhällsvetenskapliga ämnesområden, och inte innehåller naturvetenskapliga texter alls, är det emellertid något förvånande att listan för naturvetenskap inte innehåller fler naturvetenskapliga termer.

Eftersom samhällsvetenskapliga texter ingår i SveAk torde man med indexlistor över samhällsvetenskapliga lärobokstexter jämförda med akademiska texter bättre kunna isolera förändringen i typiskt ordförråd i stadiövergången till eftergymnasiala utbildningar. Topp 20-resultat av en sådan körning visas i

⁴⁰I denna undersökning har jag använt de indexlistor i naturvetenskap som faller ut i jämförelsen med samhällsvetenskapliga texter. Eftersom listorna i naturvetenskap och samhällsvetenskap innehåller ungefär lika många ord bedöms dessa täckningsgrader vara jämförbara med varandra.

tabell 7.22. Jämfört med sammanställningen av naturvetenskapliga och akademiska texter blir index för de typiska samhällsvetenskapliga orden förhållandevis låga, vilket bekräftar att texttyperna ligger närmare varandra lexikalt.

	Samhällsvetenskap	Index	SveAk	Index
1	<i>den</i> DT	141,10	<i>som</i> HP	189,16
2	<i>många</i> JJ	31,50	<i>som</i> KN	91,93
3	<i>än</i> KN	18,48	<i>andra</i> JJ	23,69
4	<i>börja</i> VB	12,91	<i>vilken</i> HP	19,09
5	<i>varje</i> DT	11,40	<i>både</i> KN	11,38
6	<i>flera</i> JJ	11,00	<i>studie</i> NN	7,01
7	<i>lång</i> AB	10,35	<i>avhandling</i> NN	6,02
8	<i>land</i> NN	8,94	<i>relation</i> NN	5,90
9	<i>stat</i> NN	7,84	<i>analys</i> NN	5,45
10	<i>jord</i> NN	7,56	<i>jag</i> PN	5,20
11	<i>ekonomisk</i> JJ	7,28	<i>min</i> PS	4,75
12	<i>fler</i> JJ	7,15	<i>såväl</i> KN	4,71
13	<i>helt</i> AB	6,97	<i>utifrån</i> PP	4,68
14	<i>liten</i> AB	6,83	<i>båda</i> JJ	4,63
15	<i>miljon</i> NN	6,12	<i>respektive</i> AB	4,62
16	<i>köpa</i> VB	6,07	<i>dock</i> AB	4,55
17	<i>vatten</i> NN	5,81	<i>perspektiv</i> NN	4,47
18	<i>stad</i> NN	5,75	<i>erfarenhet</i> NN	4,42
19	<i>krig</i> NN	5,64	<i>dels</i> KN	4,09
20	<i>riksdag</i> NN	5,55	<i>aspekt</i> NN	4,00

Tabell 7.22: Indexlista över ord i läroböcker i samhällsvetenskap/akademiska texter, topp 20

På listorna över typiska akademiska ord är topp 20 mycket lika varandra i jämförelserna med naturvetenskap och samhällsvetenskap. Formella konjunktioner, som *både*, *såväl* och *dels*, samt relativa pronomen, som *respektive* och *dock*, utmärker de akademiska texterna.

Listan i samhällsvetenskap skiljer sig tydligt mot tidigare jämförelser med naturvetenskapliga texter, i det att de typiskt ämnesrelaterade orden inte syns till (som *riksdag*, *regering* och *företag*). Detta visar också att ämnesmässiga likheter mellan texttyperna tydligt slår igenom på indexlistor över lexikala variabler. Listan i samhällsvetenskap innehåller genomgående enkla, vardagliga

ord. Denna tendens märks redan i jämförelse mellan läroböcker i naturvetenskap och samhällsvetenskap, men framkommer ännu tydligare i denna jämförelse med akademiska texter.

Ser man till informationen i tabell 7.23 visar sig några ytterligare skillnader mellan hur det typiska lexikonet i naturvetenskap respektive samhällsvetenskap förhåller sig till det typiska akademiska lexikonet. För det första är den akademiska ordlistan i jämförelse med samhällsvetenskap kortare, d.v.s. fler av de vanliga orden i akademiska texter förekommer i samhällsvetenskapliga lärobokstexter.

	Samhällsvetenskap	SveAk
Antal ord	2075	1645
Långa ord %	60,6	78,0
Långa ord på topp 100	39	67
<i>Ordklasser %</i>		
Substantiv	47,9	38,6
Adjektiv	11,6	15,1
Pronomen	1,1	0,9
Verb	22,1	19,3
Particip	2,7	11,2
Adverb	7,2	8,9
Prepositioner	3,4	1,8
Konjunktioner	0,3	0,8
Subjunktioner	0,2	0,3
Interjektioner	0,0	0,1

Tabell 7.23: Information om indexlistan över ord i läroböcker i samhällsvetenskap/akademiska texter

Vad gäller ordklassfördelningen innehåller den akademiska listan i jämförelse med samhällsvetenskap en högre andel particip och adverb, men en mindre andel substantiv, än listan mot naturvetenskap. Eftersom de akademiska ordlistorna i de båda fallen är av avsevärt olika längd är det emellertid svårt att dra några slutsatser av ordklassfördelningen över listorna. Eftersom indexprincipen av naturen missgynnar ovanliga ord i korpusarna, borde andelen typiska innehållsordklasser, t.ex. substantiv, vara extra känsliga mot förändringar i listans längd.

7.2.2.3 *Typiska ämnesfackliga ord i förhållande till styrdokument*

En relevant fråga är huruvida några av de lexikala skillnader mellan lärobokstexterna i ämnesfacken (naturvetenskap, matematik och samhällsvetenskap), som denna studie påvisar, kan kopplas till de aktuella ämnesplanerna för högstadiet respektive gymnasiet. En inte alltför djärv gissning är att läroboksförfattarna på något vis har utgått ifrån målbeskrivningar med utbildningen i utarbetandet av texterna. Kanske kan en del av ordförrådet i böckerna förklaras av styrdokumentens formuleringar.

För att undersöka om det finns något samband mellan läroböckernas språkdräkt och föreskrivna ämnesmässiga utvecklingsmål för eleverna har jag gjort en mindre studie av ämnesplanerna för grundskolan och gymnasiet, så som de beskrivs i Lgr2011 (Skolverket 2011b) och Gy2011 (Skolverket 2010, 2011a). Syftet är att analysera vilket slags verksamhet läroplanen uppmuntrar i de olika ämnesfacken, och om det föreligger några påtagliga skillnader ämnesfacken emellan, för att sedan eventuellt förklara lexikala skillnader i läroböckerna.⁴¹

Till att börja med har jag valt att i ämnesplanerna notera vilka verb som förekommer i utvecklingsmålen. Verben har valts i egenskap av viktiga innehållsord med predikativ funktion; det är med hjälp av verb som ett aktivt lärande kan beskrivas (jfr Blooms taxonomi med kognitiva lärandemål, som baseras på s.k. *aktiva verb*, i t.ex. Bloom och Krathwohl 1956 och Gronlund 2000). Om syftet hade varit att undersöka vilket ämnesinnehåll som fastslås i planerna hade viktiga substantiv, med referentiell funktion, varit naturligare att fokusera.

De verb som förekommer i utvecklingsmålen för högstadiet och gymnasiet listas i det följande, där verb som ämnesfacket är ensamt om på stadiet har markerats med fetstil.⁴²

I ämnesplanerna för högstadiet förekommer följande verb:

- Naturvetenskap: *använda*, **beskriva**, **förklara**, *genomföra*, *granska*, *ta ställning*
- Matematik: *analysera*, *använda*, *argumentera*, **formulera**, **följa**, **föra resonemang**, *göra*, *lösa*, *redogöra*, **samtala**, *välja*, *värdera*
- Samhällsvetenskap: *analysera*, *granska*, **reflektera**, *värdera*

⁴¹Det kan vara värt att påminna om att en mindre del av högstadietexterna kommer från läroböcker som skrevs i början av 2000-talet, då andra läroplaner var rådande. Som konstaterades i delstudien av traditionella läsbarhetsmått är dock den språkliga utformningen av läroböckerna i mitt material tillräckligt konstant över tid för att betraktas som ett synkront register (se avsnitt 6.5.1).

⁴²Undantagsvis förekommer kollokationer, i vilka verbets betydelse inte kan förstås på rätt sätt utan argumentet (som t.ex. i *dra slutsatser*) i uppställningen.

I ämnesplanerna för gymnasiet förekommer följande verb:

- Naturvetenskap: *analysera, använda, formulera, genomföra, granska, identifiera, kommunicera, lösa, planera, redovisa, reflektera, söka, tolka, värdera*
- Matematik: *analysera, använda, bedöma, beskriva, formulera, följa, föra resonemang, hantera, kommunicera, lösa, relatera, tolka, utforma, utvärdera, värdera*
- Samhällsvetenskap: *analysera, använda, söka/samla, granska/tolka, värdera, presentera, förklara/utreda, dra slutsatser/förstå/identifiera*

Som ämnesfack är naturvetenskap unikt i att uppmuntra elever att beskriva och förklara fenomen i naturen, vilket är tydligt språkligt förankrade aktiviteter. Undervisningen utmärks också av att man vill lära elever utföra och redogöra för experiment, alltså både att tillämpa ämneskunskap och att tala om den. Det faktum att listorna i naturvetenskap domineras av ämnesrelaterade substantiv och verb kan nog påstås vara en direkt konsekvens av dessa påbjudna fokusområden för den naturvetenskapliga undervisningen.

Fokus på den språkliga komponenten i ämnesfacken är dock allra störst i matematik, där elever ska kunna resonera om matematik med matematiska begrepp, vilka också ska kunna beskrivas och struktureras. En annan kunskapsgrän som är speciell för matematikämnet är att eleverna ska lära sig att tänka i former av modeller.

Till skillnad från övriga ämnesfack, där ämneskunskapen inte ska ifrågasättas, handlar undervisningen i samhällsvetenskap om att eleverna ska kunna reflektera över vad de lär sig, genom ett kritiskt förhållningssätt till olika informationskällor. De ska lära sig att själva identifiera och förstå samhälleliga konflikter.

I de samhällsvetenskapliga målen är de centrala verben inte lika förbundna med språkliga aktiviteter, som i målen för naturvetenskap och matematik, utan lika mycket inriktade mot kognitiva processer. Denna skillnad i målbeskrivningen uppenbaras genom ett relativt alldagligt ordförråd i de samhällsvetenskapliga lärobokstexterna, vilket denna delstudie har visat. I naturvetenskap och matematik används ett mer specialiserat uttryckssätt, vilket tydligt ytrar sig i indexlistorna. Detta ordförråd tillhör inte vardagsspråket och måste därför läras in, övas och prövas i undervisningen.

De mål som är specifika för samhällsvetenskap borde bidra till att texterna i samhällsvetenskap är mer resonerande och reflekterande till sin karaktär än naturvetenskapliga texter. Samhällsvetenskapliga texter borde därför mer likna berättelser till sin utformning. Fokusstudie 7.2.2.1 pekar mycket riktigt mot att ordförrådet i samhällsvetenskap också förekommer i romaner.

7.3 Typiskt ordförråd i naturvetenskap: sammanfattning och diskussion

7.3.1 Förhållandet till ordförråd i matematik och samhällsvetenskap

Det tydligaste resultatet i jämförelsen mellan läroböckerna i naturvetenskap och läroböckerna i de andra ämnesfacken är att de mest typiska orden i naturvetenskap till största delen är ämnesrelaterade ord och termer, som är relativt främmande i vardagsspråket. Dessutom varierar den allra mest karakteristiska naturvetenskapliga vokabulären i större utsträckning mellan högstadiet och gymnasiet, d.v.s. de högst placerade orden på indexlistorna för naturvetenskap är mer knutna till stadierna än orden på listorna i matematik och samhällsvetenskap. Eftersom det typiska naturvetenskapliga ordförrådet växer mer i stadiövergången till gymnasiet kan man tänka sig att nya ämnesområden och strategier läggs till på gymnasiet. I övriga ämnesfack återkommer i stället oftare ämnesområden och talas om på liknande vis på båda stadierna.

I jämförelsen med matematik framkommer att det typiska ordförrådet i naturvetenskap är betydligt mindre än det i matematik. Dessutom visar indexen att de allra mest typiska orden i matematik i högre grad återkommer i texterna än motsvarande ord i naturvetenskap. Matematikorden är också oftare längre än de naturvetenskapliga orden.

Den mycket höga andelen typiska substantiv på matematiklistorna talar för att böckerna innehåller få utredande ämnestexter med predicerande utsagor om deltagarna i texterna. Ämneskunskapen som lärs ut är praktiskt inriktad och föregås av illustrerande exempeluppgifter, där eleverna visas hur uppgifterna löses. Läroböckerna i matematik fungerar som facit, och är inte underlag för diskussion och argumentation.

När matematik ska läras ut som ett ämne i sig, måste det appliceras på något konkret som elever är bekanta med. Läroböckerna innehåller därför såväl ämnesspråk, d.v.s. språk som beskriver och formar ämneskunskaper, som språk från det fält som det matematiska ämnesspråket appliceras på, d.v.s. vardagen.

I jämförelsen med samhällsvetenskap kan konstateras att den typiska vokabulären i naturvetenskap och samhällsvetenskap består av i stort sett lika många ord, både på högstadiet och på gymnasiet. De ordlistor som faller ut i denna jämförelse är förhållandevis korta, vilket betyder att de båda ämnesfacken har en stor gemensam vokabulär, utanför indexlistornas sfär (som ju påvisar skillnader). De typiska orden för ämnesfacken skiljer sig inte heller särskilt mycket åt vad gäller ordklassstillhörighet.

I styrdokumentet (se avsnitt 7.2.2.3) beskrivs de olika ämnesfacken på olika vis. I målen för samhällsvetenskap är t.ex. de centrala verben inte lika förbundna med språkliga aktiviteter som i naturvetenskap och matematik, utan lika mycket inriktade mot kognitiva processer. Denna skillnad i målbeskriv-

ningen slår igenom på indexlistorna genom ett relativt alldagligt ordförråd i de samhällsvetenskapliga lärobokstexterna. Genom fokusstudie 7.2.2.1 bekräftas att kunskap om centrala samhällsvetenskapliga begrepp kan inhämtas genom att läsa dagstidningar, se på nyheterna, delta i debatter och till och med genom att läsa romaner.

Det typiska naturvetenskapliga ordförrådet utökas också mycket mer än den samhällsvetenskapliga vokabulären i stadiövergången till gymnasiet. Dessutom visar fokusstudie 7.2.2.1 att det typiska naturvetenskapliga ordförrådet blir mycket mer specifikt i stadiövergången, i förhållande till språket i A-texter. Inräknat att böckerna i samhällsvetenskap redan på högstadiet uppvisar en förhållandevis vardaglig lexikal repertoar, ligger det nära till hands att sluta sig till att de naturvetenskapliga ämnena är betydligt mer lexikalt krävande, såväl på högstadiet som på gymnasiet.

7.3.2 Stadiövergången mellan högstadiet och gymnasiet

Ordlistorna för högstadiet och gymnasiet i de naturvetenskapliga ämnena visar att andelen adjektiv och particip som tillhör det typiska ordförrådet är större på gymnasiet i samtliga ämnen. Denna utveckling stämmer överens med undersökningarna som pekar på högre nominalkvot och mer utbyggda nominalfraser i gymnasietexterna (se avsnitt 6.4.2 och 8.4). Dessutom är fler av orden på gymnasielistorna långa.

De separata ordlistorna för de naturvetenskapliga ämnena visar också på några ämnesspecifika stadieskillnader. I biologi verkar exempelvis högstadiéböckerna innehålla ett tydligt läsartilltal, med första och andra personens pronomen, vilket avtar i och med övergången till gymnasiet. I både biologi och fysik innebär stadiövergången i många fall att vardagliga ordval ersätts av en mer teknisk vokabulär, och i fysik och kemi visar ordlistorna på gymnasiet en tydlig orientering mot matematiska beräkningar.

7.3.3 Förhållandet till ordförråd i akademiska texter

I jämförelsen mellan naturvetenskapliga och akademiska texter placerar sig många vanliga formord, som *som*/HP, *den*/DT och *än*/KN, högt upp på naturvetenskapslistan. Även de mest typiska akademiska orden är i många fall både korta och vanliga, som *andra*/JJ, *jag* och *min*/PS.

Ser man till hela den akademiska listan innehåller den dock en betydligt högre andel långa ord, samt förhållandevis många particip, en ordklass som brukar förknippas med ett informationstätt och skriftspråkligt språkbruk (jfr

Halliday 1998). Inom SFL räknas particip till grammatiska metaforer, eller inkongruent språkbruk, eftersom de övertar andra ordklassers prototypiska användningsområden, vare sig de används substantiviskt eller adjektiviskt (jfr resonemang i Magnusson 2010: 39f. och Nygård Larsson 2011: 138ff.). Particip bidrar också till ökande nominalitet i texter och ingår i täljaren i beräkningen av nominalkvot. Enligt Josephsson, Melin och Oliv (1990: 40) är particip ”den ordklass som har störst förtätningsförmåga”.

Eftersom den akademiska korpusen saknar naturvetenskapliga texter är det svårt att veta om resultaten av jämförelsen beror på stadieskillnaden eller ämnesskillnaden mellan texterna. I fokusstudie 7.2.2.2, där samhällsvetenskapliga lärobokstexter jämförs med akademiska texter förändras dock inte listan över typiska akademiska ord särskilt mycket. Participen är vanligare än i lärobokstexterna även i denna jämförelse.

Ett tydligt resultat är att listan över typiska ord i samhällsvetenskapliga lärobokstexter, i förhållande till akademiska texter, är kortare än motsvarande lista för naturvetenskapliga texter. Det betyder att fler av de vanliga orden i akademiska texter förekommer i samhällsvetenskapliga lärobokstexter, och att elever på samhällsskligt program har en bättre chans att möta det akademiska ordförrådet i skolans ämnestexter.

7.3.4 Förhållandet till ordförråd i berättande texter

När lärobokstexterna i naturvetenskap jämförs med berättande texter utmärker de sig på flera sätt. Naturvetenskapstexterna innehåller många ovanliga, långa, ämnesrelaterade ord, något som Danielsson (1975: 92) kopplar samman med en hög andel substantiv. Mycket riktigt är också substantiv (och particip) mer utmärkande för de naturvetenskapliga texterna. Dessutom är verben typiskt statiska med icke-agentiva subjekt.

De berättande texterna kännetecknas i stället av korta och vanliga ord. Särskilt tydligt är de många pronomen och interjektioner som hamnar högt på listan och samsas med dynamiska, viljestyrda verb och substantiv knutna till människors utseende och uttryck. De berättande texterna är mer uppbyggda kring typiska verb än övriga texter i studien. Även interjektioner är betydligt mer prominenta än i övriga texttyper.

8

DELSTUDIE: TYPISKA NOMINALFRASER

Under lång tid har forskare framhållit olika nominala uttrycks avgörande roll i skolspråket (se t.ex. Halliday och Martin 1993; Halliday 1998; Schleppegrell 2004; Magnusson 2008, 2010; Biber och Gray 2010). Inom traditionen för SFL förklaras detta med att undervisningsspråket genom åren har utvecklat särskilda resurser för att synliggöra kunskap och strukturera information. Resurserna för att uppnå dessa mål är framför allt informationspackning, abstraktion och teknikalisering, vilka alla språkligt åstadkoms genom utbyggda och komplexa nominalfraser.

För att strukturera information betonas särskilt nominaliseringars betydelse för utbyggnad av nominalfraser. Nominaliseringar bidrar till ett ännu mer komprimerat språk än prototypiska substantiv, eftersom de medför att andra saker än ting kan kvantifieras, klassificeras och beskrivas som substantiv, vilket är fördelaktigt för att bygga upp vetenskapliga teorier. I jämförelse med andra texttyper noterar Westman (1974) ett överskott på verbalsubstantiv (som är ett slags nominaliseringar) i lärobokstexter, och Magnusson (2010) visar på ett tydligt samband mellan en hög andel nominaliseringar och långa nominalfraser i gymnasieelevtexter.

Bland olika fraser beskrivs nominalfrasen av Melin och Lange (2000) som den viktigaste i textanalytiskt hänseende, eftersom den är ”byggbar” (se avsnitt 8.1 för nominalfrasers uppbyggnad). Vidare har Biber och Gray (2010) påvisat registerskillnader i hur olika nominalfraser byggs ut med olika bestämmningar.

Om en nominalfras i en mening innehåller mycket information kommer hela meningen att innehålla mycket information. Sådana förtätade meningar är mer krävande att processa än meningar där samma informationsmängd överförs på ett mer mångordigt vis (Melin 2007). Informationstäta nominalfraser bidrar till ökade värden på de stilistiska skalorna *svårighet* och *specifikation* (Melin och Lange 2000). Dessutom kräver de en viss förtrogenhet med en skriftlig framställningsform, eftersom de är vanligare i skriftspråk (Einarsson 1978; Hellspong och Ledin 1997; Biber 1988; Magnusson 2008).

Inom stilistiken sammanfattas annars ofta textuella särdrag som kännetecknar informationspackning under etiketten *nominal stil*, ett stildrag som anses ge texter ett stelt, opersonligt och distanserat intryck (Einarsson 1978). I denna anda talar exempelvis Hellspong och Ledin (1997: 78) om nominalitet och verbalitet som en *lexikogrammatisk dimension*, i vilken olika språkdrag samsas och tillsammans kan bidra till att karakterisera en text på en skala med motpoler. Följande lexikogrammatiska särdrag utmärker nominala texter: (i) många substantiv på bekostnad av verb och pronomen, (ii) många långa nominalfraser, (iii) många attributiva prepositionsfraser, (iv) få bisatser och (v) långa meningar i regel orsakade av långa nominalfraser, samt många samordningar och uppräkningsfraser.

8.1 Nominalfrasers funktion och uppbyggnad

Enligt såväl SAG (del 3) som *Svenska Akademiens språklära* (Hultman 2003) är nominalfraser centrala för grammatiken, och därför en viktig del i beskrivningen av språket. Huvudordet i en nominalfras är typiskt sett ett substantiv (NN), men det kan även utgöras av ett egennamn (PM) eller av ett (substantiviskt) pronomen (PN). Gemensamt för nominalfraser är att de fungerar på samma sätt som substantiv i syntaktiskt avseende (Hultman 2003: 204). De används också, precis som substantiv, till att peka ut referenter: levande varelser, föremål, substanser eller abstrakta företeelser som händelser, tillstånd eller egenskaper (SAG, del 3: 11).

Enligt SAG (del 3: 132ff.) har nominalfraser följande huvudsakliga syntaktiska funktioner:

1. Primärt satsled i sats eller verbfras (subjekt, objekt, direkt objekt, indirekt objekt, egentligt subjekt, subjektspredikativ, objektspredikativ),
2. Rektion i prepositionsfras (som i sin tur oftast är adverbial eller attribut) eller subjunktionsfras (som i sin tur oftast är predikativ),
3. Attribut i nominalfras (possessivattribut, kvantitetsattribut, appositionellt attribut).

Ur strukturell synpunkt är en väsentlig egenskap hos nominalfraser att de kan konstrueras rekursivt; deras attribut är ofta själva nominalfraser, eller innehåller nominalfraser.

Attributen till nominalfraser som har ett substantiv som huvudord, NN-nominalfraser, kan vara framförställda eller efterställda huvudordet. I de flesta NN-nominalfraser ordnas bestämningarna enligt det schema som visas i tabell 8.1, och som kommer från SAG (del 3: 13). I tabellen har några exempel på nominalfraser ur läroboksmaterialet placerats in i modellen.

Framförställda attribut			Substantiv	Efterställda attribut
Definit attribut	Kvantitetsattribut	Adjektivattribut		
	<i>alla</i>		<i>aktiviteter</i>	<i>som sker där</i>
<i>denna</i>		<i>enorma</i>	<i>mängd</i>	<i>av småkrafter</i>
<i>de</i>	<i>två</i>	<i>gemensamma</i>	<i>elektronerna</i>	<i>i området omkring och mellan kärnorna</i>

Tabell 8.1: Nominalfrasens vanliga ordföljd med exempel ur gymnasieläroböcker i biologi, fysik och kemi

De olika typerna av attribut fyller olika funktioner. Närmare bestämt bidrar definitiva attribut, som t.ex. possessivattribut, till att mottagaren unikt kan identifiera referenten i fråga (antingen i verkligheten eller i textvärlden),⁴³ medan kvantitetsattribut finns till för att ange antal eller mängd. Resten av leden, adjektivattribut, huvudord och efterställda attribut, har tillsammans till uppgift att närmare beskriva referenten, för att mottagaren ska kunna föreställa sig den för sitt inre (SAG, del 3: 13). De vanligaste typerna av efterställda attribut är prepositionsfraser och relativa bisatser.

NN-nominalfraser kan växa sig komplexare än nominalfraser med egennamn eller pronomen som huvudord (PM- respektive PN-nominalfraser). Vad gäller PM-nominalfraser är det enligt SAG (del 3: 36) ”ojämförligt vanligast (...) att egennamnet bildar nominalfras ensam, utan bestämningar”. Men när PM-nominalfraser byggs ut med bestämningar, görs det på liknande vis som för NN-nominalfraser. Skillnaderna i typschemat för PM-nominalfraser består dels i att adjektiviska framförställda attribut kallas ”deskriptiva attribut”, dels i att kvantitetsattribut inte tas upp (egennamn betecknar ju oftast en unik referent, som inte behöver kvantifieras).

⁴³Definita nominalfraser kan även ha samma struktur som nominalfraser med substantiv som huvudord, men sakna själva substantivet, som t.ex. ”de särskilt utvalda”. Se vidare SAG (del 3: 41ff.).

Även när det gäller PN-nominalfraser är det vanligast att de konstrueras utan bestämmingar. I den mån det substantiviska pronomenet tar bestämmingar sker det i samma ordning som för NN-nominalfraser, men med skillnaden att kvantitets- och adjektivattribut placeras efter huvudordet. Dessutom benämns definitiva attribut ”emfatiska attribut”.

8.2 Nominalfraser av olika komplexitet

Inom forskningsområdet svenska som andraspråk har s.k. *performansanalys* av inlärares språk bidragit till en modell för hur förmågan att bygga ut nominalfraser vanligen utvecklas, från nybörjarnivå till avancerad inlärnivå. Tabell 8.2 är en sammanställning av nominalfrasutvecklingen som presenteras av Flyman Mattsson och Håkansson (2011), där interimspråkliga drag utelämnats.

	→	→
Huvudord	Huvudord + utbyggnadsord	Fler utbyggnadsord
Enkla NP	Utbyggda NP	Även komplexa NP
Få, överanvända pronomen	Få pronomen	Växling mellan pronomen och substantiv
Få adjektiv	Fler adjektiv	Ännu fler adjektiv
Få prepositioner	Mer prepositions- variation	Ännu mer prepositionsvariation

Tabell 8.2: Performansanalys av nominalfraser. Utvecklingsgång från vänster till höger

Anward (1989) presenterar några uppställningar över i vilken ordning förstaspråkstalare utvecklar en ”repertoar av uttrycksmedel”. Utvecklingsgången för vilka konstruktioner som konstituerar satsdelarna subjekt och objekt ser ut så här:

1. (a) Pronomen
(b) Substantiv
2. Nominalfras
3. Bisats

Dessa båda modeller över språkutveckling representerar ett slags komplexitetsskala som kan vara nyttig att ha i bakhuvudet vid tolkningen av resultaten av denna delstudie.

8.3 Typiska nominalfraser: studiens upplägg

I undersökningen analyseras nominalfraser ur tre aspekter, vilka kan beskrivas som täthet, längd och struktur. I ett första steg beräknas antal nominalfraser per mening i texterna, d.v.s. nominalfrastäthet.⁴⁴ Detta normaliserade frekvensmått anger alltså hur nominalfrastunga meningarna i de undersökta texterna generellt sett är och är ett sätt att undersöka syntaktisk komplexitet i en text. Ju mer packad informationen i en mening är, desto mer komplex tenderar texten att vara (Biber och Gray 2010).

Ett annat mått på packningsgrad i texter är nominalfrasernas medellängd (Melin och Lange 2000).⁴⁵ Nästa steg är således att beräkna nominalfrasernas längd, d.v.s. bestämma hur många ord som ingår i nominalfraserna (små skiljetecken räknas inte).

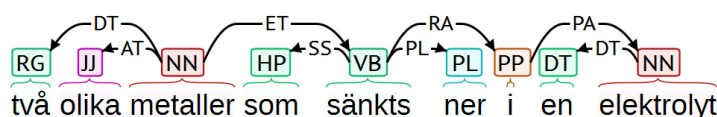
När man studerar nominalfraser kan man även notera om de är höger- eller vänsterförgrenade, d.v.s. om bestämningarna står till höger eller vänster om huvudordet. Bestämningar till höger ger en mer lättläst text och ledigare stil

⁴⁴I studier av engelska används ibland ”nounphrase density” för antal attribut till nominalfraser (jfr t.ex. Bell, McCarthy och McNamara 2012). Jag använder alltså motsvarande svenska term på annat vis.

⁴⁵Författarna anser att endast nominalfraser ”inom satsen” ska beräknas för att få ett rättvist mått på packning. Relativsätser som bestämning till en nominal i en överordnad sats ska enligt deras modell alltså inte räknas. Ökad packningsgrad (eller informationstäthet eller nominalitet) är en konsekvens av flera samverkande lexikogrammatiska drag och ligger alltså på en högre stilistisk abstraktionsnivå än enskilda lexikogrammatiska stildrag. Eftersom jag försöker hålla mig på en lexikogrammatisk nivå i beskrivningen av det naturvetenskapliga språket, väljer jag att primärt koncentrera mig på nominalfraser, och inte på packning. Av denna anledning väljer jag också att inte följa rådet för beräkning av packningsgrad, utan i stället hålla mig till SAG:s beskrivning av nominalfrasers uppbyggnad.

än bestämningar som kommer före huvudordet, vilka mer bidrar till att packa information (Melin och Lange 2000). I ett tredje steg har jag valt att undersöka nominalfrasernas syntaktiska uppbyggnad med hjälp av indexprincipen, eftersom mer detaljerade strukturer säger mer om nominalfrasernas verkliga utseende.

Eftersom nominalfraser kan innehålla andra, mindre nominalfraser t.ex. i prepositionsfrasbestämningar, har jag valt att bara räkna nominalfraser med maximal vidd, och inte sådana som på ett eller annat sätt är avhängiga andra nominala huvudord. Nominalfraser som ingår i bestämningar till nominala huvudord räknas alltså inte. I figur 29 visas ett förtydligande exempel på en nominalfras ur en lärobok i kemi för högstadiet. Som synes är hela frasens huvudord "metaller" (eftersom detta inte är beroende av någon annan nod i frasen). Dessutom är "elektrolyt" huvudord i en nominalfras som är rektion i en prepositionsfras, som i sin tur ingår i ett efterställt bisatsattribut till den överordnade nominalfrasens huvudord. Denna underordnade nominalfras räknas alltså som ingående i den större nominalfrasen, men inte för sig.



Figur 29: Exempel på en nominalfras från en lärobok i kemi för högstadiet

Jag har valt detta angreppssätt av två anledningar. För det första vill jag undvika att räkna attributiva nominalfraser flera gånger. På det sätt jag räknar ingår alla nominalfraser i beräkningarna av längd och struktur *en* gång, då underordnade påverkar såväl längden som strukturen av överordnade. För det andra är nominalfraser i attributiv ställning relativt korta. Om sådana, korta, underordnade nominalfraser skulle tillskrivas lika stor vikt i längdberäkningen som de nominalfraser de är en del av, skulle den genomsnittliga nominalfraslängden förkortas. I så fall finns en risk att underskatta texttypens benägenhet att bygga ut nominalfraser, vilket i själva verket kan vara något utmärkande för ämnet eller stadiet.

8.4 Typiska nominalfraser: resultat

8.4.1 Nominalfrastäthet

Alla exakta mätresultat för nominalfrastäthet återfinns i tabell E.24 i appendix.

Figur 30 visar till att börja med en översikt av det genomsnittliga antalet nominalfraser per mening i hela läroboksmaterialet och i ämnesfacken. I figuren har värden för berättande och akademiska texter lagts in som referenslinjer. Som synes ökar NP-graden mellan högstadiet och gymnasiet i samtliga fall, en ökning som är starkt statistiskt signifikant ($p < 0,001$).⁴⁶

Jämför man ämnesfacket naturvetenskap med hela läroboksmaterialet ser man att nominalfrastätheten både är konstant högre i naturvetenskap och ökar kraftigare mellan högstadiet och gymnasiet. Den kraftigare ökningen i naturvetenskap gäller även i jämförelse med samhällsvetenskap och matematik, även om värdena för naturvetenskap och samhällsvetenskap egentligen inte är så olika. På högstadiet är nominalfrastätheten i samhällsvetenskapliga texter något högre än i naturvetenskapliga texter ($p = 0,046$), men på gymnasiet är förhållandet det omvända ($p = 0,010$), vilket medför att nominalfrastätheten på gymnasiet är allra högst i det naturvetenskapliga ämnesfacket.

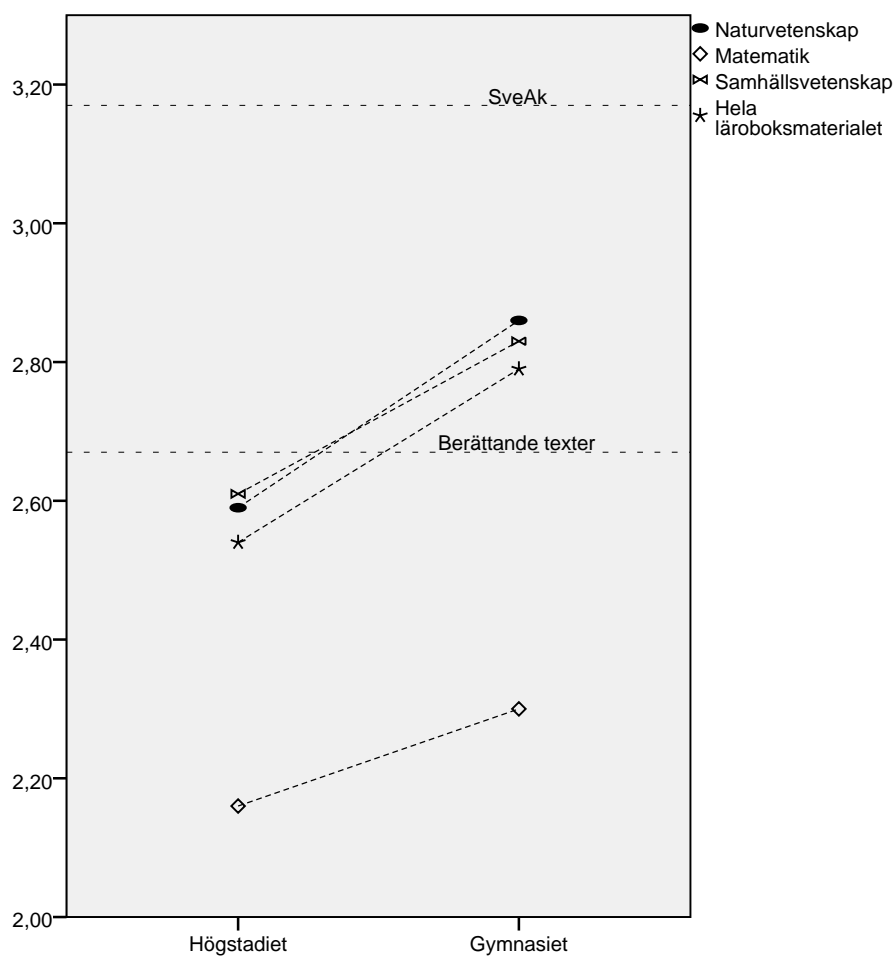
Matematik utmärker sig väsentligt i jämförelsen med mycket låga värden över stadierna, ständigt under värdet för de berättande texterna. För övriga ämnesgrupperingar gäller att nominalfrastätheten för de berättande texterna passeras i övergången från högstadiet till gymnasiet.

Figur 31 visar den ämnesspecifika utvecklingen av nominalfrastätheten i naturvetenskapliga och samhällsvetenskapliga ämnen. I de naturvetenskapliga ämnena sker en signifikant ökning i stadiövergången i samtliga ämnen. Brantast sker ökningen i biologi, som är det naturvetenskapliga ämne som har klart flest NP per mening på gymnasiet. Faktum är att biologi på gymnasiet ligger anmärkningsvärt mycket närmare värdet för de akademiska texterna än något annat ämne. Den relativt branta ökningen i biologi bidrar till en tydligare differentiering mellan de naturvetenskapliga ämnena på gymnasiet än på högstadiet.

På högstadiet är alltså inte skillnaden mellan de naturvetenskapliga ämnena lika påtaglig som på gymnasiet. Skillnaden mellan fysik och kemi är inte ens statistiskt belagd ($p = 0,104$), och skillnaden mellan biologi och kemi är inte starkt signifikant ($p = 0,006$).

För de enskilda naturvetenskapliga ämnena gäller vidare att biologi ligger starkt signifikant över hela läroboksmaterialet både på högstadiet och på gymnasiet (se figur 31a och 30). Kemi ligger signifikant över hela läroboksmate-

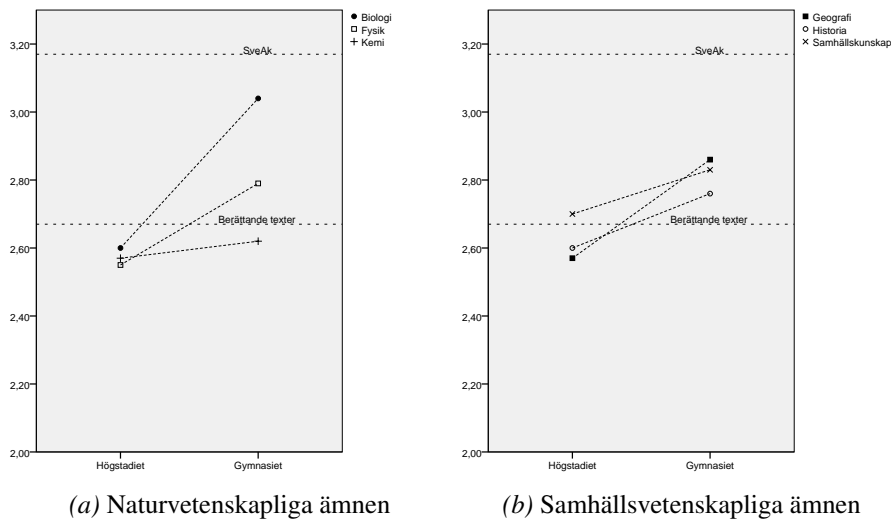
⁴⁶Som redan nämnts beräknas p -värdet med ett t-test. Gränsen för 'statistisk signifikans' ligger på 0,05 och gränsen för 'stark statistisk signifikans' på 0,001.



Figur 30: Översiktlig utveckling av genomsnittligt antal nominalfraser per mening i över stadierna

rialet på högstadiet och starkt signifikant under på gymnasiet. På gymnasiet överstiger värdet för kemi faktiskt inte det för de berättande texterna. Fysik är det ämne som följer utvecklingen i hela läroboksmaterialet tättast; varken på högstadiet eller på gymnasiet särskiljer sig ämnet åt något håll.

Går man vidare till de samhällsvetenskapliga ämnena enskilt finner man att samtliga ämnen ökar i nominalfrastäthet mellan högstadiet och gymnasiet (se figur 31b). Vidare är samhällskunskap ensam om att överskrida värdet för berättande texter redan på högstadiet, och den relativt branta utvecklingslinjen



Figur 31: Ämnesspecifik utveckling av genomsnittligt antal nominalfraser per mening över stadierna

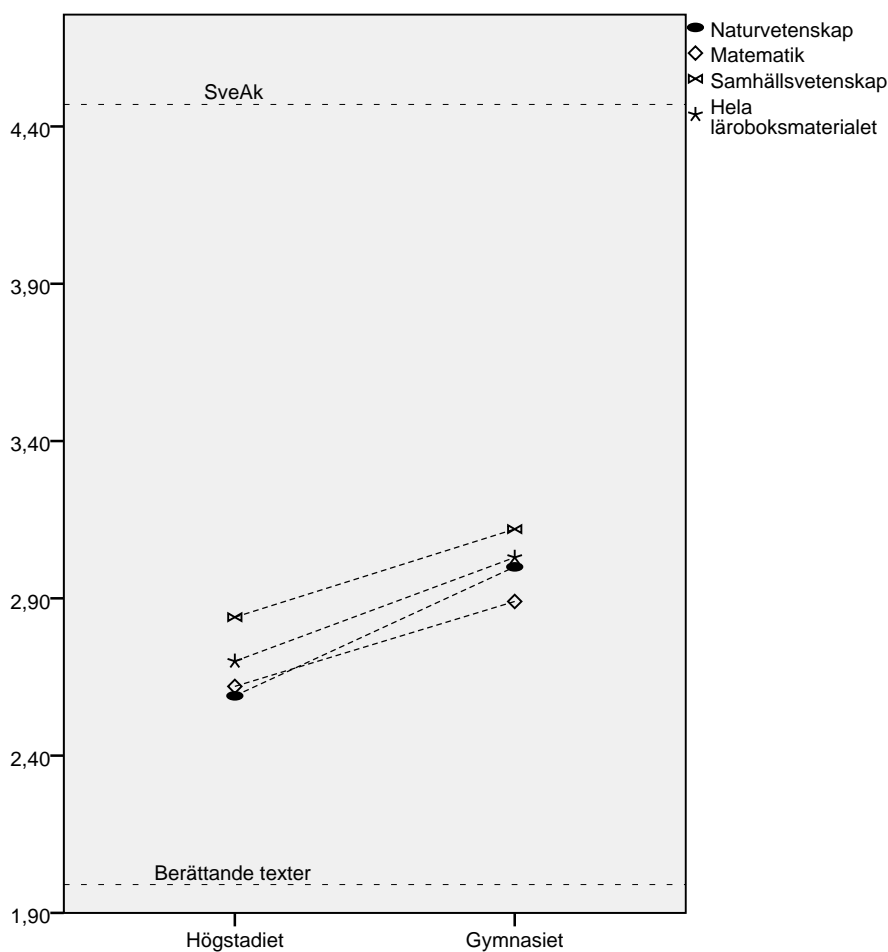
för geografi påminner om den för biologi. Geografi arbetar sig också uppåt mot värdet för de akademiska texterna, om än inte med samma tydlighet som biologi.

En viktig observation i jämförelsen mellan de naturvetenskapliga och samhällsvetenskapliga ämnena är att de naturvetenskapliga ämnena ligger närmare varandra på högstadiet och längre ifrån varandra på gymnasiet. De samhällsvetenskapliga ämnena liknar snarare varandra lika mycket på högstadiet som på gymnasiet. Denna mer konstanta ämnesspråkliga spridning i de samhällsvetenskapliga ämnena är signifikant högre än den som de naturvetenskapliga ämnena uppvisar på högstadiet.

8.4.2 Nominalfraslängd

Alla exakta mätresultat för nominalfraslängd återfinns i tabell E.25 i appendix.

En översikt över utvecklingen av NP-längder i hela läroboksmaterialet och i de olika ämnesfacken syns i figur 32. I samtliga ämnesfack ökar nominalfrasernas längd starkt signifikant mellan högstadiet och gymnasiet, men denna ökning är, precis som i fråga om NP-täthet, kraftigast i naturvetenskap.



Figur 32: Översiktlig utveckling av genomsnittliga NP-längder över stadierna

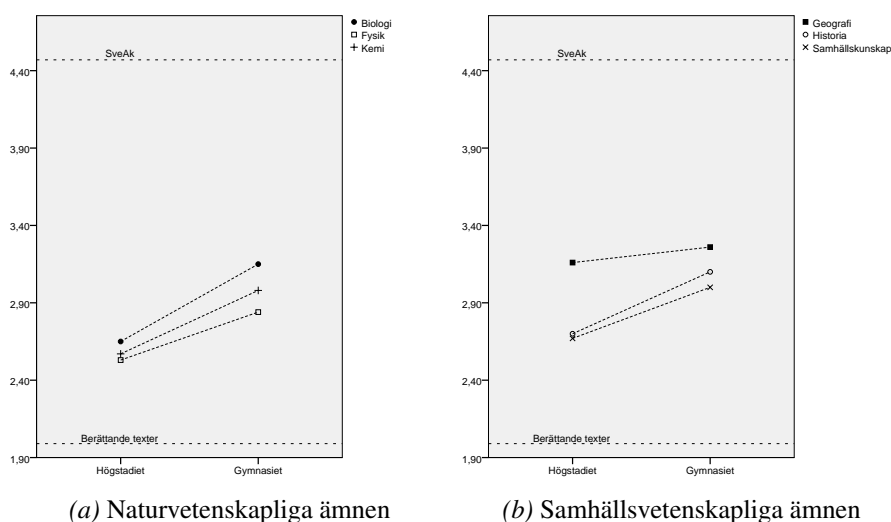
På högstadiet är nominalfraserna i naturvetenskap kortare än de i matematik ($p = 0,023$) och samhällsvetenskap. Detta förhållande mellan nominalfraslängden i naturvetenskap och samhällsvetenskap kvarstår på gymnasiet. Däremot passerar naturvetenskapliga texter nominalfraslängden i matematik på gymnasiet.

Som synes skiljer sig naturvetenskap och samhällsvetenskap mer åt i fråga om NP-längd än om NP-täthet. Nominalfraserna i samhällsvetenskap är faktiskt avsevärt mycket längre, framför allt på högstadiet, men också på gymnasiet.

Man kan också notera att matematikvärdena inte är lika slående låga i fråga om längd som i täthet; nominalfraslängden i matematik ligger snarare i paritet med värdena för övriga ämnesfack.

Nominalfraserna i de berättande texterna är betydligt kortare än i lärobokstexterna, även på högstadiet. De akademiska texterna innehåller i stället avsevärt längre nominalfraser än läroböckerna.

I figur 33 visas den ämnesspecifika utvecklingen av NP-längder i naturvetenskap och samhällsvetenskap. Liksom för tätheten ökar värdena starkt signifikant för samtliga ämnen mellan högstadiet och gymnasiet.



Figur 33: Ämnesspecifik utveckling av genomsnittliga NP-längder över stadierna

Bland de naturvetenskapliga ämnena noteras att värdena för biologi ständigt ligger högre än värdena för fysik och kemi. Precis som för tätheten har kurvan för biologi starkast lutning, vilket bidrar att nominalfrasernas längd också skiljer sig mer mellan ämnena på gymnasiet än på högstadiet. På högstadiet är visserligen skillnaden mellan fysik och kemi signifikant, men inte starkt så ($p = 0,009$). Det är den däremot mellan samtliga naturvetenskapliga ämnen på gymnasiet.

Man kan också notera att fysik och kemi har bytt plats från tidigare täthetsjämförelse; i fråga om längd är fysik det naturvetenskapliga ämne som ligger lägst, både på högstadiet och på gymnasiet. I jämförelse med hela läroboksmaterialet gäller för samtliga de naturvetenskapliga ämnena att de ligger starkt signifikant under hela läroboksmaterialet på högstadiet (detta bekräftas enklast genom att betrakta figur 33a och 32 tillsammans). Förhållandet står sig

på gymnasiet för fysik och kemi. Biologi däremot klättrar i stadiövergången förbi hela läroboksmaterialet och ligger på gymnasiet starkt signifikant över, precis som i fråga om täthet.

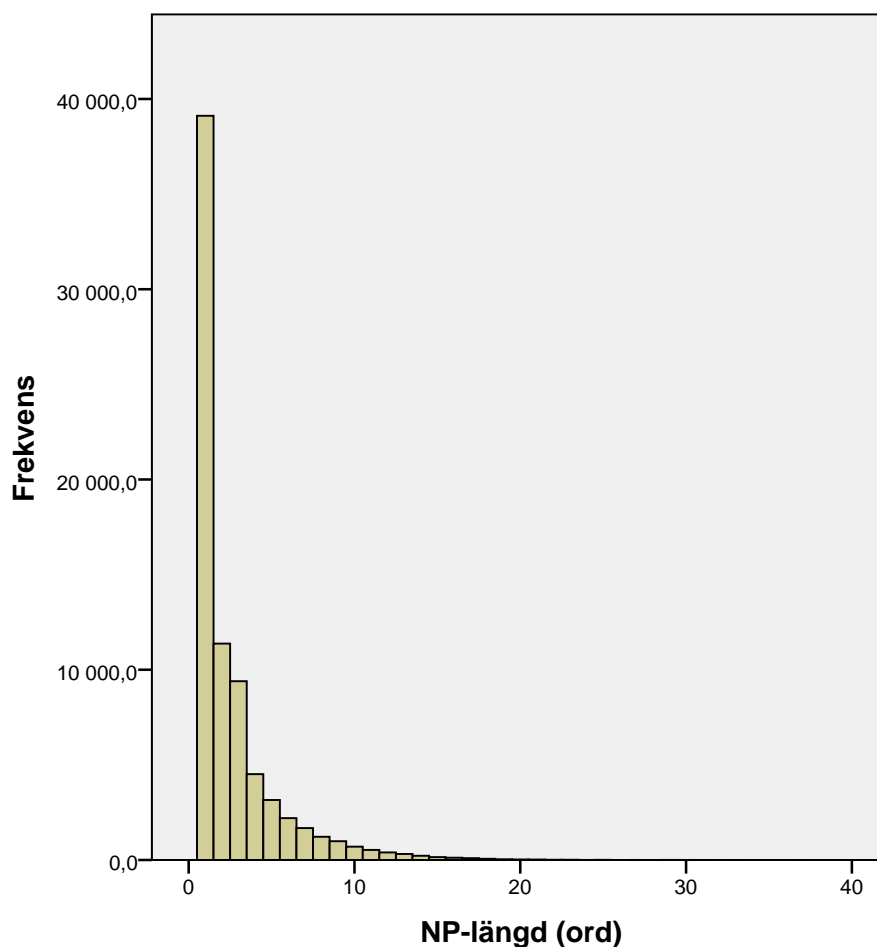
Nominalfraslängden ökar också i samtliga samhällsvetenskapliga ämnen mellan högstadiet och gymnasiet. Historia utmärker sig som ett ämne med relativt kraftigt förhöjda värden i gymnasiet. Geografi ligger väsentligt högre än övriga samhällsvetenskapliga ämnen genom hela skolgången. Särskilt noteras att nominalfraserna i geografi är betydligt längre än i övriga ämnen på högstadiet. Däremot är längdökningen i stadiövergången förhållandevis låg i geografi jämfört med i övriga ämnen.

De samhällsvetenskapliga ämnena följer inte heller i fråga om NP-längd det naturvetenskapliga mönstret med en tydligare individuell ämnesspråklig karakteristik på gymnasiet. Tvärtom ligger värdena för genomsnittlig nominalfraslängd i de samhällsvetenskapliga ämnena närmare varandra på gymnasiet än på högstadiet.

När man betraktar värdena för nominalfraslängderna i tabell E.25 (i appendix), lägger man märke till att standardavvikelseerna genomgående är högre än medelvärdena, vilket bara kan ske om distributionen av längderna inte är normalfördelad. I tabell 8.3 och figur 34 visas därför mer detaljerat hur längderna är fördelade. Exemplet visar visserligen bara längdfördelningen i biologitexter på högstadiet, men mönstret har konstaterats vara detsamma i samtliga ämnen.

