

Lexikal komplexitet i gymnasistexter på olika CEFR- nivåer

Maiju Haapanen
Avhandling pro gradu
Nordiska språk, Examensprogrammet för språkinläring och
språkundervisning
Institutionen för språk- och översättningsvetenskap
Humanistiska fakulteten
Åbo universitet
Maj 2020

Turun yliopiston laatujärjestelmän mukaisesti tämän julkaisun alkuperäisyys on tarkastettu Turnitin OriginalityCheck -järjestelmällä.

ÅBO UNIVERSITET

Institutionen för språk- och översättningsvetenskap / Humanistiska fakulteten

HAAPANEN, MAIJU: Lexikal komplexitet i gymnasisttexter på olika CEFR-nivåer

Avhandling pro gradu, 63 s., bilagor 7 s.

Nordiska språk, Examensprogrammet för språkinläring och språkundervisning

Maj 2020

Syftet med avhandlingen har varit att undersöka lexikal komplexitet i texter skrivna av finska inlärare av svenska på olika CEFR-nivåer. Den lexikala komplexiteten har studerats med hjälp av lexikal variation, dvs. hur pass varierat ordförråd inlärare använder i sina texter, och lexikal täthet, dvs. hur mycket innehållsord inlärare använder i sina texter. Därtill har jag studerat frekvensen av olika adjektiv och hur informanterna använder adjektiv i sina texter. Jag har undersökt om det finns ett samband mellan den lexikala variationen, den lexikala tätheten samt användningen av adjektiv och de olika CEFR-nivåerna.

Som material har jag använt 15 texter skrivna av finska gymnasister. Materialet ingår i projektet *Helsekvenser i tre morfologiskt olika inlärarespråk* vid Åbo universitet som finansieras av Svenska kulturfonden. Texterna har bedömts enligt en DIALUKI-skala som baserar sig på CEFR. Metoderna har varit både kvantitativa och kvalitativa. Jag har mätt den lexikala variationen med ett s.k. OVIX-värde som anger förhållandet mellan antalet lemmor och antalet löpord i texterna. Den lexikala tätheten har jag mätt med ett s.k. SAVKVOT-värde som räknas så att antalet innehållsord divideras med antalet löpord i texterna. Användningen av adjektiv har analyserats med hjälp av Nations (2001) modell av djup ordkunskap.

Analysen visar att gymnasisterna på högre färdighetsnivåer har lexikalt mer varierande texter än gymnasisterna på lägre färdighetsnivåer men texterna på lägre färdighetsnivåer är mer täta än texterna på högre färdighetsnivåer. På lägre färdighetsnivåer upprepas dock samma innehållsord mer vilket är anledningen till att den lexikala tätheten är hög på dessa nivåer.

Gymnasisterna använder adjektiv i lika hög grad som infödda talare av svenska. På lägre färdighetsnivåer används olika adjektiv mer än på högre färdighetsnivåer, men gymnasisterna på lägre färdighetsnivåer använder mer olämpliga adjektiv med tanke på kontexten där de uttrycker sina åsikter. De tre mest frekventa adjektiven är *social*, *bra* och *okej* och gymnasisterna kan stava och böja dessa tre adjektiv rätt. Adjektiven *bra* och *okej* används i relativt enkla ordkombinationer som ofta upprepas i texterna men på högre färdighetsnivåer används t.ex. förstärkningsprefixet *jätte* och adverb *hel* med adjektiven i större utsträckning än på lägre färdighetsnivåer. Användningen av dessa två adjektiv är således mer avancerad på högre färdighetsnivåer.

Nyckelord: Svenska som andraspråk, lexikal komplexitet, adjektiv, den europeiska referensramen

Innehåll

Abstract

1	Inledning	6
1.1	Syfte	6
1.2	Material och metod.....	7
1.3	Avhandlingens disposition	8
2	Ord och ordförråd.....	9
2.1	Allmänt om ord	9
2.2	Adjektiv	11
2.3	Ordförråd	12
3	Ordinläring.....	14
3.1	Lexikal kompetens	14
3.1.1	Ordförrådets bredd.....	15
3.1.2	Ordförrådets djup.....	16
3.2	Lexikal komplexitet.....	17
3.3	Ordinlärningsprocess.....	19
4	Gemensam europeisk referensram för språk	22
5	Tidigare forskning.....	24
6	Material och metod	29
6.1	Material	29
6.2	Metod	30
7	Resultat.....	33
7.1	Lexikal variation i texterna	33
7.2	Lexikal täthet i texterna.....	35
7.3	Adjektiv i texterna	36
7.3.1	Mest frekventa adjektiv	38
7.3.2	Användning av adjektiv i texterna.....	41

7.3.2.1	Adjektivet <i>social</i>	41
7.3.2.2	Adjektivet <i>bra</i>	44
7.3.2.3	Adjektivet <i>okej</i>	50
7.3.2.4	Andra adjektiv	52
8	Sammanfattning och diskussion	56
	Litteratur.....	61
	Bilagor.....	64
	Lyhennelmä.....	71
	Figurer	
	Figur 1 Frekvens av adjektiven <i>social</i> , <i>bra</i> och <i>okej</i> i de två grupperna.	40
	Tabeller	
	Tabell 1 Modell av produktiv djup ordkunskap enligt Nation (2001: 27) (min egen översättning till svenska).....	16
	Tabell 2 Medelvärden av antalet löpord samt antalet lemman och de lägsta och högsta antalen löpord samt lemman i de två grupperna.....	33
	Tabell 3 Medelvärdet av OVIX samt de lägsta och högsta värdena av OVIX i de två grupperna.....	34
	Tabell 4 Medelvärdet av SAVKVOT och de lägsta och högsta värdena av SAVKVOT i de två grupperna.....	35
	Tabell 5 Medelvärden av olika innehållsord samt alla innehållsord och deras förhållande i de två grupperna.....	36
	Tabell 6 Medelvärden av antalet löpord och antalet alla adjektiv i textmaterialet samt deras förhållande i de två grupperna.....	37
	Tabell 7 Medelvärden av antalet olika adjektiv samt alla adjektiv och deras förhållande i de två grupperna.....	37
	Tabell 8 Antalet informanter (f) samt procentandelen av alla informanter (%) som har använt de mest frekventa adjektiven i de två grupperna.....	38

1 Inledning

Ordförråd är en viktig del av språket. Utan lämpliga ord är det omöjligt att uttrycka vad man vill säga eller förstå vad andra säger. Det är väsentligt att språkinlärare får ett brett ordförråd men också att de lär sig att använda ordförrådet mångsidigt. Därför har många forskare börjat fästa avseende vid ordförrådets komplexitet hos språkinlärare. Exempelvis är det i skriftlig produktion viktigt att kunna använda olika ordtyper i stället för att alltid upprepa samma ord. Texten förväntas också innehålla rikligt med s.k. innehållsord (substantiv, verb, adjektiv och vissa adverb) som gör texten informationsrik. Det första fallet kallas *lexikal variation* och det andra fallet *lexikal täthet*. Dessa två begrepp används mycket i undersökningar för att beskriva ordförrådets eller textens komplexitet (se t.ex. Agebjörn 2017; De Clercq 2015; Haikala 2012; Hultman & Westman 1977).

I denna studie undersöker jag ordförrådets komplexitet i skriftliga uppsatser på olika färdighetsnivåer (CEFR-nivåer, se kapitel 4). Uppsatserna är skrivna av finska språkinlärare som studerar svenska som det andra inhemska språket. För att få en mer omfattande bild om användning av ordförråd i texterna undersöker jag också närmare hur gymnasisterna använder adjektiv i sina texter. Även om det redan finns forskning om användning av ordförråd hos språkinlärare (se kapitel 5 för tidigare forskning), valde jag temat för att vidare undersöka sambandet mellan ordförrådets komplexitet hos finska språkinlärare av svenska och de olika CEFR-nivåerna. Därtill finns det inte mycket forskning där den lexikala variationen, den lexikala tätheten och användningen av adjektiv hos andraspråksinlärare har undersökts.

1.1 Syfte

Syftet med föreliggande studie är att undersöka lexikal komplexitet i uppsatser skrivna av finska gymnasister. Jag tar reda på om det finns ett samband mellan lexikal variation samt lexikal täthet och de olika CEFR-nivåerna. Därtill utreder jag om det finns skillnader i användningen av adjektiv på de olika färdighetsnivåerna. Mina forskningsfrågor är följande:

- Hur varierande är ordförrådet på de olika färdighetsnivåerna?
- Hur tät är ordförrådet på de olika färdighetsnivåerna?
- Vilka adjektiv använder gymnasisterna på de olika färdighetsnivåerna?
- Hur använder gymnasisterna adjektiv på de olika färdighetsnivåerna?

De två första forskningsfrågorna anknyter till den lexikala variationen och den lexikala tätheten som jag nämnde i början av detta kapitel. Jag valde att närmare undersöka adjektiv eftersom språkinlärare på högre färdighetsnivåer enligt Bergman och Abrahamsson (2004: 608–611) använder mera adjektiv än inlärare på lägre färdighetsnivåer. Användningen av adjektiv kan således säga relativt mycket om andraspråksinlärares färdighetsnivå. Jag analyserar olika adjektivens frekvens i gymnasisternas texter, hur vanliga de mest frekventa adjektiven är och hur gymnasisterna använder adjektiv i sina texter. I den djupare analysen kommer jag att ta reda på om gymnasisterna kan böja och stava adjektiven rätt och i vilka grammatiska funktioner dessa adjektiv förekommer. Därtill redogör jag för i vilka kontexter gymnasisterna använder adjektiv och om dessa adjektiv är lämpliga med tanke på kontexten.

1.2 Material och metod

Mitt material består av 15 uppsatser som gymnasisterna har skrivit som en del av ett större projekt *Helsekvenser i tre morfologiskt olika inlärarspråk* vid Åbo universitet som finansieras av Svenska kulturfonden. Uppgiften var att skriva en argumenterande text på 150–250 ord om egna åsikter om de sociala mediernas sätt att samla information. Uppsatserna har skrivits med tangentloggningssystemet ScriptLog men jag använder endast de färdigskrivna texterna i analysen. Uppsatserna har bedömts enligt en s.k. DIALUKI-skala som baserar sig på *Gemensam europeisk referensram för språk* (CEFR). Texterna placerar sig på färdighetsnivåerna A2 (2 texter), B1 (7 texter), B2 (5 texter) och C1 (1 text). Eftersom det finns bara två texter på nivå A1 och en text på nivå C1 delar jag uppsatserna i två grupper: grupp 1 som består av uppsatserna på nivåerna A2 och B1 (9 texter) och grupp 2 som består av uppsatserna på nivåerna B2 och C1 (6 texter). Jag analyserar resultaten enligt dessa två grupper.

Metoderna i denna studie är kvantitativa och kvalitativa. I den kvantitativa analysen räknar jag alla löpord och lemman i varje uppsats och klassificerar dessa ord enligt ordklasser. Enligt Read (2000: 203) kan den lexikala variationen mätas så att antalet lemman divideras med antalet löpord (*type/token ratio*) men jag använder ett s.k. ordvariationsindex (OVIX) eftersom det inte är beroende av textens längd. För att mäta den lexikala tätheten använder jag ett s.k. SAVKVOT-värde där antalet innehållsord (substantiv, huvudverb och adjektiv) divideras med löpord (se närmare kapitel 5 för OVIX och SAVKVOT). Jag exciperar också alla adjektiv från texterna och räknar dem. Jag jämför alla värden mellan de två ovannämnda grupperna. Den kvantitativa analysen presenteras i form av tabeller. I den kvalitativa analysen analyserar jag mer ingående användningen av adjektiv med hjälp av Nations (2001) modell av djup ordkunskap (se också avsnitt 1.1 och avsnitt 3.1.2). Jag presenterar närmare mitt material och mina metoder i kapitel 6.

1.3 Avhandlingens disposition

Härnäst kommer jag att presentera de teoretiska utgångspunkterna för min undersökning. I kapitel 2 behandlas allmänt vad det avses med begreppen ord och ordförråd. Därutöver presenteras i detta kapitel ordklassen adjektiv som är relevant i denna undersökning. I kapitel 3 kommer jag att betrakta ordinläring och presentera de mest centrala begrepp: *lexikal kompetens* och *lexikal komplexitet*. I slutet av kapitel 3 behandlas också ordinlärningsprocess. I kapitel 4 presenteras *Gemensam europeisk referensram för språk* (CEFR), dess syfte och DIALUKI-skalan som baserar sig på CEFR. I kapitel 5 behandlas tidigare forskning kring den lexikala komplexiteten hos andraspråksinlärare. Efter att ha presenterat teoridelen i denna avhandling behandlas närmare materialet och metoden i kapitel 6. Därefter, i kapitel 7, följer analysdelen där jag presenterar mina resultat. Avhandlingen avslutas med kapitel 8 där jag sammanfattar de mest centrala resultaten och diskuterar didaktiska implikationer samt möjliga framtida studier om temat.

2 Ord och ordförråd

I detta kapitel kommer jag att presentera den centrala teori som anknyter till ord och ordförråd allmänt. Först, i avsnitt 2.1, redogör jag för vilka slags definitioner av begreppet ord det finns i litteraturen. Jag presenterar också de ordklasser som kan urskiljas i svenska språket. I avsnitt 2.2 betraktas den ordklass som är relevant i min studie, nämligen adjektiv. Till slut, i avsnitt 2.3, fokuserar jag på definitionen av ordförråd.

2.1 Allmänt om ord

Enligt Svenska Akademiens språklära (SAS) (2013: 28) har det varit svårt att definiera begreppet ord. SAS presenterar ändå ett alternativ för definitionen: ”Ord är den minsta språkbit som ensam kan utgöra ett yttrande” (SAS 2013: 29). Enligt denna definition kan ord således stå ensam till exempel i svar på en fråga. Ett kännetecken på ord är också att olika ord i skrift kan skiljas med ett tomrum och i tal med pauser. (SAS 2013: 28–29.)

Det finns definitioner av ord som baserar sig på hur man räknar ord exempelvis i en text. Ord kan klassificeras som *löpord*, *graford* eller *lexikonord (lemma)*. Begreppet löpord syftar på alla ord som förekommer i en text. Med begreppet graford avses däremot olika ord i texten. (SAS 2013: 30.) Lemma innehåller ordets grundform samt dess böjningar och dessa ord används som uppslagsord i ordböcker (Niitemaa 2015: 140). Exempelvis innehåller följande två meningar 13 löpord, 11 graford och 10 lemman:

Han har ett hus. Huset är svart men han vill måla huset rött.

