

GÖTEBORGS UNIVERSITET
Institutionen för svenska språket

Om Sverige

En kvantitativ textanalys av ett material för samhällsorientering
med fokus på ordförråd och läsbarhet ur ett andraspråksperspektiv

Emma Johansson

Specialarbete, SSA136, 15hp

Ämne: Svenska som andraspråk

Termin: VT 2017

Handledare: Sofie Johansson

Sammandrag

Den här uppsatsen undersöker läsbarhet ur ett andraspråksperspektiv med fokus på ordförråd. Studien tittar på om ordförrådet i boken *Om Sverige* är anpassat till nyanlända läsare, på hur stort ordförråd man behöver för att tillägna sig texten och på när man som inlärare kan förväntas ha tillförskaffat sig ett sådant ordförråd. I ett vidare perspektiv vill studien visa på hur man med hjälp av lexikal profil och ordvariation kan bedöma en texts lämplighetsgrad i relation till läsarens språkkunskaper

Med hjälp av de kvantitativa textanalysmåten *lexikal profil* och *ordvariation* undersöks bokens ordförråd med fokus på antal vanliga och ovanliga ord samt ordförrådets storlek. Denna undersökning sätts sedan i relation till den *Gemensamma europeiska referensramen för språks (CEFR)* riktlinjer för andraspråksinlärande. Studien tittar på hur stort ordförråd man kan förväntas ha på de olika CEFR-nivåerna samt på hur många timmar som en inlärare förväntas behöva studera ett språk för att nå de olika färdighetsnivåerna.

Undersökningen visar att det krävs ett betydande ordförråd även för att kunna läsa material som betecknas som lättläst och att det inte är möjligt för en nyanländ person vars språkkunskaper motsvarar CEFR-nivå A2 eller lägre att läsa boken *Om Sverige* utan mycket stor stöttning från lärare. Det behövs ett betydligt större ordförråd än man kan tillförskaffa sig på nybörjarnivå för att obehindrat kunna läsa och förstå texten. Enligt *Förordning om samhällsorientering* skall samhällsorienteringen i möjligaste mån ges på deltagarnas modersmål. Undersökningen visar att detta är mycket viktigt, då nyanlända på grund av sitt begränsade ordförråd inte kan ta del av det svenska materialet på ett tillfredsställande sätt.

Nyckelord: *lexikal profil, ordvariation, CEFR, kvantitativ analys, ordförråd, läsbarhet, läsförståelse, nyanlända, samhällsorientering*

Innehållsförteckning

1. Inledning	1
2. Syfte och frågeställningar	2
3. Teoretisk utgångspunkt och tidigare forskning	3
3.1. Teoretisk utgångspunkt.....	3
3.1.1. Språklagen, språkvård och språkpolitik.....	3
3.1.2. Förordningen om samhällsorientering.....	3
3.1.3. Gemensam europeisk referensram för språk	4
3.1.4. Språkvetenskaplig utgångspunkt	7
3.1.5. Kvantitativa läsbarhetsmått	9
3.1.6. Terminologi	12
3.2. Tidigare forskning	13
3.2.1. Läsbarhet.....	13
3.2.2. Ordförråd, Ordinlärning och ordfrekvenser.....	15
3.2.3. Gemensam referensram för språk.....	17
3.2.4. Tidigare forskning på ordförrådet i läromedelstexter och informationsmaterial från myndigheter	19
4. Material.....	20
4.1. Urval av text	20
4.1.1. Urvalskriterier.....	20
4.1.2. Omfattning.....	21
4.2. Referensmaterial.....	21
5. Metod.....	23
5.1. Bearbetning av texten	23
5.2. Analys.....	23
5.2.1. Lexikal profil	23
5.2.2. Jämförelse med CEFR skalor	24
5.2.3. Ordvariationsindex (OVIX).....	24
5.2.4. Kvalitativ analys	25
6. Resultat	25
6.1. Om Sverige	25

6.1.1.	Lexikal profil	25
6.1.2.	Omräkning till lemman.....	26
6.1.3.	Jämförelse med CEFR-nivåer.....	26
6.1.4.	Ordvariation.....	27
6.1.5.	Kategorisering av ord	28
6.2.	Jämförelse med referenstexter	28
6.2.1.	Jämförelse täckningsgrad.....	28
6.2.2.	Jämförelse vanliga och ovanliga ord	30
6.2.3.	Jämförelse ordvariation	31
7.	Diskussion och slutsatser	32
8.	Förslag till vidare forskning.....	35
9.	Litteraturförteckning.....	37
	Bilaga 1: Referensmaterial	41

1. Inledning

I den här uppsatsen undersöks en texts läsbarhet ur ett andraspråksperspektiv med hjälp av kvantitativ textanalys. Den svenskspråkiga versionen av boken *Om Sverige – samhällsorientering på svenska* (Göteborgs Stad och Länsstyrelsen Västra Götalands län 2015 [www]), undersöks med hjälp av lexikal profil och ordvariation med fokus på textens ordförråd och begriplighet ur ett L2-perspektiv. Studien tittar på om ordförrådet i *Om Sverige* är anpassat till nyanlända läsare, på hur stort ordförråd man behöver för att tillägna sig texten och på när man som inlärare kan förväntas ha tillförskaffat sig ett sådant ordförråd. I ett vidare perspektiv vill studien visa på hur man med hjälp av lexikal profil och ordvariation kan bedöma en texts lämplighetsgrad i relation till läsarens språkkunskaper

En läsares förståelse av en text är beroende både av textens egenskaper och av läsarens förmåga och språkkunskaper. Till textegenskaper som kan undersökas ur ett läsbarhetsperspektiv räknas t.ex. syntax, lexikalt urval och ordfrekvens. Bland egenskaper hos läsaren som påverkar dennes möjligheter att förstå en text kan nämnas t.ex. tidigare kunskap om ämnet, lexikala kunskaper, bekantskap med språket, motivation och engagemang.

Att förfoga över en metod för att ta reda på hur väl en text är lämpad för personer med en viss språknivå har flera fördelar. För läroboksförfattare är det viktigt att veta om ordförrådet i materialet lämpar sig för den nivå det är tänkt att användas på. När det gäller information från myndigheter finns det lagar och förordningar som styr hur kommunikation ska utformas med hänsyn till människors språkliga behov. Förutom att det finns lagstiftning är det naturligtvis av intresse för den som vill nå ut med information att förfoga över ett sätt att mäta och kontrollera att den information man vill nå ut med är språkligt anpassad till målgruppen och därmed begripbar och fyller sin funktion.

Boken som undersöks i denna uppsats, *Om Sverige*, finns översatt till flera språk. Trots att samhällsorientering främst ges på deltagarnas modersmål är det intressant att undersöka den svenskspråkiga versionen ur ett L2-perspektiv. Hur stort ordförråd behöver man för att tillägna sig texten och när kan man som inlärare

förväntas ha tillförskaffat sig ett sådant ordförråd? Visar undersökningen att boken är anpassad för att kunna användas på svenska av nyanlända? Eventuellt kan svaren på dessa frågor stärka argumentet att samhällsorientering till nyanlända bör, eller kanske till och med måste ges på modersmålet.

2. Syfte och frågeställningar

Undersökningens syfte är att med hjälp av kvantitativ textanalys undersöka hur begriplig boken *Om Sverige* är ur ett L2-perspektiv och genom detta visa på hur man med hjälp av lexikal profil och ordvariation kan bedöma en texts lämplighetsgrad i relation till läsarens språkkunskaper. Den text som har valts ut syftar till att på lättläst svenska informera om Sverige och det svenska samhället och undersökningen kommer att titta närmare på om ordförrådet i texten ligger på en nivå som gör den lättillgänglig för nyanlända med begränsade språkkunskaper.

Forskningsfrågor:

1. Hur stort ordförråd behövs för att kunna tillägna sig texten i boken *Om Sverige* och när kan man som inlärare förväntas ha tillförskaffat sig ett sådant ordförråd?
2. Är ordförrådet i *Om Sverige* anpassat till målgruppen nyanlända läsare och är det ur ordkunskapsperspektiv möjligt för nyanlända invandrare att ta del av samhällsinformationen i boken på ett tillfredsställande sätt?
3. Är en analys av lexikal profil i relation till CEFR-skalan ett användbart verktyg för att undersöka en texts lämplighetsgrad i relation till läsarens språkkunskaper?

3. Teoretisk utgångspunkt och tidigare forskning

3.1. Teoretisk utgångspunkt

3.1.1. Språklagen, språkvård och språkpolitik

Enligt *Språklagen* (§ 11) ska språket i offentlig verksamhet vara vårdat, enkelt och begripligt (Språklag 2009:600 [www]).

Institutet för språk och folkminnen [www] arbetar bland annat med språkvård och språkpolitik och följer svenskans och minoritetsspråkens utveckling i Sverige. *Språkrådet*, som är en del av institutet för språk och folkminnen, ansvarar för språkvården i Sverige och arbetar bland annat med *klarspråk*. Klarspråk står för klara och tydliga myndighetstexter skrivna på ett vårdat, enkelt och begripligt språk. Klarspråk handlar ytterst om demokrati, att alla ska ha tillgång till och förstå vad som står i texter som skrivs av myndigheterna. Språkrådet arbetar med att främja klarspråk på myndigheter, kommuner, landsting, universitet, organisationer och företag. Språkrådet har givit ut *Myndigheternas skrivregler* (Institutet för språk och folkminnen 2014) som bland annat ger handledning i hur man skriver klarspråk.

Myndigheten för tillgängliga medier (MTM) [www] tillhör kulturdepartementet och har som uppdrag att vara ett nationellt kunskapscentrum för tillgängliga medier och arbeta för att alla ska ha tillgång till litteratur och samhällsinformation utifrån var och ens förutsättningar, oavsett läsförmåga eller funktionsnedsättning.

3.1.2. Förordningen om samhällsorientering

År 2010 kom *Förordning (2010:1138) om Samhällsorientering för vissa nyanlända invandrare* [www]. Enligt denna förordning ska varje nyanländ som omfattas av etableringslagen erbjudas samhällsorientering på minst 60 timmar. Genom *Lag (2013:156) om samhällsorientering för vissa nyanlända invandrare* [www], utökades målgruppen år 2013 målgruppen till att även omfatta vissa anhöriginvandrare.

Samhällsorienteringen ska ges i ett tidigt skede efter ankomsten till Sverige och därför ska den ges på ett språk som den nyanlände behärskar. I *Länsstyrelsernas rapport om samhällsorientering 2016* [www] uppgav 91% av kommunerna att de främst ger samhällsorientering på deltagarnas modersmål men det utesluter dock inte att en del av kommunerna använder sig av andra alternativ, såsom samhällsorientering med tolk eller på lätt svenska.

3.1.3. *Gemensam europeisk referensram för språk*

För att kunna undersöka hur lätt eller svår en text är att förstå för en läsare med ett begränsat ordförråd behöver man ett sätt att uppskatta hur stort ordförråd en L2-talare som befinner sig på en viss nivå i sin språkutveckling kan tänkas ha. Denna uppskattning kommer i den här undersökningen att göras med hjälp av CEFR-skalan. CEFR står för *Common European Framework of Reference for Languages* och översätts på svenska som *Gemensam europeisk referensram för språk*. I denna uppsats kommer den engelska förkortningen CEFR att användas.

Den gemensamma europeiska referensramen för språk innefattar sex nivåer vilka presenteras i tabell 1.

Tabell 1: CEFR-nivåer (Skolverket 2007 [www]).