Längd	Frekvens	Procent	Kumulativ procent
1	39 116	51,20	51,20
2	11 374	14,89	66,09
3	9 397	12,30	78,40
4	4 513	5,91	84,30
5	3 154	4,13	88,43
6	2 200	2,88	91,31
7	1 678	2,20	93,51
8	1 221	1,60	95,11
9	987	1,29	96,40
10	701	0,92	97,32
11	530	0,69	98,01
12	397	0,52	98,53
13	317	0,41	98,94
14	222	0,29	99,24
15	152	0,20	99,43
16	121	0,16	99,59
17	97	0,13	99,72
18	64	0,08	99,80
19	39	0,05	99,85
20	29	0,04	99,89
21	24	0,03	99,92
22	14	0,02	99,94
23	14	0,02	99,96
24	4	0,01	99,97
25	9	0,01	99,98
27	4	0,01	99,98
28	2	0,00	99,99
29	5	0,01	99,99
30	3	0,00	100,00
31	1	0,00	100,00
33	1	0,00	100,00
Totalt	76 390	100,00	100,00

Tabell 8.3: Distribution av NP-längder i biologiläroböcker för högstadiet



Figur 34: Distribution av NP-längder i biologiläroböcker för högstadiet

Exemplet från biologi bekräftar ett icke normalfördelat utfall, och histogrammet visar att längderna i stället i det närmaste är Zipffördelade. Av tabellen framgår att mer än hälften av nominalfraserna är av längden 1 ord och 95 % är mellan 1 och 8 ord. Även om det förekommer några mycket långa nominalfraser i ämnet, som den rekordlånga 33-ordsfras som visas i (62), är bara 2,7 % av nominalfraserna över 10 ord långa.

- (62) *Det finns* intressanta egenskaper i vilda arter som inte går att föra över till våra jordbruksgrödor med vanlig växtförädling eftersom det kan vara svårt eller omöjligt att korsade de vilda arterna med våra odlade växter.

I exemplet står nominalfrasen i rak stil och visas i den mening den förekommer. För exempel på fraser av mer typiska längder hänvisas till avsnitt 8.4.3 om nominalfrasuppbyggnad.

8.4.3 Nominalfrasuppbyggnad

Uppbyggnaden av nominalfraser presenteras i indexlistor över strukturer som utmärker ett visst material jämfört med ett annat. Metoden väger in strukturernas frekvens och distribution (där drag med jämn distribution gynnas). I de fall där undersökningsmaterialet är ett ämnesfack ser metoden till att spridningen över de ingående ämnena också är god (se vidare avsnitt 5.3 i metodkapitlet).

I exempel 63 visas ett exempel på hur en post i en indexlista över nominalfrasstrukturer kan se ut.

(63)	Index	F-ställda attr.	Huvudord	E-ställda attr.
	1,11	AT	NN	PP

För varje fras bokförs den ordklass till vilken huvudordet hör, vilken (primär) satsdel frasen utgör eller vilken satsdel frasen är en del av,⁴⁷ samt vilka olika typer av attribut huvudordet tar.⁴⁸ Bland efterställda attribut urskiljs t.ex. subjunktionsfraser, prepositionsfraser och infinitivfraser.⁴⁹ I exemplet visas strukturen för en substantivfras med adjektiv- och prepositionsattribut, som t.ex. nominalfrasen *vanliga ord med mycket hög frekvens*.

⁴⁷Av hänsyn till läsarna har jag valt att inte inkludera satsdelsinformationen i indexlistorna, då dessa uppgifter är svåröverskådliga. I stället kommenteras nominalfrastypernas syntaktiska bruk i löpande text då det bedöms vara befogat.

⁴⁸Då den syntaktiska annoteringen i Språkbanken inte skiljer mellan definitiva attribut och kvantitetsattribut, blir antalet analyskategorier för PM- och PN-nominalfraser lika många som för NN-nominalfraser. Den attributkategori som i SAG kallas "deskriptivt attribut" taggas i Språkbanken som AT (adjektivattribut), och "emfatiska attribut" taggas som DT (definita attribut).

⁴⁹Då mängden efterställda bisatser pekar på en viss komplexitet, ingår även en komplexitetsfaktor i undersökningen, och inte bara packningsvariabler. Visserligen analyseras inte bisatsernas interna struktur vidare, men tidigare studier har kommit fram till att nominala bestämningar på satsnivå inte tillnärmelsevis är lika avgörande för den totala syntaktiska komplexiteten i läroböcker och akademiska texter som bestämningar på frasstrukturell nivå (jfr Westman 1974; Danielsson 1975; Biber och Gray 2010).

Resultaten som presenteras här har erhållits med en tröskelnivå på minst 1 förekomst per 200 meningar. Detta gränsvärde valdes eftersom 200 meningar någorlunda motsvarar ett kapitel i en lärobok. Strukturerna i indexlistorna förekommer alltså förmodligen minst 1 gång i varje kapitel. De presenteras i fallande indexordning, vilket speglar en fallande typiskhet för det material som undersöks mot jämförelsematerialet.

I avsnittet jämförs till att börja med naturvetenskap med övriga ämnesfack i läroboksmaterialet, d.v.s. matematik och samhällsvetenskap. Dessa jämförelser görs stadievis, först för högstadiet och sedan för gymnasiet. Därefter jämförs naturvetenskap och de där ingående ämnena i stadieövergången mellan högstadiet och gymnasiet. Slutligen jämförs naturvetenskap med akademiska och berättande texter. Alla skillnader som redovisas i detta avsnitt är starkt statistiskt signifikanta, om inget annat anges.

8.4.3.1 *Naturvetenskapliga texter jämförda med matematiktexter*

Indexlistan för naturvetenskapliga texter jämfört med matematiktexter på högstadiet kan ses i tabell 8.4. Listan visar alltså konstruktioner som är mer utbredda i naturvetenskap än i matematik.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	2,39		NN	VB
2	2,09	AT	NN	
3	1,74	DT, AT	NN	PP
4	1,63		NN	
5	1,48	DT, AT	NN	
6	1,41		PN	
7	1,24	DT	NN	VB
8	1,19		NN	PP

Tabell 8.4: Indexlista över nominalfrasstrukturer i naturvetenskapliga lärobokstexter för högstadiet (jämfört med matematiktexter)

Med undantag för det enkla pronomenet (PN) har alla utmärkande konstruktioner substantiviskt huvud (NN). Samtliga konstruktioner används huvudsakligen som subjekt, förutom substantivet som bestäms av ett adjektivattribut, som främst används i objektsadverbial. Bland attributen återfinns såväl framförställda definitiva attribut (DT) och adjektivattribut (AT) som efterställda prepositionsattribut (PP) och bisatsattribut (VB).

I exempel (64) och (65) visas prov på de olika konstruktionerna med bisatsattribut, nummer 1 och 7.

- (64) Glykol, som har två OH-grupper, används som frostskyddsmedel och är giftig. (kemi)
- (65) Lerrörmasken är en havsborstmask som sitter fast på havsbotten. (biologi)

I (66) och (67) exemplifieras de olika konstruktionerna med prepositionsattribut, nummer 3 och 8.

- (66) Den hade en elektrisk ledning mellan sändare och mottagare. (fysik)
- (67) Kraven på större skördar ökade i takt med att befolkningen blev större. (kemi)

Exempel (68) visar ett substantiv bestämt av adjektivattribut (struktur nummer 2) som ingår i ett adverbial.

- (68) Det genetiska materialet förändras ständigt genom mutationer och naturligt urval. (biologi)

I jämförelse med den kompletterande omvända indexlistan (se tabell 8.5), som visar konstruktioner mer utbredda i matematik än i naturvetenskap, noteras att utbyggda nominalfraser är mer förekommande i de naturvetenskapliga ämnena. Matematiktexterna innehåller förvisso flera substantivfraser med definitiva attribut, men fraser innehållande adjektivattribut och bisatsattribut finns endast med på indexlistan för naturvetenskap.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	2,43		PM	
2	1,56	DT, DT	NN	
3	1,52	DT	NN	
4	1,27	DT	NN	PP

Tabell 8.5: Indexlista över nominalfrasstrukturer i matematiktexter för högstadiet (jämfört med naturvetenskapliga lärobokstexter)

I exempel (69), (70) och (71) visas prov de olika konstruktionerna med definitiva attribut (nummer 2–4) från matematiktexter för högstadiet. I exemplet som illustrerar två definitiva attribut kan påpekas att det andra i följd är ett kvantitetsattribut, enligt SAG:s modell.

(69) *Vilka är de fyra talen?*

(70) *Hur stora är dessa vinklar?*

(71) *Hyran höjdes med 2,50 kr per kvadratmeter.*

Konstruktionen med högst index i matematik består av ett enkelt egennamn (PM), och exemplifieras i (72) och (73).

(72) *Hur mycket ska Liam ge Theo för att de ska få lika mycket?*

(73) *Emma köper en läsk och en muffins till sig själv och till sina tre kompisar.*

Indexlistan för naturvetenskapliga texter jämfört med matematiktexter på gymnasiet kan ses i tabell 8.6. Faktum är att gymnasielistan uteslutande består av identiska poster som den för högstadiet, med undantag för ordningen på konstruktionerna från och med fjärde plats.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	2,31		NN	VB
2	1,89	AT	NN	
3	1,67	DT, AT	NN	PP
4	1,52	DT	NN	VB
5	1,41		PN	
6	1,38		NN	
7	1,29	DT, AT	NN	
8	1,26		NN	PP

Tabell 8.6: Indexlista över nominalfrasstrukturer i naturvetenskapliga lärobokstexter för gymnasiet (jämfört med matematiktexter)

Även indexlistan för matematiktexter jämfört med naturvetenskapliga texter på gymnasiet, som kan ses i tabell 8.7, bär många likheter med motsvarande högstadielista. Samtliga högstadietkonstruktioner finns kvar, och en enda konstruktion har tillkommit, nämligen det definitiva egennamnet på plats nummer 2. Fraser av denna typ kan ses i exempel (74) och (75).

(74) *Detta a -värde ger linjen L3.*

(75) *Vinkeln C är rät.*

Som framgår av exemplen rör det sig om hur man i matematik tilldelar variabelnamn åt fenomen. Dessa namn är visserligen inte lika prototypiska egennamn som t.ex. namn på personer, men likväl fullgoda egennamn enligt den definition som ges för ordklassen i SAG (del 2: 116): "[Ett egennamn] har ingen betydelse: det beskriver inte en typ eller en klass av referenter utifrån dessas beskaffenhet eller funktion. Egennamnet är i stället principiellt endast en beteckning, en etikett, på en unik referent". Den osäkerhet man eventuellt kan känna beträffande ordklasstilldelningen av variabelnamn hänger troligen samman med att fenomenet hör till ett specialiserat språkbruk.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	1,54		PM	
2	1,38	DT	PM	
3	1,37	DT	NN	
4	1,25	DT, DT	NN	
5	1,17	DT	NN	PP

Tabell 8.7: Indexlista över nominalfrasstrukturer i matematiktexter för gymnasiet (jämfört med naturvetenskapliga lärobokstexter)

8.4.3.2 Naturvetenskapliga texter jämförda med samhällsvetenskapliga texter

I detta avsnitt jämförs strukturer i naturvetenskap med strukturer i samhällsvetenskap. Nominalfrasstrukturer som i denna jämförelse är typiska för naturvetenskap på högstadiet presenteras i tabell 8.8. Denna indexlista innehåller tre nominalfrasstrukturer. Mest särskiljande är enkla pronomen, därefter kommer två substantivfraser, den ena bestämd av en efterställd bisats och den andra av ett framförställt adjektivattribut. De enkla pronomina används nästan uteslutande som subjekt, medan substantivfraserna även flitigt förekommer i objektiv ställning.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	1,21		PN	
2	1,19		NN	VB
3	1,14	AT	NN	

Tabell 8.8: Indexlista över nominalfrasstrukturer i naturvetenskapliga lärobokstexter för högstadiet (jämfört med samhällsvetenskapliga lärobokstexter)

Exempel (76) och (77) visar bisatsbestämda substantivfraser, enligt struktur nummer 2. Den första nominalfrasen ingår i ett adverbial och den andra utgör ett direkt objekt.

(76) *Förklaringen är att det blod som passerat tunntarmens kapillärnät har fyllts på med näringsämnen som har frigjorts under matspjälkningen.* (biologi)

(77) *Med kunskaperna som grund kan man åstadkomma modeller, som tar fasta på det intressanta i ett fenomen eller ett förlopp.* (fysik)

Adjektivbestämda substantivfraser, enligt struktur nummer 3, i kan ses i exempel (78) och (79).

(78) *I motsats till de lätta atomernas kärnor kräver de tyngre atomkärnorna tillförsel av energi när de steg för steg bildas av lätta atomkärnor.* (kemi)

(79) *Med laserljus kan man lokalisera ytliga tumörer.* (fysik)

Den kompletterande indexlistan för samhällsvetenskapliga högstadietexter jämfört med naturvetenskapliga kan ses i tabell 8.9. Samhällsvetenskapliga texter innehåller klart mer utbyggda nominalfraser än naturvetenskapliga texter, vilket yttrar sig i fler nominalfraser bestämda av definitiva attribut, adjektivattribut och prepositionsfrasattribut. Endast bisatsattribut faller ut till fördel för naturvetenskap i högstadijämförelsen.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	1,97		PM	
2	1,30	DT, AT	NN	
3	1,29	DT, AT	NN	PP
4	1,12	DT	NN	PP

Tabell 8.9: Indexlista över nominalfrasstrukturer i samhällsvetenskapliga lärobokstexter för högstadiet (jämfört med naturvetenskapliga lärobokstexter)

I exempel (80), (81) och (82) visas prov på de tre vanliga substantivfras-typerna ur samhällsvetenskapliga läroböcker för högstadiet. De motsvarar alltså posterna 2–4 på listan i tabell 8.9.

- (80) *Strejken varade i tolv dagar och hade invånarnas fulla stöd.*
(samhällskunskap)
- (81) *Det betydde att den största delen av landets befolkning var hungrig och missnöjd vid en avgörande tidpunkt i landets historia.* (historia)
- (82) *Längre mot öster bildar Karpaterna en båge mellan Alperna och Svarta havet.* (geografi)

Indexlistan för naturvetenskap på gymnasiet ser något annorlunda ut än högstadielistan (se tabell 8.10). Enkla pronomen och adjektivbestämda substantiv är inte längre typiska. Däremot finns den bisatsbestämda substantivfrasen kvar, och den har fått sällskap av ännu en substantivfras med bisatsattribut, av bestämd typ. Dessutom finns definitiva substantivfraser på listan.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	1,34		NN	VB
2	1,31	DT	NN	VB
3	1,11	DT	NN	

Tabell 8.10: Indexlista över nominalfrasstrukturer i naturvetenskapliga lärobokstexter för gymnasiet (jämfört med samhällsvetenskapliga lärobokstexter)

Vad gäller bruket kan påpekas att frasen som bestäms av både definit attribut och bisatsattribut oftast används i objektsadverbial, medan övriga två främst förekommer i subjektställning. (83) och (84) visar exempel på hur de två bisatsbestämda substantivfrastyperna, nummer 1 och 2, används i naturvetenskapliga läroböcker för gymnasiet.

- (83) *När gasen kommer ner i de fuktiga luftrören, bildas svavelsyrlighet som fräter på slemhinnorna.* (kemi)
- (84) *De har ritats så att en ruta motsvarar 1 N, och kraftpilarna är ritade i den riktning som dynamometrarna drar i gummibandet.* (fysik)

Indexlistan för samhällsvetenskapliga gymnasietexter jämfört med naturvetenskapliga kan ses i tabell 8.11. Majoriteten av frastyperna på den samhällsvetenskapliga gymnasielistan finns också på motsvarande högstadielista, men två konstruktioner har tillkommit: en substantivfras bestämd av två definitiva attribut (post 3), och en substantivfras bestämd av adjektiv- och prepositionsattribut (post 4). Fortfarande är de typiskt samhällsvetenskapliga nominalfraserna mer utbyggda än de naturvetenskapliga; frastyper med adjektiv- och prepositionsattribut finns endast på den samhällsvetenskapliga indexlistan. Men även olika slags definitiva nominalfraser är vanligare i samhällsvetenskapliga texter.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	2,13		PM	
2	1,30	DT, AT	NN	PP
3	1,13	DT, DT	NN	
4	1,11	AT	NN	PP
5	1,11	DT, AT	NN	

Tabell 8.11: Indexlista över nominalfrasstrukturer i samhällsvetenskapliga lärobokstexter för gymnasiet (jämfört med naturvetenskapliga lärobokstexter)

Det enkla egennamnet används lika ofta som subjekt som i platsadverbial. Den senare användningen exemplifieras i (85).

- (85) *Redan 1945, först i Jalta på Krim och senare i Potsdam söder om Berlin, träffades ledarna för USA, Storbritannien och Sovjetunionen för att kunna diskutera det framtida Europa.* (historia)

Den bestämda respektive den obestämda substantivfrasen med adjektiv- och prepositionsattribut (post 2 respektive 4) visas i exempel (86) och (87).

- (86) *De är en viktig del i nyhetsförmedlingen.* (samhällskunskap)
- (87) *Idag är landets urbaniseringsgrad ca 85 %, dvs så stor andel av vårt lands befolkning bor i tätorter (orter med fler än 200 invånare).*
(geografi)

Slutligen exemplifieras de två substantivfrastyper, på post 3 och 5, som saknar efterställda attribut i (88) och (89).

- (88) *Efter valet 1936 bildade de båda partierna också regering med ett omfattande social-politiskt program, folkhemspolitiken.* (historia)
- (89) *Mina klasskompisar började se mig som "feministen", som om det var min främsta egenskap, och det ansågs inte vara en speciellt positiv egenskap.* (samhällskunskap)

8.4.3.3 Biologitexter på högstadiet och gymnasiet

För att se om det finns några ämnesspecifika skillnader bland de naturvetenskapliga ämnena jämförs stadiövergångarna även inom ämnena. Bland nominalfrasstrukturerna i biologi visar det sig att pronomina nominalfraser med efterställt bisatsattribut är vanligare på högstadiet än på gymnasiet (se tabell 8.12).

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	1,22		PN	VB
2	1,12		NN	

Tabell 8.12: Indexlista över nominalfrasstrukturer i biologitexter för högstadiet (jämfört med för gymnasiet)

Konstruktionen förekommer oftast som subjekt, därefter även som direkt objekt och i objektsadverbial. Ett exempel på en nominalfras med denna struktur från biologi på högstadiet kan ses i (90).

- (90) *Det är t.ex. knappast någon som tror att en skog slutar fungera om den sista vitryggiga hackspetten i Sverige dör ut.*

Även substantiv som saknar bestämningar är mer vanliga i biologispråket på högstadiet än på gymnasiet ($p = 0,003$).

I jämförelsen mellan biologi på gymnasiet och på högstadiet finner man betydligt fler typiska strukturer (se tabell 8.13). Förutom det ensamma egenamnet som toppar listan, består listan nästan uteslutande av nominalfrasstrukturer med substantiviska huvudord och olika slags bestämningar. Till att börja med finns de flesta tänkbara kombinationer av framförställda attribut, men utan efterställda attribut, representerade. Bland dessa används strukturen där nominalfrasen bestäms av ett ensamt adjektivattribut oftast i objektsadverbial. Därefter följer funktionerna direkt objekt, subjekt och i platsadverbial. Övriga substantivfrasstrukturer på listan utan efterställda bestämningar intar oftast subjektposition.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	1,95		PM	
2	1,44	DT	NN	VB
3	1,41	AT	NN	PP
4	1,40	AT	NN	
5	1,36	DT, AT	NN	PP
6	1,36		NN	VB
7	1,36		PN	PP
8	1,27	DT	NN	PP
9	1,25		NN	PP
10	1,23	DT, AT	NN	
11	1,18	DT, AT	NN	VB
12	1,17	DT	NN	

Tabell 8.13: Indexlista över nominalfrasstrukturer i biologitexter för gymnasiet (jämfört med för högstadiet)

Efterställd bestämning i form av bisats återfinns också på flera platser på listan. När dessa innehåller framförställda attribut förekommer de i högre utsträckning i objektsadverbial än som subjekt. Exempel på var och en av konstruktionerna med efterställt bisatsattribut (post 2, 6 och 11) från biologi på gymnasiet visas i (91), (92) och (93).

- (91) *Våra celler har därför ett "bromssystem" som ser till att de i normalfallet inte delar sig.*
- (92) *Förklaringen är att det blod som passerat tunntarmens kapillärnät har fyllt på med näringsämnen som har frigjorts under matspjälkningen.*
- (93) *Hypofysen är en endokrin körtel som står i förbindelse med hypotalamus.*

Flera mönster hör till fraser som bestäms av en efterställd prepositionsfras. Denna struktur förekommer ensam, tillsammans med ett definit attribut, tillsammans med ett adjektivattribut, samt med både ett definit attribut och ett adjektivattribut. Den mest typiska är substantivfrasen med ett framförställt adjektivattribut, på post 3. Denna används oftast i objektsadverbial, därefter som objekt, och exemplifieras i (94).

- (94) *Aorta viker sedan av nedåt och förgrenar sig i artärer och arterioler som leder till olika kapillärnät i den nedre delen av kroppen.*

En annan typisk prepositionsfrasbestämd fras, på post 7, innehåller ett pronominellt huvud och används övervägande som subjekt. Denna fras exemplifieras i (95).

- (95) *Många av bakterierna i havet kan utföra fotosyntes, de kallas för cyanobakterier (tidigare kallades de för blågröna alger).*

På det hela taget märks alltså utvecklingen mot allt mer utbyggda nominalfraser på gymnasiet mycket tydligt i biologispråket.

8.4.3.4 Fysiktexter på högstadiet och gymnasiet

Fysiktexterna på högstadiet utmärks av fler nominalfraser med pronominellt huvud, antingen ensamma eller med en efterställd bisatsbestämmare (se tabell 8.14). Båda konstruktionerna fungerar huvudsakligen som subjekt, och den med bisatsbestämmare är även ofta objekt.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	1,55		PN	VB
2	1,18		PN	

Tabell 8.14: Indexlista över nominalfrasstrukturer i fysiktexter för högstadiet (jämfört med för gymnasiet)

Två exempel på nominalfraser med pronomen och bisatsbestämmare visas i (96) och (97), det första i subjektsposition och det andra i objektsställning.

- (96) *De som släpper fram ström bra kallas elektriska ledare och de som inte släpper fram ström kallas elektriska isolatorer.*

- (97) *De hörde något som liknade fågelsång.*

Även om indexlistan för fysikspråket på gymnasiet (se tabell 8.15) är ganska lik den för gymnasiespråket i biologi, kan vissa ämnesspecifika drag noteras.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	2,17		PM	
2	2,10	DT	PM	
3	1,26	DT, AT	NN	PP
4	1,35	DT	NN	PP
5	1,29		NN	PP
6	1,23	DT, AT	NN	
7	1,22	DT	NN	
8	1,20	DT, AT	NN	VB
9	1,18	DT	NN	VB
10	1,13	DT, DT	NN	

Tabell 8.15: Indexlista över nominalfrasstrukturer i fysiktexter för gymnasiet (jämfört med för högstadiet)

För det första finns inte bara det ensamma egennamnet med på listan, utan även egennamn med definit attribut. Denna konstruktion, på post 2, exemplifieras i (98) och (99), vilka visar att det, i analogi med i matematik på gymnasiet, handlar om hur man i ämnet benämner variabler.

(98) *Av en vägvisare framgår att orten A ligger 5 km söderut och orten B 7 km norrut.*

(99) *Lokets dragkraft på första vagnen är F , och varje vagn har massan m .*

Vidare innehåller listan för fysik en substantivfras som bestäms av två framförställda definita attribut, på post 10. Ett exempel på denna konstruktion visas i (100). Av exemplet framgår att det ena av attributen snarast är det SAG benämner kvantitetsattribut.

(100) *Längs nodlinjerna 1, 2 och 3 är de två vågrörelserna i motfas.*

En nominalfraskonstruktion, som också finns på gymnasielistan för biologi, är substantivfrasen med bisatsattribut och både ett framförställt definit attribut och ett adjektivattribut. Denna frastyp, på post 8, står oftast i objektposition, vilket visas exempel på i (101).

(101) *Den konkava spegeln i fig. 9 har en speciell form som kallas parabolisk.*

Även fysikspråket uppvisar många fler typiska typer av utbyggda nominalfraser på gymnasiet än på högstadiet.

8.4.3.5 Kemitexter på högstadiet och gymnasiet

Liksom i fysik används i kemi enkla pronomen i högre utsträckning på högstadiet än på gymnasiet (se tabell 8.16). Kemispråket på högstadiet utmärks dessutom av enkla substantiv.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	1,43		PN	
2	1,24		NN	

Tabell 8.16: Indexlista över nominalfrasstrukturer i kemitexter för högstadiet (jämfört med för gymnasiet)

Sammanfattningsvis kan konstateras att det naturvetenskapliga högstadiespråket i jämförelse med gymnasiespråket framför allt kännetecknas av mycket enkla konstruktioner, i huvudsak pronomen och substantiv utan några bestämmningar, även om biologi- och fysikspråkets indexlistor visserligen inkluderar pronomen med bisatsattribut.

Högt upp bland typiska strukturer för kemispråket (se tabell 8.17) på gymnasiet hamnar två substantivfraser med appositionsattribut (AN), varav en även innehåller ett definit attribut.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	2,02		PM	
2	1,55	AN	NN	
3	1,37		NN	PP
4	1,30	DT	NN	VB
5	1,30	DT	NN	PP
6	1,17	DT, DT	NN	
7	1,17	DT, AT	NN	
8	1,14	DT	NN	
9	1,14	DT, AT	NN	PP
10	1,11	AT	NN	PP

Tabell 8.17: Indexlista över nominalfrasstrukturer i kemitexter för gymnasiet (jämfört med för högstadiet)

Vid en närläsning av de fraser som taggats på detta vis konstaterades att flertalet av fraserna med definit attribut var av två slag. Antingen hade avstavning i originaltexterna gjort att parsern misstagit ett bestämt substantiv för en bestämd artikel med två på varandra följande substantiv, eller också hade parsern taggat fria adverbial som inleddes med "t.ex.", "främst" eller "särskilt" som appositioner. Eftersom ingen av dessa former borde klassas som appositioner har jag valt att justera indexlistan för kemispråket på gymnasiet och låtit denna till största del feltaggade appositionskonstruktion utgå, trots att enstaka korrekt taggade exempel, som (102) kunde hittas.

(102) *Man ska alltså väga upp 0,100 mol bariumklorid, BaCl₂.*

Däremot har den enkla substantivfrasen med appositionsbestämning, på post 2, behållits. Denna konstruktion avspeglar kemiämnets benägenhet att omnämna grundämnen och kemiska föreningar på flera sätt: med vardagliga namn, kemiska namn eller med kemiska formler, se exempel (103) och (104).

(103) *Dikväveoxid (lustgas) är inte särskilt giftig, det är däremot NO och NO₂.*

(104) *Bauxiten överförs först till aluminiumoxid, Al₂O₃.*

Denna utvidgning med appositioner mellan högstadiespråket och gymnasiespråket märks endast för kemi. En annan konstruktion som blir vanligare på gymnasiet, endast i kemi, är substantivfrasen med adjektiv- och prepositionsattribut, på post 10, ($p = 0,001$). Denna används framför allt i objektsadverbial, och exemplifieras i (105).

- (105) *Man kan lyfta upp protoner från det parallella tillståndet till det antiparallella genom att tillföra energi i form av elektromagnetisk strålning med exakt rätt frekvens.*

Kemispråket på gymnasiet utmärks också av att substantiv byggs ut med två definitiva attribut (post 6), varav det ena oftast bestämmer kvantitet. Detta särdrag delar kemi med såväl fysik som matematik. Ett exempel på en dylik fras i kemi visas i (106).

- (106) *Kolatomernas fyra valenselektroner gör att varje kolatom får ädelgasstruktur genom att dela fyra elektronpar med andra atomer; vanligtvis väteatomer eller andra kolatomer.*

Sammantaget är de flesta former av utbyggda nominalfraser mycket vanligare i de naturvetenskapliga ämnesspråken på gymnasiet än på högstadiet. Eventuellt stöder resultaten också att definitiva attribut är mer utmärkande för naturvetenskapliga gymnasiespråk än adjektiviska attribut, och att prepositionsattribut är mer utmärkande än bisattsattribut.

8.4.3.6 Naturvetenskapliga texter jämförda med akademiska texter

I tabell 8.18 redovisas indexlistan för naturvetenskap som helhet jämfört med akademiska texter. Som synes innehåller listan bara två poster. Båda är dessutom relativt enkla och har förhållandevis låga index.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	1,26		NN	VB
2	1,15		NN	

Tabell 8.18: Indexlista över nominalfrasstrukturer i naturvetenskapliga lärobokstexter (jämfört med akademiska texter)

Indexlistan för akademiska texter jämfört med naturvetenskapliga texter kan ses i tabell 8.19.

Denna lista är betydligt mer utförlig; här finns fraser med alla tre typer av huvuden, samt en hel del utbyggda frastyper av olika slag. Vid närmare inspektion av listan framkommer att samtliga frastyper med substantiviskt huvud redan är bekanta från de naturvetenskapliga ämnenas olika indexlistor för gymnasiet. Utbyggnadsmönstren för de substantivfraser som flitigt används i akademiska texter utvecklas alltså redan på gymnasiet, om än något olika i olika specifika

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	2,98		PM	
2	2,12	HD	PM	
3	1,78		PN	VB
4	1,76	DT, AT	NN	PP
5	1,43	DT, AT	NN	
6	1,43	DT, AT	NN	VB
7	1,43		PN	PP
8	1,34	DT	NN	PP
9	1,29	DT	NN	VB
10	1,26		PN	
11	1,22		NN	PP
12	1,21	DT, DT	NN	
13	1,19	AT	NN	PP

Tabell 8.19: Indexlista över nominalfrasstrukturer i akademiska texter (jämfört med naturvetenskapliga lärobokstexter)

ämnen. Inga nya kombinationer av attribut särskiljer de akademiska texterna, vilket betyder att dessa konstruktioners positionering på listan endast beror på att de är ännu vanligare i de akademiska texterna.

På listan återfinns emellertid några särdrag som inte finns på gymnasie-listorna för naturvetenskap. Mest förvånande är att det enkla pronomenet finns bland posterna, det som tidigare skilde högstadietexterna från gymnasietexterna. Även pronomina nominalfraser med efterställda attribut finns med på den akademiska listan, på post 3 och 7. Dessa frastyper exemplifieras i (107) och (108).

(107) Något som Lyon Clark bara nämner i förbifarten är *nykritikens genombrott mot mitten av 1900-talet*.

(108) En av förespråkarna för att socialisationsforskningen bör inriktas mot mottagliga år är *Niemi som både var med i tidig socialisationsforskning (1970-tal) och även har kritiserat den inriktning denna forskning tog*.

Några tänkbara förklaringar till att pronomen är vanligare i akademiska texter tas upp i fokusstudie 8.4.4.1. Helt klart är emellertid att också egennamn används i högre utsträckning i akademiska texter än i naturvetenskapliga lärobokstexter (post 1 och 2), och att dessa förekommer främst rör det akademiska referenssystemet och källhanteringen. Allra vanligast är ett enkelt egennamn

(oftast ett efternamn) och därefter en kombination av för- och efternamn (i listan betecknat som frasens huvud, HD). Dessa två användningar exemplifieras i (109) och (110).

(109) Progovac (1994) redovisar goda skäl för att se polaritet som ett formellt fenomen och har stöd från empiri i flera språk för att man ska göra skillnad mellan polaritet i negerade och icke-negerade kontexter.

(110) Här kan man skönja den rädsla som av Boel Westin rubricerats "fruktan för fiktionen" och som ständigt ledsagat barnlitteraturen.

8.4.3.7 Naturvetenskapliga texter jämförda med berättande texter

Precis som i jämförelsen med akademiska texter presenteras inga stadiespecifika indexlistor för naturvetenskap i jämförelse med berättande texter, av anledningen att inga anmärkningsvärda skillnader mellan stadierna står att finna. Indexlistan för naturvetenskapliga texter jämförda med berättande texter visas i tabell 8.20.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	2,95	AT	NN	
2	2,28		NN	PP
3	2,16		NN	VB
4	2,15	AT	NN	PP
5	1,88	DT	NN	VB
6	1,86	DT	NN	PP
7	1,72		NN	
8	1,51	DT	NN	
9	1,46	DT, AT	NN	PP
10	1,45	DT, AT	NN	
11	1,40	DT, DT	NN	

Tabell 8.20: Indexlista över nominalfrasstrukturer i naturvetenskapliga lärobokstexter (jämfört med berättande texter)

För det första kan påpekas att indexen för de mest typiska frastyperna är något högre än för jämförelserna mellan lärobokstexter, vilket skulle kunna förklaras av att denna jämförelse rör olika texttyper. För det andra innehåller indexlistan hela 11 olika nominalfrastyper med substantiviskt huvud som är

mer utbredda i naturvetenskapliga lärobokstexter än i berättande texter. Listan innehåller allt från ett enkelt substantiv till substantiv med en, två och tre bestämmingar.

Den kompletterande indexlistan, för berättande texter jämfört med naturvetenskapliga, kan ses i tabell 8.21.

	Index	F-ställda attr.	Huvudord	E-ställda attr.
1	5,19		PM	
2	2,75		PN	
3	1,78		PN	VB

Tabell 8.21: Indexlista över nominalfrasstrukturer i berättande texter (jämfört med naturvetenskapliga lärobokstexter)

Denna blygsamma lista domineras av det enkla egennamnet, som är mer än fem gånger så utbrett i berättande texter. Därefter faller två nominalfrastyper med pronominellt huvud ut i jämförelsen, varav den sista innehåller ett bisatsattribut och exemplifieras i (111), (112) och (113).

(111) Den som inte får tala, *är död*. (barnböcker)

(112) *Är det* du som är Stockholms Maradona? (ungdomsromaner)

(113) *Och sagt* det som skulle sägas. (vuxenromaner)

8.4.4 Fokusstudier

8.4.4.1 Pronomentyper

Ett relativt oväntat resultat från jämförelsen med akademiska texter (se avsnitt 8.4.3.6) är att nominalfraser med pronomen som huvudord utmärker de akademiska texterna. Ett ymnigare bruk av pronomen brukar sägas tyda på dels en mer kontextbunden texttyp, dels en mer deltagande texttyp. Akademiska texter torde generellt sett inte vara särskilt kontextbundna, men möjligen mer deltagande än gymnasietexterna. En mer detaljerad uppställning över vilka pronomen och pronomentyper som används i akademiska texter och naturvetenskapliga lärobokstexter presenteras i tabell 8.22.

I tabellen listas de pronomen som utgör minst en procentandel av det totala antalet pronomen i någon av texttyperna. De fall där de akademiska texterna utmärker sig positivt gäller framför allt de personliga pronomina ”jag/mig”,

	NV	SveAk
<i>Första och andra personens personliga pronomen</i>		
jag	0,27	8,31
mig	0,05	1,35
du	7,03	1,25
dig	1,15	0,23
vi	9,55	4,80
oss	1,37	0,98
<i>Tredje personens personliga pronomen</i>		
han	1,84	5,08
hon	0,64	3,68
den	7,71	3,98
det	25,74	26,67
sig	8,08	9,49
de	9,15	8,15
dem	1,96	1,87
<i>Demonstrativa pronomen</i>		
detta	2,19	4,54
dessa	1,06	0,87
<i>Generaliserande pronomen</i>		
man	13,52	7,03
<i>Reciprokt pronomen</i>		
varandra	2,16	0,90
<i>Allmänt obestämt pronomen</i>		
något	0,82	1,62

Tabell 8.22: Översikt över pronomenbruket i naturvetenskapliga lärobokstexter och akademiska texter. Siffrorna anger procentandel av totalt antal pronomen

samt ”han” och ”hon. Den högre andelen första personens personliga pronomen speglar troligtvis det faktum att författare till akademiska texter faktiskt tillåts vara synligare i texterna än läroboksförfattare. Den förhållandevis rika användningen av tredje personens personliga pronomen beror förmodligen på

att de akademiska texterna innehåller mycket referenser till tidigare forskares verk och tankar. Båda dessa resultat ligger i linje med tidigare resultat för humanistiska akademiska texter (se avsnitt 2.2.2). Möjligtvis hade resultatet sett annorlunda ut om den akademiska korpusen hade innehållit naturvetenskapliga texter.

Ser man till pronomentyper är den mest påfallande skillnaden att det generaliserande pronomenet ”man” utgör dubbelt så stor andel av pronomina i naturvetenskapliga lärobokstexter. Detta beror antagligen på att läroböckerna i högre utsträckning presenterar ämnesstoffet som vore det allmänt vedertaget, medan de akademiska texterna mer uttalat behandlar olika teorier. Att naturvetenskapliga lärobokstexter tillhör ett särskilt generaliserande register antyder också Edling (2006). I hennes studie innehåller naturvetenskaplig text fler generaliserande substantiv än texter i svenska.

Även de personliga pronomina ”du” och ”vi” utgör en betydligt större andel i lärobokstexterna. Av allt att döma har alltså läroböckerna ett mer läsarinkluderande tilltal än de akademiska texterna.

8.4.4.2 Om nominalfrastäthet och nominalfraslängd som komplexitetsmått

Vad gäller den kvantitativa metoden kan förtydligas att beräkningarna grundar sig på nominalfraser som utgör primära satsled och nominalfraser som ingår i fria adverbial. Antalet nominalfraser i meningar kan således variera beroende på vilka verbtyper som förekommer i texterna, samt på hur många huvudsatser, bisatser och fria adverbial som ingår.

Beslutet att inte särskilt räkna nominalfraser som ingår i attribut till nominala huvudord bör dock kommenteras något mer utförligt. När man jämför resultaten av NP-täthet och NP-längd kan det nämligen vara nyttigt att ha i åtanke att olika texttyper är olika benägna att ta nominala attribut. För lärobokstexterna gäller att variationen mellan ämnena ligger på 0,25–0,30 nominalt attribut/nominalfras på högstadiet, och på 0,95–1,17 nominalt attribut/nominalfras på gymnasiet. Värdet för de akademiska texterna är 0,34 och för de berättande texterna 0,13.⁵⁰

Eftersom nominalfraserna på gymnasiet oftare än på högstadiet innehåller andra, kortare nominalfraser, kan man, med en mer strikt kvantitativ och mindre pragmatisk syn på nominalfraser, hävda att på gymnasiet underskattas NP-tätheten, medan NP-längden i stället överskattas. Denna omständighet gör att andel substantiv faktiskt är ett bättre mått på texternas nominalitetsgrad än nominalfrastätheten per mening. Men då ska man samtidigt komma ihåg att substantivfrekvensen bara är *en* lexikogrammatisk parameter som bidrar till ökad nominalitet.

Även det faktum att NP-tätheten är beräknad per mening kan diskuteras ytterligare. Att räkna fraser per mening är något problematiskt eftersom meningslängder i materialet varierar; en mening är inte en konstant enhet i det av-

⁵⁰Eftersom det rör sig om så många förekomster av nominalfraser i texterna är såklart också de allra minsta skillnader signifikanta.

seendet. Dock motsvarar en mening en viktig syntaktisk enhet, som den största textenhet som inte är underordnad någon annan. En mening är också typiskt ett uttryck för en ”självständig språkhandling” (SAG, del 1: 197). Nominalfraser har också beräknats per mening i tidigare forskning (se t.ex. Feng, Elhadad och Huenerfauth 2009) och då hävdats vara relevant för syftet; ju fler nominalfraser – referenter – en mening innehåller, desto mer information måste läsaren hålla i huvudet för att kunna tolka den (Barzilay och Lapata 2008). Men man skulle också kunna räkna nominalfraser per sats eller per 200 ord eller liknande, mer fixerade enheter.

Frågan som infinner sig är hur stark korrelationen mellan antal nominalfraser per mening och meningslängd är. Är denna stark är det ytliga standardmålet på syntaktisk komplexitet, meningslängd, nog för att uppskatta den komplexitetsökning som en ökad NP-täthet innebär. Denna fråga är dessutom även aktuell för NP-längd, ju starkare korrelation mellan NP-längd och meningslängd, desto mindre anledning att beräkna NP-längd.

En undersökning av korrelationen mellan meningslängd och NP-täthet respektive NP-längd i de naturvetenskapliga ämnena presenteras i tabell 8.23. Av tabellen kan utläsas att de båda korrelationsvärdena är relativt lika på högstadiet och på gymnasiet. Korrelationen med meningslängd är dock betydligt starkare för NP-längd än för NP-täthet. I genomsnitt ligger styrkan på korrelationen mellan meningslängd och täthet på 0,74. En förändrad meningslängd påverkar alltså bara NP-tätheten med en faktor $0,74^2 = 0,55$, medan NP-längden påverkas med en faktor $0,92^2 = 0,85$.

		NP-täthet	NP-längd
Biologi	Högstadiet	$r_s = 0,74$	$r_s = 0,90$
	Gymnasiet	$r_s = 0,72$	$r_s = 0,92$
Fysik	Högstadiet	$r_s = 0,75$	$r_s = 0,91$
	Gymnasiet	$r_s = 0,76$	$r_s = 0,93$
Kemi	Högstadiet	$r_s = 0,74$	$r_s = 0,90$
	Gymnasiet	$r_s = 0,72$	$r_s = 0,93$

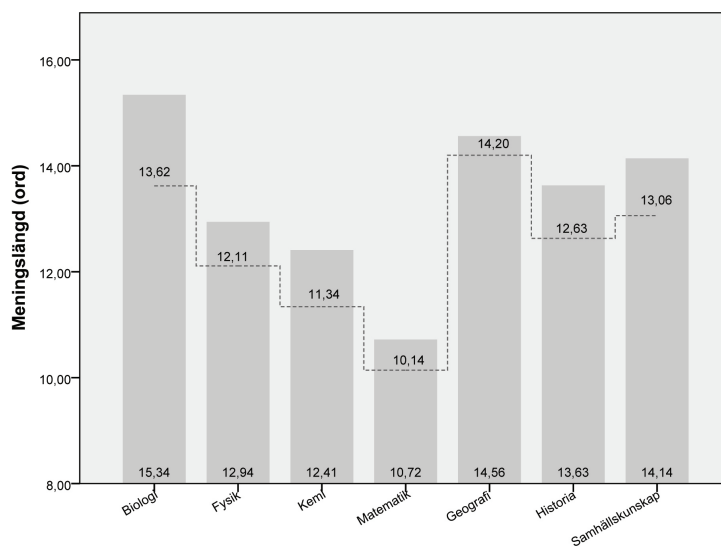
Tabell 8.23: Spearmans rangkorrelation mellan meningslängd och NP-täthet respektive NP-längd i lärobokstexter i naturvetenskapliga ämnen

För de akademiska texterna är styrkan på korrelationen mellan meningslängd och NP-täthet 0,70 och mellan meningslängd och NP-längd 0,92. Dessa siffror är mycket lika siffrorna för läroböckerna. Motsvarande värden för de be-

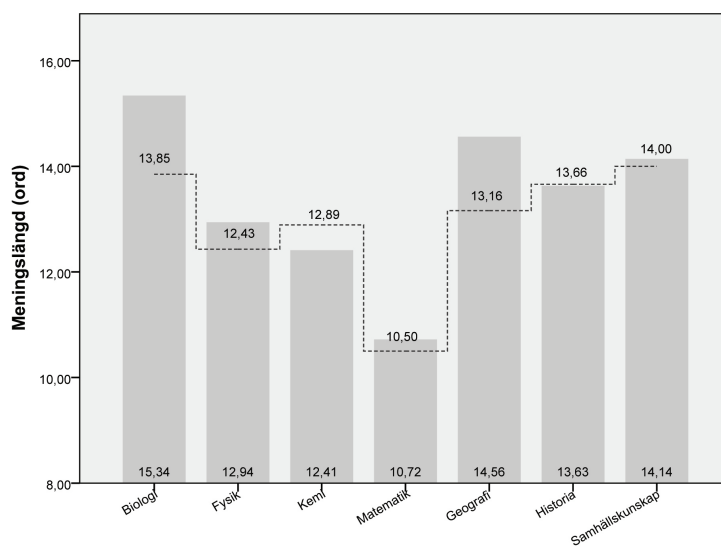
rättande texterna är dock 0,84 och 0,93. Måtten meningslängd och NP-täthet ger alltså mer lika resultat för de berättande texterna än för övriga texttyper i undersökningen.

Ett annat sätt att illustrera samvariationen är att visa hur väl ökningen av NP-täthet och NP-längd förutser ökningen av meningslängd mellan högstadiet och gymnasiet. I figur 35 visas de faktiska, observerade meningslängderna i hela läroboksmaterialet, tillsammans med uppskattade meningslängder, utifrån observerad NP-täthet respektive NP-längd.

I figurerna syns att mätningarna av NP-längd och meningslängd visar på mer liknande resultat än NP-täthet och meningslängd. Detta förhållande tycks gälla för samtliga ämnen utom för geografi.



(a) Estimerad meningslängd utgående från observerad NP-täthet



(b) Estimerad meningslängd utgående från observerad NP-längd

Figur 35: Förhållandet mellan observerad meningslängd (staplad) och estimerad meningslängd (streckad) i gymnasieläroböcker

8.5 Typiska nominalfraser i naturvetenskap: sammanfattning och diskussion

Studien visar att både nominalfrastätheten och nominalfraslängden ökar i samtliga ämnestexter i övergången mellan högstadiet och gymnasiet, samt att de akademiska texterna befinner sig betydligt högre på båda dessa skalor. Båda dessa resultat pekar mot en ökad nominalitetsgrad i texterna över skolåren. Indexlistorna för gymnasiet och för de akademiska texterna bekräftar också att attributiva prepositionsfraser ökar i stadieövergångarna, ytterligare en lexikogrammatisk variabel som bidrar till en ökad nominalitet.

8.5.1 Förhållandet till nominalfraser i matematik och samhällsvetenskap

I jämförelse med matematiska texter är nominalfraserna i naturvetenskapliga texter vanligare, längre och mer utbyggda, särskilt ofta med adjektiv- eller bisatsattribut. De enda attribut som utmärker matematiska lärobokstexter är de definitiva, vilket till stor del förklaras genom att kvantitetsattribut ingår bland de attribut som taggas som sådana.

I jämförelse med samhällsvetenskapliga texter visar det sig att nominalfraserna i naturvetenskapliga texter dels är kortare, dels oftare bestäms av bisatser. Nominalfraserna i samhällsvetenskap är längre och mer utbyggda med definitiva, adjektiviska och prepositionella attribut. Samhällsvetenskapstexterna är tydligen än mer nominala än de naturvetenskapliga texterna, på både högstadiet och gymnasiet. Detta resultat ligger i linje med resultatet av NQ-analysen (se avsnitt 6.4.2).

Eftersom fem av de sex ämnen som ingår i Westmans studie är samhällsvetenskapliga, är det inte särskilt förvånande att beskrivningarna av nominalfraser i läroböckerna på 1970-talet bäst stämmer överens med mina resultat för just de samhällsvetenskapliga lärobokstexterna (jfr avsnitt 2.4.2.1).

8.5.2 Stadieövergången mellan högstadiet och gymnasiet

När naturvetenskap bryts ner i sina ämnesmässiga beståndsdelar faller mer olika nominalfrasstrukturer ut för de olika ämnena på gymnasiet än på högstadiet, vilket kan tolkas som att det naturvetenskapliga ämnesfacket är mer språkligt homogent på högstadiet. På gymnasiet följer de olika ämnesspråken skilda utvecklingslinjer, vilket resulterar i att läroböckerna på gymnasiet har en tydligare individuell ämnesspråklig karakteristik. Denna tendens märks också tydligt i undersökningarna av nominalfrasernas täthet och längd.

Medan de naturvetenskapliga ämnena har ett mycket mer homogent kärnspråk på högstadiet, verkar de olika samhällsvetenskapliga ämnena utvecklas mer i samma takt mellan högstadiet och gymnasiet, vilket gör att samhällsvetenskap som ämnesfack tycks mer sammanhållet än naturvetenskap på gymnasiet.⁵¹ Detta skulle kunna betyda att de samhällsvetenskapliga ämnena på gymnasiet inte utvecklar en lika tydlig enskild ämnesspråklig karakteristik som de naturvetenskapliga ämnena, vars gemensamma språkliga kärna splittras i stadiövergången till gymnasiet.

I Danielsson 1975 framkommer att nominalfraserna i gymnasieläroböcker i fysik (och samhällskunskap) innehåller betydligt fler adjektivattribut än motsvarande högstadietexter. Tendensen mot mer utbyggda nominalfraser på gymnasiet är tydlig för lärobokstexterna i samtliga naturvetenskapliga ämnen i mitt material. Strukturerna som faller ut för högstadiespråken är förvånansvärt enkla: för det mesta bara enkla pronomen och substantiv, som i enstaka fall kombineras med ett efterställt bisatsattribut. Inga framförställda bestämmingar eller efterställda prepositionsattribut hör särskilt högstadiespråket till. Nominalfrasutbyggnaden går i enda riktning, mot en påtagligt rikare och mer komplex struktur, i stadiövergången. Litet annorlunda uttryckt är ämnesspråken på gymnasiet tydligt utvecklade ur ämnesspråken på högstadiet.

8.5.3 Förhållandet till nominalfraser i akademiska texter

NQ-undersökningen (i 6.4.2) pekar på att lärobokstexterna för gymnasiet i alla ämnesfack faktiskt är mer informationspackade än akademiska texter. Fler pronomina nominalfraser i akademiska texter än i gymnasietexter talar också för att gymnasietexter är allra mest nominala. Nominalfraserna i de akademiska texterna är visserligen generellt sett längre än i gymnasietexterna, men indexlistorna visar att de inte är mer utbyggda. Längdökningen måste alltså bero på att en högre andel nominalfraser byggs ut enligt de mönster som etableras i stadiövergången mellan högstadiet och gymnasiet. Det faktum att de akademiska texterna också har fler nominalfraser per mening än gymnasietexterna hänger förmodligen samman med att meningarna i akademiska texter generellt sett är längre.

På liknande vis som det naturvetenskapliga gymnasiespråket förefaller vara en utveckling av högstadiespråket, tycks det akademiska språket till största del vara en vidareutveckling av skolspråket, än en fristående texttyp. Detta märks i första hand genom att det i det närmaste inte förekommer några starkt utmär-

⁵¹Denna tendens syns också tydligt i mätningarna av meningslängd, ordlängd och LIX (i avsnitt 6.3).

kande nominalfrastyper i naturvetenskapliga lärobokstexter i jämförelse med akademiska texter. Dessutom förekommer de utbyggda NN-nominalfraserna som kännetecknar de akademiska texterna redan flitigt på gymnasiet.

Resultatet av fokusstudien om pronomentyper (se 8.4.4.1) tyder dock på att författarna till de akademiska texterna är mer synliga i texterna än författarna till lärobokstexterna. Dessutom kretsar de akademiska texterna mer kring namngivna individer än lärobokstexterna. Eftersom den akademiska korpusen inte innehåller naturvetenskapliga texter är det dock svårt att veta om resultatet beror på ämnesskillnader eller stadieskillnader i jämförelsen.

8.5.4 Förhållandet till nominalfraser i berättande texter

Den största skillnaden mellan nominalfraserna i de naturvetenskapliga texterna och de berättande texterna rör huvudfigurerna i texterna. I naturvetenskap talas det om substantiviska företeelser, medan det i berättande texter refereras till namngivna individer, antingen med namn eller med pronomen.

Nominalfraserna i berättande texter är också förhållandevis korta, medan nominalfraserna i naturvetenskapliga texter är betydligt mer utbyggda, med olika slags både framför- och efterställda bestämningar.