Meningarna innehåller 13 löpord eftersom det finns sammanlagt 13 ord i dessa två meningar. Orden *han* och *huset* förekommer två gånger i meningarna och därför räknar man dessa endast en gång när det är fråga om graford. Vad gäller lemman är *hus* och *huset* böjningsformer av samma ord vilket betyder att man räknar dessa som ett lemma. Enligt Niitemaa (2015) kan orden ytterligare indelas i *ordfamiljer*. Ordfamiljen innehåller ordets grundform, dess böjningar men också dess avledningar. Det råder ändå oenighet om definitionen av begreppet ordfamilj och därför är begreppet vagt. (Niitemaa 2015:

140–141.) I min undersökning räknar jag både löpord och lemman i informanternas texter. I avsnitt 6.2 kommer jag att presentera mer ingående hur jag har räknat och klassificerat dessa ord.

Enligt SAS (2013) indelas ord i ordklasser enligt den grammatiska funktionen som ordet har i fraser, hur ordet böjs och vilken slags betydelse ordet har. Dessa ordklasser är substantiv, verb, adjektiv, adverb, pronomen, prepositioner, konjunktioner, subjunktioner, interjektioner och räkneord. Det finns dock ord som delar samma egenskaper med mer än en ordklass och kan således höra till flera ordklasser. Därför är en heltäckande ordklassindelning av alla ord nästan omöjlig att göra. Ett exempel på sådana ord är particip som delar likadana egenskaper med åtminstone verb och adjektiv (t.ex. *den skrattande flickan, en tappad plånbok*). (SAS 2013: 36.)

Ordklasserna kan vidare klassificeras på basis av två kriterier. För det första kan en skillnad bland ordklasserna göras mellan *öppna* och *slutna* ordklasser. Öppna ordklasser får hela tiden nya ord genom lån, avledning och sammansättning medan slutna ordklasser sällan får nya ord. Öppna ordklasser är substantiv, verb och adjektiv medan slutna ordklasser är en del av adverb, prepositioner, pronomen, konjunktioner och subjunktioner. För det andra kan en skillnad göras mellan *oböjliga* och *böjliga* ord. Substantiv, adjektiv, verb, pronomen och räkneord hör till böjliga ord och har således egna böjningsformer. Prepositioner, konjunktioner, subjunktioner, interjektioner och adverb hör till oböjliga ord och kan huvudsakligen inte böjas. (SAS 2013: 36–38.)

Ett annat sätt att definiera ord är att skilja mellan *innehållsord* och *formord*. Innehållsord syftar på de ord som bär betydelse. Sådana ord är substantiv, verb och adjektiv. (Niitemaa 2015: 140.) Med formord avses däremot ord som har en mer grammatisk betydelse. Formord innehåller t.ex. artiklar, pronomen, konjunktioner och prepositioner. Vanligtvis har dessa ord ingen stor betydelse i sig själva utan de skapar bl.a. förbindelse mellan satser eller formar innehållsordens betydelse. (Read 2000: 17.) Formord är mer frekventa i språket än innehållsord (Niitemaa 2015: 140).

2.2 Adjektiv

Som redan konstaterats i avsnitt 2.1 hör adjektiv till de öppna ordklasserna. Detta betyder således att nya adjektiv bildas ständigt varför det är oklart hur många adjektiv det finns i svenska språket. Adjektiven kan beskriva stabila egenskaper (t.ex. *ett stort hus*) eller tillfälliga tillstånd (t.ex. *ett öppet fönster*). (SAS 2013: 76.) Enligt SAS (2013) kan adjektiven användas i två betydelser: Adjektiven kan *karakterisera* saker som de syftar på eller *klassificera* saker enligt en viss klass. I det tidigare exemplet *ett stort hus* karakteriserar adjektivet *stort* substantivet *hus* och i exemplet *ett öppet fönster* karakteriserar adjektivet *öppet* substantivet *fönster*. Ett exempel på klassificerande adjektiv är adjektivet *fransk* i frasen *ett franskt vin*, där adjektivet beskriver ett vin som kommer från Frankrike, dvs. klassificerar vinet enligt en klass. (SAS 2013: 77–78.) Adjektiven kan kompareras i positiv, komparativ och superlativ (t.ex. *stor, större, störst*) och de kongruensböjs i genus (t.ex. *ett stort hus, en stor hund*), bestämdhet (t.ex. *ett stort hus, det stora huset*), sexus (t.ex. *den gamle mannen, den gamla kvinnan*) och numerus (t.ex. *en stor hund, tre stora hundar*) (SAS 2013: 82).

Adjektiv har tre huvudfunktioner. Adjektiven kan fungera som *attribut* till substantiv och då kongruensböjs adjektivet med huvudordet i obestämda nominalfraser (t.ex. *ett stort hus, en stor hund*). I bestämda nominalfraser har adjektivet däremot ingen kongruensböjning i genus eller numerus utan då har adjektivet oftast ändelsen *-a* (t.ex. *det stora huset, den stora hunden*). Adjektiven kan också fungera som *predikativ* och då kongruensböjs adjektivet också med huvudordet i nominalfrasen (t.ex. i satsen *mitt hus är stort*). Adjektivet kan också vara *adverbial* med ändelsen *-t* och fungerar då som en bestämning till verb (t.ex. i satsen *hon springer snabbt*). Utöver dessa tre huvud användningar kan adjektiven fungera som *huvudord i en nominalfras* som inte har något substantiviskt huvudord (t.ex. *de unga*). (SAS 2013: 76–77.)

Som framgick i avsnitt 2.1 kan ord dela egenskaper med mer än en ordklass och därför kan det vara svårt att definiera vilken ordklass dessa ord tillhör. Ett exempel på sådana ord är *egen* och *själv* som ibland har definierats som adjektiv eftersom de böjs som adjektiv. I SAS (2013: 132) definieras ändå dessa ord som adjektiviska pronomen. Ibland klassificeras också adjektiv med ändelsen *-t* som adverb men enligt SAS (2013: 77) hör

sådana ord till ordklassen adjektiv. Particip delar också likadana egenskaper med både verb och adjektiv men SAS (2013: 79) definierar particip som verbavledda adjektiv. I denna avhandling följer jag samma ordklassdelning som SAS. Jag kommer att presentera min ordklassdelning närmare i avsnitt 6.2.

2.3 Ordförråd

Puro (1999) konstaterar att ordförråd är den struktur som håller språket samman. Denna struktur är ändå inget isolerat förråd var språkanvändare hämtar ord utan ordets processning kräver också andra delområden av språket. Dessa områden är t.ex. morfologi, fonologi, syntax och semantik. Detta betyder att kunskap i dessa delområden finns i ordförrådet och de skapar och formar därigenom också ordförrådet. (Puro 1999: 2.)

När det talas om ordförråd är ett s.k. *mentalt lexikon* ett viktigt begrepp. Enligt Niitemaa (2015) betyder det mentala lexikonet ett individuellt ordförråd som bildar ett nätverk. Detta nätverk förändrar sig ständigt. (Niitemaa 2015: 145.) Det är vanligt att språkanvändaren snabbt kan hitta rätta ord vilket indikerar att det mentala lexikonet är ett välorganiserat system (Enström 2016: 49). Enligt Agustín Llach (2011: 8) är orden i ordförrådet associativt lagrade. Enström (2016: 50) påpekar också att orden i det mentala lexikonet är relaterade till andra liknande ord. Enligt Enström (2016: 56–65) är de viktigaste relationerna synonymi (t.ex. *liten – småväxt*), överordnade och underordnade begrepp, dvs. hyponymi (t.ex. *djur – husdjur – hund*), motsatser, dvs. antonymi (t.ex. *varm – kall*) och helhet- eller delrelationer, dvs. meronymi (t.ex. *kropp – arm – hand – finger*).

Niitemaa (2015) konstaterar att speciellt de ord som ofta används hör till många olika knutar eller s.k. *noder* i det mentala lexikonet. Exempelvis kan begreppet häst ligga i noderna *djur*, *hobbyer* eller *skrämmande saker*. (Niitemaa 2015: 145.) Som ovan redan framkommit ligger sådana ord som har liknande betydelser ofta nära varandra. Enligt Niitemaa (2015) gäller detta också sådana ord som har ett likartat uttal. Därtill sammanhänger orden i förstaspråk och andraspråk med varandra i nätverket. Detta

betyder att om man hämtar ett ord från det mentala lexikonet aktiveras alla ordets länkar också i de andra språken som ligger i minnet. (Niitemaa 2015: 145.)

3 Ordinläring

Forskare har försökt skapa en omfattande bild av vad det betyder att kunna ett ord. När man talar om lexikal kunskap betyder det inte bara att kunna länka ordets talade eller skrivna form med dess betydelse. Undersökningar inom ordförrådsforskning har ofta fokuserat på ordförrådets bredd, ordförrådets djup eller hur passivt (receptivt) och aktivt (produktivt) ordförråd skiljer sig från varandra (Agustín Llach 2011: 3). Read (2000: 3–4) påpekar vidare att det inte räcker att kunna många ord, ordens olika former eller betydelser utan språkinlärare behöver också kunna använda dessa ord i olika kommunikativa situationer. Många forskare talar om *lexikal kompetens* (se t.ex. Meara 1996; Henriksen 1999) som enligt CEFR (2009: 107) betyder ”kunskap om och förmåga att använda ordförrådet i ett språk”.

I det följande presenterar jag de olika aspekterna i ordinläring. I avsnitt 3.1 redogör jag närmare för lexikal kompetens. Lexikal kompetens omfattar såväl receptiv som produktiv ordkunskap men eftersom jag i min avhandling undersöker texterna som gymnasisterna själva har skrivit kommer jag i synnerhet att behandla lexikal kompetens från produktivt perspektiv. Därtill kommer jag också att diskutera vad det menas med *lexikal komplexitet* som står i fokus i denna avhandling. Avslutningsvis, i avsnitt 3.3, behandlar jag ordinlärningsprocessen, dvs. hur ord lärs in i praktiken.

3.1 Lexikal kompetens

Begreppet lexikal kompetens har definierats på många olika sätt. Henriksen (1999) konstaterar att forskare ofta har använt samma termer för olika begrepp som ingår i den lexikala kompetensen. Därför borde man precisera de termer och komponenter som bildar den lexikala kompetensen. (Henriksen 1999: 303.) När man i litteraturen talar om den lexikala kompetensen kommer det dock ofta fram två komponenter: ordförrådets *bredd* och *djup*. Som redan ovan framgått, har dessa två komponenter ofta stått i fokus inom ordförrådsforskning. Härnäst kommer jag att presentera vad det menas med ordförrådets bredd och djup.

3.1.1 Ordförrådets bredd

Med ordförrådets bredd avses hur många ord en individ kan, dvs. hur stort ordförråd en individ har. Meara (1996: 37) konstaterar att språkinlärare med ett mer omfattande ordförråd även har bättre kunskaper i andra delområdena av språket än språkinlärare med ett mer begränsat ordförråd. Enligt Enström (2013) finns det mycket forskning om ordförrådets bredd hos modersmålstalare men resultaten har varierat i hög grad. Av den anledningen har det varit svårt att definiera hur brett ordförråd individen vanligtvis har. (Enström 2013: 170.)

Enligt Flyman Mattsson och Håkansson (2010) kan en individ ca 8000–10000 ord i sitt modersmål före skolstarten och antalet ord ökar med 3000 ord för varje skolår. Ordförrådet indelas ofta i *basordförråd* och i *utbyggt ordförråd*. Det ovannämnda antalet ord som individen kan före skolstarten sägs vara basordförråd. Basordförrådet består av ord som har en allmängiltig betydelse och som ofta är s.k. överordnade begrepp, t.ex. *skor* i stället för *sandaler*. Ord i basordförrådet används också för att ersätta specifika ord, t.ex. *gå* i stället för *promenera*. De 3000 ord som individen lär sig varje skolår hör till utbyggt ordförråd. Detta ordförråd består av synonymer och fackord av olika slag. (Flyman Mattsson & Håkansson 2010: 35, 37.)

En intressant fråga är dock hur många ord en språkinlärare måste kunna. Schmitt (2010) poängterar att det beror på hurdana inlärningsmål inläraren har. För att nå språkkunskaper som liknar infödda talares språkkunskaper måste inlärares ordförråd vara lika stort som ordförrådet hos infödda talare. För många språkinlärare räcker dock ett ordförråd med vilket de klarar sig i olika kommunikativa situationer. (Schmitt 2010: 6–7.) Enligt Viberg (1988: 219) är det i början av språkinläring viktigt att lära sig basordförrådet som man måste kunna i alla situationer. Viberg (1988) konstaterar att 85 % av ord i en text vanligtvis hör till de 1000 mest frekventa orden. Man skulle kunna dra den slutsatsen att ett rätt litet ordförråd krävs för att förstå text. Detta stämmer dock inte för att de mindre vanliga orden ger mest betydelse till innehållet. (Viberg 1988: 216.) Vidare påpekar Viberg (1988: 219) att när språkinläring har varit på gång en längre tid måste inläraren lära sig ordförråd som passar i olika stilar (t.ex. vetenskaplig, talspråklig och teknisk stil), dvs. utbyggt ordförråd.

3.1.2 Ordförrådets djup

Ordförrådets djup innebär hur mycket information en individ har om de enskilda orden. Enligt Enström (2013) finns det skillnader vilka kunskaper om orden som krävs för att förstå ord (receptiv kunskap) eller för att använda orden själv (produktiv kunskap). Hon konstaterar vidare att det är svårare att använda ord i egen produktion än det är att förstå orden i olika sammanhang. I produktiva situationer måste inläraren behärska hur ordet stavas eller uttalas, vilka grammatiska former ordet har, vilka syntaktiska konstruktioner ordet ingår i (t.ex. vilka möjligheter det finns att kombinera ordet) samt hurdana betydelser och stilistiska aspekter ordet har. I receptiva situationer behöver inläraren inte nödvändigtvis kunna ett ord fullkomligt utan hen behöver kunna förstå ungefärligt vad ordet betyder. Inläraren kan också dra nytta av olika gissningsstrategier för att ta reda på ordets betydelse. (Enström 2013: 175.)

Nation (2001) har gjort en modell som definierar djup ordkunskap. Som framgått i början av detta kapitel fokuserar jag på produktiv ordkunskap varför jag endast presenterar Nations modell för produktiv ordkunskap (Tabell 1).

Tabell 1 Modell av produktiv djup ordkunskap enligt Nation (2001: 27) (min egen översättning till svenska)

Form	<i>Uttal</i>	Hur uttalas ordet?
	<i>Skriftlig form</i>	Hur skrivs ordet?
	<i>Orddelar</i>	Vilka orddelar behövs för att uttrycka betydelsen?
Betydelse	<i>Form och betydelse</i>	Vilken ordform kan användas för att uttrycka betydelsen?
	<i>Begrepp och referenser</i>	Vilka enheter kan begreppet syfta på?
	<i>Associationer</i>	Vilka andra ord kan användas i stället för detta ord?

