

GÖTEBORGS UNIVERSITET
Institutionen för svenska språket

Förutsättningar för ordförrådsutveckling på SFI:
en frekvensbaserad lexikalisk analys av ord i undervisningen

Leila Swartling

C-uppsats, SSA136 15 hp
Ämne: Svenska som andraspråk
Termin: HT18
Handledare: Julia Prentice

Sammandrag

Syftet med uppsatsen var att undersöka hur ord som undervisas på svenska för invandrare (SFI) kurs C och D korrelerar med frekvensordlistor över det svenska språket samt om kurs D innehåller fler ovanliga ord än kurs C. Under fyra observationstillfällen hos två olika lärare samlade jag in vilka ord som de fokuserade på i hög grad, det vill säga ord som fick *rich instruction*, och i lägre grad, det vill säga ord som fick *grundläggande definition*. Utöver detta material samlade jag även in allt textmaterial som eleverna mötte under lektionstiden. Resultatet visade att 87% av orden som i undervisningsmaterialet, 80% av de ord som fick *grundläggande definition* och 87% av de ord som fick *rich instruction* tillhörde svenskans vanligaste ord. Resultatet visade ingen progression mellan kurs C och kurs D med avseende på ordfrekvens.

Nyckelord: *svenska som andraspråk, ordundervisning, frekvensbaserad lexikalisk analys, rich instruction, frekvensordlista, svenska för invandrare*

Innehållsförteckning

1.	Introduktion	1
1.1.	Bakgrund	1
2.	Syfte och frågeställningar	2
2.1.	Frågeställningar	3
3.	Tidigare forskning	3
3.1.	Frekvensbaserad lexikalisk analys i svensk forskning	3
3.2.	Muntligt språkligt inflöde i andraspråksundervisning	4
3.3.	Progression mellan språkkurser med avseende på ordens frekvensnivå	5
4.	Teori	6
4.1.	Ordförråd och dess struktur	6
4.2.	Fokus på högfrekventa ord för effektiv ordförrådsutveckling	7
4.3.	Högfrekventa ord i svenskan	8
4.4.	Grad av fokus på ord i undervisningen	11
5.	Material och metod	13
5.1.	Urval	14
5.2.	Material	15
5.2.1.	Beskrivning av lektionerna	15
5.3.	Genomförande	16
5.3.1.	Pilotstudie	17
5.4.	Analys och bearbetning av material	18
5.4.1.	Bearbetning av frekvensordlista	18
5.4.2.	Bearbetning av material	19
5.5.	Forskningsetiska aspekter	21
5.6.	Reliabilitet och validitet	21
6.	Resultat och analys	23

6.1.	Resultat	23
6.1.1.	Grad av fokus på de högfrekventa orden	23
6.1.2.	Progression mellan kurs C och D	25
6.1.3.	Partikelverb och reflexiva verb	27
6.1.3.1.	Partikelverb	27
6.1.3.2.	Reflexiva verb	28
6.2.	Analys av resultat	28
7.	Diskussion	30
7.1.	Metoddiskussion	30
7.2.	Resultatdiskussion	33
7.3.	Slutsatser och förslag på vidare forskning	34
8.	Litteraturförteckning	36
	Bilaga 1	39
	Bilaga 2	40

Figurförteckning

Figur 1.	Det mentala lexikonet i utveckling (Järborg 2007:64)	7
Figur 2.	Exempel på frågekort	15
Figur 3.	Observationsschema	17
Figur 4.	Andel av materialkategorierna som täcks av frekvensordlistorna	24
Figur 5.	Jämförelse av täckningsgrad på Camillas lektioner	26
Figur 6.	Jämförelse av täckningsgrad på Sivs lektioner	26
Figur 7.	Exempel på taveltext från pilotobservation.	39
Figur 8.	Exempel på stencil	40

Tabellförteckning

Tabell 1. Täckningsgrad utifrån frekvensband med 1404 lemman	10
Tabell 2. Förekomst av partikelverb fördelat i materialet	27
Tabell 3. Förekomst av reflexiva verb fördelat i materialet	28

1. Introduktion

Att snabbt lära sig det svenska språket är ofta en prioritet för personer som nyligen flyttat eller flytt till Sverige. Det är av stor vikt att utbildningen är effektiv och högkvalitativ eftersom språkkunskaper är en nyckel till arbetsmarknad, vidare studier och integration.

En viktig del av att lära sig ett språk är att utveckla ett stort ordförråd. För att förstå en text utan hjälpmedel måste läsaren förstå mellan 95–98% av orden och en lägre andel kända ord minskar läsförståelsen markant (Nation 2013:205f). Svenska för invandrare (SFI) syftar till att ge eleverna det basordförråd som krävs för att delta i vardags-, samhälls- och arbetsliv samt fortsatta studier. Ett sätt att förhålla sig till den stora mängd ord som behövs för ett funktionellt språk i en sådan stor bredd av situationer är att fokusera inläringen på de vanligaste orden i språket (Nation 2013:94).

Är ordundervisningen inom SFI effektiv i bemärkelsen att det är de mest frekventa orden i vardags-, samhälls- och arbetsliv samt fortsatta studier som får fokus i utbildningen? Bör lärare vara mer uppmärksamma på vilka ord som är de vanligaste eller sker det automatiskt att de vanligaste orden undervisas om först? Det är nog få lärare som konsulterar en frekvensbaserad ordlista inför varje lektion, men en möjlighet är att dessa ord kommer i fokus utan att läraren behöver vara medveten om det; kanske kommer det naturligt att tidigt i utbildningen fokusera på de vanligaste orden i språket. Om så inte är fallet kan det indikera att SFI-undervisningen har utrymme att effektivisera elevernas ordinläring med avseende på vilka ord som fokuseras.

1.1. Bakgrund

Svenska för invandrare (SFI) beskrivs i Statens skolverks författningssamling på följande vis:

Kommunal vuxenutbildning i svenska för invandrare är en kvalificerad språkutbildning som syftar till att ge vuxna invandrare grundläggande

kunskaper i svenska språket. En elev med ett annat modersmål än svenska ska inom utbildningen få lära sig och utveckla ett funktionellt andraspråk. Utbildningen ska ge språkliga redskap för kommunikation och aktivt deltagande i vardags-, samhälls- och arbetsliv samt fortsatta studier. (SKOLFS 2017)

SFI består av fyra olika kurser: A, B, C och D. Beroende på elevernas studiebakgrund blir de tilldelade en av tre studievägar. Studieväg 1 består av alla fyra kurser; Studieväg 2 består av kurs B, C och D; och studieväg 3 består av kurs C och D. Kurs B och kurs C kan alltså utformas som antingen nybörjarkurser eller fortsättningskurser beroende på elevernas tidigare kunskaper (SKOLFS 2017).

Fram till januari 2018 var kurserna i SFI explicit knutna till den gemensamma europeiska referensramen för språk (GERS) (SKOLFS 2009). GERS är en gemensam europeisk standard för att bedöma olika nivåer av språklig kompetens med syfte att kunna jämföra färdigheter mellan olika språk. GERS innehåller kriterier för sex olika nivåer av språkfärdighet, från nybörjare till avancerad språkanvändare: A1, A2, B1, B2, C1 och C2 (Skolverket 2009). Kurserna på SFI motsvarade då tillsammans nivåerna A1 till B1+. Kurs A var relaterad till nivå A1-/A1, kurs B var relaterad till A1/A2, kurs C var relaterad till nivå A2/A2+ och kurs D var relaterad till nivå B1/B1+ (SKOLFS 2009). I den senaste förordningen har denna explicita koppling till GERS tagits bort, men kunskapskraven för kurs C och D har inte ändrats (SKOLFS 2017).

2. Syfte och frågeställningar

Syftet med uppsatsen är att undersöka om ordundervisningen på SFI är effektiv i den bemärkelsen att det är de mest frekventa orden i språket som undervisas. Vidare är syftet att undersöka om det går att observera en progression mellan kurs C och D med avseende på ordfrekvens samt hur frekvensnivån förhåller sig till hur mycket fokus orden fått i undervisningen.

2.1. Frågeställningar

- I hur stor utsträckning lägger lärare i SFI C och D fokus på svenskans högfrekventa ord i klassrumsundervisningen?
- Finns det en skillnad på andelen högfrekventa ord mellan de ord som får mer fokus och de ord som får mindre fokus i klassrumsundervisningen?
- Syns en progression mellan SFI C och D på de undervisade orden från mer frekventa till mindre frekventa ord?

3. Tidigare forskning

Nedan följer en genomgång av tidigare forskning om vilka ord som andraspråksinlärare möter i språkundervisningen. De flesta studier handlar om vilka ord elever möter i *engelsk* andraspråksundervisning och använder så kallad *frekvensbaserad lexikalisk analys* som analysmetod. Metoden går ut på att man jämför det textmaterial man vill analysera med ordlistor som är ordnade efter hur vanliga orden är i språket. Som resultat får man en täckningsgrad, alltså hur många procent av textmaterialet som består av språkets vanligaste ord (Nation 2014:16ff). Under nästa rubrik följer först en kort överblick över metoden och över hur den använts i svensk forskning.

3.1. Frekvensbaserad lexikalisk analys i svensk forskning

Frekvensbaserad lexikalisk analys är en metod som konstruerades av Laufer och Nation (1995) med syfte att kunna analysera inlärares produktiva ordförråd. De använde datorprogrammet VocabProfile för att jämföra inläratexter med två listor över de 1000 vanligaste engelska orden och de 1000 vanligaste orden om följer därefter (Laufer och Nation 1995:315). På så sätt kunde de se skillnader i täckningsgrad av de vanligaste orden mellan texter producerade av inlärare på olika nivåer (Ibid.:316).

Sedan liknande analysprogram publicerats på nätet har metoden använts i en bredd olika syften (Lindberg 2006:100). I svensk forskning har det bland annat använts för att utforma ett test av skolspråkskunskaper (Ibid.:100), analysera

sammansättningen i en läromedelstext i kemi (Kokkinakis och Frändberg 2013), analysera hur gymnasieelever använder sin vokabulär i texter (Ohlsson 2018) och i en studentuppsats undersöktes hur stort ordförråd som krävdes för att läsa den samhällsorienterande läroboken *Om Sverige* (Johansson 2017).

3.2. Muntligt språkligt inflöde i andraspråksundervisning

Nedan följer en redogörelse för två studier som undersökte vilket muntligt språkligt inflöde som inlärare möter på andraspråkslektioner. Den första studien adresserar frågan om hur många nya ord som en inlärare brukar möta under en vanlig språklektion. Studien genomfördes av Meara, Lightbown och Halter med material från undervisning i engelska som andraspråk för 11–12 åringar i Quebec (1997:33). Det analyserade materialet bestod av transkriberingar av 10 inspelningar á 30 min av lärares tal på lektioner (Ibid.:33). Som definition för vad som skulle räknas som ”nya ord” valde Meara et al. att jämföra de transkriberade inspelningarna med en ordlista av Nation från 1986. Nations ordlista, skriver de, är främst frekvensbaserad men har även andra inslag med syfte att den ska innehålla de 2500 *ordfamiljer* som inlärare med stor sannolikhet möter tidigt i sin språkinlärning. En ordfamilj består av ett huvudord, grammatiska varianter av ordet och även avledningar med affix (för mer ingående förklaring se rubrik 4.3.). De ord som inte täcktes av denna ordlista definierade forskarna som ”nya” (Ibid.:34f). Resultatet visade att 0,55% av orden som lärarna använde på lektionerna var nya för inlärarna (Ibid.:38). Forskarna skriver att de inledningsvis tolkade resultatet som att andelen ovanliga ord i undervisningen var låg, men att siffrorna ser annorlunda ut om man tänker att eleverna sammanslaget under en dag med fem timmars undervisning möter hela 50 ovanliga ord i genomsnitt. De ifrågasätter även sin egna definition av vad som är ”nya ord” för eleverna, eftersom annan forskning visar att inlärare på samma nivå som de i studien ofta har stora luckor i sitt basordförråd, vilket indikerar att eleverna i realiteten möter många fler nya ord per dag än vad resultaten visar (Ibid.:40).