Avancerad språkanvändare	C2	Kan med lätthet förstå praktiskt taget allt som han/hon hör eller läser. Kan sammanfatta information från olika muntliga eller skriftliga källor och därvid återge argument och redogörelser i en sammanhängande presentation. Kan uttrycka sig spontant, mycket flytande och precist samt därvid skilja mellan finare nyanser även i mer komplexa situationer.
	C1	Kan förstå ett brett urval av krävande, längre texter och urskilja underförstådda betydelser. Kan uttrycka sig flytande och spontant utan att i större utsträckning söka efter ord och uttryck. Kan använda språket flexibelt och effektivt för såväl sociala som studie- och arbetsrelaterade ändamål. Kan producera tydliga, välstrukturerade och detaljerade texter om komplexa ämnen. Kan på ett planerat sätt med olika medel organisera och binda samman en text till en sammanhängande helhet.
Självständig språkanvändare	B2	Kan förstå huvudinnehållet i komplexa texter om både konkreta och abstrakta ämnen, inklusive fackmässiga diskussioner inom det egna specialområdet. Kan delta i samtal så pass flytande och spontant att ett normalt umgänge med modersmålstalare blir fullt möjligt, utan ansträngning för någondera parten. Kan producera tydlig och detaljerad text inom ett brett fält av ämnen, förklara en ståndpunkt i ett aktuellt ämne samt framhålla såväl för- som nackdelar vid olika valmöjligheter.
	B1	Kan förstå huvudinnehållet i vad han/hon hört eller läst om välkända förhållanden som han/hon regelbundet möter i arbetet, i skolan, på fritiden, etc. när det presenteras på tydligt standardspråk. Kan hantera de flesta situationer som vanligtvis uppstår under resor i områden där språket talas. Kan producera enkla, sammanhängande texter om ämnen som är välkända eller av personligt intresse. Kan beskriva erfarenheter och händelser, berätta om drömmar, förhoppningar och framtidsplaner och kortfattat ge skäl för och förklaringar till åsikter och planer.
Användare på nybörjarnivå	A2	Kan förstå meningar och vanliga uttryck som rör omedelbara behov (t.ex. mycket grundläggande information om individ och familj, inköp, lokal geografi, sysselsättning). Kan kommunicera i enkla och rutinmässiga sammanhang som kräver ett enkelt och direkt utbyte av information. Kan i enkla termer beskriva delar av sin egen bakgrund, sin närmaste omgivning samt förhållanden inom områden som rör omedelbara behov.
	A1	Kan förstå och använda välkända, vardagliga uttryck och mycket grundläggande fraser som syftar till att möta konkreta behov. Kan presentera sig själv och andra samt ställa och besvara frågor om personliga förhållanden som var han/hon bor, personer han/hon känner och saker han/hon har. Kan samtala på ett enkelt sätt under förutsättning att samtalspartnern talar långsamt och tydligt och är beredd att hjälpa till.

© Europarådet

Enligt Skolverkets publikation *Gemensam referensram för språk: lärande undervisning och bedömning* (Skolverket 2007:145 [www]), är ordförrådets storlek, omfång samt behärskning av ordförrådet viktiga parametrar av språktillägnande och därmed viktiga för bedömningen av en språkinlärares färdighet och för planeringen av språkinläring och språkundervisning.

CEFR och ordförråd

För ordförråd och behärskning av ordförråd på de olika nivåerna anges följande färdighetsdeskriptorer som presenteras i tabell 2 och 3.

Tabell 2: CEFR färdighetsdeskriptorer ordförråd (Skolverket 2007 [www]).

	ORDFÖRRÅD
C2	Har ett mycket stort ordförråd, inbegripet vardagliga och idiomatiska uttryck, och vet vilka associationer de ger.
C1	Har ett stort ordförråd och kan då han/hon saknar ord kringgå detta genom att uttrycka sig på ett annat sätt. Söker inte alltför tydligt efter ord och uttryck eller använder sig av undvikandestrategier. Kan effektivt välja vardagliga och idiomatiska uttryck.
B2	Har ett stort ordförråd inom sitt område och när det gäller de flesta allmänna ämnen. Kan variera formuleringarna för att undvika alltför många upprepningar, men luckor i ordförrådet kan fortfarande leda till tvekan och omformuleringar.
B1	Har ett tillräckligt stort ordförråd för att uttrycka sig, om än ibland med hjälp av omformuleringar, om de flesta vardagliga ämnen, såsom familj, hobbyer, intressen, arbete, resor och aktuella händelser.
A2	Har ett tillräckligt stort ordförråd för att utföra rutinmässiga vardagshandlingar som rör kända situationer och ämnen.
	Har ett tillräckligt stort ordförråd för att kunna uttrycka grundläggande kommunikativa behov. Har ett tillräckligt stort ordförråd för att klara av enkla grundläggande behov.
A1	Har ett grundläggande förråd av enstaka ord och fraser som rör vissa konkreta situationer.

Tabell 3: CEFR färdighetsdeskriptorer behärskning av ordförråd. (Skolverket 2007 [www]).

	BEHÄRSKNING AV ORDFÖRRÅD
C2	Genomgående korrekt och lämplig användning av ordförrådet.
C1	Enstaka mindre misstag, men inga större fel med anknytning till ordförrådet.
B2	Generellt stor lexikal korrektet, men viss hopblandning och felaktiga ordval kan förekomma utan att det hindrar kommunikationen.
B1	Behärskar ett grundläggande ordförråd väl, men gör fortfarande grova fel när han/hon uttrycker mer komplexa tankegångar eller hanterar okända ämnen eller situationer.
A2	Behärskar ett begränsat ordförråd som rör konkreta vardagsbehov.
A1	Deskriptor saknas.

Skolverket skriver vidare att referensramens användare bör överväga och om lämpligt ange:

- hur stort ordförråd (d.v.s. antalet ord och fasta fraser) som inläraren behöver/ har förutsättningar att/ måste behärska;
- hur omfångsrikt ordförråd (d.v.s. vilka domäner, teman etc. som ingår) som inläraren behöver/ har förutsättningar att/ måste behärska;
- vilken behärskning av ordförrådet som inläraren behöver/har förutsättningar att/ måste ha (Skolverket 2007:145 [www]).

Referensramens färdighetsdeskriptorer anger inte specifikt hur många ord en inlärare bör kunna på varje CEFR-nivå, det har dock gjorts viss forskning på ordförråd i relation till CEFR-nivåer. Denna forskning presenteras nedan under rubriken

tidigare forskning. Färdighetsdeskriptorerna ger oss inte heller någon information om vad det innebär att kunna ett ord. Definitioner av detta presenteras nedan i avsnittet *språkvetenskaplig utgångspunkt*.

Hur lång tid tar det att nå de olika färdighetsnivåerna på CEFR-skalan?

Hur snabbt en inlärare gör framsteg beror på många faktorer såsom exponering för språket och kulturen, kunskaper i andra språk, studievana, motivation och så vidare. Därför är det inte möjligt att exakt fastställa antalet lektionstimmar som behövs för att nå de olika nivåerna. Som en fingervisning kan dock nämnas *Cambridge exams* (Cambridge English support site [www]) uppskattning. *Cambridge exams* uppskattar att det tar ungefär 200 lektionstimmar för en inlärare att nå nästa nivå på skalan. En uppskattning av antal lektionstimmar för att nå CEFR- nivå A2-C2 visas i tabell 4.

Tabell 4: Lektionstimmar per CEFR-nivå (Cambridge English support site [www]).

Common European Framework Guided Learning Hours (from beginner level)

CEFR Level	Cambridge English Exam	Number of Hours (approximate)
C2	<i>Cambridge English: Proficiency (CPE)</i>	1,000—1,200
C1	<i>Cambridge English: Advanced (CAE)</i>	700—800
B2	<i>Cambridge English: First (FCE)</i>	500—600
B1	<i>Cambridge English: Preliminary (PET)</i>	350—400
A2	<i>Cambridge English: Key (KET)</i>	180—200

CEFR- nivåer och Sfi

Skolverket uppger att de olika CEFR-nivåerna motsvarar följande nivåer inom Sfi-utbildningen: Kurs A motsvarar A1-/A1, Kurs B motsvarar A1/A2, Kurs C motsvarar A2/A2+ och Kurs D motsvarar B1/B1+ (Skolverket Kursplan för Sfi)

3.1.4. Språkvetenskaplig utgångspunkt

Förutsättningar för lärande

Gibbons (2013) skriver i sin bok *Stärk språket stärk lärandet* om vikten av stöttning (scaffolding) och om hur elever och lärare kan samarbeta för att eleven ska utveckla nya färdigheter och en högre förståelse och på egen hand kunna utföra uppgifter som

de tidigare behövde hjälp med. Gibbons framför sin modell för språk- och kunskapsutvecklande arbetssätt med stöd av Lev Vygotskys teorier om lärande. Vygotskys (1978) ser mänsklig utveckling som en social företeelse. Som en del av sin sociokulturella teori beskriver han zonen för närmaste utveckling. Denna kan beskrivas som avståndet eller den kognitiva klyftan mellan vad ett barn kan göra utan hjälp och vad ett barn kan göra med hjälp av en vuxen, eller vad en elev kan göra utan stöttning och med stöttning från läraren. Enligt Gibbons hävdar Vygotsky att det enda ”nyttiga” lärandet är det som ligger steget före, i nästa utvecklingsfas (Gibbons 2013:40). När det gäller förutsättningar för lärande genom textläsning menar Nation (2006:147) att det behövs 95-98% täckningsgrad, dvs. 95-98% av orden måste vara kända för läsaren för att man denne obehindrat och utan stöd ska kunna läsa och förstå en text. Om färre ord är kända måste därmed någon form av stöttning ske för att läsaren ska kunna tillägna sig texten.

Vad innebär det att kunna ett ord?

Inom språkforskning är det vanligt att dela upp ordförståelse i receptiv eller passiv och produktiv eller aktiv kunskap. Med receptiv kunskap menas hörförståelse och läsning medan man vid produktiv kunskap talar om skriv- och talfärdigheter. Nation (2001:24) påpekar dock att terminologin inte är helt korrekt eftersom att det finns en produktiv del även i receptiv ordförståelse, när vi lyssnar eller läser producerar vi mening. Ett annat sätt är att dela upp efter ordkunskapens storlek eller bredd och styrka eller djup. Med bredd menas antalet ord en inlärare kan och med djup menar man vad inläraren vet om dessa ord (Milton 2009:22). Ordförrådets bredd undersöks kvantitativt medan ordförrådets djup blir föremål för kvalitativa undersökningar. För att få en så heltäckande bild som möjligt bör man mäta både storleken och styrkan i ordförrådet. (Laufer et al 2004:202). Nation (2001:27) delar in ordkunskap i tre olika delar; kunskap om form, kunskap om betydelse och kunskap om användning. Dessa delar presenteras i tabell 5.

Tabell 5: *What is involved in knowing a word (Nation 2001:27)*

Form	spoken	R	What does the word sound like?
		P	How is the word produced?
	written	R	What does the word look like?
		P	How is the word written and spelled?
	word parts	R	What parts are recognisable in this word?
		P	What word parts are needed to express the meaning?
Meaning	form and meaning	R	What meaning does this word form signal?
		P	What word form can be used to express this meaning?
	concept and referents	R	What is included in the concept?
		P	What items can the concept refer to?
	associations	R	What other words does this make you think of?
		P	What other words could we use instead of this one?
Use	grammatical functions	R	In what patterns does the word occur?
		P	In what patterns must we use this word?
	collocations	R	What words or types of words occur with this one?
		P	What word or types of words must we use with this one?
	Constraints on use (register, frequency...)	R	Where, when, and how often would we expect to meet this word?
		P	Where, when and how often can we use this word?

Note: In column 3, R=receptive knowledge, P= productive knowledge.

3.1.5. *Kvantitativa läsbarhetsmått*

Det finns en rad metoder för kvantitativ textanalys, nedan presenteras några av de vanligaste:

I Sverige har man sedan 1968 förlitat sig på **Läsbarhetsindex** (LIX) som ett mått på läsbarhet. LIX utvecklades av Björnsson (1968) och mäter längden på meningar och ord i en text. Läsbarhetsforskning har dock visat att det finns en rad andra egenskaper som bidrar till komplexiteten och därmed läsbarheten i texter (Heimann Mühlenbock 2013:5) LIX räknas ut genom att man delar antalet ord med antalet meningar för att få fram den genomsnittliga meningslängden. Sedan räknas antal långa ord ut och när man adderar genomsnittlig meningslängd och antalet

långa ord får man en texts LIX-värde. Tabell 6 visar LIX-värde för olika texttyper, antal ord delat i antal meningar, antal långa ord samt tolkning.