Användning	<i>Grammatiska funktioner</i>	I vilka mönster behöver detta ord användas?
	<i>Kollokationer</i>	Vilka ord eller ordtyper behöver användas med detta ord?
	<i>Begränsning av användning (register, frekvens...)</i>	Var, när och hur ofta kan detta ord användas?

Tabellen visar att språkanvändaren måste ha en bred förståelse av ordets form, betydelse och hur detta ord används i praktiken för att kunna ett ord fullkomligt. González-Fernández och Schmitt (2017: 286) poängterar dock att Nations modell beskriver den maximala kunskapen och inte ens infödda talare kan alla aspekter av alla ord. Henriksen (1999: 311) konstaterar att ordkunskap utvecklas i stadier: från en bristande ordkunskap till en precis och djup ordkunskap. I följande avsnitt kommer jag att betrakta begreppet lexikal komplexitet som också anknyter till ordförrådets bredd och djup.

3.2 Lexikal komplexitet

För att ta reda på vad det menas med begreppet lexikal komplexitet kommer jag först att presentera kort hur begreppet *komplexitet* har definierats i litteraturen. Forskare inom andraspråksinlärning har skapat många olika modeller för att beskriva språkinlärares kunskaper i L2. Housen, Kuiken och Vedder (2012) konstaterar att kunskaper i L2 kan beskrivas med tre viktiga komponenter: *komplexitet*, *korrekthet* och *flyt* (eng. *complexity*, *accuracy* och *fluency*, dvs. CAF). Dessa tre komponenter har använts mycket i andraspråksforskning för att få mera information om språkinlärares egen produktion, kunskaper och utveckling i L2. (Housen, Kuiken & Vedder 2012: 1.)

Enligt Bulté och Housen (2012) betyder komplexitet i allmänhet hur svårt eller lätt det är att lära sig eller processa vissa egenskaper av ett språk. Några egenskaper kan vara svårare eller lättare att lära sig. (Bulté & Housen 2012: 23.) Exempelvis konstaterar Buss (2002) att språkinlärare först lär sig att använda innehållsord för att bilda enkla meningar. När

språkinlärningsprocessen fortskrider blir syntaxen mer komplex vilket betyder att inläraren måste lära sig att använda formord. (Buss 2002: 122.)

Enligt Bulté och Housen (2012: 26) hör lexikal komplexitet till s.k. *lingvistisk komplexitet*. Bulté och Housen (2012) konstaterar vidare att lingvistisk komplexitet kan ses från två olika synpunkter: den kan beskriva inlärares L2 system som helhet eller den kan beskriva enstaka lingvistiska strukturer i inlärares L2 system. Det förra kallas *systemkomplexitet* (eng. *system complexity*) och det andra *strukturkomplexitet* (eng. *structure complexity*). Systemkomplexitet beskriver storleken, bredden, eller variationen av inlärares L2 system, dvs. de olika strukturer och ord som inläraren känner till eller kan använda. Strukturkomplexitet beskriver mer den djupa kunskapen i L2, dvs. hur komplexa enstaka strukturer, ord eller regler är. Med strukturkomplexitet avses t.ex. hur många olika betydelser ett ord bär eller hur många komponenter en lingvistisk form innehåller (t.ex. en nominalfras jämfört med en adjektivfras i svenska språket). (Bulté & Housen 2012: 25.)

När det talas om lexikal komplexitet använder Bulté och Housen (2012) begreppen *systemisk lexikal komplexitet* (eng. *systemic lexical complexity*) för att beskriva systemkomplexitet och *strukturell lexikal komplexitet* (eng. *structural lexical complexity*) för att beskriva strukturkomplexitet. Systemisk lexikal komplexitet kan ytterligare beskrivas med begreppet *lexikal variation* (Bulté & Housen 2012: 28). Med lexikal variation avses hur stor variation en språkinlärare har i sitt ordförråd, dvs. hur många olika ord inläraren kan använda i sin egen produktion (Haikala 2012: 63). Detta har operationaliserats på många olika sätt, men enligt Read (2000: 203) kan variationen mätas med s.k. *type/token ratio*, där alla lemmorna divideras med alla löpord t.ex. i en text.

Systemisk lexikal komplexitet kan också beskrivas med begreppet *lexikal täthet* (Bulté & Housen 2012: 28). Med lexikal täthet avses hur mycket information det finns t.ex. i en text (DeClercq 2015: 72). Detta kan mätas så att antalet innehållsord divideras med antalet löpord eller med antalet formord (Bulté & Housen 2012: 31). Med lexikal täthet kan det således studeras hur många innehållsord en språkinlärare har använt i sin produktion. (se avsnitt 2.1 för olika definitioner av ord). Laurén (2003: 143) påpekar dock att den lexikala tätheten också kan vara ett inlärarspråkligt drag vilket betyder att andraspråksinlärare kan

uppnå hög lexikal täthet om hen upprepar samma innehållsord i sin text. Enligt Haikala (2012: 64) är det därför viktigt att mäta både den lexikala variationen och den lexikala tätheten.

Strukturell lexikal komplexitet kan enligt Bulté och Housen (2012: 28) beskrivas med begreppet *lexikal sofistikerings*. Lexikal sofistikerings syftar på hur mycket en språkinlärare använder mindre frekventa ord (Read 2000: 200). Detta betyder att de ord som språkinläraren har använt jämförs med hur ofta infödda talare använder dessa ord. Detta är möjligt med hjälp av olika frekvenslistor av de vanligaste orden i ett visst språk. (Flyman Mattsson & Håkansson 2010: 38.) Vidare kan lexikal sofistikerings mätas genom att dividera antalet mindre frekventa ord med antalet löpord (Bulté & Housen 2012: 31).

I min avhandling är både systemisk lexikal komplexitet och strukturell lexikal komplexitet viktiga. Jag undersöker lexikal variation och lexikal täthet i språkinlärares texter samt hur många adjektiv dessa inlärare använder. Detta beskriver systemisk lexikal komplexitet. Jag betraktar också hur pass vanliga adjektiv språkinlärare använder, vilket beskriver den strukturella lexikala komplexiteten. Även om det traditionellt har använts kvantitativa metoder för att mäta lexikal komplexitet, kommer jag i denna avhandling att använda också kvalitativa metoder (se närmare avsnitt 6.2). Jag redogör för hur informanterna använder adjektiv i sina texter, dvs. hur adjektiv stavas och böjs, i vilka ordkombinationer adjektiv används och i vilka kontexter adjektiv förekommer, vilket ytterligare hör till strukturell lexikal komplexitet.

3.3 Ordinlärningsprocess

Enligt Agustín Llach (2011) finns det olika synpunkter på hur ord lärs in. Ordinlärning kan ses som en process med olika utvecklingsstadier. Man har bl.a. undersökt i vilken ordning olika aspekter av ordkunskap (t.ex. morfologi, semantik och syntax) lärs in, i vilken ordning olika ordklasser lärs in och i vilken ordning vissa lexikala enheter lärs in. (Agustín Llach 2011: 4–5.) Exempelvis konstaterar Flyman Mattson och Håkansson (2010) att det är ordklassen substantiv som dominerar när barn börjar lära sig sitt

modersmål. Barn lär sig innehållsrika substantiv så att de kan kommunicera något med få ord (Flyman Mattson & Håkansson 2010: 33.)

Enligt Bergman och Abrahamsson (2004) gäller det ovannämnda också andraspråksinlärare: på nybörjarnivån används mest konkreta substantiv och vanliga verb. Däremot använder andraspråksinlärare på nybörjarnivån adjektiv sparsamt. Hos andraspråksinlärare på mellannivå är det fortfarande substantiv och verb som dominerar. Användningen av adjektiv ökar men den är ännu relativt sparsamt. På avancerad nivå börjar inläraren använda fler adjektiv. (Bergman & Abrahamsson 2004: 608–611.) Användningen av adjektiv kan således säga relativt mycket om andraspråksinlärares färdighetsnivå och därför analyserar jag närmare adjektiv i denna undersökning.

En annan synpunkt på hur ord lärs in anknyter till det mentala lexikonet som presenterades i avsnitt 2.3. Ur denna synpunkt sker ordinläring med hjälp av associationer som formar olika nätverk. Detta betyder att man bildar relationer mellan de ord som redan finns i ordförrådet och de nya ord som lärs in. (Agustín Llach 2011: 8.) Enström (2016) påpekar att det därför är viktigt för språkinläraren att veta hur orden i det mentala lexikonet är organiserade. Detta kan hjälpa språkinläraren i ordinlärningsprocessen. (Enström 2016: 50.)

Språkinlärare kan lära sig nya ord både medvetet och omedvetet (Enström 2016: 25). Enligt Viberg (2004) kallas det förra *intentionell inläring*. Språkinläraren kan t.ex. upprepa medvetet ord för att lära sig dem bättre. (Viberg 2004: 197.) Oftast lär man sig ändå nya ord omedvetet (Enström 2016: 25). Detta kallas *incidentiell inläring*. Incidentiell inläring sker som en biprodukt när man använder språket, t.ex. när man läser. (Schmitt 2010: 29.) Ordinlärningsprocessen i sådana situationer sker enligt Viberg (2004) så att inlärare isolerar en ordform och lagrar denna form samt ordets betydelse i minnet. Isoleringen av en ordform, s.k. *segmentering*, är relativt lätt då det gäller skrivet språk medan isoleringen är svårare i talat språk. Därför är det viktigt för språkinläraren att lära sig segmenteringen. (Viberg 2004: 197.)

Utöver segmenteringen är det viktigt för språkinlärare att kunna tolka betydelsen av okända ord. (Viberg 2004: 198). Detta kan göras med hjälp av kontexten och en allmän

kunskap om världen (Enström 2016: 25–26). Enligt Viberg (2004) kan språkinläraren använda olika *språkliga ledtrådar* för att gissa ordets betydelse. Inläraren kan exempelvis använda ord i sitt modersmål eller i andra inlärd språk för att tolka betydelsen av okända ord i ett annat språk (t.ex. verbet *se* i svenska språket och *see* i engelska språket). Språkinläraren kan också använda s.k. *referentiella ledtrådar*, dvs. ledtrådar i situationen. Språkinläraren kan exempelvis ta reda på vad en annan individ talar om utifrån kontexten. (Viberg 2004: 199.) Det är också viktigt att språkinläraren möter de ord som hen redan känner till i olika kontexter. På detta sätt lär inläraren sig ordets olika betydelser och i vilka situationer dessa ord kan användas. Ofta behöver inläraren möta ordet många gånger i olika kontexter för att komplettera sin kunskap om detta ord. (Enström 2016: 25–26.)

Som redan framkommit i avsnitt 3.1.2 är det lättare att förstå ord än att använda dem produktivt. Enligt Enström (2013) är det därför vanligt att det receptiva ordförrådet är större än det produktiva ordförrådet och förståelsen av ett ord sker oftast före användningen. Detta betyder dock inte att de ord som andraspråksinlärare känner till receptivt alltid kommer att senare bli en del av det produktiva ordförrådet. (Enström 2013: 175.) Som redan konstaterats i avsnitt 3.1.2 utvecklas ordkunskap i stadier och Enström (2013: 175) poängterar att andraspråksinlärare känner till olika ord i olika grad.

4 Gemensam europeisk referensram för språk

Informanternas texter i denna avhandling har bedömts enligt en s.k. DIALUKI-skala som baserar sig på *Gemensam europeisk referensram för språk* (eng. *Common European Framework of Reference for Languages*, dvs. CEFR, se bilaga 1). Därför presenterar jag referensramen och dess syfte allmänt i detta kapitel. Eftersom min avhandling fokuserar på ordförråd kommer jag att betrakta också de aspekter i DIALUKI-skalan som anknyter till ordförråd. Jag kommer att använda förkortningen CEFR i denna avhandling.

CEFR har utarbetats av Europarådet för att bilda en gemensam grund för språkinlärning i Europa. Detta innefattar exempelvis kurs- och läroplaner samt läromedel för språkinlärning, vilka skulle kunna skilja sig mycket från varandra i olika länder i Europa utan någon gemensam referensram. Det är viktigt att ha en gemensam grund för språkinlärning för att människor som arbetar med språk ska kunna kommunicera effektivt med varandra. Dessa människor, t.ex. lärare och kursutvecklare, kan använda referensramen som hjälpmedel i sitt eget arbete. Referensramen innehåller en beskrivning av hurdana kunskaper och färdigheter språkanvändaren behöver tillägna sig så att hen kan kommunicera på ett visst språk. Referensramen innehåller också färdighetsnivåer som kan användas för att följa språkinlärningsprocessen hos språkinlärare. (CEFR 2009: 1.)

Som ovan framgått innehåller referensramen olika färdighetsnivåerna. Enligt CEFR (2009) består nivåskalan av tre breda färdighetsnivåer: A, B och C. Inlärare på nivå A kallas *användare på nybörjarnivå*, inlärare på nivå B kallas *självständiga användare* och inlärare på nivå C kallas *avancerade användare*. Dessa tre färdighetsnivåer kan vidare indelas i två olika nivåer så att nivå A indelas i A1 och A2, nivå B indelas i B1 och B2 och nivå C indelas i C1 och C2. Således består nivåskalan av totalt sex nivåer (A1, A2, B1, B2, C1 och C2). (CEFR 2009: 23.)

De texter som jag använder i denna undersökning placerar sig på nivåerna A2, B1, B2 och C1. I det följande kommer jag att presentera närmare dessa nivåer och redogöra för hur ordkunskap framkommer i beskrivningarna. Givetvis innehåller DIALUKI-skalan

inga direkta instruktioner eller listor om vilka ord inläraren måste kunna använda på olika färdighetsnivåerna. Däremot nämner DIALUKI-skalan olika teman som inläraren måste kunna berätta om (se bilaga 1 för DIALUKI-skalan).

Enligt DIALUKI-skalan kan inlärare på nivå A2 använda enkla uttryck om sin egen familj, sina levnadsförhållanden, sin utbildning och sitt arbete. Dessutom kan inlärare på nivå A2 uttrycka egna erfarenheter och ta del av allmänna vardagliga diskussioner. Detta betyder att inläraren behöver kunna använda ordförråd som anknyter till dessa områden, t.ex. vanliga ord om familjemedlemmarna eller om skolvärlden. Enligt DIALUKI-skalan kan inlärare på nivå B1 berätta om vanliga saker som intresserar honom eller henne. Det är inte klart vilka teman dessa intressen ska innehålla. Detta betyder ändå att inläraren ska kunna använda sådant ordförråd som anknyter till de saker som hen är intresserad av. Hen kan uttrycka sina åsikter om konkreta och abstrakta saker, t.ex. om kultur, musik och filmer. Hen kan också beskriva sina känslor och reaktioner om händelser. Detta betyder att inläraren behöver kunna använda bl.a. adjektiv för att uttrycka sina känslor.