En liknande undersökning gjorde Horst (2010:164) av transkriberingar av lärare från 32 undervisningstimmar i engelska som andraspråk på nivån ”high intermediate - advanced” i Montreal. Hon undrade (Ibid.:162) liksom Meara et al. (1997) vilka möjligheter inlärare har att öka sitt ordförråd utifrån det muntliga språkliga inflöde de möter under lektionstid. Resultaten visar att lärare i engelska som andraspråk

förenklar sitt språk, och detta speciellt första dagen på kursen (Horst 2010:173). Hon kom även fram till att ovanliga ord förekommer sporadiskt och utan tillräckligt många upprepningar för att inlärarna ska kunna lära sig orden utan explicit undervisning av betydelsen (Ibid.:176).

Båda studierna analyserar material från undervisning som har ett uttalat fokus på muntlig kommunikation. I undervisning som inte har en speciell pedagogisk inriktning mot muntlig kommunikation är det dock viktigt att även ta hänsyn till de skriftliga texter som eleverna möter i undervisningen eftersom man annars missar en stor del av det språkliga inflöde som eleverna möter. I mitt materialval hade jag detta i åtanke. Studierna utgår också ifrån att en större andel ovanliga ord är önskvärdt, vilket skiljer sig från min teoretiska utgångspunkt som är att det är positivt om läraren fokuserar på de vanligaste orden i språket.

3.3. Progression mellan språkkurser med avseende på ordens frekvensnivå

I en konferensartikel presenterar Meara (1993) ett projekt där syftet var att undersöka ifall andelen ovanliga ord växer i takt med att L2-undervisning blir mer avancerad. Projektet genomfördes på uppdrag av BBC som ville få en analys av ordförrådet i språkprogrammen på BBC English radio. Hypotesen som projektet utgick från var att andelen ovanliga ord skulle öka varefter svårighetsnivån på kurserna ökade (Ibid.:2). Studiens resultat motbevisade dock hypotesen. Genom att jämföra de transkriberade programmen med Nations ordlista från 1986 kunde de se hur många ord som inte räknades till de vanligaste orden i engelskan. Denna frekvensbaserade lexikaliska analys visade inte på någon skillnad mellan de program som var svåra eller lätta utan andelen svåra ord visade sig variera mer mellan olika avsnitt av samma program än mellan olika program (Ibid.:7).

I ett svenskt projekt, SVALex, undersökte François, Volodina, Pilán och Tack (2016) ordförrådet i läroböcker i svenska som andraspråk. Genom att utgå från en uppdelning av läroböckerna utifrån GERS kunde de sammanställa en ordlista som redovisade ords förekomster på de olika GERS-nivåerna. SVALex har bland annat använts i forskning om automatisk nivåbestämning av inlärares texter (Volodina, Pilán och Alfter 2016) och för att skapa datagenererade språkövningar (Pilán, Volodina och Borin 2017). François et al. (2016:217) skriver att de hoppas på forskning som kan svara på frågorna om hur många ord en inlärare bör lära sig per

GERS-nivå och vilka ord som är passande att lära sig på vilken GERS-nivå. SVALex är potentiellt ett bra verktyg för många studier av ordförråd i svenska som andraspråksundervisning men på grund av att SVALex utgår från ordfrekvens i läroböcker, och inte från språket i stort, är den inte användbar för att svara på mina frågeställningar.

4. Teori

Nedan presenteras studiens teoretiska utgångspunkter. Inledningsvis beskriver jag en teori om hur ordförrådet är strukturerat. Därefter går jag igenom Nations teori om vilka ord som bör fokuseras för att utveckla ordförrådet och vilka ord dessa skulle kunna vara i en svensk kontext. Till sist beskriver jag olika sätt för lärare att fokusera på ord i undervisningen.

4.1. Ordförråd och dess struktur

En språkbrukares ordförråd består av två delar: det *receptiva* ordförrådet och det *produktiva* ordförrådet (Nation 2013:47). Det receptiva ordförrådet består av ord som språkbrukaren förstår men ännu inte lärt sig använda i egenproducerat tal och text. Det produktiva ordförrådet består i sin tur av ord som språkbrukaren är så väl förtrogen med att hen själv kan använda dem i tal och skrift. För att ett ord ska anses vara en del av det produktiva ordförrådet måste språkbrukaren kunna många olika aspekter av ordet:

- Hur ordet stavas eller uttalas.
- Hur ordet böjs och var det kan placeras i en sats.
- Vilka syntaktiska egenskaper ordet har och vilka ord som brukar kombineras med det, exempelvis vilka prepositioner som ofta hör samman med vilka verb.
- Vilka olika betydelser och associationer ett ord har och vilken stilnivå det tillskrivs (Ibid.:49).

Ett begrepp som beskriver en språkbrukares ordförråd är *det mentala lexikonet* (Järborg 2007:62). Figur 1 visar hur det mentala lexikonet utvecklas och växer; nya ord blir först en del av det receptiva ordförrådet för att senare införlivas i det produktiva ordförrådet.



Figur 1. Det mentala lexikonet i utveckling (Järborg 2007:64)

4.2. Fokus på höfrekventa ord för effektiv ordförrådsutveckling

Det råder sämre förutsättningar för en person som lär sig ett andraspråk, jämfört med ett förstaspråk, gällande möjligheten att tillägna sig ett stort ordförråd menar Nation (2013:94). Jämfört med förstaspråkstalare har andraspråksinlärare både mindre tid att utveckla ett grundläggande ordförråd och också ofta färre möjligheter till naturlig kontakt med språket, vilket ökar behovet av en effektiv ordförrådsutveckling inom språkundervisningens ramar (Ibid.:94). Utifrån forskning om ordförråd och ordundervisning föreslår därför Nation att ordundervisning bör:

1. fokusera på de mest frekventa orden i målspråket.
2. utöver de höfrekventa orden fokusera på domänspecifika ord som är användbara för inlärares unika syften.

Nation skriver till och med att ”the high-frequency words of the language are so important that considerable time should be spent on these words by both teachers and learners.” (2013:24). Var gränsen ska dras för vilka ord som är tillräckligt högfrekventa för att det ska löna sig för inläraren att fokusera på dem låter Nation vara en bedömningsfråga för enskilda lärare (Ibid.:22). Han lyfter också upp två viktiga begränsningar för explicit ordundervisning. För det första måste läraren ha en väl avvägd tidsfördelning mellan explicit ordundervisning och annan undervisning eftersom undervisningen måste lämna utrymme för andra aspekter av språket under den begränsade tiden som finns att tillgå. För det andra går det endast att erhålla kunskap om vissa aspekter av ord genom att möta ordet många gånger i naturlig kontext, bland annat för att få en känsla för nyanser i stilnivå (Ibid.:94).

4.3. Högfrekventa ord i svenskan

Att fastställa vilka ord som tillhör de mest frekventa i ett språk är dock minst sagt en komplex process som kräver beslut om vilken korpus som bäst representerar språket, om vad som ska räknas som ett ord och även om hur många ord som ska räknas som högfrekventa.

En *korpus* är enligt Johansson Kokkinakis (2007:102) ”en textsamling eller en textmängd” som kan användas för att i vetenskapligt syfte belägga teorier. För att skapa en *frekvensordlista*, alltså en ordlista som är sorterad efter hur ofta orden förekommer, som speglar det svenska språket i sin helhet måste korpusen bestå av texter som tillsammans representerar språkbruket i stort. Till exempel är tidningsspråk, enligt Järborg (2007:89), ett korpusmaterial som är representativt för skriftspråk, men samtidigt är det inte representativt för hela svenska språket eftersom konkreta och vardagliga ord är underrepresenterade i tidningsgenrerna. Trots att frekvensordlistor alltså skiljer sig åt beroende på vilka korpusar som ligger till grund för dem brukar ändå fördelningen av ordklasser vara lika; formord (t.ex. prepositioner, konjunktioner och pronomen) återfinns i högre grad bland de vanligaste orden (Enström 2010:36).

För att skapa en frekvensordlista måste forskaren även skapa en exakt definition för vad som räknas som ett ord, för att därefter kunna räkna antalet förekomster. *Ordfamiljer* är en ordenhet som består av ett huvudord, dess böjningsformer samt avledningar med affix. Exempelvis skulle en ordfamilj som utgår från huvudordet

lojal förutom *lojala* och *lojalt* även kunna innehålla avledningen *illojal* med böjningsformer. Enligt McLean (2017:2) blev ordfamiljer en populär enhet i engelsk korpuslingvistik efter en tongivande artikel av Nation och Bauer (1993). Vad som ingår i en ordfamilj är dock delvis en fråga om subjektiv bedömning vilket utgör ett stort problem då enheten används för att räkna ord, anser Nation (2013:11). Även Järborg (2007:79) ser problem i hur ordfamiljer bör avgränsas speciellt i ett språk som svenskan med dess många sammansättningar. I den svenska litteraturen är det vanligt att utgå från *lemma* vid studier av ordförråd. Det finns även flera engelska lexikon som utgår från lemman (exv. Gardner och Davies 2013). Lemman består av ett huvudord och dess böjningsformer. Exempelvis tillhör *bil*, *bilar*, *bilen*, *bilarna*, *bils*, *bilars*, *bilens* och *bilarnas* samma lemma (Järborg 2007:73).

Det har bedrivits flera projekt med syftet att skapa en frekvensordlista över svenska språket som helhet. Ett pionjärprojekt inom frekvensordlistor i Sverige är Nusvensk frekvensordbok 1–4 (Allén 1970–1980). Ordböckerna bygger på en korpus bestående av pressmaterial från 1965. En senare frekvensordlista, Base Vocabulary Pool, skapades av Eva Forsbom (2006). Trots namnet är det en svensk frekvensordlista som är skapad utifrån en korpus som är tänkt att representera vad en genomsnittlig person läser på ett år. Den utgår från en miljon ord av skriftligt material från 90-talet (Ibid.:1). Det senaste större projektet som resulterat i en frekvensordlista över svenska språket är KELLY-listan (Volodina och Johansson Kokkinakis 2012a). Listan är skapad utifrån en korpus med 114 miljoner ord från webbtexter. En av listans styrkor jämfört med de som presenterats ovan är att den representerar ett modernt språkbruk eftersom korpusen består av en mix av genrer som inkluderar bloggar, forum och chattar vilka har en mer talspråkslik stil än exempelvis pressmaterial (Ibid.:1041). KELLY-listan syftar till att utgöra ett verktyg för språkinlärare genom att orden i listan även finns översatta till åtta andra språk som också ingår i projektet (Ibid.:1040). För att öka användbarheten för inlärare delades orden i listan upp med vägledning av nivåbeskrivningarna för Gemensam Europeisk Referensram för Språk (GERS, eng. CEFR; Kilgarriff et al. 2014:137). Den svenska listan är uppdelad i sex lika stora delar, 1404 huvudord i varje del, och varje lista är tilldelad en GERS-nivå utifrån ordens frekvens (se tabell 1; Volodina och Johansson Kokkinakis 2012b:37).