Tabell 6: Referensvärden LIX (Melin & Lange 2000:166)

Texttyp	LIX	Ord/men	LO	tolkning
Barn- och ungdomsböcker	27	12	15	Mycket lätt
Skönlitteratur	33	15	18	lätt
Dags- och veckopress	39	14	25	medelsvår
Saklitteratur	47	18	27	svår
Facklitteratur	56	20	35	Mycket svår

Nominalkvot (NQ) är ett mått på informationspackning i texten och räknas ut genom att antal substantiv, prepositioner och particip i en text delas med antal pronomen, adverb och verb. Normalvärdet för NQ är 1.0. Där ligger morgontidningar och läroböcker för högstadiet (Melin & Lange 2000:48). Tabell 7 visar referensvärden för nominalkvot.

Tabell 7: Referensvärden NQ (Melin & Lange 2000:168)

Broschyr	1.19
Lärobok	1.18
Morgontidning	1.04
Kvällstidning	0.99
Debatt	0.89
Veckotidning	0.85
Gymnasistprosa	0.72
Talspråk	0.25

Type/token ratio (TTR) mäter variation i ordförrådet genom att mäta antalet unika ord (types) i förhållande till det totala antalet ord (tokens) i en text. Detta görs genom att antal unika ord delas med det totala antalet ord (Heimann Mühlenbock 2013:35).

I den här uppsatsen begränsas undersökningen till följande två metoder:

Lexikal profil (Lexical frequency profiles) utvecklades av Laufer och Nation (1995) som ett sätt att undersöka antalet ordfamiljer på en viss frekvensnivå i en skriven text (Nation 2001:179). På svenska talar man om frekvensbaserad lexikal analys och med hjälp av denna analys kan man analysera storleken på ordförrådet i texter. Denna kunskap kan användas både för att mäta ordförrådet i en elevtext och få en bild av storleken på författarens ordförråd, och som i denna undersökning, för att uppskatta hur lätt eller svår en text är att förstå för en person med begränsat ordförråd. Lindberg (2007:46) skriver att detta verktyg har validerats i förhållande till andra lexikala test och uppvisat hög korrelation med etablerade mått på lexikal kunskap.

Ordvariationsindex (OVIX) utarbetades av Hultman och Westman (1977) och är ett vanligt mått för att mäta ordvariation inom svensk språkforskning. Likt TTR (type/token ratio) visar OVIX antalet unika ord i förhållande till det totala antalet ord i en text, till skillnad från TTR är OVIX dock konstruerat så att det kompenserar för textens längd. Ett högt OVIX gör texten mer svårläst då det finns många olika ord i texten. Ett OVIX värde på över 70 innebär en hög variation medan ett värde på under 60 innebär ett väldigt enkelt ordförråd. (Heimann Mühlenbock 2013:35). Referensvärden för OVIX presenteras i tabell 8.

Tabell 8: Referensvärden ordvariationsindex (Melin & Lange 2000:167)

Textgrupp	ovix
Morgontidning	89.3
Kvällstidning	85.8
Veckotidning	85.1
Lärobok	78.6
Broschyr	74.3
Debatt	72.9
Reportage	87.7
Bruksprosa	79.5

Det finns ett flertal datorprogram för att beräkna värden för textmått, i den här studien används programmet *AntWordProfiler* för den lexika frekvensprofilen medan ordvariation räknas ut med hjälp av tjänsten *lix.se*, som även kan användas för att beräkna en texts läsbarhetsindex. Dessa program presenteras närmare under *Metod*.

3.1.6. Terminologi

När man talar om ord och ordförråd är det viktigt att man tänker på att ord kan definieras på olika vis. Nedan följer en förteckning av några vanliga definitioner av ord:

Graford är den term som brukar användas om man med ord menar ”alla enheter i en text som förekommer mellan två mellanrum”. (Allén et al. 1989:190).

Med begreppet **lexikalisk enhet** eller **lexem** menas ett språkligt tecken som är lagrat som en enhet i minnet, till exempel *bok*, *röd*, *läsa*. Dessa lexikaliska enheter utgör de grundläggande enheterna i ett språks ordförråd vilka bildar grundvalen för olika ordformer, det vill säga grammatiska varianter som *boken*, *rödare*, *läste*. Olika böjningsformer av ett ord, till exempel *läser* och *läste*, räknas också som en enda lexikalisk enhet (Enström 2010:16). Svensén (2004:117) beskriver lexem som ”en lexikalisk enhet bestående av en viss, avgränsad betydelse på innehållssidan och en viss form på uttryckssidan”.

Token definieras på engelska av Dee Gardner som ”simply the frequency (number of occurrences) of types or other word units within a text, set of data, conversation, and so forth” (Gardner 2013:31). Antalet tokens i en text är samma som det totala antalet ord i texten och därmed också samma som antalet graford.

Type definieras av Dee Gardner som ”one or more contiguous letters of the roman alphabet that form a distinct word” (Gardner 2013:31). Antalet types i en text är samma som antalet unika ord i texten.

Den engelska definitionen av **lemma** lyder: ”Another form-based concept of ‘word’ is the lemma (plural lemmas), which essentially assumes that different types are considered the same word if they have the same base form, if they are from the same part of speech (all are nouns, all are verbs, etc.) and if they vary only in their spelling for grammatical purposes. This relationship is considered to be one of

inflection” (Gardner 2012:32). I den svenska definitionen av lemma skiljer sig ett lemma åt både i ord och i böjningsform. Man måste här ta hänsyn till homonymi och polysemi. Frågan är om identiska lexem med olika betydelser ska betraktas som representanter för olika betydelser hos ”samma ord” (polysemi) eller om de ska betraktas som två olika ord (homonymi). Svensén (2004:118) skriver att den traditionella metoden bygger på språkhistoriska kriterier där lexem med samma form men olika uttryck behandlas som homonymer. Med detta synsätt är homonyma ord två olika lemman.

Ordfamilj eller **word family** definieras på engelska av Dee Gardner som ”word family expands the definition of ‘word’ to include not only types related by inflection (as in lemmas) but also types related by derivation – i.e. same base form but with suffixes and prefixes that change parts of speech and/or meaning (Gardner 2013:33). Ordfamiljer inkluderar alltså olika ordklasser medan lemman består av olika former av ord inom samma ordklass.

3.2. Tidigare forskning

Forskare har försökt besvara hur många ord man behöver kunna för att läsa och förstå en text, vilka ord man behöver kunna och när under inlärningsprocessen man som L2 inlärare lär sig dessa ord. Nedan redogörs för forskning inom detta område.

Inledningsvis redogörs för läsbarhetsforskning i ett bredare perspektiv, därefter för ord och ordinläring med fokus på ordfrekvenser. Därefter presenteras forskning på ordförråd och CEFR-skalan. Vidare följer en sammanfattning av tidigare forskning om myndighetstexter och läromedelstexter med fokus på ord och läsbarhet.

3.2.1. *Läsbarhet*

Reichenberg (2000:17) skriver i sin avhandling *Röst och kausalitet i lärobokstexter* att beteckningen läsbarhet vanligen används för att beteckna egenskaper i texter som har med deras lätlästhets och svårlästhets att göra. Ibland har termen även används som ett sammanfattande namn för innehållsliga, typografiska och språkliga aspekter, men oftast har den använts för enbart språkliga aspekter. Björnsson (1979:9) ger

följande definition av läsbarhet: ”Summan av sådana språkliga egenskaper hos en text, vilka gör den mer eller mindre svårtillgänglig för läsaren”.

Forskning om läsbarhet kan delas upp i två huvudgrenar: Den del som syftar till att kunna kvantifiera läsbarheten och den del som syftar till att utveckla metoder för att undersöka den mer kvalitativa förståelsen. I den första av dessa grenar är man ute efter ett så tillförlitligt mätvärde som möjligt på läsbarhet, i det andra letar man efter en djupare förklaring till varför en text är lätt- eller svårläst (Reichenberg 2000:17-18).

För att läsning ska vara givande behöver textens svårighetsgrad motsvara läsarens språkkunskaper. Läsbarhetsmått är ytliga mått på hur lätt en text är att förstå. Dessa mått har framtagits inom internationell läsbarhetsforskning under de senaste 100 åren (Heimann Mühlenbock 2013:26). I USA startade forskningen om läsbarhet på 1920 talet och hade sin höjdpunkt 1930-1960 (Heimann Mühlenbock 2013:36). Flertalet läsbarhetsformler utvecklades enligt Heimann Mühlenbock under denna tid. (Dale & Chall, 1948; Flesch, 1951; Fry, 1968; Gunning, 1968; Spache, 1953). De flesta läsbarhetsformler mäter kvantitativa aspekter såsom ordförråd och syntax. Data från dessa former kan användas för att kombinera text med läsarens språknivå och läsförmåga (Oakland and Lane 2009: 244).

Globalt sett har forskning om läsbarhet främst handlat om den stilistiska svårigheten i skriven engelska. På svenska har läsbarhetsforskning främst varit av intresse för pedagoger och lärare men i takt med den ökade immigrationen har det vuxit fram en ökad efterfrågan på förenklade texter och ett fokus på att göra information mer lättåtkomlig (Heimann Mühlenbock 2013:6)

Inom svensk läsbarhetsforskning är Björnsson (1968) pionjären och han är även upphovsmannen till LIX, som man i Sverige länge har förlitat sig på som ett mått på läsbarhet (Heimann Mühlenbock 2013:5,7).

Oakland och Lane (2009:248) menar att en läsares förståelse av en text är beroende både av textens egenskaper och av egenskaper hos läsaren. Bland läsarens egenskaper nämns tidigare kunskaper om ämnet, lexikala kunskaper, bekantskap med språket, motivation och engagemang. Bland textegenskaper nämns syntax, lexikal selektion och ordfrekvens. Att läsbarhet inte enbart beror på textaspekter

utan även på läsarens förmåga är viktigt att tänka på när man diskuterar en texts läsbarhet.

3.2.2. Ordförråd, Ordinlärning och ordfrekvenser

”Ord är språkets byggstenar och utan dem finns det inget språk” (Milton 2009:12)

Ordförrådet anses vara den enskilt viktigaste faktorn för skolframgång för den som studerar på sitt andraspråk och forskning har visat att förståelsen av läromedelstexter i hög grad påverkas av den enskilda orden i texten. En högre andel okända ord i en text påverkar läsförståelsen markant och minskar möjligheten att sluta sig till okända ord från kontexten, vilket generellt är svårare för andraspråksinlärare. (Lindberg 2007: 34).

Milton menar att en av fördelarna med att undersöka ordinlärning är att ordkunskap, åtminstone vid första anblicken, kan räknas och mätas på ett meningsfullt sätt. Det går att räkna antalet ord i en text och komma fram till att en text på 400 ord är dubbelt så lång som en text på 200 ord. Man kan säga att en inlärare som kan 2000 ord i ett främmande språk kan dubbelt så många ord som en inlärare som kan 1000 ord. Men det som vid första anblicken verkar enkelt att mäta är i själva verket förenat med en hel del svårigheter. Till exempel är det inte alltid självklart vad som är ett ord, och en enkel uppgift som att räkna antalet ord i en text kan därför resultera i flera olika svar (Milton 2009:15-16).

Med ord kan man mena både type och token och idag väljer forskare i en allt större utsträckning andra definitioner som räknar olika former av ett ord som en enhet (såsom lemma och ordfamilj). Anledningen till detta är hur ord böjs och härleds. Inlärare lär sig inte varje ord som en separat enhet. När man har lärt sig en form av ett ord kan man använda sig av regler för t.ex. pluralbildning, verbformer och adjektivets komparation och därmed få hela familjer av ord till sitt förfogande (Milton 2009:18-19).