Nivåerna B1 och B2 ser ganska likadana ut vad gäller olika teman. Enligt DIALUKI-skalan kan inlärare på nivå B2 dock berätta om olika saker som intresserar honom eller henne. Inläraren på nivå B2 ska således ha ett relativt brett ordförråd vad gäller de saker som hen är intresserad av. Hen kan också beskriva mångsidigt sina känslor vilket indikerar att inläraren behöver kunna använda bl.a. olika adjektiv för att uttrycka sina känslor. På nivå C1 ska inläraren allmänt ha ett relativt brett ordförråd eftersom inlärare enligt DIALUKI-skalan kan på nivå C1 skriva texter om varierande teman.

5 Tidigare forskning

I detta kapitel presenterar jag tre tidigare studier som är relevanta för min undersökning. Haikala (2012) har undersökt lexikal kvalitet i abiturientuppsatser på olika färdighetsnivåer. Agebjörns (2017) undersökning behandlar lexikal komplexitet i andraspråksinlärares muntliga produktion. I slutet av detta kapitel presenterar jag Lauréns (2003) undersökning där hon har studerat språkbadelevs berättelsestruktur samt syntaktisk och lexikal kompetens i muntlig och skriftlig produktion.

Haikala (2012) har undersökt sambandet mellan lexikal kvalitet och olika färdighetsnivåer. Som material använder hon 123 texter skrivna av finska abiturienter i studentexamen i B1-svenska. Längden för uppsatsen är 150–200 ord. Färdighetsnivåerna i hennes undersökning baserade sig på Studentexamensnämndens bedömning och det högsta poängtalet var 99 poäng. Enligt poängtalet indelade Haikala texterna i tre betygsgrupper: betyget 3 (god) fick texterna med 99–80 poäng, betyget 2 (medelmåttig) texterna med 78–60 poäng och betyget 1 (svag) texterna med 58–0 poäng. Det fanns 29 texter i grupp 3, 48 texter i grupp 2 och 46 texter i grupp 1. (Haikala 2012: 61.)

Haikala (2012) analyserade lexikal kvalitet i abiturienternas texter med hjälp av lexikal variation och lexikal täthet (se avsnitt 3.2 för definition av lexikal variation och lexikal täthet). Hon använde ett s.k. OVIX-värde (ordvariationsindex) för att mäta den lexikala variationen och ett s.k. SAVKVOT-värde för att mäta den lexikala tätheten. OVIX-värdet räknades med formeln $\ln(N)/\ln(2-\ln(V)/\ln(N))$, där N beskriver löpord och V lemman. Hon använde OVIX eftersom det inte är beroende på textens längd. SAVKVOT räknades så att andelen substantiv, huvudverb och adjektiv dividerades med löpord i en text. Haikala räknade t.ex. alla graford som löpord och betraktade böjningsformer av samma ord som ett lemma. Adjektivadverb med ändelsen *-t* betraktade hon som adjektiv. (Haikala 2012: 64.)

Resultaten i Haikalas (2012) studie visar att betygsgrupp 3 i snitt har det högsta antalet löpord och lexikala ord i sina texter. Betygsgrupp 1 däremot uppvisar det lägsta antalet löpord och lexikala ord i medeltal. Detta betyder att betygsgrupp 1 skriver kortare texter

än de två andra grupperna. (Haikala 2012: 65.) Resultaten visar vidare att betygsgrupp 3 genomsnittligt har de högsta OVIX-värdena och betygsgrupp 1 de lägsta OVIX-värdena. Gymnasister med högre betyg har således bredare ordförråd än gymnasister med lägre betyg. Haikala jämför sitt resultat av OVIX-värden med tidigare undersökningar och noterar att gymnasister i hennes undersökning uppvisar ett relativt begränsat ordförråd. (Haikala 2012: 66.) Vad gäller SAVKVOT visar resultaten att skillnaderna i värdena mellan de tre betygsgrupperna var små. SAVKVOT berättar således inte mycket om ordförrådets kvalitet i gymnasisternas texter. (Haikala 2012: 68.) Enligt Haikala (2012: 66, 68) krävs det en närmare lexikal eller syntaktisk analys för att få en mer omfattande bild av de faktorer som påverkar bedömningen.

Syftet med Agebjörns (2017) studie var att undersöka relationen mellan lexikal komplexitet och bedömning av muntlig språkfärdighet. Därtill betraktade han också relationen mellan de olika mått som användes för att mäta den lexikala komplexiteten. (Agebjörn 2017: 37.) Hans material bestod av inspelningar av 20 vuxna andraspråksinlärare som gjordes när de deltog i den muntliga delen av Swedex B2. Swedex B2 är ett språktest som testar om inlärare uppnått B2 i CEFR. Testet innehåller fyra delar: läsförståelse, hörförståelse, skriftlig del och muntlig del. I den muntliga delen diskuterade testtagarna parvis i 20 minuter om ett givet tema och detta bedömdes av två bedömare. Det fanns fyra bedömningskriterier: flyt och interaktion, kommunikativ design, grammatisk korrekthet och komplexitet samt uttal och intonation. Enligt Agebjörn var kommunikativ design det viktigaste kriteriet med hänsyn till hans undersökning eftersom ordförråd ingick i detta. Tre betyg användes för varje kriterium: A och B (godkända) och C (underkänd). (Agebjörn 2017: 39.)

Enligt Agebjörn (2017) fanns det 6 informanter som fick betyget C vad gäller kriteriet kommunikativ design, 7 informanter som fick betyget B och 7 informanter som fick betyget A. Vad gäller hela Swedex B2 (där man tog alla fyra delar av språkfärdigheten i beaktande) fanns det 11 informanter som blev godkända och 9 informanter som blev underkända. Han använde också sammanvägningspoängen för de fyra kriterierna (A=2, B=1 och C=0) så att t.ex. den informant som fick fyra C fick 0 poäng och den informant som fick A enligt dessa fyra kriterier fick 8 poäng. Han använde betyget i kriteriet kommunikativ design, slutbetyget i hela Swedex B2 och sammanvägningspoängen i sin

analys då han testade korrelationer mellan den lexikala komplexiteten och bedömningen. (Agebjörn 2017: 40.)

Agebjörn (2017) använde sammanlagt 11 mått för att mäta den lexikala komplexiteten: andelen innehållsord av alla ord (lexikal täthet), antalet graford, antalet ordtyper, OVIX (lexikal variation), antalet innehållsord, antalet ordtyper som är innehållsord, innehållsordvariation, högfrekventa ord (andelen ord som ingår i de 5000 mest frekventa ord i korpus Bloggmix i språkbanken), högfrekventa innehållsord, ordlängd och prepositionsfel. Dessa använde han för att mäta informationstätheten och ordförrådets bredd samt djup. (Agebjörn 2017: 40–41.) Därtill analyserade Agebjörn närmare två informanter: En vars produktion var lexikalt mindre komplex än informanternas produktion genomsnittligt men fick det högsta betyget i de fyra kriterierna och blev godkänd på hela Swedex B2 och en vars produktion var lexikalt mer komplex än informanternas produktion genomsnittligt men blev underkänd både vad gäller kriteriet kommunikativ design och hela Swedex B2. Informanternas betyg var således oväntat positivt respektive negativt. I denna närstudie analyserade han bl.a. lågfrekventa ord och begriplighet i informanternas produktion. (Agebjörn 2017: 45–46.)

Av resultaten i Agebjörns (2017) undersökning framgick att endast de mått som mäter ordförrådets bredd korrelerade signifikant med bedömningen. Måttet högfrekventa ord korrelerade med bedömning men så att de informanter som använde fler högfrekventa ord bedömdes positivt. Därtill visade resultaten i den kvalitativa delen av studien att produktionen var begripligare hos den informant som hade bedömts positivt än hos den informant som hade bedömts negativt. Detta innebär, enligt Agebjörn, att språkinlärare behöver kunna använda ord i sitt basordförråd flytande och inte bara skapa ett brett ordförråd. Resultaten i både den kvantitativa och den kvalitativa delen av studien visade att den grammatiska kompetensen hade betydelse i bedömningen även vad gäller kriteriet kommunikativ design där ordförrådet togs i beaktande. (Agebjörn 2017: 47–48.)

Laurén (2003) har undersökt hur finskspråkiga språkbads elever producerar skriftliga och muntliga berättelser på svenska. Hon fokuserade på hur dessa elever skapar sammanhängande berättelsestrukturer. Därtill analyserade hon den syntaktiska och lexikala nivån i dessa berättelser. (Laurén 2003: 11.) Eftersom min undersökning

behandlar ordförråd presenterar jag endast den delen av Lauréns studie som är relevant med hänsyn till min undersökning, dvs. lexikala drag i elevernas skriftliga produktion. Det skriftliga materialet i Lauréns (2003) undersökning insamlades longitudinellt, i årskurserna 8 och 9. Materialet bestod av två språkbadsgrupper, två grupper som hade svenska som förstaspråk och två finskspråkiga grupper som hade lärt sig A-svenska i traditionell undervisning. (Laurén 2003: 35, 39.) Hon studerade språkbadselevernars utveckling och jämförde deras utveckling med de elever som hade svenska som förstaspråk och med de finskspråkiga elever som inte hade fått språkbadsundervisning. (Laurén 2003: 34.)

Laurén (2003) använde både lexikal täthet och lexikal variation när hon analyserade den lexikala kompetensen i de skrivna berättelserna. Som innehållsord räknades substantiv, huvudverb, adjektiv och adjektivadverb med ändelsen *-t* (Laurén 2003: 90). Detta gjorde också Haikala (2012) i sin undersökning. Laurén (2003) mätte den lexikala variationen med OVIX och räknade olika böjningsformer av ett ord som olika ord. Exempelvis räknades verben *vara* och *var* som två olika ord. (Laurén 2003: 92.) Hon använde således samma mått för den lexikala variationen och för den lexikala tätheten som Haikala (2012) och Agebjörn (2017) använde i sina undersökningar. I motsats till Laurén räknade Haikala dock olika böjningsformer av ett ord som samma ord (se ovan Haikalas undersökning).

Resultaten i Lauréns (2003) studie visar att de svenskspråkiga grupperna i de två årskurserna skrev längre texter än språkbadsgrupperna. Därtill skrev språkbadseleverna längre texter i årskurs 8 än de finskspråkiga i traditionell undervisning. Däremot fanns det inga stora skillnader i årskurs 9 mellan språkbadseleverna och de finskspråkiga i traditionell undervisning vad gäller längden i texterna. (Laurén 2003: 140.)

Den lexikala tätheten var högre hos språkbadseleverna och hos den finska jämförelsegruppen än hos de svenska eleverna. Laurén (2003) konstaterar därför att den höga lexikala tätheten kan vara ett drag som förekommer hos språkinlärare. Finskspråkiga i åk 8 i traditionell undervisning uppvisade den högsta lexikala tätheten och det fanns en stor skillnad i den mellan dem och språkbadseleverna. Däremot fanns det inga tydliga skillnader i den lexikala tätheten mellan språkbadseleverna och den finskspråkiga jämförelsegruppen i årskurs 9. Vad gäller den lexikala variationen visar resultaten att den

lexikala variationen var signifikant mindre hos den finskspråkiga jämförelsegruppen än språkbadseleverna i årskurs 8. I årskurs 9 fanns det inga stora skillnader i den lexikala variationen mellan de finskspråkiga i traditionell undervisning och språkbadseleverna. De svenskspråkiga eleverna hade den största lexikala variationen i båda årskurserna. (Laurén 2003: 143.)

Haikalas (2012) undersökning är den relevantaste av dessa tre studier med tanke på min undersökning eftersom hon undersökte ordförråd i skriftlig produktion hos gymnasister. Agebjörn (2017) betraktade muntlig produktion hos vuxna andraspråksinlärare och därför är mina resultat inte helt jämförbara med hans resultat. Detsamma gäller Lauréns (2003) studie eftersom hon undersökte språkbadselever. Dessa elever samt eleverna i kontrollgruppen i traditionell undervisning var också något yngre än informanterna i min undersökning. Jag kommer dock att använda samma mått som användes i dessa tre studier, nämligen lexikal täthet och lexikal variation för att mäta den lexikala komplexiteten. I avsnitt 3.2 konstaterades att den lexikala variationen ofta mäts med s.k. type/token ratio. I dessa tre studier användes ändå OVIX som inte är beroende av textens längd. Därför kommer jag också att använda OVIX i denna undersökning. I följande kapitel behandlar jag närmare mitt material och mina metoder.

6 Material och metod

I detta kapitel presenterar jag närmare materialet och metoderna för denna undersökning. I avsnitt 6.1 presenteras materialet, dvs. projektet *Helsekvenser i tre morfologiskt olika inlärarespråk* som materialet ingår i, informanterna som har skrivit texterna, uppgiften och hur dessa texter har bedömts enligt CEFR. I avsnitt 6.2 behandlas metoderna, dvs. de principer enligt vilka jag har gjort analysen.

6.1 Material

Materialet i denna undersökning består av 15 uppsatser skrivna av finska gymnasister. Materialet ingår i projektet *Helsekvenser i tre morfologiskt olika inlärarespråk* vid Åbo universitet som finansieras av Svenska kulturfonden. Syftet med projektet är att studera helsekvenser, dvs. sekvenser av två eller flera ord eller morfem som lärs in som oanalyserade enheter. Informanterna i denna undersökning är finsk- och svenskspråkiga gymnasister som skrev texterna på svenska (första- eller andraspråk), finska (första- eller andraspråk) och engelska (andraspråk). (Helsekvenser i tre morfologiskt olika inlärarespråk.) I min undersökning använder jag endast de uppsatser som finskspråkiga gymnasister har skrivit på svenska. Dessa gymnasister har studerat svenska som det andra inhemska språket.

Uppgiften för den svenska texten var att skriva en argumenterande text på 150–250 ord om egna åsikter om de sociala mediernas sätt att samla information om användarna (se bilaga 2). Sammanlagt 16 finskspråkiga gymnasister skrev texten på svenska. Trots instruktionen om längden, återfinns en text som är mycket kort (ca 50 ord). Då dess korthet kunde påverka resultaten i denna undersökning har jag valt att utelämna texten ur analysen. Gymnasisterna hade 30 minuter tid att skriva uppsatsen på svenska. Texterna har skrivits med tangentloggningsprogrammet ScriptLog som sparar t.ex. korrigeringar och pauser och som används för att analysera skrivprocessen. I denna undersökning använder jag endast de färdigskrivna texterna eftersom mitt syfte inte är att studera skrivprocessen.