Hur många av orden i en frekvensordlista som ska ses som högfrekventa, eller som de vanligaste orden, är en fråga som kan ges olika svar. För engelskan definieras oftast de vanligaste 2000 ordfamiljerna som högfrekventa (Nation 2013:23) men Schmitt och Schmitt (2014:486) menar att den siffran utgår från gammal forskning och därför behöver omvärderas. De argumenterar istället för att de 3000 vanligaste ordfamiljerna ska räknas som högfrekventa. En metod som Schmitt och Schmitt använder för att avgöra vilka ord som ska räknas som högfrekventa är att undersöka hur frekvensnivåerna för orden är fördelade. En relativt liten andel av de ord som finns utgör nämligen en mycket stor andel av de ord vi använder, vilket man kan använda som indikator för att avgöra vilka ord som ska räknas som högfrekventa (Ibid.:486f). I en svensk kontext med KELLY-listan som korpus kan vi se att de 1404 mest frekventa lemmarna (i KELLY markerade att tillhöra GERS-nivå A1) utgör hela 69,6% av orden i korpusen (se tabell 1). Därefter minskar täckningsgraden drastiskt. Att lära sig nästa sjok av 1404 lemman innebär att man som inlärare förstår ytterligare 5,4% av texterna i den korpus som KELLY-listan bygger på. Nästa 1404 lemman gör att man förstår ytterligare 2,3% och så vidare. Täckningsgraden blir allt mindre för lemman längre ned i KELLY-listan.

Tabell 1. Täckningsgrad utifrån frekvensband med 1404 lemman

Lemman	GERS	Täckningsgrad
0–1404	A1	69,6%
1405–2808	A2	5,4%
2809–4212	B1	2,3%
4213–5616	B2	1,3%
5617–7020	C1	0,8%
7021–8425	C2	0,2%

Metoden ger inga självklara svar på var gränsen ska dras mellan vilka ord som är högfrekventa och inte. Men en utgångspunkt kan vara att gränsen mellan hög- och lågfrekventa ord bör dras där behållningen av att lära sig orden i frekvensordning inte längre är speciellt stor, jämfört med att exempelvis fokusera på domänspecifika ord. Schmitt och Schmitt (2014:487) anser att täckningsgraden blivit mycket liten då 2000 ordfamiljer inte täcker mer än 3% av de korpusar som de utgår från. Siffrorna

är inte direkt jämförbara eftersom det handlar om olika språk och olika orddefinitioner, men från den information som finns att tillgå skulle en försvarbar mängd ord kanske som mest kunna ligga runt B1 eller B2 innan täckningsgraden blir mycket liten. I denna uppsats definierar jag därför svenskans högfrekventa, eller vanligaste, ord som de 4212 mest frekventa lemmarna från KELLY-listan, motsvarande B1.

Det finns dock andra möjliga uppdelningar av hur många lemmarna som inlärare bör kunna på varje GERS-nivå. Milton (2010) har undersökt hur många engelska lemmarna som inlärare kan beroende på vilken GERS-nivå de anses ha uppnått och har mätt detta genom ett ordtest av de 5000 vanligaste orden i engelskan. Enligt hans studie (Ibid.:224) motsvarar B1 de 2750–3250 vanligaste lemmarna vilket å ena sidan i stort sammanfaller med KELLY-listans uppdelning av lemmarna men å andra sidan är ett mycket mindre intervall än KELLY-listans uppdelning. Elever med kunskap om 4212 lemmarna ligger enligt Milton snarare på C1-nivå. Miltons resultat visar dock på att hur många lemmarna som en inlärare kan vid olika GERS-nivåer skiljer sig mellan olika språk och därför kan det vara vanskligt att utgå från Miltons uppdelning i en undersökning av svenska (Ibid.:226). Följaktligen utgår definitionen av svenskans högfrekventa ord från KELLY-listans uppdelning.

4.4. Grad av fokus på ord i undervisningen

I undervisningen kan specifika ord få mer eller mindre fokus och undervisningstid. Ett ord kan få väldigt lite fokus men ändå vara ett ord som läraren använder när hen pratar, eller få väldigt mycket uppmärksamhet genom att vara det huvudsakliga målet för undervisningen. Enström skriver att det är viktigt för inlärare att möta ord i deras vanliga kontext eftersom det är så man oftast möter orden (2010:23). Hon menar även att rikt språkligt inflöde skapar goda förutsättningar för inlärare att bygga upp ett stort ordförråd (Ibid.:24). Detta indikerar att det är viktigt att lärare använder ett brett ordförråd i uppgifter och när de talar med eleverna. Schmitt menar dock att det inte går att förlita sig på att få ett stort ordförråd endast genom språkligt inflöde (2008:348). I en forskningssammanställning skriver han att ordinläring genom läsning ger viss förståelse för orden men sällan gör dem till en del av läsarens produktiva vokabulär. Dessutom, skriver Schmitt, behövs hela 8–10 möten med ordet i kontext bara för att få en grundläggande förståelse för betydelsen

(Ibid.:347f). För att öka sitt ordförråd krävs istället mycket fokus på de specifika orden, fortsätter han (Ibid.:341). Den forskning som han har sammanställt indikerar att alla metoder som ”leder till mer exponering, uppmärksamhet, manipulation, eller tid spenderad på en lexikalisk enhet bidrar till inläringen” (min översättning, 2008:339). Med andra ord leder mer fokus på ett ord till en större chans att eleven införlivar ordet i sitt ordförråd.

Ord som förklaras av läraren men som får en låg grad av fokus kallar jag för ord som får *grundläggande definition*, vilket kan liknas vid Flowerdews begrepp *embedded definition* (1992). Embedded definitions används enligt Flowerdew av läraren för att se till att studenterna förstår de ord och begrepp som används under föreläsningarna utan att låta ordförklaringarna ta mer tid än behövt (1992:209). Enligt Nation (2013:117) är det här sättet idealt att använda då inlärnarna måste förstå ordet för att kunna ta sig vidare i den uppgift de har framför sig. Exempelvis om eleverna får en skriftlig instruktion som innehåller ord som de inte förstår behöver dessa förklaras så att eleverna förstår vad de förväntas göra. Nation (2013:126) skriver också att korta, lättförståeliga och tydliga ordförklaringar ofta är det bästa för att gynna ordinläringen, speciellt om inlärnaren möter ordet för första gången. Hans slutsats bygger delvis på att både Chaudron (1982) och senare Ellis (1995) menade att mer elaborerade förklaringar inte betyder att eleverna förstår bättre, utan att det snarare kan göra att eleverna tappar fokus på vad det är som förklaras.

I andra fall är det berättigat att lägga mer fokus på olika aspekter av ett ord förutom enbart den semantiska sidan. Nation (2013:117) använder begreppet *rich instruction* för att förklara hur läraren lägger mer fokus på ett ord än vad som rimligen behövs för att endast hjälpa inlärnarna att komma vidare i undervisningen, samtalet eller läsningen för stunden. Detta innebär oftast att det läggs mer tid på ordet, att olika aspekter av ordet undersöks, att ordet får explicit uppmärksamhet och att eleverna får bearbeta ordet genom övningar eller annan tankeverksamhet (Ibid.:117). Nation (Ibid.:117) menar att *rich instruction* är passande för ord som är högfrekventa i språket eller som tillhör en domän som eleven behöver lära sig ord ifrån. Vidare menar Nation (Ibid.:117) att *rich instruction* är lämpligt för ord som eleven redan mött ett antal gånger och därför har i sin receptiva vokabulär, med syfte att nu hjälpa eleven att även införliva det i hans produktiva vokabulär.

I föreliggande uppsats definieras *rich instruction* som de fall där ett ord har fått fokus bortom en *grundläggande definition*. En *grundläggande definition* kan ha getts i form av bilder, översättning till modersmål, charader, exempel eller genom en definition på svenska. *Rich instruction* uppstår då någon eller några av följande aspekter av ordet har fokuserats utöver den grundläggande betydelsen (inspirerad av Nation 2013:118):

- Uttal
- Stavning
- Orddelar
- Grammatisk funktion
- Kollokationer
- Synonymer
- Antonymer
- Hyponymi
- Nyansskillnader från andra ord (exempelvis stilnivå och stilvalör)
- Etymologi
- Eleverna får gissa ordets betydelse utifrån kontext

Sammantaget kan vi konstatera att för att ordförrådet ska utvecklas effektivt behöver läraren hjälpa eleverna att fokusera på de mest högfrekventa orden och utöver dem fokusera på domäner som är viktiga för eleven. Förutom att välja ut vilka ord som eleverna bör möta i undervisningen är det även viktigt för ordförrådsutvecklingen vilken typ av ordundervisning inlärarna får – *grundläggande definition* eller *rich instruction* – och om instruktionen ges innan eller efter att eleven införlivat ordet i det receptiva ordförrådet.

5. Material och metod

Studien har en kvantitativ ansats och det huvudsakliga tillvägagångssättet för materialinsamling till uppsatsen var systematisk observation. För att analysera materialet användes därefter programmet AntWordProfiler (Anthony 2014) och den

frekvensbaserade ordlistan KELLY. Nedan följer först en beskrivning av urvalet, materialet och observationerna. Därefter beskriver jag hur jag förberedde materialet och ordlistan inför analysen. Till sist följer några forskningsetiska reflektioner och en diskussion om metodens reliabilitet och validitet.

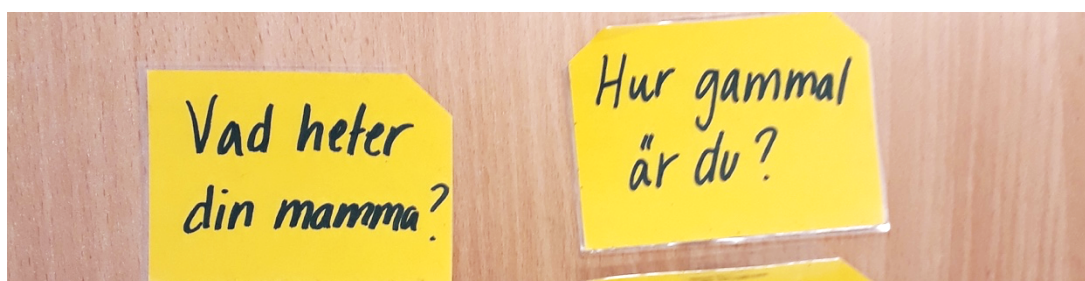
5.1. Urval

Urvalet utgick från populationen behöriga SFI-lärare i Göteborgsområdet som undervisade i kurs C och D. Kurserna valdes eftersom en övervägande del av eleverna på SFI går studieväg 3 och går alltså endast kurs C och D (Skolverket 2018:136). I dessa fall är C utformad som en nybörjarkurs. Mitt slutgiltiga urval bestod av två lärare från två olika skolor i Göteborgsområdet som erbjuder SFI-undervisning. Den ena läraren nåddes genom hennes kollega som var en bekant till mig. Den andra läraren föreslogs av en rektor jag kontaktat. Jag tog kontakt med lärarna via mail och de fick själva föreslå lektionstillfällen som passade dem. De gav sitt skriftliga samtycke inför den första observationen. Jag fick muntligt samtycke från eleverna i början av varje lektion. Mer om samtycke finns under rubrik 5.5..

I uppsatsen har lärarna fått fingerade namn: Camilla och Siv. Båda lärarna är behöriga SFI-lärare och hade vid tidpunkten för observationerna både kurser på C- och D-nivå med övervägande andel elever från studieväg 3. Camilla har 3 års och Siv har 13 års arbetslivserfarenhet på SFI. Storleken på urvalet begränsades av svårigheten att hitta lärare som undervisade både C och D kurser samtidigt.

5.2. Material

Det material som används i analysen samlades in under fyra olika SFI-lektioner i kurs C och D. Materialet består av de ord som eleverna mötte under lektionerna indelade i tre olika kategorier baserat på vilket fokus orden fick. Den första kategorin består av de ord som förekom i det skriftliga *undervisningsmaterialet*. Undervisningsmaterialet samlades in i samband med lektionen eller i efterhand och bestod av taveltext, stenciler, texter, frågekort och frågor till hörövningar. I figur 2 finns exempel på hur frågekorten såg ut. Exempel på taveltext och stencil finns i bilaga 1.



Figur 2. Exempel på frågekort

Taveltexten dokumenterades genom att jag skrev av den under lektionen. Den andra kategorin, *grundläggande definition*, består av de ord som fick sin betydelse förklarad av läraren. Den tredje kategorin är de ord som läraren undervisade om på fler sätt än att bara förklara vad de betyder. Dessa ord hör till kategorin *rich instruction*. Orden som fick *grundläggande definition* eller *rich instruction* samlades in genom observation. Observationerna beskrivs ingående under 5.3.. I uppsatsen refererar jag till de olika kategorierna som *materialkategorier*.