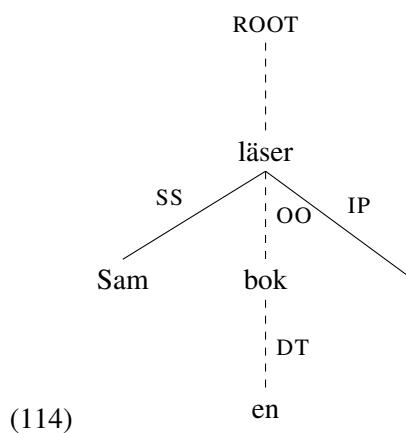
9

DELSTUDIE: TYPISKA MENINGAR

I denna delstudie undersöks djupet och strukturen av dependensträd för meningar, i syfte att säga något om enskilda meningars och, i förlängningen, texters komplexitet. Eftersom analysen i denna delstudie baseras på lingvistisk kunskap utgör den en mer direkt och detaljerad utredning av meningars komplexitet än det traditionella, ytliga läsbarhetsmättet meningslängd, som brukar användas som indikator på syntaktisk komplexitet.

9.1 Typiska meningar: studiens upplägg

I studien beräknas två parametrar: meningars parsningsdjup och typiska uppbyggnad. Parsningsdjupet för en mening definieras som trädets maxdjup, d.v.s. antalet noder från roten till det mest avlägsna lövet (jfr Heimann Mühlenbock 2013: 41f.). I exempel (114) visas ett parsträd för meningen *Sam läser en bok*, med djupet 3 noder. Maxdjupet har markerats med streckade grenar i trädets.



De syntaktiska kategorier som används kommer från MAMBA-opsättning-
en och listas i tabell 9.1.⁵²

Tagg	Betydelse
+A	Konjunktionellt adverb
+F	Koordination på huvudsatsnivå
AA	Annat adverbial
AT	Framförställt attribut
CA	Kontrastivt adverbial
CJ	Samordnat led
DT	Determinerare bestämningsord
EF	Relativ bisats
ET	Efterställd bestämning
FV	Finit verb predikatsverb
HD	Huvud
I?	Frågetecken
IF	Infinitivfras utom infinitivmärke
IK	Kommatecken
IO	Indirekt objekt (dativobjekt)
IP	Punkt
IT	Divis (bindestreck)
MS	Makrosyntagm
NA	Negerande adverbial
OA	Objektsadverbial (prepositionsobjekt)
OO	Direkt objekt (ackusativobjekt)
PA	Prepositionskomplement
PL	Verbpartikel
PT	Predikativt attribut (reaktion)
RA	Platsadverbial
SP	Subjektspredikativ (subjektiv predikatsfyllnad)
SS	Subjekt (övrigt subjekt)
TA	Tidsadverbial
UA	Underordnad sats (bisats) utom subjunktion
VA	Korrelativt adverbial
VG	Verbgrupp

Tabell 9.1: Opsättning av syntaktiska taggar och deras betydelse

⁵² Annoteringen av texterna beskrivs närmare i avsnitt 4.1.

9.2 Typiska meningar: resultat

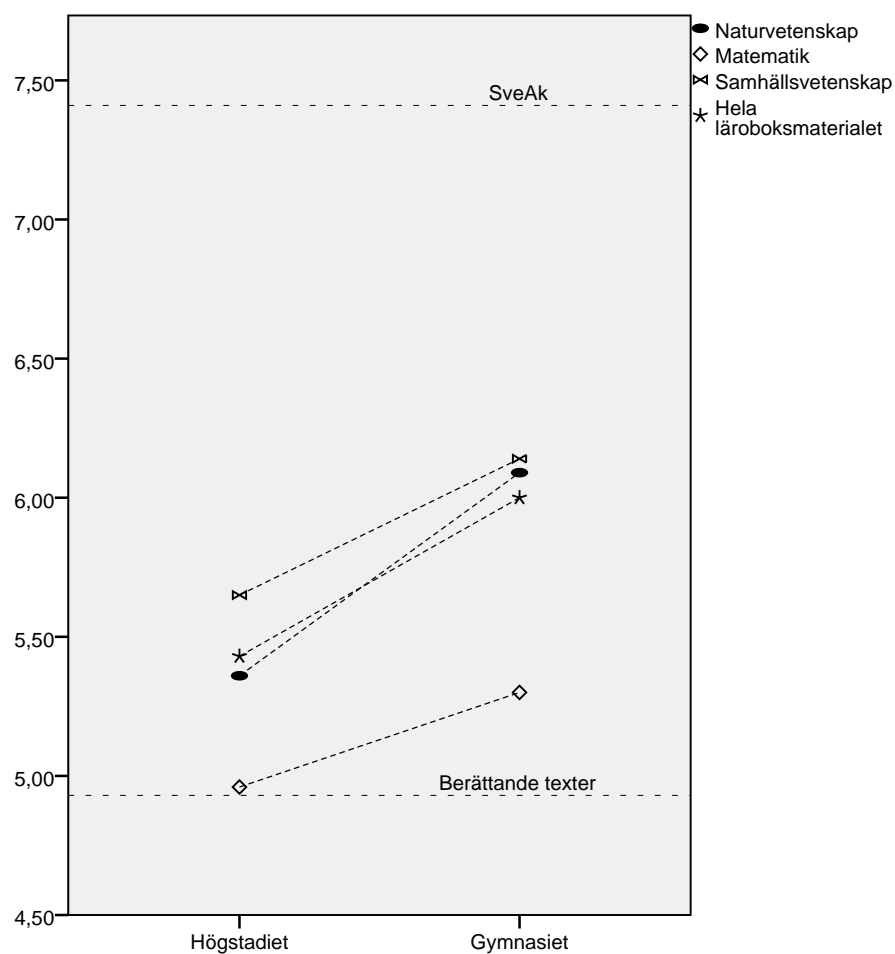
9.2.1 Parsningsdjup

För parsningsdjupen presenteras i första hand linjediagram med medelvärden för de olika materialen för högstadiet och gymnasiet. I dessa diagram har en streckad trendlinje ritats in för att synliggöra utvecklingen i stadiövergången. Medelvärdena ger dock en begränsad bild av fördelningen av parsningsdjup i texterna. Därför presenteras även parsningsdjupen i låddiagram, som bättre beskriver meningskomplexiteten i materialen som helhet.

Detaljerade mätvärden för parsningsdjup finns i appendix E.4.

9.2.1.1 Parsningsdjup i de olika ämnesfacken

Parsningsdjupet för läroboksmaterialet uppdelat på ämnesfack åskådliggörs i figur 36 och 37. Av linjediagrammet kan utläsas att de berättande texterna har ett genomsnittligt parsningsdjup på knappt 5 noder, vilket samtliga ämnesfack överstiger redan på högstadiet, om än i olika utsträckning.



Figur 36: Översiktlig utveckling av genomsnittligt parsningsdjup över stadierna

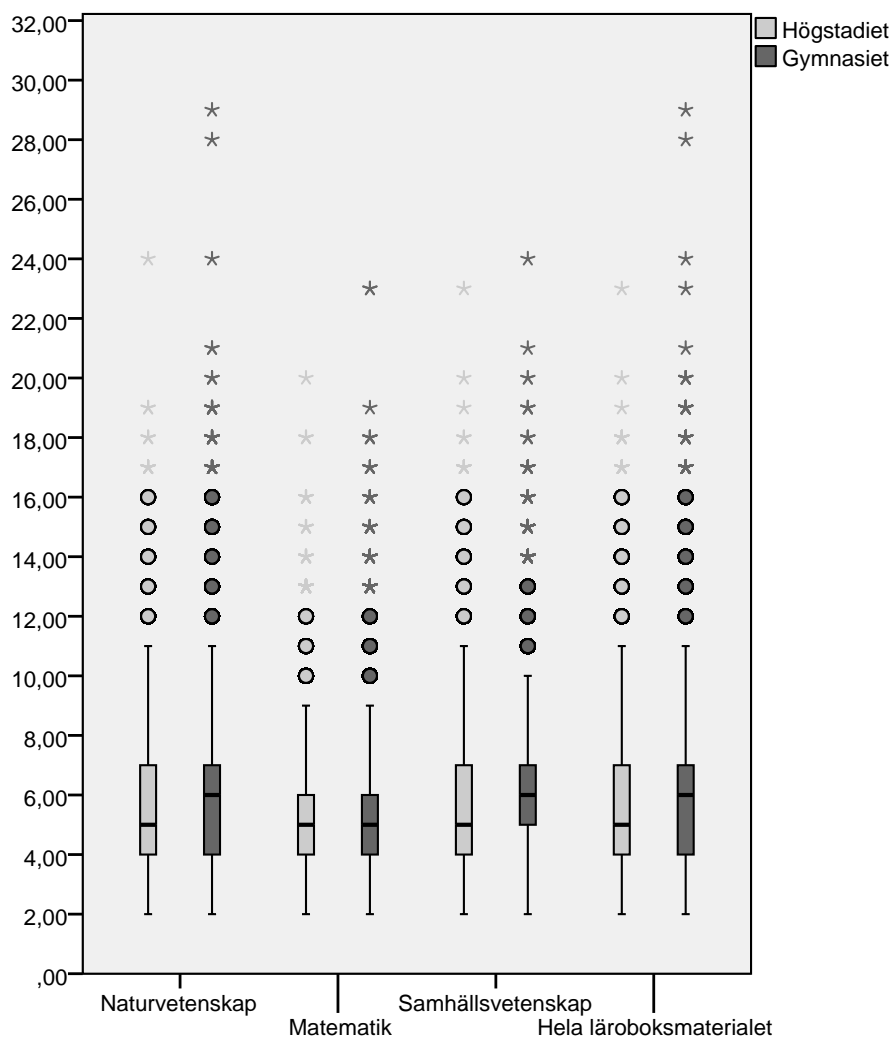
I samtliga ämnesfack stiger parsningsdjupet i stadieövergången till gymnasiet. Skillnaden mellan stadierna är högst i naturvetenskap och minst i matematik. Inget av ämnesfacken kommer dock i närheten av parsningsdjupet i de akademiska texterna, som överstiger 7 noder.

Låddiagrammet (figur 37) synliggör några väsentliga tillägg till den bild av materialet som medelvärdena tecknar.⁵³ Meningarna i naturvetenskap befinner sig på båda stadierna i intervallet 2–11, och hälften av meningarna ligger mellan 4 och 7. Skillnaden mellan stadierna är att på högstadiet mäter fler mellan 4 och 5, och på gymnasiet ligger fler meningar mellan 6 och 7, samt att gymnasietexterna innehåller fler extremvärden.

Fördelningen av djupen i meningarna i matematik visar att de typiska meningarna i ämnet inte skiljer sig åt alls i fråga om djup mellan stadierna, som varierar mellan 2 och 9 och främst samlas mellan 4 och 6. Ökningen i medelvärde beror endast på fler extremvärden på gymnasiet.

I samhällsvetenskap visar fördelningen av meningarna att ämnesfacket på högstadiet är mycket likt naturvetenskap; värdena varierar mellan 2 och 11, med hälften av meningarna inom 4–7, varav de allra flesta på 4 eller 5. Där- emot avviker gymnasiefördelningen på så vis att både maxvärdet och kvartilavståndet minskar; de mest komplexa meningarna har nu djupet 10 och meningar med djup 4 tillhör inte längre de mest typiska.

⁵³I ett låddiagram sammanfattas spridningen av observationsvärden med hjälp av fem värden: medianvärdet, undre och övre kvartilen (fjärdedelen) samt minimum- och maximumvärdet. Eventuella avvikande värden markeras med egna symboler. Själva lådan representerar det kvartilspann inom vilket 50 % av värdena ligger, och medianvärdet är markerat med en vertikal linje inuti lådan. Avståndet mellan övre och undre kvartilen, *kvartilavståndet*, motsvarar längden på lådan. Från lådans respektive sidor sträcker sig en vågrät linje ut till minimum- respektive maximumvärdet i serien (avvikande värden exkluderade). Observationsvärden mellan 1,5 och 3 lådlängder från lådans över- eller underkant kallas *utliggare* och markeras i diagrammet med ringar. Observationsvärden som ligger mer än 3 lådlängder från lådans kanter kallas *extremvärden* och markeras med stjärnor.



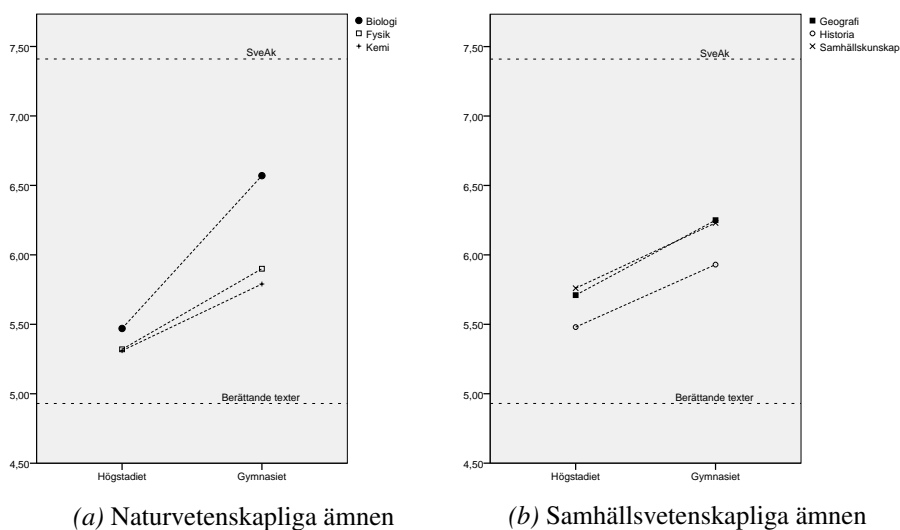
Figur 37: Översiktlig fördelning av parsningsdjup över stadierna

9.2.1.2 Parsningsdjup i de olika naturvetenskapliga ämnena

Betraktar man figur 38 och 39, som visar parsningsdjupen för de enskilda ämnena i naturvetenskap och samhällsvetenskap, ser man liknande tendenser som dem i meningslängdsundersökningen (se avsnitt 6.3 och figur 17). Bland de naturvetenskapliga ämnena utmärker sig biologi med en förhållandevis kraftig stegring mellan högstadiet och gymnasiet, medan värdena för

fysik och kemi i stort sett är identiska på högstadiet, och därefter i stort sett följer varandra åt. Utvecklingen resulterar i att de naturvetenskapliga ämnena uppvisar en större spridning på gymnasiet än på högstadiet.

En sådan spridning sker inte i stadiövergången för de samhällsvetenskapliga ämnena. Det mest påtagliga resultatet för de samhällsvetenskapliga ämnena är de lika nivåerna för geografi och samhällskunskap.



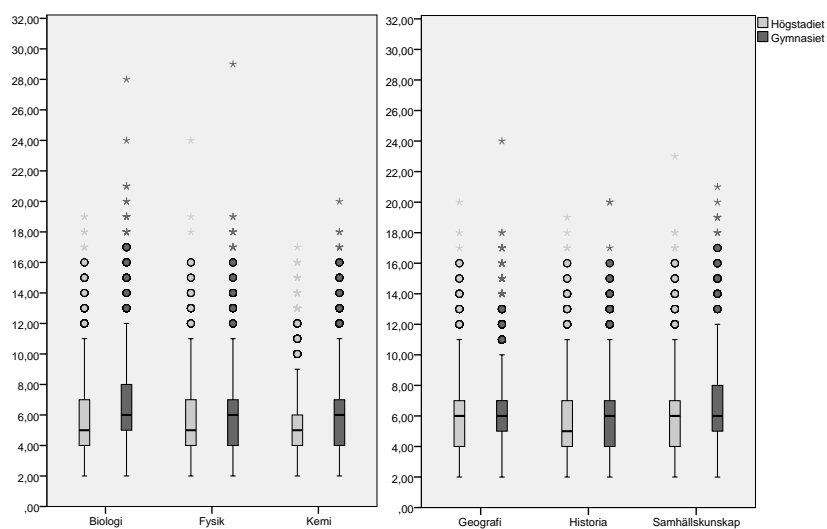
Figur 38: Ämnesspecifik utveckling av genomsnittligt parsningsdjup över stadierna

Fördelningen av djupen i de naturvetenskapliga ämnena avslöjar att den dramatiska ökningen av komplexitet i biologiböckerna för gymnasiet i stora drag kan förklaras med att meningar med djup 4 blir mindre vanliga, medan meningar med djup 6–8 blir fler. Dessutom ökar maxvärdet från 11 till 12.

Meningarna i fysik fördelar sig på liknande vis som meningarna i naturvetenskap som ämnesfack. På båda stadierna befinner sig meningarna på djup mellan 2 och 11, med flest inom spannet 4–7, och i stadiövergången stiger medianvärdet från 5 till 6.

Meningarna i kemi på högstadiet samlas inom ett relativt litet intervall, 2–9, och hälften av meningarna har ett djup på 4–6. I stadiövergången tillkommer många meningar med djupet 10 och 11.

En högre andel av meningarna i geografi och samhällskunskap är av djup från och med 5 på gymnasiet, d.v.s. andelen meningar med djup 4 minskar i stadiövergången. En stor skillnad mellan geografi och samhällskunskap på gymnasiet, som medelvärdena inte avslöjar, är att maxdjupet i geografi är 10



(a) Naturvetenskapliga ämnen (b) Samhällsvetenskapliga ämnen

Figur 39: Ämnesspecifik fördelning av parsningsdjup över stadierna

(det minskar i stadiövergången), medan maxdjupet i samhällskunskap är hela 12. Meningarna i historieböckerna fördelar sig relativt lika på högstadiet och på gymnasiet, även om medianvärdet ökar från 5 till 6.

9.2.2 Meningsuppbyggnad

Uppbyggnaden av meningarna presenteras i indexlistor, d.v.s. listor som plockar ut strukturer som utmärker ett visst material jämfört med ett annat (se vidare avsnitt 5.3 i metodkapitlet).

Indexlistorna över typiska meningsstrukturer har konstruerats för de parsningsdjup som i denna studie (avsnitt 9.2.1) har konstaterats vara typiska för de olika texttyper som jämförs. För ämnesfacken på högstadiet är detta djup 5 noder. För naturvetenskap och samhällsvetenskap ökar mediandjupet till 6 noder i gymnasietexterna, medan djupet i matematikexterna för gymnasiet ligger kvar på 5 noder. För samtliga naturvetenskapliga ämnen är medianvärdet för parsningsdjupet i högstadietexterna 5 noder och i gymnasietexterna 6 noder.

Resultaten som redovisas här har erhållits med en tröskelnivå på minst 1 förekomst per 2 000 meningar. Detta gränsvärde valdes utifrån materialets omfång; det gav indexlistor av sådant omfång att resultatet gick att överblicka manuellt. Ett högre tröskelvärde, som det i nominalfrasundersökningen (se avsnitt 8.4.3), visade sig inte resultera i några strukturer, vilket troligtvis beror på att varje mening (vanligtvis) bara innehåller ett maxdjupsmönster, till skillnad från nominalfrasmonster. En annan möjlig förklaring är att materialet är i minsta laget för att undersöka distributionen av ”vanliga” strukturella mönster, eftersom sådana kan se mycket olika ut.

I listorna presenteras strukturerna i fallande indexordning, vilket speglar en fallande typiskhet för det material som undersöks mot jämförelsematerialet. Alla skillnader som redovisas i detta avsnitt är starkt statistiskt signifikanta.

I beräkningen av strukturernas indexvärde skiljs mellan vänsterställda och högerställda mönster i förhållande till det finita verbet (rotnoden). Om strukturen med maximal komplexitet är placerad före roten anges strukturen i indexlistorna i kolumnen märkt ”vänster”, och om strukturen är placerad efter roten anges den i kolumnen märkt ”höger”. Motiveringen för denna åtskillnad är att s.k. vänster- eller högertyngd är etablerade stildrag inom textforskningen (se t.ex. Melin och Lange 2000; Lagerholm 2008). När man läser sägs man orientera sig efter det finita verbet och ju senare detta introduceras, desto mer information måste man hålla i minnet för att kunna tolka satsen. Långa fundament bidrar till vänstertyngd, d.v.s. att viktiga verb kommer sent i meningen. Vänstertyngd bidrar således till att texten blir mer svårläst. Enligt (Lagerholm 2008: 131) förekommer vänstertunga meningar framför allt i författningstexter och specialiserade facktexter. Bestämningar till höger anses i stället ge en mer lättläst text och en ledigare, talspråkligare stil.

9.2.2.1 *Naturvetenskapliga texter jämförda med matematiktexter*

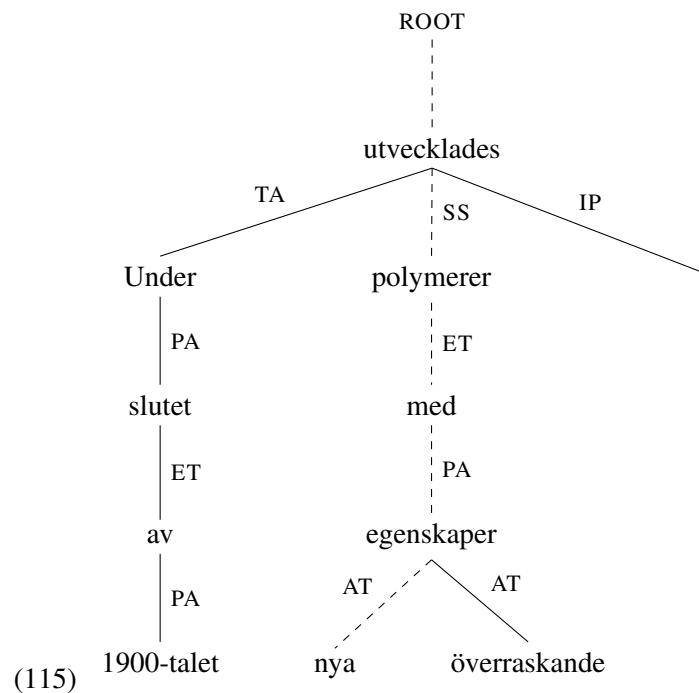
I detta avsnitt beskrivs resultaten av jämförelsen mellan naturvetenskapliga texter och matematiktexter för högstadiet. Indexlistan för typiska naturvetenskapliga meningsstrukturer kan ses i tabell 9.2. Eftersom den tidigare analysen av typiska parsningsdjup visar att mediandjupet i naturvetenskapliga högstadietexter är 5 noder, är det strukturer med detta omfång som listan omfattar.

	Index	Vänster	Höger
1	1,16		SS: ET, PA, AT
2	1,11		SP: ET, OA, PA
3	1,11		SP: ET, PA, CJ
4	1,11		OA: PA, ET, PA
5	1,10		SP: ET, PA, AT
6	1,10	SS: ET, RA, PA	

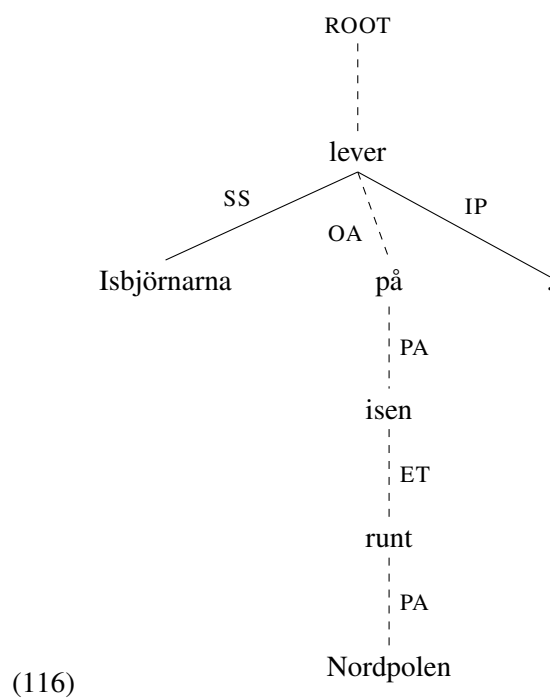
Tabell 9.2: Indexlista över meningsstrukturer i naturvetenskapliga lärobokstexter för högstadiet (jämfört med matematiktexter)

Komplexiteten i naturvetenskapliga meningar kan oftast lokaliseras till nominalfraser, vilka fungerar som subjekt (SS) eller subjektspredikativ (SP). Fraserna är också oftast högerställda predikatet, vilket tyder på att meningarna är relativt lättlästa.

Ett exempel på en mening med den allra mest typiska strukturen visas i (115). Meningen lyder: *Under slutet av 1900-talet utvecklades polymerer med nya överraskande egenskaper* och är hämtad ur en kemibok. Strukturen utgörs av ett högerställt subjekt, som bestäms av en efterställd prepositionsfras (ET), i vilken rektionen (PA) har ett adjektivattribut (AT).



Också prepositionsfraser som bestämmer predikatet medverkar typiskt till meningskomplexiteten i de naturvetenskapliga lärobokstexterna. Meningen i exempel (116), *Isbjörnarna lever på isen runt Nordpolen*, kommer från en lärobok i biologi och kan exemplifiera en sådan struktur (post 4 på listan).

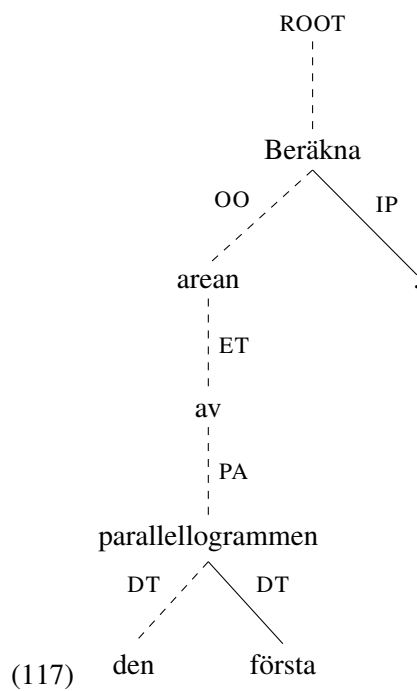


Den kompletterande indexlistan över syntaktiska strukturer i matematiktexter för högstadiet visas i tabell 9.3. Denna lista, som också har konstruerats för strukturer med djupet 5, innehåller relativt många mönster. Ungefär två tredjedelar av dessa är efterställda predikatsverbet, medan resten är framförställda.

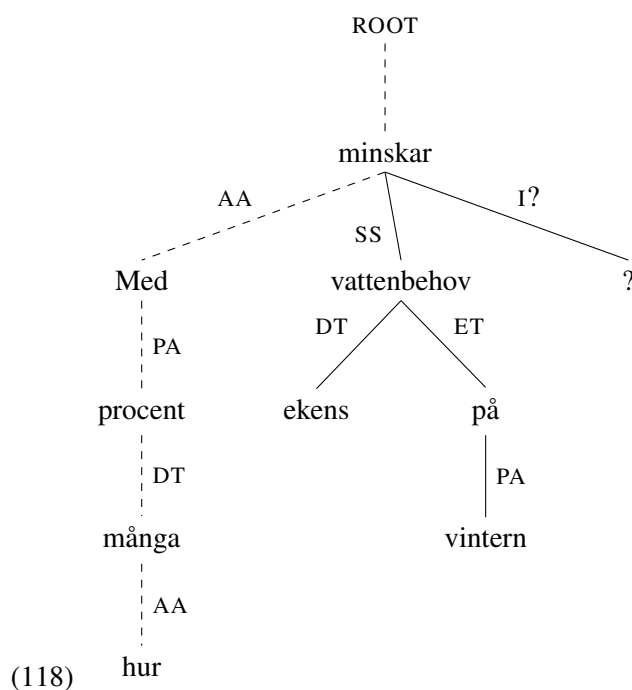
	Index	Vänster	Höger
1	1,75		OO: ET, PA, DT
2	1,35		SS: ET, PA, DT
3	1,27	SS: DT, DT, AA	
4	1,23	AA: PA, DT, AA	
5	1,22	SS: ET, PA, DT	
6	1,17		SP: AA, AA, DT
7	1,16		VG: OA, PA, DT
8	1,13		VG: OO, DT, AA
9	1,13		+F: OA, PA, DT
10	1,12		OO: UA, OO, DT

Tabell 9.3: Indexlista över meningsstrukturer i matematiktexter för högstadiet (jämfört med naturvetenskapliga lärobokstexter)

De mest typiska strukturerna förekommer i nominalfraser i direkt objektiv funktion eller i subjektsställning. Exempelmeningen i (117), *Beräkna arean av den första parallelogrammen*, har komplexiteten koncentrerad kring ett direkt objekt, på samma vis som i struktur nummer 1 på listan.



Resten av listan visar att komplexiteten i matematikmeningarna är av ganska olika karaktär. Visserligen pekar flera mönster ut nominala satsdelar, men det maximala parsningsdjupet knyts också ofta till meningens predikat, i form av komplexa adverbial och/eller verbgrupper. Parsträdet som anges i exempel (118) visar prov på adverbiell komplexitet till vänster om det finita verbet (struktur 4 på listan). Meningen utläses: *Med hur många procent minskar ekens vattenbehov på vintern?*



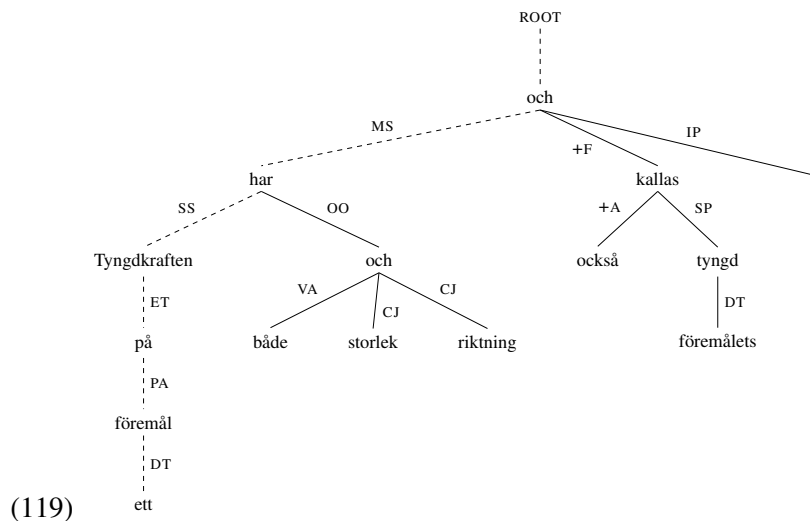
Just denna konstruktion verkar i samtliga fall tillhöra frågesatser av typen "PP hur många NN". Några fler meningar av detta slag är: *På hur många sätt kan du rita triangeln?* *Med hur många centimeter slog hon sitt gamla rekord?* och *Vid hur många träningsstillfällen kostar de båda gymmen lika mycket?*

De meningsstrukturer som faller ut i jämförelsen mellan naturvetenskapliga texter och matematiktexter för gymnasiet syns i tabell 9.4. Listan är relativt kort och indexen för mönstren är förhållandevis låga. Denna gång är det 6 noder långa typiska mönster som har extraherats.

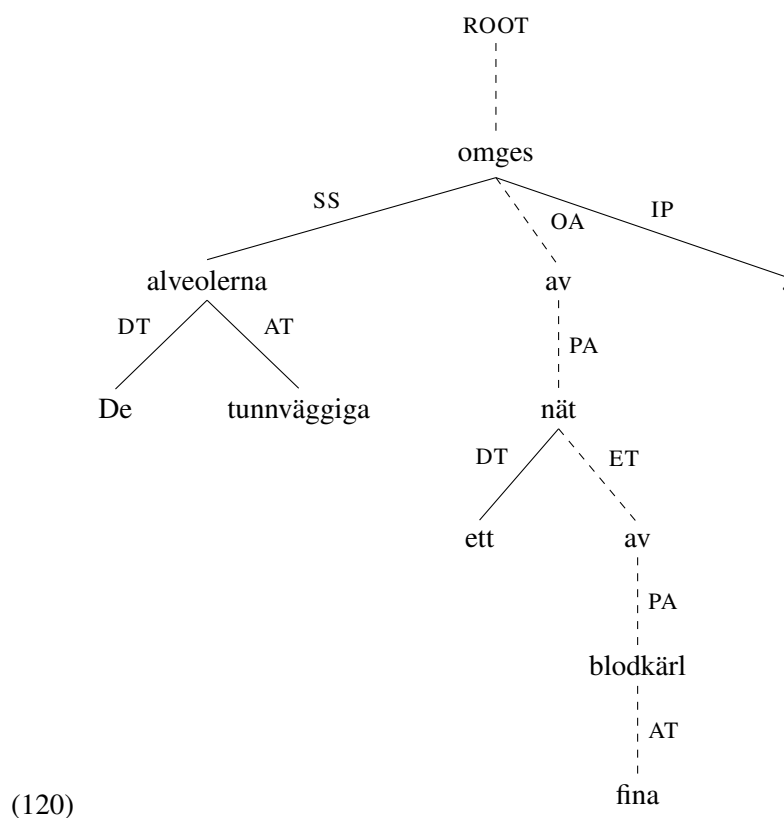
	Index	Vänster	Höger
1	1,12	MS: SS, ET, PA, DT	
2	1,12		OA: PA, ET, PA, AT
3	1,10		VG: OO, ET, PA, AT
4	1,10	SS: ET, PA, ET, PA	

Tabell 9.4: Indexlista över meningsstrukturer i naturvetenskapliga lärobokstexter för gymnasiet (jämfört med matematiktexter)

Den översta strukturen inkluderar makrosyntagmer till vänster om roten, vilket gör att meningarna blir långa. Parträdet i exempel (119) visar en mening med denna struktur. Meningen utläses: *Tyngdkraften på ett föremål har både storlek och riktning och kallas också föremålets tyngd* och kommer från en fysikbok.



Nummer 2 på listan beskriver meningar vars komplexitet är uppbyggd kring ett efterställt prepositionsobjekt tillhörande predikatet. En sådan mening från en biologibok visas i exempel (120) och utläses: *De tunnväggiga alveolerna omges av ett nät av fina blodkärl.*

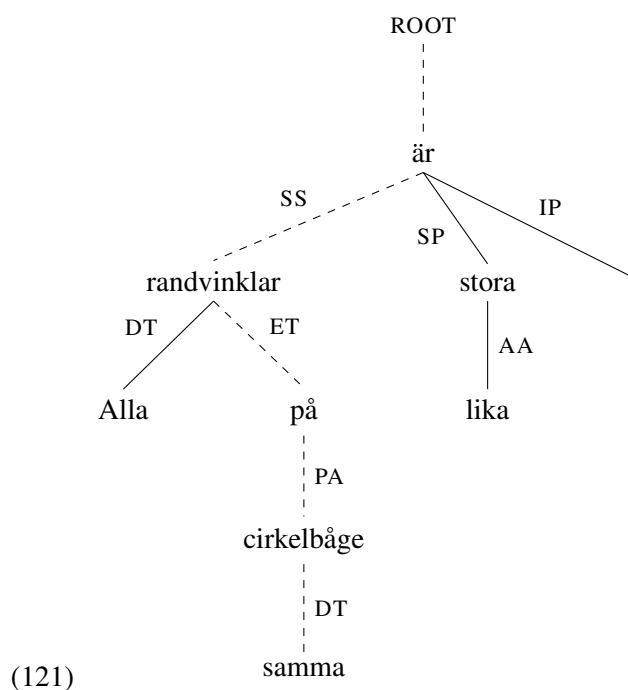


Den sista indexlistan i jämförelsen mellan naturvetenskap och matematik beskriver typiska meningsstrukturer med djupet 5 i matematiktexter för gymnasiet (se tabell 9.5). Liksom för högstadiet är matematiklistan betydligt längre och innehåller många olika slags strukturer. Den största delen av komplexiteten är högerställd och ungefär hälften av posterna pekar ut utbyggda nominalfraser.

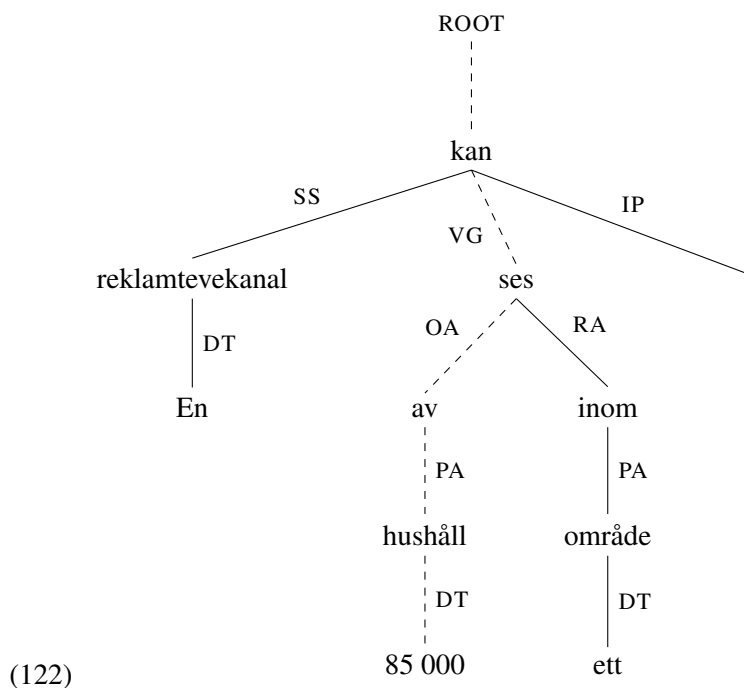
	Index	Vänster	Höger
1	1,62		OO: ET, PA, DT
2	1,24		SP: ET, PA, DT
3	1,23	MS: VG, OO, DT	
4	1,22	SS: ET, PA, DT	
5	1,18		VG: OA, PA, DT
6	1,18		SS: ET, PA, DT
7	1,17		VG: VG, OO, DT
8	1,13		SS: ET, PA, CJ
9	1,11		OA: PA, IF, OO
10	1,10		AA: PA, IF, OO

Tabell 9.5: Indexlista över meningsstrukturer i matematiktexter för gymnasiet (jämfört med naturvetenskapliga lärobokstexter)

Nummer 1 på listan kvarstår från högstadiet. Meningar med denna typiska struktur förekommer i uppmaningar i räkneuppgifter. Några exempel är: *Beräkna summan av de tre avbetalningarna*, *Ge också ett närmevärde med tre decimaler* och *Bestäm summan av de fyra talen*. En annan struktur som återkommer från högstadielistan är nummer 4 på listan. Denna exemplifieras med meningen *Alla randvinklar på samma cirkelbåge är lika stora* i exempel (121). Meningar som byggs upp enligt detta mönster tycks användas för att beskriva antingen scenarion i räkneuppgifter eller matematiska samband. Några fler exempel lyder: *Antalet gula trianglar i hela konstruktionen är summan av radnumren*, *Priset på en dator var 5 995 kr* och *Kvadraten av ett udda tal är udda*.



Också nästföljande mönster på listan, adverbialt bestämda verbgrupper, svarar mot en vanlig meningstyp som ofta förekommer i syfte att förklara olika enheters beskaffenhet. Exempelmeningen *En reklamtevekanal kan ses av 85 000 hushåll inom ett område* i (122) har denna struktur.⁵⁴ Ytterligare exempel är: *Lånet ska betalas tillbaka på 4 år*, *En godtycklig fyrrhörning kan alltid delas upp i två trianglar* och *Arean under en graf kan beskrivas med begreppet integral*.



⁵⁴Observera att dubbeltydiga meningar, i bästa fall, resulterar i flera alternativa träd. I detta fall har den avslutande prepositionsfrasen analyserats som rumsadverbial. En analys av prepositionsfrasen som attribut till "hushåll" hade eventuellt resulterat i en mer lättolkad analys av meningen. En sådan analys hade också förändrat maxdjupets värde. Det kan dock vara värt att hålla i minnet att en mening inte med nödvändighet endast har **en** syntaktisk tolkning eller *ett* parsträd. Det parsträd som väljs är det som beräknas vara mest sannolikt utifrån parametrar som bestäms av lexikonet, parsningsalgoritmen och träningsdata.

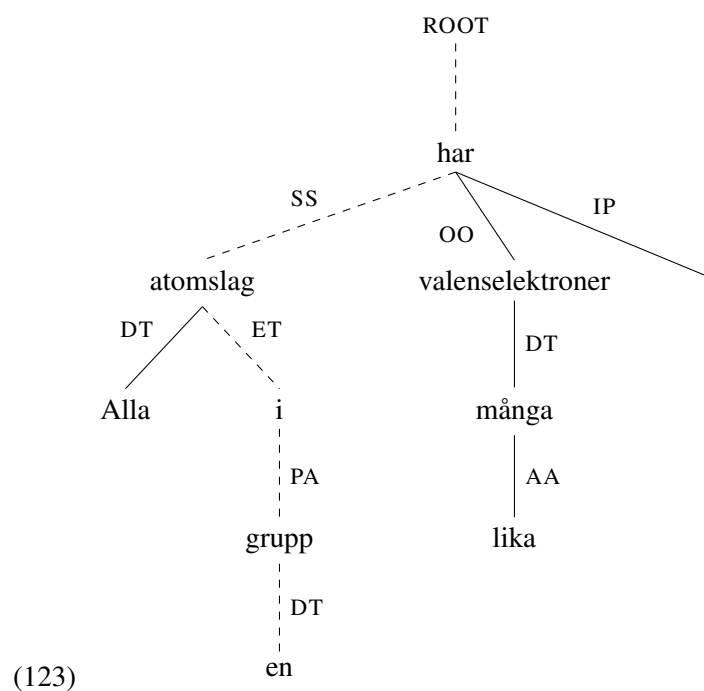
9.2.2.2 Naturvetenskapliga texter jämförda med samhällsvetenskapliga texter

När meningar i naturvetenskapliga läroböcker för högstadiet jämförs med meningar i samhällsvetenskapliga läroböcker faller endast ett mönster ut för parningsdjup 5 (se tabell 9.6).

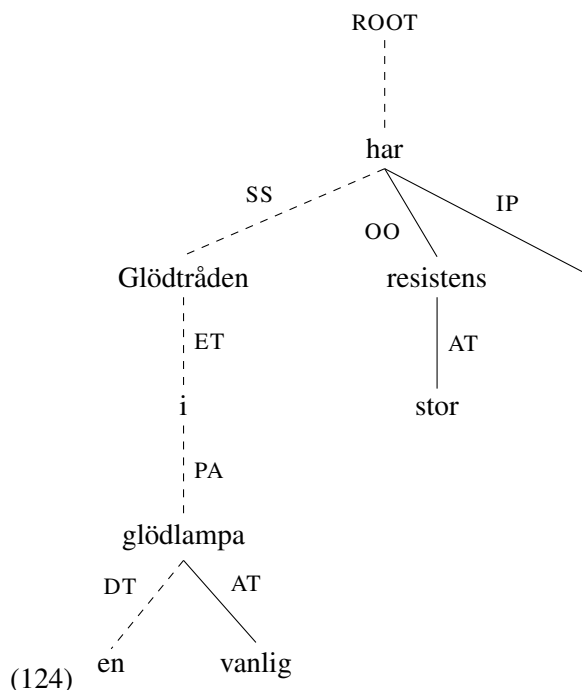
	Index	Vänster	Höger
1	1,11	SS: ET, PA, DT	

Tabell 9.6: Indexlista över meningsstrukturer i naturvetenskapliga lärobokstexter för högstadiet (jämfört med samhällsvetenskapliga lärobokstexter)

Denna struktur gäller meningar med komplexiteten lokaliserad till ett initialt placerat subjekt som bestäms med en prepositionsfras och visas närmare i exempel (123) och (124). Exempelmeningen i (123) utläses: *Alla atomslag i en grupp har lika många valenselektroner* och är hämtad ur en lärobok i kemi.



Exempelmeningen i (124), *Glödtråden i en vanlig glödlampa har stor resistans*, kommer från en lärobok i fysik.



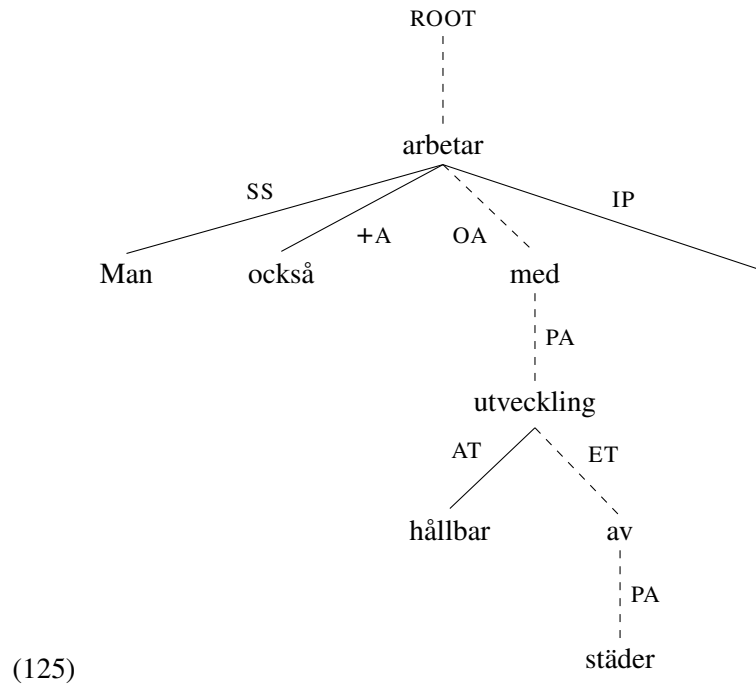
Listan över typiska strukturer med djupet 5 i samhällsvetenskapliga texter för högstadiet kan ses i tabell 9.7. Denna innehåller något fler poster, om än inte alls lika många som matematiklistorna. Tydligt liknar naturvetenskap syntaktiskt sett mer samhällsvetenskap än matematik.

	Index	Vänster	Höger
1	1,13		OA: PA, ET, PA
2	1,12		MS: RA, PA, DT
3	1,12		SS: ET, PA, CJ
4	1,10	TA: PA, ET, PA	

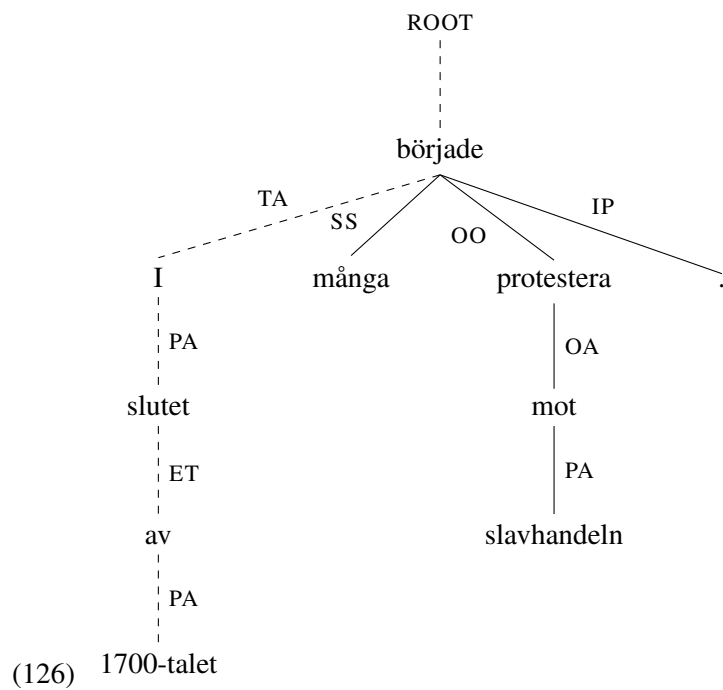
Tabell 9.7: Indexlista över meningsstrukturer i samhällsvetenskapliga lärobokstexter för högstadiet (jämfört med naturvetenskapliga lärobokstexter)

Värt att notera är att bara ett av mönstren på indexlistan utgörs av en nominalfras. Komplexiteten i de samhällsvetenskapliga texterna bör alltså inte huvudsakligen kopplas till utbyggda nominalfraser, utan i stället till utbyggda

adverbial (prepositionsfraser). Exempel (125) visar en mening med ett komplext objektsadverbial, vars struktur motsvarar listans översta post. Meningen kommer från en geografibok och utläses: *Man arbetar också med hållbar utveckling av städer.*



Exempel (126), *I slutet av 1700-talet började många protestera mot slavhandeln*, är hämtat ur en historiebok och visar den typiska meningsstrukturen med ett utbyggt tidsadverbial i fundamentet.



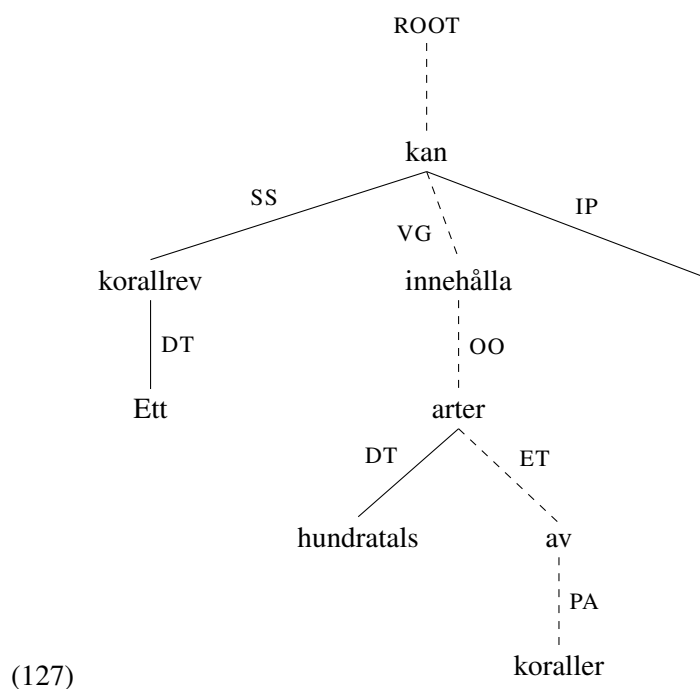
9.2.2.3 Biologitexter på högstadiet och gymnasiet

Tabell 9.8 visar meningsstrukturer med djupet 5 som är typiska för högstadie-
texter i biologi. Samtliga mönster förekommer efter predikatsverbet och är
jämnt fördelade över verb- och nominalfraser.

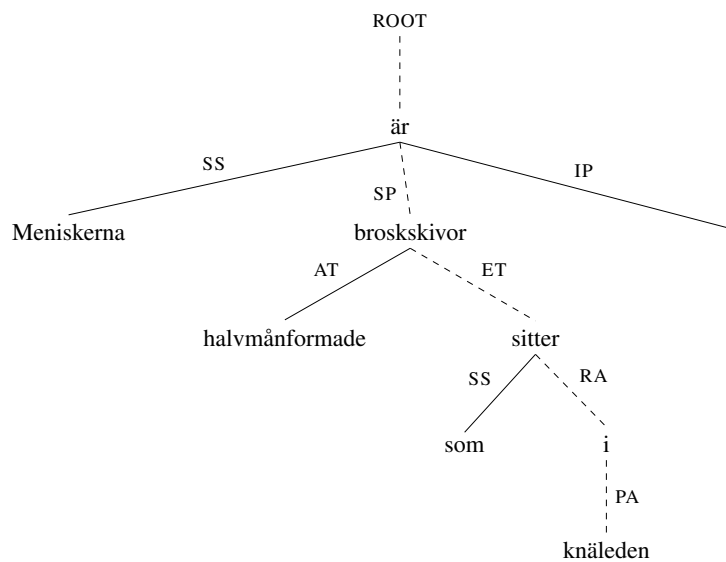
	Index	Vänster	Höger
1	1,17		VG: OO, ET, PA
2	1,11		SP: ET, PA, DT
3	1,11		SP: ET, RA, PA
4	1,11		VG: OA, PA, CJ

Tabell 9.8: Indexlista över meningsstrukturer i biologitexter för högstadiet (jämfört
med för gymnasiet)

Den mest typiska konstruktionen exemplifieras i (127) med meningen *Ett
korallrev kan innehålla hundratals arter av koraller.*



Exempel (128) visar hur nominalfrasmonstret på plats nummer 3 realiseras i meningen *Meniskerna är halvmånformade broskskivor som sitter i knäleden.*



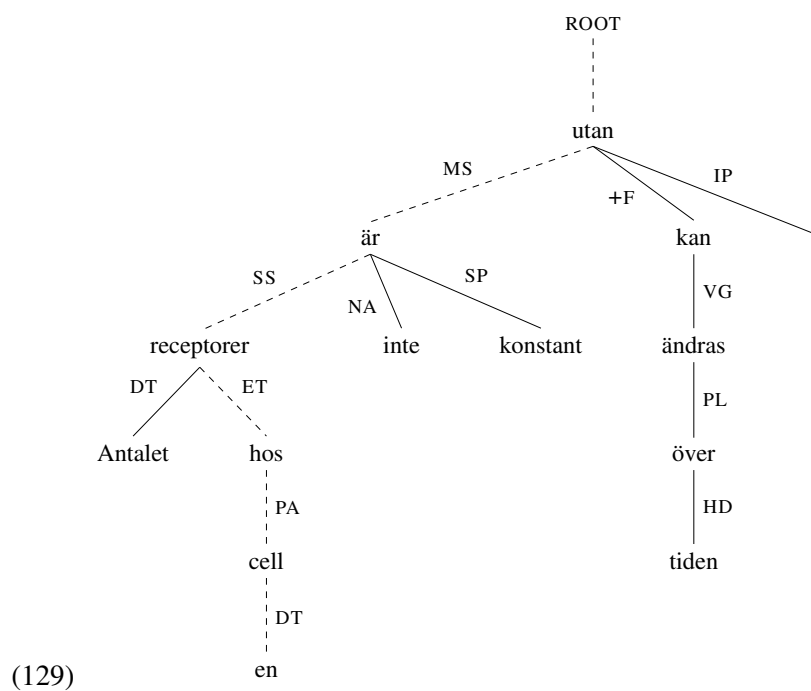
(128)

Gymnasielistan för 6 noder djupa meningsstrukturer i biologi kan ses i tabell 9.9.