Som redan framkommit i kapitel 4 har texterna bedömts enligt DIALUKI-skalan som baserar sig på CEFR (se också bilaga 1). Texterna placerar sig på färdighetsnivåerna A2 (2 texter), B1 (7 texter), B2 (5 texter) och C1 (1 text). Nivåerna B1 och B2 är således de vanligaste färdighetsnivåerna. Eftersom det endast finns två texter på nivå A2 och en text på nivå C1 har jag indelat texterna i två grupper: grupp 1 består av texterna på nivåerna A2 och B1 och grupp 2 består av texterna på nivåerna B2 och C1. Det finns sålunda sammanlagt 9 texter i grupp 1 och 6 texter i grupp 2. Jag kommer att använda dessa grupper i min analys.

6.2 Metod

Metoderna i denna undersökning är både kvantitativa och kvalitativa. De tre första forskningsfrågorna utformar den kvantitativa delen och den fjärde forskningsfrågan den kvalitativa delen i denna undersökning (se avsnitt 1.1). Analysen inleds med att jag räknar varje ord i informanternas texter med hjälp av *LIX räknare* (<https://www.lix.se>). Texten kan kopieras till nätsidan och räknaren visar t.ex. antalet löpord, olika ord samt deras frekvens, ordvariation, antalet meningar och andelen långa ord i texten. Nackdelen med LIX räknare är dock att den inte räknar ordets böjda former som samma ord. Jag har valt att använda samma princip som Haikala (2012) i sin undersökning och räknar olika böjningsformer av ett ord som ett lemma (se kapitel 5). Därför räknar jag lemmen själv i stället för de siffror som LIX räknaren ger. Efter detta gör jag en Excel-tabell för varje text där jag skriver in varje lemma (t.ex. *ett medium*) och dess böjda former (t.ex. *medier*, *mediernas*, *medierna*) samt frekvensen av dessa böjda former. Därefter räknar jag antalet alla lemmen och antalet olika ordformer i varje text. På detta sätt får jag antalet löpord och antalet lemmen i varje text (se avsnitt 2.1 för definitionen av olika ord).

För att få svar på den första forskningsfrågan räknar jag först medelvärden av löpord och lemmen enligt de två grupperna som presenterades i avsnitt 6.1. Efter detta räknar jag den lexikala variationen i varje text för att se hur varierat ordförråd informanterna använder i sina texter (se definitionen av lexikal variation i avsnitt 3.2). Detta gör jag med hjälp av OVIX som även Haikala (2012) använde i sin undersökning (se kapitel 5). Formeln är

$\ln(N)/\ln(2-\ln(V)/\ln(N))$, där N beskriver löpord och V lemman. Därefter räknar jag medelvärdet av OVIX-värden i de två grupperna och jämför dessa värden med varandra.

Den andra forskningsfrågan handlar om den lexikala tätheten i informanternas texter. Den lexikala tätheten kan mätas så att innehållsord divideras med löpord (se även avsnitt 3.2). Jag kommer att använda termen SAVKVOT för detta i enlighet med Haikala (2012) (se kapitel 5). För att mäta den lexikala tätheten klassificerar jag först alla ord i de ovannämnda Excel-tabellerna enligt fyra grupper: substantiv, huvudverb, adjektiv och andra. Jag räknar också ihop alla substantiv, huvudverb och adjektiv i varje text. Därefter räknar jag SAVKVOT i varje text, dvs. dividerar andelen substantiv, huvudverb och adjektiv med löpord i en text. Avslutningsvis räknar jag medelvärdet av SAVKVOT-värden i båda grupperna och jämför dessa värden med varandra. Härnäst presenterar jag hur jag gör ordklassindelningen i denna undersökning.

Som redan framgått i avsnitt 2.2 följer jag *Svenska Akademiens Språklära* vad gäller ordklassindelning. Detta innebär att jag klassificerar t.ex. particip (t.ex. *intresserad*) och adjektivadverb med ändelsen *-t* (t.ex. *snabbt*) som adjektiv. Jag använder också *Svenska Akademiens ordbok* (SAOB) för att klassificera enstaka ord. Till gruppen övriga klassificerar jag alla ord som inte är substantiv, huvudverb eller adjektiv. I gruppen huvudverb ingår alla andra verb förutom hjälpverb. Det finns dock ett undantag, nämligen verbet *få* som används som huvudverb i informanternas texter. Vad gäller egennamn följer jag samma princip som Haikala (2012) och utesluter dessa från analysen.

För att få svar på den tredje forskningsfrågan räknar jag först alla adjektiv i varje ovannämnd tabell varefter jag räknar medelvärdet av antalet adjektiv i de två grupperna. Jag dividerar medelvärdet av antalet adjektiv med medelvärdet av antalet löpord i de två grupperna för att redogöra för hur mycket informanterna använder adjektiv i sina texter. Jag jämför denna andel med samma andel hos infödda talare av svenska i Hultmans och Westmans (1977) studie. Efter detta räknar jag antalet olika adjektiv i varje tabell och medelvärdet av antalet olika adjektiv i de två grupperna. Jag dividerar medelvärdet av antalet olika adjektiv med medelvärdet av antalet alla adjektiv i båda grupperna för att redovisa hur mycket informanterna använder olika adjektiv i sina texter. Jag jämför denna siffra mellan de två grupperna.

Frekvensen av adjektiv studeras från två synpunkter: hur många informanter som har använt olika adjektiv och hur många gånger dessa adjektiv förekommer i texterna. Först gör jag två tabeller (för grupp 1 och grupp 2) där jag samlar alla adjektiv som förekommer hos respektive grupp, antalet informanter som har använt dessa adjektiv och hur många gånger dessa informanter har använt de olika adjektiven i sina texter. Därefter räknar jag det sammanlagda antalet informanter enligt varje adjektiv. Jag jämför de adjektiv som är mest frekventa i respektive grupp med en frekvenslista av de 50 mest frekventa adjektiven i svenska språket enligt Allén (1971) för att redogöra för hur lexikalt sofistikerade adjektiv informanterna använder i sina texter (se avsnitt 3.2 för definitionen av lexikal sofistisering). Vidare jämför jag användningen mellan de två grupperna varefter jag redovisar om det finns samma adjektiv som är frekventa i båda grupperna och räknar hur många gånger dessa adjektiv förekommer i respektive grupp. Jag jämför detta antal med antalet informanter som har använt dessa adjektiv i respektive grupp för att redogöra för hur mycket informanterna upprepar dessa adjektiv i texterna. Till sist jämförs användningen mellan de två grupperna.

Avslutningsvis behandlas den kvalitativa delen av denna undersökning där jag närmare analyserar de adjektiv som är frekventa hos båda grupperna. För att redovisa hur informanterna använder adjektiv i sina texter exciperar jag först varje mening som innehåller dessa adjektiv ur texterna. Jag analyserar adjektiven med hjälp av Nations (2001) modell av produktiv djup ordkunskap (se avsnitt 3.1.2). Detta betyder att jag analyserar stavning och böjning av adjektiven, grammatiska funktioner, ord som ofta används med dessa adjektiv och i vilka kontexter dessa adjektiv används. Därefter redogör jag för om det finns andra adjektiv som används i liknande kontexter som dessa adjektiv. Jag jämför även användningen mellan de två grupperna.

7 Resultat

I detta kapitel presenterar jag mina resultat. Kapitlet inleds med avsnitt 7.1 där jag presenterar den lexikala variationen i gymnasisternas texter. Efter detta, i avsnitt 7.2, redogör jag för den lexikala tätheten i texterna. I avsnitt 7.3 behandlar jag resultaten om antalet adjektiv i allmänhet, de mest frekventa adjektiven och hur dessa adjektiv används i texterna. Jag presenterar alla resultat med hänsyn till de två grupperna som jag redogjort för i avsnitt 6.1.

7.1 Lexikal variation i texterna

Som redan framkommit kan den lexikala variationen mätas med OVIX-värdet (se 6.2). Innan jag presenterar resultaten av lexikal variation i textmaterialet redovisar jag resultaten av värdena som behövs för att mäta OVIX-värdet: antalet löpord och antalet lemman i materialet. För att se hur antalet löpord och lemman varierar i texterna redogör jag också för de lägsta och högsta antalen löpord och lemman hos båda grupperna. De genomsnittliga värdena av antalet löpord samt antalet lemman och de lägsta och högsta antalen löpord och lemman i båda grupperna presenteras i tabell 2.

Tabell 2 Medelvärden av antalet löpord samt antalet lemman och de lägsta och högsta antalen löpord samt lemman i de två grupperna

Grupp	Löpord	Löpord (lägsta)	Löpord (högsta)	Lemman	Lemman (lägsta)	Lemman (högsta)
1 (A2-B1)	161,22	113	207	62,44	42	77
2 (B2-C1)	204,83	156	194	77,33	49	100

Tabellen visar att grupp 2 i genomsnitt använder både löpord (204,83) och lemman (77,33) i högre grad i sina texter än grupp 1. Värdena för löpord är betydligt högre i texterna hos grupp 2 än i texterna hos grupp 1. Medelvärdet för antalet löpord pekar således på att grupp 2 i medeltal har skrivit längre texter än grupp 1. Tabellen visar vidare att både det lägsta och det högsta antalet lemman är mindre i grupp 1 än i grupp 2. Det högsta antalet löpord är dock större i grupp 1 än i grupp 2 vilket visar att det även förekommer långa

texter i den lägre färdighetsgruppen. Det finns en relativt stor skillnad mellan det lägsta och det högsta antalet löpord i grupp 1, i grupp 2 är skillnaden mindre. Däremot finns det en större skillnad i antalet lemman i grupp 2 än i grupp 1.

Jag fortsätter med OVIX-värdena som anger lexikal variation i texterna. I tabell 3 presenteras medelvärdet för OVIX och de lägsta och högsta värdena av OVIX i båda grupperna.

Tabell 3 Medelvärdet av OVIX samt de lägsta och högsta värdena av OVIX i de två grupperna

Grupp	OVIX	OVIX (lägsta)	OVIX (högsta)
1 (A2-B1)	29,79	24,87	36,15
2 (B2-C1)	31,80	24,46	39,77

Av Tabell 3 framgår att grupp 2 genomsnittligt har något högre OVIX-värde (31,80) än grupp 1 (29,79). Skillnaden är dock relativt liten. Resultaten indikerar att grupp 1 upprepar mera samma ord i texterna än grupp 2. Grupp 2 har således lexikalt mer varierande texter än grupp 1. Detta kan tolkas så att informanterna i grupp 2 har bredare ordförråd än informanterna i grupp 1 och kan således variera mera mellan olika ord.

Tabellen visar vidare att skillnaden mellan det lägsta och det högsta OVIX-värdet är stor hos båda grupperna. Detta kan påverka resultaten vad gäller OVIX-medelvärdena. Grupp 2 uppvisar dock en större skillnad mellan det lägsta och det högsta OVIX-värdet. Det högsta OVIX-värdet är högre i grupp 2 än i grupp 1 men intressant nog är det lägsta värdet något lägre i grupp 2 än det lägsta värdet i grupp 1. Detta betyder att det i grupp 2 förekommer texter med både den största och den minsta lexikala variationen. Den stora skillnaden mellan det lägsta och det högsta värdet av OVIX i de två grupperna kan betyda att det finns variation i ordförrådets bredd hos informanterna i respektive grupp.

När resultaten jämförs med tidigare forskning, kan det noteras att den lexikala variationen ändå är ganska låg i båda grupperna. I Haikalas (2012) undersökning är det lägsta OVIX-värdet i medeltal 37,0 (betygsgrupp 1) och det högsta värdet 50,5 (betygsgrupp 3). Skillnaderna mellan OVIX-värden hos mina informanter och OVIX-värden hos informanterna i Haikalas studie är således ganska stora. Resultaten indikerar att båda

grupperna i denna undersökning har relativt begränsat ordförråd jämfört med informanterna i Haikalas undersökning.

7.2 Lexikal täthet i texterna

Lexikal täthet kan räknas med s.k. SAVKVOT (se 6.2) så att antalet innehållsord divideras med antalet löpord i texten. Grupp 1 har genomsnittligt 72,00 innehållsord i sina texter medan medelvärdet av innehållsord är 84,17 i grupp 2. Grupp 2 använder således ett större antal innehållsord vilket är enligt förväntningen eftersom grupp 2 också genomsnittligt har flera löpord i texterna. Medelvärdet av SAVKVOT samt de lägsta och högsta värdena av SAVKVOT hos båda grupperna presenteras i tabell 4.

Tabell 4 Medelvärdet av SAVKVOT och de lägsta och högsta värdena av SAVKVOT i de två grupperna

Grupp	SAVKVOT	SAVKVOT (lägsta)	SAVKVOT (högsta)
1 (A2-B1)	45,14	39,61	51,41
2 (B2-C1)	41,68	36,18	48,45

Tabellen visar att grupp 1 har ett högre SAVKVOT-värde (45,14) än grupp 2 (41,68). Detta betyder att även om grupp 2 använder flera innehållsord än grupp 1 är förhållandet mellan antalet innehållsord och antalet löpord mindre än hos grupp 1. Resultaten pekar på att texterna i grupp 1 är mer informationsrika och täta än texterna i grupp 2. Detta är ett överraskande resultat då texterna i grupp 1 har bedömts vara på lägre färdighetsnivåer. Tabellen visar vidare att skillnaden mellan det lägsta och det högsta värdet är relativt stor i båda grupperna. Både det lägsta och det högsta värdet är mindre hos grupp 2 än hos grupp 1. Den stora skillnaden mellan det lägsta och det högsta värdet i SAVKVOT kan innebära att det förekommer variation i hur tät ordförråd informanterna i respektive grupp har.

När resultaten jämförs med tidigare undersökning, kan det observeras att den lexikala tätheten är relativt hög i båda grupperna och högre än i Haikalas (2012) undersökning där det lägsta SAVKVOT-medelvärdet är 34,9 (betygsgrupp 2) och det högsta medelvärdet 36,0 (betygsgrupp 3). Anledning till detta kan vara att SAVKVOT-värdet inte berättar

hur mycket olika innehållsord upprepas i texterna. Som redan framgått i avsnitt 3.2 kan den lexikala tätheten bli hög när samma innehållsord upprepas i texten. Det tidigare resultatet av lexikal variation (se ovan 7.1) visade att den lexikala variationen i genomsnitt är lägre i min undersökning än i Haikalas undersökning. Detta indikerar att även om texterna i min undersökning har högre lexikal täthet än texterna i Haikalas undersökning, upprepar gymnasisterna i min undersökning mer samma ord än gymnasisterna i Haikalas undersökning.