5.2.1. Beskrivning av lektionerna

Alla observerade lektioner utspelade sig i traditionellt möblerade klassrum med bänkar vända mot tavlan. Antalet elever varierade mellan 13–22 personer varav alltid ca 4 personer kom in efter lektionens start. Generellt sett använde lärarna inte någon lärobok utan använde istället texter och uppgifter från olika källor, vilket

styrker mitt beslut om att använda material från klassrumspraktik istället för läroböcker. Jag uppfattade det som att ingen av lektionerna ingick i något större övergripande tema utan att de framför allt knöt an till tidigare lektioner genom att de tog vid från ämnet eller uppgifterna på dessa.

Eleverna på Camillas lektion i C-kursen hade endast studerat på SFI i två veckor. Lektionen var innehållsmässigt indelad i tre delar. Först fokuserades årstider i helklass och aktiviteter som man kan utföra på olika årstider. Därefter fick eleverna en skrivuppgift som gick ut på att svara på ett brev. Till sist satt de i grupper och ställde varandra frågor som fanns nedskrivna på lappar (figur 2).

På Camillas lektion i D-kursen fokuserade hon på bisatser och underordnade konjunktioner. Lektionen inleddes med tavelgenomgång och därefter övade de med hjälp av stenciler samt varvade övningarna med att gå igenom facit i helklass.

Sivs lektion i C-kursen inleddes med springdiktamen på två korta texter som handlade om en kvinnas morgonrutin. Därefter skrev Siv texten på tavlan och den analyserades i helklass. Till sist hade läraren en genomgång om pronomen vilket eleverna sen fick öva på i en stencil.

Sivs lektion i D-kursen inleddes med två hörövningar från ett gammalt nationellt prov som handlade om att gå på bio. Därefter övade de på att sätta bisatser först i meningar och till sist gjorde de en stencil om placering av satsadverbial.

5.3. Genomförande

Jag använde systematisk observation med observationsschema för att samla in mitt material. Observationerna utfördes 25–30 oktober 2018 under totalt 8,3 timmar och varje observation varade i genomsnitt 103 minuter. Ett lektionstillfälle användes som pilotstudie för att utvärdera metoden. Utöver pilotstudien gjorde jag en observation på kurs C och en på kurs D per lärare.

I observationsschemat antecknade jag ord som på ett eller annat sätt förklarades under lektionen. Efter varje ord markerade jag på vems initiativ förklaringen skedde och vilken grad av fokus som ordet fick. Initiativet kategoriserades som antingen lärarens eller en elevs. Vilket fokus ordet fick markerades som antingen *grundläggande definition* eller *rich instruction*. Slutligen lämnades utrymme för kommentarer för att exempelvis skriva en förklaring till ord som upplevdes som gränsfall för att kunna anpassa kategoriseringen i efterhand.

Ord/fras/konstruktion	Initiativ		Fokus		Kommentarer
	Lärare	Elev	Grundläggande def.	Rich instruction	
girighet	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
överdrifter	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	
där	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	grammatisk funktion

Figur 3. Observationsschema

Oftast var gränsdragningen mellan kategorierna tydlig. En svårighet var dock tillfällena då lärarna fokuserade på uttal. Undervisning i uttal räknade jag som en aspekt av *rich instruction*, dock undervisades uttal ibland implicit genom körläsning eller genom att läraren uttalade ordet tydligt flera gånger. Jag bestämde att körläsning inte var tillräckligt för att räknas som *rich instruction* eftersom uttalsundervisningen i de fallen är mycket implicit och inte säkert fokuserar på de enskilda orden utan kanske snarare på betoning i meningen. I de fall där läraren förklarade någon aspekt av uttalet räknade jag det dock som *rich instruction*, som då Camilla fokuserade på ordet *tycker* och explicit fokuserade skillnaden mellan ljuden *i* och *y*.

I några få fall förklarades orden av eleverna själva. De gångerna behandlade jag det på samma sätt som om läraren förklarade ordet eftersom lärarna i alla dessa fall godkände förklaringarna genom att inte ifrågasätta dem.

För att komplettera observationsschemat förde jag fältanteckningar om hur lektionerna fortlöpte. Fältanteckningar, menar Denscombe (2018:303), är viktiga för att förstå det insamlade materialet och påverkande omständigheter. Jag antecknade även allt som läraren skrev på tavlan för att kunna analysera taveltexten tillsammans med övrigt undervisningsmaterial.

5.3.1. Pilotstudie

Den första observationen var en pilotstudie för att testa metoden och följaktligen utelämnade jag det insamlade materialet från analysen.

Samtycket fortlöpte utan problem. Eleverna accepterade min närvaro efter att jag bett om muntligt samtycke på svenska och engelska. Jag beslutade därför att

fortsätta med det muntliga samtycket på de båda språken. Även observationsschemat fungerade bra och det var oftast lätt att avgöra vilken kategori förklaringarna tillhörde.

På observationen placerade jag mig längst fram ute i kanten och fotograferade tavlan för att senare transkribera från bild till text, men jag insåg att det var bättre att placera mig längst bak och skriva av tavlan direkt istället för att ta bild på den. Enligt Denscombe (2018:303f) är det viktigt att försöka smälta in i miljön för att inte påverka den under observationer. Detta kan göras genom att undvika kontakt med deltagarna och genom att välja en strategisk placering. Den nya placeringen gjorde min närvaro mindre påtaglig i klassrummet.

5.4. Analys och bearbetning av material

Det insamlade materialet analyserades med hjälp av programmet AntWordProfiler (Anthony 2014). Programmet jämför textfiler med varandra och analyserar vilka ord som förekommer i båda filerna. Jag använde programmet för att analysera hur stor andel av orden i mitt insamlade material som täcktes av ordlistor över svenskans vanligaste ord, en så kallad *frekvensbaserad lexikalisk analys*. Programmet kan analysera flerordsenheter men det kräver en bearbetning av materialet som inte rymdes inom uppsatsens ramar. Programmet kan också endast läsa filer av typen .txt. Nedan beskriver jag hur jag bearbetat frekvensordlistan och materialet inför analysen.

5.4.1. Bearbetning av frekvensordlista

Jag har valt att utgå från KELLY-listan eftersom den är baserad på en modern korpus, utgår från lemman och har ett uttalat syfte att vända sig mot inlärare av språk. Som förklarar ovan är orden i KELLY-listan markerade efter GERS-nivå och jag valde att utgå från den uppdelningen för att göra en uppskattning av vad som kan ses som svenskans vanligaste ord (se rubrik 4.3.). Uppskattningen landade på omkring de 4212 första posterna i KELLY-listan, det vill säga de ord som markerats som A1, A2 eller B1, och därför utgick jag från dessa ord i min analys. Därtill, som nämnt i bakgrunden, motsvaras SFI-kurserna C och D av GERS-nivåerna A1 till B1 vilket bidrar till att avgränsningen lämpar sig för syftet i denna uppsats. Jag

bearbetade listan med utgångspunkt att programmet inte skulle tolka flerordsenheter eller förkortningar med punkter eftersom det kräver mer avancerad bearbetning som jag inte hade möjlighet att göra. Jag skapade tre separata listor för vardera GERS-nivå enligt följande kriterier:

- I de fall då två olika exempel förekom, där det ena var särskrivet och det andra var sammansatt, exempelvis *i morgon/imorgon*, raderade jag den särskrivna versionen.
- Förkortningar hanterade jag genom att radera punkter och mellanslag exempelvis ändrade jag *t. ex.* till *tex.*
- Jag raderade flerordsenheter, vilket var 2,6% av de 4212 första posterna i KELLY-listan.

För att ge en bild av vilka typer av ord som listorna innehöll efter bearbetningen följer här några slumpmässiga exempel: A1-listan innehåller ord som *tredje, kvinna, dag, oss*; A2-listan ord som *uttalande, fiende, ögonblick, ursäkt*; och B1-listan ord som *stabilitet, klimatförändring, mördare, storm*. Det är dessa listor som jag utgår ifrån för att besvara mina frågeställningar.

Utöver frekvensordlistorna skapade jag även en lista av alla egennamn som förekommer i materialet, och en lista över grammatisk terminologi. Jag använde de två listorna för att nyansera analysen av de ord som inte täcktes av frekvensordlistorna. Egennamn tillhör inte ordböcker och liknande eftersom de, som Järborg skriver (2007:67), räknas som omvärldskunskap och inte som ordkunskap. De är även exkluderade ur KELLY-listan vilket betyder att egennamn i mitt material inte kommer täckas av frekvensordlistorna. Listan över grammatisk terminologi är viktig eftersom sådana ord används frekvent i språkundervisning fastän de är ovanliga i språket i stort.

5.4.2. Bearbetning av material

I ganska stor utsträckning bearbetade jag materialet på samma sätt som Volodina och Johansson Kokkinakis (2012b) bearbetade sin korpus för att skapa KELLY-listan; i KELLY-listan uteslöts egennamn, skiljetecken, siffror, particip och ord på

andra språk än svenska (Ibid.:16). Eftersom KELLY-listan baseras på lemman lemmatiserade även jag mitt material. På grund av att AntWordProfiler inte kan analysera flerordsenheter utan speciell bearbetning av materialet och därför felaktigt tolkar dem som enskilda ord exkluderade jag en del flerordsenheter ur listorna som programmet analyserade. Nedan följer alla ändringar punkt för punkt:

- Jag uteslöt skiljetecken, siffror, particip och ord på andra språk än svenska. Liksom Volodina och Johansson Kokkinakis tog jag bort siffror (8, 34, 0) men behöll de nummer och ordningstal som var skrivna med bokstäver från början (exempelvis *tredje* och *två*).
- Jag behöll egennamn och grammatiska termer i materialet men urskiljer dem istället med hjälp av separata ordlistor (se 5.4.1.).
- Jag lemmatiserade mitt material med hjälp av verktyget *Sparv* (Borin et al. 2016). Sparv lemmatiserar dock inte pronomen och artiklar på samma sätt som KELLY-listan och därför ändrade jag dessa manuellt för att efterlikna KELLY-listan.¹ Exempelvis lemmatiserade Sparv *dig* → *du* och *mitt* → *jag* medan KELLY-listan innehåller både subjekts och objektsform men inte kongruens så att *dig* → *dig* och *mitt* → *min*.
- Jag tog bort *har/hade* i perfekt och pluskvamperfekt och markerade dem som samma lemma (t.ex. *har ägt* → *äga*).
- Partikelverb och reflexiva verb togs bort ur materialet och redovisas separat under 6.1.3.. Järborg (2007:78) skriver att flerordsenheter ska redovisas skilt från den lexikala profileringen. Den glidande skalan mellan vad som kan bedömas vara en flerordsenhet och inte gjorde att jag endast skilde ut partikelverb och reflexiva verb eftersom de är lätta att identifiera och klassificera. Alltså behandlas övriga flerordsuttryck som exempelvis kollokationer som enskilda ord. En annan anledning till att jag lämnade kvar övriga flerordsenheter i materialet var att KELLY-listan endast innehåller ett fåtal flerordsuttryck och därför likt mig behandlade de flesta flerordsuttryck i korpusen som enskilda ord. Jag diskuterar implikationerna av detta val i metoddiskussionen (7.1.).

¹ Kelly-listan har ej lemmatiserat pronomen eller *den/det/ett/en*, förutom *vilket/vilka* → *vilken*, *några/något* → *någon*, *inget/inga* → *ingen*, *mina/mitt* → *min* och *dessa* → *denna*.

- Förkortningar anpassade jag på samma sätt som jag gjorde i bearbetningen av KELLY-listan, dvs. tog bort punkter och mellanslag mellan bokstäverna.