Ordfrekvenser

En av de viktigaste skillnaderna på ord ur ett inlärarperspektiv är med vilken frekvens de förekommer. Detta är viktigt eftersom det bestämmer vilka ord som en inlärare troligtvis träffar på samt hur ofta de påträffas (Milton 2009:31). Det är allmänt antaget att det finns en stark relation emellan ett ords frekvens och hur troligt det är att en inlärare kommer i kontakt med ordet och lär sig det (Milton 2009:34). Milton menar att man inom ordinlärning kan urskilja ett system som är baserat på ordfrekvens. Forskning pekar på att ett språks mest frekventa ord tenderar att vara dem som är lättast att lära sig och de ord som L2 inlärare tillägnar sig snabbast tenderar att återfinnas i de i de mest frekventa banden (Milton 2009:51). Viktigt är att påpeka är att information om ett ords frekvens inte ger oss information om dess svårighet, vad det däremot ger oss information om är hur troligt det är att en inlärare stöter på ordet och får det repeterat för sig tillräckligt många gånger för att lära sig det (Milton 2009:38).

Hur stort ordförråd behövs för textförståelse?

En viktig fråga inom studier av hur stort ordförråd som behövs för att läsa en text är vilken täckningsgrad som behövs för obehindrad förståelse. Hur många ord måste man känna till för att förstå texten, hur många ord kan vara okända utan att det påverkar förståelsen?

Nation (2006:147) menar att det behövs 95-98% täckningsgrad, dvs. 95-98% av orden måste vara kända för läsaren för att man denne obehindrat och utan stöd ska kunna läsa och förstå en text. Enligt Nation behöver en kännedom om 8000-9000 ordfamiljer för att uppnå denna nivå. Detta kan jämföras av att vuxna välutbildade modersmålstalare beräknas ha ett ordförråd på ca 20000 ordfamiljer. Det är viktigt att nämna att de siffror som Nation hänvisar till gäller forskning om engelska. Det finns inga motsvarande siffror för svenska. Som en jämförelse skriver Viberg (1993) enligt Lindberg (2007) att ordförrådets storlek hos enspråkiga barn vid skolstarten anses ligga på 8000 -10000 ord och sedan öka under skolåldern med ca 3000 ord per år under gynnsamma omständigheter.

3.2.3. *Gemensam referensram för språk*

Skolverkets riktlinjer för användning av gemensam referensram för språk har presenterats ovan under avsnittet *teoretisk bakgrund*. I dessa riktlinjer anges dock inte hur många ord en inlärare bör lära sig per nivå och inte heller för hur dessa ord bör väljas ut. Med tanke på Miltons forskning om att de mest frekventa orden tenderar att vara dem som är lättast att lära sig och som en L2-inlärare tillägnar sig snabbast är det rimligt att man kan ange hur många ord en man kan och bör lära sig för de olika nivåerna på samt inom vilka frekvensband dessa ord befinner sig. Viss forskning inom detta område har gjorts, både i Sverige och internationellt, främst när det gäller att ta fram frekvensbaserade ordlistor fördelade över de olika frekvensnivåerna.

I Sverige har sådan forskning skett genom EU-projektet KELLY (KEywords for Language Learning for Young and adults alike). KELLY genomfördes 2009-2012 och syftade till att göra ordlistor till 9 språk (svenska, engelska, norska, grekiska, italienska, polska, arabiska, kinesiska och ryska) och anpassa dessa efter CEFR-nivåerna (Johansson Kokkinakis, Volodina 2011).

Milton (2010) beskriver in sin artikel *The development of vocabulary breadth across the ce-fr-levels* försök att mäta ordförrådets bredd, det vill säga antalet ord som en inlärare kan i ett främmande språk på de olika CEFR-nivåerna. Milton menar att den gemensamma referensramen för språk indikerar att ordförrådets bredd är en användbar måttstock i beskrivningen av nivåerna och att inlärare kan förväntas få ett ökat ordförråd allt eftersom de rör sig uppåt genom de olika nivåerna. Undersökningar med Mearas och Miltons (2003) ordförrådstest XLex har visat att detta är fallet och har även indikerat hur stort ordförråd en inlärare förväntas ha på varje CEFR- nivå.

Mearas och Miltons XLex är ett ordfrekvenstest där orden räknas som lemmatiserade ord. XLex estimerar en inlärares kunskap om de 5000 mest frekventa orden i ett språk. Milton och Meara testade studenter som tog Cambridge examen på varje CEFR-nivå och estimerade deras ordförråd med hjälp av XLex. Dessa resultat återfinns i tabell 9. Milton och Alexio (2009) använde tre olika språkversioner av XLex och data från 500 inlärare av engelska, grekiska och franska som andra- eller främmande språk. Resultatet, som presenteras i tabell 10, visar på ett progressivt

ökande ordförråd för varje CEFR-nivå. Milton och Alexios forskning visar även indikationer på att ordförrådets bredd på varje nivå varierar mellan olika språk. Det är dock ej klart om detta beror på språken i sig eller på variationer i de korpusar som har används för att konstruera testerna av ordförrådets storlek.

Tabell 9 : Ordförråd (antal lemman) per CEFR-nivå (Milton 2010).

CEFR Levels	Cambridge exams	XLex (5000 max)
A1	Starters, Movers and Flyers	<1,500
A2	Kernel English Test	1,500 – 2,500
B1	Preliminary English Test	2,500 – 3,250
B2	First Certificate in English	3,250 – 3,750
C1	Cambridge Advanced English	3,750 – 4,500
C2	Cambridge Proficiency in English	4,500 – 5,000

Tabell 10: Ordförråd (antal lemman) per CEFR-nivå engelska, franska och grekiska (Milton 2010).

CEFR Level	EFL in Hungary	EFL in Greece	French in Spain	French in Greece	Greek in Greece
A1		1477 (580)	894 (604)	1125 (620)	1492 (705)
A2		2156 (664)	1700 (841)	1756 (398)	2237 (538)
B1	3135 (434)	3263 (434)	2194 (717)	2422 (517)	3338 (701)
B2	3668 (666)	3304 (666)	2450*	2630 (251)	4012 (415)
C1	4340 (471)	3690 (471)	2675 (643)	3212 (473)	
C2		4068 (261)	3721 (416)	3525 (883)	

* in this field there is only one student score and no SD can be calculated

3.2.4. Tidigare forskning på ordförrådet i läromedelstexter och informationsmaterial från myndigheter

Läromedelstexter

Vid studier av läromedelstexter dominerade länge studier av ordförråd och läsbarhet såväl i Sverige som internationellt men på senare år har studier av läromedelsspråk kännetecknats av en mer textlingvistik inriktning (Lindberg 2007:18-20).

Golden & Hvenekilde (1983) kartlade ordförrådet i 30 läromedel i historia, fysik och geografi för år 4-9 i Norge. Genom frekvensundersökningar kom de fram till att den mest intressanta ordgruppen utgjordes av ord som varken tillhörde den högfrekventa eller lågfrekventa gruppen. I stället pekade de ut en grupp av ord som är så vanliga att man kan räkna med att de behärskas av andraspråkselever men som ändå är centrala för förståelsen av texterna. Golden & Hvenekilde kallade dessa ord för icke-fackliga ord och menar att det är dessa ord som ofta vållar stort problem för andraspråkselever eftersom de är centrala för textförståelsen men många lärare inte tänker på att de kan vara okända och därför inte förklarar dem.

Inlärartexter har varit föremål för flera studier både på svenska och andra språk. Johansson Kokkinakis och Frändberg (2013) studerade högstadielävers användning av naturvetenskapligt språkbruk i kemiämnet i TIMSS. I sin studie studerade de hur språket används och realiseras i läromedelstexter och elevtexter i kemiämnet i årskurs 7-9. Som en del av denna studie jämfördes läromedelstexten med de 8000 vanligaste orden i det svenska språket med hjälp av en frekvensbandsanalys. De kom där fram till att de 1000 vanligaste orden utgjorde 60% av texten samt att de 21.3% som inte fanns med i frekvensordlistan som användes i studien till största delen utgjordes av tekniska och akademiska ord. Studier av texter i språkinlärningssammanhang har genomförts av flertalet andra forskare (Johansson Kokkinakis & Magnusson, 2011; Hultman & Westman, 1977; Lindberg & Johansson Kokkinakis, 2007, Nygård Larsson 2011).

Informationsmaterial från myndigheter.

Rickard Domeij diskuterar i *Webbtillgänglighet ur ett flerspråkigt perspektiv* (2011) hur tillgänglig informationen på myndigheters webbsidor är för nationella minoriteter och personer med annat modersmål än svenska. Domeij beskriver här att

ytterst lite gjorts för att förbättra situationen för personer med annat modersmål än svenska men att man börjar se exempel på att det håller på att utvecklas ett flerspråkigt tänkande kring kommunikationen mellan myndigheter och medborgare på internet.

Inom forskningsprojektet *Effekter av klarspråksarbete* [www] undersöks vilka effekter olika typer av klarspråksinsatser får på texter i offentlig förvaltning. Svenska myndigheter måste enligt lag skriva ”vårdat, enkelt och begripligt”. Andreas Nord och Lena Lind Palicki har inom ramen för ovan nämnda projekt undersökt hur arbetet för att genomföra detta ser ut (Nord & Palicki 2014).

4. Material

4.1. Urval av text

I den här uppsatsen genomförs en analys av boken *Om Sverige* som är ett material för samhällsorientering framtaget av Göteborgs Stad och Länsstyrelsen i Västra Götalands län. Boken kan laddas ner gratis på 12 olika språk. Den kan även beställas som tryckt material på svenska och arabiska. Första upplagan av boken kom ut 2010 efter att regeringen beslutat att landets kommuner skulle erbjuda samhällsorientering för nyanlända invandrare. Det material som analyseras i denna uppsats är femte upplagan från 2015. Boken beskrivs av Länsstyrelsen som ”knappt 200 lättlästa sidor om svensk historia och traditioner, om hur samhället är organiserat, om svenskt vardagsliv, och om vilka rättigheter och skyldigheter man har” och den främsta målgruppen är nyanlända invandrare (Sveriges Radio- En handbok om att leva i Sverige [www]).

4.1.1. Urvalskriterier

Materialet valdes enligt följande kriterier: det skulle vara ett informationsmaterial utgivet av en myndighet, såsom kommun eller landsting samt rikta sig till personer som är nya i Sverige och ha ett informativt syfte. Det skulle även vara ett relativt omfattande material som lämpar sig för en kvantitativ datorbaserad textanalys.

4.1.2. *Omfattning*

Den analyserade texten från boken *Om Sverige* omfattar 46832 graford och 5755 unika ord.

4.2. Referensmaterial

Som referensmaterial valdes kortare textavsnitt ut. Referensmaterialet presenteras i tabell 11.

Tabell 11: Referensmaterial.

Referenstext 1: *Informationsverige.se*

Referenstext 1 består av textavsnitt från informationsverige.se som är länsstyrelsernas gemensamma portal för samhällsinformation för invandrare. Den analyserade texten består av avsnitten Boende, Trivselregler och Fritid.

Urvalskriterier: Referenstext valdes för att en jämförelse med ytterligare ett informationsmaterial riktad till nyanlända invandrare skulle kunna göras.

Omfattning: 1460 graford och 562 unika ord.

Referenstext 2: *Göteborgs Stad - Information om skola, skolkort och skolskjuts*

Referenstext 2 består av information från Göteborgs stad riktad till medborgarna i kommunen. Den består av information om skola, skolkort och skolskjuts från Göteborgs stads hemsida.

Urvalskriterier: Referenstext 2 valdes som ett referensmaterial bestående av information av myndigheter riktad till allmänheten i kommunen (ej specifikt riktad till nyanlända eller personer med begränsade språkkunskaper).

Omfattning: 1501 graford och 399 unika ord.

Referenstext 3: *Nyhetsrapportering Världen/Sverige från GP/TT*

Referenstext 3 består av ett antal nyhetsnotiser från GP/TT. Det är ett urval av en dags förstasidesnyheter både från Sverige och Världen.

Urvalskriterier: Referenstext 3 valdes som referensmaterial med ett förväntat svårare språk.

Omfattning: 1589 graford och 696 unika ord.