	Index	Vänster	Höger
1	1,12	MS: SS, ET, PA, DT	

Tabell 9.9: Indexlista över meningsstrukturer i biologitexter för gymnasiet (jämfört med för högstadiet)

Listan innehåller bara ett mönster som ingår i en makrosyntaxtagm på huvudsatsnivå. En mening med denna struktur visas i exempel (129). Ordalydelsen är: *Antalet receptorer hos en cell är inte konstant utan kan ändras över tiden.*



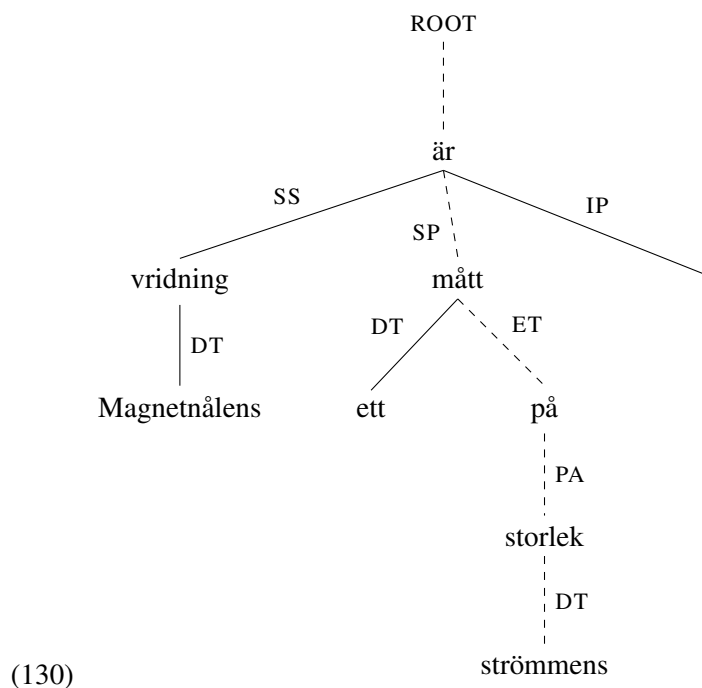
9.2.2.4 *Fysiktexter på högstadiet och gymnasiet*

I tabell 9.10 presenteras indexlistan över typiska syntaktiska mönster med djupet 5 i fysiktexter för högstadiet. Förteckningen innehåller tre poster, varav två utgörs av nominalfrasstrukturer.

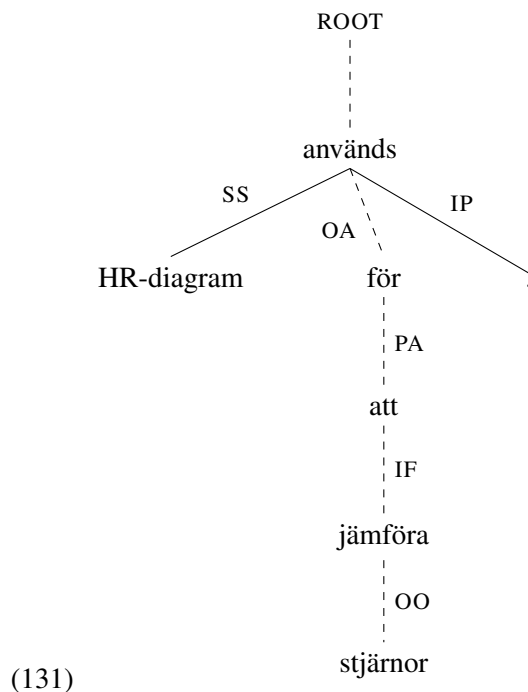
	Index	Vänster	Höger
1	1,15		SP: ET, PA, DT
2	1,14		OA: PA, IF, OO
3	1,11	SS: ET, RA, PA	

Tabell 9.10: Indexlista över meningsstrukturer i fysiktexter för högstadiet (jämfört med för gymnasiet)

I den mest typiska meningsstrukturen återfinns den maximala komplexiteten i subjektspredikativ ställning, ett mönster som även fanns med på biologilistan för högstadiet. Strukturen visas i exempel (130) i meningen *Magnetnålens vridning är ett mått på strömmens storlek*.



Den näst mest typiska strukturen bestämmer predikatsverbet med hjälp av en infinitivfras och exemplifieras i (131) med meningen *HR-diagram används för att jämföra stjärnor*.



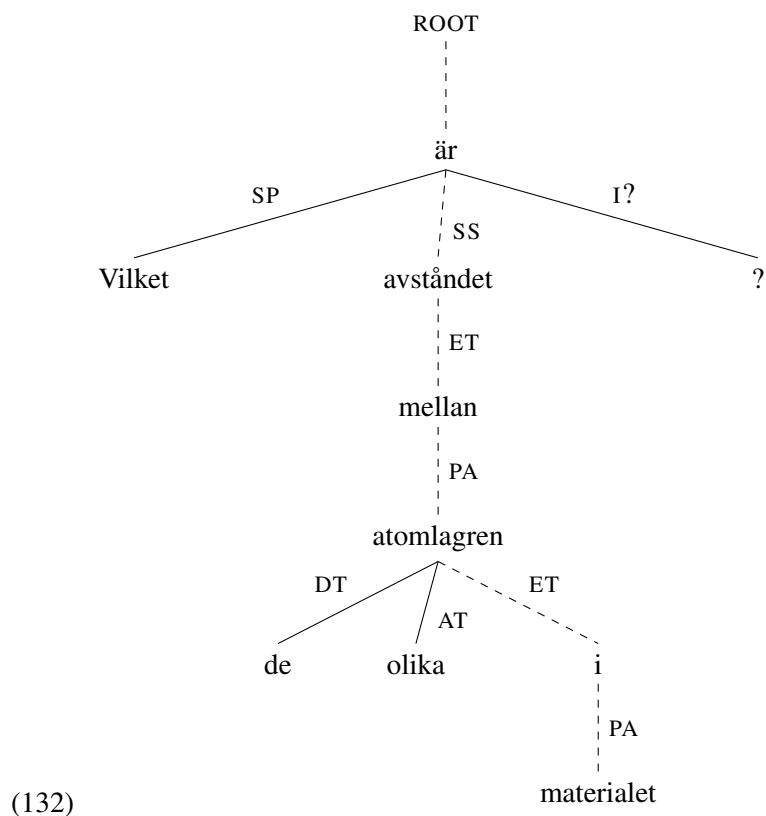
Gymnasielistan över 6 noder djupa mönster i fysik (se tabell 9.11) är relativt lång. Den domineras av ett flertal makrosyntagmer på huvudsatsnivå, i vilka komplexiteten återfinns i olika satsfunktioner: subjekt, objekt och predikat. Dessutom finns både framför- och efterställda nominalfraser i subjektspostion med två prepositionsfrasattribut representerade på listans två översta poster.

210 *Delstudie: typiska meningar*

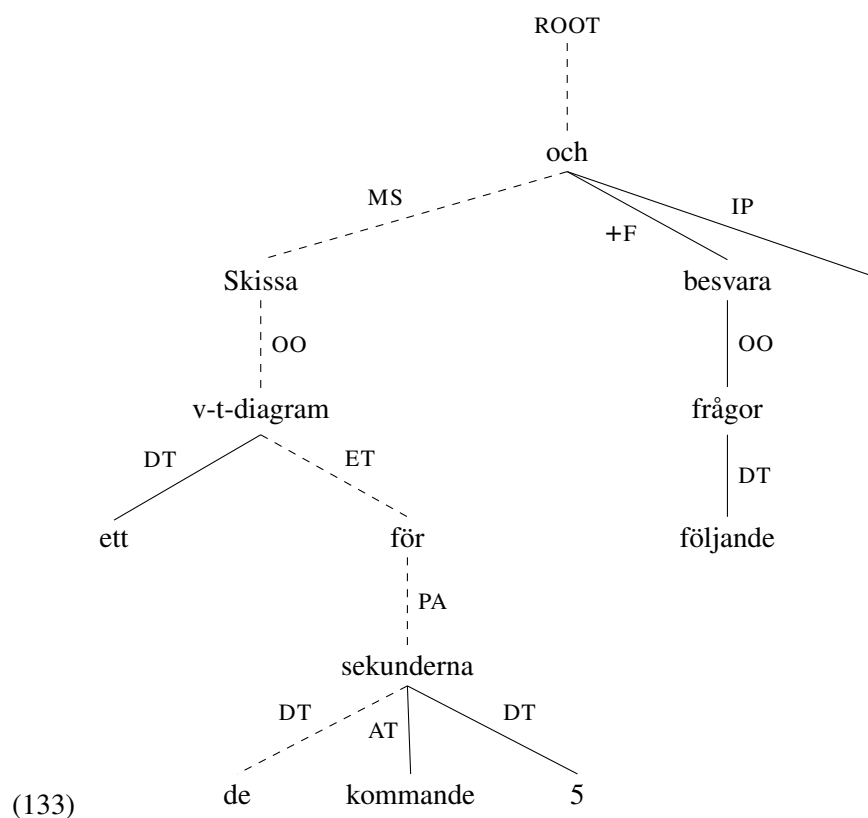
	Index	Vänster	Höger
1	1,19		SS: ET, PA, ET, PA
2	1,16	SS: ET, PA, ET, PA	
3	1,13	MS: OO, ET, PA, DT	
4	1,13		OO: ET, PA, DT, DT
5	1,12	MS: SS, ET, PA, AT	
6	1,11	MS: VG, OA, PA, AT	
7	1,10	MS: SS, ET, PA, DT	

Tabell 9.11: Indexlista över meningsstrukturer i fysiktexter för gymnasiet (jämfört med för högstadiet)

Den översta strukturen visas i exempel (132) med meningen *Vilket är avståndet mellan de olika atomlagren i materialet?*



Exempel (133) visar den mest typiska strukturen bland makrosyntagmerna (post 3), där komplexiteten finns inom det direkta objektet. Meningen lyder: *Skissa ett v-t-diagram för de kommande 5 sekunderna och besvara följande frågor.*



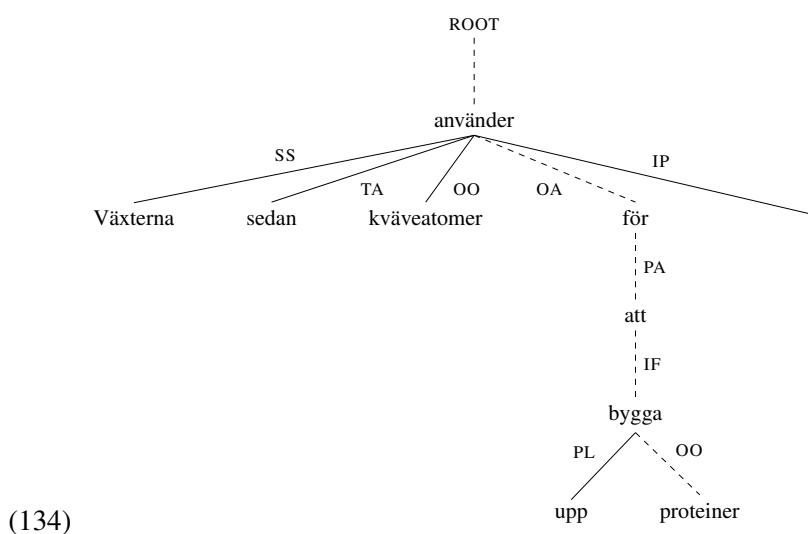
9.2.2.5 *Kemitexter på högstadiet och gymnasiet*

Kemitexterna för högstadiet innehåller tre typiska meningsstrukturer med djupet 5, som beskrivs i tabell 9.12. Som synes återfinns samtliga strukturer till höger om roten i meningen. Samtliga innehåller också adverbiala bestämningar. De två översta mönstren tillhör verbfraser och den sista strukturen, som för övrigt också fanns med på högstadielistan i biologi, utgörs av en nominalfras.

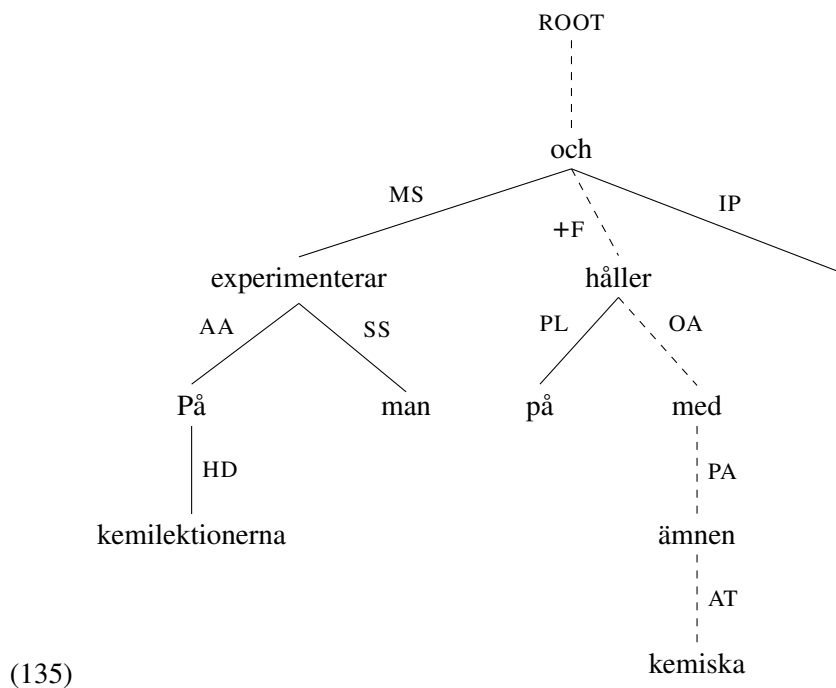
	Index	Vänster	Höger
1	1,14		OA: PA, IF, OO
2	1,11		+F: OA, PA, AT
3	1,11		SP: ET, RA, PA

Tabell 9.12: Indexlista över meningsstrukturer i kemitexter för högstadiet (jämfört med för gymnasiet)

Det översta mönstret finns i meningen *Växterna använder sedan kväveatomer för att bygga upp proteiner*, vars struktur visas i exempel (134).



Exempelmeningen i (135), *På kemilektionerna experimenterar man och håller på med kemiska ämnen*, åskådliggör den näst mest typiska syntaktiska strukturen på listan.

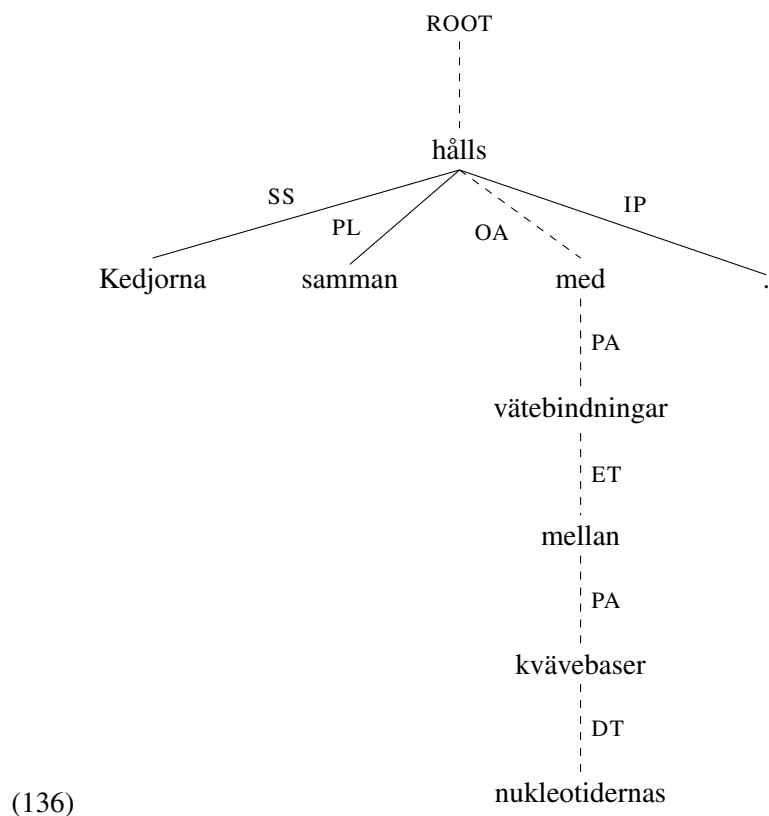


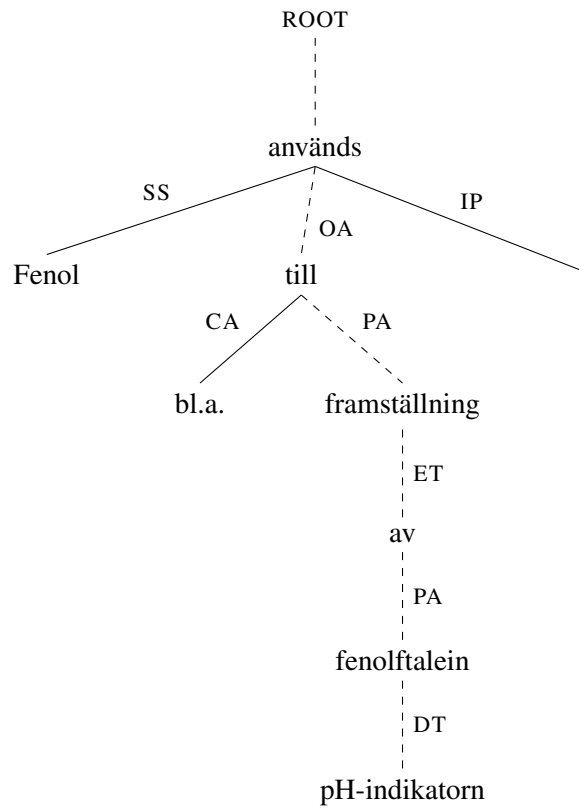
Indexlistan över 6 noder djupa syntaktiska mönster i gymnasietexter i kemi visas i tabell 9.13. Som synes är det bara objektsadverbial av en viss typ som utmärker sig särskilt.

	Index	Vänster	Höger
1	1,20		OA: PA, ET, PA, DT

Tabell 9.13: Indexlista över meningsstrukturer i kemitexter för gymnasiet (jämfört med för högstadiet)

Detta mönster kan ses i användning i parsträden för meningarna *Kedjorna hålls samman med vätebindningar mellan nukleotidernas kvävebaser* respektive *Fenol används till bl.a. framställning av pH-indikatorn fenolftalein*, vilka visas i exempel (136) och (137).





(137)

9.2.2.6 *Naturvetenskapliga texter jämförda med akademiska texter*

Meningsstrukturer som är typiska för naturvetenskapliga lärobokstexter, i jämförelse med akademiska texter, indexlistas i tabell 9.14. Eftersom såväl högstadie- som gymnasietexter ingår i jämförelsen har mönster av både längd 5 och 6 tillåtits på listan för naturvetenskap.

	Index	Vänster	Höger
1	1,24		OO: ET, PA, DT
2	1,22	SS: ET, PA, DT	
3	1,15		OA: PA, ET, PA
4	1,15		SS: ET, PA, DT
5	1,15		VG: OA, PA, DT
6	1,14		VG: OO, ET, PA
7	1,13	MS: SS, ET, PA	
8	1,12		OO: ET, PA, AT
9	1,12		VG: OO, ET, PA, AT
10	1,12		VG: OO, ET, PA, DT
11	1,11		SP: ET, PA, DT
12	1,12		OA: PA, UA, SS, DT
13	1,10		OA: PA, ET, SS

Tabell 9.14: Indexlista över meningsstrukturer i naturvetenskapliga lärobokstexter (jämfört med akademiska texter)

Flera av strukturerna på indexlistan i fråga har redan demonstrerats med exempel. Därför visas inga nya exempel i direkt anslutning till just denna lista. Den intresserade läsaren hänvisas i stället till exempel (117) för struktur nummer 1 på listan, till exempel (123) och (124) för nummer 2, till exempel (116) för nummer 3, till exempel (127) för nummer 6, och slutligen till exempel (130) för nummer 11.

De strukturer som är typiska i de akademiska texterna anges i tabell 9.15. Tabellen listar mönster som innehåller 7 noder, vilket är medianvärdet för meningarnas parsningsdjup i SveAk-korpusen.

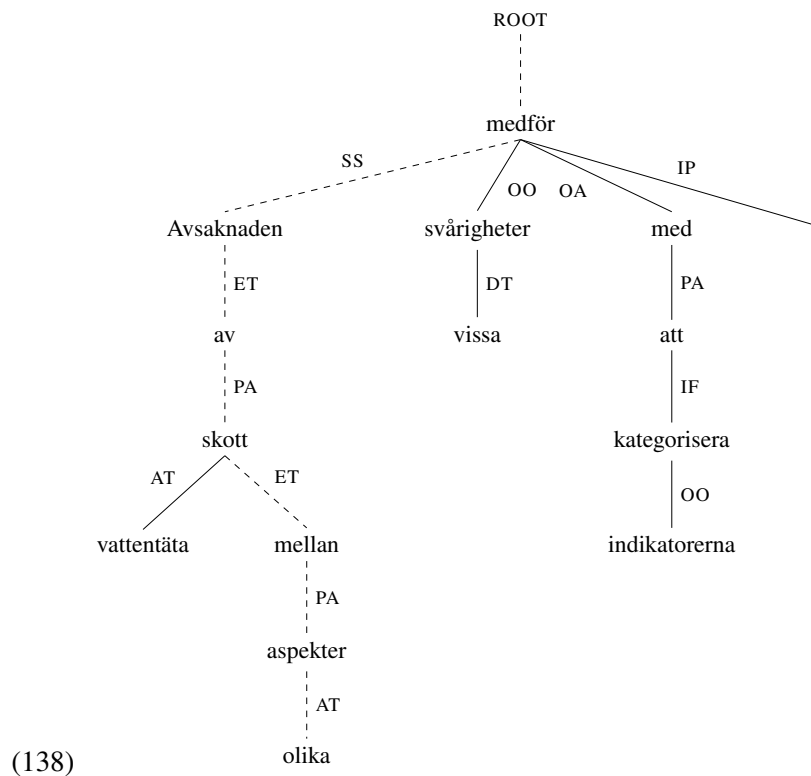
	Index	Vänster	Höger
1	1,18		VG: SP, *UA, ET, PA, DT
2	1,15	SS: ET, PA, ET, PA, AT	
3	1,14	SS: ET, PA, ET, PA, DT	
4	1,14		VG: SP, *UA, ET, PA, AT
5	1,13		OA: PA, ET, PA, CJ, AT
6	1,12		SS: UA, SS, ET, PA, DT
7	1,12		OO: UA, SS, ET, PA, AT
8	1,11		VG: OA, PA, ET, PA, AT
9	1,11		SS: ET, PA, ET, PA, AT
10	1,11		OO: ET, PA, ET, PA, CJ

Tabell 9.15: Indexlista över meningsstrukturer i akademiska texter (jämfört med naturvetenskapliga lärobokstexter)

Den översta strukturen visar i stora drag samma typ av meningsstruktur som nummer 4 på listan. Några exempel på meningar med denna struktur är: *Människors sociala system kan därmed förstås som en strukturell koppling mellan människans eget medvetande och psyken i hennes omvärld* och *Avhandlingens insatsteori kan inte läsas som en introduktion till Vygotskijs teorier*. Exempelen följer alla mallen "SS, FV, VB i passivum, som, NP", vilket uppenbarligen är en typisk meningsuppbyggnad för just akademiska texter. Däremot analyseras meningarna systematiskt fel; bisatsanalysen stämmer inte eftersom inget finit verb kan sättas in efter "som" (denna felaktighet i analysen markeras med en asterisk i listan). I själva verket rör det sig om en prepositionsfras med inledande "som". Denna fras kan dock med fördel sägas fungera som just subjektspredikativ (och inte som fritt adverbial), eftersom den kongruerar med meningens subjekt.

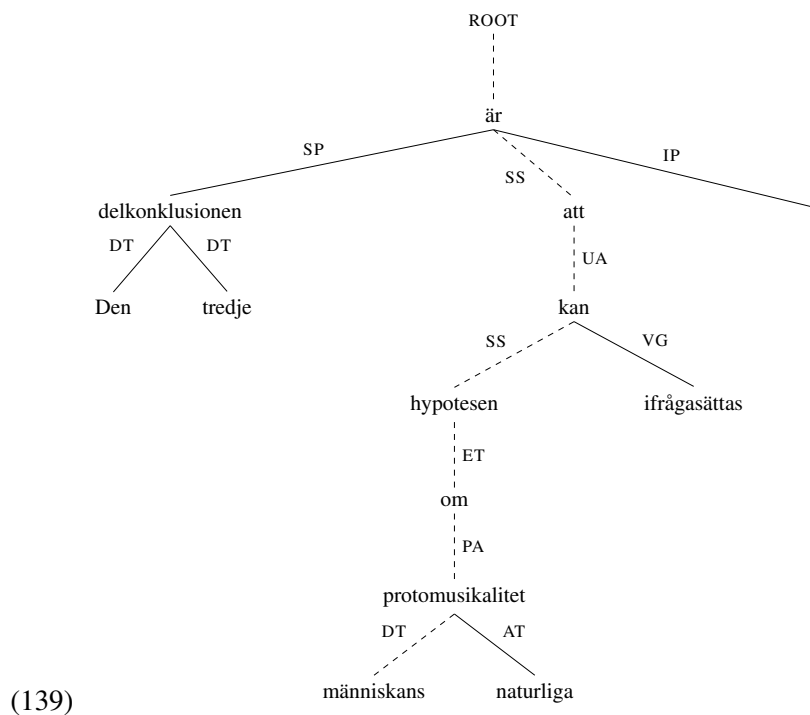
Ett annat tydligt resultat speglas av mönster nummer 2 och 3 på listan. Dessa beskriver kraftigt utbyggda vänsterställda subjekt, som endast skiljer sig åt med avseende på den djupast liggande noden. Ett av de främsta kännetecknen för akademiska texter är uppenbarligen förhållandevis svårlästa meningar.

En mening som är uppbyggd på detta vis är: *Avsaknaden av vattentäta skott mellan olika aspekter medför vissa svårigheter med att kategorisera indikatorerna*, som visas i exempel (138).⁵⁵ Strukturen återfinns för övrigt också (i högerställd position) på plats 9 på listan.



⁵⁵I exemplet har den avslutande prepositionsfrasen analyserats som objektsadverbial, vilket resulterar i en svårtolkad läsning. Prepositionsfrasen borde hellre attributivt bestämma "svårigheter", en analys som dock inte förändrar maxdjupets värde.

Två av mönstren på indexlistan för akademiska texter innehåller bisatser (UA). Mönstren för bisatsernas uppbyggnad är likadana, men de skiljer sig åt genom att stå i subjekt- eller objektsposition (post 6 och 7). Exempel (139) visar meningen *Den tredje delkonklusionen är att hypotesen om människans naturliga protomusikalitet kan ifrågasättas* med ett bisatsbestämt subjekt.



9.2.2.7 *Naturvetenskapliga texter jämförda med berättande texter*

I jämförelsen mellan naturvetenskapliga och berättande texter ingår lärobokstexter från både högstadie- och gymnasieböcker. Därför tas mönster av både längd 5 och 6 upp på listan för naturvetenskap. Denna lista visas i tabell 9.16.

	Index	Vänster	Höger
1	1,78	SS: ET, PA, DT	
2	1,40		SS: ET, PA, DT
3	1,38		OO: ET, PA, DT
4	1,27	SS: ET, PA, AT	
5	1,26		OO: ET, PA, AT
6	1,24		SS: ET, PA, AT
7	1,20		SP: ET, PA, DT
8	1,20		OA: PA, DT, DT
9	1,19	MS: SS, ET, PA	
10	1,18		OA: PA, ET, PA, DT
11	1,18		OO: ET, PA, CJ
12	1,17		VG: OA, PA, AT
13	1,16		SP: ET, PA, AT
14	1,15		SS: ET, PA, ET, PA
15	1,14	SP: ET, PA, DT	
16	1,14		SS: ET, PA, CJ
17	1,13		VG: OA, PA, DT
18	1,13	SS: ET, PA, ET, PA	
19	1,13	MS: SS, ET, PA, DT	
20	1,12		OO: ET, PA, ET, PA
21	1,12		VG: OO, ET, PA, AT
22	1,12		OO: UA, SS, DT
23	1,11		OA: PA, ET, PA, AT
24	1,11	SS: ET, PA, CJ	
25	1,11	AA: PA, ET, PA	
26	1,11		SP: ET, PA, CJ

Tabell 9.16: Indexlista över meningsstrukturer i naturvetenskapliga lärobokstexter (jämfört med berättande texter)

Även om mönstren ter sig ganska mångskiftande vid en första anblick, kan man lägga märke till att komplexitet i nominalfraser ständigt återkommer. För det första representerar de allra flesta strukturerna nominala satsled, men även adverbialen och verbgrupperna på listan blir komplexa med hjälp av nominalfraser (d.v.s. rektionerna, PA, följs av olika attribut i listan).

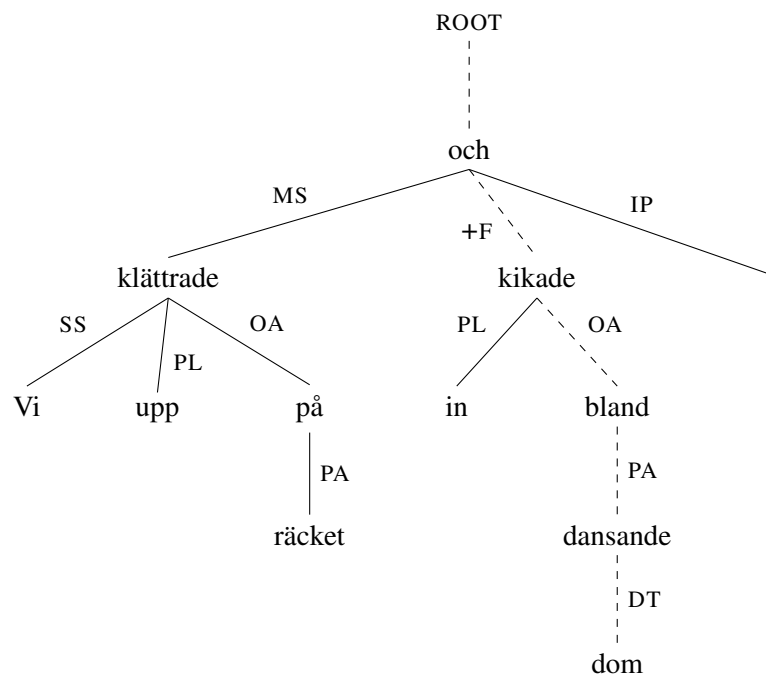
Även denna gång visar det sig att flera av strukturerna redan har exemplifierats. Jag hänvisar därmed till exempel (123) och (124) för nummer 1 på listan, till exempel (115) för nummer 6, till exempel (130) för nummer 7, till exempel (136) och (137) för nummer 10, till exempel (132) för nummer 18, till exempel (119) och (129) för nummer 19, och till exempel (120) för nummer 23.

Betraktar man resultatet för den omvända jämförelsen, i tabell 9.17, blir överskottet av komplexa nominalfraser i läroböckerna ännu tydligare. På indexlistan över typiska meningsstrukturer av längd 5 i berättande texter finns, till att börja med, inte en enda nominalfras. I stället gäller samtliga mönster koordinerade huvudsatser. Vidare är inte nominalfraserna som ingår i mönstren på listan tillnärmelsevis lika utbyggda.

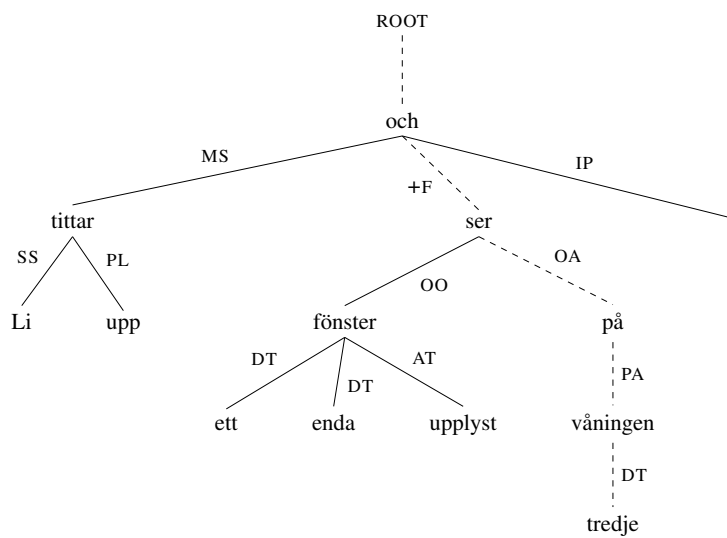
	Index	Vänster	Höger
1	1,16		+F: OA, PA, DT
2	1,13		MS: +F, OA, PA
3	1,13		MS: VG, OA, PA
4	1,11		MS: OA, PA, DT
5	1,11		MS: VG, OO, DT

Tabell 9.17: Indexlista över meningsstrukturer i berättande texter (jämfört med naturvetenskapliga lärobokstexter)

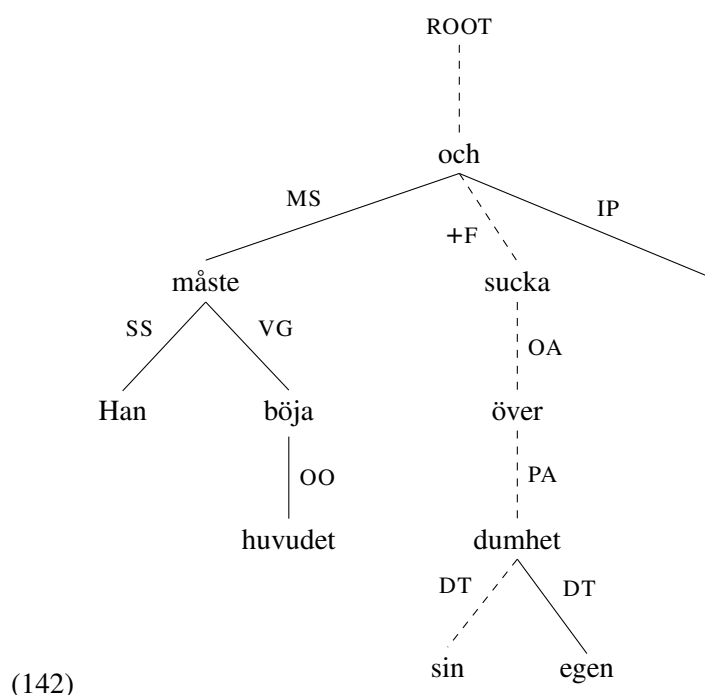
Det mest typiska syntaktiska draget för de berättande texterna exemplifieras i (140), (141) och (142). Meningarna kommer, i tur och ordning, från en barnbok, en ungdomsroman och en vuxenroman och lyder: *Vi klättrade upp på räcket och kikade in bland dom dansande*, *Li tittar upp och ser ett enda upplyst fönster på tredje våningen* och *Han måste böja huvudet och sucka över sin egen dumhet*.



(140)



(141)



9.2.3 Fokusstudier

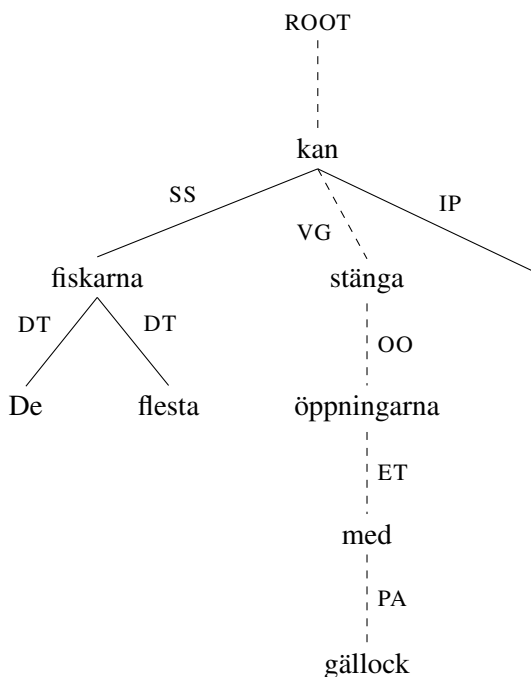
9.2.3.1 Parsning av prepositionsfrasers syntaktiska funktion

Eftersom den automatiska syntaxanalysen är det mest sofistikerade steget i den analysapparat som jag tillämpar på mina texter måste korrektheten i analyserna utvärderas, innan några definitiva slutsatser kan dras. En relativt enkel kontroll kan göras genom att manuellt läsa igenom alla meningar som har tilldelats ett visst mönster och notera och kommentera ifall betänkligheter upptäcks. Sådana kontroller har gjorts löpande och på det hela taget uppskattar jag att parserns korrekthet är tillräckligt god för att låta de automatiska analyserna ligga till grund för slutsatser.

En speciell problematik, som är värd att kommenteras, gäller en välkänd utmaning inom språkvetenskaplig databehandling. Det handlar om att lära en dator att entydigt och korrekt avgöra prepositionsfrasers beroendeställning, d.v.s. huruvida de ska tolkas som attribut eller som adverbial, en uppgift som sällan vållar problem för oss människor.

För att tydliggöra vad problemet handlar om visas i exempel (143) den analys som har gjorts av meningen *De flesta fiskarna kan stänga öppningarna med gällock*. Meningen kommer från en högstadiesbok i biologi och påstås matcha

det mest typiska mönstret, "VG: OO, ET, PA". Som synes har prepositionsfrasen analyserats som attribut till "öppningarna", i stället för den troligare tolkningen med prepositionsfrasen som adverbial till "stänga".



(143)

För att närmare utreda omfattningen av problemet har jag gjort en stickprovsundersökning av 100 slumpmässigt utvalda meningar från naturvetenskapliga lärobokstexter, vars syntaktiska analys innehåller en prepositionsfras i maxdjupet. Detta kriterium uppfylls till att börja med av 41 % av alla meningar i materialet. Den sammanlagda korrektheten för parsern i att avgöra om prepositionsfrasen är ett attribut eller ett adverbial uppskattas till omkring 84 %. Det visar sig vidare att korrektheten i att tagga prepositionsfraser som adverbial är mycket högre (uppskattningsvis omkring 100 %) än korrektheten att tagga prepositionsfraser som attribut (uppskattningsvis omkring 80 %). Dessutom är det något enklare för parsern att korrekt tagga vänsterställda prepositionsfraser i attributiv funktion (uppskattningsvis omkring 87 %) än högerställda (uppskattningsvis omkring 77 %).

Exemplet ovan (i 143) visar sig alltså vara typiskt för det sätt på vilket felen oftast görs; parsern verkar föredra en analys med attributiva prepositionsfraser framför adverbiala. Som har antytts tidigare beror parserns val på flera olika

parametrar, men just för detta ändamål spelar träningsdata en stor roll. I de texter som parsern har tränats på är det förmodligen vanligare att prepositionsfraser ingår i nominalfraser än i verbfraser.

Jag har också utfört några tester för att försäkra mig om att de indexlistor som presenteras i detta kapitel inte påverkas av den felmarginal som parsningen av prepositionsfraser och adverbial rör sig inom. I testerna räknades frekvensen för prepositionsfraser ner och frekvensen för olika typer av adverbial upp, i enlighet med stickprovsindikationerna om parserns över- respektive underskattning av dessa olika syntaktiska funktioner. Sammanlagt visar testerna att samtliga strukturer på indexlistorna i detta kapitel ändå hamnar på listorna. Värt att påpeka är också att problematiken inte inverkar på indexvärdena, eftersom felet antas göras i samma utsträckning för båda materialen i jämförelserna. Endast frekvensvärdena för de olika strukturerna påverkas, och sådana värden redovisas inte i avhandlingen. Däremot borde värdena för parsningsdjup överskattas något, eftersom attribut befinner sig en nod längre ner i trädstrukturen än adverbial.

9.2.3.2 Om parsningsdjup som komplexitetsmått

Stadieövergångarna i ämnesfacken och i ämnena tycks kunna beskrivas på liknande vis genom beräkning av parsningsdjup och meningslängd, i alla fall om man ser till medelvärdena. Också i förhållande till de akademiska texterna ligger lärobokstexterna på samma avstånd oavsett vilket syntaktiskt mått som används: meningslängd eller parsningsdjup (eller nominalfrastäthet eller nominalfraslängd).

En tydlig skillnad syns dock i jämförelsen mellan lärobokstexterna och de berättande texterna. När det gäller meningslängd ligger lärobokstexterna mycket närmare värdet för de berättande texterna, vilket gör att matematiktexterna hamnar under denna nivå på högstadiet (se 6.3.1). I fråga om parsningsdjup förhåller sig lärobokstexterna på större avstånd från de berättande texterna, och matematik ligger över gränsen för de berättande texterna redan på högstadiet.

Tydligt föreligger en skillnad i korrelationen mellan meningslängdsmåttet och parsningsdjupsmåttet för å ena sidan lärobokstexterna och de akademiska texterna, å andra sidan de berättande texterna.

Resultatet av en korrelationsanalys mellan meningslängd och parsningsdjup i de naturvetenskapliga läroböckerna visas i tabell 9.18. Den genomsnittliga korrelationsstyrkan mellan dessa variabler 0,88, vilket betyder att en förändring i meningslängd påverkar parsningsdjupet med en faktor $0,88^2 = 0,78$. I de

akademiska texterna korrelerar meningslängd och parsningsdjup med $0,87^2 = 0,76$, alltså i likartad utsträckning som i lärobokstexterna. Samma värde för de berättande texterna är emellertid hela $0,92^2 = 0,84$.

Biologi	Högstadiet	$r_s = 0,88$
	Gymnasiet	$r_s = 0,87$
Fysik	Högstadiet	$r_s = 0,89$
	Gymnasiet	$r_s = 0,90$
Kemi	Högstadiet	$r_s = 0,88$
	Gymnasiet	$r_s = 0,88$

Tabell 9.18: Spearmans rangkorrelation mellan meningslängd och parsningsdjup i lärobokstexter i naturvetenskapliga ämnen

9.3 Typiska meningar i naturvetenskap: sammanfattning och diskussion

9.3.1 Förhållandet till meningar i matematik och samhällsvetenskap

Studien visar att medelvärdet för parsningsdjupet i texterna stiger i stadiövergången till gymnasiet för samtliga ämnesfack och ämnen. Det framgår också att lärobokstexterna befinner sig på en nivå mellan de berättande och de akademiska texterna, om än närmare de berättande texterna, med undantag för matematik på högstadiet, som ligger i nivå med de berättande texterna.

Bland ämnesfacken sker ökningen av genomsnittligt parsningsdjup kraftigt i naturvetenskap, vilket på ämnesnivå beror på att biologi har en relativt brant utvecklingskurva. På meningsnivå kan ökningen till stor del beskrivas med att andelen meningar med djup på 4 och 5 blir färre på gymnasiet. Dessutom innehåller gymnasietexterna fler instanser av meningar med väsentligt djupare syntaktiska analyser, som avviker från de typiska värdena.

Matematiktexterna får lägst värden av såväl ämnesfacken som ämnena, och det högre genomsnittliga djupet i matematiktexterna för gymnasiet visar sig uteslutande kunna härledas till extremvärden.

I jämförelsen mellan naturvetenskap och samhällsvetenskap ser fördelningen av meningarnas djup mycket lika ut på högstadiet, men medelvärdet är något högre i samhällsvetenskap. På gymnasiet skiljer sig dock fördelningen av meningsdjup mellan ämnesfacken; meningarna i naturvetenskap placerar sig inom ett bredare spann, som innefattar fler både grundare och djupare syntaktiska meningsanalyser än i samhällsvetenskap.

Det tydligaste resultatet sett till meningsstrukturer är det att naturvetenskapliga och samhällsvetenskapliga texter liknar varandra till stor del. På gymnasiet kan inga ämnesfackligt utmärkande mönster över huvud taget identifieras, och i texterna för högstadiet märks bara ett speciellt mönster som är vanligare i naturvetenskap. Den struktur som pekats ut innehåller ett prepositionsfrasbestämt subjekt i fundamentposition, som i meningen *Glödtråden i en vanlig glödlampa har stor resistans*.

I de samhällsvetenskapliga texterna för högstadiet är framför allt komplexa adverbial vanligare. En mening som *I slutet av 1700-talet började många protestera mot slavhandeln*, med ett utbyggt tidsadverbial i fundamentet, är konstruerad enligt ett typiskt samhällsvetenskapligt mönster.

I förhållande till matematik utmärker sig meningarna i naturvetenskap med att komplexiteten oftare kan lokaliseras till nominala subjekt eller subjektspredikativ. Matematiklistorna innehåller, å andra sidan, betydligt fler poster, men framför allt fler vänsterställda mönster, vilket indikerar att matematiktexterna är mer svårlästa än de naturvetenskapliga texterna. Tidigare traditionella indikatorer på läsbarhet har dock i de flesta fall pekats i motsatt riktning (jfr resultaten i kapitel 6).

Ungefär hälften av posterna på matematiklistorna pekar ut utbyggda nominalfraser. Den allra mest typiska strukturen, med ett högerställt prepositionsfrasbestämt direkt objekt, återfinns överst både på högstadie- och gymnasie-listan. Meningar med denna typiska struktur fungerar ofta som uppmaningar och inleds med en imperativ, som i meningen *Bestäm summan av de fyra talen*.

Över huvud taget kännetecknas texterna i naturvetenskap och matematik mer av utbyggda nominalfraser än texterna i samhällsvetenskap.

9.3.2 Stadieövergången mellan högstadiet och gymnasiet

För de enskilda naturvetenskapliga ämnena visar parsningsdjupsundersökningen att fysik, kemi och historia följer varandra tätt över stadierna, på en relativt låg komplexitetsnivå. På liknande sätt följer geografi och samhällskunskap varandra över stadierna, men på en något högre nivå. Den mest säregna utvecklingskurvan står biologi för, som är det ämne i vilket meningarnas komplexitet mest dramatiskt ökar i övergången till gymnasiet.

Ser man till meningsstrukturerna i biologi märks att komplexiteten i gymnasietexterna ofta är högre p.g.a. koordination på huvudsatsnivå, som i meningen *Antalet receptorer hos en cell är inte konstant utan kan ändras över tiden*. I fysiktexterna framkommer en liknande tendens; i övergången till gymnasiet ökar komplexiteten i och med det att maktrosyntagmatiska mönster blir allt vanligare. Meningen *Skissa ett v-t-diagram för de kommande 5 sekunderna*

och besvara följande frågor visar prov på en sådan konstruktion på gymnasiet. Högstadietexterna i fysik innehåller generellt sett kortare meningar, i vilka komplexiteten främst knyts till nominalfraser som utgör subjekt eller subjektspredikativ.

Stadieövergången i kemi kan i stora drag beskrivas genom att objektsadverbial (i verbfraser) byggs ut och blir djupare. En typisk högstadiemening är: *Växterna använder sedan kväveatomer för att bygga upp proteiner* och en typisk gymnasiemening är: *Kedjorna hålls samman med vätebindningar mellan nukleotidernas kvävebaser*.

Utöver jämförelserna med akademiska och berättande texter är resultaten för lärobokstexterna i denna undersökning direkt jämförbara med de olika värden som Heimann Mühlenbock (2013: 110) anger. Dessa inkluderar genomsnittligt parsningsdjup för s.k. informationstext (från SUC), och nyhetstext (från SUC och GP). Värdena för dessa texttyper ligger på 5,60 respektive 5,92. Läroböckerna i naturvetenskap för högstadiet, som i genomsnitt mäter 5,36, placerar sig därmed något under nivån för informationstext, medan läroböckerna för gymnasiet, med ett genomsnittsdjup på 6,09, i stället hamnar något över nivån för nyhetstext.

9.3.3 Förhållandet till meningar i akademiska texter

I förhållande till meningarna i de naturvetenskapliga lärobokstexterna är meningarna i de akademiska texterna förhållandevis svårlästa. Detta beror främst på att många meningar inleds med långa, utbyggda subjekt, som meningen *Avsaknaden av vattentäta skott mellan olika aspekter medför vissa svårigheter med att kategorisera indikatorerna*. Dessutom innehåller många meningar i de akademiska texterna bisatsbestämda subjekt eller direkta objekt, vilket är relativt ovanligt i lärobokstexterna.

9.3.4 Förhållandet till meningar i berättande texter

I jämförelsen mellan naturvetenskapliga lärobokstexter och berättande texter är det tydligaste resultatet att komplexiteten i lärobokstexterna främst orsakas av komplexa nominalfraser. Nominalfraserna förekommer ofta som enskilda satsled, men ingår också ofta i långa attribut och adverbial.

Komplexiteten i de berättande texterna är inte tillnärmelsevis lika koncentrerad till nominalfraser, utan byggs i stället främst upp av samordning på huvudsatsnivå (jfr ”komplexa meningar” i SAG, del 4:832ff.). Detta typiska syntaktiska drag kan tydligt relateras till A-texters syfte att redogöra för händelseförlopp (jfr avsnitt 2.1), som i meningen *Han måste böja huvudet och sucka över sin egen dumhet.*

10

SAMMANFATTNING OCH DISKUSSION

En viktig uppgift för skolan är att stötta elevers språkutveckling och därigenom deras sociala utveckling. En del i det arbetet är att lära ut hur olika texttyper byggs upp. Min studie tar avstamp i ett statiskt betraktelsesätt av genrer, där arbete inom en viss genre innebär att producera och konsumera texter med mer eller mindre konventionaliserad form eller struktur. Enligt detta synsätt är alltså genrens form synlig i materialet, texterna.

När målet är att beskriva systematisk variation i texter är empiriska observationsanalyser av nytta. Även om olika kritiska synsätt ligger mer i tiden har renodlat beskrivande data fortfarande många förtjänster i egenskap av grundforskning. I bästa fall är deskriptiv data tidlös, eftersom inga tolkningar som riskerar att bli omoderna tillförs analysen. Observationerna kan därför i ett långsiktigt perspektiv ligga till grund för mer subjektiva tolkningar av orsaksförhållanden och till förklaringar.