Det ovan nämnda fenomenet kan också iaktas när jag jämför den lexikala variationen och den lexikala tätheten mellan de två grupperna. Grupp 1 har högre lexikal täthet i texterna men den lexikala variationen är lägre hos grupp 1 än hos grupp 2. Detta kan tolkas så att grupp 1 använder mer samma innehållsord och på det sättet uppnår högre lexikal täthet. För att vidare redovisa detta presenterar jag förhållandet mellan medelvärdet av olika innehållsord och medelvärdet av alla innehållsord. Resultaten presenteras i tabell 5.

Tabell 5 Medelvärden av olika innehållsord samt alla innehållsord och deras förhållande i de två grupperna

Grupp	olika innehållsord	alla innehållsord	andel olika innehållsord av alla innehållsord (%)
1 (A2-B1)	32,11	72,00	44,73
2 (B2-C1)	41,33	84,17	49,00

Tabell 5 visar att grupp 2 använder fler olika innehållsord (41,33) i texterna än grupp 1 (32,11). Resultaten indikerar att grupp 1 upprepar mer samma innehållsord än grupp 2. Resultatet kan delvis förklara varför grupp 1 har högre SAVKVOT-värde än grupp 2 (se tabell 4).

7.3 Adjektiv i texterna

Grupp 1 använder genomsnittligt 13,00 adjektiv i texterna medan det motsvarande värdet är 16,67 för grupp 2. Antalet i genomsnitt är således större hos grupp 2 än hos grupp 1.

För att få veta hur stor andel av ord är adjektiv i materialet har jag dividerat medelvärdet av adjektiv med medelvärdet av löpord i texterna. Resultaten presenteras i tabell 6.

Tabell 6 Medelvärden av antalet alla adjektiv och antalet löpord i textmaterialet samt deras förhållande i de två grupperna

Grupp	Adjektiv	Löpord	Andel adjektiv av löpord (%)
1 (A2-B1)	13,00	161,22	8,29
2 (B2-C1)	16,67	204,83	8,58

Tabellen visar att 8,29 % av ord i texterna är adjektiv i grupp 1 medan motsvarande andel är 8,58 % i grupp 2. Andelarna är relativt lika i de två grupperna. Enligt SAS (2013: 76) är andelen adjektiv av löpord oftast 5–10 % i skrift hos förstaspråksskribenter. I Hultman och Westmans (1977: 126) undersökning var andelen adjektiv av löpord 6,65 % i L1-gymnasisttexter. Andelen adjektiv är något högre i båda grupperna i min undersökning men resultaten stämmer relativt bra med de ovan nämnda andelarna.

I avsnitt 7.2 konstaterade jag att gymnasisterna upprepar samma innehållsord i sina texter och att grupp 1 använder mer samma innehållsord än grupp 2. I det följande redogör jag närmare för hur mycket samma adjektiv upprepas i båda gruppernas texter. Resultaten presenteras i tabell 7.

Tabell 7 Medelvärden av antalet olika adjektiv samt alla adjektiv och deras förhållande i de två grupperna

Grupp	olika adjektiv	alla adjektiv	andel olika adjektiv av alla adjektiv (%)
1 (A2-B1)	8,33	13,00	64,10
2 (B2-C1)	8,00	16,67	48,00

Tabell 7 visar att även om grupp 2 använder flera adjektiv (16,67) i genomsnitt i sina texter än grupp 1 (13,00) är användningen av olika adjektiv mindre i grupp 2 än i grupp 1. Förhållandet mellan olika adjektiv och alla adjektiv är således lägre i grupp 2 (48,00) än i grupp 1 (64,10). Skillnaderna mellan de två grupperna är relativt stora. Resultaten indikerar att adjektivbruket är mer varierande i grupp 1 än i grupp 2. Det är intressant att grupp 2 i genomsnitt använder mer olika innehållsord än grupp 1 (se tabell 5 ovan) men

användningen av olika adjektiv är ändå mycket mindre än i grupp 1. Detta kan innebära att informanterna i grupp 1 har bredare ordförråd än grupp 2 vad gäller adjektiv och kan därför använda och variera mera mellan dessa adjektiv. Detta är ett överraskande resultat då texterna i grupp 1 har bedömts vara på lägre färdighetsnivåer. Detta berättar dock inte hur informanterna använder adjektiv i sina texter, t.ex. om informanterna kan använda dessa adjektiv i lämpliga kontexter. Detta kommer jag att analysera närmare i avsnitt 7.3.2.

7.3.1 Mest frekventa adjektiv

Som redan framgått i avsnitt 6.2 betraktar jag frekvensen av adjektiv från två synpunkter: hur många informanter som har använt olika adjektiv och hur många gånger dessa adjektiv förekommer i texterna. Enligt den första synpunkten är de mest frekventa adjektiven, med vilka jag avser här de som minst en tredjedel av informanterna har använt, *social, bra, okej, intresserad, god, viktig* och *dålig* i grupp 1 och *social, okej, bra, olik, gammal, riktig, ny, intresserad, kul, normal* och *rolig* i grupp 2. Antalet mest frekventa adjektiv är således 7 hos grupp 1 medan samma antal är 11 hos grupp 2. Antalet och andelen informanter som har använt dessa adjektiv i respektive grupp presenteras i tabell 8.

Tabell 8 Antalet informanter (f) samt procentandelen av alla informanter (%) som har använt de mest frekventa adjektiven i de två grupperna

Grupp 1 (n=9)			Grupp 2 (n=6)		
Adjektiv	f	%	Adjektiv	f	%
social	8	88	social	6	100
bra	6	67	okej	5	83
okej	5	56	bra	3	50
intresserad	3	34	olik	3	50
god	3	34	gammal	3	50
viktig	3	34	riktig	3	50
dålig	3	34	ny	2	34
			intresserad	2	34
			kul	2	34
			normal	2	34
			rolig	2	34

Tabell 8 visar att det mest frekventa adjektivet i båda grupperna är *social*: 88 % av informanterna i grupp 1 och 100 % av informanterna i grupp 2 har använt det i sin text. Det är inte överraskande att gymnasisterna använder mycket adjektivet *social* eftersom uppsatsernas tema anknyter till de sociala medierna. Därtill är instruktionen för uppgiften skriven på svenska och uttrycket *de sociala medierna* förekommer i instruktionen (se bilaga 2).

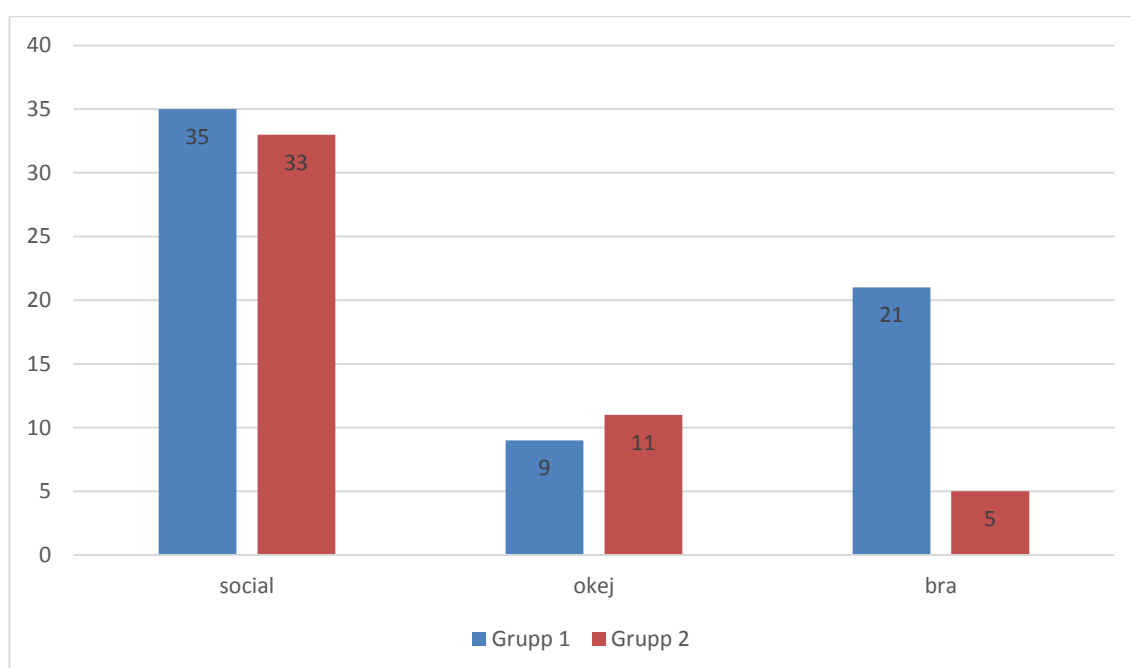
De två näst mest frekventa adjektiven efter adjektivet *social* i grupp 1 är *bra* (67 % av informanterna) och *okej* (56 % av informanterna). Resten av adjektiv, dvs. *intresserad*, *god*, *viktig* och *dålig* förekommer hos tre (34 %) informanter i grupp 1. I grupp 2 är det näst mest frekventa adjektivet *okej* (83 % av informanterna) efter adjektivet *social*. Adjektiven *bra*, *olik*, *gammal* och *riktig* förekommer hos hälften av informanterna. Resten av adjektiv, dvs. *ny*, *intresserad*, *kul*, *normal* och *rolig* förekommer hos två (34 %) informanter. Därmed kan det konstateras att de mest frekventa adjektiven totalt sett är *social*, *bra* och *okej*.

Allén (1971) har gjort en frekvensordlista över de mest frekventa adjektiven i svenska språket. Jag jämför de 50 mest frekventa adjektiven i Alléns lista med de mest frekventa adjektiven i båda gruppernas texter (se bilaga 4 för de 50 adjektiven i Alléns lista). Allén har dock inkluderat sådana adjektiv som jag inte klassificerar som adjektiv i denna undersökning (t.ex. *själv*, *egen*, *mer*, *fler*). Därför tar jag med de första 50 adjektiv som jag klassificerar som adjektiv från Alléns lista. I ordlistan finns det också olika böjningsformer av ett och samma adjektiv men jag betraktar dessa som ett ord och tar med dessa ord endast en gång.

När de mest frekventa adjektiven i Alléns lista jämförs med de mest frekventa adjektiven i båda gruppernas texter kan det noteras att adjektiven *god* och *bra* förekommer i texterna hos grupp 1 och adjektiven *bra*, *olik*, *gammal*, *riktig* och *ny* i texterna hos grupp 2. Det är intressant att grupp 2 använder fler adjektiv som hör till de vanligaste adjektiven enligt Allén (1971) än grupp 1. Detta skulle kunna betyda att informanterna i grupp 1 kan använda mer sofistikerat ordförråd än grupp 2, vilket är ett överraskande resultat. Antalet mest frekventa adjektiv i de två grupperna kan dock påverka resultaten eftersom antalet är större i grupp 2 än i grupp 1. Som redan konstaterades i avsnitt 7.3 är det också oklart

om gymnasisterna i grupp 1 kan använda dessa adjektiv i lämpliga kontexter och känner till betydelsen av dessa adjektiv.

I det följande betraktar jag adjektivens frekvens från den andra synpunkten, dvs. hur många gånger båda grupperna använder de tre mest frekventa adjektiven (*social*, *bra* och *okej*). Resultaten presenteras i figur 1.



Figur 1 Frekvens av adjektiven *social*, *bra* och *okej* i de två grupperna.

Figuren visar att frekvensen av adjektivet *social* är högst av de tre adjektiven. Grupp 1 använder adjektivet *social* 35 gånger i sina texter och grupp 2 33 gånger. Som framkommit ovan, är detta dock inte överraskande på grund av uppgiftens tema och anvisningar. Resultaten indikerar ändå att de 8 gymnasister som använder adjektivet *social* i grupp 1 och de 6 gymnasister som använder det i grupp 2 (se tabell 8) upprepar adjektivet i sina texter. Detta kan bero på att det kan vara svårt att hitta någon annan term för att beskriva begreppet sociala medier.

Frekvensen av adjektivet *okej* är relativt likadan i båda grupperna. Grupp 1 använder adjektivet 9 gånger i sina texter och grupp 2 11 gånger. Även om grupp 2 upprepar adjektivet *okej* mer än grupp 1 finns det procentuellt fler informanter i grupp 2 som

använder adjektivet *än* i grupp 1. Relationen är således relativt likadan i båda grupperna vad gäller användningen av adjektivet *okej*.

Det finns däremot en tämligen stor skillnad i frekvensen av adjektivet *bra* mellan de två grupperna. Grupp 1 använder adjektivet *bra* 21 gånger i sina texter medan grupp 2 använder samma adjektiv endast 5 gånger. När detta jämförs med de tidigare presenterade resultaten (se tabell 8), kan det noteras att grupp 1 upprepar adjektivet *bra* i sina texter. I grupp 2 är användningen av adjektivet *bra* mer sparsamt. I det följande avsnittet, 7.3.2, kommer jag att analysera närmare dessa tre mest frekventa adjektiv för att få en möjlig förklaring för användningen.

7.3.2 Användning av adjektiv i texterna

I detta avsnitt betraktar jag närmare de mest frekventa adjektiven, dvs. *social*, *bra* och *okej*, som båda grupperna ofta har använt i sina texter. Jag kommer att analysera användningen med hjälp av Nations (2001) modell av produktiv djup ordkunskap (se avsnitt 3.1.2). Jag behandlar ett adjektiv åt gången och jämför användningen mellan de två grupperna. Avslutningsvis, i avsnitt 7.3.2.4, presenterar jag andra adjektiv som informanterna använder i liknande kontexter som adjektiven *bra* och *okej*. Resultaten redovisas främst på gruppnivå. Jag presenterar också exempel från informanternas texter för att belysa användningen. I varje exempel är adjektiven understruken och informanternas ID-nummer, CEFR-nivå och grupp ges inom parentes.

7.3.2.1 Adjektivet *social*

Som redan konstaterades i avsnitt 7.3.1 använder informanterna mycket adjektivet *social* eftersom uppsatsens tema anknyter till de sociala medierna. Alla de 8 informanter som använder adjektivet *social* i grupp 1 stavar adjektivet rätt. I det följande ges tre exempel på användningen av adjektivet *social* hos grupp 1.