Referenstext 4: *Samhällskunskapsbok åk 1-3*

Referenstext 4 består av ett utdrag ur en samhällskunskapsbok för åk 1-3.

Urvalskriterier: Referenstext 4 valdes som ett referensmaterial i form av läromedel riktat till barn i åk 1-3.

Omfattning: 1906 graford och 610 unika ord.

Referenstext 5: *Geografibok åk 4-6*

Referenstext 5 består av ett utdrag ur en geografibok för åk 4-6.

Urvalskriterier: Referenstext 5 valdes som ett referensmaterial i form av läromedel riktat till barn i åk 4-6.

Omfattning: 1833 graford och 633 unika ord.

Referenstext 6: *Samhällskunskapsbok åk 7-9*

Referenstext 6 består av ett utdrag ur en samhällskunskapsbok för åk 7-9.

Urvalskriterier: Referenstext 6 valdes som ett referensmaterial i form av läromedel riktat till barn i åk 7-9.

Omfattning: 1418 graford och 569 unika ord.

Referenstext 4, 5 och 6 består av utdrag ut läromedel. Detta material analyserades för att kunna göra en jämförelse av boken *Om Sverige* med läromedel för olika stadier. Läromedel i samhällskunskap valdes då det ligger nära ett material för samhällsorientering. Då en samhällskunskapsbok inte fanns att tillgå för åk 4-6 analyseras även en geografibok. Analysen av de tre läromedlen fyller även en annan funktion. Då man kan anta att det finns flest antal vanliga ord i boken för åk 1-3 och minst i boken för åk 7-9 kan ett förväntat resultat vid analysen av läromedlen användas för att validera testmetoden. På grund av copyrightskalet kommer texten från läromedlen inte att citeras.

5. Metod

Metoden i denna undersökning är en kvantitativ textanalys. I denna studie undersöks boken *Om Sverige* på ordnivå för att studera hur många vanliga respektive ovanliga ord texten innehåller. Resultaten av analysen jämförs sedan med tidigare forskning på ordförråd i relation till CEFR-skalor och en bedömning av vilken nivå en inlärare behöver ligga på för att kunna tillgodogöra sig texten görs. Detta sätts i relation till hur lång tid det förväntas ta att uppnå de olika CEFR-nivåerna. Dessutom analyseras bokens ordvariation för att bedöma storleken på ordförrådet i texten.

Någon omfattande kvalitativ analys av texten görs inte. Däremot görs en övergripande kategorisering av de ord utanför frekvenslistorna (dvs. ord som inte finns med bland de 5000 vanligaste svenska graforden) som förekommer flest gånger i boken. Detta görs för att få en uppfattning om vilka dess ord är och kunna diskutera möjligheten till vidare forskning för att undersöka hur man kan stötta läsaren genom att lyfta fram och presentera de vanligaste ovanliga orden. Efter analysen av boken *Om Sverige* analyseras även de sex referenstexterna med metoderna lexikalisk frekvensprofil och ordvariationsindex för att jämförelser ska kunna göras.

5.1. Bearbetning av texten

För att en lexikal profil med hjälp av datorprogrammet *AntWordProfiler* ska kunna göras måste texten som analyseras bestå av oformatterad text. Därför konverterades texten i boken *Om Sverige* först från PDF till textformat. Detta genomfördes med hjälp av datorprogrammet *AntFileConverter* [www].

5.2. Analys

5.2.1. Lexikal profil

En lexikal frekvensprofil (Lexical frequency profile) genomfördes med hjälp av datorprogrammet *AntWordProfiler* [www]. I denna undersökning analyserades texten och referenstexterna mot ordlistor bestående av de 5000 vanligaste svenska

graforden. Graforden i frekvenslistorna har extraherats från korpusen Bloggmix 2012 [www] och listorna används i undervisning på institutionen för svenska språket vid Göteborgs universitet. Analysen med lexikal profil tittar på följande aspekter i texten och referenstexterna:

- **Andel ord per frekvensband**

Hur många av orden i boken finns bland de 1000 vanligaste svenska graforden? Hur många finns med bland de 2000 vanligaste svenska graforden, hur många finns med bland de 3000 vanligaste osv.

- **Andel ord i frekvensband 1000 jämfört med andel ord som inte finns med i frekvenslistorna i procent.**

Jämförande analys av texten och referenstexterna för att se vilken text som har flest antal vanliga ord (ord i frekvenslista 1000) och vilken text som har flest ovanliga ord (ord som inte finns med i någon av frekvenslistorna och därmed inte bland 5000 av de vanligaste svenska orden).

- **Täckningsgrad i procent.**

Täckningsgraden visar hur stor procent av orden i de olika texterna som finns med bland 5000 av de vanligaste svenska graforden, det vill säga hur stor del av orden som täcks av frekvenslistorna.

5.2.2. *Jämförelse med CEFR skolor*

Resultatet av analysen med lexikal profil jämförs sedan med tidigare forskning på ordförråd i relation till CEFR-skolor för att bedöma på vilken nivå man bör ligga för att kunna tillgodogöra sig texten i boken *Om Sverige*. Detta sätts i relation till hur lång tid det förväntas ta att nå de olika nivåerna.

5.2.3. *Ordvariationsindex (OVIX)*

För att studera textens ordvariation har textmättet ”ordvariationsindex (OVIX)” använts. Denna analys har genomförts med hjälp av hemsidan [lix.se](http://www.lix.se) [www].

Analysen med ordvariationsindex tittar på följande aspekter i texten:

- **Antalet unika ord i förhållande till det totala antalet ord**

Ett högt OVIX-värde gör texten mer svårläst då det betyder att det finns många olika ord i texten. Ett lågt OVIX värde innebär ett enkelt ordförråd.

5.2.4. *Kvalitativ analys*

Utöver de ovan nämnda analyserna görs slutligen en översiktlig kvalitativ analys av ord som inte finns med bland 5000 av de vanligaste svenska graforden men som förekommer mer än 10 gånger i boken *Om Sverige*. Denna grupp består av totalt 87 ord som analyseras översiktligt genom att indelas i grupperna *samhällsfunktioner och myndigheter, förkortningar och namn, samhällsorienterade ord* och *allmänna abstrakta skriftspråkliga ord*.

6. Resultat

6.1. Om Sverige

6.1.1. *Lexikal profil*

Inledningsvis genomfördes den lexikala frekvensprofilen av boken *Om Sverige*. Tabell 12 visar resultatet av denna analys. Vi kan se att 66.98% av orden i texten finns med i frekvensband 1000, 7.16% av orden finns med i frekvensband 2000, 3.52% av orden finns med i frekvensband 3000, 2.54% av orden finns med i frekvensband 4000, 2.44% av orden finns med i frekvensband 5000 och 17.36% finns inte med i någon av frekvenslistorna och därmed inte bland de 5000 vanligaste orden.

Tabell 12: Andel ord per frekvensband i %.

Frekvens- band 1000	Frekvens- band 2000	Frekvens- band 3000	Frekvens- Band 4000	Frekvens- Band 5000	Ord utanför listorna
66.98 %	7.16%	3.52%	2.54%	2.44%	17.36%

6.1.2. Omräkning till lemman

Efter analysen summerades antal unika lemman per frekvensband i en lemmabaserad frekvensordlista. Denna omräkning gjordes för att en jämförelse med internationell forskning lättare skulle kunna göras. Beräkningen visade att graforden i listorna motsvarade 4266 lemman fördelade på följande sätt över de olika frekvensbanden. Tabell 13 visar resultatet av omräkningen till lemman.

Tabell 13: Omräkning till lemman.

Frekvensband	Antal lemman	Totalt
1000	844	844
2000	868	1712
3000	842	2554
4000	856	3410
5000	856	4266

6.1.3. Jämförelse med CEFR-nivåer

Resultatet av den lexikala frekvensprofilen jämförs sedan med tidigare forskning på hur många ord en inlärare förväntas kunna per CEFR-nivå och hur lång tid det förväntas ta att uppnå de olika nivåerna. Tabell 14 visar denna jämförelse.

Tabell 14: Jämförelse med CEFR-nivåer

CEFR -nivå	A1 (Sfi A/B)	A2 (Sfi B/C)	B1 (Sfi D)	B2	C1	C2
Antal lektionstimmar (Cambridge exams)		180-200	350-400	500-600	700-800	1000- 1200
Förväntad storlek på ordförråd (lemman) (Milton 2010)	<1500	1500- 2500	2750- 3250	3250- 3750	3750- 4500	4500- 5000
Jämförelse frekvensband (Fb)	Fb 1000 (844 lemman)	Fb 2000: (1712 lemman)		Fb:4000: (3410 lemman)	Fb 5000 (4266 lemman)	

		Fb 3000: (2554 lemman)		
Täckningsgrad Om Sverige %	66.98	74.14- 76-66	80.2	82.64

Enligt Miltons forskning kan man på A2 nivå förväntas ha ett ordförråd på 1500-2500 lemmen, tabell 14 visar att detta motsvarar någonstans mellan de 2000 och 3000 vanligaste svenska graforden. Om man räknar med att inläraren på A2 nivå har hunnit lära sig den 2000 vanligaste svenska graforden motsvarar dessa 1712 lemmen och 74.14% av orden i boken *Om Sverige*. Om man i stället räknar med att inläraren har hunnit lära sig de 3000 vanligaste svenska graforden motsvarar dessa 2552 lemmen och 77.66 % av orden i boken. Forskning visar dock att kunskap om 74.14 – 77.66% av orden är långt ifrån tillräckligt för att obehindrat kunna läsa och förstå en text. För kunskap om de 4266 vanligaste lemmatiserade orden (vilket motsvarar de 5000 vanligaste svenska graforden) behöver man enligt Milton komma upp i C1 nivå, vilket enligt *Cambridge exams* tar ca 700-800 lektionstimmar. Men även om en inlärare har kommit upp på C1 nivå och lärt sig alla de 5000 vanligaste svenska graforden i frekvenslistorna, motsvarar det bara 82.64% av orden i *Om Sverige*, även detta en för låg ordkunskapsnivå för att självständigt kunna läsa och förstå texten.

6.1.4. *Ordvariation*

Boken *Om Sverige* har ett ordvariationsindex på 61.7. Enligt Heimann Mühlenbock (2013:35) innebär ett värde på över 70 en hög variation medan ett värde på under 60 innebär ett väldigt enkelt ordförråd. Boken *Om Sverige* har alltså ett lågt ordvariationsindex. (Se: *Tabell 8: Referensvärden ordvariationsindex (Melin & Lange 2000:167)*)