En förutsättning för att kunna uttala sig om de naturvetenskapliga ämnesspråkens egenheter är att först ha kännedom om hur dessa ämnesspråk är uppbyggda. Just om språket i naturvetenskap har många forskare framfört olika åsikter, men deras uppgifter bygger sällan på faktiska resultat. Jag anser att ett mer naturvetenskapligt förhållningssätt, där fakta bygger på observationer, skulle vara att föredra. Här finns en stor potential i att ta hjälp av språkteknologiska verktyg vid sammanställning av information om olika texttyper.

Det övergripande syftet med avhandlingen är att redovisa resultatet av en undersökning av de naturvetenskapliga ämnesspråken, så som de förekommer i svenska läroböcker för högstadiet och gymnasiet. Fokus ligger på språket som system, och på läsbarhet. Delstudierna omfattar traditionella textanalysmått, som t.ex. meningslängd, ordlängd, LIX och OVIX, typiskt ordförråd, typiska nominalfrasstrukturer och typiska meningsstrukturer. I samtliga analyser sätts det naturvetenskapliga läroboksspråket i relation till språket i läroböcker i matematik och samhällsvetenskapliga ämnen, samt till språket i berättande och akademiska texter. I detta avslutande kapitel sammanfattar jag de viktigaste resultaten av mina undersökningar.

Ett ytterligare mål med avhandlingsarbetet är att utveckla och demonstrera en metodologi för att med språkteknologiska hjälpmedel karakterisera språk i text. I detta kapitel utvärderar jag även hur väl denna metodologi fungerar för mina syften.

10.1 Språkliga drag i naturvetenskapliga ämnesspråk

De lärobokstexter som ingår i undersökningarna delas upp utifrån två grundläggande dimensioner – ämne och stadium. Dessutom grupperas läroböcker i ämnena biologi, fysik och kemi under ämnesfacksbeteckningen ”naturvetenskap” och läroböcker i geografi, historia och samhällskunskap under ämnesfacksbeteckningen ”samhällsvetenskap”. Genom att betrakta läroböcker i ett visst ämnesfack eller ett visst ämne som olika register, blir det möjligt att jämföra dels olika ämnesfackliga språk, dels olika ämnesspråk med varandra.

I delstudierna ställs de naturvetenskapliga lärobokstexterna mot olika delmaterial, enligt en analysgång som närmare beskrivs i 5.4. Jag har dock valt att här sammanfatta resultaten i den ordning som jämförelserna introduceras i anslutning till syftesformuleringen i 1.2. Enligt den strukturen jämförs ämnesspråk i naturvetenskapliga ämnen, på högstadie- och gymnasienivå, med (1) varandra, (2) ämnesspråk i matematik på motsvarande stadier, (3) ämnesspråk i samhällsvetenskapliga ämnen på motsvarande stadier, (4) akademiskt språk och (5) språk i berättande texter.

10.1.1 Stadieövergången mellan högstadiet och gymnasiet

De typiska orden i de naturvetenskapliga ämnena är längre på gymnasiet än på högstadiet. Dessutom innehåller det typiska ordförrådet i samtliga naturvetenskapliga ämnen på gymnasiet en högre andel adjektiv och particip än på högstadiet, vilket bidrar till de ökade värdena för NQ och nominalfraslängd.

De typiska nominalfraserna i de naturvetenskapliga ämnesspråken på högstadiet är förvånansvärt enkla. De utgörs av enkla pronomen eller substantiv, som ibland bestäms av en efterställd bisats. I ämnesspråken på gymnasiet tillkommer fler typiska nominalfrasstrukturer, med prepositionsattribut och framförställda bestämningar, som t.ex. *en mycket viktig uppgift för kärnkraftsindustrin*, som kommer från en fysikbok. I stadiövergången går nominalfrasutbyggnaden således i enda riktning, mot en påtagligt rikare och mer komplex struktur. Av allt att döma så långt verkar alltså de naturvetenskapliga ämnesspråken på gymnasiet vara tydligt utvecklade ur ämnesspråken på högstadiet.

Sammantaget är de flesta former av utbyggda nominalfraser mycket vanligare på gymnasiet än på högstadiet. Resultaten för gymnasietexterna antyder också att framförställda attribut oftare är definitiva än adjektiviska, och efterställda attribut oftare har formen av prepositionsfraser än bisatser. Ett exempel på en typisk nominalfras ur en biologibok med både definit attribut och prepositionsfrasattribut är *samma typer av läkemedel*.

Undersökningarna visar att de naturvetenskapliga ämnesspråken genomgår en kraftigare utveckling mellan stadierna än ämnesspråken i övriga ämnesfack. De naturvetenskapliga ämnesspråken tycks alltså bli särskilt svåra på gymnasiet i förhållande till på högstadiet.

I flera avseenden skiljer sig de naturvetenskapliga ämnena åt vad gäller lexikala och syntaktiska skillnader mellan stadierna. Särskilt ämnesspråken i biologi och kemi spelar viktiga roller för att språkutvecklingen i naturvetenskap sker i snabbare takt än i övriga ämnesfack. Biologitexterna blir syntaktiskt mycket mer komplexa på gymnasiet, vilket framgår av kraftigt ökade värden för meningslängd, nominalfrästäthet och parsningsdjup. Kemitexterna blir i stället lexikalt mycket mer komplexa på gymnasiet, vilket framgår av kraftigt ökade värden för NQ och ordlängd.

Vidare verkar högstadietexterna i biologi innehålla ett mer tydligt läsartilltal, med första och andra personens pronomen, än gymnasietexterna. I både biologi och fysik innebär stadieövergången i många fall att vardagliga ordval ersätts av en mer teknisk vokabulär, och i fysik och kemi visar ordlistorna på gymnasiet en tydlig orientering mot matematiska beräkningar.

Ämnesspråken i biologi och fysik utmärks på högstadiet av att en stor del av den syntaktiska komplexiteten knyts till nominalfraser. På gymnasiet är i stället koordination på huvudsatsnivå utmärkande. Komplexitetsökningen i kemitexterna kan i stora drag beskrivas genom att objektsadverbial (i verbfraser) byggs ut och blir djupare.

10.1.2 Förhållandet till ämnesspråk i matematik

I en av de mer ingående fokusstudierna (7.2.2.1) konstateras att många av de typiska naturvetenskapliga orden, som t.ex. *cell*, *molekyl* och *elektron*, är ämnesrelaterade termer som är mycket främmande i vardagsspråket. Detsamma gäller för de mest typiska orden i matematik, t.ex. *area*, *avrunda* och *ekvation*.

I jämförelse med matematiska texter är nominalfraserna i naturvetenskapliga texter vanligare, längre och mer utbyggda, särskilt ofta med adjektiv- eller bisatsattribut. Däremot är det typiska ordförrådet i naturvetenskap både mindre omfattande och består av fler korta ord. De naturvetenskapliga texterna inne-

håller också färre typiska meningsstrukturer och relativt få av dessa är vänsterställda. Trots att i stort sett alla de kvantitativa mätningarna tyder på att matematikspråket är förhållandevis enkelt pekar sammanlagt dessa resultat mot att matematikspråket också har sina svårigheter.

I ordförrådsstudien (se 7.2.1.1) visas att typiska matematikord kommer från två olika domäner: den inomvetenskapliga, ämnesspråkligt matematiska, och den vardagliga. Häri tror jag en stor del av förklaringen till att matematiktexterna uppvisar motsägelsefulla resultat ligger. Det typiska matematiska ämnesspråket är förhållandevis mycket komplext, främst lexikalt sett. Men då matematiktexter också till stor del innehåller vardagsspråk, sänks värdena avsevärt i flera undersökningar. Dessutom innehåller matematiktexterna mycket formler och symboler, som i sig kan utgöra svårigheter, men som inte ingår i min undersökning av ämnesspråk. Konsekvenser av rensningen av texterna kommenteras vidare i avsnitt 10.2.1.

10.1.3 Förhållandet till ämnesspråk i samhällsvetenskapliga ämnen

I fokusstudien av typiska ämnesfackliga ords vardagsspråklighet (7.2.2.1) visas att de mest typiska samhällsvetenskapliga orden är förhållandevis vanliga i vardagsspråket, i jämförelse med de mest typiska orden i naturvetenskap och matematik. Eftersom samhällsvetenskapen rör människors levnadsvillor, så som de upplevs, kan man bli bekant med de typiska ämnesspråkliga termerna genom att läsa dagstidningar, se på nyheterna, följa debatter och till och med genom att läsa romaner. Naturvetenskapen och matematiken handlar om ett alternativt, formaliserat synsätt på vår omvärld, som måste läras in på ett annat vis än att bara delta i allmänna sociala sammanhang.

Det typiska naturvetenskapliga ordförrådet utökas dessutom mycket mer i stadiövergången till gymnasiet än den samhällsvetenskapliga vokabulären. Då läroböckerna i naturvetenskap också innehåller förhållandevis ovanliga ord kan man anta att de lexikalt sett kräver mer av eleverna. Men de borde också kräva mer uppmärksamhet av eleverna, eftersom ämnesinnehållet knappt går att erövra på annat håll. Slutsatsen av detta är att de naturvetenskapliga ämnena (och matematik) lexikalt sett har ett mer utpräglat ämnesspråk, som kräver en mer specifikt språkinriktad undervisning.

Nominalfraserna i samhällsvetenskap är både längre och mer utbyggda, med definitiva, adjektiviska och prepositionella attribut. Bisatsattribut är dock också i jämförelse med samhällsvetenskap mer utmärkande för de naturvetenskapliga nominalfraserna.

Sett till typiska meningsstrukturer liknar naturvetenskapliga och samhällsvetenskapliga texter till stor del varandra. På gymnasiet kan inga ämnesfackligt utmärkande mönster över huvud taget identifieras, och i texterna för högstadiet märks bara ett speciellt mönster som är vanligare i naturvetenskap. Den struktur som pekas ut innehåller ett prepositionsfrasbestämt subjekt i fundamentposition, som i meningen *Glödtråden i en vanlig glödlampa har stor resistans*.

I de samhällsvetenskapliga texterna för högstadiet är framför allt komplexa adverbial vanligare. En mening som *I slutet av 1700-talet började många protestera mot slavhandeln*, med ett utbyggt tidsadverbial i fundamentet, är konstruerad enligt ett typiskt samhällsvetenskapligt mönster. Sammanfattningsvis kan den typiska meningskomplexiteten i naturvetenskap (och i matematik) främst härledas till utbyggda nominalfraser. För samhällsvetenskap återfinns den mest typiska komplexiteten i olika slags adverbial.

I jämförelserna mellan de naturvetenskapliga ämnena och de samhällsvetenskapliga finns emellertid en mycket tydlig skillnad. Även om de ämnesspecifika kurvorna alltid visar på en språkutveckling i stadiövergången, syns ett återkommande mönster i det att värdena för de naturvetenskapliga ämnena på högstadiet ligger mycket närmare varandra än värdena för gymnasiet. Dessa tendenser märks kanske framför allt i undersökningarna av meningslängd, ordlängd, LIX, nominalfrastäthet, nominalfraslängd och parsningsdjup.

Det naturvetenskapliga ämnesfacket är alltså mer språkligt homogent på högstadiet. På gymnasiet följer de olika ämnesspråken skilda utvecklingslinjer, vilket resulterar i att de får en tydligare individuell karakteristik. De olika samhällsvetenskapliga ämnena utvecklas mer i samma takt mellan högstadiet och gymnasiet, vilket gör att samhällsvetenskap som ämnesfack tycks mer sammanhållet än naturvetenskap på gymnasiet. Slutsatsen att de naturvetenskapliga ämnesspråkens gemensamma språkliga kärna splittras i stadiövergången till gymnasiet stärks av det faktum att den allra mest karakteristiska vokabulären är mer knuten till stadierna än i matematik och samhällsvetenskap.

10.1.4 Förhållandet till språket i akademiska texter

I delstudien av nominalfraser framkommer att såväl nominalfrastätheten som nominalfraslängden är högre i de akademiska texterna än i gymnasietexterna. Båda dessa resultat pekar mot en ökad nominalitetsgrad i akademiska texter. Indexlistorna för gymnasiet och för de akademiska texterna bekräftar också att attributiva prepositionsfraser ökar i stadiövergångarna, ytterligare en lexikogrammatisk variabel som bidrar till en ökad nominalitet.

NQ-undersökningen pekar däremot på att lärobokstexterna för gymnasiet i alla ämnesfack är mer informationspackade än akademiska texter. Fler pronominella nominalfraser i akademiska texter än i gymnasietexter talar också för att gymnasietexter är allra mest nominala. Nominalfraserna i de akademiska texterna är visserligen generellt sett längre än i gymnasietexterna, men indexlistorna visar att de inte är mer utbyggda. Längdökningen måste alltså bero på att en högre andel nominalfraser byggs ut enligt de mönster som etableras i stadiövergången mellan högstadiet och gymnasiet. Det faktum att de akademiska texterna också har fler nominalfraser per mening än gymnasietexterna hänger förmodligen samman med att meningarna i akademiska texter generellt sett är längre.

I förhållande till de naturvetenskapliga lärobokstexterna är meningarna i de akademiska texterna förhållandevis svårlästa. Detta beror främst på att många meningar inleds med långa, utbyggda subjekt, som meningen *Avsaknaden av vattentäta skott mellan olika aspekter medför vissa svårigheter med att kategorisera indikatorerna*.

Betraktar man texterna från ett elevperspektiv, i den ordning eleverna möter texterna, förefaller det naturvetenskapliga gymnasiespråket, som sagt, vara en vidareutveckling av högstadiespråket. På liknande vis tycks det akademiska språket till största del vara en vidareutveckling av skolspråket än en texttyp som står på en egen grund. Detta märks både genom att det, i det närmaste, inte förekommer några starkt utmärkande nominalfrastyper i naturvetenskapliga lärobokstexter i jämförelse med akademiska texter, och genom att de utbyggda NN-nominalfraserna som kännetecknar de akademiska texterna redan förekommer flitigt på gymnasiet. En del av förklaringen till utvecklingen i texterna kan dock vara att produktionsförhållanden av texterna är omvända. Det är troligt att ämnesspecialister förenklar akademiska texter för gymnasiet, och sedan förenklar gymnasietexter för högstadiet.

I jämförelserna mellan naturvetenskapliga och akademiska texter är det viktigt att komma ihåg att texterna skiljer sig åt både ämnesmässigt och stadiemässigt. Den ämnesmässiga diskrepansen mellan de naturvetenskapliga texterna och de akademiska texterna märks nog allra mest i delstudien om typiskt ordförråd (jfr resultaten av fokusstudie 7.2.2.2). Ämnesinnehållets inverkan på resultaten är åtminstone inte lika uppenbar i studier av t.ex. nominalfrasstruktur eller syntaktisk komplexitet.

Topikskillnader mellan texter verkar också slå igenom i fokusstudien av pronomen (8.4.4.1). Resultaten härifrån visar att andra personens personliga pronomen utgör en betydligt större andel i lärobokstexterna. Av allt att döma har alltså läroböckerna ett mer läsarinkluderande tilltal än de akademiska texterna. Å andra sidan innehåller mina akademiska texter en högre andel första och tredje personens personliga pronomen, vilket dels tyder på att författare till

akademiska texter är synligare i texterna, dels att akademiska texter innehåller förhållandevis mycket referenser till tidigare forskares verk och tankar. Båda dessa resultat ligger i linje med tidigare resultat för humanistiska akademiska texter (se avsnitt 2.2.2).

Eftersom den akademiska korpusen saknar naturvetenskapliga texter kan jag inte uttala mig om huruvida resultaten beror på stadieskillnaden eller ämnesskillnaden mellan texterna. Det är därför vanskligt att dra några generella slutsatser om ämnesspråklig progression mellan skolspråket och det akademiska språket.

10.1.5 Förhållandet till språket i berättande texter

I jämförelse med berättande texter utmärker sig lärobokstexterna i naturvetenskap på flera sätt. De allra flesta mätningar tyder på att lärobokstexterna är svårare redan på högstadiet och innehåller en mer utvecklad repertoar av uttrycksmedel (jfr utvecklingsgången för referenshantering i 8.2).

Ordklassmässigt är substantiv och particip mycket vanligare i naturvetenskapstexterna. Dessa innehåller också förhållandevis många ovanliga, långa, ämnesrelaterade substantiv och verb. Dessutom är verben typiskt statiska, som t.ex. *innehålla* och *bestå (av)*, med icke-agentiva subjekt. I de naturvetenskapliga texterna refereras oftare till substantiviska företeelser, med hjälp av utbyggda nominalfraser. Just utbyggda nominalfraser är utmärkande för den typiska strukturella komplexiteten i lärobokstexterna jämfört med de berättande texterna.

I de berättande texterna refereras mycket oftare till namngivna individer, antingen med egennamn eller med pronomen. Ordlistan över de berättande texterna domineras av personliga och possessiva pronomen, som *hon/PN* och *min/PS*, och vanliga interjektioner, som *ja* och *nej*. De typiska substantiven i berättande texter refererar till människors kroppsdelar, som t.ex. *ansikte* och *huvud*, eller till mänskliga uttrycksmedel, som *blick* och *röst*, medan de typiska verben, som *säga*, *fråga* och *undra*, hör till ett agentivt subjekt. Slutligen gäller den typiska komplexiteten i berättande texter koordination på huvudsatsnivå.

I mätningarna av ordvariation (OVIX och LVIX) får de berättande texterna förhållandevis högre värden än i övriga mätningar. LVIX-värdet för berättande texter är betydligt högre än värdena för lärobokstexterna, men även något högre än värdet för de akademiska texterna. Det visar sig vidare att det främst är vuxenromanerna som bidrar till dessa höga värden.

Av alla ämnesfack som ingår i undersökningarna verkar naturvetenskap vara det som har minst gemensamt med berättande texter. Naturvetenskapliga ämnesspråkliga ord är mycket avskilda i förhållande till vokabulären i A-texter,

särskilt de typiska orden på gymnasiet. De typiska samhällsvetenskapliga orden överlappar mycket mer ordförrådet i A-texter, eftersom den samhällsvetenskapliga ämneskunskapen till stor del berör vardagen. På liknande vis är matematiktexter fulla av vardagsspråk, i syfte att förankra ämneskunskapen i vardagliga situationer. Naturvetenskapliga texter tycks i mångt och mycket sakna beskrivningen av tillämpad och konkretiserad ämneskunskap.

10.1.6 Slutsatser

Under hela 1900-talet har skriftspråket successivt förenklats och skrivna offentliga texter präglas i dag av ett klarspråksideal. I *Språklagen* formuleras budskapet på följande vis: ”Språket i offentlig verksamhet ska vara vårdat, enkelt och begripligt” (Kulturdepartementet 2009: §11). För att myndigheter ska kunna uppfylla dessa krav, och aktivt kunna arbeta med att förenkla texter, finns flera handböcker med råd och riktlinjer (se t.ex. Sundin 2007; Språkrådet 2014). Ett sådant råd lyder: ”Ligger [LIX] under 30 kan du känna dig säker. Då är nämligen texten mycket lättläst. Ligger du högre bör du bearbeta texten ytterligare: korta av för långa meningar och välj enklare och kortare ord” (Sundin 2007: 256).

Naturligtvis har också läroboksförfattare ett ansvar för att uppfylla dessa krav på lättillgänglig information, som ett led i demokratiseringen av skolan. Forskare menar också att olika läsbarhetsmått påverkar utformningen av moderna läroböcker (jfr 2.4.2.3).

Mina resultat, i fråga om meningslängd, ordlängd, LIX och OVIX, visar också mycket riktigt att läroböckerna i dag är något enklare ur läsbarhetsvinkel än på 1970-talet. Men vad betyder det egentligen? Som vi har sett kännetecknas A-texter av en hög ordvariation, och att OVIX är lägre beror kanske på att berättande inslag i läroböckerna har minskat under årens lopp (jfr Sandqvist 1995). Berättande text sägs samtidigt inte utgöra något problem för elever.

Kopplingen mellan läsbarhetsmått och förståelse är alltså sammansatt. Också antagandet att korta satser skulle vara mer lättlästa kan ifrågasättas, eftersom studier har visat att korta texter och satser ofta utgör hinder för läsförståelsen och är symptomatiska för bristande logisk uppbyggnad av innehållet (jfr 6.1).

Frågan är alltså hur man ska tolka dessa förändringar över tid. Skolresultaten går ner och läroböckerna är språkligt enklare till sin utformning. Är det möjligen så att dagens lärobokstexter har bristande textuell koherens p.g.a. en strävan att skriva lättlästa texter? Eller kan det vara så att språket i läroböckerna inte är tillräckligt komplext för att eleverna ska kunna tillägna sig avancerade ämneskunskaper i skolan?

Samtliga resultat visar visserligen på en språkutveckling som sker i stadiövergången mellan högstadiet och gymnasiet, både inom de olika ämnesfacken och inom de olika separata ämnena. På gymnasiet är meningarna och orden längre, och ordförrådet mer varierat. Vidare ökar nominalfrastätheten och nominalfraserna bli längre, vilket pekar mot en ökad nominalitetsgrad. Dessutom blir de typiska meningsstrukturerna mer komplexa. Men av mina resultat framgår också att glappet mellan gymnasietexterna och de akademiska texterna är förhållandevis stort, i jämförelse med övergången mellan högstadiet och gymnasiet. Detta gäller också för de samhällsvetenskapliga texterna i förhållande till de akademiska texterna (mellan vilka ämnesinnehållet är mer överensstämmande). Är möjligen gymnasieböckerna för enkla för att förbereda eleverna för högskolestudier?

Dessa frågor kan inte min forskning besvara, bara väcka. Ser man däremot förbi resultaten från läsbarhetsdelstudien bekräftas till viss del hypotesen att de naturvetenskapliga ämnesspråken är speciella i många avseenden. I förhållande till orden i samhällsvetenskapliga texter är de typiska naturvetenskapliga orden mycket mer sällsynta i vardagsspråket. Dessutom sker en häftigare utveckling av de naturvetenskapliga ämnesspråken i stadiövergången till gymnasiet, och på gymnasiet uppvisar de olika naturvetenskapliga ämnesspråken en tydligare individuell karakteristik.

I förhållande till berättande texter är de naturvetenskapliga lärobokstexterna mycket avvikande. Redan på högstadiet innehåller de en mer utvecklad repertoar av uttrycksmedel, med förhållandevis många ovanliga, långa, ämnesrelaterade substantiv och verb. De naturvetenskapliga lärobokstexterna utmärks också av att mycket av den strukturella komplexiteten knyts till utbyggda nominalfraser, medan den typiska komplexiteten i berättande texter gäller kordination på huvudsatsnivå. Av alla ämnesfack som ingår i undersökningarna är naturvetenskap det som har minst gemensamt med berättande texter.

De olika avgränsningarna som har gjorts under arbetets gång innebär att jag endast kan beskriva en del av ämnesspråken som helhet. Icke desto mindre bidrar min forskning till utforskningen av svenskt ämnesspråk i stort. Sammantaget antyder mina resultat att det kan finnas ett samband mellan de naturvetenskapliga ämnesspråken och svårigheter för elever att tillgodogöra sig innehållet i de naturvetenskapliga läroböckerna.

10.2 Metodologiska reflektioner

Den största metodologiska fördelen med att använda språkvetenskaplig databehandling för att besvara språkvetenskapliga frågor är att betydligt större material kan analyseras än vid manuell textanalys. Materialunderlaget i Westmans läroboksstudie från 1974, vars resultat alltså ligger till grund för referensvärden i stilistiska handböcker och läroboksforskning, uppgår exempelvis till 23 000 ord (jfr 2.4.2.1). Mina resultat baseras på en ungefär 200 gånger så stor lärobokskorpus på drygt 5,2 miljoner token (jfr 3.1).

Stilistisk data som är framtagen automatiskt kan ibland dokumentera vad människor redan vet, men, i alla fall när det gäller stora textmängder, ibland även påvisa typiska drag som är osynliga för ögat. Tekniken kan alltså föra stilistiken framåt, samtidigt som den i vissa avseenden begränsar den, då de variabler som undersöks måste kunna identifieras automatiskt. Till viss del riskerar man därför att leta efter intressanta resultat på redan känd mark, när man kanske egentligen skulle vilja analysera texterna mer förutsättningslöst.

I detta stycke reflekterar jag över några olika metodologiska aspekter i mina undersökningar. Först berörs den rensning som genomförts för att isolera språket i texterna och sortera bort felsegmenterade meningar, därefter indexprincipen som en metod för att identifiera texttyptypiska språkliga variabler, och till sist de olika komplexitetsmått som ingår i undersökningarna.

10.2.1 Rensningen

En viktig avgränsning som jag har gjort, i förhållande till många andra studier av texter, är att fokusera språket i texterna. Jag har därför behövt rensa texterna på meningar som i hög grad innehåller matematiska formler eller andra formella symboler (se 4.2).

Denna rensning kan dock ses som ett resultat i sig vad gäller utformningen av lärobokstexterna (jfr 4.2.1). Utvinningen av det naturliga språket innebär en avsevärt större decimering av hela textmängden för matematik än för naturvetenskap, vilket implicerar att formler och symboler ges större utrymme och därmed också har större betydelse i just matematik. Men rensningen har inte bara betydelse för matematikämnet. Enligt Lemke (2000) är det inte möjligt att koda alla idéer med endast språkliga medel inom naturvetenskapliga ämnen. Också kemi- och fysikspråket innehåller formler/symboler, vilket också bekräftas i rensningsprocessen. Däremot innehåller inte biologispråket lika mycket formler/symboler (jfr även Ribeck 2014).

Beträffande ämnesspråken i naturvetenskap visar också rensningen på några tendenser i stadiövergången. Andelen formler/symboler i fysik och kemi är avsevärt högre på gymnasiet än på högstadiet. Också gymnasietexterna i biologi innehåller mer formler/symboler än högstadietexterna. I matematik ligger andelen på en relativt konstant, mycket hög nivå. I de samhällsvetenskapliga ämnena märks ingen särskild trend gällande förhållandet mellan formler/symboler och språk på de olika stadierna.

Det största problemet jag ser med rensningen är att några fullt naturliga långa meningar har rensats bort. När sedan textuell komplexitet, som meningslängd, LIX, nominalfrastäthet och parsningsdjup beräknas, inverkar naturligtvis det på resultatet. Men alla texter har åtminstone rensats på samma sätt, och jag är övertygad om att resultaten hade varit mer osäkra ifall inte längdkontrollen hade ingått i rensningsreglerna. Just den regeln eliminerar mycket brus som meningssegmenteraren taggat som meningar, men som egentligen är uppräknningar, förteckningar eller liknande. Också kravet på ett finit verb i meningen är inte renodlat riktat mot för mycket formler/symboler i meningarna, utan tjänar också som en kontroll för att meningarna är korrekt segmenterade. Denna kontroll gör också mycket nytta.

Mina ämnesspråkliga resultat säger alltså olika mycket om de olika ämnestexterna. När det gäller matematik utgör den del jag undersöker en mindre del av de lärobokstexter som eleverna möter i undervisningen. Naturligtvis finns en risk att sådant som ovana läsare tycker är svårt tas bort i rensningen.⁵⁶ Jag vill därför vara extra försiktig att uttala mig om svårighetsgrader för olika ämnestexter. Mina språkliga variabler underskattar t.ex. förmodligen den sammanlagda komplexiteten i matematiktexter, som till stor del utgörs av att avkoda och processa formella symboler.

10.2.2 Indexprincipen

I och med denna avhandling har jag visat att indexprincipen är en metod som kan användas för att påvisa språkliga drag som är utmärkande för olika texttyper. För mina syften har metoden dock fungerat olika väl i olika delstudier. I delstudien av typiskt ordförråd bedöms utfallet vara mycket positivt. Ordlistemetoden kan definitivt användas för att ta fram underlag som i förlängningen kan utgöra stöd i skolans arbete med ordinläring (jfr t.ex. AWL i 2.2.3).

⁵⁶Flera olika representationssätt, i form av t.ex. siffror och symboler, nämns ofta som en kritisk egenskap för matematikämnet (jfr Duval 2006; Sfard 2007). Studier framhåller förekomsten av symboler i matematiktexter som den mest troliga anledningen till att man behöver en ämnesspecifik läskunnighet för att förstå matematiska texter, medan matematiska texter skrivna med naturligt språk inte verkar kräva något mer än en allmän läsförståelse (Österholm 2005).

Också typisk nominalfras- och meningsuppbyggnad kan identifieras med indexprincipen. Några jämförelser resulterar dock i mycket korta indexlistor när det kommer till dessa variabler, och för gymnasietexterna i naturvetenskap och samhällsvetenskap lyckas inte någon enda typiska parsträdsstruktur extraheras.

Eftersom inget facit finns, och inte heller kan konstrueras, är det omöjligt att med säkerhet förklara varför vissa indexlistor är kortare än andra. Man kan inte utesluta möjligheten att två olika texttyper i en jämförelse inte utmärker sig i förhållande till varandra vad gäller en viss variabel. Jag tror dock att storleken på mitt material kan ha inverkat på utfallet när indexprincipen ska känna igen flerordiga strukturer. Extraktion av längre strukturer kräver mer text. För nominalfraser är utgångsläget emellertid något bättre än för parsträdsstrukturerna. Varje mening innehåller i genomsnitt 2–3 nominalfraser med en genomsnittlig längd på 2–3 ord. Varje mening innehåller däremot (oftast) bara ett maxdjup. Dessutom är detta i genomsnitt hela 5–6 noder, och i varje steg står valet mellan fler möjliga noder.

För att vara en modern korpus är min läroboktextsamling inte särskilt stor. Detta beror både på att utbudet av moderna läroböcker har sin begränsning och på att lärobokstexter är relativt svåråtkomliga, särskilt i digitalt format. Det är t.ex. betydligt mer tacksamt att samla in material från tidningar eller bloggar, som dagligen lägger ut nya texter på nätet för alla att läsa. Å andra sidan är min samling av läroböcker betydligt mer omfattande än det materialunderlag som de flesta andra, både kvalitativa och kvantitativa, läroboksstudier grundar sig på.

10.2.3 Komplexitetsmått

I delstudierna förekommer olika mått på syntaktisk komplexitet: å ena sidan det ytliga meningslängdsmåttet, å andra sidan nominalfrastäthet, nominalfraslängd och parsningsdjup. Mätningarna visar att stadieövergångarna i de naturvetenskapliga ämnena tycks kunna beskrivas på liknande vis genom beräkning av vilket som helst av dessa mått. Också i förhållande till de akademiska texterna ligger lärobokstexterna på samma avstånd oavsett vilket syntaktiskt mått som används (jfr 8.4.4.2 och 9.2.3.2).

Skillnader mellan de olika måtten märks dock i jämförelsen mellan de naturvetenskapliga lärobokstexterna och å ena sidan matematiktexterna, å andra sidan de berättande texterna. Med nominalfraslängdsmåttet får matematiktexterna högre värden relativt övriga lärobokstexter, och hamnar inte långt under dessa. Tydligt skiljer sig komplexiteten i matematiktexterna på något vis från övriga lärobokstexters. Vidare samvarierar måtten nominalfrastäthet

och parsningsdjup i högre grad med meningslängdsmåttet för de berättande texterna än för övriga texttyper i undersökningen. Det är således nyttigt att beräkna syntaktisk komplexitet i termer av t.ex. parsningsdjup och nominalfrastäthet, eftersom korrelationen mellan sådana mått och meningslängd skiljer sig åt mellan olika texttyper, och då i synnerhet mellan A- och B-texter. Men även nominalfraslängdsmåttet verkar kunna påvisa ett slags komplexitet som inte övriga mått kan fånga. En sådan typ av komplexitet finns i matematiktexterna.

När det gäller lexikal komplexitet och informationspackning tycks det också kunna vara mödan värt att utnyttja viss lingvistisk information för att uppskatta svårighetsgraden; både NQ- och LVIX-undersökningarna ger särskilda bilder av de olika materialens egenheter, som inte framkommer med hjälp av något ytligt analysmått. En ytterligare vinst med mer lingvistiskt sofistikerade analysmetoder är förmågan att kunna påvisa vilka faktiska strukturer som kännetecknar olika texter. Med indexprincipen kan jag peka på betydligt mer detaljerade resultat än bara ”långa ord”, ”långa meningar” eller ”hög ordvariation”. För den som vill träna elever på att avkoda nya strukturer är detta ett redskap.

10.3 Framtiden

Än så länge saknar de svenska läroplanerna tydliga beskrivningar av den förväntade ämnesspråkliga progressionen inom olika ämnen och olika stadier. I ett framtida arbete med att utforma dylika mål kan mina indexlistor bidra med konkreta exempel på typiska, ämnesspråkliga konstruktioner för olika ämnen och stadier. Mina resultat kan också användas för att kontrollera i vilken utsträckning de typiska ämnesspråkliga dragen förekommer i elevtexter (jfr 2.4.2.1 och *Gymnasistsvenska* från 1977). På så vis skulle man kunna identifiera inom vilka områden elever behöver särskilt stöd i att utveckla erforderad ämnesspråklig kompetens.

Ett annat förslag är att inte stanna vid t.ex. typiska nominalfrasmonster med en viss abstrakt syntaktisk form, utan att även undersöka ifall de typiska strukturerna motsvarar återkommande språkliga formuleringar. Läsförståelse beror inte bara på språklig komplexitet, utan också på vilka lexikala uttryck man läser. Vanliga eller välkända flerordiga uttryck borde t.ex. inte vara lika svår-förståeliga varje gång de påträffas i texterna.

Olika typer av flerordingar, som t.ex. kollokationer och lexikala knippen, behöver därför uppmärksammas inom ämnesspråklig forskning. Ett annat, innovativt, sätt att identifiera lexikalt återkommande flerordiga mönster är att undersöka s.k. *hybrida n-gram*. Till skillnad från traditionella n-gram, består

hybrida dito inte uteslutande av textord, utan utgörs av enheter på varierande lingvistisk nivå, som textord, lemman eller ordklasser (Wible och Tsao 2011). Fördelen med hybrida n-gram är att man med hjälp av lingvistisk information kan höja abstraktionsnivån för de uttryck man identifierar (jfr Forsberg m.fl. 2014).

Vidare tror jag att information om särskilt texttypiska drag genererade med indexprincipen kan användas av forskare för att få en bättre inblick i sitt material. Sådana språkliga profiler skulle också kunna utnyttjas i kvalitetssäkrande syfte, t.ex. i valet mellan olika läromedel.

Avhandlingsarbetet har också resulterat i värdefulla textunderlag vad gäller skolspråk och akademiskt språk, som framtida forskare från olika fält kan utnyttja. Till att börja med finns en rad ytterligare ämnesspråkliga språkdrag som kan kartläggas i materialen, antingen med hjälp av indexprincipen eller med någon annan metod. Några exempel på angelägna variabler är verbfras- och fundamentstrukturer. Härutöver är multimodal och diskursiv/textuell komplexitet viktiga ämnesspråkliga områden som behöver beskrivas närmare.

LITTERATURFÖRTECKNING

- Allén, Sture 1970. *Nusvensk frekvensordbok baserad på tidningstext, Vol. 1, Graford*. Stockholm: Almqvist & Wiksell.
- Anward, Jan 1984. *Språkbruk och språkutveckling i skolan*. Lund: Liber.
- Anward, Jan 1989. Om form och funktion i språkutvecklingen. Carin Sandqvist och Ulf Teleman (red.), *Språkutveckling under skoltiden*, 95–115. Lund: Studentlitteratur.
- Ask, Sofia 2005. *Tillgång till framgång. Lärare och studenter om studieövergången till högre utbildning*. Växjö: Växjö universitet, Institutionen för humaniora.
- Ask, Sofia 2006. *Hållbara texter. Grunderna i formellt skrivande*. Stockholm: Liber.
- Ask, Sofia 2007. *Vägar till ett akademiskt skriftspråk*. Växjö: Växjö University Press.
- Bailey, Alison L., Frances A. Butler, Malka Borrego, Charmien LaFramenta och Christine Ong 2002. Towards the characterization of academic language in upper elementary classrooms. *Language Testing Update* 31: 45–52.
- Baker, Paul, Costas Gabrielatos, Majid Khosravini, Michał Krzyżanowski, Tony McEnery och Ruth Wodak 2008. A useful methodological synergy? Combining critical discourse analysis and corpus linguistics to examine discourses of refugees and asylum seekers in the UK press. *Discourse & Society* 19 (3): 273–306.
- Baptista, Jorge, Neuza Costa, Joaquim Guerra, Marcos Zampieri, Maria Cabral och Nuno Mamede 2010. P-AWL: academic word list for Portuguese. Thiago Alexandre Salgueiro Pardo, António Branco, Aldebaro Klautau, Renata Vieira och Vera Lúcia Strube de Lima (red.), *Computational processing of the Portuguese language, lecture notes in computer science, 6001/2010*, 120–123. Berlin, Heidelberg: Springer.
- Barzilay, Regina och Mirella Lapata 2008. Modeling local coherence: an entity-based approach. *Computational Linguistics* 34 (1): 1–34.
- Bauer, Laurie och Paul Nation 1993. Word families. *International Journal of Lexicography* 6: 253–279.

- Bell, Courtney M., Philip M. McCarthy och Danielle S. McNamara 2012. Using LIWC and Coh-Metrix to investigate gender differences in linguistic styles. Philip M. McCarthy och Chutima Boonthum-Denecke (red.), *Applied natural language processing: identification, investigation, and resolution*, 545–556. Hershey, PA: IGI Global.
- Bergstrand Grenner, Ingrid och Lena Ekblad 2011. Att förstå eller inte förstå – det är frågan. En studie om läroboksspråkets betydelse för läsförståelsen hos elever i skolår 4. Malmö högskola, Lärarutbildningen. Examensarbete, 15 hp.
- Bernstein, Basil 1971. *Class, codes and control. Theoretical studies towards a sociology of language. Vol. 1, Theoretical studies towards a sociology of language*. London: Routledge.
- Biber, Douglas 1988. *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, Douglas 1995. *Dimensions of register variations: a cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Biber, Douglas 2006. *University language. A corpus-based study of spoken and written registers*. Amsterdam/Philadelphia: John Benjamins Publishing Company.
- Biber, Douglas 2011. Corpus linguistics and the study of literature. Back to the future? *Scientific Study of Literature* 1 (1): 15–23.
- Biber, Douglas och Federica Barbieri 2007. Lexical bundles in university spoken and written registers. *English for Specific Purposes* 26: 263–286.
- Biber, Douglas och Susan Conrad 1999. Lexical bundles in conversation and academic prose. Hilde Hasselgard och Signe Oksefjell (red.), *Out of corpora: studies in honor of Stig Johansson*, 181–189. Amsterdam: Rodopi.
- Biber, Douglas och Susan Conrad 2009. *Register, genre and style*. Cambridge: Cambridge University Press.
- Biber, Douglas, Susan Conrad, Randi Reppen, Pat Byrd, Marie Helt, Victoria Clark, Viviana Cortes, Eniko Csomay och Alfredo Urzua 2004. *Representing language use in the university: analysis of the TOEFL 2000 spoken and written academic language corpus*. Princeton, NJ: Educational Testing Service.
- Biber, Douglas och Bethany Gray 2010. Challenging stereotypes about academic writing: complexity, elaboration, and explicitness. *Linguistics and Education* 9 (1): 2–20.

- Bickel, Steffen, Peter Haider och Tobias Scheffer 2005. Predicting sentences using n-gram language models. *Proceedings of the conference on human language technology and empirical methods in natural language processing*, 193–200. Stroudsburg, PA: ACL.
- Björkqvall, Anders 2010. *Den visuella texten. Multimodal analys i praktiken*. Uppsala: Hallgren och Fallgren.
- Björnsson, Carl-Hugo 1968. *Läsbarhet*. Stockholm: Bokförlaget Liber.
- Bloom, Benjamin S. och David R. Krathwohl 1956. *Taxonomy of educational objectives: the classification of educational goals, by a committee of college and university examiners. Handbook 1: cognitive domain*. New York: Longmans.
- Blåsjö, Mona 2004. *Studenters skrivande i två kunskapsbyggande miljöer*. Stockholm: Stockholm Studies in Scandinavian Philology 37, Stockholms universitet.
- Blåsjö, Mona 2009. Appropriering av fackkunskap – förutsättning för ett ”godkänt” akademiskt skrivande? Gunhild Åm Vatn, Ingvild Folkvord och Jon Smidt (red.), *Skrivning i kunnskapssamfunnet*, 143–159. Trondheim: Tapir Akademisk Forlag.
- Borin, Lars 2010. Med Zipf mot framtiden – en integrerad lexikonresurs för svensk språkteknologi. *LexicoNordica* 17: 35–54.
- Borin, Lars, Markus Forsberg och Lennart Lönngrén 2013. Saldo: a touch of yin to WordNet’s yang. *Language resources and evaluation* 47 (4): 1191–1211.
- Borin, Lars, Markus Forsberg och Johan Roxendal 2012. Korp – the corpus infrastructure of Språkbanken. *Proceedings of the 8th international conference on language resources and evaluation*. Istanbul: ELRA.
- Bratt, Bengt 1985. *Språkutveckling – vägen till kunskap*. Lund: Studentlitteratur.
- Brown, Peter F., John Cocke, Vincent J. Della Pietra, Fredrick Jelinek, John D. Lafferty, Robert L. Mercer och Paul S. Roossin 1990. A statistical approach to machine translation. *Computational Linguistics* 16 (2): 79–85.
- Burrows, John F. 2007. All the way through: testing for authorship in different frequency strata. *Literary and Linguistic Computing* 22 (1): 27–47.
- Campion, Mary E. och Warwick B. Elley 1971. *An academic vocabulary list*. Wellington: NZCER.
- Carlsson, Marie och Kerstin von Brömssen 2011. Kritisk läsning av pedagogiska texter – en introduktion. Marie Carlsson och Kerstin von Brömssen (red.), *Kritisk läsning av pedagogiska texter. Genus, etnicitet och andra kategoriseringar*, 15–46. Lund: Studentlitteratur.

- Carlund, Carina, Håkan Jansson, Sofie Johansson Kokkinakis, Julia Prentice och Judy Ribeck 2012. An academic word list for Swedish – a support for language learners in higher education. *Proceedings of the SLTC 2012 workshop on NLP for CALL, 25 October, 2012*. Lund: SLTC.
- Cavaglià, Gabriela 2002. Measuring corpus homogeneity using a range of measures for inter-document distance. *Proceedings of the third international conference on language resources and evaluation*. Las Palmas: ELRA.
- Chamot, Anna U. och Michael O'Malley 1994. *The CALLA handbook: implementing the cognitive academic language learning approach*. Reading, MA: Addison-Wesley Publishing Company.
- Chandler, Daniel 1995. *The act of writing. A media theory approach*. Aberystwyth: University of Wales.
- Chelba, Ciprian 2013. Statistical language modeling. Alexander Clark, Chris Fox och Shalom Lappin (red.), *The handbook of computational linguistics and natural language processing*, 74–104. Oxford: Wiley-Blackwell.
- Chung, Teresa Mihwa och Paul Nation 2003. Technical vocabulary in specialised texts. *Reading in a Foreign Language* 15 (2): 103–116.
- Clark, Alexander och Shalom Lappin 2013. Unsupervised learning and grammar induction. Alexander Clark, Chris Fox och Shalom Lappin (red.), *The handbook of computational linguistics and natural language processing*, 197–220. Oxford: Wiley-Blackwell.
- Cobb, Tom och Marlise Horst 2001. Reading academic English: carrying learners across the lexical threshold. John Flowerdew och Matthew Peacock (red.), *Research perspectives on English for academic purposes*, 315–329. Cambridge: Cambridge University Press.
- Cobb, Tom och Marlise Horst 2004. Is there room for an academic word list in French? Paul Bogaards och Batia Laufer (red.), *Vocabulary in a second language: selection, acquisition, and testing*, 15–38. Amsterdam: John Benjamins.
- Collins, Kevyn 2014. Computational assessment of text readability: a survey of current and future research. <http://www-personal.umich.edu/~kevynct/pubs/ITL-readability-invited-article-v10-camera.pdf>. Working draft.
- Coxhead, Averil 2000. A new academic word list. *TESOL Quarterly* 34 (2): 213–238.

- Coxhead, Averil 2002. The academic word list: a corpus-based word list for academic purposes. Bernhard Kettemann och Marko Georg (red.), *Teaching and learning by doing corpus analysis. Proceedings of the 4th international conference on teaching and language corpora, Graz 19–24 July, 2000*. Amsterdam: Rodopi.
- Coxhead, Averil 2011. The academic word list 10 years on: research and teaching implications. *TESOL Quarterly* 45 (2): 355–362.
- Coxhead, Averil, Stevens Liesje och Jenna Tinkle 2010. Why might secondary school textbooks be difficult to read? *New Zealand Studies in Applied Linguistics* 16 (2): 35–52.
- Coxhead, Averil och Paul Nation 2001. The specialized vocabulary of English for academic purposes. John Flowerdew och Matthew Peacock (red.), *Research perspectives on English for academic purposes*, 185–216. Cambridge: Cambridge University Press.
- Craig, Hugh 2004. Stylistic analysis and authorship studies. Susan Schreibman, Ray Siemens och John Unsworth (red.), *A companion to digital humanities*. Oxford: Blackwell.
- Cummins, Jim P. 1981. Age on arrival and immigrant second language learning in Canada. A reassessment. *Applied Linguistics* 2: 132–149.
- Dagens Nyheter 2013. Ingen ljusning i sikte. 4 december 2013. <http://www.dn.se/ledare/huvudledare/ingen-ljusning-i-sikte/>.
- Danielsson, Sylvia 1975. *Läroboksspråk*. Umeå: Acta Universitatis Umensis.
- Dunning, Ted 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics* 19 (1): 61–74.
- Duval, Raymond 2006. A cognitive analysis of problems of comprehension in a learning of mathematics. *Educational Studies in Mathematics* 61: 103–131.
- Dysthe, Olga, Siri Breistein, Jens Kjeldsen och Liv Ingeborg Lied 2006. Studentperspektiv på rettleiing. Olga Dysthe och Akylina Samara (red.), *Forskningsveiledning på master- og doktorgradsnivå*, 207–227. Oslo: Abstrakt Forlag.
- Dysthe, Olga, Frøydis Hertzberg och Torlaug Løkesgaard Hoel 2011. *Skriva för att lära*. Lund: Studentlitteratur.
- Edling, Agnes 2006. *Abstraction and authority in textbooks. The textual paths towards specialized language*. Uppsala: Acta Universitatis Upsaliensis.
- Ehn, Billy och Orvar Löfgren 2004. *Hur blir man klok på universitetet?* Lund: Studentlitteratur.

- Einarsson, Jan 1978. *Talad och skriven svenska. Sociolingvistiska studier*. Lund: Ekstrand.
- Einarsson, Jan 2005. Studenter i en akademisk språkvärld. Maria Lindgren (red.), *Den skrivande studenten. Idéer, erfarenheter och forskning från Testverkstaden vid Växjö universitet*, 71–85. Växjö: Växjö universitet.
- Ejerhed, Eva, Gunnel Källgren, Ola Wennstedt och Magnus Åström 1992. *The linguistic annotation system of the Stockholm-Umeå corpus project*. Umeå: Department of general linguistics, University of Umeå.
- Ekvall, Ulla 1995. Läroboken – begriplig och intressant? Siv Strömquist (red.), *Läroboksspråk. Om språk och layout i svenska läroböcker*, 47–76. Uppsala: Hallgren och Fallgren.
- Ekvall, Ulla 1997. *Formativt, figurativt, operativt i läroböcker för barn. Hur syn på kunskap och kunskapsställande påverkat naturkunskapsböckers innehåll, språk och struktur. Del 1. Utvecklingen under 1800-talet*. Lund: Institutionen för nordiska språk.
- Ekvall, Ulla 2001. Den styrda och styrande läroboken. Björn Melander och Björn Olsson (red.), *Verklighetens texter. Sjutton fallstudier*, 43–80. Lund: Studentlitteratur.
- Ekvall, Ulla 2010. Läroboken i ett kemiklassrum. Cecilia Falk, Andreas Nord och Rune Palm (red.), *Svenskans beskrivning 30*, 70–81. Stockholm: Institutionen för nordiska språk.
- Ekvall, Ulla 2011. Lärobokstext och kemilektion. Niklas Ammert (red.), *Att spegla världen. Läromedelsstudier i teori och praktik*, 117–137. Lund: Studentlitteratur.
- Eldridge, John 2008. ”No, there isn’t an ‘academic vocabulary’, but...”. *TESOL Quarterly* 42 (1): 109–113.
- Englund, Boel 2011. Vad gör läroböcker? Niklas Ammert (red.), *Att spegla världen. Läromedelsstudier i teori och praktik*, 279–294. Lund: Studentlitteratur.
- Englund, Boel och Per Ledin 2003. Inledning. Boel Englund och Per Ledin (red.), *Teoretiska perspektiv på sakprosa*, 13–31. Lund: Studentlitteratur.
- Englund, Boel, Per Ledin och Jan Svensson 2003. Sakprosa – vad är det? Boel Englund och Per Ledin (red.), *Teoretiska perspektiv på sakprosa*, 35–59. Lund: Studentlitteratur.
- Fabricius, Susanne, Fredrik Holm och Anders Nystrand 2013. *Spektrum biologi, grundbok*. Stockholm: Liber.

- Fagerström, Emelie och Lizette Karlsson 2014. Ett utökat urval av litteratur till elever med lässvårigheter. Kan elever med lässvårigheter läsa icke lättläst barnlitteratur? Linnéuniversitetet. Examensarbete, 15 hp.
- Feng, Lijun, Noémie Elhadad och Matt Huenerfauth 2009. Cognitively motivated features for readability assessment. *Proceedings of the 12th conference of the European chapter of the ACL*, 229–237. Aten: ACL.
- Fischer-Starcke, Bettina 2009. Keywords and frequent phrases of Jane Austen's 'Pride and prejudice': a corpus-stylistic analysis. *International Journal of Corpus Linguistics* 14 (4): 492–523.
- Fischer-Starcke, Bettina 2010. *Corpus linguistics in literary analysis: Jane Austen and her contemporaries*. London: Continuum.
- Flowerdew, John 2002. Corpus-based analyses in EAP. John Flowerdew (red.), *Academic discourse*, 95–114. London: Pearson.
- Flyman Mattsson, Anna och Gisela Håkansson 2011. *Bedömning av svenska som andraspråk. En analysmodell baserad på grammatiska utvecklingsstadier*. Lund: Studentlitteratur.
- Forsberg, Markus, Richard Johansson, Linnéa Bäckström, Lars Borin, Benjamin Lyngfelt, Joel Olofsson och Julia Prentice 2014. From construction candidates to construction entries: an experiment using semi-automatic methods for identifying constructions in corpora. *Constructions and Frames* 6 (1): 114–135.
- Gardner, Dee och Mark Davies 2014. A new academic vocabulary list. *Applied Linguistics* 35 (3): 305–327.
- af Geijerstam, Åsa 2006. *Att skriva i naturorienterade ämnen i skolan*. Uppsala: Uppsala universitet.
- af Geijerstam, Åsa och Jenny Wiksten Folkeryd 2006. Elevtexter i grundskolan. *Textvård. Att läsa, skriva och bedöma texter*, 36–44. Stockholm: Norstedts.
- Gellerstam, Martin 1989. Om svenskan i översättningar från engelskan. Sture Allén, Martin Gellerstam och Sven-Göran Malmgren (red.), *Orden speglar samhället*, 105–117. Stockholm: Allmänna förlaget.
- Gellerstam, Martin 1996. Anföringens estetik. Olle Josephson (red.), *Stilstudier*, 12–29. Uppsala: Hallgren och Fallgren.
- Ghadessy, M 1979. Frequency counts, word lists, and materials preparation: a new approach. *English Teaching Forum* 17 (1): 24–27.
- Gibbons, Pauline 2014. Fem steg för planering av ett språk- och kunskapsutvecklande arbetsområde. Skolverket. <http://www.skolverket.se>.