Vid en ordanalys av det här slaget måste man skilja mellan tokens och types (Nation 2013:9). Om man räknar *tokens* räknar man varje enskilt ord oavsett om ordet förekommer flera gånger. Meningen *det är inte lätt när det är svårt* innehåller 8 tokens. Om man räknar *types* räknar man inte dubletter utan räknar bara de unika orden (Ibid.:9). Exempelmeningen ovan innehåller därför 6 types. Jag har valt att utgå från tokens i min analys. I metoddiskussionen (7.1.) diskuterar jag implikationer av valet.

5.5. Forskningsetiska aspekter

I min metod och materialinsamling har jag följt de forskningsetiska huvudprinciperna; deltagarnas intressen skyddas, deltagarna ger sitt informerade samtycke, nationell lagstiftning följs och inga falska löften har getts (Denscombe 2018:438). Lärarna gav sitt skriftliga samtycke (se Bilaga 2) efter att ha delgetts information om studiens syfte, genomförande och behandling av det insamlade materialet. Lärarna fick även information om att de själva, eleverna och deras skolor anonymiseras i uppsatsen och att deras deltagande är helt frivilligt och kan avbrytas när som helst.

Eleverna blev informerade muntligen om uppsatsens syfte och genomförande på engelska och på svenska utifrån ett fritt hållet manus. Jag meddelade dem att de kunde fråga mig eller deras lärare om de hade frågor eller ville att jag skulle utesluta någonting de sagt. Det var ingen elev som valde att göra det. Denscombe (2018:442) skriver att verbalt samtycke är tillräckligt vid observationer av lågriskkaraktär där det inte är realistiskt att erhålla samtycke från alla deltagare. Studien kan betraktas vara av lågriskkaraktär eftersom eleverna är helt anonymiserade, ingen inspelning sker och då jag inte återger elevernas yttranden i uppsatsen.

5.6. Reliabilitet och validitet

Begreppen reliabilitet och validitet handlar om till vilken grad en studie är fri från slumpmässiga och systematiska fel. Hög *reliabilitet*, alltså en undersökning utan

slumpmässiga fel, uppnås om undersökningen är noga genomförd och dokumenterad utan att oklarheter uppstår (Esaiasson et al. 2017:64). Hög *validitet*, alltså en undersökning utan systematiska fel, blir ett resultat av noga definierade begrepp och att metoden leder till att uppsatsen faktiskt svarar på frågeställningarna (Ibid.:58).

En fördel med observationsscheman är att de stärker studiens reliabilitet genom att minska subjektiv påverkan på vilka aspekter som fokuseras under observationen (Ibid.:305f). Genom att endast samla in skriftligt material hade reliabiliteten kunnat stärkas ytterligare, men då hade jag inte kunnat besvara min andra frågeställning som handlar om vilken grad av fokus orden får i undervisningen; lärarens pedagogiska val i klassrumssituationen går inte att få syn på genom att analysera läromedel. Denscombe (2018:297f) menar att observation tar oss mycket nära de riktiga händelserna och att vi med metoden undviker det avstånd som uppstår mellan forskaren och verkligheten vid andra metoder så som textanalys, intervjuer eller frågeundersökningar.

En möjlig problematik för studiens validitet går att finna i metoden frekvensbaserad lexikalisk analys. Meara (2005:40) testade metoden med hjälp av simulerade inlärttexter. Han kom fram till att verktyget är pålitligt så länge man undersöker stora skillnader i storlek på ordförråd, men att det är ett trubbigt verktyg för att analysera mindre skillnader. I en replik av Laufer (2005) kritiserar hon dock Mearas metod att använda simulerade inlärttexter. Sammanfattningsvis har alltså metoden fått kritik på basis av att den kanske inte är tillräckligt nyanserad för att analysera små skillnader i ordförråd. Trots detta tvivel är det ett välanvänt verktyg som idag är den bästa metoden att tillgå för uppsatsens syfte. Lindberg (2007:46) menar också att resultat från frekvensbaserad lexikalisk analys visat på stor korrelation med andra ordförrådstest, vilket styrker dess validitet.

Frågan är också om mina resultat har extern validitet, alltså är generaliserbara. Försiktighet krävs vid antaganden om generaliserbarhet. Mitt urval är mycket litet, så resultaten bör snarare ses som indikationer och en möjlig utgångspunkt för vidare studier.

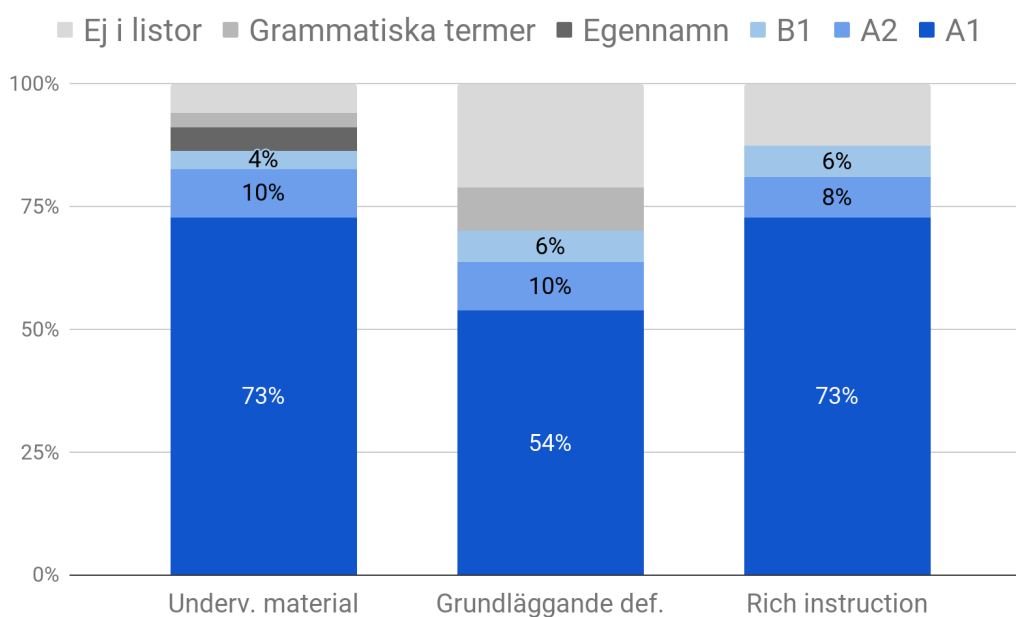
6. Resultat och analys

Nedan redovisas resultat och analys. Först redovisas resultatet för materialet som helhet för att besvara de två första frågeställningarna, därefter redovisas resultat för skillnader mellan kurs C och D för att besvara den tredje frågeställningen. Till sist i resultatdelen redovisas partikelverb och reflexiva verb vilka inte inkluderades i den frekvensbaserade lexikaliska analysen. Därefter följer en analys av resultaten.

6.1. Resultat

6.1.1. *Grad av fokus på de högfrekventa orden*

Nedan följer resultatet av den frekvensbaserade lexikala analysen av materialet. Listorna som jag bearbetade utifrån KELLY-listan (med namnen A1, A2 och B1) refererar jag till som *frekvensordlistorna*. I figur 4 nedan redovisar jag alla tre materialkategorier och hur stor del av orden i dem som täcktes av de olika listorna. För att underlätta analysen av vilka ord som inte täcks av frekvensordlistorna inkluderar jag även hur många ord som bestod av egennamn och hur många ord som kan betraktas som grammatiska termer i figur 4. Notera att försiktighet krävs vid jämförelser mellan undervisningsmaterialet och de andra två kategorierna på grund av att jag samlade in undervisningsmaterialet på ett annat sätt än de andra kategorierna. Det innebär att undervisningsmaterialet mest består av hela meningar medan de andra kategorierna består av listor på lösryckta ord vilket påverkar hur det går att tolka täckningsgraden.



Figur 4. Andel av materialkategorierna som täcks av frekvensordlistorna

Undervisningsmaterialet innehåller 2751 tokens men bara 618 types.² Frekvensordlistorna täcker 87%³ av materialet. Med egennamn och grammatiska termer täcks hela 94%. De ord som inte täcks av någon av listorna omfattar 93 types. Några av dessa 93 types är genomskinliga sammansättningar som troligen skulle täckas av frekvensordlistorna om orden stod var för sig (t.ex. *favoritfärg, dubbelsäng, golvlampa, chokladglass*). Några andra ord täcks inte av frekvensordlistorna men är ändå väntade i ett skolsammanhang på grund av att de är starkt förknippade med skolan som domän (t.ex. *sportlovsvecka, höstlovsvecka, hörförståelseuppgift, studiedag, klassrum, klasskamrat, läxa*). De flesta ord som inte täcks av någon av listorna är dock varken genomskinliga sammansättningar eller ord från skoldomänen. Exempel på övriga ord är *tvättmaskin, äpple, hen, intro, utstrålning, pannkaka* och *snarka*.

Listan med ord som fick *grundläggande definition* innehåller totalt 190 tokens och 148 types. Orden täcks till 70% av frekvensordlistorna, och tillsammans med de

² Partikelverb och reflexiva verb är exkluderade ur beräkningen av token och types.

³ Alla procenttal i resultatdelen är avrundade till heltal.

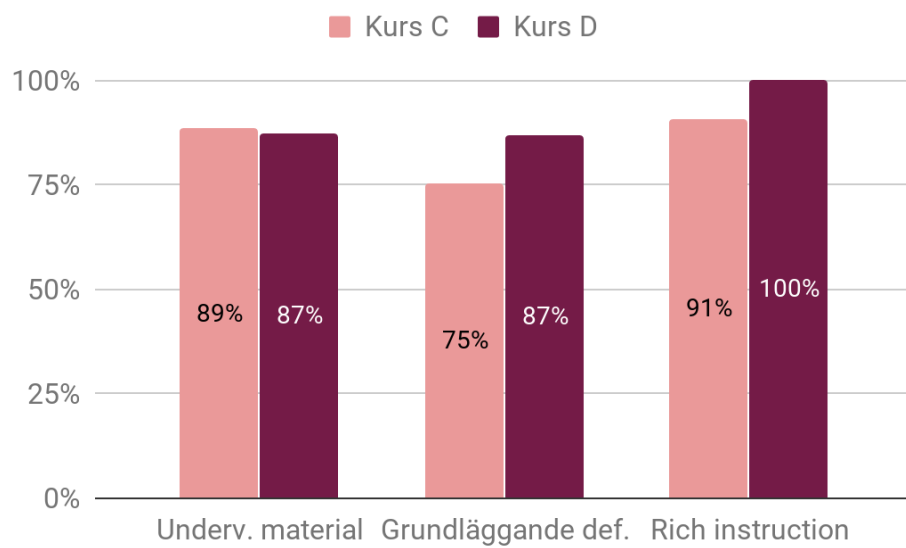
grammatiska termerna täcks 79%. De övriga orden består av 38 types. Bara ett ord är förknippat med skoldomänen (*höstlovsvecka*) och några få ord kan betraktas som genomskinliga sammansättningar.

Slutligen innehåller listan av ord som fick *rich instruction* 95 tokens och 70 types. Dessa ord täcks till 87% av frekvensordlistorna. De som inte täcks är 12 types. Av dessa är inga specifika för skoldomänen och inget ord är genomskinligt på ett sånt sätt att man kan anta att inlärarna förstår ordet utan instruktion. Exempel på ord som inte täcks av frekvensordlistorna är *baka*, *fat*, *trivsamt* och *vandrarhem*.