1. Men jag tycker att sociala medierna skulle inte samla information. (3002, B1, grupp 1)
2. Sociala mediernas sätt att samla information om människor som använder... (3009, A2, grupp 1)
3. Å ena sidan det kan göra lättare att använda sociala medier... (3015, B1, grupp 1)

I grupp 2 använder samtliga informanter adjektivet *social* i sina texter. Såsom i grupp 1 kan även informanterna i grupp 2 stava adjektivet rätt. I följande exempel (4–6) illustreras hur adjektivet *social* används i grupp 2.

4. Sociala medier ser ut som vanliga företag men de är olika. (3005, B2, grupp 2)
5. Jag tycker att de sätt de sociala medierna använder att samla information från denna användaren är helt okej, (3006, B2, grupp 2)
6. Jag tycker att det är inte så roligt att sociala mediernas sätt att samla information om mig. (3008, B2, grupp 2)

Exemplen visar att informanterna i båda grupperna använder den rätta formen av adjektivet, dvs. med ändelsen *-a* som här markerar pluralis. Man kan dock inte tolka detta så att informanterna själva kan stava eller böja adjektivet *social* rätt eftersom informanterna troligen har kopierat det direkt från instruktionen för uppgiften. I grupp 1 finns det ändå en informant som använder adjektivet *social* utan ändelsen *-a*. Detta exempel diskuterar jag i det följande.

De 8 informanterna i grupp 1 och de 6 informanterna i grupp 2 som använder adjektivet *social* använder det i korrekt grammatisk funktion, dvs. som attribut till substantivet *medium*. Som redan framgått i avsnitt 7.3.1 upprepar informanterna nominalfrasen *sociala medier* i sina texter eftersom det inte finns många alternativ till detta begrepp. I båda grupperna finns det också informanter som använder andra framförställda attribut i nominalfraser utöver adjektivet *social*. I det följande ges tre exempel på satser där dessa bestämmingar förekommer hos grupp 1 (de övriga bestämmingarna är kursiverade).

7. ...när man får reklam som är grund av information från *ers* sociala medierna. (3002, B1, grupp1)
8. Sociala mediernas sätt att samla information om människor som använder *annolunda* sociala medier till exempel instagram eller snapchat. (3009, A2, grupp1)
9. Det är okej bara om *någon* social media användare är inte privat... (3026, B1, grupp 1)

Exempel 7 och 9 visar att informanterna har placerat pronomen rätt i nominalfrasen, dvs. före adjektivet. I exempel 7 använder informanten adjektivet *social* rätt, dvs. med ändelsen *-a* som krävs efter genitiv (**ers* i exemplet). Substantivet *medium* borde dock stå i obestämd form efter possessivt pronomen men i exemplet står det i bestämd form. Detta kan indikera att informanten inte känner till eller kommer ihåg denna regel utan använder adjektivet *social* med ändelsen *-a* bara eftersom substantivet *medium* står i pluralis.

Exempel 7–9 visar vidare att informanterna kan forma relativt komplexa nominalfraser även om det finns några grammatiska fel eller skrivfel i exemplen. Den intressantaste frasen återfinns ändå i exempel 9 där informanten använder pronomenet *någon* och efter detta står adjektivet *social* i singularis men substantivet *medium* i pluralis (*media*). Denna informant är den enda som använder adjektivet *social* utan ändelsen *-a*. I sin text använder informanten också adjektivet med ändelsen *-a* i pluralis vilket kan innebära att informanten vet skillnaden mellan adjektivet *social* med ändelsen *-a* och utan ändelsen *-a*.

Det finns också informanter i grupp 2 som använder andra framförställda attribut i nominalfraser utöver adjektivet *social*. Följande exempel illustrerar användningen (de övriga bestämningarna är kursiverade).

10. Men nufördagen det är så svart att använda *ingen sociala* medier (3006, B2, grupp 2)
11. Så min åsikt är att det är inte okej att samla information från *mina sociala* medier. (3007, B2, grupp2)
12. Jag måste ha *olika sociala* medier t. ex. att skola och när jag pratar med mina kompisar. (3008, B2, grupp 2)

Exemplen visar att informanterna i grupp 2 också kan använda relativt komplexa nominalfraser och i exempel 10, 11 och 12 böjer informanterna adjektivet *social* rätt med hänsyn till nominalfrasens huvudord. På basis av detta material är det omöjligt att redogöra för om informanten i exempel 11 känner till regeln att adjektivet har ändelsen *-a* och substantivet är i obestämd form efter genitiv (*mina* i exemplet). I exempel 10

använder informanten adjektivet *social* med ändelsen *-a* men före adjektivet använder hen ändå pronomenet *ingen* i singularis. Användning av pronomenet i singularis indikerar att informanten troligen har velat använda begreppet *sociala medier* i singularis. Således tycks informanten inte känna till skillnaden mellan adjektivet *social* med ändelsen *-a* och utan ändelsen *-a*. Detta indikerar vidare att informanten har tagit adjektivet *social* direkt från instruktionen för uppgiften och inte kan använda det kreativt i denna kontext. En annan möjlig förklaring är att informanten inte kan böja pronomenet *ingen* rätt i pluralis. Pronomenet *inga* förekommer dock i informantens text i en annan kontext vilket betyder att denna förklaring är mindre sannolik.

Eftersom informanterna använder adjektivet *social* endast i ordkombinationen *social + medium* är det oklart om informanterna känner till de olika betydelseerna av adjektivet och i vilka kontexter detta adjektiv kan användas. Som redan framgått har informanterna sannolikt tagit adjektivet *social* direkt från instruktionen för uppgiften och därför är det också oklart om informanterna kan använda adjektivet produktivt i olika kontexter utan hjälp. Resultaten visar dock att det inte finns stora skillnader i användningen av adjektivet *social* mellan de två grupperna.

7.3.2.2 Adjektivet *bra*

Det andra adjektivet som är frekvent hos båda grupperna är adjektivet *bra*. De 6 informanter i grupp 1 och de 3 informanter i grupp 2 som använder adjektivet *bra* kan stava det rätt. I det följande presenteras exempel på en vanlig användning av adjektivet *bra* hos båda grupperna.

13. Jag tycker det är bra. (3009, A2, grupp 1)
14. Det skulle vara bra att vi kan välja att vilka information... (3015, B1, grupp 1)
15. Självklart, det är ett jättestor bra sätt att följa vad människor vill köpa... (3007, B2, grupp 2)
16. Men jag också tycker att enligt reklamen är bättre än reklamerna inte enligt för mig. (3008, B2, grupp 2)

Som framgår av exemplen vet informanterna i båda grupperna att adjektivet *bra* inte kongruensböjs efter huvudordet. Det är intressant att informanterna i grupp 2 inte

använder adjektivet som sådant utan adjektivet *bra* förekommer antingen med förstärkningsprefixet *jätte* (exempel 15) eller i komparativformen *bättre* (exempel 16). Förstärkningsprefixet *jätte* förekommer inte i texterna skrivna av informanterna i grupp 1 och komparativformen *bättre* förekommer bara hos en informant i grupp 1. Detta betyder naturligtvis inte att informanterna i grupp 1 inte kan använda dessa former av *bra* men det pekar ändå på att informanterna i grupp 2 varierar mer mellan olika former av adjektivet *bra* än informanterna i grupp 1.

I båda grupperna använder informanterna adjektivet *bra* i korrekta grammatiska funktioner. I texterna fungerar adjektivet *bra* antingen som predikativ eller som attribut till huvudordet. I det följande ges tre exempel på meningar där adjektiven fungerar som predikativ hos båda grupperna.

17. Jag tycker det är bra. (3009, A2, grupp 1)
18. Å andra sidan det skulle också vara bra att annonser... (3015, B1, grupp 1)
19. Även om sociala medier kan veta jättemycket om en människa är det bättre att de vet någonting... (3010, C1, grupp 2)

Exemplen visar att informanterna ofta använder ordkombinationen *det + vara + bra* när adjektivet *bra* fungerar som predikativ. Åtminstone i texterna hos grupp 1 är ordkombinationen *det + vara + bra* en av de vanligaste ordkombinationerna. Sammanlagt fyra informanter (45 %) i grupp 1 använder denna ordkombination. Tre av dessa informanter upprepar ordkombinationen två eller flera gånger i texten. Det finns endast en informant i grupp 2 som använder denna kombination (exemplet 19). Informanterna kan ha lärt denna ordkombination som en helhet, s.k. *helsekvens* (Sundman 2011) och därför använder de den många gånger i materialet.

Som ovan framgått fungerar adjektivet *bra* också som attribut i informanternas texter. Följande fyra exempel visar användningen hos båda grupperna.

20. Jag tycker att de är inte en bra sak. (3002, B1, grupp 1)
21. ...och rekommendera bra reklam för mig. (3027, A2, grupp 1)
22. Självklart, det är ett jättebra sätt att följa vad människor vill köpa... (3007, B2, grupp 2)

23. Det är en jättebra sak att sociala medier kan samla information om deras användare... (3010, C1, grupp 2)

Användning av ordkombinationen *en + bra + sak* i exemplen 20 och 23 är relativt vanlig åtminstone i grupp 1. En tredjedel av alla informanter i grupp 1 använder denna ordkombination i sina texter. Dessa informanter upprepar också kombinationen två eller flera gånger i sina texter. Däremot återfinns endast en informant i grupp 2 som använder denna kombination (exempel 23) och hen har ytterligare förstärkningsprefixet *jätte* med adjektivet *bra*. Ordkombinationen *en + bra + sak* kan också vara ett exempel på *helsekvens* eftersom den är så frekvent i texterna. Vidare förekommer det endast ett par nominalfraser som innehåller adjektivet *bra* och vars huvudord inte är substantivet *sak* i respektive grupp, dvs. *reklam* (exempel 21) och *sätt* (exempel 22). Resultaten indikerar således att informanterna i båda grupperna använder adjektivet *bra* i relativt enkla satser och nominalfraser och åtminstone informanterna i grupp 1 upprepar samma strukturer i sin text. Upprepning sker i mindre grad hos informanterna i grupp 2.

Som redan framkommit fick informanterna skriva om sina egna åsikter om de sociala mediernas sätt att samla information. Informanterna i båda grupperna använder adjektivet *bra* när de berättar om de accepterar eller ogillar de sociala mediernas informationssamling eller riktad reklam. I det följande ges fyra exempel på sådan användning av adjektivet *bra* i grupp 1.

24. Jag tycker det är bra. (3009, A2, grupp 1)
25. Jag tycker om att de är en bra sak att sociala mediernas anpassa reklam enligt information de har samlat. (3015, B1, grupp 1)
26. Jag tycker att det är bra att jag få reklamer som är bra till mig. (3016, B1, grupp 1)
27. Information samlar är bra saker. (3027, A2, grupp 1)

Exemplen visar hur informanterna uttrycker att de accepterar de sociala mediernas informationssamling (exempel 24 och 27) eller riktad reklam (exempel 25 och 26). I exempel 26 använder informanten adjektivet *bra* två gånger men i det understrukna fallet avser informanten med *bra* att hen accepterar riktad reklam. Detta sätt att använda adjektivet *bra* är relativt vanligt hos grupp 1. Sammanlagt fyra informanter (45 %) i grupp

1 använder adjektivet *bra* i denna kontext. Detta innebär att informanterna i grupp 1 använder ett relativt enkelt sätt att uttrycka sina åsikter.

Som redan framkommit använder grupp 2 också adjektivet *bra* när de uttrycker att de accepterar informationssamlingen eller riktad reklam. I följande ges alla meningar där informanterna i grupp 2 använder adjektivet *bra* i denna kontext.

28. Men jag också tycker att enligt reklamen är bättre än reklamerna inte enligt för mig. (3008, B2, grupp 2)
29. Det är en jättebra sak att sociala medier kan samla information om deras användare... (3010, C1, grupp 2)
30. Även om sociala medier kan veta jättemycket om en människa är det bättre att de vet någonting eftersom det hjälper deras användare att hitta saker... (3010, C1, grupp 2)

Exemplen visar att informanten i exempel 29 och 30 uttrycker att informationssamlingen är acceptabel och i exempel 28 uttrycker informanten att hen accepterar riktad reklam. Sammanlagt två informanter (35 % av alla informanter) i grupp 2 använder adjektivet *bra* i denna kontext. Detta betyder att användningen av adjektivet *bra* i denna kontext är mindre frekvent i grupp 2 än i grupp 1 men skillnaden är inte stor. Detta är dock inte överraskande eftersom informanterna i grupp 2 allmänt använder adjektivet *bra* i mindre utsträckning än informanterna i grupp 1. I exempel 28 använder informanten adjektivet för att jämföra om hen tycker mer om riktad eller oriktad reklam. I exempel 30 använder informanten adjektivet för att uttrycka sin åsikt att de sociala medierna borde veta någonting om sina användare. Användning av detta slag förekommer inte hos informanterna i grupp 1.

Resultaten indikerar således att bruket av adjektivet *bra* är relativt enkelt hos grupp 1 åtminstone i kontexten där de uttrycker att de accepterar informationssamlingen och riktad reklam. Resultaten visar vidare att användningen av adjektivet är mångsidigare hos grupp 2 än hos grupp 1. Härnäst kommer jag att betrakta hur informanterna använder adjektivet *bra* när de uttrycker att de inte accepterar informationssamlingen.

Informanterna uttrycker även sitt ogillande mot informationssamlingen med hjälp av adjektivet *bra* med negation och åtminstone i grupp 1 är detta bruk relativt frekvent. I det följande ges alla sådana satser hos grupp 1 (negationen är kursiverad).

31. Jag tycker att de är *inte* en bra sak. (3002, B1, grupp 1)
32. Men jag också tycker att det är *inte* så bra för barn eller undåmar... (3009, A2, grupp 1)
33. Jag tycker om att de är *inte* så bra sak. (3015, B1, grupp 1)

Exemplen visar att informanterna i grupp 1 använder adjektivet *bra* med negationen *inte* när de vill uttrycka att de inte tycker om att sociala medier samlar information om sina användare. Totalt tre (33 %) informanter använder denna ordkombination. Endast en informant av dessa använder både adjektivet *dålig* och kombinationen *inte + bra* i denna kontext. Detta kan betyda att informanterna i grupp 1 har ett relativt begränsat ordförråd vad gäller åtminstone negativa adjektiv i kontexter där de uttrycker sina åsikter och kan därför inte variera mellan olika adjektiv. Jag kommer att behandla de övriga adjektiven som informanterna använder i stället för adjektivet *bra* eller ordkombinationen *inte + bra* i avsnitt 7.3.2.4.