- Gledhill, Chris 2000. The discourse function of collocation in research article introductions. *English for Specific Purposes* 19: 115–135.
- Golden, Anne 2005. *Å gripe poenget. Förståelse av metaforiske uttrykk fra lærebøker i samfunnskunnskap hos minoritetslever i ungdomsskolen*. Oslo: Acta Humaniora 227, UniPub forlag.
- Government of South Australia 2015. Australian curriculum: Literacy. <http://www.australiancurriculum.edu.au/GeneralCapabilities/Pdf/Literacy>. Informationen hämtad 10 maj 2015.
- Granger, Sylviane och Magali Paquot 2009. In search of general academic English: a corpus-driven study. Kallia Katsampoxaki-Hodgetts (red.), *Options and practices of L.S.P practitioners conference proceedings*, 94–108. University of Crete Publications.
- Gries, Stefan Th. 2006. Exploring variability within and between corpora: some methodological considerations. *Corpora* 1 (2): 109–151.
- Gronlund, Norman E. 2000. *How to write and use instructional objectives*. Needham heights: Allyn and Bacon.
- Halácsy, Péter, András Kornai och Csaba Oravecz 2007. Hunpos – an open source trigram tagger. *Proceedings of the 45th annual meeting of the association for computational linguistics, Proceedings of the demo and poster sessions*, 209–212. Prag: ACL.
- Halliday, Michael Alexander Kirkwood 1993. Some grammatical problems in scientific English. Michael Alexander Kirkwood Halliday och Jim R. Martin (red.), *Writing science: literacy and discursive power*, 69–85. Pittsburgh: University of Pittsburgh Press.
- Halliday, Michael Alexander Kirkwood 1998. Things and relations. Regrammaticising experience as technical knowledge. Jim R. Martin och Robert Veel (red.), *Reading science. Critical and functional perspectives on discourse of science*, 185–235. New York: Routledge.
- Halliday, Michael Alexander Kirkwood och Ruqaiya Hasan 1985. *Context and text: aspects of language in a social-semiotic perspective*. Geelong, Victoria: Deakin University Press.
- Halliday, Michael Alexander Kirkwood och Jim R. Martin 1993. *Writing science: literacy and discursive power*. Pittsburgh: University of Pittsburgh Press.
- Heimann Mühlenbock, Katarina 2013. *I see what you mean. Assessing readability for specific target groups*. Göteborg: University of Gothenburg.
- Hellspong, Lennart och Per Ledin 1997. *Vägar genom texten: handbok i brukstextanalys*. Lund: Studentlitteratur.

- Hipkiss, Anna Maria 2014. *Klassrummets semiotiska resurser: en språkdidaktisk studie av skolämnena hem- och konsumentkunskap, biologi och kemi*. Umeå: Umeå universitet, Studier i språk och litteratur från Umeå universitet 23.
- Hirsh, David och Paul Nation 1992. What vocabulary size is needed to read unsimplified texts for pleasure? *Reading in a Foreign Language* 8: 689–696.
- Hlaváčová, Jaroslava 2000. Rarity of words in a language and in a corpus. *Proceedings of the 2nd international conference on language resources and evaluation*, 1595–1598. Aten: ELRA.
- Holmberg, Per, Anna-Malin Karlsson och Andreas Nord 2011. Funktionell grammatik och textanalys. Per Holmberg, Anna-Malin Karlsson och Andreas Nord (red.), *Funktionell textanalys*, 7–18. Stockholm: Norstedts.
- Holmegaard, Margareta, Sofie Johansson Kokkinakis, Jerker Järborg, Inger Lindberg och Karin Sandvall 2005. Projektet Ord i läroböcker (OrdiL). Inger Lindberg och Karin Sandvall (red.), *Språket och kunskapen – att lära på sitt andraspråk i skola och högskola*, 149–182. Göteborg: Göteborgs universitet. Institutet för svenska som andraspråk.
- Holmlund, Helena, Josefin Häggblom, Erica Lindahl, Sara Martinson, Anna Sjögren, Ulrika Vikman och Björn Öckert 2014. *Decentralisering, skolval och fristående skolor: resultat och likvärdighet i svensk skola, Rapport 2014:25*. Uppsala: IFAU.
- Hoover, David L. 2001. Statistical stylistics and authorship attribution: an empirical investigation. *Literary and Linguistic computing* 16 (4): 421–444.
- Hoover, David L. 2008. Quantitative analysis and literary studies. Ray Siemens och Susan Schreibman (red.), *A companion to digital literary studies*, 69–85. Oxford: Blackwell.
- Hori, Masahiro 2004. *Investigating Dickens' style: a collocational analysis*. Basingstoke: Palgrave Macmillan.
- Hulstijn, Jan H. 2011. Language proficiency in native and non-native speakers: an agenda for research and suggestions for second-language assessment. *Language Assessment Quarterly* 8: 229–249.
- Hultman, Tor G. 1988. Skrivandet i skolan sett ur ett utvecklingsperspektiv. Britt-Louise Gunnarsson, Caroline Liberg och Staffan Wahlén (red.), *Skrivande. Rapport från ASLA:s nordiska symposium, Uppsala, 10–12 november 1988*, 69–89. Uppsala: ASLA.

- Hultman, Tor G. 1989. Skrivutveckling i ett pragmatiskt perspektiv. Carin Sandqvist och Ulf Teleman (red.), *Språkutveckling under skoltiden*, 117–142. Lund: Studentlitteratur.
- Hultman, Tor G. 2003. *Svenska Akademiens språklära*. Stockholm: Norstedts Akademiska Förlag.
- Hultman, Tor G. och Margareta Westman 1977. *Gymnasistsvenska*. Lund: Liber.
- Huntley, Helen 2006. *Essential academic vocabulary: mastering the complete academic word list*. Boston: Houghton Mifflin Company.
- Hvenekilde, Anne 1993. *Matte på ett språk vi förstår*. Stockholm: Skriptor/-Almqvist & Wiksell.
- Hyland, Ken och Polly Tse 2007. Is there an "academic vocabulary"? *TESOL Quarterly* 41 (2): 235–253.
- Högskoleverket 2011. Standard för svensk indelning av forskningsämnen 2011. Högskoleverket.
- Ingo, Rune 2007. *Konsten att översätta*. Lund: Studentlitteratur.
- Iversen Kulbrandstad, Lise 1996. *Lesing på et andrespråk. En studie av fire innvandrerdommers lesing av lærebokstekster på norsk*. Oslo: Acta Humaniora 30, UniPub forlag.
- Jansson, Håkan, Sofie Johansson Kokkinakis, Judy Ribbeck och Emma Sköldberg 2012. A Swedish academic word list: methods and data. Ruth Vatvedt Fjeld och Julie Matilde Torjusen (red.), *Proceedings of the XV Euralex international congress*, 955–960. Oslo: University of Oslo.
- Johansson Kokkinakis, Sofie och Ulrika Magnusson 2009. Quantitative measures on student texts. Päivi Juvonen (red.), *Språk och lärande. Rapport från ASLA:s höstsymposium, Stockholm, 7–9 november 2008*, 43–56. Stockholm: Stockholms universitets förlag.
- Johansson Kokkinakis, Sofie, Emma Sköldberg, Birgit Henriksen, Kari Kinn och Janne Bondi Johannessen 2012. Developing academic word lists for Swedish, Norwegian and Danish – a joint research project. *Proceedings of the XV Euralex international congress*, 563–569. Oslo: University of Oslo.
- Josephson, Olle 2004. *Ju: ifrågasatta självklarheter om svenskan, engelskan och alla andra språk i Sverige*. Stockholm: Norstedts Akademiska Förlag.
- Josephson, Olle 2006. Ibland är engelskan ordfattigare än svenskan. Svenska Dagbladet, 28 oktober. <http://www.svd.se/ibland-ar-engelskan-ordfattigare-an-svenskan>.

- Josephsson, Olle, Lars Melin och Tomas Oliv 1990. *Elevtext*. Lund: Studentlitteratur.
- Jurafsky, Daniel och James H. Martin 2009. *Speech and language processing*. New Jersey: Prentice Hall. 2nd ed.
- Karlsson, Anna-Malin 2006. *En arbetsdag i skriftsamhället. Ett etnografiskt perspektiv på skriftenvändning i vanliga yrken*. Stockholm: Norstedts.
- Karlsson, Anna-Malin och Hans Strand 2012. Text i verksamhet: mot en samlad förståelse. *Språk och stil. Tidskrift för svensk språkforskning. Tema text* 22 (1): 110–134.
- Kiland, Erika 2012. Textbedömning och kvantitativa mått. En undersökning av textlängd, ordförråd och syntaktisk komplexitet i elevtexter i förhållande till betyg. Göteborgs universitet, Institutionen för svenska språket. Examensarbete, 15 hp.
- Kilgarriff, Adam 2001. Comparing corpora. *International Journal of Corpus Linguistics* 6 (1): 1–37.
- Kress, Gunther, Carey Jewitt, Jon Ogborn och Charalampos Tsatsarelis 2001. *Multimodal teaching and learning: the rhetorics of the science classroom*. London: Continuum.
- Kuisma, Anu Emilia 2013. *Undersökning av läsbarheten i fem svenskspråkiga EU-broschyrer*. Östra Finlands universitet: Svenska språket.
- Kulturdepartementet, Svensk författningssamling 2009. Språklag (2009:600).
- Kungliga biblioteket 2012. I Nationalbibliografin redovisad utgivning (monografier). [http://www.kb.se/Dokument/Soka/I Nationalbibliografin redovisad utgivning 2012.pdf](http://www.kb.se/Dokument/Soka/I%20Nationalbibliografin%20redovisad%20utgivning%202012.pdf).
- Källgren, Gunnel 2006. SUC 2.0. Department of Linguistics, Stockholm University.
- Lagerholm, Per 2008. *Stilistik*. Lund: Studentlitteratur.
- Laufer, Batia 1992. How much lexis is necessary for reading comprehension? Pierre J. L. Arnaud och Henri Béjoint (red.), *Vocabulary and applied linguistics*, 126–132. London: Macmillan.
- Laufer, Batia 1996. The lexical threshold of L2 reading: where it is and how it relates to L1 reading ability. Kari Sajavaara och Courtney Fairweather (red.), *Approaches to second language acquisition*, 55–62. University of Jyväskylä: Jyväskylä.
- Ledin, Per och Staffan Selander 2003. Institution, text och genre. Boel Englund och Per Ledin (red.), *Teoretiska perspektiv på sakprosa*, 91–122. Lund: Studentlitteratur.

- Lemke, Jay L. 2000. Multimedia literacy demands of the scientific curriculum. *Linguistics and Education* 10 (3): 247–271.
- Li, Yifan, Petr Musilek, Marek Reformat och Loren Wyard-Scott 2009. Identification of pleonastic *it* using the web. *Journal of Artificial Intelligence Research* 34: 339–389.
- Liberg, Caroline 2009. Genrepedagogik i ett didaktiskt perspektiv. Päivi Juonen (red.), *Rapport från ASLA:s höstsymposium 2008*. Stockholm: ASLA:s skriftserie.
- Liberg, Caroline, Jenny Wiksten Folkeryd, Åsa af Geijerstam och Agnes Edling 1999. Students' encounters with different texts in school – a short presentation. www.did.uu.se/carolineliberg/forskotext.html.
- Liljestrand, Birger 1993. *Språk i text. Handbok i stilistik*. Lund: Studentlitteratur.
- Lin, Dekang 1996. On the structural complexity of natural language sentences. *Coling 1996, Volume 2: the 16th international conference on computational linguistics*. Köpenhamn: ACL.
- Lindberg, Inger 2007. Bedömning av skolrelaterat ordförråd. Mikael Olofsson (red.), *Bedömning, flerspråkighet och lärande*. Stockholm: LHS förlag.
- Lindberg, Inger 2009. Conceptualizing school-related, academic language – theoretical and empirical approaches. Bjørg-Karin Ringen, Ole Kolbjørn Kjørven och Antoinette Gagné (red.), *Teacher diversity in a diverse school – challenges and opportunities for teacher education*, 178–192. Vallset: Opplandske forlag.
- Lindberg, Inger och Sofie Johansson Kokkinakis 2007. *OrdiL – en korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år*. Göteborg: Göteborgs universitet.
- Lindberg, Inger och Sofie Johansson Kokkinakis 2008. Word type grouping in Swedish secondary school textbooks – an inventory of words from a second language perspective. *Teaching and language corpora (TALC) conference*. Lissabon.
- Lindberg, Inger och Sofie Johansson Kokkinakis 2011. Identification of lexical cohesive ties in secondary school text books. *AILA 2011. The 16th world congress of applied linguistics*. Peking.
- Loman, Bengt och Nils Jörgensen 1971. *Manual för analys och beskrivning av makrosyntaxer*. Lund: Studentlitteratur.
- Lundberg, Ingvar 2002. The child's route into reading and what can go wrong. *Dyslexia* 8: 1–13.

- Lundberg, Ingvar och Monica Reichenberg 2008. *Vad är lättläst?* Härnösand: Specialpedagogiska skolmyndigheten.
- Løkensgard Hoel, Torlaug 2010. *Skriva på universitet och högskolor*. Lund: Studentlitteratur.
- Macken-Horarik, Mary 1996. Literacy and learning across the curriculum: towards a model of register for secondary school teachers. Rugalya Hasan och Geoffrey Williams (red.), *Literacy in society*, 232–279. London: Longman.
- Magnusson, Ulrika 2008. *Språk i ämnet*. Stockholm: Skolverket.
- Magnusson, Ulrika 2010. *Skolspråk i utveckling. En- och flerspråkiga elevers bruk av grammatiska metaforer i senare skolår*. Göteborg: Institutionen för svenska språket, Göteborgs universitet.
- Mahlberg, Michaela 2007. Corpus stylistics: bridging the gap between linguistic and literary studies. Michael Hoey, Michaela Mahlberg, Michael Stubbs och Wolfgang Teubert (red.), *Text, discourse and corpora. Theory and analysis*, 217–246. London: Continuum.
- Mahlberg, Michaela 2015. *Corpus stylistics and Dickens's fiction*. London: Routledge.
- Mair, Christian 2006. *Twentieth-century English. History, variation and standardization*. Cambridge: Cambridge University Press.
- Malvern, David D., Brian J. Richards, Ngoni Chipere och Pilar Durán 2004. *Lexical diversity and language development: quantification and assessment*. Basingstoke: Palgrave Macmillan.
- Marco, Maria J. L. 2000. Collocational frameworks in medical research papers: a genre-based study. *English for Specific Purposes* 19: 63–83.
- Martin, Jim R. 1993. Technicality and abstraction: language for the creation of specialized texts. Michael Alexander Kirkwood Halliday och Jim R. Martin (red.), *Writing science: literacy and discursive power*, 203–220. Pittsburgh: University of Pittsburgh Press.
- Martin, Jim R. och David Rose 2008. *Genre relations. Mapping culture*. London: Equinox Publishing Ltd.
- Martin, Jim R. och Robert Veel 1998. *Reading science. Critical and functional perspectives on discourse of science*. London: Routledge.
- Mays, Eric, Fred J. Damerau och Robert L. Mercer 1991. Context based spelling correction. *Information Processing and Management* 27 (5): 517–522.

- McKenna, C. Wayne F. och Alexis Antonia 2001. The statistical analysis of style: reflections on form, meaning, and ideology in the 'Nausicaa' episode of *Ulysses*. *Literary and Linguistic Computing* 16 (4): 353–373.
- Megyesi, Beáta B. 2009. The open source tagger HunPoS for Swedish. Kristiina Jokinen och Eckhard Bick (red.), *Proceedings of the 17th Nordic conference of computational linguistics (NODALIDA)*. Odense: NEALT.
- Melander, Björn 1991. *Innehållsmönster i svenska facktexter*. Uppsala universitet: Institutionen för nordiska språk.
- Melander, Björn 1995. Läsebokssvenska, bruksprosa och begreppslighet. En översikt över svensk språkforskning kring läroböcker. Siv Strömquist (red.), *Läroboksspråk. Om språk och layout i svenska läroböcker*, 12–46. Uppsala: Hallgren och Fallgren.
- Melander, Björn 2003. Läroboksspråket – en flintskallig primadonna? Staffan Selander (red.), *Kobran, nallen och majjen. Tradition och förnyelse i svensk skola och skolforskning*, 133–153. Stockholm: Myndigheten för skolutveckling & Liber.
- Mel'čuk, Igor 1988. *Dependency syntax: theory and practice*. Albany: State University of New York Press.
- Melin, Lars 1992. Textbindning och läsbarhet. Ulla-Britt Kotsinas, Inger Lindell och Per Ledin (red.), *Svenskans beskrivning 19*, 225–233. Lund: Lund University Press.
- Melin, Lars 1995. Grafisk pyttipanna. Om text och grafisk form i läroböcker. Siv Strömquist (red.), *Läroboksspråk. Om språk och layout i svenska läroböcker*, 77–123. Uppsala: Hallgren och Fallgren.
- Melin, Lars 2007. Packat och oklart. Informationspackningens roll för begriplighet i text. Lotta Collin, Saara Haapamäki, Marketta Sundman och Lars Wollin (red.), *Under språkets hud. Festskrift till Erik Andersson på 60-årsdagen*, 147–155. Åbo: Åbo Akademis förlag.
- Melin, Lars och Sven Lange 2000. *Att analysera text*. Lund: Studentlitteratur.
- Moens, Marie-Francine, Caroline Uyttendaele och Jos Dumortier 2000. Intelligent information extraction from legal texts. *Information and Communications Technology Law* 9 (1): 17–26.
- Mühlenbock, Katarina 2009. Readable, legible or plain words – presentation of an easy-to-read Swedish corpus. *Multilingualism, Proceedings of the 23rd Scandinavian conference of linguistics*, 325–327. Uppsala: Uppsala University.
- Nation, Paul 1997. Vocabulary size, text coverage and word lists. Norbert Schmitt och Michael McCarthy (red.), *Vocabulary: description, acquisition, pedagogy*, 6–19. New York: Cambridge University Press.

- Nation, Paul 2001. *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Nation, Paul och Kyongho Hwang 1995. Where would general service vocabulary stop and special purposes vocabulary begin? *System* 23 (1): 35–41.
- Nettelblad, Folke och Karin Nettelblad 2013. *Spektrum kemi, grundbok*. Stockholm: Liber.
- Nilsson Björkenstam, Kristina, Sofia Gustafson Capková och Mats Wirén 2014. The Stockholm University Strindberg corpus: content and possibilities. Roland Lysell (red.), *Strindberg on international stages/Strindberg in translation*. Cambridge: Cambridge Scholars Publishing.
- Nivre, Joakim, Johan Hall och Jens Nilsson 2006. MaltParser: a data-driven parser-generator for dependency parsing. *Proceedings of the 5th international conference on language resources and evaluation, May 24–26, 2006*. Genua: ELRA.
- Nivre, Joakim, Jens Nilsson och Johan Hall 2006. A Swedish treebank with phrase structure and dependency annotation. *Proceedings of the 5th international conference on language resources and evaluation, May 24–26, 2006*, 1392–1395. Genua: ELRA.
- Nord, Andreas 2012. Varför textanalys? Den språkvetenskapliga textanalysen i högre utbildning ur ett textforskningsperspektiv. *Språk och stil. Tidskrift för svensk språkforskning. Tema text* 22 (1): 187–212.
- Nordenfors, Mikael 2011. *Skriftspråsutveckling under högstadiet*. Göteborg: Göteborgs universitet, GNS.
- Nygård Larsson, Pia 2009. Hur systematisk är systematiken? Några semiotiska resursers betydelse för skapandet av explicita tematiska mönster i lärartext och lärobok. Päivi Juvonen (red.), *Språk och lärande. Rapport från ASLA:s höstsymposium, Stockholm, 7–9 november 2008*, 71–83. Stockholm: Stockholm universitets förlag.
- Nygård Larsson, Pia 2011. *Biologiämnets texter: text, språk och lärande i en språkligt heterogen gymnasieklass*. Malmö: Lärarutbildningen, Malmö högskola.
- Nyström, Catharina 2000. *Gymnasisters skrivande. En studie av genre, textstruktur och sammanhang*. Uppsala: Institutionen för nordiska språk vid Uppsala universitet.
- Nyström Höög, Catharina 2006. Gymnasisttexter. *Textvård. Att läsa, skriva och bedöma texter*, 36–44. Stockholm: Norstedts.
- Nyström Höög, Catharina 2010. *Mot ökad diskursivitet? Skrivutveckling speglad i provtexter från årskurs 5 och årskurs 9*. Uppsala: Institutionen för nordiska språk, Arbetsrapporter om modern svenska.

- Oakey, David 2002. Formulaic language in English academic writing: a corpus-based study of the formal and functional variation of a lexical phrase in different academic disciplines. Randi Reppen, Susan M. Fitzmaurice och Douglas Biber (red.), *Using corpora to explore linguistic variation*, 111–129. Amsterdam: John Benjamins Publishing Company.
- O'Halloran, Kieran 2007. The subconscious in James Joyces 'Eveline': a corpus stylistic analysis that chews on the 'fish hook'. *Language and Literature* 16 (3): 227–244.
- Paquot, Magali 2007. Towards a productively-oriented academic word list. Jacek Walinski, Krzysztof Kredens och Stanislaw Gozdz-Roszkowski (red.), *Corpora and ICT in language studies, PALC 2005*, 127–140. Frankfurt am Main: Peter Lang.
- Paquot, Magali 2010. *Academic vocabulary in learner writing. From extraction to analysis*. London & New York: Continuum.
- Partanen, Ilpo 2006. *Jämförelse av webbtidningstexters kvantifierbara läsbarhet i Dagens Nyheter och Hufvudstadsbladet*. Jyväskylä universitet: Jyväskylä.
- Platzack, Christer 1973. *Språket och läsbarheten*. Lund: Gleerups Bokförlag.
- Praninskas, Jean 1972. *American university word list*. London: Longman.
- Prentice, Julia 2010. *På rak sak. Om ordförbindelser och konventionaliserade uttryck bland unga språkbrukare i flerspråkiga miljöer*. Göteborg: Institutionen för svenska språket, Göteborgs universitet, GNS.
- Rabiner, Lawrence och Biing-Hwang Juang 1993. *Fundamentals of speech recognition*. Upper Saddle River, NJ: PTR Prentice Hall.
- Rayson, Paul och Rayson Garside 2000. Comparing corpora using frequency profiling. *Proceedings of the workshop on comparing corpora, the 38th annual meeting of the association for computational linguistics*, 1–6. Aten: ACL.
- Read, John 2000. *Assessing vocabulary*. Cambridge: Cambridge University Press.
- Reichenberg, Monica 2000. *Röst och kausalitet i lärobokstexter*. Göteborg: Göteborg Studies in Educational Sciences 149.
- Reichenberg, Monica 2010. Forskning om läromedel. http://www.slff.se/global/forskningsbibliografi/forskningsbibliografi_2010.pdf. Sveriges Läromedelsförfattares Förbund – SLFF.
- Reichenberg, Monica och Ingvar Lundberg 2011. Varför behövs lättlästa läroböcker? Niklas Ammert (red.), *Att spegla världen. Läromedelsstudier i teori och praktik*, 317–341. Lund: Studentlitteratur.

- Ribeck, Judy 2011. Identifying lexical bundles in secondary school textbooks. Zygmunt Vetulani (red.), *Human language technologies as a challenge for computer science and linguists. Proceedings of the 5th language and technology conference*, 202–206. Poznań: Fundacja Uniwersytetu im. A. Mickiewicza.
- Ribeck, Judy 2014. I magen på piktofagen. Visuella kommunikationsmedel i gymnasieläroböcker i biologi, fysik, kemi och matematik. Opublicerad uppsats i forskarkursen ”Textteori och textanalys”, Göteborgs universitet.
- Ribeck, Judy och Lars Borin 2014. Lexical bundles in Swedish secondary school textbooks. *Human language technology challenges for computer science and linguistics. Lecture notes in computer science*, 238–249. Berlin: Springer International Publishing.
- Ribeck, Judy, Håkan Jansson och Emma Sköldberg 2014. Från aspekt till övergripande – en ordlista över svensk akademisk vokabulär. Ruth Vatvedt Fjeld och Marit Hovdenak (red.), *Nordiske studier i leksikografi 12. Rapport fra Konferanse om leksikografi i Norden, Oslo, 2013*, 370–384.
- Ritter, Alan, Sam Clark, Mausam och Oren Etzioni 2011. Named entity recognition in tweets: an experimental study. *Proceedings of the 2011 conference on empirical methods in natural language processing*. Edinburgh: ACL.
- Romaine, Suzanne 1984. *The language of children and adolescents: the acquisition of communicative competence*. Oxford och New York: Blackwell.
- Rongen Breivega, Kjersti 2003. *Vitskaplege argumentasjonsstrategiar*. Oslo: Norsk sakprosa i samarbeid med Norsk faglitterær forfatter- og oversetterforening.
- Rose, Ted och Adam Kilgarrieff 1998. Measures of corpus similarity and homogeneity between corpora. *Proceedings of the 3rd conference on empirical methods in natural language processing (EMNLP-3)*, 46–52. Granada: ACL.
- Salö, Linus 2010. Engelska eller svenska? En kartläggning av språksituationen inom högre utbildning och forskning. Rapporter från Språkrådet 1. Språkrådet.
- Sandqvist, Carin 1995. Om form och innehåll i historieläroböcker. Siv Strömquist (red.), *Läroboksspråk. Om språk och layout i svenska läroböcker*, 124–188. Uppsala: Hallgren och Fallgren.
- Savický, Petr och Jaroslava Hlaváčová 2002. Measure of word commonness. *Journal of Quantitative Linguistics* 9: 215–231.
- Saville-Troike, Muriel 1984. What really matters in second language learning for academic achievement? *TESOL Quarterly* 18 (2): 199–219.

- Schleppegrell, Mary J. 2004. *The language of schooling. A functional linguistics perspective*. New Jersey: Routledge.
- Schmitt, Norbert och Diane Schmitt 2005. *Focus on vocabulary: mastering the academic word list*. London: Longman.
- Schmitt, Norbert, Diane Schmitt och Caroline Clapham 2001. Developing and exploring the behaviour of two new versions of the vocabulary levels test. *Language Testing* 18: 55–88.
- Schwarm, Sara E. och Mari Ostendorf 2005. Reading level assessment using support vector machines and statistical language models. *Proceedings of the 43rd annual meeting of association for computational linguistics*, 523–530. Stroudsburg: ACL.
- Scott, Mike 1997. PC analysis of key words – and key key words. *System* 25 (2): 233–245.
- Sebastiani, Fabrizio 2002. Machine learning in automated text categorization. *ACM Computing Surveys* 34 (1): 1–47.
- Selander, Staffan 1988. *Lärobokskunskap. Pedagogisk textanalys med exempel från läroböcker i historia 1841–1985*. Lund: Studentlitteratur.
- Selander, Staffan 2001. Läroplan och sociala intressen. Björn Melander och Björn Olsson (red.), *Verklighetens texter. Sjutton fallstudier*, 121–147. Lund: Studentlitteratur.
- Sfard, Anna 2007. When the rules of discourse change, but nobody tells you: making sense of mathematics learning from a commognitive standpoint. *The Journal of the Learning Sciences* 16 (4): 565–613.
- Siljan, Henriette 2011. *Metaforisering, nominalisering og normering. En teoretisk studie av grammatisk metafor og to empiriske undersøkelser av språktrekket i læreboktekster*. Oslo: Universitetet i Oslo.
- Simpson-Vlach, Rita och Nick C. Ellis 2010. An academic formulas list: new methods in phraseology research. *Applied Linguistics* 31 (4): 487–512.
- Skolverket 2010. Ämnesplaner och kurser för gymnasieskolan 2011. <http://www.skolverket.se/laroplaner-amnen-och-kurser/gymnasie-utbildning/gymnasieskola>.
- Skolverket 2011a. Gymnasieskola 2011.
- Skolverket 2011b. Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011.
- Skolverket 2013. PISA 2012. 15-åringars kunskaper i matematik, läsförståelse och naturvetenskap. Rapport 398. Skolverket.
- Skolverket 2014a. Grundskolan i internationella kunskapsmätningar – kunskap, skolmiljö och attityder till lärande. Rapport 407. Skolverket.

- Skolverket 2014b. Vad innebär en språkutvecklande undervisning? <http://www.skolverket.se/skolutveckling/lorande/nt/grundskoleutbildning/>.
- Skolverket 2015. Språkutvecklande arbetssätt. <http://www.skolverket.se/loroplaner-amnen-och-kurser/grundskoleutbildning/diskutera-och-utveckla/sprakutvecklande-arbetsatt>.
- Sköldberg, Emma och Sofie Johansson Kokkinakis 2012. A och O om akademiska ord. Om framtagning av en svensk akademisk ordlista. *Nordiska studier i lexikografi 11. Rapport från konferensen om lexikografi i Norden, maj 2011*, 575–585. Lund: NFL.
- Snow, Catherine, E. 2010. Academic language and the challenge of reading for learning about science. *Science* 328: 450–452.
- Språkrådet 2014. *Myndigheternas skrivregler*. Stockholm: Norstedts Juridik AB/Fritzes.
- Strömquist, Siv 1995. Läroboksspråk – mönster eller monster? Siv Strömquist (red.), *Läroboksspråk. Om språk och layout i svenska läroböcker*, 7–11. Uppsala: Hallgren och Fallgren.
- Stubbs, Michael 1986. Language development, lexical competence and nuclear vocabulary. Michael Stubbs (red.), *Educational Linguistics*, 98–115. Oxford & New York: Blackwell.
- Stubbs, Michael 2005. Conrad in the computer: examples of quantitative stylistic methods. *Language and Literature* 14 (1): 5–24.
- Sundin, Maria 2007. *Lättläst – så funkar det*. Stockholm: Jure Förlag.
- Sutcliffe, Richard F.E. och Bronwyn E.A. Slater 1995. Disambiguation by association as a practical method: experiments and findings. *Journal of Quantitative Linguistics* 2 (1): 43–52.
- Svensk ordbok 2009. *Svensk ordbok utgiven av Svenska Akademien*. Stockholm: Svenska Akademien/Norstedts i distribution).
- Svenska Dagbladet 2013. Björklund: kommuner saknar kompetens. 3 december 2013. http://www.svd.se/nyheter/inrikes/bjorklund-kommuner-saknar-kompetens_8786760.svd.
- Svensk akademisk ordlista*, version 2.0 2015. <http://spraakbanken.gu.se/ao/>. Sidan uppdaterad 17 februari 2015.
- Svensén, Bo 2004. *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.
- Svensson, Ann-Christine 2011. Hur används läroboken? Niklas Ammert (red.), *Att spegla världen. Läromedelsstudier i teori och praktik*, 295–315. Lund: Studentlitteratur.

- Svensson, Jan och Anna-Malin Karlsson 2012. Inledning: text, textforskning och textteori. *Språk och stil. Tidskrift för svensk språkforskning. Tema text* 22 (1): 5–30.
- Säljö, Roger 2014. *Lärande i praktiken. Ett sociokulturellt perspektiv*. Lund: Studentlitteratur.
- Teich, Elke 2003. *Cross-linguistic variation in system and text. A methodology for the investigation of translations and comparable texts*. Berlin/New York: Mouton de Gruyter.
- Teleman, Ulf 1974. *Manual för grammatisk beskrivning av talad och skriven svenska*. Lund: Studentlitteratur.
- Teleman, Ulf 1988. Veta och kunna: om metakunskapens roll vid produktion av skriftliga texter. Britt-Louise Gunnarsson, Caroline Liberg och Staffan Wahlén (red.), *Skrivande. Rapport från ASLA:s nordiska symposium, Uppsala, 10–12 november 1988*, 111–122. Uppsala: ASLA.
- Teleman, Ulf 1989. Språkutveckling under skoltiden: ett problemfält. Carin Sandqvist och Ulf Teleman (red.), *Språkutveckling under skoltiden*, 7–13. Lund: Studentlitteratur.
- Teleman, Ulf, Staffan Hellberg och Erik Andersson 1999. *Svenska Akademiens grammatik (SAG)*. Stockholm: Svenska Akademien/Norstedts i distribution.
- Thelander, Mats 1970. *Fiktionsprosa och sakprosa. Jämförande stilanalys med kvantitativ metod*. Lund: Studentlitteratur.
- Trondman, Mats 1994. *Bilden av en klassresa: sexton arbetarbarn på väg till och i högskolan*. Stockholm: Carlsson.
- Tønnesson, Johan L. 2008. *Hva er sakprosa?* Oslo: Universitetsforlaget.
- Undvall, Lennart och Anders Karlsson 2013. *Spektrum fysik, grundbok*. Stockholm: Liber.
- Unsworth, Len 2001. *Teaching multiliteracies across the curriculum. Changing contexts of text and image in classroom practice*. Buckingham: Open University Press.
- Veel, Robert 1997. Learning how to mean – scientifically speaking: apprenticeship into scientific discourse in the secondary school. Frances Christie och Jim R. Martin (red.), *Genre and institutions. Social processes in the workplace and school*, 161–195. London: Continuum.
- Viberg, Åke 1993. Andraspråksinläring i olika åldrar. Eva Cerú (red.), *Svenska som andraspråk. Mera om språket och inläringen. Lärarbok 2*, 13–83. Stockholm: Natur och Kultur.

- Virtanen, Tuija 2010. Variation across text and discourses: theoretical and methodological perspectives on text type and genre. Heidrun Dorgeloh och Anja Wanner (red.), *Syntactic variation and genre*, 53–84. Berlin/New York: Walter de Gruyter.
- Vygotskij, Lev S. 1962. *Thought and language*. Cambridge: Mass: MIT Press.
- Wallin Victorin, Margareta 2011. Bilder i läroböcker. Niklas Ammert (red.), *Att spegla världen. Läromedelsstudier i teori och praktik*, 219–233. Lund: Studentlitteratur.
- Wang, Karen Ming-Tzu och Paul Nation 2004. Word meaning in academic English: homography in the academic word list. *Applied Linguistics* 25 (3): 291–314.
- West, Michael 1953. *A general service list of English words*. London: Longman.
- Westman, Margareta 1974. *Bruksprosa*. Lund: LiberLäromedel.
- Wible, David och Nai-lung Tsao 2011. The StringNet lexico-grammatical knowledgebase and its application. *Proceedings of the workshop on multiword expressions: from parsing and generation to the real world (MWE 2011)*, 128–130. Portland, OR: ACL.
- Xue, Guoyi och Paul Nation 1984. A university word list. *Language Learning and Communication* 3: 215–229.
- Zhang, HuaRui, Chu-Ren Huang och Francesca Quattri 2012. SMR-Cmp: square-mean-root approach to comparison of monolingual contrastive corpora. *Proceedings of COLING 2012: demonstration papers*, 527–534. Mumbai: ACL.
- Österholm, Magnus 2005. Characterizing reading comprehension of mathematical texts. *Educational Studies in Mathematics* 63: 325–346.
- Östling, Robert 2013. Stagger: an open-source part of speech tagger for Swedish. *Northern European Journal of Language Technology* 3: 1–18.

Appendix



FÖRTECKNING ÖVER TERMINOLOGI

A.1 Återkommande förkortningar

AWL	The Academic Word List
BI	Biologi
FY	Fysik
GE	Geografi
HI	Historia
KE	Kemi
LIX	Läsbarhetsindex
LVIX	Lemmavariationsindex
MA	Matematik
NQ	Nominalkvot
NV	Naturvetenskap
OrdiL	Ord i Läromedel (forskningsprojekt)
OVIX	Ordvariationsindex
SFL	Systemisk funktionell lingvistik
SUC	Stockholm-Umeå corpus
SK	Samhällskunskap
SV	Samhällsvetenskap
SveAk	Svensk Akademisk korpus

A.2 Lingvistisk notation

A.2.1 Ordklasstaggar

AB	adverb
DT	bestämningsord
HA	frågande/relativt adverb
HD	frågande/relativt bestämning
HP	frågande/relativt pronomen
HS	frågande/relativt possessivuttryck
IN	interjektion
JJ	adjektiv
KN	konjunktion
NN	substantiv
PC	particip
PL	partikel
PM	egennamn
PN	pronomen
PP	preposition
PS	possessivuttryck
SN	subjunktion
VB	verb

A.2.2 Frastaggar

AN	appositionsfras
AT	adjektivfras
DT	bestämningsordsfras
HD	huvudordsfras
KN	konjunktionsfras
NP	nominalfras
PP	prepositionsfras
VB	bisats

A.2.3 Syntaktiska taggar

+A	Konjunktionellt adverb
+F	Koordination på huvudsatsnivå
AA	Annat adverbial
AT	Framförställt attribut
CA	Kontrastivt adverbial
CJ	Samordnat led
DT	Determinerare bestämningsord
EF	Relativ bisats
ET	Efterställd bestämning
FV	Finit verb predikatsverb
HD	Huvud
I?	Frågetecken
IF	Infinitivfras utom infinitivmärke
IK	Kommatecken
IO	Indirekt objekt (dativobjekt)
IP	Punkt
IT	Divis (bindestreck)
MS	Makrosyntagm
NA	Negerande adverbial
OA	Objektsadverbial (prepositionsobjekt)
OO	Direkt objekt (akusativobjekt)
PA	Prepositionskomplement
PL	Verbpartikel
PT	Predikativt attribut (reaktion)
RA	Platsadverbial
SP	Subjektspredikativ (subjektiv predikatsfyllnad)
SS	Subjekt (övrigt subjekt)
TA	Tidsadverbial
UA	Underordnad sats (bisats) utom subjunktion
VA	Korrelativt adverbial
VG	Verbgrupp

B

MATERIALFÖRTECKNING

B.1 Undersökta läroböcker

- Biologi

- Högstadiet

- * *Biologi*, Fabricius, Susanne (2001). Stockholm: Almqvist & Wiksell.
 - * *PULS Biologi*, Andréasson, Berth (2001). Stockholm: Natur och Kultur.
 - * *PULS Biologi*, Andréasson, Berth m.fl. (2011). Stockholm: Natur och Kultur.
 - * *TitaNO Biologi*, Henriksson, Anders (2012). Interaktiv lärobok. Gleerups Utbildning AB.

- Gymnasiet

- * *Insikt 1*, Brynhildsen, Lena, Brändén, Henrik och Magnus Ehinger (2011). Stockholm: Natur och Kultur.
 - * *Insikt 2*, Brynhildsen, Lena och Henrik Brändén (2012). Stockholm: Natur och Kultur.
 - * *Iris biologi 1*, Henriksson, Anders (2011). Interaktiv lärobok. Gleerups Utbildning AB.
 - * *Iris biologi 2*, Henriksson, Anders (2013). Interaktiv lärobok. Gleerups Utbildning AB.

- Fysik

- Högstadiet

- * *Fysik*, Undvall, Lennart och Anders Karlsson (2001). Stockholm: Almqvist & Wiksell.
 - * *PULS Fysik*, Sjöberg, Staffan och Börje Ekstig (2001). Stockholm: Natur och Kultur.

- * *PULS Fysik*, Sjöberg, Staffan och Börje Ekstig (2011). Stockholm: Natur och Kultur.
- * *TitaNO FYSIK*, Monthan, Ingrid (2011). Interaktiv lärobok. Gleerups Utbildning AB.
- Gymnasiet
 - * *Heureka 1*, Alphonse, Rune (2011). Stockholm: Natur och Kultur.
 - * *Heureka 2*, Alphonse, Rune (2011). Stockholm: Natur och Kultur.
 - * *Impuls Fysik 1*, Fraenkel, Lars m.fl. (2011). Interaktiv lärobok. Gleerups Utbildning AB.
 - * *Impuls Fysik 2*, Fraenkel, Lars m.fl. (2013). Interaktiv lärobok. Gleerups Utbildning AB.
- Kemi
 - Högstadiet
 - * *Kemi*, Nettelblad, Folke A. och Christer Ekdahl (2002). Stockholm: Almqvist & Wiksell.
 - * *PULS Kemi*, Andréasson, Berth (2002). Stockholm: Natur och Kultur.
 - * *PULS Kemi*, Andréasson, Berth m.fl. (2011). Stockholm: Natur och Kultur.
 - * *Spektrum Kemi*, Netteblad, Folke och Karin Netteblad (2013). Stockholm: Liber.
 - * *TitaNO Kemi*, Sterning, Göran (2012). Interaktiv lärobok. Gleerups Utbildning AB.
 - Gymnasiet
 - * *Gymnasiekemi 1*, Andersson, Stig m.fl. (2013). Stockholm: Liber.
 - * *Gymnasiekemi 2*, Ellervik, Ulf m.fl. (2013). Stockholm: Liber.
 - * *Kemiboken 1*, Borén, Hans, Börner m.fl. (2011). Stockholm: Liber.
 - * *Kemiboken 2*, Borén, Hans, Börner m.fl. (2012). Stockholm: Liber.
 - * *Syntes kemi 1*, Henriksson, Anders (2011). Interaktiv lärobok. Gleerups Utbildning AB.
 - * *Syntes kemi 2*, Henriksson, Anders (2013). Interaktiv lärobok. Gleerups Utbildning AB.

- Matematik

- Högstadiet

- * *Formula 7*, Mårtensson, Gert m.fl. (2012). Interaktiv lärobok. Gleerups Utbildning AB.
 - * *Matematikboken X*, Undvall, Lennart m.fl. (2011). Stockholm: Liber.
 - * *Matematikboken Y, Röd*, Undvall, Lennart m.fl. (2002). Stockholm: Almqvist & Wiksell.
 - * *Matematikboken Y*, Undvall, Lennart m.fl. (2012). Stockholm: Liber.
 - * *Matematikboken Z*, Undvall, Lennart m.fl. (2013). Stockholm: Liber.
 - * *Mattestegen. C steg 9–12*, Andrén, Cecilia m.fl. (2002). Stockholm: Natur och Kultur.
 - * *Uppdrag Matte 7*, Wedbjer Rambell, Olga och Magnus Hansson (2010). Stockholm: Liber.
 - * *Uppdrag Matte 8*, Wedbjer Rambell, Olga och Magnus Hansson (2011). Stockholm: Liber.
 - * *Uppdrag Matte 9*, Wedbjer Rambell, Olga och Magnus Hansson (2012). Stockholm: Liber.

- Gymnasiet

- * *Exponent 1a*, Johansson, Lars-Göran och Tommy Olsson (2011). Interaktiv lärobok. Gleerups Utbildning AB.
 - * *Exponent 1b*, Silborn, Bo m.fl. (2011). Interaktiv lärobok. Gleerups Utbildning AB.
 - * *Exponent 1c*, Silborn, Bo m.fl. (2011). Interaktiv lärobok. Gleerups Utbildning AB.
 - * *Matematik 5000, 1abc, 2bc, 3bc, 4*, Alfredsson, Lena m.fl. (2011). Stockholm: Natur och Kultur.

- Geografi

- Högstadiet

- * *Geografi 8*, Jonasson, Christer m.fl. (2002). Stockholm: Natur och Kultur.
 - * *Geografi, Punkt SO, Del 2*, Andersson, Bo (2003). Malmö: Gleerups Utbildning AB.

- * *Utkik Geografi*, Andersson, Bo (2013). Interaktiv lärobok. Gleerups Utbildning AB.
- Gymnasiet
 - * *Geografi 1*, Östman, Peter (2010). Stockholm: Liber.
 - * *Geografi 2*, Östman, Peter (2010). Stockholm: Liber.
- Historia
 - Högstadiet
 - * *Historia, Punkt SO, Grundbok, del 2*, Nilsson, Erik m.fl. (2002). Malmö: Gleerups Utbildning AB.
 - * *Levande historia. 8*, Hildingson, Kaj och Lars Hildingson (2002). Stockholm: Natur och Kultur.
 - * *Utkik Historia*, Nilsson, Erik m.fl. (2013). Interaktiv lärobok. Gleerups Utbildning AB.
 - Gymnasiet
 - * *Alla tiders historia, Bas*, Almgren, Hans och Hans Albin Larsson (2013). Interaktiv lärobok. Gleerups Utbildning AB.
 - * *Perspektiv på historien*, Nyström, Hans m.fl. (2011). Interaktiv lärobok. Gleerups Utbildning AB.
- Samhällskunskap
 - Högstadiet
 - * *Samhälle i dag. 8*, Hildingson, Kaj m.fl. (2002). Stockholm: Natur och Kultur.
 - * *Samhällskunskap, Punkt SO*, Andersson, Göran (2003). Malmö: Gleerups.
 - * *Samhällskunskap – Så klart!* Leijon, Peter (2013). Interaktiv lärobok. Gleerups Utbildning AB.
 - Gymnasiet
 - * *Kompass till samhällskunskap*, Eliasson, Maria och Gunilla Nolervik (2011). Interaktiv lärobok. Gleerups Utbildning AB.
 - * *Reflex plus*, Almgren, Hans m.fl. (2011). Interaktiv lärobok. Gleerups Utbildning AB.
 - * *Spela roll*, Harstad, Fredrik m.fl. (2011). Interaktiv lärobok. Gleerups Utbildning AB.

B.2 Tidskrifter i den akademiska korpusen

Ämne	Tidskrift	Token
Etnologi	<i>Kulturella perspektiv – svensk etnologisk tidskrift</i>	103 306
Filosofi	<i>Tidskrift för politisk filosofi</i>	328 538
Historia	<i>Historisk tidskrift</i>	1 343 167
Konst	<i>Svensk tidskrift för musikforskning</i>	73 723
Litteraturv.	<i>Barnboken</i>	227 355
Psykologi	<i>Mellanrummet – nordisk tidskrift för barn- och ungdomspsykoterapi</i>	43 072
Statsv.	<i>Offentlig förvaltning</i>	41 983
	<i>Didaktisk tidskrift</i>	68 451
Utbildningsv.	<i>Högre utbildning</i>	166 244
	<i>Pedagogisk forskning i Sverige</i>	120 926

Tabell B.1: Tidskrifter i den akademiska korpusen, SveAk

B.3 Romaner i korpusen med berättande texter

- Barnböcker
 - 6–9 år
 - * Lidbeck, Petter (2004), *En dag i prinsessan Victorias liv*.
 - * Lidbeck, Petter (2005), *När Johan vaknar en morgon är han stark*.
 - * Stark, Ulf (1992), *Kan du vissla Johanna*.
 - * Stark, Ulf (2003), *Fullt med flugor i klassen*.
 - * Stark, Ulf (2003), *När jag besökte himlen*.
 - * Stark, Ulf (2005), *Märklin och Turbin*.
 - * Thor, Annika (2002), *Rött hjärta blå fjärl*.
 - * Thor, Annika (2003), *Pirr i magen och klump i halsen*.
 - * Åkerblom, Gull (2005), *Inez och spökhuset*.
 - * Åkerblom, Gull (2005), *Inez värsta vecka*.
 - * Åkerblom, Gull (2006), *Inez långa väg till mormor*.
 - * Åkerblom, Gull (2006), *Trolleri, Inez*.
 - 9–12 år
 - * Hagmar, Pia (2001), *Drömmen om en häst*.
 - * Hagmar, Pia (2003), *Drömponnyn*.
 - * Hagmar, Pia (2003), *Vänner*.
 - * Johansson, Elsie (2004), *Mormorsmysteriet*.
 - * Nilsson, Ulf (2006), *Den döde talar*.
 - * Nilsson, Ulf (2005), *En halv tusenlapp*.
 - * Stark, Ulf (2004), *Min vän Percy, Buffalo Bill och jag*.
 - * Wägner, Veronica (2004), *Kråkös hemlighet*.
 - * Wägner, Veronica (2005), *Tillträde förbjudet*.
- Ungdomsromaner, 12–15 år
 - Lidbeck, Petter (2002), *Bäst i laget*.
 - Lidbeck, Petter (2003), *Vargön*.
 - Thor, Annika (2005), *Nu, imorgon!*
 - Åkerblom, Gull (2003), *Fritt fall*.
 - Åkerblom, Gull (2006), *De andra*.

- Vuxenromaner
 - Norstedtsromaner
 - * Claesson, Stewe (1999), *Stämman i havet: bidrag till allmänna kölldäran*.
 - * Connelly, Michael (1999), *Svart eko*.
 - * Edelfeldt, Inger (1999), *Det hemliga namnet*.
 - * Francis, Clare (1999), *Mörk passion*.
 - * Guterson, David (1999), *Öster om bergen*.
 - * James, Erica (1999), *Kärlek eller förnuft*.
 - * Mosley, Walter (1999), *En liten gul hund*.
 - * Odensten, Per (1999), *En lampa som gör mörker*.
 - * Schulze, Ingo (1999), *Simple stories*.
 - * Wassmo, Herbjørg (1999), *Karnas arv*.
 - * Wolfe, Tom (1999), *En riktig man*.
 - StorSUC-romaner
 - * Adelswärd, Viveka (1991), *Prat skratt skvaller och gräl*.
 - * Alfvén, Inger (1992), *Elefantens öga*.
 - * Backström, Gösta (1991), *Oäktingarna*.
 - * Claesson, Stig (1991), *Rosine*.
 - * Dagsås, Johanna (1993), *Riddaren i mina drömmar*.
 - * Eklöf, Mats (1991), *Pojken som blev Simson*.
 - * Ekman, Kerstin (1993), *Händelser vid vatten*.
 - * Ekström, Margareta (1994), *Olga om Olga*.
 - * Erkelius, Lina (1993), *Matti*.
 - * Erkelius, Per Agne (1991), *Himlavagnen*.
 - * Falk, Göran (1990), *Kopparmåne*.
 - * Fogelström, Per Anders (1993), *Hem, till sist*.
 - * Fröjdh, Maria (1993), *Blåeld*.
 - * Fuchs, Rickard (1993), *Släkt!*
 - * Guillou, Jan (1993), *Den enda segern*.
 - * Gustafsson, Lars (1991), *En kakelsättares eftermiddag*.
 - * Hagman, Per (1991), *Cigarett*.
 - * Hylinger, Claes (1990), *Den stora sammankomsten*.
 - * Hägg, Göran (1993), *Money money*.
 - * Kullenberg, Annette (1993), *Diamanten som log i skymning-
en*.

- * Larsmo, Ola (1993), *Himmel och jord må brinna.*
- * Lillpers, Birgitta (1993), *Medan de ännu hade hästar.*
- * Olin, Elisabeth (1993), *Färdemän.*
- * Rydsjö, Elsi (1993), *Kommer aldrig solen?*
- * Sjögren-Olsson, Margareta (1993), *Nomadliv.*
- * Stenberg, Birgitta (1994), *Arabesk.*
- * Thorvall, Kerstin (1993), *När man skjuter arbetare.*
- * Tornborg, Rita (1991), *Rosalie.*
- * Tunström, Göran (1991), *Det sanna livet.*

C

RENSNINGSGREGLER

Ordklassbeteckningar

MID små skiljetecken som +, =

NN substantiv

PAD parenteser

RG räkneord (grundtal)

VB verb

Meningen rensas bort ifall den:

- innehåller en ensam parentes
- innehåller > 4 RG
- börjar med RG eller a, b, c, d, e eller f
- slutar med RG
- innehåller = eller +
- innehåller en siffra eller bokstav följd av . eller)
- saknar VB
- innehåller > 53 ord.
- innehåller > 33 ord och $\text{antal NN} / \text{antal VB} \geq 2$.
- innehåller $3 \leq$ förekomster av MID, PAD eller RG och $2 \leq$ olika typer av dessa tecken.