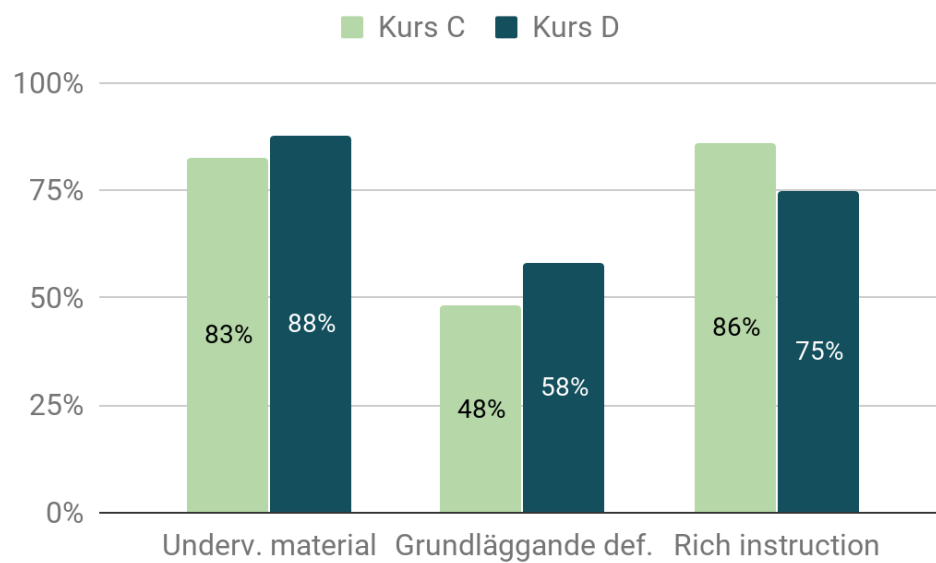
I figur 4 kan man se att de ord som fick *rich instruction* i högre grad tillhör svenskans vanligaste ord än de ord som fick *grundläggande definition*. Skillnaderna i täckningsgrad mellan de två materialkategorierna är ett resultat av att täckningsgraden av A1-listan skiljer sig med 19 procentenheter mellan materialkategorierna. Samtidigt täcker A2 och B1 ungefär lika mycket av orden i *grundläggande definition* och *rich instruction*. Den stora skillnaden har alltså främst att göra med skillnader i täckningsgrad av A1, det vill säga de allra vanligaste orden.

6.1.2. Progression mellan kurs C och D

Resultaten visar att det inte fanns någon trend från mer vanliga till mindre vanliga ord mellan de observerade lektionerna på kurs C och kurs D. I figur 5 och figur 6 avser procenttalen hur stor del av materialet som täcks av alla frekvensordlistorna (A1, A2 och B1) sammanslaget. I endast två av sex fall är täckningsgraden högre på kurs C än på kurs D: i undervisningsmaterialet på Camillas lektioner och bland de ord som fått *rich instruction* på Sivs lektioner. Det betyder att eleverna på kurs C *inte* mötte mer frekventa ord än eleverna på kurs D vid de observerade tillfällena. Alltså erbjuder mitt material inga bevis för att det skulle skett en progression mot allt ovanligare ord vid de observerade tillfällena.



Figur 5. Jämförelse av täckningsgrad på Camillas lektioner



Figur 6. Jämförelse av täckningsgrad på Sivs lektioner

6.1.3. Partikelverb och reflexiva verb

Partikelverb och reflexiva verb sorterades ut manuellt innan analysen i AntWordProfiler med anledning av att programmet inte kan analysera flerordsenheter. Jag redovisar dem istället under nedanstående rubriker.

6.1.3.1. Partikelverb

18 unika partikelverb förekommer i materialet. Endast ett partikelverb fick *rich instruction: ställer in*. Sex ord fick *grundläggande definition: tycker om, ser ut, torkar av, koppla av, slappna av* och *turas om*. Dessa partikelverb förekom även i undervisningsmaterialet. Förutom dessa förekom även partikelverb som ej berördes explicit: *fylla år, gå ut, gå igenom, kopplas ihop (med), resa bort, sätta ut, hälsa på* och *bli av*.

Tabell 2. Förekomst av partikelverb fördelat i materialet

Underv. material	Antal (av 2829 tokens)	Grundläggande def.	Antal (av 210 tokens)	Rich instr. tokens)	Antal (av 97 tokens)
<i>tycka om</i>	7	<i>tycka om</i>	1	<i>ställa in</i>	1
<i>fylla år</i>	4	<i>se ut</i>	1		
<i>se ut</i>	3	<i>torka av</i>	1		
<i>hälsa på</i>	2	<i>koppla av</i>	1		
<i>ställa in</i>	2	<i>slappna av</i>	1		
<i>torka av</i>	2	<i>turas om</i>	1		
<i>bli av</i>	2				
<i>turas om</i>	2				
<i>gå ut</i>	1				
<i>gå igenom</i>	1				
<i>kopplas ihop</i>	1				
<i>resa bort</i>	1				
<i>sätta ut</i>	1				
<i>koppla av</i>	1				
<i>slappna av</i>	1				

6.1.3.2. Reflexiva verb

Inga reflexiva verb fick *rich instruction*. Fyra reflexiva verb fick *grundläggande definition*: *kamma sig*, *känna sig*, *säga sig* och *lära sig*. Dessa förekom även i undervisningsmaterialet. Förutom dessa förekom *lägga sig* två gånger i undervisningsmaterialet.

Tabell 3. Förekomst av reflexiva verb fördelat i materialet

Underv. material	Antal (av 2829 tokens)	Grundläggande def.	Antal (av 210 tokens)
<i>lägga sig</i>	2	<i>kamma sig</i>	1
<i>kamma sig</i>	2	<i>känna sig</i>	1
<i>säga sig</i>	2	<i>säga sig</i>	1
<i>känna sig</i>	1	<i>lära sig</i>	1
<i>lära sig</i>	1		

6.2. Analys av resultat

Uppsatsens första frågeställning lyder: I hur stor utsträckning lägger lärare i SFI C och D fokus på svenskans högfrekventa ord i klassrumsundervisningen? Mitt resultat visar att det inte bara är högfrekventa ord som får fokus på lektionerna, utan att materialet även innehåller en del ord som inte är med i frekvensordlistorna. Bland de ord som inte var med i frekvensordlistorna fanns en del grammatiska termer. Jag tog med dem i undersökningen eftersom det är en tydligt avgränsad grupp av ord som även har ett tydligt syfte i andraspråkelevernas ordförråd fastän de inte tillhör de vanligaste i språket i stort. Tillsammans med egennamn och grammatiska termer täcker frekvensordlistorna hela 94% av undervisningsmaterialet, vilket är lika många procent som KELLY-listans 4212 översta lemman täcker sin egna korpus (inklusive egennamn och skiljetecken; Volodina & Johansson Kokkinakis 2012b:43). Alltså är sammansättningen mellan ovanliga och vanliga ord ungefär likadan i undervisningsmaterialet som i den korpus KELLY-listan utgick ifrån – ett material som inte är anpassat för inlärare. En del av de ord som inte täcks av

frekvensordlistorna är vardagsnära ord som kan vara viktiga för inlärare men som är underrepresenterade i textkorpusar, exempelvis den som KELLY skapades utifrån. Exempel på vardagsord från materialet är *baka*, *äpple*, *tandkräm* och liknande. Men även utan att räkna med de här orden är det många ord som fortfarande kan betraktas som ovanliga och inte tillhörande de högfrekventa orden. Utifrån detta resultat verkar det finnas utrymme att ha ännu mer fokus på de vanligaste orden i undervisningen.

Den andra frågeställningen söker svar på om det går att se om ord som får mer fokus består av vanligare ord än de som får mindre fokus eller vice versa. För att söka svar på den här frågan går det inte att göra jämförelser mellan undervisningsmaterialet och de två andra materialkategorierna på grund av att de har olika karaktär. Undervisningsmaterialet består nämligen i stor utsträckning av hela meningar vilket betyder att formord förekommer många gånger medan de andra kategorierna består av ordlistor borttryckta från sitt sammanhang. Det är dock möjligt att jämföra frekvensnivån på ord som fick *grundläggande definition* och de som fick *rich instruction*. Resultatet visar att de ord som fick *rich instruction*, och som alltså fick mer fokus än de andra orden, i högre utsträckning tillhör svenskans högfrekventa ord än vad de ord som fick *grundläggande definition* gör. Sambandet syns även i figur 5 och 6 där resultatet delats upp efter lärare och kurs. En förklarande faktor kan vara att många av orden som fick *rich instruction* var formord, och att formord ofta får mycket uppmärksamhet i samband med grammatikundervisning. Till exempel hade Siv en genomgång om pronomen där hon beskrev pronomens grammatiska funktion och lät eleverna förstå dem med hjälp av exempelmeningar, och gav således orden *rich instruction*. Även Camilla hade liknande övningar med underordnade konjunktioner. De ord som fick *grundläggande definition* förklarades i sin tur ofta i situationer där de hindrade eleverna från att förstå innehållet i undervisningsmaterialet, och bestod alltså oftare av innehållsord så som verb och substantiv. Denna förklaringsmodell förstärks av att den största skillnaden mellan *grundläggande definition* och *rich instruction* gick att finna i den lista som innehåller flest formord, A1-listan (se figur 4). Att de ord som fick mest fokus också oftare var högfrekventa i svenskan syns tydligt i materialet eftersom mönstret är detsamma på alla fyra observerade lektioner (se figur 5 och 6)

trots att de högfrekventa ordens täckningsgrad är lägre på Sivs lektioner än på Camillas.

Uppsatsens tredje och sista frågeställning handlar om det finns en progression mellan kurserna med avseende på om lärarna i större utsträckning fokuserar på högfrekventa ord i C-kursen än i D-kursen. Resultatet erbjuder inga bevis på att så är fallet. Resultatet skulle till viss del kunna förklaras av att många formord används i mer komplicerade satser vilka ofta får större fokus i kurs D än i kurs C, exempelvis ägnade Camillas elever på D-nivå en hel lektion till underordnade konjunktioner i bisatser. I kontrast till lektionerna på D-nivå var lektioner på C-nivå mer inriktade på konkreta situationer och berättelser. Speciellt Camillas C-lektion hade mycket fokus på verb och konkreta innehållsord istället för på formord eftersom de övade på helfraser och pratade om fritidsaktiviteter som hör till olika årstider. En annan orsak till att det är ungefär samma täckningsgrad på kurs C och D kan vara att lektionerna har olika teman. Exempelvis hade det kanske syntts en progression mellan Sivs lektioner om de hade bearbetat texter på tema morgonrutiner på båda kurserna. En sista möjlig förklaring är att lärarna naturligtvis kan ha valt material som inte är adekvat för målgruppen. Inom varje grupp kan det dock finnas en stor variation på elevernas nivå som lärarna måste förhålla sig till, vilket gör att skillnaderna mellan kurserna kanske inte alltid blir så markant. Ordförråd är dessutom inte det enda i undervisningen som kan visa progression, alltså presenterar jag den här förklaringen med viss försiktighet.

7. Diskussion

I metoddiskussionen behandlar jag framför allt metodologiska val som kan ha påverkat utfallet, därefter diskuterar jag resultaten för att till sist föreslå möjliga utgångspunkter för vidare forskning.

7.1. Metoddiskussion

Besluten om vad som ska räknas som högfrekventa ord och om hur KELLY-listan skulle bearbetas inför analysen spelar roll för uppsatsens resultat. Annorlunda val hade kunnat ändra resultatet, eller påverka hur resultaten framställdes.

För det första hade avgränsningen av högfrekventa ord kunnat göras utifrån andra eller fler parametrar. En vanlig enhet i frekvensbaserad lexikalisk analys (se exempelvis Laufer & Nation 1995) är att utgå från frekvensband på 1000 ordenheter. Jag valde dock att istället utgå från KELLY-listans GERS-uppdelning, där varje nivå innehöll 1404 lemmen, eftersom det är samma GERS-nivåer som SFI-kurserna avser lära ut. Här finns det dock anledning att kritisera mängden lemmen per GERS-nivå utifrån Miltons undersökning (2010) vilken indikerar att antalet lemmen per nivå inte är så stort som KELLY-listan föreslår. Valet mellan frekvensband på 1000 eller 1404 lemmen var dock av mindre betydelse på grund av den ringa förekomst av svenska studier av lärares ordförråd i språkundervisning att jämföra resultaten med. Men att följa GERS-uppdelningen gav åtminstone möjligheten att i analysen jämföra KELLY-listans täckningsgrad av mitt material med täckningsgraden av KELLY-listan på dess korpus.