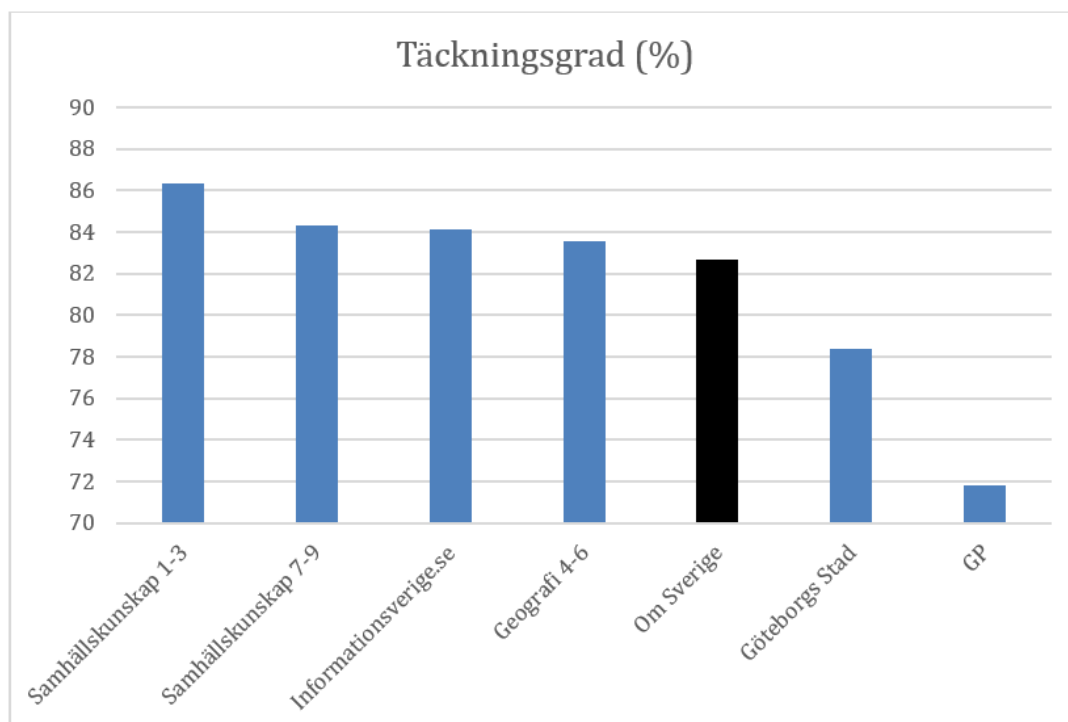
6.1.5. Kategorisering av ord

Analysen visar att 17.36% (8129 graford) av graforden i boken inte förekommer bland de 5000 vanligaste orden. Detta motsvarar 63.04% av alla unika ord i boken (3623 unika ord). För att få en överblick av vilka dessa ord är gjordes en översiktlig kvalitativ analys av de 87 ord utanför frekvenslistorna som förekommer mer än 10 gånger i texten. Bland dessa ord förekommer namn på samhällsfunktioner och myndigheter, exempelvis *Socialtjänsten* och *Skatteverket*. Det förekommer även en del förkortningar såsom *ID*, *FN*, *BVC*, *CSN*, *Sfi* och namn såsom *Norge*, *Finland*, *Stockholm* och *Gustav*. Vanligast förekommande är dock samhällsorienterade ord exempelvis *förening*, *funktionsnedsättning*, *invånarna*, *tjänstepension*, *folkbokförd*, *anställning* och *jämlikhet*. Vidare förekommer allmänna abstrakta skriftspråkliga ord såsom *betalas*, *ansvarar*, *allmänna*, *synsätt* och *sådant*.

6.2. Jämförelse med referenstexter

6.2.1. Jämförelse täckningsgrad

Figur 1 visar täckningsgraden för *Om Sverige* och referensmaterialen. Notera att Y-axeln har klippts av för att göra skillnader mellan texterna tydligare.

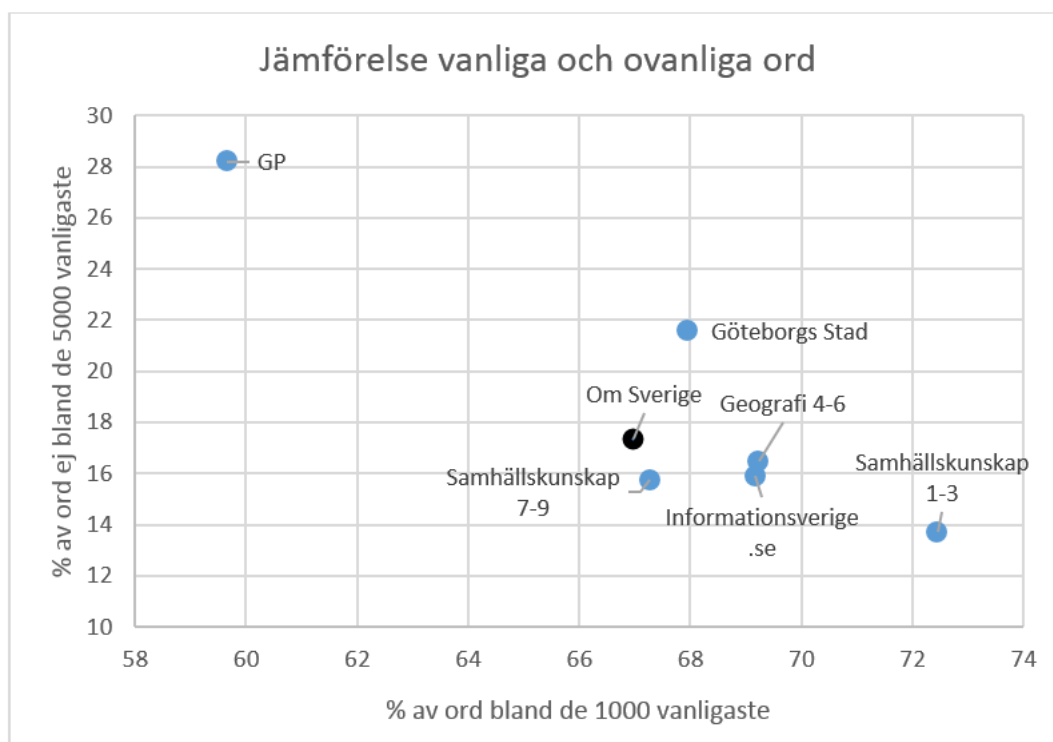


Figur 1: Täckningsgrad

Figur 1 visar täckningsgraden för de olika texterna. Vi kan se att *Om Sverige* har en täckningsgrad på 82.64%. Den andra texten som är riktad till nyanlända Informationsverige.se har en något högre täckningsgrad på 84.11%. Informationstexten från Göteborgs Stad har en något lägre täckningsgrad med 78.41%. Det är ett förväntat resultat att de texter som riktas specifikt till nyanlända har fler antal vanliga ord än den text som inte specifikt riktas till denna målgrupp. I jämförelse med texten från GP som har en täckningsgrad på 71.8% kan man konstatera att en morgontidningstext har betydligt fler antal ovanliga ord än informationsmaterial från myndigheter. I jämförelse med de tre läromedlen kan man se att samhällskunskapsboken för åk 1-3 har allra flest vanliga ord med 86.31%. Täckningsgraden i läromedlen för åk 4-6 (83.53%) och åk 7-9 (84.28%) ligger båda väldigt nära täckningsgraden i *Om Sverige*. Enligt denna jämförelse skulle en nyanländ invandrare, för att kunna läsa dessa texter, skulle behöva ha ett lika stort ordförråd som en elev på mellan- eller högstadiet.

6.2.2. Jämförelse vanliga och ovanliga ord

Figur 2 visar en jämförelse av andel vanliga och ovanliga ord i boken *Om Sverige* och referenstexterna. Diagrammet visar hur många ord i varje text som är med bland de 1000 vanligaste orden samt hur många ord i varje text som inte är med bland de 5000 vanligaste orden.



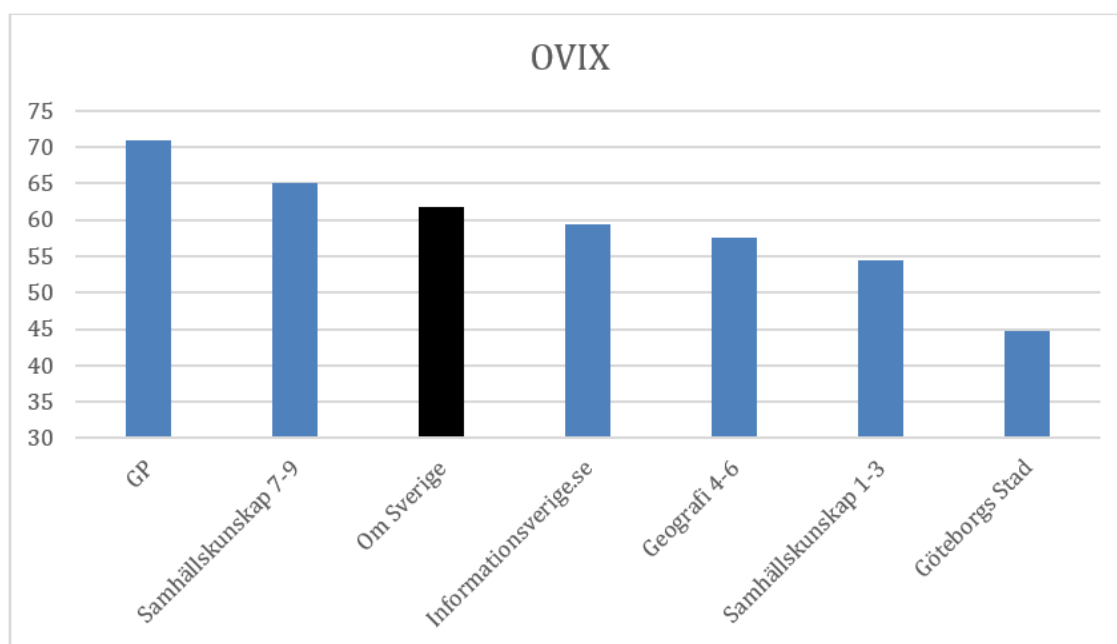
Figur 2: Andel vanliga och ovanliga ord, jämförelse med referenstexter.

I figur 2 jämförs vilka texter som har flest antal vanliga ord respektive vilka texter som har flest antal ovanliga ord. Om vi börjar med att titta på *Om Sverige* finns 66.98% orden i boken i frekvenslistan med de 1000 vanligaste svenska orden medan 17.36% av orden inte med i någon av frekvenslistorna. Informationsverige.se har något fler vanliga ord (69.18%) och något färre ovanliga ord (15.89%). Även informationstexten från Göteborgs Stad har en stor andel vanliga ord (67.95%) men den har ett högre antal ovanliga ord än de två texterna riktade specifikt till nyanlända invandrare (21.59%). Detta kan antas bero på att denna text är riktad till en bredare allmänhet och inte specifikt till invandrare. Den största avvikelser ser vi på texten från GP, som har ett betydligt färre antal vanliga ord (59.66%) och betydligt fler ovanliga ord (28.19%) än de andra texterna.

En jämförelse av de tre läromedlen visar att de ligger på en rak linje där åk 1-3 har flest antal vanliga ord (72.46%) och minst antal ovanliga ord (13.69%). Åk 4-6 och 7-9 ligger väldigt nära varandra med 69.23% (åk 4-6) respektive 67.28% (åk 7-9) vanliga ord och 16.48% (åk 4-6) respektive 15.73% (åk 7-9) ovanliga ord. Att texten för åk 4-6 innehåller något mer ovanliga ord än den för åk 7-9 kan troligtvis förklaras med att den är från en geografibok och innehåller många namn som inte finns med på frekvenslistorna. Förutom detta visar jämförelsen av läromedlen förväntat resultat då texten för åk 1-3 har flest antal vanliga ord, texten för åk 7-9 har minst antal vanliga ord och texten för åk 4-6 befinner sig mitt i mellan de andra två texterna.

6.2.3. Jämförelse ordvariation

Figur 3 jämför *Om Sveriges* OVIX-värde med de olika referenstexterna.



Figur 3: ordvariation

Om Sverige har ett OVIX-värde på 61.7 vilket betyder att boken har en lägre ordvariation än samhällskunskapsboken för åk 7-9 men en högre ordvariation än Geografiboken för åk 4-6. Analysen av ordvariationsindex visar precis som den lexikaliska frekvensprofilen att ordförrådet i *Om Sverige* ligger på en nivå som kan jämföras med läromedel för mellan- eller högstadiet. Informationsverige.se ligger

strax under *Om Sverige* med ett OVIX värde på 59.46. Intressant nog har informationstexten från Göteborgs stad en låg ordvariation (44.77) i förhållande till alla de andra texterna. Det är möjligt att denna text har bearbetats för att göra den mer lättläst, och att det låga OVIX värdet är ett resultat av detta. Texten från GP har ett OVIX värde på 70.98. Med ett värde på över 70 har texten en hög ordvariation. (Se: *Tabell 8: Referensvärden ordvariationsindex (Melin & Lange 2000:167)*).

7. Diskussion och slutsatser

Denna studie har med hjälp av kvantitativ analys undersökt hur begriplig boken *Om Sverige* är ur ett L2-perspektiv. Studien har tittat på om ordförrådet i boken *Om Sverige* är anpassat till nyanlända läsare, på hur stort ordförråd man behöver för att tillägna sig texten samt när man som inlärare kan förväntas ha ett sådant ordförråd. I ett vidare perspektiv har studien tittat på om kvantitativ analys av lexikal profil och ordvariation är ett lämpligt verktyg för att bedöma en texts lämplighetsgrad i relation till läsarens språkkunskaper. De tre forskningsfrågorna diskuteras och besvaras nedan.