Undantag från ovan:

- Följande tecken räknas inte som MID-förekomster: , : ; & %
- Parenteser räknas inte om de innehåller något alfabetiskt ord
- Ett parentespar (vänster- och högerparentes) räknas som 1 tecken
- - och / räknas inte som MID-förekomster om de följs av ett alfabetiskt ord
- Inledande RG som följs av *procent* räknas inte
- Avslutande RG bestående av 4 siffror räknas inte om de föregås av ett månadsnamn eller *år*
- Avslutande RG räknas inte om det föregås av *s./sid./sidan*

D

PSEUDOKOD FÖR INDEXPRINCIPEN

```
def index_principle:
    """ The program presents you with an index list
        on basis of the input corpora and the
        variable chosen for extraction. The corpus of
        investigation (the target corpus) preferably
        belongs to some all-embracing genre, e.g.,
        academic language. This corpus is then
        evaluated against a reference corpus (with
        language from another field, e.g., narrative
        texts), in that the final index list contains
        variables specific to the field of
        investigation. Both the target corpus and the
        reference corpus can be composed of an
        unlimited number of subcorpora. The selection
        criteria applied are: frequency, range,
        dispersion and keyness. """
    parser.add_argument("--target_corpus_folder",
        required=True, help=Name of the folder with
        corpora to be examined. Each line in the corpora
        files should contain: word form, lemgram, POS,
        dependency)
    parser.add_argument("--reference_corpus_folder",
        required=True, help=Reference corpus to define
        keywords against. Each line in the corpora files
        should contain: word form, lemgram, POS,
        dependency)
    parser.add_argument("--parameter", required=True,
        help=Parameter to be examined, e.g., words, np,
        parse_tree)
    parser.add_argument("--occurrences", default='15',
        help=Number that each parameter has to occur in
        all subcorpora per --norm_value. #set to 50 for
        np and parse_tree)
```

```

parser.add_argument("--key", default=1.1, help=
    Keynes threshold)
parser.add_argument("--smoothing", default=50, help=
    Smoothing constant.)
parser.add_argument("--norm_value", default=1e6,
    help=Normalizing value. #set to 1e4 (sentences)
    for np and parse_tree)
    program_handler(target_corpus_folder,
        reference_corpus_folder, parameter,
        occurrences, key, smoothing, norm_value)

def program_handler(target_corpus_folder,
    reference_corpus_folder, parameter, occurrences, key,
    smoothing, norm_value):
    """ The program handler """
    #contains FreqDists with frequencies for the
    subcorpora
    target_fd_list = []
    reference_fd_list = []
    #the whole corpus, used to calculate reduced
    frequency
    whole_target_corpus = []
    whole_reference_corpus = []
    # used differently for different parameters
    target_cfd = nltk.ConditionalFreqDist()
    reference_cfd = nltk.ConditionalFreqDist()
    for corpus in target_corpus_folder:
        read_from_corpus(corpus, parameter,
            target_cfd, target_fd_list,
            whole_target_corpus)
    for corpus in reference_corpus_folder:
        read_from_corpus(corpus, parameter,
            reference_cfd, reference_fd_list,
            whole_reference_corpus)
    index_list = extraction_handler(parameter,
        occurrences, norm_value, key, smoothing,
        target_fd_list, whole_target_corpus,
        reference_fd_list, whole_reference_corpus)

def read_from_corpus(corpus, parameter, cfd, fd_list,
    whole_corpus):
    """ Reads the corpus and saves frequencies of
    variables """
    fd_subcorpus = nltk.FreqDist()
    if parameter == "words":
        for line in target_corpus:

```

```

variable_to_store = lemgram + "|" + POS
# word form distributions
cfd[variable_to_store].inc(word from)
elif parameter == "np":
    for sentence in target_corpus:
        # This function is omitted in
        # this pseudo-code
        nps = find_all_nps(sentence)
        for variable_to_store in nps:
            # clause_element
            # distributions
            cfd[variable_to_store].
                inc(variable_to_store
                    (clause_element))
elif parameter == "parse_tree":
    for sentence in target_corpus:
        # This function is omitted in
        # this pseudo-code
        pts = find_all_pts(sentence)
        for variable_to_store in pts:
            # part of speech of the
            # root-node
            # distributions
            cfd[variable_to_store].
                inc(rot_pos)

corpus.append(variable_to_store)
fd.subcorpus.inc(variable_to_store)
fd_list.append(fd_subcorpus)

def extraction_handler(parameter, occurrences,
norm_value, key, smoothing, target_fd_list,
whole_target_corpus, reference_fd_list,
whole_reference_corpus):
    """ Constructs a list with variables that occur at
        least n times in each subcorpus and that are
        keywords for the target corpus against the
        reference corpus """
    target_rf_dict = {} # dictionary with reduced
        frequencies
    reference_rf_dict = {}
    subtarget_rf_dict = {} # dictionary with reduced
        frequencies for subcorpora
    candidates=set() # for variables with enough range
    keywords = {} # the final word list with keywordness
    # calculation of reduced frequencies for subcorpora
    for fd in target_fd_list:

```

```

for variable in fd.keys():
    frequency = fd[variable]
    chunk_storlek = fd.size() / frequency
    # This function is omitted in this
    # pseudo-code
    chunk_lista = chunks(target_subcorpus,
        chunk_storlek)
    count = 0
    for chunk in chunk_lista:
        if variable in chunk:
            count++
            subtarget_rf_dict[variable] = float(
                count) / len(whole_target_corpus) *
                norm_value + smoothing
    # range handling
    if subtarget_rf_dict[variable] * norm_value >=
        occurrences for all subcorpora:
        candidates.add(variable)
# calculation of reduced frequencies
for candidate in candidates:
    frequency = the sum of all fd[candidate] in
        target_fd_list
    chunk_storlek = len(whole_target_corpus) /
        frequency
    chunk_lista = chunks(whole_target_corpus,
        chunk_storlek)
    count = 0
    for chunk in chunk_lista:
        if candidate in chunk:
            count++
target_rf_dict[candidate] = float(count) / len(
    whole_target_corpus) * norm_value + smoothing
    if candidate in any fd.keys() in
        reference_fd_list():
        frequency = the sum of all fd[
            candidate] in
            reference_fd_list
        chunk_storlek = len(
            whole_reference_corpus) /
            frequency
        chunk_lista = chunks(
            whole_reference_corpus,
            chunk_storlek)
        count = 0
        for chunk in chunk_lista:
            if candidate in chunk:

```

```
        count++
reference_rf_dict[candidate] = float(count)
    / len(whole_reference_corpus) *
    norm_value + smoothing
else:
    reference_rf_dict[candidate] = smoothing
# key word handling
for candidate in target_rf_dict:
    keyness = target_rf_dict[candidate] /
        reference_rf_dict[candidate]
    if keyness >= key:
        keywords[candidate] = keyness
return keywords sorted according to keyness
```


E

SAMLADE RESULTAT

E.1 Delstudie: traditionella textanalysmått

E.1.1 Ytliga textanalysmått

Ämne/Ämnesfack	Stadium	Medel	Median	Std.av.
Naturvetenskap	Högstadiet	11,27	11	5,19
	Gymnasiet	13,56	13	6,57
Matematik	Högstadiet	9,52	8	5,24
	Gymnasiet	10,72	10	6,19
Samhällsvetenskap	Högstadiet	12,37	12	5,82
	Gymnasiet	14,12	13	6,45
Hela läroboksmaterialet	Högstadiet	11,48	11	5,54
	Gymnasiet	13,40	12	6,57

Biologi	Högstadiet	11,65	11	5,28
	Gymnasiet	15,34	14	7,03
Fysik	Högstadiet	11,65	11	5,28
	Gymnasiet	12,94	12	6,45
Kemi	Högstadiet	11,12	10	4,96
	Gymnasiet	12,41	12	5,80
Geografi	Högstadiet	12,76	12	5,82
	Gymnasiet	14,56	14	6,16
Historia	Högstadiet	11,90	11	5,54
	Gymnasiet	13,63	13	6,25
Samhällskunskap	Högstadiet	12,46	12	6,04
	Gymnasiet	14,14	13	6,84

Tabell E.1: Meningslängd i lärobokstexterna

Ämne/Ämnesfack	Stadium	Medel	Median	Std.av.
Naturvetenskap	Högstadiet	5,01	4	2,99
	Gymnasiet	5,31	4	3,36
Matematik	Högstadiet	4,55	4	2,80
	Gymnasiet	4,87	4	3,16
Samhällsvetenskap	Högstadiet	5,19	4	3,13
	Gymnasiet	5,44	5	3,43
Hela läroboksmaterialet	Högstadiet	5,02	4	3,01
	Gymnasiet	5,30	4	3,37

Biologi	Högstadiet	5,02	4	2,90
	Gymnasiet	5,32	5	3,19
Fysik	Högstadiet	4,90	4	2,89
	Gymnasiet	5,07	4	3,21
Kemi	Högstadiet	5,11	4	3,14
	Gymnasiet	5,53	5	3,64
Geografi	Högstadiet	5,27	5	3,21
	Gymnasiet	5,35	5	3,39
Historia	Högstadiet	5,23	4	3,02
	Gymnasiet	5,57	5	3,38
Samhällskunskap	Högstadiet	5,07	4	3,15
	Gymnasiet	5,40	4	3,51

Tabell E.2: Ordlängd i lärobokstexterna

Ämne/Ämnesfack	Stadium	LIX
Naturvetenskap	Högstadiet	35,98
	Gymnasiet	42,06
Matematik	Högstadiet	29,29
	Gymnasiet	35,51
Samhällsvetenskap	Högstadiet	39,03
	Gymnasiet	43,75
Hela läroboksmaterialet	Högstadiet	36,33
	Gymnasiet	41,85

Biologi	Högstadiet	36,47
	Gymnasiet	43,63
Fysik	Högstadiet	34,40
	Gymnasiet	39,17
Kemi	Högstadiet	37,07
	Gymnasiet	43,39
Geografi	Högstadiet	41,01
	Gymnasiet	43,63
Historia	Högstadiet	38,64
	Gymnasiet	44,61
Samhällskunskap	Högstadiet	37,43
	Gymnasiet	43,01

Tabell E.3: LIX i lärobokstexterna

Ämne/Ämnesfack	Stadium	OVIX
Naturvetenskap	Högstadiet	62,29
	Gymnasiet	69,19
Matematik	Högstadiet	59,53
	Gymnasiet	63,35
Samhällsvetenskap	Högstadiet	65,92
	Gymnasiet	69,69
Hela läroboksmaterialet	Högstadiet	63,45
	Gymnasiet	68,57

Biologi	Högstadiet	64,69
	Gymnasiet	70,77
Fysik	Högstadiet	61,18
	Gymnasiet	68,55
Kemi	Högstadiet	61,01
	Gymnasiet	68,23
Geografi	Högstadiet	67,52
	Gymnasiet	68,59
Historia	Högstadiet	66,76
	Gymnasiet	72,48
Samhällskunskap	Högstadiet	63,47
	Gymnasiet	68,00

Tabell E.4: OVIX i lärobokstexterna

E.1.2 Djupare textanalysmått

Ordklass	NV	MA	SV	Hela
adjektiv	6,34	5,40	6,90	6,49
adverb	6,17	4,31	5,92	5,79
bestämningsord	4,55	4,95	4,69	4,67
egennamn	0,73	1,91	1,54	1,25
frågande/relativt adverb	1,19	2,36	0,93	1,24
frågande/relativt adverb	0,15	0,39	0,16	0,19
frågande/relativt possessivuttryck	0,00	0,01	0,00	0,00
frågande/relativt pronomen	1,74	1,40	1,75	1,70
grundtal	1,17	4,96	1,28	1,74
infinitivmärke	0,94	0,77	1,10	0,98
interjektion	0,02	0,06	0,04	0,03
konjunktion	4,01	3,01	4,66	4,14
litet skiljetecken	2,17	1,51	2,44	2,19
ordningstal	0,09	0,21	0,15	0,14
parentes	0,72	0,69	0,44	0,60
particip	1,02	0,59	0,97	0,94
partikel	1,33	1,18	1,08	1,19
possessivuttryck	0,46	0,54	0,63	0,54
preposition	9,79	9,13	10,73	10,06
pronomen	5,70	5,12	4,89	5,27
stort skiljetecken	7,96	9,41	7,28	7,89
subjunktion	1,61	1,25	1,26	1,40
substantiv	24,95	24,19	24,57	24,7
verb	17,03	16,59	16,38	16,68

Tabell E.5: Översiktlig ordklassfördelning (i procent av totalt antal token) i lärobokstexterna för högstadiet i naturvetenskap, matematik, samhällsvetenskap och hela läroboksmaterialet

Ordklass	BI	FY	KE
adjektiv	6,46	5,97	6,55
adverb	6,05	6,57	5,92
bestämningsord	4,14	5,18	4,32
egennamn	0,58	0,86	0,73
frågande/relativt adverb	1,08	1,37	1,11
frågande/relativt adverb	0,13	0,15	0,16
frågande/relativt possessivuttryck	0,00	0,01	0,00
frågande/relativt pronomen	1,87	1,62	1,72
grundtal	0,99	1,49	1,03
infinitivmärke	1,00	0,90	0,91
interjektion	0,01	0,02	0,04
konjunktion	4,28	3,40	4,34
litet skiljetecken	2,05	2,26	2,21
ordningstal	0,11	0,12	0,05
parentes	0,73	0,57	0,88
particip	1,00	0,95	1,09
partikel	1,31	1,33	1,37
possessivuttryck	0,66	0,39	0,33
preposition	10,16	9,55	9,61
pronomen	5,38	6,28	5,46
stort skiljetecken	7,68	8,05	8,15
subjunktion	1,48	1,83	1,50
substantiv	25,48	23,92	25,51
verb	17,21	17,10	16,86

Tabell E.6: Ämnesspecifik ordklassfördelning (i procent av totalt antal token) i naturvetenskapliga ämnen på högstadiet: biologi, fysik och kemi

Ordklass	GE	HI	SK
adjektiv	7,53	6,81	6,40
adverb	5,65	5,92	6,15
bestämningsord	4,50	4,83	4,76
egennamn	1,37	2,56	0,71
frågande/relativt adverb	0,92	0,89	0,97
frågande/relativt adverb	0,20	0,11	0,18
frågande/relativt possessivuttryck	0,00	0,01	0,00
frågande/relativt pronomen	1,70	1,55	1,99
grundtal	1,31	1,40	1,14
infinitivmärke	0,92	0,99	1,40
interjektion	0,01	0,03	0,07
konjunktion	4,92	4,65	4,41
litet skiljetecken	3,25	1,93	2,20
ordningstal	0,09	0,27	0,13
parentes	0,52	0,49	0,31
particip	1,02	1,05	0,86
partikel	0,90	1,19	1,13
possessivuttryck	0,46	0,67	0,74
preposition	11,59	10,82	9,77
pronomen	3,67	4,73	6,28
stort skiljetecken	6,99	7,59	7,25
subjunktion	0,97	1,14	1,69
substantiv	26,88	23,92	22,87
verb	14,43	16,35	18,37

Tabell E.7: Ämnesspecifik ordklassfördelning (i procent av totalt antal token) i samhällsvetenskapliga ämnen på högstadiet: geografi, historia och samhällskunskap

Ordklass	NV	MA	SV	Hela
adjektiv	6,15	5,43	7,04	6,44
adverb	5,79	3,47	5,61	5,39
bestämningsord	5,12	5,47	4,78	5,04
egennamn	1,64	2,23	3,20	2,40
frågande/relativt adverb	1,02	1,37	0,90	1,02
frågande/relativt adverb	0,13	0,25	0,16	0,16
frågande/relativt possessivuttryck	0,02	0,02	0,01	0,01
frågande/relativt pronomen	1,65	1,10	1,54	1,54
grundtal	1,86	5,04	1,23	2,06
infinitivmärke	0,93	0,72	1,18	1,00
interjektion	0,03	0,03	0,02	0,02
konjunktion	3,88	3,30	4,73	4,14
litet skiljetecken	2,46	2,62	2,94	2,67
ordningstal	0,10	0,19	0,18	0,15
parentes	1,31	1,48	0,82	1,11
particip	1,35	0,99	1,18	1,23
partikel	0,96	0,60	0,98	0,92
possessivuttryck	0,44	0,53	0,63	0,53
preposition	10,86	9,90	11,40	10,89
pronomen	4,39	3,74	4,06	4,16
stort skiljetecken	6,71	8,20	6,44	6,82
subjunktion	1,67	1,28	1,38	1,49
substantiv	25,82	26,14	24,17	25,16
verb	15,64	15,66	15,15	15,44

Tabell E.8: Översiktlig ordklassfördelning (i procent av totalt antal token) i lärobokstexterna för gymnasiet i naturvetenskap, matematik, samhällsvetenskap och hela läroboksmaterialet

Ordklass	BI	FY	KE
adjektiv	6,56	5,75	6,14
adverb	6,41	5,56	5,40
bestämningsord	4,68	5,41	5,27
egennamn	1,20	2,01	1,73
frågande/relativt adverb	0,88	1,34	0,85
frågande/relativt adverb	0,08	0,21	0,11
frågande/relativt possessivuttryck	0,02	0,02	0,03
frågande/relativt pronomen	2,06	1,35	1,56
grundtal	0,85	2,80	1,92
infinitivmärke	1,13	0,96	0,68
interjektion	0,02	0,02	0,04
konjunktion	4,37	3,45	3,77
litet skiljetecken	2,20	2,32	2,83
ordningstal	0,09	0,12	0,08
parentes	1,55	0,74	1,64
particip	1,41	1,19	1,46
partikel	1,05	1,07	0,79
possessivuttryck	0,53	0,47	0,30
preposition	11,23	10,51	10,81
pronomen	4,69	4,87	3,56
stort skiljetecken	5,88	7,01	7,28
subjunktion	1,78	1,79	1,45
substantiv	25,14	25,16	27,22
verb	16,14	15,79	15,01

Tabell E.9: Ämnesspecifik ordklassfördelning (i procent av totalt antal token) i naturvetenskapliga ämnen på gymnasiet: biologi, fysik och kemi

Ordklass	GE	HI	SK
adjektiv	7,54	7,03	6,58
adverb	5,90	5,38	5,54
bestämningsord	4,53	5,13	4,73
egennamn	2,63	4,61	2,41
frågande/relativt adverb	0,77	1,03	0,91
frågande/relativt adverb	0,10	0,16	0,21
frågande/relativt possessivuttryck	0,00	0,01	0,01
frågande/relativt pronomen	1,21	1,57	1,86
grundtal	1,18	1,41	1,13
infinitivmärke	1,08	1,03	1,39
interjektion	0,01	0,01	0,03
konjunktion	4,90	4,58	4,73
litet skiljetecken	3,12	2,78	2,93
ordningstal	0,09	0,32	0,13
parentes	0,96	0,60	0,86
particip	1,17	1,24	1,12
partikel	0,82	1,14	0,97
possessivuttryck	0,57	0,69	0,65
preposition	11,98	11,86	10,29
pronomen	3,39	3,82	4,98
stort skiljetecken	6,35	6,62	6,38
subjunktion	1,13	1,17	1,86
substantiv	25,98	22,90	23,40
verb	13,82	14,85	16,83

Tabell E.10: Ämnesspecifik ordklassfördelning (i procent av totalt antal token) i samhällsvetenskapliga ämnen på gymnasiet: geografi, historia och samhällskunskap

Ordklass	Berättande	SveAk
adjektiv	4,39	6,77
adverb	8,68	6,00
bestämningsord	3,26	5,34
egennamn	3,17	2,97
frågande/relativt adverb	0,65	0,83
frågande/relativt adverb	0,02	0,09
frågande/relativt possessivuttryck	0,01	0,01
frågande/relativt pronomen	1,23	1,81
grundtal	0,28	1,43
infinitivmärke	0,93	1,38
interjektion	0,36	0,04
konjunktion	4,24	4,76
litet skiljetecken	4,59	3,28
ordningstal	0,08	0,16
parentes	0,57	1,91
particip	0,77	1,72
partikel	2,16	0,61
possessivuttryck	1,17	0,78
preposition	8,09	11,13
pronomen	12,67	4,73
stort skiljetecken	8,38	4,95
subjunktion	1,90	1,77
substantiv	12,63	21,74
verb	19,75	14,88

Tabell E.11: Ordklassfördelning (i procent av totalt antal token) i referenskorpusarna med berättande respektive akademiska texter

Ämne/Ämnesfack	Stadium	NQ
Naturvetenskap	Högstadiet	1,24
	Gymnasiet	1,48
Matematik	Högstadiet	1,30
	Gymnasiet	1,62
Samhällsvetenskap	Högstadiet	1,36
	Gymnasiet	1,49
Hela läroboksmaterialet	Högstadiet	1,30
	Gymnasiet	1,50

Biologi	Högstadiet	1,28
	Gymnasiet	1,39
Fysik	Högstadiet	1,15
	Gymnasiet	1,41
Kemi	Högstadiet	1,28
	Gymnasiet	1,65
Geografi	Högstadiet	1,66
	Gymnasiet	1,69
Historia	Högstadiet	1,33
	Gymnasiet	1,50
Samhällskunskap	Högstadiet	1,09
	Gymnasiet	1,27

Tabell E.12: NQ i lärobokstexterna

Ämne/Ämnesfack	Stadium	LVIX
Naturvetenskap	Högstadiet	50,97
	Gymnasiet	51,60
Matematik	Högstadiet	48,14
	Gymnasiet	48,01
Samhällsvetenskap	Högstadiet	51,09
	Gymnasiet	52,82
Hela läroboksmaterialet	Högstadiet	50,62
	Gymnasiet	51,61

Biologi	Högstadiet	52,37
	Gymnasiet	53,13
Fysik	Högstadiet	50,61
	Gymnasiet	51,02
Kemi	Högstadiet	49,94
	Gymnasiet	50,64
Geografi	Högstadiet	50,74
	Gymnasiet	51,37
Historia	Högstadiet	52,07
	Gymnasiet	54,08
Samhällskunskap	Högstadiet	50,46
	Gymnasiet	53,01

Tabell E.13: LVIX i lärobokstexterna

E.2 Delstudie: typiska ord

E.2.1 Naturvetenskapliga texter jämförda med matematiktexter

E.2.1.1 Högstadiet

	Naturvetenskap		Matematik	
	UNIKT	Index	UNIKT	Index
1	<i>ämne</i> NN	23,04	<i>area</i> NN	33,71
2	<i>växt</i> NN	17,50	<i>tal</i> NN	33,12
3	<i>energi</i> NN	13,44	<i>avrunda</i> VB	30,57
4	<i>atom</i> NN	12,22	<i>triangel</i> NN	24,19
5	<i>kemisk</i> JJ	11,91	<i>uttryck</i> NN	18,59
6	<i>syre</i> NN	11,26	<i>kosta</i> VB	18,31
7	<i>cell</i> NN	8,92	<i>värde</i> NN	15,74
8	<i>koldioxid</i> NN	8,48	<i>omkrets</i> NN	15,54
9	<i>molekyl</i> NN	8,42	<i>vinkel</i> NN	14,53
10	<i>metall</i> NN	8,16	<i>ekvation</i> NN	14,51
11	<i>luft</i> NN	7,82	<i>pris</i> NN	14,49
12	<i>egenskap</i> NN	7,52	<i>kvadrat</i> NN	14,46
13	<i>stark</i> JJ	7,36	<i>betala</i> VB	14,38
14	<i>vätska</i> NN	7,28	<i>rektangel</i> NN	14,28
15	<i>djur</i> NN	7,13	<i>bråk</i> NN	13,81
16	<i>gas</i> NN	7,01	<i>procent</i> NN	13,11
17	<i>ström</i> NN	6,54	<i>svara</i> VB	13,07
18	<i>natur</i> NN	6,53	<i>teckna</i> VB	12,63
19	<i>blod</i> NN	6,51	<i>räkna ut</i> VB	11,90
20	<i>sprida</i> VB	6,51	<i>figur</i> NN	11,38
21	<i>bakterie</i> NN	6,31	<i>heltal</i> NN	10,16
22	<i>elektrisk</i> JJ	6,26	<i>sträcka</i> NN	10,13
23	<i>kropp</i> NN	6,24	<i>svar</i> NN	9,12
24	<i>bränsle</i> NN	6,17	<i>andel</i> NN	9,08
25	<i>liv</i> NN	6,05	<i>nämnare</i> NN	8,87
26	<i>fungera</i> VB	6,03	<i>sannolikhet</i> NN	8,58
27	<i>värme</i> NN	6,00	<i>räkna</i> VB	8,52
28	<i>människa</i> NN	5,83	<i>diagram</i> NN	8,50
29	<i>järn</i> NN	5,63	<i>multiplitera</i> VB	8,39
30	<i>utveckla</i> VB	5,54	<i>graf</i> NN	8,19
31	<i>forskare</i> NN	5,49	<i>dividera</i> VB	7,82
32	<i>vatten</i> NN	5,41	<i>elev</i> NN	7,70

Naturvetenskap			Matematik	
	UNIKT	Index	UNIKT	Index
33	<i>muskel</i> NN	5,38	<i>parentes</i> NN	7,30
34	<i>kol</i> NN	5,33	<i>beräkna</i> VB	7,29
35	<i>hud</i> NN	5,27	<i>linje</i> NN	7,10
36	<i>orsaka</i> VB	5,17	<i>krona</i> NN	6,90
37	<i>reaktion</i> NN	5,13	<i>köpa</i> VB	6,86
38	<i>varm</i> JJ	5,02	<i>vilket</i> HP	6,82
39	<i>skydda</i> VB	4,89	<i>cirkel</i> NN	6,80
40	<i>sjukdom</i> NN	4,71	<i>höjd</i> NN	6,69
41	<i>atmosfär</i> NN	4,68	<i>addera</i> VB	6,61
42	<i>hjärna</i> NN	4,65	<i>längd</i> NN	6,55
43	<i>utvecklas</i> VB	4,64	<i>rymma</i> VB	6,44
44	<i>mark</i> NN	4,63	<i>diameter</i> NN	6,41
45	<i>mat</i> NN	4,60	<i>sammanlagd</i> AB	6,35
46	<i>miljö</i> NN	4,58	<i>multiplikation</i> NN	6,22
47	<i>bilda</i> VB	4,57	<i>decimal</i> NN	6,17
48	<i>skal</i> NN	4,48	<i>volym</i> NN	6,16
49	<i>kraft</i> NN	4,48	<i>sida</i> NN	6,12
50	<i>hård</i> JJ	4,47	<i>tärning</i> NN	6,08
51	<i>uppstå</i> VB	4,45	<i>hon</i> PN	6,07
52	<i>nästan</i> AB	4,37	<i>per</i> PP	5,98
53	<i>tillverka</i> VB	4,34	<i>ränta</i> NN	5,93
54	<i>organism</i> NN	4,29	<i>pengar</i> NN	5,87
55	<i>hända</i> VB	4,28	<i>peng</i> NN	5,87
56	<i>reagera</i> VB	4,27	<i>variabel</i> NN	5,78
57	<i>partikel</i> NN	4,27	<i>täljare</i> NN	5,58
58	<i>viktig</i> JJ	4,27	<i>timme</i> NN	5,48
59	<i>kärna</i> NN	4,22	<i>hur</i> HD	5,44
60	<i>jord</i> NN	4,21	<i>division</i> NN	5,34
61	<i>vår</i> PS	4,21	<i>skriva</i> VB	5,33
62	<i>själva</i> JJ	4,14	<i>rät</i> JJ	5,32
63	<i>skada</i> VB	4,12	<i>summa</i> NN	5,26
64	<i>leva</i> VB	4,08	<i>bråkform</i> NN	5,18
65	<i>förbränning</i> NN	4,01	<i>förenkla</i> VB	5,18
66	<i>därför</i> AB	3,96	<i>mått</i> NN	5,14
67	<i>hav</i> NN	3,95	<i>kort</i> NN	5,08
68	<i>ske</i> VB	3,93	<i>vinna</i> VB	5,04
69	<i>kemi</i> NN	3,93	<i>total</i> NN	5,03
70	<i>rörelse</i> NN	3,89	<i>tabell</i> NN	4,98
71	<i>bryta</i> VB	3,88	<i>medelhastighet</i> NN	4,95

Naturvetenskap			Matematik		
	UNIKT	Index		UNIKT	Index
72	<i>tunn</i> JJ	3,78	<i>medelvärde</i> NN		4,91
73	<i>svamp</i> NN	3,78	<i>liter</i> NN		4,85
74	<i>lätt</i> AB	3,69	<i>rätblock</i> NN		4,82
75	<i>andas</i> VB	3,68	<i>parallelogram</i> NN		4,77
76	<i>typ</i> NN	3,68	<i>jag</i> PN		4,73
77	<i>levande</i> PC	3,67	<i>tiondel</i> NN		4,73
78	<i>plast</i> NN	3,66	<i>ni</i> PN		4,71
79	<i>väte</i> NN	3,65	<i>cylinder</i> NN		4,71
80	<i>kretslopp</i> NN	3,63	<i>rätt</i> NN		4,63
81	<i>experiment</i> NN	3,62	<i>halv</i> JJ		4,53
82	<i>lagra</i> VB	3,62	<i>styck</i> NN		4,47
83	<i>fotosyntes</i> NN	3,60	<i>siffra</i> NN		4,46
84	<i>värma</i> VB	3,59	<i>subtraktion</i> NN		4,45
85	<i>ljus</i> NN	3,58	<i>addition</i> NN		4,45
86	<i>försvinna</i> VB	3,56	<i>kub</i> NN		4,44
87	<i>tryck</i> NN	3,55	<i>tiondels</i> JJ		4,37
88	<i>farlig</i> JJ	3,55	<i>subtrahera</i> VB		4,37
89	<i>utnyttja</i> VB	3,52	<i>median</i> NN		4,37
90	<i>smälta</i> VB	3,49	<i>rabatt</i> NN		4,36
91	<i>ändå</i> AB	3,49	<i>mycket</i> PN		4,34
92	<i>strålning</i> NN	3,48	<i>alternativ</i> NN		4,33
93	<i>påverka</i> VB	3,48	<i>åka</i> VB		4,32
94	<i>öga</i> NN	3,47	<i>hastighet</i> NN		4,31
95	<i>olja</i> NN	3,46	<i>y-axel</i> NN		4,29
96	<i>sig</i> PN	3,46	<i>grundpotensform</i> NN		4,28
97	<i>släppa</i> VB	3,39	<i>motsvara</i> VB		4,25
98	<i>eld</i> NN	3,37	<i>minut</i> NN		4,23
99	<i>leda</i> VB	3,29	<i>lön</i> NN		4,21
100	<i>salt</i> NN	3,28	<i>väga</i> VB		4,20

Tabell E.14: Indexlista över ord i läroböcker i naturvetenskap/matematik för högstadiet, topp 100

E.2.1.2 *Gymnasiet*

	Naturvetenskap		Matematik	
	UNIKT	Index	UNIKT	Index
1	<i>molekyl</i> NN	19,38	<i>triangel</i> NN	22,07
2	<i>reaktion</i> NN	16,10	<i>ekvation</i> NN	20,27
3	<i>cell</i> NN	12,20	<i>tal</i> NN	20,07
4	<i>elektron</i> NN	11,80	<i>räknare</i> NN	18,37
5	<i>jon</i> NN	10,08	<i>vinkel</i> NN	16,62
6	<i>reagera</i> VB	9,92	<i>motivera</i> VB	11,78
7	<i>protein</i> NN	9,61	<i>graf</i> NN	11,23
8	<i>atom</i> NN	9,48	<i>area</i> NN	10,56
9	<i>kemisk</i> JJ	8,82	<i>sann</i> JJ	10,55
10	<i>binda</i> VB	8,68	<i>multiplitera</i> VB	9,90
11	<i>energi</i> NN	8,62	<i>värde</i> NN	9,65
12	<i>ämne</i> NN	8,08	<i>variabel</i> NN	9,47
13	<i>bilda</i> VB	8,07	<i>funktion</i> NN	9,45
14	<i>luft</i> NN	7,41	<i>pris</i> NN	9,43
15	<i>växt</i> NN	7,38	<i>betala</i> VB	9,29
16	<i>gas</i> NN	7,11	<i>uttryck</i> NN	9,06
17	<i>vatten</i> NN	7,09	<i>kvadrat</i> NN	8,97
18	<i>syre</i> NN	6,89	<i>procent</i> NN	8,79
19	<i>ske</i> VB	6,88	<i>nämnare</i> NN	8,57
20	<i>metall</i> NN	6,73	<i>kosta</i> VB	8,54
21	<i>partikel</i> NN	6,44	<i>linje</i> NN	8,24
22	<i>bindning</i> NN	6,37	<i>decimal</i> NN	8,19
23	<i>vätska</i> NN	5,91	<i>sannolikhet</i> NN	8,14
24	<i>laddning</i> NN	5,70	<i>heltal</i> NN	7,85
25	<i>stark</i> JJ	5,60	<i>kostnad</i> NN	7,70
26	<i>avge</i> VB	5,53	<i>led</i> NN	7,69
27	<i>miljö</i> NN	5,29	<i>ränta</i> NN	7,63
28	<i>väteatom</i> NN	5,15	<i>bråk</i> NN	7,37
29	<i>koncentration</i> NN	5,01	<i>rektangel</i> NN	7,34
30	<i>grundämne</i> NN	4,96	<i>derivata</i> NN	7,33
31	<i>påverka</i> VB	4,94	<i>beräkning</i> NN	7,30
32	<i>massa</i> NN	4,92	<i>beräkna</i> VB	7,16
33	<i>koldioxid</i> NN	4,75	<i>rät</i> JJ	7,05
34	<i>uppstå</i> VB	4,69	<i>term</i> NN	6,92
35	<i>kraft</i> NN	4,60	<i>lån</i> NN	6,91
36	<i>elektrisk</i> JJ	4,52	<i>timme</i> NN	6,87

Naturvetenskap			Matematik		
	UNIKT	Index		UNIKT	Index
37	<i>egenskap</i> NN	4,34		<i>sträcka</i> NN	6,80
38	<i>DNA-molekyl</i> NN	4,34		<i>svar</i> NN	6,69
39	<i>frigöra</i> VB	4,34		<i>köpa</i> VB	6,68
40	<i>kropp</i> NN	4,32		<i>procentuell</i> JJ	6,60
41	<i>ta upp</i> VB	4,32		<i>förenkla</i> VB	6,53
42	<i>laddad</i> PC	4,24		<i>din</i> PS	6,44
43	<i>låg</i> JJ	4,22		<i>parentes</i> NN	6,32
44	<i>blandning</i> NN	4,17		<i>reell</i> JJ	6,32
45	<i>process</i> NN	4,17		<i>längd</i> NN	6,24
46	<i>lätt</i> AB	4,17		<i>månad</i> NN	6,24
47	<i>liv</i> NN	4,16		<i>sats</i> NN	6,20
48	<i>röra</i> VB	4,14		<i>dividera</i> VB	6,20
49	<i>trots</i> PP	4,06		<i>summa</i> NN	6,19
50	<i>väte</i> NN	4,06		<i>kurva</i> NN	6,17
51	<i>dessutom</i> AB	4,02		<i>vinst</i> NN	6,15
52	<i>proton</i> NN	3,97		<i>hon</i> PN	6,00
53	<i>därmed</i> AB	3,95		<i>närmevärde</i> NN	5,89
54	<i>däremot</i> AB	3,91		<i>skriva</i> VB	5,87
55	<i>värme</i> NN	3,91		<i>punkt</i> NN	5,87
56	<i>järn</i> NN	3,90		<i>medelvärde</i> NN	5,81
57	<i>framställa</i> VB	3,85		<i>ökning</i> NN	5,62
58	<i>bunden</i> PC	3,78		<i>vektor</i> NN	5,59
59	<i>fri</i> JJ	3,76		<i>höjd</i> NN	5,58
60	<i>människa</i> NN	3,74		<i>matematik</i> NN	5,53
61	<i>därför</i> AB	3,72		<i>omkrets</i> NN	5,53
62	<i>svag</i> JJ	3,68		<i>tärning</i> NN	5,48
63	<i>röra sig</i> VB	3,63		<i>tolka</i> VB	5,46
64	<i>kol</i> NN	3,54		<i>belopp</i> NN	5,46
65	<i>forskare</i> NN	3,51		<i>falsk</i> JJ	5,36
66	<i>omgivning</i> NN	3,50		<i>matematisk</i> JJ	5,33
67	<i>hud</i> NN	3,47		<i>elev</i> NN	5,28
68	<i>transportera</i> VB	3,46		<i>median</i> NN	5,28
69	<i>enda</i> JJ	3,46		<i>påstående</i> NN	5,23
70	<i>ljus</i> NN	3,41		<i>svara</i> VB	5,20
71	<i>natur</i> NN	3,37		<i>rätvinklig</i> JJ	5,17
72	<i>experiment</i> NN	3,33		<i>sida</i> NN	5,12
73	<i>hos</i> PP	3,32		<i>likformig</i> JJ	5,07
74	<i>ju</i> AB	3,25		<i>koordinat</i> NN	5,07
75	<i>alltså</i> AB	3,22		<i>cirkel</i> NN	4,98

Naturvetenskap			Matematik	
	UNIKT	Index	UNIKT	Index
76	<i>våglängd</i> NN	3,21	<i>polynom</i> NN	4,97
77	<i>fas</i> NN	3,20	<i>exponent</i> NN	4,88
78	<i>tryck</i> NN	3,19	<i>utfall</i> NN	4,87
79	<i>kärna</i> NN	3,18	<i>sälja</i> VB	4,85
80	<i>lagra</i> VB	3,16	<i>redovisa</i> VB	4,85
81	<i>absorbera</i> VB	3,07	<i>lös</i> JJ	4,83
82	<i>sitta</i> VB	3,03	<i>multiplikation</i> NN	4,76
83	<i>förmåga</i> NN	3,02	<i>kontrollera</i> VB	4,73
84	<i>utnyttja</i> VB	3,01	<i>per</i> PP	4,63
85	<i>ställ</i> NN	3,00	<i>siffra</i> NN	4,60
86	<i>fungera</i> VB	2,99	<i>diagram</i> NN	4,60
87	<i>djur</i> NN	2,95	<i>term</i> NN	4,57
88	<i>bränsle</i> NN	2,95	<i>lutning</i> NN	4,57
89	<i>hända</i> VB	2,94	<i>bank</i> NN	4,56
90	<i>bestå</i> VB	2,94	<i>subtrahera</i> VB	4,56
91	<i>skikt</i> NN	2,92	<i>minut</i> NN	4,52
92	<i>tunn</i> JJ	2,92	<i>täljare</i> NN	4,52
93	<i>strålning</i> NN	2,90	<i>avrunda</i> VB	4,50
94	<i>styra</i> VB	2,88	<i>krona</i> NN	4,50
95	<i>struktur</i> NN	2,86	<i>olikhet</i> NN	4,44
96	<i>energirik</i> JJ	2,83	<i>bevisa</i> VB	4,38
97	<i>kväve</i> NN	2,80	<i>rätt</i> NN	4,36
98	<i>membran</i> NN	2,80	<i>bevis</i> NN	4,34
99	<i>skydda</i> VB	2,80	<i>matematiker</i> NN	4,28
100	<i>rörelse</i> NN	2,80	<i>nollställe</i> NN	4,24

Tabell E.15: Indexlista över ord i läroböcker i naturvetenskap/matematik för gymnasiet, topp 100

E.2.2 Naturvetenskapliga texter jämförda med samhällsvetenskapliga texter

E.2.2.1 Högstadiet

Naturvetenskap			Samhällsvetenskap		
	UNIKT	Index		UNIKT	Index
1	ämne NN	14,85	företag NN		11,16
2	atom NN	13,50	regering NN		8,91
3	cell NN	12,00	riksdag NN		8,56
4	kropp NN	11,14	krig NN		7,78
5	syre NN	10,86	parti NN		7,41
6	bakterie NN	10,33	makt NN		7,31
7	blod NN	9,17	stad NN		7,23
8	molekyl NN	8,42	land NN		7,12
9	kemisk JJ	7,99	vara NN		6,83
10	vätska NN	7,48	bo VB		6,50
11	växt NN	6,40	stat NN		6,47
12	inhålla VB	6,23	Sverige NN		6,45
13	energi NN	6,01	pengar NN		6,27
14	egenskap NN	5,80	karta NN		6,26
15	hjärna NN	5,66	politisk JJ		6,16
16	metall NN	5,63	handel NN		5,62
17	ljus NN	5,62	invånare NN		5,60
18	koldioxid NN	5,43	folk NN		5,60
19	hud NN	5,42	befolkning NN		5,43
20	muskel NN	5,28	betala VB		5,37
21	vatten NN	5,25	lön NN		5,27
22	luft NN	5,17	svensk JJ		5,11
23	salt NN	4,92	område NN		5,08
24	reaktion NN	4,89	Danmark NN		4,82
25	elektrisk JJ	4,81	samhälle NN		4,77
26	öga NN	4,79	skatt NN		4,69
27	ägg NN	4,68	fattig JJ		4,66
28	skal NN	4,61	nord NN		4,60
29	ström NN	4,57	politiker NN		4,58
30	föremål NN	4,38	Finland NN		4,58
31	omvandla VB	4,29	demokrati NN		4,41
32	djur NN	4,27	jordbruk NN		4,14
33	partikel NN	4,06	inkomst NN		4,14
34	kärna NN	4,02	ekonomisk JJ		4,12

Naturvetenskap			Samhällsvetenskap		
	UNIKT	Index		UNIKT	Index
35	<i>gas</i> NN	3,97		<i>beslut</i> NN	4,10
36	<i>temperatur</i> NN	3,93		<i>arbetare</i> NN	4,10
37	<i>material</i> NN	3,85		<i>region</i> NN	4,07
38	<i>kemi</i> NN	3,85		<i>bonde</i> NN	4,04
39	<i>reagera</i> VB	3,84		<i>skola</i> NN	4,00
40	<i>glas</i> NN	3,83		<i>industri</i> NN	3,99
41	<i>sort</i> NN	3,69		<i>ekonomi</i> NN	3,97
42	<i>brinna</i> VB	3,61		<i>gård</i> NN	3,96
43	<i>förbränning</i> NN	3,60		<i>val</i> NN	3,95
44	<i>andas</i> VB	3,54		<i>Frankrike</i> NN	3,86
45	<i>värme</i> NN	3,51		<i>medlem</i> NN	3,76
46	<i>lätt</i> AB	3,50		<i>procent</i> NN	3,72
47	<i>skelett</i> NN	3,46		<i>Norge</i> NN	3,72
48	<i>plast</i> NN	3,43		<i>Storbritannien</i> NN	3,67
49	<i>tunn</i> JJ	3,42		<i>verksamhet</i> NN	3,61
50	<i>organism</i> NN	3,42		<i>huvudstad</i> NN	3,57
51	<i>experiment</i> NN	3,40		<i>nordisk</i> JJ	3,54
52	<i>svamp</i> NN	3,39		<i>värld</i> NN	3,54
53	<i>tryck</i> NN	3,38		<i>jobb</i> NN	3,52
54	<i>insekt</i> NN	3,38		<i>politik</i> NN	3,52
55	<i>strålning</i> NN	3,36		<i>europaisk</i> JJ	3,51
56	<i>ihop</i> PL	3,35		<i>arbete</i> VB	3,48
57	<i>atmosfär</i> NN	3,33		<i>Sveriges</i> NN	3,43
58	<i>äta</i> VB	3,32		<i>konflikt</i> NN	3,43
59	<i>blandning</i> NN	3,32		<i>USA</i> NN	3,37
60	<i>väte</i> NN	3,31		<i>norr</i> NN	3,36
61	<i>hjärta</i> NN	3,30		<i>välja</i> VB	3,35
62	<i>väga</i> VB	3,11		<i>anställda</i> NN	3,32
63	<i>signal</i> NN	3,10		<i>same</i> NN	3,26
64	<i>bilda</i> VB	3,10		<i>Europa</i> NN	3,24
65	<i>kväve</i> NN	3,04		<i>storstad</i> NN	3,21
66	<i>planet</i> NN	3,01		<i>besluta</i> VB	3,20
67	<i>bränsle</i> NN	3,00		<i>förslag</i> NN	3,20
68	<i>fotosyntes</i> NN	2,99		<i>tysk</i> JJ	3,18
69	<i>färg</i> NN	2,97		<i>man</i> NN	3,18
70	<i>vattenånga</i> NN	2,94		<i>norra</i> JJ	3,12
71	<i>lunga</i> NN	2,93		<i>rik</i> JJ	3,09
72	<i>eld</i> NN	2,91		<i>amerikansk</i> JJ	3,08
73	<i>lagra</i> VB	2,89		<i>söder</i> NN	3,07

Naturvetenskap			Samhällsvetenskap		
	UNIKT	Index		UNIKT	Index
74	<i>suga</i> VB	2,86		<i>rike</i> NN	3,07
75	<i>bestå</i> VB	2,86		<i>kust</i> NN	3,02
76	<i>levande</i> PC	2,83		<i>sälja</i> VB	3,00
77	<i>släppa</i> VB	2,81		<i>text</i> NN	2,99
78	<i>tråd</i> NN	2,79		<i>stöd</i> NN	2,95
79	<i>orsaka</i> VB	2,77		<i>jag</i> PN	2,94
80	<i>röra sig</i> VB	2,77		<i>diskutera</i> VB	2,93
81	<i>fågel</i> NN	2,76		<i>medborgare</i> NN	2,93
82	<i>stjärna</i> NN	2,75		<i>språk</i> NN	2,93
83	<i>testa</i> VB	2,75		<i>köpa</i> VB	2,92
84	<i>ta upp</i> VB	2,74		<i>kultur</i> NN	2,92
85	<i>fast</i> JJ	2,72		<i>mål</i> NN	2,90
86	<i>koppla</i> VB	2,71		<i>utbildning</i> NN	2,90
87	<i>energirik</i> JJ	2,70		<i>fred</i> NN	2,86
88	<i>röd</i> JJ	2,68		<i>vilja</i> VB	2,86
89	<i>uppbyggd</i> PC	2,68		<i>tvinga</i> VB	2,85
90	<i>skada</i> VB	2,64		<i>arbetskraft</i> NN	2,82
91	<i>ben</i> NN	2,63		<i>rättighet</i> NN	2,82
92	<i>inuti</i> PP	2,63		<i>kvinn</i> NN	2,81
93	<i>fånga</i> VB	2,61		<i>fabrik</i> NN	2,81
94	<i>röra</i> VB	2,61		<i>möjlighet</i> NN	2,80
95	<i>blanda</i> VB	2,60		<i>internet</i> NN	2,80
96	<i>effekt</i> NN	2,59		<i>svensk</i> NN	2,80
97	<i>faktisk</i> AB	2,58		<i>krav</i> NN	2,77
98	<i>kretslopp</i> NN	2,58		<i>ö</i> NN	2,76
99	<i>rör</i> NN	2,57		<i>Stockholm</i> NN	2,75
100	<i>hälla</i> VB	2,56		<i>tjänst</i> NN	2,73

Tabell E.16: Indexlista över ord i läroböcker i naturvetenskap/samhällsvetenskap för högstadiet, topp 100

E.2.2.2 *Gymnasiet*

Naturvetenskap		Samhällsvetenskap		
	UNIKT	Index	UNIKT	Index
1	<i>cell</i> NN	20,69	<i>land</i> NN	11,61
2	<i>molekyl</i> NN	19,38	<i>stad</i> NN	11,51
3	<i>elektron</i> NN	15,50	<i>regering</i> NN	11,22
4	<i>ämne</i> NN	13,93	<i>ekonomisk</i> JJ	9,48
5	<i>atom</i> NN	12,91	<i>politisk</i> JJ	9,43
6	<i>reaktion</i> NN	12,51	<i>riksdag</i> NN	8,71
7	<i>protein</i> NN	11,16	<i>krig</i> NN	8,70
8	<i>jon</i> NN	10,08	<i>stat</i> NN	8,43
9	<i>energi</i> NN	8,32	<i>företag</i> NN	8,13
10	<i>vätska</i> NN	7,14	<i>parti</i> NN	8,08
11	<i>reagera</i> VB	7,03	<i>makt</i> NN	7,93
12	<i>partikel</i> NN	6,93	<i>samhälle</i> NN	7,36
13	<i>laddning</i> NN	6,62	<i>svensk</i> JJ	7,35
14	<i>syre</i> NN	6,61	<i>kommun</i> NN	6,95
15	<i>kropp</i> NN	6,22	<i>folk</i> NN	6,77
16	<i>avge</i> VB	6,18	<i>vara</i> NN	6,67
17	<i>bindning</i> NN	6,12	<i>demokrati</i> NN	6,33
18	<i>koncentration</i> NN	6,10	<i>befolkning</i> NN	6,06
19	<i>kemisk</i> JJ	5,99	<i>beslut</i> NN	5,89
20	<i>binda</i> VB	5,93	<i>pengar</i> NN	5,82
21	<i>ljus</i> NN	5,92	<i>fattig</i> JJ	5,78
22	<i>innehålla</i> VB	5,84	<i>rättighet</i> NN	5,52
23	<i>väteatom</i> NN	5,69	<i>kvinnor</i> NN	5,46
24	<i>figur</i> NN	5,67	<i>ekonomi</i> NN	5,42
25	<i>lösning</i> NN	5,66	<i>konflikt</i> NN	5,39
26	<i>proton</i> NN	5,63	<i>köpa</i> VB	5,23
27	<i>grundämne</i> NN	5,54	<i>betala</i> VB	5,17
28	<i>gas</i> NN	5,32	<i>värld</i> NN	4,96
29	<i>temperatur</i> NN	5,25	<i>bo</i> VB	4,84
30	<i>egenskap</i> NN	5,22	<i>politik</i> NN	4,74
31	<i>hastighet</i> NN	4,78	<i>handel</i> NN	4,74
32	<i>massa</i> NN	4,57	<i>familj</i> NN	4,42
33	<i>metall</i> NN	4,48	<i>inkomst</i> NN	4,38
34	<i>elektrisk</i> JJ	4,38	<i>politiker</i> NN	4,32
35	<i>experiment</i> NN	4,35	<i>demokratisk</i> JJ	4,32
36	<i>DNA-molekyl</i> NN	4,34	<i>social</i> JJ	4,31

Naturvetenskap			Samhällsvetenskap		
	UNIKT	Index		UNIKT	Index
37	<i>väte</i> NN	4,31	<i>lön</i> NN		4,27
38	<i>färg</i> NN	4,30	<i>region</i> NN		4,27
39	<i>volym</i> NN	4,20	<i>möjlighet</i> NN		4,18
40	<i>laddad</i> PC	4,07	<i>utbildning</i> NN		4,17
41	<i>tabell</i> NN	4,07	<i>mena</i> VB		4,15
42	<i>blandning</i> NN	4,05	<i>internationell</i> JJ		4,09
43	<i>vatten</i> NN	4,04	<i>samarbete</i> NN		4,09
44	<i>koldioxid</i> NN	3,97	<i>uropeisk</i> JJ		3,99
45	<i>tryck</i> NN	3,93	<i>marknad</i> NN		3,92
46	<i>halt</i> NN	3,92	<i>man</i> NN		3,89
47	<i>ström</i> NN	3,90	<i>jordbruk</i> NN		3,88
48	<i>frigöra</i> VB	3,84	<i>arbetslöshet</i> NN		3,87
49	<i>våglängd</i> NN	3,73	<i>val</i> NN		3,83
50	<i>prov</i> NN	3,61	<i>område</i> NN		3,77
51	<i>bilda</i> VB	3,56	<i>världskrig</i> NN		3,76
52	<i>enhet</i> NN	3,53	<i>offentlig</i> JJ		3,72
53	<i>föremål</i> NN	3,44	<i>besluta</i> VB		3,70
54	<i>strålning</i> NN	3,44	<i>internet</i> NN		3,68
55	<i>blod</i> NN	3,42	<i>kultur</i> NN		3,67
56	<i>omvandla</i> VB	3,38	<i>rätt</i> NN		3,63
57	<i>bunden</i> PC	3,24	<i>barn</i> NN		3,56
58	<i>luft</i> NN	3,18	<i>handla</i> VB		3,54
59	<i>hos</i> PP	3,16	<i>arbetskraft</i> NN		3,50
60	<i>beräkna</i> VB	3,14	<i>näringsliv</i> NN		3,50
61	<i>densitet</i> NN	3,09	<i>allt</i> AB		3,48
62	<i>hud</i> NN	3,09	<i>jag</i> PN		3,47
63	<i>kemi</i> NN	3,09	<i>tidning</i> NN		3,46
64	<i>bestå</i> VB	3,09	<i>skatt</i> NN		3,41
65	<i>kväve</i> NN	3,08	<i>invånare</i> NN		3,38
66	<i>trycke</i> NN	3,08	<i>procent</i> NN		3,37
67	<i>ägg</i> NN	3,03	<i>1800-tal</i> NN		3,37
68	<i>signal</i> NN	3,01	<i>utveckling</i> NN		3,37
69	<i>kärna</i> NN	3,01	<i>idag</i> AB		3,36
70	<i>växt</i> NN	2,92	<i>fråga</i> NN		3,36
71	<i>konstant</i> JJ	2,89	<i>verksamhet</i> NN		3,30
72	<i>absorbera</i> VB	2,84	<i>nyhet</i> NN		3,28
73	<i>kraft</i> NN	2,80	<i>mänsklig</i> JJ		3,26
74	<i>membran</i> NN	2,80	<i>fatta</i> VB		3,24
75	<i>batteri</i> NN	2,80	<i>järnväg</i> NN		3,23