Viktigare var valet av gränstal för hur många ord som skulle räknas som högfrekventa och därmed användas för att analysera materialkategorierna. Nation (2013:23) skriver att det föreslagits att de 2000–3000 vanligaste ordfamiljerna i engelskan kan räknas till de högfrekventa. Detta tal är inte helt jämförbart med svenskan eftersom språken har olika struktur vilket påverkar ordfamiljernas storlek. Dessutom utgår KELLY-listan från lemmen och inte från ordfamiljer. Samtidigt är 2000–3000 ordfamiljer den bästa referenspunkt jag har för en definition av högfrekventa ord i svenskan. Listorna A1, A2 och B1 täcker tillsammans 4212 lemmen vilket är fler än vad Nation föreslår för ordfamiljer, vilket är brukligt då ordfamiljer innehåller fler ”ord” än lemmen. En annan utgångspunkt för avgränsningen var, som beskrivet ovan i 4.3., när behållningen av att lära sig ord i frekvensordning inte längre är lika märkbar. De 1404 översta orden i KELLY-listan täcker 69,6% av korpusen, de 1404 som kommer därefter täcker 5,4%, därefter 2,3% men efter det kryper täckningsgraden ner på mycket små tal. Om jag även hade räknat med B2 listan, med en täckningsgrad på 1,3% av korpusen, hade resultatet visat att lite fler ord täckts av frekvensordlistorna och lite färre ord hade räknats som ovanliga.

För det andra förutsätter metoden att KELLY-listan är en välkomponerad frekvensordlista som faktiskt lämpar sig för andraspråksinlärare. Volodina och Johansson Kokkinakis (2012b:29) uppmärksammade under deras arbete med listan

att den saknade en del vardagsord och andra ord som är viktiga för språkinlärare, även om fler ord från vardagsdomänen kom med i listan senare i processen. De hade två förslag på förklaringar till varför vardagsord var underrepresenterade: antingen att dessa ord var underrepresenterade i korpusen som de använde jämfört med vardaglig kommunikation eller att vardagsord helt enkelt inte förtjänar en plats i en bas-vokabulär på grund av att de hör till specifika domäner (Ibid.:29). De överlämnar till läraren att anpassa listan genom att lägga till de domänspecifika ord som är viktiga för den specifika språkkursen (Ibid.:30). Anledningen till att exempelvis Camillas lektion på C-nivå, som handlade om fritidsaktiviteter och årstider, hade en ganska låg andel högfrekventa ord kan alltså bero på att orden som lärdes ut var mer domänspecifika än allmänna. Nation skriver att man utöver de högfrekventa orden bör fokusera på domänspecifika ord som är av värde för den individuella inläraren (2013:94). Alltså är det upp till lärarens bedömning, i samråd med eleverna, att avgöra om exempelvis fritidsaktiviteter är någonting som är viktigt för inlärarna att kunna samtala om.

För det tredje påverkas utfallet av hur jag valde att hantera flerordsuttryck. I valet mellan att ta med eller exkludera flerordsenheter uppstår en konflikt mellan validitet och reliabilitet. Jag exkluderade partikelverb och reflexiva verb ur analysen i AntWordProfiler men lät andra flerordsuttryck vara kvar i materialet. Detta var en typ av kompromiss mellan att särbehandla alla flerordsuttryck, på bekostnad av reliabilitet på grund av att flerordsenheter inte alltid är objektivt identifierbara, och att behandla alla flerordsuttryck som enskilda ord, på bekostnad av validiteten. I ett första försök till analys testade jag att exkludera fler flerordsuttryck och när jag sedan ändrade till att bara exkludera partikelverb och reflexiva verb visade det sig att skillnaden inte blev så stor, vilket är värt att notera.

För det fjärde och sista baseras resultaten på tokens och inte på types. Meara et al. (1997:35f) understryker att valet i hög grad påverkar resultaten eftersom de ovanliga orden oftast har färre upprepningar i en text än de vanligare orden. Om jag hade använt types istället för tokens hade mitt resultat visat att en mindre andel av orden skulle räknas till de högfrekventa. Detta skulle ge störst utslag på undervisningsmaterialet eftersom det innehåller hela meningar där formord ofta upprepas, till skillnad från listorna med *rich instruction* och *grundläggande definition* på grund av att de orden sällan förklaras flera gånger. En anledning till att

jag valde tokens istället för types var att det skulle försvåra uppdelningen av materialet baserat på lärare och kurs. Ibland förekom samma ord både på Camillas kurs C och på Sivs kurs C, exempelvis *favorit*. Om jag hade baserat min analys på types hade ordet *favorit* bara räknats en gång i det sammanlagda resultatet. I delen där jag delade upp resultatet på lärare och kurs hade dock ordet *favorit* behövt vara en del av resultatet i flera olika analyser. Det hade gjort resultatet mer svårtolkat.

7.2. Resultatdiskussion

Som jag konstaterade i analysen finns det alltså utrymme för lärarna att fokusera mer på de vanligaste orden. Dock är det inte en självklarhet att undervisningen vore bättre om alla ord täcktes av frekvensordlistorna. Meara et al. (1997:2) utgick i deras studie från premissen att eleverna inte möter sitt andraspråk så ofta i vardagen och att de därför är i stort behov av en ”rich lexical environment” i undervisningen. Utifrån en sådan synvinkel kan man betrakta resultatet som mer positivt ju fler ovanliga ord som används. Jag utgår dock istället från Nations rekommendation att lärare först och främst bör lära ut de högfrekventa orden och utöver dem endast fokusera på ord från domäner som är viktiga för inlärarna (2013:94). Det innebär att de ord i *grundläggande definition* och *rich instruction* som inte hör till de högfrekventa orden bör tillhöra domäner som eleverna vill kunna kommunicera inom. Ur den synvinkeln visar resultatet att det kan finnas förbättringspotential för att effektivisera ordundervisningen på SFI, beroende på om man bedömer att de ”ovanliga” orden kommer från domäner som är viktiga för eleverna eller ej. En möjlig kompromiss mellan synsätten är att lärare skulle kunna skapa en ”rich lexical environment” genom att använda högfrekventa ord i högre grad men samtidigt utnyttja bredden av de 4212 ord som ingår i listan.

Resultatet visar också att de ord som fick *rich instruction* generellt sett tillhör de vanligare orden jämfört med de ord som fick *grundläggande definition*. Nation (2013:117) skriver att *rich instruction* är bäst att använda först efter att eleverna stött på ordet några gånger eftersom den typen av ordundervisning syftar till att få eleverna att införliva ordet i sitt produktiva ordförråd. Alltså visar resultatet positiva tendenser som går i linje med syftet med *rich instruction*. Samtidigt är det inte självklart om en tydligare tendens hade varit bättre eller sämre eftersom det inte finns någon tidigare forskning att jämföra resultatet med på den här punkten.

I analysen skrev jag att det kan finnas flera förklaringar till att det inte syntes någon progression mellan kurserna. Dock stärks mitt resultat av att inte heller Meara (1993) fann någon progression mellan kurserna på BBC English. Trots att det är olika språk och olika frekvensordlistor borde frekvensskillnader fortfarande vara jämförbara och båda resultaten visar på att det inte verkar ske någon progression naturligt. Detta får implikationen att lärare som vill fokusera på de högfrekventa orden först aktivt måste arbeta för att lära ut ord i frekvensordning. Utifrån ett sådant antagande har lektionerna vid de observerade tillfällena en förbättringspotential med avseende på ordundervisning. Värt att notera är dock att det finns en risk att mätverktyget, som Meara (2005) föreslår, inte är tillräckligt känsligt för att med säkerhet upptäcka skillnader i ordförråd i mindre textutdrag. I textutdrag bestående av 300 ord kunde hans simulering endast i 67% av fallen med säkerhet upptäcka en skillnad mellan ordförråd på 3000 och 3500 ord. Alltså skulle det kunna vara slumpen som styr ifall det handlar om mindre skillnader i det ordförråd som används, speciellt om underlaget är litet, vilket skulle kunna vara fallet i jämförelserna mellan kurs C och kurs D.

7.3. Slutsatser och förslag på vidare forskning

Mitt material kommer från nedslag i enskilda lärares undervisning som inte säkert är representativa för hela kurserna eller för lärare i allmänhet. För att kunna dra några egentliga slutsatser krävs alltså ett större material och ett större urval. Ett större material hade kunnat ge svar på några av de osäkerheter som jag skrev om i analysen, exempelvis om den uteblivna progressionen hade att göra med vilka teman de hade på lektionerna.

Mitt resultat visar dock indikationer på att lärare måste göra aktiva val för att fokusera på de högfrekventa orden och för att skapa en progression i ordundervisningen från mer frekventa till mindre frekventa ord. Dock är det viktigt att samtidigt se att det finns andra värden i ordundervisningen som kan visa progression mellan kurser, exempelvis fokus på ordförrådets djup eller domänspecifikt ordförråd. Någoting som verkar ske naturligt i undervisningen är att de ord som får *rich instruction* generellt brukar vara vanligare ord än de som får *grundläggande definition*. Detta går i linje med det Nation skriver om att *rich*

instruction är bäst när det syftar till att införliva ordet i det produktiva ordförrådet (2013:117).

Förslag på vidare forskning är för det första att göra en liknande undersökning med ett större material. En annan väg framåt vore att undersöka spridningen på de ord som används. Möter inlärarna många olika ord eller är det ett mer begränsat urval som de möter i undervisningen? Återkommer orden tillräckligt många gånger för att eleverna ska ha en chans att lära sig orden genom språkligt inflöde? Lär lärarna ut en stor bredd av de högfrekventa orden eller är det någon kategori inom de högfrekventa orden som konsekvent försummas? Ett annat alternativ vore att undersöka de ord som *inte* täcks av frekvensordlistorna. Tillhör dessa ord domäner som är relevanta för eleverna? Ett tredje förslag på vidare forskning vore att fokusera mer på läromedel och undersöka om läroböckerna på tidigare nivåer innehåller en större andel högfrekventa ord än läroböcker med högre svårighetsgrad.