1. Hur stort ordförråd behövs för att kunna tillägna sig texten i boken *Om Sverige* och när kan man som inlärare förväntas ha tillförskaffat sig ett sådant ordförråd?

Enligt Milton kan man inom ordinlärning urskilja ett system som är baserat på ordfrekvens. Forskning pekar på att ett språks mest frekventa ord tenderar att vara de som är lättast att lära sig och de ord som L2 inlärare tillägnar sig snabbast tenderar att återfinnas i de i de mest frekventa banden (Milton 2009:51). Meara och Milton (2003) har uppskattat att en inlärare som nått nivå A2 har ett ordförråd på 1500-2500 lemmatiserade ord. *Cambridge exams* uppskattar att det tar ca 200 lektionstimmar att nå nivå A2. Det är dock viktigt att komma ihåg att det finns en stor individuell variation i hur snabbt man lär sig ett språk. Miltons listor består av lemman medan de svenska frekvenslistorna som har används i denna undersökning består av graford. Därför har det under arbetets gång visat sig vara svårt att genom jämförelse med

Miltons forskning dra klara slutsatser om hur många av de 5000 vanligaste svenska orden en inlärare på A2 nivå kan förväntas kunna. En omräkning av graforden till lemman har dock visat att de 5000 graford i frekvenslistorna som har använts för den lexikaliska frekvensprofilen i denna studie motsvarar 4266 lemman.

Miltons forskning visar att man på A2 nivå kan förväntas ha ett ordförråd på 1500-2500 lemman, vilket motsvarar någonstans mellan de 2000 och 3000 vanligaste svenska graforden. Om man räknar med att inläraren på A2 nivå har hunnit lära sig den 2000 vanligaste svenska graforden motsvarar dessa enligt denna omräkning 1712 lemman och 74.14% av orden i boken *Om Sverige*. Om man i stället räknar med att inläraren har hunnit lära sig de 3000 vanligaste svenska graforden motsvarar dessa ca 2552 lemman och 77.66 % av orden i boken. Forskning visar dock att kunskap om 74.14 – 77.66% av orden är långt ifrån tillräckligt för att obehindrat kunna läsa och förstå en text. Därmed är det klart att det för en nyanländ invandrare som har läst upp till A2-nivå inte är möjligt att genom självständig läsning tillägna sig informationen i boken *Om Sverige*.

För kunskap om de 5000 vanligaste svenska orden vilka motsvarar 4266 lemmatiserade ord behöver man enligt Milton komma upp i C1 nivå, vilket enligt *Cambridge exams* tar ca 700-800 lektionstimmar. Men även om en inlärare har kommit upp på C1 nivå och lärt sig alla de 5000 vanligaste svenska graforden i frekvenslistorna, motsvarar det bara 82.64% av orden i *Om Sverige*, även detta en för låg ordkunskapsnivå för att självständigt kunna läsa och förstå texten. Nation (2006) menar att det behövs kunskap om 95-98% av orden i en text för att kunna läsa den obehindrat utan stöd. Enligt Nation behövs ett ordförråd på ca 9000 ordfamiljer för att komma upp i den nivån. 9000 ordfamiljer motsvarar ungefär 22500 lemman. Även om de olika definitionerna av ord vållat svårigheter och dessa jämförelser bara kan bli uppskattningar är det klart att det krävs ett betydligt större ordförråd än de ca 1500-2500 lemman en inlärare på A2 nivå förväntas kunna för att komma upp i den täckningsgraden för boken *Om Sverige*.

Undersökningen har därmed visat att täckningsgraden i *Om Sverige* är för liten för en inlärare på A2 nivå men även att den är för liten för att en inlärare som befinner sig på en betydligt högre nivå självständigt och problemfritt ska kunna tillägna sig texten. Jämförelsen med referenstexternas täckningsgrad (*Figur 1:*

Täckningsgrad), visar att täckningsgraden i *Om Sverige* ligger nära täckningsgraden i läromedel för å 4-6 och 7-9. Enligt denna jämförelse skulle en nyanländ invandrare, för att kunna läsa *Om Sverige*, behöva ha ett lika stort ordförråd som en elev på mellan- eller högstadiet. Enligt Viberg (1993) anses ordförrådets storlek hos enspråkiga barn vid skolstarten ligga på 8000-10000 ord och sedan öka under skolåldern med ca 3000 ord per år under gynnsamma omständigheter. Det är därför inte rimligt att en nyanländ invandrare som nått A2-nivå skulle kunna ha ett lika stort ordförråd som en enspråkig elev på dessa stadier. Textens ordvariationsindex ligger på 61.7. Enligt Heimann Mühlenbock (2013:35) innebär ett värde på över 70 en hög variation medan ett värde på under 60 innebär ett väldigt enkelt ordförråd. Detta betyder att bokens ordvariationsindex är lågt, och vi kan därmed dra slutsatsen att texten försvåras främst av att den innehåller många ovanliga ord och inte av att de innehåller många olika ord.

2. Är ordförrådet i *Om Sverige* anpassat till målgruppen nyanlända läsare och är det ur ordkunskapsperspektiv möjligt för nyanlända invandrare att ta del av samhällsinformationen i boken på ett tillfredsställande sätt?

Analysen har visat att det inte är möjligt för en nyanländ person vars språkkunskaper motsvarar A2 nivå eller lägre att läsa boken *Om Sverige* och ta till sig innehållet utan mycket stor stöttning från lärare. Stöttning är en viktig del av lärarens roll, men enligt Vygotsky bör lärandet ske inom den närmaste utvecklingsfasen. Denna studie har inte undersökt hur stöttning skulle kunna underlätta förståelsen av texten eller hur mycket stöttning som behövs för att en nyanländ ska ha möjlighet att förstå innehållet i texten. Den stora skillnaden mellan bokens ordförråd och en nyanländs förväntade ordförråd gör det dock rimligt att anta att mycket stor stöttning skulle behövas och att lärandet i detta fall inte skulle ske inom den närmaste utvecklingsfasen. Enligt *Förordning om samhällsorientering* ska samhällsorientering i möjligaste mån ges på deltagarnas modersmål. Undersökningen har visat att detta är mycket viktigt, då nyanlända på grund av sitt begränsade ordförråd inte kan ta del av det svenska materialet på ett tillfredsställande sätt. Det behövs ett betydligt större ordförråd än man kan tillförskaffa sig på nybörjarnivå för att obehindrat kunna läsa och förstå boken *Om*

Sverige. Vi kan därmed dra slutsatsen att ordförrådet i boken *Om Sverige* inte är anpassat till gruppen nyanlända läsare.

3. Är en analys av lexikal profil i relation till CEFR-skalan ett användbart verktyg för att undersöka en texts lämplighetsgrad i relation till läsarens språkkunskaper?

Undersökningen har visat på hur man med hjälp av lexikal profil och CEFR-skalan kan undersöka hur väl en text är lämpad för personer med en viss språknivå. Det har visat sig möjligt att göra intressanta observationer om en texts lämplighetsgrad i relation till läsarens språkkunskaper genom att jämföra storleken på ordförrådet i en text med förväntad storlek på inlärares ordförråd då de befinner sig på de olika CEFR-nivåerna. Undersökningen har även visat på att det finns vissa begränsningar i detta tillvägagångssätt; tiden en inlärare behöver för att nå en viss språknivå är mycket individuell, vilket kan göra det svårt att dra generella slutsatser. Jämförelser med tidigare forskning inom området har komplicerats av de olika definitionerna av *ord*. En kvantitativ studie som denna undersöker dessutom enbart ordförrådets bredd eller storlek och inte ordförrådets djup eller styrka. Enligt Laufer (2004) bör man mäta både storleken och styrkan i ordförrådet för att få en så heltäckande bild som möjligt. Detta skulle kunna göras genom att utveckla den översiktliga kvalitativa analys som har presenterats i resultatdelen. Förslag på hur detta skulle kunna göras presenteras i kapitel 8.

8. Förslag till vidare forskning

För att studera hur informationen i boken *Om Sverige* skulle kunna förenklas, skulle det ur ett ordkunskapsperspektiv vara intressant att kvalitativt undersöka och kategorisera de ord som faller utanför de 5000 vanligaste för att se vilka av dessa ord som skulle kunna bytas ut eller förklaras närmare för att underlätta läsningen. Läsning av texten skulle kunna underlättas genom att man tillhandahåller ordlistor/förklaringar av de ovanliga ord som är vanligt förekommande i texten. Den översiktliga kategoriseringen av de ord utanför frekvensbanden som förekommer flest gånger ger en fingervisning av vilka ord som skulle vara aktuella att ha med i

sådana listor men detta skulle kunna undersökas mer detaljerat. Analysen av de 87 ord utanför frekvenslistorna som förekommer fler än 10 gånger visade att dessa ord kan delas in i 4 huvudkategorier: Namn på samhällsfunktioner och myndigheter (*Skatteverket*), förkortningar och namn (*ID, Norge*); samhällsorienterade ord (*förening*) och allmänna abstrakta skriftspråkliga ord (*ansvarar, synsätt*). Egennamn och namn på länder borde inte vålla läsaren några större problem och man får förutsätta att förkortningar och namn på samhällsfunktioner förklaras i texten. Då återstår samhällsorienterade ord och allmänna abstrakta skriftspråkliga ord som de ord som utmärker sig som vanligt förekommande ovanliga ord som skulle kunna vara de ord som kommer innebära störst svårigheter för läsaren. Enligt Golden & Hvenekilde (1983) är de allmänna abstrakta skriftspråkliga orden (eller icke-fackliga) oftast de som vållar störst problem för läsaren. Man bör därför fundera på om dessa ord kan bytas ut eller förklaras för att göra texten mer lättläst och detta vore en intressant infallsvinkel för vidare forskning.

Jämförelsen med internationell forskning för att associera en viss storlek på ordförrådet med en viss CEFR-nivå har försvårats av att man speciellt i engelsk forskning ofta talar om lemman eller ordfamiljer medan frekvenslistorna i denna undersökning består av graford. Efter analysen gjordes en omräkning för att uppskatta antal unika lemman per frekvensband. Det skulle dock vara intressant att göra om samma analys med lemmatiserade svenska frekvenslistor och sedan jämföra resultaten. Det skulle också vara intressant att analysera och jämföra ordförrådet i flera hela läroböcker för svenska som andraspråk på de olika CEFR-nivåerna för att se om storleken på ordförrådet som undervisas där stämmer överens med de uppskattningar som finns angående förväntat ordförråd på de olika CEFR-nivåerna. Man skulle då kunna undersöka hur många av orden i läroböckerna som täcks av frekvenslistorna och inom vilket frekvensband de återfinns. Vidare skulle man kunna analysera inlärartexter producerade av inlärare som studerar på de olika CEFR-nivåerna för att undersöka hur ordförrådet i dessa texter utvecklas. Speciellt intressant skulle det vara att analysera texter av samma inlärare när denne befinner sig på de olika CEFR-nivåerna i en longitudinell studie.