Naturvetenskap			Samhällsvetenskap		
	UNIKT	Index	UNIKT	Index	
76	<i>muskel</i> NN	2,78	<i>skola</i> NN	3,23	
77	<i>levande</i> PC	2,75	<i>äga</i> VB	3,22	
78	<i>ta upp</i> VB	2,75	<i>arbete</i> NN	3,22	
79	<i>väga</i> VB	2,75	<i>tvinga</i> VB	3,21	
80	<i>vandra</i> VB	2,73	<i>gräns</i> NN	3,20	
81	<i>negativ</i> JJ	2,72	<i>rik</i> JJ	3,18	
82	<i>energirik</i> JJ	2,71	<i>råd</i> NN	3,15	
83	<i>djur</i> NN	2,70	<i>hot</i> NN	3,13	
84	<i>fas</i> NN	2,66	<i>medlem</i> NN	3,12	
85	<i>kedja</i> NN	2,65	<i>lokal</i> JJ	3,11	
86	<i>alltså</i> AB	2,65	<i>arbetsmarknad</i> NN	3,11	
87	<i>ner</i> PL	2,61	<i>källa</i> NN	3,11	
88	<i>riktning</i> NN	2,61	<i>plats</i> NN	3,10	
89	<i>spänning</i> NN	2,59	<i>diskutera</i> VB	3,09	
90	<i>tillföra</i> VB	2,59	<i>sälja</i> VB	3,06	
91	<i>vattenmolekyler</i> NN	2,58	<i>vem</i> HP	3,06	
92	<i>tråd</i> NN	2,58	<i>efterfrågan</i> NN	3,05	
93	<i>röd</i> JJ	2,56	<i>tjänst</i> NN	3,03	
94	<i>positiv</i> JJ	2,56	<i>idé</i> NN	3,03	
95	<i>struktur</i> NN	2,53	<i>fabrik</i> NN	3,02	
96	<i>upptäcka</i> VB	2,53	<i>amerikansk</i> JJ	2,98	
97	<i>lagra</i> VB	2,53	<i>syfte</i> NN	2,96	
98	<i>varandra</i> PN	2,52	<i>statlig</i> JJ	2,95	
99	<i>skikt</i> NN	2,50	<i>konkurrens</i> NN	2,94	
100	<i>tunn</i> JJ	2,50	<i>bostad</i> NN	2,92	

Tabell E.17: Indexlista över ord i läroböcker i naturvetenskap/samhällsvetenskap för gymnasiet, topp 100

E.2.3 Biogitexter för högstadiet och gymnasiet

	Högstadiet	Index	Gymnasiet	Index
1	<i>drog</i> NN	5,59	<i>dock</i> AB	7,32
2	<i>alkohol</i> NN	5,16	<i>organisk</i> JJ	6,31
3	<i>druvsocker</i> NN	4,49	<i>därmed</i> AB	6,12
4	<i>dricka</i> VB	4,12	<i>vilket</i> HP	4,61
5	<i>skriva</i> VB	3,89	<i>denna</i> DT	4,33
6	<i>narkotika</i> NN	3,65	<i>glukos</i> NN	4,20
7	<i>Sverige</i> NN	3,61	<i>binda</i> VB	4,19
8	<i>hjärta</i> NN	3,61	<i>denna</i> PN	4,05
9	<i>betyda</i> VB	3,59	<i>utgöra</i> VB	3,84
10	<i>din</i> PS	3,34	<i>molekyl</i> NN	3,71
11	<i>röka</i> VB	3,30	<i>cellmembran</i> NN	3,57
12	<i>du</i> PN	3,26	<i>evolution</i> NN	3,50
13	<i>andas</i> VB	3,12	<i>receptor</i> NN	3,35
14	<i>slida</i> NN	3,12	<i>membran</i> NN	3,26
15	<i>hjärna</i> NN	3,04	<i>arké</i> NN	3,22
16	<i>jag</i> PN	3,02	<i>eukaryot</i> NN	3,20
17	<i>blodkärl</i> NN	2,96	<i>process</i> NN	3,16
18	<i>känsla</i> NN	2,91	<i>jon</i> NN	3,13
19	<i>groddjur</i> NN	2,87	<i>struktur</i> NN	3,05
20	<i>blod</i> NN	2,86	<i>respektive</i> AB	3,03
21	<i>kräldjur</i> NN	2,84	<i>avsnitt</i> NN	3,02
22	<i>uppgift</i> NN	2,83	<i>eukaryot</i> JJ	3,01
23	<i>sinnescell</i> NN	2,83	<i>nukleotid</i> NN	2,93
24	<i>tillverka</i> VB	2,81	<i>vars</i> HS	2,90
25	<i>avfall</i> NN	2,79	<i>kloroplast</i> NN	2,87
26	<i>sort</i> NN	2,77	<i>pH</i> NN	2,87
27	<i>tobak</i> NN	2,76	<i>utnyttja</i> VB	2,87
28	<i>samlag</i> NN	2,74	<i>ribosom</i> NN	2,87
29	<i>muskel</i> NN	2,74	<i>samt</i> KN	2,82
30	<i>sönderdela</i> VB	2,70	<i>cellandning</i> NN	2,81
31	<i>lungblåsa</i> NN	2,67	<i>flertal</i> NN	2,73
32	<i>njure</i> NN	2,66	<i>faktor</i> NN	2,72
33	<i>bra</i> JJ	2,62	<i>m-RNA</i> NN	2,69
34	<i>livmoder</i> NN	2,59	<i>sekvens</i> NN	2,69
35	<i>den här</i> DT	2,59	<i>produktion</i> NN	2,66
36	<i>nikotin</i> NN	2,58	<i>population</i> NN	2,65
37	<i>rita</i> VB	2,58	<i>princip</i> NN	2,60

	Högstadiet	Index	Gymnasiet	Index
38	<i>äta</i> VB	2,57	<i>halt</i> NN	2,57
39	<i>särskild</i> AB	2,54	<i>kommunicera</i> VB	2,49
40	<i>person</i> NN	2,51	<i>bunden</i> PC	2,48
41	<i>rygggrad</i> NN	2,47	<i>begrepp</i> NN	2,48
42	<i>kropp</i> NN	2,46	<i>endast</i> AB	2,47
43	<i>dinosaurie</i> NN	2,46	<i>lipid</i> NN	2,45
44	<i>hård</i> JJ	2,42	<i>upphov</i> NN	2,42
45	<i>tand</i> NN	2,39	<i>diffundera</i> VB	2,39
46	<i>mens</i> NN	2,39	<i>specifik</i> JJ	2,39
47	<i>medicin</i> NN	2,38	<i>celltyp</i> NN	2,38
48	<i>missbrukare</i> NN	2,36	<i>organell</i> NN	2,32
49	<i>ni</i> PN	2,36	<i>individ</i> NN	2,32
50	<i>kamrat</i> NN	2,36	<i>le</i> VB	2,28
51	<i>mussla</i> NN	2,35	<i>spjälka</i> VB	2,27
52	<i>minnas</i> VB	2,35	<i>protein</i> NN	2,27
53	<i>mat</i> NN	2,34	<i>omgivande</i> PC	2,24
54	<i>farlig</i> JJ	2,33	<i>krävas</i> VB	2,24
55	<i>lunga</i> NN	2,32	<i>nämligen</i> AB	2,23
56	<i>skriva ner</i> VB	2,31	<i>förhållande</i> NN	2,22
57	<i>spindel</i> NN	2,31	<i>genvarianter</i> NN	2,21
58	<i>minut</i> NN	2,31	<i>RNA</i> NN	2,21
59	<i>könssjukdom</i> NN	2,30	<i>betydlig</i> AB	2,20
60	<i>skelett</i> NN	2,27	<i>medverka</i> VB	2,18
61	<i>vad</i> HP	2,25	<i>amnosyra</i> NN	2,17
62	<i>graviditet</i> NN	2,23	<i>förmåga</i> NN	2,17
63	<i>läkare</i> NN	2,22	<i>oorganisk</i> JJ	2,15
64	<i>urdjur</i> NN	2,21	<i>evolutionär</i> JJ	2,15
65	<i>bild</i> NN	2,19	<i>anta</i> VB	2,14
66	<i>hud</i> NN	2,18	<i>andel</i> NN	2,12
67	<i>anteckna</i> VB	2,18	<i>grundämne</i> NN	2,12
68	<i>rökare</i> NN	2,17	<i>låg</i> JJ	2,12
69	<i>penis</i> NN	2,17	<i>koncentration</i> NN	2,12
70	<i>penicillin</i> NN	2,16	<i>pH-värde</i> NN	2,12
71	<i>arm</i> NN	2,16	<i>grund</i> NN	2,11
72	<i>dra ihop</i> VB	2,16	<i>utvinna</i> VB	2,11
73	<i>luftrör</i> NN	2,16	<i>ammoniak</i> NN	2,11
74	<i>klaff</i> NN	2,15	<i>embryo</i> NN	2,11
75	<i>pojke</i> NN	2,15	<i>teori</i> NN	2,10
76	<i>prata</i> VB	2,15	<i>berggrund</i> NN	2,10
77	<i>känna sig</i> VB	2,14	<i>bergart</i> NN	2,10

	Högstadiet	Index	Gymnasiet	Index
78	<i>ben</i> NN	2,14	<i>funktion</i> NN	2,09
79	<i>nervtråd</i> NN	2,13	<i>kedja</i> NN	2,08
80	<i>amfetamin</i> NN	2,13	<i>befinna</i> VB	2,07
81	<i>dålig</i> JJ	2,12	<i>mål</i> NN	2,07
82	<i>procent</i> NN	2,12	<i>syfte</i> NN	2,06
83	<i>bota</i> VB	2,09	<i>jordart</i> NN	2,06
84	<i>insekt</i> NN	2,09	<i>analysera</i> VB	2,05
85	<i>symtom</i> NN	2,08	<i>kommunikation</i> NN	2,04
86	<i>kokain</i> NN	2,08	<i>enzym</i> NN	2,04
87	<i>puls</i> NN	2,08	<i>reglerande</i> PC	2,04
88	<i>mjuk</i> JJ	2,07	<i>variant</i> NN	2,03
89	<i>snäcka</i> NN	2,07	<i>frigöra</i> VB	2,03
90	<i>må</i> VB	2,06	<i>kräva</i> VB	2,02
91	<i>ungdom</i> NN	2,06	<i>motsatt</i> PC	2,02
92	<i>barn</i> NN	2,05	<i>slutligen</i> AB	2,02
93	<i>svamptråd</i> NN	2,05	<i>befinna sig</i> VB	2,02
94	<i>näsa</i> NN	2,04	<i>dessutom</i> AB	2,01
95	<i>kondom</i> NN	2,04	<i>kunskap</i> NN	2,01
96	<i>olaglig</i> JJ	2,03	<i>flercellig</i> JJ	2,00
97	<i>flicka</i> NN	2,03	<i>tyckas</i> VB	2,00
98	<i>huvudvärk</i> NN	2,03	<i>förklaring</i> NN	2,00
99	<i>socker</i> NN	2,02	<i>hypotes</i> NN	2,00
100	<i>luft</i> NN	2,02	<i>aktivera</i> VB	2,00

Tabell E.18: Indexlista över ord i biologiläroböcker, topp 100

E.2.4 Fysiktexter för högstadiet och gymnasiet

	Högstadiet	Index	Gymnasiet	Index
1	<i>lins</i> NN	7,06	<i>konstant</i> JJ	8,83
2	<i>anteckna</i> VB	6,69	<i>fält</i> NN	6,99
3	<i>måne</i> NN	5,59	<i>accelerera</i> VB	6,80
4	<i>konvex</i> JJ	5,42	<i>acceleration</i> NN	6,02
5	<i>kikare</i> NN	5,22	<i>beräkna</i> VB	5,75
6	<i>transistor</i> NN	4,51	<i>figur</i> NN	5,64
7	<i>himmel</i> NN	4,41	<i>lösning</i> NN	5,44
8	<i>diod</i> NN	4,36	<i>enligt</i> PP	5,37
9	<i>ljud</i> NN	4,24	<i>värde</i> NN	5,27
10	<i>konkav</i> JJ	4,10	<i>area</i> NN	5,27
11	<i>ämne</i> NN	3,94	<i>ffäder</i> NN	5,18
12	<i>synas</i> VB	3,63	<i>rörelsemängd</i> NN	5,04
13	<i>hävstång</i> NN	3,51	<i>medium</i> NN	4,30
14	<i>sol</i> NN	3,42	<i>bestämna</i> VB	4,21
15	<i>elektromagnet</i> NN	3,37	<i>graf</i> NN	4,16
16	<i>stjärnbild</i> NN	3,36	<i>hastighet</i> NN	4,07
17	<i>spegel</i> NN	3,32	<i>radie</i> NN	4,04
18	<i>så här</i> AB	3,31	<i>nuklid</i> NN	4,03
19	<i>materia</i> NN	3,27	<i>positiv</i> JJ	3,87
20	<i>magnet</i> NN	3,24	<i>foton</i> NN	3,72
21	<i>stjärna</i> NN	3,20	<i>förhållande</i> NN	3,62
22	<i>öga</i> NN	3,19	<i>resulterande</i> PC	3,57
23	<i>utvidga</i> VB	3,14	<i>beteckna</i> VB	3,49
24	<i>kemisk</i> JJ	3,05	<i>därmed</i> AB	3,43
25	<i>förstås</i> AB	3,05	<i>riktad</i> PC	3,41
26	<i>planet</i> NN	3,04	<i>lag</i> NN	3,40
27	<i>höra</i> VB	3,03	<i>fältstyrka</i> NN	3,39
28	<i>glasögon</i> NN	3,03	<i>proportionell</i> JJ	3,39
29	<i>ljudvåg</i> NN	3,00	<i>komposant</i> NN	3,39
30	<i>färg</i> NN	2,95	<i>elektromagnetisk</i> JJ	3,32
31	<i>infraröd</i> JJ	2,91	<i>storhet</i> NN	3,31
32	<i>järn</i> NN	2,91	<i>samband</i> NN	3,27
33	<i>meteor</i> NN	2,88	<i>mätning</i> NN	3,24
34	<i>växthus</i> NN	2,85	<i>uppskatta</i> VB	3,22
35	<i>brännpunkt</i> NN	2,81	<i>avge</i> VB	3,19
36	<i>vit</i> JJ	2,79	<i>anta</i> VB	3,19
37	<i>behöva</i> VB	2,78	<i>ljushastighet</i> NN	3,16

	Högstadiet	Index	Gymnasiet	Index
38	<i>plats</i> NN	2,78	<i>vinkelrät</i> JJ	3,16
39	<i>kamera</i> NN	2,77	<i>horisontell</i> JJ	3,14
40	<i>tro</i> VB	2,75	<i>normalkraft</i> NN	3,14
41	<i>transformator</i> NN	2,74	<i>maximal</i> JJ	3,14
42	<i>sak</i> NN	2,74	<i>total</i> JJ	3,13
43	<i>flaska</i> NN	2,73	<i>diagram</i> NN	3,10
44	<i>ytpänning</i> NN	2,72	<i>verka</i> VB	3,07
45	<i>Sverige</i> NN	2,72	<i>densitet</i> NN	3,05
46	<i>betastrålning</i> NN	2,71	<i>storlek</i> NN	3,02
47	<i>bild</i> NN	2,68	<i>vertikal</i> JJ	3,02
48	<i>tröghet</i> NN	2,67	<i>triangel</i> NN	3,02
49	<i>vanlig</i> JJ	2,65	<i>vektor</i> NN	3,02
50	<i>batteri</i> NN	2,65	<i>dos</i> NN	3,01
51	<i>ljudkälla</i> NN	2,64	<i>svar</i> NN	3,01
52	<i>termometer</i> NN	2,63	<i>laddning</i> NN	2,99
53	<i>skugga</i> NN	2,62	<i>ögonblick</i> NN	2,98
54	<i>magnetism</i> NN	2,62	<i>övning</i> NN	2,93
55	<i>galax</i> NN	2,60	<i>formel</i> NN	2,93
56	<i>varm</i> JJ	2,58	<i>jämviktsläge</i> NN	2,93
57	<i>heta</i> VB	2,57	<i>genomsnittlig</i> JJ	2,93
58	<i>atomnummer</i> NN	2,57	<i>stav</i> NN	2,93
59	<i>apparat</i> NN	2,57	<i>våglängd</i> NN	2,88
60	<i>mera</i> AB	2,57	<i>vagn</i> NN	2,87
61	<i>slag</i> NN	2,56	<i>beräkning</i> NN	2,82
62	<i>läsa</i> VB	2,54	<i>våg</i> NN	2,80
63	<i>varv</i> NN	2,51	<i>tidpunkt</i> NN	2,77
64	<i>farlig</i> JJ	2,51	<i>avta</i> VB	2,76
65	<i>spik</i> NN	2,50	<i>analysera</i> VB	2,74
66	<i>buller</i> NN	2,50	<i>definiera</i> VB	2,74
67	<i>musik</i> NN	2,50	<i>vila</i> NN	2,69
68	<i>elmotor</i> NN	2,50	<i>lutning</i> NN	2,69
69	<i>kärnenergi</i> NN	2,50	<i>kraft</i> NN	2,67
70	<i>komet</i> NN	2,50	<i>negativ</i> JJ	2,67
71	<i>hål</i> NN	2,48	<i>potentiell</i> JJ	2,65
72	<i>ju</i> AB	2,47	<i>myon</i> NN	2,65
73	<i>växelström</i> NN	2,46	<i>spalt</i> NN	2,65
74	<i>el</i> NN	2,46	<i>prognos</i> NN	2,65
75	<i>ledning</i> NN	2,45	<i>flöde</i> NN	2,63
76	<i>lysdiod</i> NN	2,43	<i>naturligtvis</i> AB	2,61
77	<i>brännvidd</i> NN	2,42	<i>hjälp</i> NN	2,60

	Högstadiet	Index	Gymnasiet	Index
78	<i>maskin</i> NN	2,42	<i>stråldos</i> NN	2,60
79	<i>fullmåne</i> NN	2,41	<i>puls</i> NN	2,59
80	<i>radioaktiv</i> JJ	2,41	<i>producera</i> VB	2,56
81	<i>lampa</i> NN	2,40	<i>denna</i> DT	2,53
82	<i>ljus</i> JJ	2,39	<i>kollision</i> NN	2,53
83	<i>astronom</i> NN	2,39	<i>neråt</i> AB	2,51
84	<i>pröva</i> VB	2,38	<i>metod</i> NN	2,49
85	<i>tänka ut</i> VB	2,38	<i>innebära</i> VB	2,49
86	<i>lysa</i> VB	2,38	<i>homogen</i> AB	2,47
87	<i>sjö</i> NN	2,38	<i>impuls</i> NN	2,47
88	<i>strålningsenergi</i> NN	2,37	<i>konstant</i> NN	2,47
89	<i>se ut</i> VB	2,36	<i>elektroskop</i> NN	2,47
90	<i>växt</i> NN	2,35	<i>kurva</i> NN	2,47
91	<i>tv</i> NN	2,35	<i>preparat</i> NN	2,46
92	<i>elektronrör</i> NN	2,35	<i>tumör</i> NN	2,46
93	<i>någonting</i> PN	2,35	<i>längd</i> NN	2,45
94	<i>öppna</i> VB	2,34	<i>nivå</i> NN	2,43
95	<i>fokus</i> NN	2,33	<i>variera</i> VB	2,42
96	<i>näthinna</i> NN	2,32	<i>ange</i> VB	2,39
97	<i>gå upp</i> VB	2,30	<i>amplitud</i> NN	2,38
98	<i>elektronik</i> NN	2,30	<i>brytningsindex</i> NN	2,38
99	<i>kylskåp</i> NN	2,30	<i>likformig</i> JJ	2,38
100	<i>vad</i> HP	2,29	<i>coulomb</i> NN	2,38

Tabell E.19: Indexlista över ord i fysikläroböcker, topp 100

E.2.5 Kemitexter för högstadiet och gymnasiet

	Högstadiet	Index	Gymnasiet	Index
1	<i>eld</i> NN	7,98	<i>koncentration</i> NN	12,76
2	<i>sak</i> NN	5,76	<i>mol</i> NN	11,36
3	<i>brinna</i> VB	5,34	<i>figur</i> NN	9,19
4	<i>äta</i> VB	4,95	<i>fas</i> NN	7,36
5	<i>faktiskt</i> AB	4,79	<i>beräkna</i> VB	7,32
6	<i>förvandla</i> VB	4,77	<i>volym</i> NN	7,06
7	<i>mark</i> NN	4,74	<i>jämvikt</i> NN	6,21
8	<i>jord</i> NN	4,59	<i>ange</i> VB	6,17
9	<i>du</i> PN	4,58	<i>följande</i> PC	5,42
10	<i>varför</i> HD	4,52	<i>kovalent</i> JJ	5,17
11	<i>växt</i> NN	4,50	<i>bestämma</i> VB	4,95
12	<i>olja</i> NN	4,17	<i>substansmängd</i> NN	4,81
13	<i>se ut</i> VB	4,12	<i>respektive</i> AB	4,80
14	<i>kropp</i> NN	4,01	<i>oxidera</i> VB	4,63
15	<i>mat</i> NN	3,99	<i>vätebindning</i> NN	4,62
16	<i>trä</i> NN	3,99	<i>prov</i> NN	4,53
17	<i>tillverka</i> VB	3,98	<i>därmed</i> AB	4,39
18	<i>tänka</i> VB	3,98	<i>pH</i> NN	4,31
19	<i>malm</i> NN	3,95	<i>ekvivalenspunkt</i> NN	4,23
20	<i>djur</i> NN	3,88	<i>kolonn</i> NN	4,14
21	<i>väldigt</i> AB	3,88	<i>tabell</i> NN	3,91
22	<i>bränsle</i> NN	3,87	<i>dipol</i> NN	3,89
23	<i>tro</i> VB	3,81	<i>kapitel</i> NN	3,86
24	<i>vad</i> HP	3,78	<i>oxidationstal</i> NN	3,75
25	<i>ihop</i> PL	3,77	<i>värde</i> NN	3,73
26	<i>druvsocker</i> NN	3,65	<i>ske</i> VB	3,71
27	<i>sort</i> NN	3,52	<i>struktur</i> NN	3,71
28	<i>stål</i> NN	3,50	<i>oxoniumjon</i> NN	3,70
29	<i>socker</i> NN	3,45	<i>reducera</i> VB	3,70
30	<i>emulsion</i> NN	3,44	<i>våglängd</i> NN	3,65
31	<i>fossil</i> JJ	3,42	<i>titrera</i> VB	3,65
32	<i>växthuseffekt</i> NN	3,38	<i>titrering</i> NN	3,61
33	<i>fett</i> NN	3,31	<i>anta</i> VB	3,60
34	<i>glas</i> NN	3,30	<i>stationär</i> JJ	3,56
35	<i>vätejon</i> NN	3,29	<i>enligt</i> PP	3,48
36	<i>kretslopp</i> NN	3,27	<i>studera</i> VB	3,45
37	<i>gips</i> NN	3,26	<i>konstant</i> JJ	3,41

	Högstadiet	Index	Gymnasiet	Index
38	<i>sten</i> NN	3,23	<i>analys</i> NN	3,41
39	<i>stenkol</i> NN	3,23	<i>halt</i> NN	3,40
40	<i>din</i> PS	3,22	<i>vattenlösning</i> NN	3,33
41	<i>plast</i> NN	3,21	<i>beteckna</i> VB	3,30
42	<i>regn</i> NN	3,19	<i>riktning</i> NN	3,30
43	<i>atomslag</i> NN	3,19	<i>jod</i> NN	3,30
44	<i>tänka ut</i> VB	3,18	<i>enhet</i> NN	3,21
45	<i>tvål</i> NN	3,17	<i>motsvarande</i> PC	3,17
46	<i>hjälpa</i> VB	3,17	<i>dess</i> PS	3,15
47	<i>skada</i> VB	3,13	<i>hastighet</i> NN	3,03
48	<i>farlig</i> JJ	3,11	<i>funktionell</i> JJ	3,03
49	<i>mena</i> VB	3,10	<i>absorbera</i> VB	3,02
50	<i>släppa</i> VB	3,07	<i>utföra</i> VB	3,01
51	<i>brand</i> NN	3,07	<i>vars</i> HS	3,00
52	<i>giftig</i> JJ	3,07	<i>avge</i> VB	3,00
53	<i>hålla ihop</i> VB	3,06	<i>binda</i> VB	2,97
54	<i>gjord</i> PC	3,05	<i>förhållande</i> NN	2,94
55	<i>släcka</i> VB	3,04	<i>beräkning</i> NN	2,94
56	<i>full</i> JJ	3,03	<i>substansmängd</i> NN	2,93
57	<i>försvinna</i> VB	3,03	<i>reaktionshastighet</i> NN	2,93
58	<i>rosta</i> VB	3,01	<i>svar</i> NN	2,93
59	<i>biobränsle</i> NN	3,00	<i>elektronpar</i> NN	2,92
60	<i>släppa ut</i> VB	3,00	<i>reaktion</i> NN	2,88
61	<i>människa</i> NN	2,99	<i>denna</i> DT	2,81
62	<i>ozon</i> NN	2,99	<i>innebära</i> VB	2,80
63	<i>liv</i> NN	2,95	<i>endast</i> AB	2,80
64	<i>naturgas</i> NN	2,95	<i>reaktionsformel</i> NN	2,79
65	<i>icke-metall</i> NN	2,94	<i>diagram</i> NN	2,79
66	<i>mycket</i> JJ	2,92	<i>attrahera</i> VB	2,79
67	<i>gummi</i> NN	2,92	<i>löslighet</i> NN	2,79
68	<i>bil</i> NN	2,90	<i>hydrofob</i> JJ	2,79
69	<i>kväva</i> VB	2,86	<i>motsvara</i> VB	2,77
70	<i>dricka</i> VB	2,84	<i>antal</i> NN	2,76
71	<i>bra</i> AB	2,82	<i>oxidationsmedel</i> NN	2,74
72	<i>utsläpp</i> NN	2,81	<i>uppta</i> VB	2,72
73	<i>konstgjord</i> JJ	2,80	<i>överföra</i> VB	2,70
74	<i>smaka</i> VB	2,78	<i>exoterm</i> NN	2,69
75	<i>råvara</i> NN	2,78	<i>membran</i> NN	2,69
76	<i>tensid</i> NN	2,77	<i>fenol</i> NN	2,69
77	<i>återvinna</i> VB	2,76	<i>mobil</i> JJ	2,69

	Högstadiet	Index	Gymnasiet	Index
78	<i>smuts</i> NN	2,74	<i>tillstånd</i> NN	2,68
79	<i>metalloxid</i> NN	2,72	<i>tryck</i> NN	2,67
80	<i>koksalt</i> NN	2,71	<i>markera</i> VB	2,66
81	<i>sand</i> NN	2,68	<i>per</i> PP	2,65
82	<i>rena</i> VB	2,68	<i>vänster</i> NN	2,64
83	<i>batteri</i> NN	2,66	<i>entalpi</i> NN	2,64
84	<i>is</i> NN	2,65	<i>mäta</i> VB	2,63
85	<i>död</i> JJ	2,65	<i>bägare</i> NN	2,61
86	<i>sätta ihop</i> VB	2,63	<i>därefter</i> AB	2,60
87	<i>material</i> NN	2,61	<i>total</i> JJ	2,59
88	<i>skydda</i> VB	2,61	<i>råda</i> VB	2,59
89	<i>medicin</i> NN	2,61	<i>temperatur</i> NN	2,59
90	<i>problem</i> NN	2,60	<i>båda</i> JJ	2,59
91	<i>ändå</i> AB	2,59	<i>råda</i> VB	2,59
92	<i>alkemist</i> NN	2,57	<i>motsatt</i> PC	2,58
93	<i>förbränning</i> NN	2,56	<i>uppnå</i> VB	2,58
94	<i>elda</i> VB	2,56	<i>vänster</i> JJ	2,54
95	<i>förr</i> AB	2,55	<i>reaktiv</i> JJ	2,54
96	<i>värld</i> NN	2,54	<i>framgå</i> VB	2,54
97	<i>mineralämne</i> NN	2,52	<i>strukturformel</i> NN	2,51
98	<i>mjuk</i> JJ	2,51	<i>överskott</i> NN	2,51
99	<i>hålla</i> VB	2,50	<i>formel</i> NN	2,50
100	<i>råolja</i> NN	2,49	<i>protolyseras</i> VB	2,50

Tabell E.20: Indexlista över ord i kemiläroböcker, topp 100

E.2.6 Naturvetenskapliga texter jämförda med akademiska texter

	Naturvetenskap		SveAk	
1	<i>som</i> HP	193,19	<i>som</i> KN	91,93
2	<i>den</i> DT	102,18	<i>andra</i> JJ	23,06
3	<i>vatten</i> NN	27,99	<i>jag</i> PN	19,82
4	<i>än</i> KN	22,52	<i>vilken</i> HP	19,09
5	<i>cell</i> NN	16,79	<i>både</i> KN	11,38
6	<i>energi</i> NN	16,56	<i>min</i> PS	8,93
7	<i>bilda</i> VB	14,02	<i>social</i> JJ	8,69
8	<i>atom</i> NN	13,64	<i>relation</i> NN	7,97
9	<i>växt</i> NN	13,46	<i>studie</i> NN	7,39
10	<i>molekyl</i> NN	13,34	<i>perspektiv</i> NN	6,42
11	<i>ämne</i> NN	12,96	<i>utifrån</i> PP	5,87
12	<i>elektron</i> NN	12,37	<i>avhandling</i> NN	5,84
13	<i>varje</i> DT	12,16	<i>diskussion</i> NN	5,43
14	<i>luft</i> NN	11,98	<i>politisk</i> JJ	5,27
15	<i>kemisk</i> JJ	11,18	<i>erfarenhet</i> NN	4,9
16	<i>syre</i> NN	10,83	<i>syfte</i> NN	4,79
17	<i>börja</i> VB	10,74	<i>såväl</i> KN	4,71
18	<i>bakterie</i> NN	10,17	<i>respektive</i> AB	4,62
19	<i>jord</i> NN	9,97	<i>fråga</i> NN	4,62
20	<i>protein</i> NN	9,25	<i>situation</i> NN	4,45
21	<i>djur</i> NN	9,21	<i>verksamhet</i> NN	4,42
22	<i>gas</i> NN	8,99	<i>intervju</i> NN	4,37
23	<i>lång</i> AB	8,98	<i>tolkning</i> NN	4,37
24	<i>koldioxid</i> NN	8,67	<i>aspekt</i> NN	4,35
25	<i>metall</i> NN	8,49	<i>möjlighet</i> NN	4,29
26	<i>temperatur</i> NN	8,43	<i>hon</i> PN	4,28
27	<i>jon</i> NN	7,65	<i>text</i> NN	4,13
28	<i>vätska</i> NN	7,27	<i>dels</i> KN	4,09
29	<i>mycket</i> JJ	7,06	<i>handling</i> NN	4,06
30	<i>kalla</i> VB	6,81	<i>utgångspunkt</i> NN	4,05
31	<i>röra sig</i> VB	6,56	<i>snarare</i> AB	4,04
32	<i>syra</i> NN	6,52	<i>intresse</i> NN	4,02
33	<i>kropp</i> NN	6,46	<i>begrepp</i> NN	4,01
34	<i>helt</i> AB	6,44	<i>uttryck</i> NN	4
35	<i>du</i> PN	6,42	<i>samtal</i> NN	3,98
36	<i>ta upp</i> VB	6,24	<i>således</i> AB	3,92
37	<i>båda</i> JJ	6,02	<i>sammanhang</i> NN	3,89

	Naturvetenskap		SveAk	
38	<i>liten</i> AB	5,95	<i>diskutera</i> VB	3,82
39	<i>blod</i> NN	5,93	<i>handla</i> VB	3,81
40	<i>inhålla</i> VB	5,87	<i>aktör</i> NN	3,81
41	<i>elektrisk</i> JJ	5,87	<i>svensk</i> JJ	3,8
42	<i>partikel</i> NN	5,75	<i>analys</i> NN	3,74
43	<i>färg</i> NN	5,7	<i>mening</i> NN	3,68
44	<i>ljus</i> NN	5,66	<i>berättelse</i> NN	3,64
45	<i>ström</i> NN	5,62	<i>uttrycka</i> VB	3,58
46	<i>tillverka</i> VB	5,54	<i>dock</i> AB	3,56
47	<i>grundämne</i> NN	5,52	<i>kontext</i> NN	3,54
48	<i>reaktion</i> NN	5,42	<i>föreställning</i> NN	3,54
49	<i>kolatomb</i> NN	5,42	<i>betrakta</i> VB	3,54
50	<i>tryck</i> NN	5,36	<i>mena</i> VB	3,53
51	<i>se ut</i> VB	5,29	<i>språk</i> NN	3,53
52	<i>fett</i> NN	5,18	<i>praktik</i> NN	3,5
53	<i>kol</i> NN	5,16	<i>utbildning</i> NN	3,5
54	<i>värme</i> NN	5,12	<i>ekonomisk</i> JJ	3,39
55	<i>mark</i> NN	5,04	<i>fokus</i> NN	3,38
56	<i>järn</i> NN	4,98	<i>val</i> NN	3,33
57	<i>proton</i> NN	4,98	<i>elev</i> NN	3,32
58	<i>gen</i> NN	4,96	<i>emellertid</i> AB	3,29
59	<i>salt</i> NN	4,85	<i>skola</i> NN	3,29
60	<i>sol</i> NN	4,8	<i>ovan</i> AB	3,27
61	<i>omvandla</i> VB	4,78	<i>liksom</i> PP	3,25
62	<i>reagera</i> VB	4,69	<i>position</i> NN	3,23
63	<i>hav</i> NN	4,63	<i>kategori</i> NN	3,23
64	<i>laddning</i> NN	4,6	<i>betydelse</i> NN	3,22
65	<i>sjö</i> NN	4,59	<i>betona</i> VB	3,21
66	<i>organism</i> NN	4,59	<i>lärare</i> NN	3,17
67	<i>bränsle</i> NN	4,58	<i>enskild</i> JJ	3,16
68	<i>yta</i> NN	4,4	<i>tolka</i> VB	3,14
69	<i>alkohol</i> NN	4,34	<i>teoretisk</i> JJ	3,14
70	<i>väteatom</i> NN	4,33	<i>samhälle</i> NN	3,13
71	<i>binda</i> VB	4,32	<i>makt</i> NN	3,12
72	<i>massa</i> NN	4,31	<i>central</i> JJ	3,1
73	<i>enzym</i> NN	4,29	<i>ena</i> JJ	3,08
74	<i>strålning</i> NN	4,26	<i>karaktär</i> NN	3,08
75	<i>äta</i> VB	4,23	<i>kvinnor</i> NN	3,07
76	<i>organisk</i> JJ	4,23	<i>presentera</i> VB	3,06
77	<i>avge</i> VB	4,21	<i>framstå</i> VB	3,04

	Naturvetenskap		SveAk	
78	<i>hud</i> NN	4,21	<i>uppfatta</i> VB	3,04
79	<i>röd</i> JJ	4,19	<i>bakgrund</i> NN	3,04
80	<i>ägg</i> NN	4,16	<i>konsekvens</i> NN	3,03
81	<i>väte</i> NN	4,03	<i>tydlig</i> JJ	2,98
82	<i>art</i> NN	3,99	<i>förståelse</i> NN	2,98
83	<i>varm</i> JJ	3,98	<i>genomföra</i> VB	2,98
84	<i>bestå</i> VB	3,98	<i>kritik</i> NN	2,97
85	<i>ner</i> PL	3,98	<i>institution</i> NN	2,95
86	<i>muskel</i> NN	3,97	<i>forskning</i> NN	2,95
87	<i>bra</i> AB	3,97	<i>utsträckning</i> NN	2,94
88	<i>atmosfär</i> NN	3,95	<i>tidig</i> JJ	2,93
89	<i>batteri</i> NN	3,94	<i>avse</i> VB	2,93
90	<i>hjärna</i> NN	3,89	<i>möte</i> NN	2,93
91	<i>experiment</i> NN	3,87	<i>hennes</i> PS	2,89
92	<i>hastighet</i> NN	3,85	<i>kulturell</i> JJ	2,88
93	<i>tunn</i> JJ	3,84	<i>förutsättning</i> NN	2,85
94	<i>släppa</i> VB	3,83	<i>syn</i> NN	2,84
95	<i>blandning</i> NN	3,82	<i>kring</i> PP	2,84
96	<i>ungefär</i> AB	3,82	<i>ram</i> NN	2,83
97	<i>sur</i> JJ	3,77	<i>personlig</i> JJ	2,81
98	<i>aminosyra</i> NN	3,76	<i>hävda</i> VB	2,8
99	<i>kväve</i> NN	3,74	<i>ansvar</i> NN	2,79
100	<i>bindning</i> NN	3,73	<i>historisk</i> JJ	2,78

Tabell E.21: Indexlista över ord i läroböcker i naturvetenskap/akademiska texter, topp 100

E.2.7 Naturvetenskapliga texter jämförda med berättande texter

	Naturvetenskap	Index	Berättande	Index
1	<i>ämne</i> NN	24,57	<i>jag</i> PN	84,69
2	<i>bilda</i> VB	23,19	<i>hon</i> PN	40,40
3	<i>energi</i> NN	16,62	<i>han</i> PN	24,93
4	<i>cell</i> NN	15,33	<i>min</i> PS	14,73
5	<i>olik</i> JJ	14,88	<i>hennes</i> PS	12,42
6	<i>innehålla</i> VB	13,91	<i>hans</i> PS	11,87
7	<i>atom</i> NN	13,79	<i>säga</i> VB	10,53
8	<i>molekyl</i> NN	13,31	<i>ja</i> IN	9,33
9	<i>elektron</i> NN	12,36	<i>fråga</i> VB	8,90
10	<i>växt</i> NN	12,11	<i>ni</i> PN	8,84
11	<i>kemisk</i> JJ	11,17	<i>ansikte</i> NN	6,86
12	<i>syre</i> NN	10,72	<i>blick</i> NN	6,83
13	<i>bestå</i> VB	10,63	<i>dörr</i> NN	6,67
14	<i>bakterie</i> NN	10,06	<i>nej</i> IN	6,26
15	<i>använda</i> VB	9,63	<i>väl</i> AB	6,16
16	<i>exempel</i> NN	9,38	<i>röst</i> NN	5,53
17	<i>protein</i> NN	9,26	<i>nog</i> AB	5,43
18	<i>koldioxid</i> NN	8,75	<i>undra</i> VB	5,33
19	<i>temperatur</i> NN	8,29	<i>ringa</i> VB	5,33
20	<i>gas</i> NN	8,15	<i>aldrig</i> AB	5,31
21	<i>egenskap</i> NN	8,08	<i>prata</i> VB	5,12
22	<i>påverka</i> VB	8,02	<i>hem</i> PL	5,06
23	<i>metall</i> NN	7,69	<i>skratta</i> VB	5,05
24	<i>jon</i> NN	7,67	<i>vilja</i> VB	5,01
25	<i>reaktion</i> NN	7,62	<i>huvud</i> NN	5,01
26	<i>vatten</i> NN	7,39	<i>ingenting</i> PN	4,97
27	<i>kalla</i> VB	7,22	<i>vänta</i> VB	4,96
28	<i>öka</i> VB	7,13	<i>kväll</i> NN	4,88
29	<i>lösning</i> NN	6,96	<i>säng</i> NN	4,87
30	<i>minska</i> VB	6,60	<i>resa</i> VB	4,74
31	<i>djur</i> NN	6,60	<i>tycka</i> VB	4,71
32	<i>vätska</i> NN	6,56	<i>där</i> AB	4,57
33	<i>jord</i> NN	6,53	<i>sätta sig</i> VB	4,52
34	<i>syra</i> NN	6,39	<i>nu</i> AB	4,49
35	<i>fungera</i> VB	6,10	<i>veta</i> VB	4,49
36	<i>mängd</i> NN	6,07	<i>vända sig</i> VB	4,45
37	<i>omvandla</i> VB	6,02	<i>resa sig</i> VB	4,44

	Naturvetenskap	Index	Berättande	Index
38	<i>bero</i> VB	5,81	<i>folk</i> NN	4,43
39	<i>partikel</i> NN	5,77	<i>nicka</i> VB	4,42
40	<i>utveckla</i> VB	5,73	<i>berätta</i> VB	4,34
41	<i>därför</i> AB	5,72	<i>ingen</i> PN	4,29
42	<i>grundämne</i> NN	5,52	<i>höra</i> VB	4,20
43	<i>tillverka</i> VB	5,49	<i>ropa</i> VB	4,15
44	<i>material</i> NN	5,44	<i>morgon</i> NN	4,05
45	<i>kolatom</i> NN	5,43	<i>ens</i> AB	4,03
46	<i>form</i> NN	5,43	<i>be</i> VB	4,00
47	<i>viss</i> JJ	5,30	<i>stirra</i> VB	3,99
48	<i>antal</i> NN	5,21	<i>hus</i> NN	3,98
49	<i>elektrisk</i> JJ	5,18	<i>sova</i> VB	3,92
50	<i>grupp</i> NN	5,11	<i>mamma</i> NN	3,83
51	<i>forskare</i> NN	5,07	<i>stå</i> VB	3,82
52	<i>gen</i> NN	5,04	<i>vända</i> VB	3,82
53	<i>metod</i> NN	5,01	<i>tro</i> VB	3,82
54	<i>proton</i> NN	5,00	<i>rygg</i> NN	3,71
55	<i>fett</i> NN	4,95	<i>rum</i> NN	3,71
56	<i>miljö</i> NN	4,88	<i>kök</i> NN	3,70
57	<i>kol</i> NN	4,87	<i>svara</i> VB	3,70
58	<i>ske</i> VB	4,86	<i>pappa</i> NN	3,68
59	<i>laddning</i> NN	4,86	<i>känna sig</i> VB	3,68
60	<i>art</i> NN	4,86	<i>skrika</i> VB	3,64
61	<i>utvecklas</i> VB	4,83	<i>le</i> VB	3,62
62	<i>organism</i> NN	4,80	<i>hand</i> NN	3,58
63	<i>ofta</i> AB	4,80	<i>säga sig</i> VB	3,55
64	<i>reagera</i> VB	4,79	<i>lyssna</i> VB	3,51
65	<i>många</i> JJ	4,76	<i>ensam</i> JJ	3,50
66	<i>ström</i> NN	4,69	<i>vem</i> HP	3,50
67	<i>bränsle</i> NN	4,63	<i>minnas</i> VB	3,49
68	<i>salt</i> NN	4,59	<i>arm</i> NN	3,44
69	<i>organisk</i> JJ	4,57	<i>ju</i> AB	3,43
70	<i>tryck</i> NN	4,55	<i>far</i> NN	3,43
71	<i>järn</i> NN	4,53	<i>alldeles</i> AB	3,42
72	<i>ingå</i> VB	4,46	<i>plötslig</i> AB	3,39
73	<i>beskriva</i> VB	4,45	<i>bakom</i> PP	3,37
74	<i>orsaka</i> VB	4,43	<i>stund</i> NN	3,37
75	<i>alkohol</i> NN	4,35	<i>man</i> NN	3,36
76	<i>väteatom</i> NN	4,34	<i>axel</i> NN	3,34
77	<i>natur</i> NN	4,31	<i>åka</i> VB	3,33

	Naturvetenskap	Index	Berättande	Index
78	<i>binda</i> VB	4,30	<i>titta</i> VB	3,30
79	<i>även</i> AB	4,28	<i>ögonblick</i> NN	3,27
80	<i>strålning</i> NN	4,27	<i>natt</i> NN	3,27
81	<i>enzym</i> NN	4,27	<i>viska</i> VB	3,25
82	<i>viktig</i> JJ	4,26	<i>gång</i> NN	3,25
83	<i>lösa</i> VB	4,23	<i>igen</i> AB	3,25
84	<i>avge</i> VB	4,22	<i>bord</i> NN	3,25
85	<i>förening</i> NN	4,15	<i>rädd</i> JJ	3,24
86	<i>mäta</i> VB	4,09	<i>om</i> PL	3,23
87	<i>experiment</i> NN	4,09	<i>fönster</i> NN	3,21
88	<i>atmosfär</i> NN	4,08	<i>glad</i> JJ	3,18
89	<i>kapitel</i> NN	4,03	<i>tyst</i> JJ	3,17
90	<i>väte</i> NN	4,03	<i>ta sig</i> VB	3,15
91	<i>del</i> NN	3,97	<i>vän</i> NN	3,14
92	<i>yta</i> NN	3,97	<i>tänka</i> VB	3,14
93	<i>system</i> NN	3,97	<i>trappa</i> NN	3,13
94	<i>luft</i> NN	3,95	<i>glömma</i> VB	3,12
95	<i>bild</i> NN	3,90	<i>pengar</i> NN	3,12
96	<i>innebära</i> VB	3,89	<i>på</i> PL	3,11
97	<i>bindning</i> NN	3,84	<i>peng</i> NN	3,08
98	<i>blandning</i> NN	3,83	<i>någon</i> PN	3,08
99	<i>hastighet</i> NN	3,77	<i>låta</i> VB	3,08
100	<i>framställa</i> VB	3,77	<i>knä</i> NN	3,01

Tabell E.22: Indexlista över ord i läroböcker i naturvetenskap/berättande texter, topp 100

E.2.8 Procentuell fördelning av ordformer som räknas som participenheter i ordlistorna

Particippost	Procentuell fördelning av ordformer	Ursprungslista	
		Tabell	Position
<i>levande</i>	levande: 100	E.14	77
<i>laddad</i>	laddade: 55, laddad: 37, laddat: 8	E.15	42
<i>bunden</i>	bundna: 54, bunden: 36, bundet: 11	E.15	58
<i>levande</i>	levande: 100	E.16	76
<i>uppbyggd</i>	uppbyggda: 42, uppbyggd: 38, uppbyggt: 20, uppbyggande: 1	E.16	89
<i>laddad</i>	laddade: 55, laddad: 37, laddat: 8	E.17	40
<i>bunden</i>	bundna: 54, bunden: 36, bundet: 11	E.17	57
<i>levande</i>	levande: 100	E.17	77
<i>bunden</i>	bundna: 56, bunden: 25, bundet: 13, bindande: 6	E.18	40
<i>omgivande</i>	omgivande: 79, omgiven: 13, omgivna: 4, omgivet: 3	E.18	53
<i>reglerande</i>	reglerande: 91, reglerat: 6, reglerad: 3	E.18	87
<i>motsatt</i>	motsatt: 55, motsatta: 45	E.18	91
<i>resulterande</i>	resulterande: 100	E.19	22
<i>riktad</i>	riktad: 80, riktade: 16, riktat: 4	E.19	25
<i>följande</i>	följande: 100	E.20	9
<i>motsvarande</i>	motsvarande: 100	E.20	45
<i>gjord</i>	gjorda: 66, gjord: 32, gjort: 1	E.20	54
<i>motsatt</i>	motsatt: 50, motsatta: 50	E.20	92

Tabell E.23: Procentuell fördelning av ordformer som räknas som participenheter i ordlistorna

E.3 Delstudie: typiska nominalfraser

E.3.1 Nominalfraser per mening

Ämne/Ämnesfack	Stadium	Medel	Median	Std.av.
Naturvetenskap	Högstadiet	2,59	2	1,34
	Gymnasiet	2,86	3	1,53
Matematik	Högstadiet	2,16	2	1,27
	Gymnasiet	2,30	2	1,41
Samhällsvetenskap	Högstadiet	2,61	2	1,34
	Gymnasiet	2,83	3	1,46
Hela läroboksmaterialet	Högstadiet	2,54	2	1,35
	Gymnasiet	2,79	3	1,53

Biologi	Högstadiet	2,60	2	1,31
	Gymnasiet	3,04	3	1,59
Fysik	Högstadiet	2,55	2	1,36
	Gymnasiet	2,79	3	1,52
Kemi	Högstadiet	2,57	2	1,27
	Gymnasiet	2,62	2	1,35
Geografi	Högstadiet	2,57	2	1,34
	Gymnasiet	2,86	3	1,38
Historia	Högstadiet	2,60	2	1,31
	Gymnasiet	2,76	3	1,43
Samhällskunskap	Högstadiet	2,70	2	1,41
	Gymnasiet	2,83	3	1,57

Tabell E.24: NP per mening i lärobokstexterna

E.3.2 Nominalfraslängd

Ämne/Ämnesfack	Stadium	Medel	Median	Std.av.
Naturvetenskap	Högstadiet	2,59	1	2,68
	Gymnasiet	3,00	2	3,26
Matematik	Högstadiet	2,62	2	2,54
	Gymnasiet	2,89	2	2,94
Samhällsvetenskap	Högstadiet	2,84	2	2,98
	Gymnasiet	3,12	2	3,36
Hela läroboksmaterialet	Högstadiet	2,70	2	2,79
	Gymnasiet	3,03	2	3,26

Biologi	Högstadiet	2,65	1	2,75
	Gymnasiet	3,15	2	3,60
Fysik	Högstadiet	2,53	1	2,68
	Gymnasiet	2,84	2	3,04
Kemi	Högstadiet	2,57	1	2,57
	Gymnasiet	2,98	2	3,04
Geografi	Högstadiet	3,16	2	3,21
	Gymnasiet	3,26	2	3,40
Historia	Högstadiet	2,70	2	2,78
	Gymnasiet	3,10	2	3,31
Samhällskunskap	Högstadiet	2,67	1	2,92
	Gymnasiet	3,00	2	3,36

Tabell E.25: NP-längder i lärobokstexterna

E.4 Delstudie: typiska meningar

E.4.1 Parsningsdjup

Ämne/Ämnesfack	Stadium	Medel	Median	Std.av.
Naturvetenskap	Högstadiet	5,36	5	2,03
	Gymnasiet	6,09	6	2,35
Samhällsvetenskap	Högstadiet	5,65	5	2,19
	Gymnasiet	6,14	6	2,32
Matematik	Högstadiet	4,96	5	1,95
	Gymnasiet	5,30	5	2,22
Hela läroboksmaterialet	Högstadiet	5,43	5	2,10
	Gymnasiet	6,00	6	2,33

Biologi	Högstadiet	5,47	5	2,07
	Gymnasiet	6,57	6	2,51
Fysik	Högstadiet	5,32	5	2,06
	Gymnasiet	5,90	6	2,29
Kemi	Högstadiet	5,31	5	1,95
	Gymnasiet	5,79	6	2,14
Geografi	Högstadiet	5,71	5	2,16
	Gymnasiet	6,25	6	2,25
Historia	Högstadiet	5,48	5	2,09
	Gymnasiet	5,93	6	2,26
Samhällskunskap	Högstadiet	5,76	6	2,29
	Gymnasiet	6,23	6	2,43

Tabell E.26: Parsningsdjup i lärobokstexterna