8. Litteraturförteckning

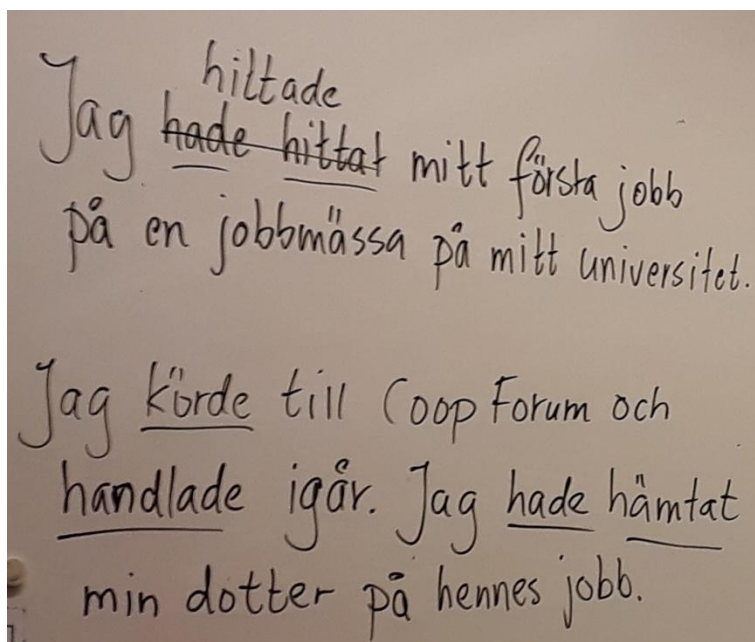
- Allén, Sture 1970–1980. *Nusvensk frekvensordbok baserad på tidningstext: Frequency dictionary of present-day Swedish based on newspaper material*. Stockholm: Almqvist & Wiksell international.
- Anthony, Laurence 2014. AntWordProfiler (Version 1.4.1) [Computer Software]. Tokyo, Japan: Waseda University. <<http://www.laurenceanthony.net/software>>.
- Bauer, Laurie & Paul Nation. Word families. *International journal of Lexicography*, 6/4: 253–279.
- Borin, Lars, Markus Forsberg, Martin Hammarstedt, Dan Rosén, Roland Schäfer, Anne Schumacher 2016. Sparv: Språkbanken's corpus annotation pipeline infrastructure. I: *Proceedings of The Sixth Swedish Language Technology Conference (STLC)*. Umeå, Sverige 17–18 November 2016.
- Chaudron, Craig 1982. Vocabulary elaboration in teachers' speech to L2 learners 1. *Studies in Second Language Acquisition*, 4/2: 170–180.
- Denscombe, Martyn 2018. *Forskningshandboken: för småskaliga forskningsprojekt inom samhällsvetenskaperna*. 4 uppl. Lund: Studentlitteratur.
- Ellis, Rod 1995. Modified Oral Input and the Acquisition of Word Meanings. *Applied Linguistics*, 16/4: 409–441.
- Enström, Ingegerd 2010. *Ordens värld: Svenska ord – struktur och inläring*. Stockholm: Hallgren & Fallgren.
- Esaiasson, Peter, Mikael Gilljam, Henrik Oscarsson, Ann E. Towns & Lena Wängnerud 2017. *Metodpraktikan: konsten att studera samhälle, individ och marknad*. 5 uppl. Stockholm: Wolters Kluwer.
- Forsbom, Eva 2006. A Swedish Base Vocabulary Pool. I: *Swedish Language Technology Conference (STLC)*. Göteborg, Sverige 27–28 Oktober 2006: 1–4.
- François, Thomas, Elena Volodina, Ildikó Pilán & Anaïs Tack 2016. SVALex: a CEFR-graded lexical resource for Swedish foreign and second language learners. I: Calzolari, Nicoletta, Khalid Choukri, Thierry Declerck, Sara Goggi, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk & Stelios Piperidis. *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*. Portorož, Slovenia: 213–219.

- Gardner, Dee & Mark Davies 2013. A new academic word list. *Applied linguistics*, 35/3: 305–327.
- Horst, Marlise 2010. How Well Does Teacher Talk Support Incidental Vocabulary Acquisition?. *Reading in a Foreign Language*, 22/1: 161–180.
- Johansson, Emma 2017. *Om Sverige: en kvantitativ textanalys av ett material för samhällsorientering med fokus på ordförråd och läsbarhet ur ett andraspråksperspektiv*. St. upps., Göteborgs universitet.
- Johansson Kokkinakis, Sofie 2007. Språkteknologiskt arbete i OrdiL-projektet. I: Lindberg, Inger & Sofie Johansson Kokkinakis (red.), *OrdiL: en korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år (ROSA 8)*. Göteborg: Institutet för svenska som andraspråk, Institutionen för svenska språket, Göteborgs universitet: 101–134.
- Johansson Kokkinakis, Sofie & Birgitta Frändberg 2013. Högstadielävers användning av naturvetenskapligt språkbruk kemiämnet i TIMSS. *Utbildning & demokrati*, 22/3: 53–68.
- Järborg, Jerker 2007. Om ord och ordkunskap. I: Lindberg, Inger & Sofie Johansson Kokkinakis (red.). *Ordil: en korpusbaserad kartläggning av ordförrådet i läromedel i grundskolans senare år (ROSA 8)*. Göteborg: Institutet för svenska som andraspråk, Institutionen för svenska språket, Göteborgs universitet: 101–134.
- Kilgarriff Adam, Charalabopoulou Frieda, Gavrilidou Maria, Bondi Johannessen Janne, Khalil Saussan, Sofie Johansson Kokkinakis, Robert Lew, Serge Sharoff, Ravikiran Vadlapudi & Elena Volodina 2014. Corpus-Based Vocabulary lists for Language Learners for Nine Languages. *Language Resources and Evaluation Journal*, 48/1: 121-163.
- Laufer, Batia 2005. Lexical frequency profiles: From Monte Carlo to the real world: A response to Meara (2005). *Applied linguistics*, 26/4: 582–588.
- Laufer, Batia & Paul Nation 1995. Vocabulary size and use: lexical richness in L2 written production. *Applied linguistics*, 16/3: 307–322.
- Lindberg, Inger 2006. Bedömning av skolrelaterat ordförråd. I: Olofsson, Mikael (Red.), *Symposium 2006: bedömning, flerspråkighet och lärande*. Stockholm: HLS förlag: 83–107.

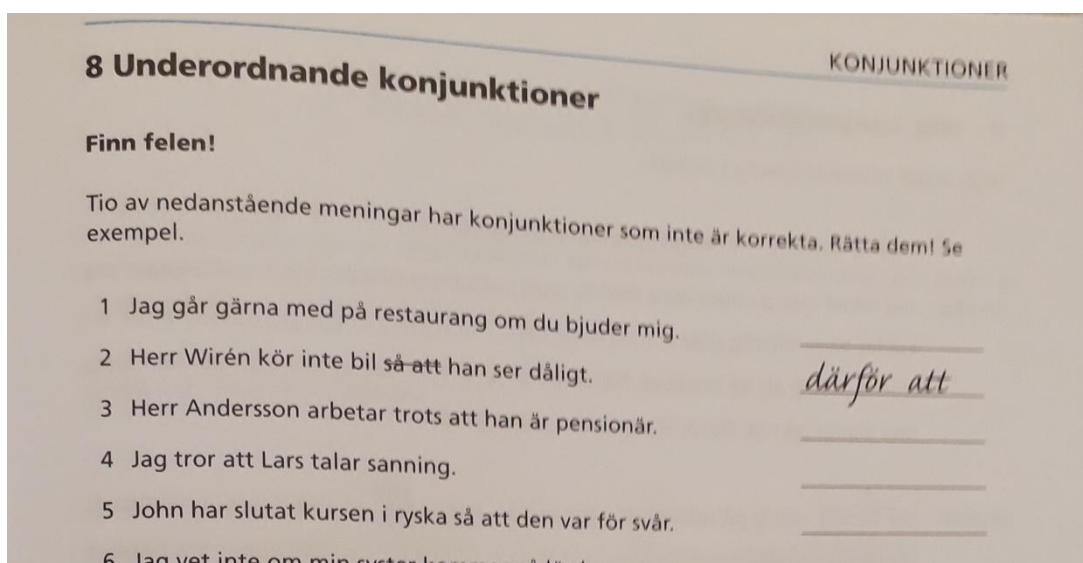
- Lindberg, Inger 2007. Forskning om läromedelsspråk och ordförrådsutveckling. I: Lindberg, Inger & Sofie Johansson Kokkinakis (red.). *Ordil: en korpusbaserad kartläggning av ordförrådet i läromedel i grundskolans senare år* (ROSA 8.) Göteborg: Institutet för svenska som andraspråk, Institutionen för svenska språket, Göteborgs universitet: 13–60.
- McLean Stuart 2017. Evidence for the adoption of the flemma as an appropriate word counting unit. *Applied Linguistics*, 39/6: 823–845.
- Meara, Paul 1993. Tintin and the world service: A look at lexical environments. I: *LATEFL: Annual Conference Report*: 32–37.
- Meara, Paul 2005. Lexical frequency profiles: A Monte Carlo analysis. *Applied linguistics*, 26: 32–47.
- Meara, Paul, Patsy M. Lightbown & Randall H. Halter 1997. Classrooms as lexical environments. *Language Teaching Research*, 1/1: 28–46.
- Milton, James 2010. The development of vocabulary breadth across the CEFR levels. I: Bartning, Inge, Maisa Martin & Ineke Vedder (red.), *Communicative proficiency and linguistic development: Intersections between SLA and language testing research*. Eurosla Monographs Series, 1: 211–232.
- Nation, Ian S. P. 2013. *Learning vocabulary in another language*. 2 uppl. New York: Cambridge University Press.
- Ohlsson, Elisabeth 2018. *Att synliggöra vokabulär: en kvantitativ studie av gymnasieelevers skrivuppgifter på svenska inom CLISS-projektets ram*. Lic.-avh., Göteborgs universitet.
- Pilán, Ildikó, Elena Volodina & Lars Borin 2017. Candidate sentence selection for language learning exercises: from a comprehensive framework to an empirical evaluation. *TAL*, 57/3: 67–91.
- Schmitt Norbert 2008. Instructed second language vocabulary learning. *Language teaching research*, 12/3: 329–363.
- Schmitt, Norbert & Diane Schmitt 2014. A reassessment of frequency and vocabulary size in L2 vocabulary teaching. *Language Teaching*, 47/4: 484–503
- SKOLFS 2009:2. *Förordning om kursplan för kommunal vuxenutbildning i svenska för invandrare*.
- SKOLFS 2017:91. *Skolverkets föreskrifter om kursplan för kommunal vuxenutbildning i svenska för invandrare*.

- Skolverket 2009. *Gemensam europeisk referensram för språk: lärande, undervisning och bedömning*. <<https://www.skolverket.se/publikationer?id=2144>>. Hämtat 2018-12-04.
- Skolverket 2018. *Beskrivande data 2017: förskola, skola och vuxenutbildning*. Rapport 468. <<https://www.skolverket.se/publikationer?id=3953>>. Hämtat 2018-01-29.
- Volodina, Elena, Ildikó Pilán & David Alfter 2016. Classification of Swedish learner essays by CEFR levels. I: *CALL communities and culture—short papers from EUROCALL 2016*: 456–461.
- Volodina, Elena & Sofie Johansson Kokkinakis 2012a. Introducing Swedish Kelly-list, a new lexical e-resource for Swedish. I *Proceedings of LREC 2012*, Istanbul, Turkiet.
- Volodina, Elena & Sofie Johansson Kokkinakis 2012b. *Swedish Kelly: Technical Report*. GU-ISS-2012-01. Språkbanken: Göteborgs universitet.

Bilaga 1



Figur 7. Exempel på taveltext från pilotobservation.



Figur 8. Exempel på stencil

Bilaga 2

Samtycke till deltagande i studie

Om studien

Syftet med studien är att undersöka vilka ord som förekommer i undervisningen på SFI kurs C och D. Den huvudsakliga metoden är klassrumsobservationer vilka kompletteras med det skriftliga material som figurerar i undervisningen. Till det skriftliga materialet räknas bland annat material som delas ut till eleverna, sidor ur läroböcker som eleverna läser under lektionen, presentationer som visas för hela klassen och vad som skrivs på tavlan. Studien resulterar förhoppningsvis i fördjupade insikter om vilka ord elever möter tidigt i svenskundervisningen.

Konfidentialitet och din medverkan

Jag närvarar i 2-3 undervisningsgrupper under ungefär en lektionstimme i vardera grupp. Under lektionerna kommer jag att fotografera tavlan ifall den används. Tavelskriften kommer överföras till ett textdokument och därefter raderas bilderna. Jag samlar även in övrigt undervisningsmaterial som eleverna mött under lektionen från dig. Inga ljud- eller filminspelningar görs. Observationerna kommer vara anonymiserade och ditt namn kommer inte figurera i uppsatsen. Ej heller skolans eller elevens namn kommer att skrivas ut.

Observationerna och det insamlade materialet kommer att användas i forskningssyfte. Studien kommer publiceras i form av en studentuppsats vilken senare går att finna på Göteborgs universitets hemsida (<https://svenska.gu.se/publikationer/studentuppsatser>).

Kontaktuppgifter

Leila Swartling

Telefonnummer

Mail

Göteborgs universitet

Samtycke

Jag har blivit informerad om projektet samt om att mitt deltagande är frivilligt och om att jag när som helst kan avbryta observationerna eller mitt deltagande.