9. Litteraturförteckning

- Allén, S et al. 1989. *Orden speglar samhället*. Stockholm: Allmänna förlaget
- Björnson, C-H. 1968. *Läsbarhet*. Stockholm: Liber.
- Cambridge English Support Site <<https://support.cambridgeenglish.org/hc/en-gb/articles/202838506-Guided-learning-hours>> Hämtat 2017-04-05
- Dale, E., & Chall, J. S. 1948. A formula for predicting readability. *Educational Research Bulletin*, 27, 37–54.
- Domeij, R. Webbtillgänglighet i flerspråkigt perspektiv. I: Domeij, R. m.fl. Språkteknologi för ökad tillgänglighet. Rapport från ett nordiskt seminarium. Linköping, 27–28 oktober 2010. Linköping University Electronic Press.
- Enström, I. 2010. *Ordens värd. Svenska ord- struktur och inläring*. Stockholm: Hallgren & Fallgren
- Flesch, R. 1951. *How to test readability*. New York: Harper & Brothers.
- Fry, E. 1968. A readability formula that saves time. *Journal of Reading*, 11, 513–516.
- Gardner, D. 2013. *Exploring Vocabulary Language in action*. New York: Routledge.
- Gibbons, P. 2013. *Stärk språket stärk lärandet*. Stockholm: Hallgren & Fallgren.
- Golden, A. & Hveneklide, A. 1983. *Rapport fra prosjektet laerebokspråk. Sentret for språkpedagogikk*, Universitetet i Oslo.
- Gunning, R. 1968. *The technique of clear writing*. New York: McGraw-Hill.
- Göteborgs Stad och Länsstyrelsen Västra Götalands län 2015. *Om Sverige-samhällsorientering på svenska*. 5 uppl. Göteborg: Göteborgs Stad <<https://www.informationsverige.se/Svenska/Samhalle/Sidor/Boken-Om-Sverige-p%C3%A5-flera-spr%C3%A5k.aspx>> Hämtat 2017-02-18.
- Heimann Mühlenbock, K. 2013. *I see what you mean: assessing readability for specific target groups*. Ph.D. diss., Data linguistica. No. 24, Institutionen för svenska språket, Göteborgs universitet.
- Hultman, Tor G. & Westman, Margareta 1977. *Gymnasistsvenska*. Lund: Liber Läromedel.
- Institutet för språk och folkminnen <<http://www.sprakochfolkminnen.se/sprak.html>> Hämtat 2017:04.29

- Institutet för språk och folkminnen 2014. *Myndigheternas skrivregler*. 8 uppl. Stockholm: Fritzes/Norstedts Juridik <<http://www.sprakochfolkminnen.se/om-oss/publikationer/institutets-publikationer/sprakvard/2016-09-14-myndigheternas-skrivregler.html>> Hämtat 2017-04-29
- Institutet för språk och folkminnen *Klarspråk- för en vårdad enkel och begriplig svenska*<<http://www.sprakochfolkminnen.se/download/18.1bc6136f1422723e4bf53fd/1398151025612/folder-klarsprak-2013.pdf>> Hämtat 2017-04-29
- Johansson Kokkinakis, Sofie & Magnusson Ulrika 2010: Computer based quantitative methods applied to first and second language student writing. I: *Young Urban Swedish Variation and change in multilingual settings*. s 105-124. Red. Roger Källström & Inger Lindberg. Göteborgsstudier i nordisk språkvetenskap 14. <http://hdl.handle.net/2077/26570>
- Johansson Kokkinakis, Sofie & Elena Volodina 2011. *Corpus-based approaches for the creation of a frequency based vocabulary list in the EU project KELLY – issues on reliability, validity and coverage*.
- Johansson Kokkinakis, S. & Frändberg, B. 2013. Högstadielära användning av naturvetenskapligt språkbruk i kemiämnet i TIMSS. *Utbildning och Demokrati*, 22(3), 53-68.
- Laufer B., Elder C., Hill K., Congdon P. 2004. Size and strength: do we need both to measure vocabulary knowledge? *Language Testing*, vol. 21, no. 2, Pp. 202-226
- Laufer, B. and Nation, I.S.P. 1995. Lexical richness in L2 written production: Can it be measured? *Applied Linguistics* 16, 3: 307-322.
- Lindberg, Inger 2007. Forskning om läromedelspråk och ordförrådsutveckling. I: Lindberg, Inger & Johansson Kokkinakis, Sofie (red.), 2007: *OrdiL – en korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år*. Rosa-rapport 8. Institutionen för svenska som andraspråk. Göteborgs universitet.
- Länsstyrelsen i Jönköpings län 2016. *Samhällsorientering för nyanlända. Länsstyrelsernas gemensamma redovisning 2016*. <<http://extra.lansstyrelsen.se/integration/SiteCollectionDocuments/rapporter/Samhallsorientering-for-nyanlanda-2016.pdf>> Hämtat 2017-04-29
- Meara, P., & Milton, J. 2003. *X_Lex, the Swansea Levels Test*. Newbury: Express.

- Melin Lars & Sven Lange 2000 [1986]. *Att analysera text. Stilanalys med exempel*. 3. Uppl. Lund: Studentlitteratur
- Milton, J. 2009. *Measuring Second Language Vocabulary Acquisition*. Bristol: Multilingual Matters.
- Milton, J., & Alexiou, T. 2009. Vocabulary size and the Common European Framework of Reference for Languages. In B. Richards et al. (Eds.), *Vocabulary studies in first and second language acquisition* (pp. 194–211). Basingstoke: Palgrave.
- Milton J. 2010. The development of vocabulary breadth across the CEFR levels. In Bartning I., Martin M., Vedder I. (Eds.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research* (pp. 211–232). Eurosla Monographs Series, 1
- Myndigheten för tillgängliga medier <<http://www.mtm.se/om-oss/>> Hämtat 2017-04-29
- Nation, I. S. P. 1990. Teaching and learning vocabulary. Boston: Heinle & Heinle
- Nation, I.S.P. 2001. *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, I. S. P. 2006. How large a vocabulary is needed for reading and listening? *The Canadian Modern Language Review*, 63, 59–82.
- Nord, Andreas & Lena Lind Palicki 2014. ”Ni ska skriva enkelt – det står i lagen”. Legitimeringar av klarspråk i en utbildning för myndighetsskribenter. I: *Språk i undervisning*. Rapport från ASLA:s vårsymposium Linköping 11-12 maj 2012. / redaktörer: Christina Rosén, Per Simfors, Ann-Kari Sundberg. Reviderad upplaga 2014. ASLA:s skriftserie, Paper i proceeding 2014
- Nygård Larsson, Pia 2011. *Biologiämnets texter: text, språk och lärande i en språkligt heterogen gymnasieklass*. Malmö: Lärarutbildningen, Malmö högskola.
- Oakland, T., & Lane, H. B. (2004). Language, Reading, and Readability Formulas: Implications for Developing and Adapting Tests. *International Journal of Testing*, 4 (3), 239-252.
- Reichenberg, Monica 2000. *Röst och kausalitet i lärobokstexter. En studie av elevers förståelse av olika textversioner*. Acta Universitatis Gothoburgensis.
- SFS 2009:600 *Språklag*. Stockholm: Kulturdepartementet

- <http://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/spraklag-2009600_sfs-2009-600> Hämtat 2017-04-29
- SFS 2010:1138 *Förordning om samhällsorientering för vissa nyanlända invandrare*. Stockholm: Arbetsmarknadsdepartementet <https://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/forordning-20101138-om-samhallsorientering-for_sfs-2010-1138> Hämtat 2017-04-29
- SFS 2013:156 *Lag om samhällsorientering för vissa nyanlända invandrare*. <http://www.riksdagen.se/sv/dokument-lagar/dokument/svensk-forfattningssamling/lag-2013156-om-samhallsorientering-for-vissa_sfs-2013-156> Hämtat 2017-04-29
- Skolverket 2007. *Gemensam referensram för språk: lärande undervisning och bedömning*. <https://www.skolverket.se/om-skolverket/publikationer/visa-enskild-publikation?_xurl_=http%3A%2F%2Fwww5.skolverket.se%2Fwtpub%2Fws%2Fskolbok%2Fwpubext%2Ftrycksak%2Fblob%2Fpdf2144.pdf%3Fk%3D2144> Hämtat 2017-04-05
- Skolverket 2012. *Kursplan för Sfi*. <<https://www.skolverket.se/laroplaner-amnen-och-kurser/vuxenutbildning/komvux/sfi>> Hämtat 2017-06-05
- Spache, G. D. 1953. A new readability formula for primary-grade reading materials. *Elementary School Journal*, 53, 410–413.
- Språkbanken. *Bloggmix 2012* <<https://spraakbanken.gu.se/swe/resurs/bloggmix2012#tabs=information>> Hämtat 2017-03-18
- Svensén, B. 2004. *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts
- Sveriges radio – *En handbok om att leva i Sverige* <<http://sverigesradio.se/sida/artikel.aspx?programid=104&artikel=5452170>> Hämtat 2017-03-18
- Viberg, Å. 2004. Lexikal utveckling i ett andraspråk. I: K Hyltenstam & I Lindberg (red.) *Svenska som andraspråk i forskning, undervisning och samhälle*. Lund: Studentlitteratur
- Vygotsky, L. 1978. *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press

Bilaga 1: Referensmaterial

Referensmaterial från Informationsverige.se

Allt hämtat 2017-04-02

<https://www.informationsverige.se/Svenska/Leva-och-bo/Sidor/Fritid.aspx>

<https://www.informationsverige.se/Svenska/Leva-och-bo/Sidor/Trivselregler.aspx>

<https://www.informationsverige.se/Svenska/Leva-och-bo/Sidor/Fritid.aspx>

Referensmaterial från Göteborgs Stad

Allt hämtat 2017-04-02

http://goteborg.se/wps/portal/start/forskola-och-utbildning/grundskola/forskoleklass/om-forskoleklass!/ut/p/z1/04_Sj9CPykssy0xPLMnMz0vMAfljo8ziAwy9Ai2cDB0N_N0t3Qw8Q7wD3Py8fdxdAs30wwkpiAJKG-AAjgb6BbmhigDIU23f/dz/d5/L2dBISEvZ0FBIS9nQSEh/

http://goteborg.se/wps/portal/start/forskola-och-utbildning/grundskola/sa-fungerar-grundskolan/skolkort-skolskjuts!/ut/p/z1/04_Sj9CPykssy0xPLMnMz0vMAfljo8ziAwy9Ai2cDB0N_N0t3Qw8Q7wD3Py8fdxdAgz1wwkpiAJKG-AAjgb6BbmhigBNtq3p/dz/d5/L2dBISEvZ0FBIS9nQSEh/

http://goteborg.se/wps/portal/start/forskola-och-utbildning/grundskola/sa-fungerar-grundskolan/skolskjuts!/ut/p/z1/04_Sj9CPykssy0xPLMnMz0vMAfljo8ziQw0NAi2cDB0NLCw9XQ08vSwTQ1DLI39LQz0wwkpiAJKG-AAjgb6BbmhigBg-Puf/dz/d5/L2dBISEvZ0FBIS9nQSEh/

http://goteborg.se/wps/portal/start/forskola-och-utbildning/grundskola/forskoleklass/regler-for-forskoleklass!/ut/p/z1/04_Sj9CPykssy0xPLMnMz0vMAfljo8ziTYzcDQy9TAy9DdwND_AwcQ8z93d0Dgo39DUz0wwkpiAJKG-AAjgb6BbmhigCisn5H/dz/d5/L2dBISEvZ0FBIS9nQSEh/

Referensmaterial från GP

Allt hämtat 2017-04-02

<http://www.gp.se/nyheter/v%C3%A4rlden/sex-gripna-f%C3%B6r-misshandel-av-asyls%C3%B6kande-1.4220366>

<http://www.gp.se/nyheter/v%C3%A4rlden/flynn-redovisade-inte-ryska-int%C3%A4kter-1.4220257>

<http://www.gp.se/nyheter/v%C3%A4rlden/presidentval-kan-f%C3%A5-betydelse-f%C3%B6r-assange-1.4220239>

<http://www.gp.se/nyheter/v%C3%A4rlden/poliser-matf%C3%B6rgiftning-p%C3%A5-tr%C3%A4ningsl%C3%A4ger-1.4220394>

<http://www.gp.se/nyheter/v%C3%A4rlden/vaktm%C3%A4stare-f%C3%B6rgiftade-och-slog-ihj%C3%A4-20-1.4220260>

<http://www.gp.se/nyheter/v%C3%A4rlden/serbien-v%C3%A4ljer-ny-president-1.4220282>

<http://www.gp.se/nyheter/sverige/polisen-vill-ha-mer-milit%C3%A4rhj%C3%A4lp-mot-terror-1.4220233>