I grupp 2 använder informanterna inte adjektivet *bra* med negationen *inte* såsom i grupp 1. Som redan framkommit använder informanterna i grupp 2 även adjektivet *bra* i mindre utsträckning än informanterna i grupp 1. I stället för adjektivet *bra* använder informanterna i grupp 2 adjektivet *okej* mer i sådana situationer där de uttrycker att de accepterar eller ogillar informationssamlingen. Jag kommer att betrakta adjektivet *okej* hos båda grupperna närmare i avsnitt 7.3.2.3.

Informanterna använder också adjektivet *bra* i andra kontexter än när de uttrycker om de accepterar eller ogillar informationssamlingen eller riktad reklam. I det följande ges fem exempel på meningar där informanterna i båda grupperna använder adjektivet *bra* i olika kontexter.

34. Olika åsikter är en bra sak... (3002, B1, grupp 1)
35. Det är också ganska bra att reklamer som t.ex Facebook skickar är bra för användarna. (3016, B1, grupp 1)

36. Mycket av det handlas också för att ge dig bättre reklam. (3018, B1, grupp 1)
37. ...de kan berättar intressanta saker och rekommendera bra reklam för mig. (3027, A2, grupp 1)
38. Självklart, det är ett jättebra sätt att följa vad människor vill köpa... (3007, B2, grupp 2)

Av exemplen framgår att endast en informant i grupp 2 använder adjektivet *bra* i en annan kontext (exempel 38). Detta exempel kunde också tolkas så att informanten accepterar informationssamlingen. I exemplet använder hen dock adjektivet *bra* för att allmänt beskriva att informationssamlingen är nyttig för företag varför jag presenterar det här och inte i samband med de andra exemplen där informanterna uttrycker att de accepterar informationssamlingen.

Hos grupp 1 förekommer adjektivet *bra* ofta när informanterna beskriver substantivet *reklam* (exempel 35, 36 och 37). En tredjedel av informanterna använder adjektivet i denna kontext. Adjektivet *bra* används för att framföra att reklam är på något sätt lämplig för användaren. Anledningen till att informanterna i grupp 1 relativt ofta använder adjektivet *bra* i denna kontext kan vara att de inte känner till andra adjektiv som de skulle kunna använda. Sammanlagt 7 informanter i grupp 1 uttrycker ändå på något sätt att reklam är eller borde vara lämplig för användaren. Det finns således också andra strukturer som informanterna använder i stället för adjektivet *bra* i denna kontext. I det följande presenteras exempel på sådana meningar (strukturerna är understrukna).

39. Å andra sidan, kan man få reklam av intresserade varar... (3002, B1, grupp 1)
40. ...en ska få till exempel reklam som en är intresserad av. (3009, A2, grupp 1)
41. ...är det bar att jag se annonser på vanor som jag tycker om. (3015, B1, grupp 1)
42. På detta sätt det är möjligt att skicka riktiga reklamer också för barnen. (3016, B1, grupp 1)
43. På det sätt kan vi få reklam, som intresserar oss och produkter som vi behöver. (3018, B1, grupp 1)
44. Jag får reklamer som är riktade till mammor. (3025, B1, grupp 1)

Exemplen visar att endast en informant använder något annat adjektiv i stället för adjektivet *bra*, dvs. adjektivet *riktig* i exemplet 42. Adjektivet är ändå inte lämpligt i

denna kontext. Exempler visar vidare att andra informanter använder varierande strukturer för att uttrycka att reklam är eller borde vara lämplig för användaren. Sammanlagt fem informanter uttrycker endast en gång att reklam borde vara lämplig för användaren. De två informanter som använder detta uttryck två eller flera gånger varierar mellan olika strukturer. En av dessa två informanter använder adjektivet *bra* (exempel 36) och strukturen *reklam som intresserar oss* (exempel 43). Den andra informanten använder adjektivet *bra* (exempel 35), adjektivet *riktig* (exempel 42) och strukturen *reklamer som jag är inte intresserad av*.

7.3.2.3 Adjektivet *okej*

Det tredje adjektivet som är frekvent hos båda grupperna är adjektivet *okej*. Som i avsnitt 7.3.1 konstaterades använder grupp 2 mera adjektivet *okej* än grupp 1. I det följande ges tre exempel på användningen av adjektivet *okej* hos båda grupperna.

45. Jag tycker om att de är okej att sociala medierna... (3013, B1, grupp 1)
46. Det är inte ok för mig, att... (3007, B2, grupp 2)
47. Jag tycker att det är okej att samla information om man använder det riktigt. (3022, B2, grupp 2)

Utifrån exemplen kan det konstateras att informanterna kan stava adjektivet rätt. I exempel 46 använder informanten formen *ok* som vanligtvis används som en interjektion. Informanten använder dock också formen *okej* i sin text vilket betyder att hen också känner till detta alternativ. Informanterna i båda grupperna vet ytterligare att adjektivet är oböjligt. Exempler visar att användningen av adjektivet *okej* är likartad som användningen av adjektivet *bra* i båda grupperna (se avsnitt 7.3.2.2). Adjektivet *okej* används av informanterna för att uttrycka om de accepterar eller ogillar de sociala mediernas informationssamling. Adjektivet används i inga andra kontexter. Det är intressant att även om adjektiven *okej* och *bra* används av båda grupperna, finns det bara en informant i grupp 1 och en informant i grupp 2 som använder båda adjektiven i denna kontext. Detta betyder att de flesta informanter använder antingen adjektivet *bra* eller adjektivet *okej* i denna kontext och inte varierar mellan dessa två adjektiv.

I avsnitt 7.3.2.2 konstaterades att grupp 1 i relativt stor utsträckning använder ordkombinationen *inte + bra* i sina texter när de uttrycker sitt ogillande mot informationsinsamling. Detsamma gäller ordkombinationen *inte + okej* hos informanterna i grupp 2. I det följande ges två exempel på sådan användning hos grupp 2 (negationen *inte* är kursiverad).

48. Men om sociala medier följa någons plats var man är det är *inte okej* och... (3006, B2, grupp 2)

49. Fast det är lättare, är det *inte okej* att följa andra människors snapchat eller whatsapp. (3007, B2, grupp 2)

Exemplen visar att informanterna i grupp 2 använder ordkombinationen *inte + okej* på samma sätt som grupp 1 använder ordkombinationen *inte + bra*. Totalt tre (50 %) informanter i grupp 2 använder denna kombination. Det finns bara en informant i grupp 1 som använder kombinationen *inte + okej* i sin text. Det är intressant att endast en informant i grupp 2 använder ett negativt adjektiv i stället för denna kombination i denna kontext (se närmare avsnitt 7.3.2.4). Andra informanter i grupp 2 använder inte några andra negativa adjektiv som kunde ersätta denna ordkombination. Som redan framkommit i avsnitt 7.3.2.2 finns det endast en informant i grupp 1 som använder ett negativt adjektiv, *dålig*, i denna kontext. Detta betyder således att båda grupperna möjligen har ett bristande ordförråd, åtminstone vad gäller negativa adjektiv i kontexter där de uttrycker sina åsikter.

Alla informanter i båda grupperna använder adjektivet *okej* i den grammatiska funktion som adjektiv används, dvs. som predikativ. Adjektivet *okej* förekommer i ordkombinationen *det + vara + okej* hos alla fem informanter som har använt adjektivet i båda grupperna. Adjektivet *bra* förekommer även i liknande ordkombinationer när det fungerar som predikativ (se avsnitt 7.3.2.2). I det följande ges tre exempel på sådan användning av adjektivet *okej* hos båda grupperna.

50. Det är felt att använda och anpassa information förren man har frågat om det är okej och om... (3025, B1, grupp 1)

51. Det är okej att sociala medier samlar... (3007, B2, grupp 2)

52. Sociala medier samlar information om deras användare och det är helt okej. (3010, C1, grupp 2)

Exemplen visar hur ordkombinationen *det + vara + okej* vanligtvis används. I exempel 52 använder informanten adverbet *helt* framför adjektivet *okej* för att förstärka adjektivet. Det finns totalt två informanter som använder adverbet *helt* framför *okej* i sina texter. Sådana adverb förekommer inte i grupp 1. Det finns endast en informant i grupp 1 och en informant i grupp 2 som använder adjektivet *okej* i någon annan ordkombination än i denna. Resultaten visar således att de flesta informanter i båda grupperna använder adjektivet *okej* i den enkla ordkombinationen *det + vara + okej* som de upprepar i sina texter. Eftersom denna ordkombination förekommer relativt många gånger är det möjligt att det är en *helsekvens* som informanterna har lärt som en helhet (se också 7.3.2.2). Grupp 2 använder dock adverb för att förstärka adjektivet *okej* vilket indikerar att användningen är mer avancerad i grupp 2 än i grupp 1.

7.3.2.4 Andra adjektiv

Enligt Nations (2001) modell av djup ordkunskap är det viktigt att veta ordets betydelse, t.ex. vilka enheter ett specifikt ord kan syfta på eller vilka andra ord kan användas i stället för detta ord (se närmare avsnitt 3.1.2). Som redan framgått var syftet med uppgiften att skriva en argumenterande text där informanterna uttrycker sina åsikter med motiveringar. I avsnitten 7.3.2.2 och 7.3.2.3 konstaterade jag att informanterna ofta använder adjektiven *bra* och *okej* när de uttrycker om de accepterar eller ogillar de sociala mediernas informationssamling eller riktad reklam. I detta avsnitt betraktar jag de övriga adjektiven som de använder i liknande kontext och om dessa adjektiv är lämpliga med tanke på denna kontext. Detta betyder inte att dessa andra adjektiv är synonyma till adjektiven *bra* eller *okej* utan den gemensamma faktorn mellan dessa och adjektiven *bra* samt *okej* är att de används i samma kontext.

Informanterna i grupp 1 använder relativt sparsamt andra adjektiv än adjektiven *bra* och *okej* för att uttrycka att de accepterar informationssamlingen. I det följande ges exempel på dem.

53. Sociala mediernas sätt att samla information om användarna är god eftersom barn vill också använda sociala medierna. (3016, B1, grupp 1)
54. Jag tycker att det är viktigt att samla information om sociala mediernas användare. (3018, B1, grupp 1)
55. Jag tycker att det är rätt om till exempel facebook samlar information... (3027, A2, grupp 1)

Exemplen visar att informanterna använder tre andra adjektiv, dvs. *god*, *viktig* och *rätt*, i denna kontext. Adjektiven *viktig* och *rätt* förekommer båda endast en gång i denna kontext hos grupp 1. Adjektivet *god* som används i exempel 53 förekommer i denna kontext hos två informanter i grupp 1. Adjektivet är dock inte lämpligt med hänsyn till kontexten. Adjektiven *rätt* i exempel 55 och *viktig* i exempel 54 är lämpliga i denna kontext. Resultaten visar således att informanterna inte använder många andra adjektiv i en kontext där de uttrycker att de accepterar de sociala mediernas informationssamling. De kan ändå välja sådana adjektiv som är lämpliga med tanke på kontexten.

Det förekommer även färre andra adjektiv än *bra* eller *okej* hos grupp 2 än hos grupp 1. Endast ett adjektiv används i stället för dessa när informanterna i grupp 2 uttrycker att informationssamlingen är acceptabel. Detta adjektiv presenteras i exempel 56.

56. och ibland kan det vara kul att ha en reklam enligt för mig. (3007, B2, grupp 2)

Adjektivet *kul* förekommer hos två informanter i grupp 2. Adjektivet är lämpligt i denna kontext. Det är intressant att informanterna i grupp 2 i motsats till grupp 1 inte alls använder adjektivet *god* när de uttrycker att de accepterar informationssamlingen. Således förstår informanterna i grupp 2 sannolikt att adjektivet *god* inte är lämpligt i denna kontext. Anledningen till att informanterna i båda grupperna inte använder många andra adjektiv i denna kontext kan vara att informanterna i större utsträckning uttrycker negativa åsikter om de sociala mediernas informationssamling. Användningen av andra adjektiv i denna kontext betraktar jag härnäst.

Grupp 1 använder mera andra adjektiv när de uttrycker sitt ogillande mot de sociala mediernas informationssamling. Dessa adjektiv presenteras i följande exempel.

57. Det är inte sociala mediernas sak att veta och det är inte kul! (3002, B1, grupp 1)
58. Men de kan också vara dålig att till exempel Google vet så mycket om mig. (3015, B1, grupp 1)
59. ... det är inte riktigt att se hans/hennes informationer för alla. (3016, B1, grupp 1)
60. Men det känner inte god att veta, att man samlar min information... (3018, B1, grupp 1)
61. Det är felt att använda och anpassa information förrän man har frågat om det är okej... (3025, B1, grupp 1)
62. Ibland det går jätte lång om de samlar information och det är inte rätt, jag tycker att. (3027, A2, grupp 1)

Exemplen visar att det finns sammanlagt 6 adjektiv som informanterna i grupp 1 använder i stället för ordkombinationen *inte* + antingen *bra* eller *okej* när de uttrycker att de inte accepterar informationssamlingen. Det finns endast två negativa adjektiv, dvs. *dålig* och *fel* (exempel 58 och 61). Andra adjektiv förekommer i satserna med negationen *inte*. Jag konstaterade tidigare att det är möjligt att grupp 1 har ett bristande ordförråd vad gäller negativa adjektiv (se avsnitten 7.3.2.2 och 7.3.2.3) och detta förstärker tolkningen.

Som kan noteras i exempel 60 förekommer adjektivet *god* även i denna kontext. Som redan framgått är adjektivet dock inte lämpligt med tanke på kontexten. Adjektivet *god* förekommer hos totalt tre informanter (35 % av alla informanter) i grupp 1 med och utan negationen *inte*. Det är intressant att dessa tre informanter också använder adjektivet *bra* i sina texter vilket betyder att de känner till detta alternativ. Det är möjligt att de har velat variera mellan adjektiv och försökt hitta en synonym till adjektivet *bra*. Informanterna har ändå inte känt till skillnaderna i de betydelser som adjektiven *bra* och *god* bär och i vilka kontexter dessa adjektiv kan användas. De andra adjektiven som inte passar i denna kontext är adjektivet *dålig* i exempel 58 och adjektivet *riktig* i exempel 59. De andra adjektiven, dvs. *kul*, *fel* och *rätt* är lämpliga i denna kontext vilket betyder att informanterna sannolikt känner till betydelsen av dessa ord. Resultaten visar således att informanterna i grupp 1 har svårigheter att välja adjektiv som är lämpliga med tanke på denna kontext.

Grupp 2 använder också flera andra adjektiv när de uttrycker sitt ogillande mot informationssamlingen. Dessa adjektiv presenteras i följande exempel.

63. Men jag tycker att de är lite oärlig att smussla reklamer för gamla människor som förstår inte denna marknadsförning. (3006, B2, grupp 2)
64. Det skulle vara trevlig om de samla inga information från oss, men kanske det är omöjligt. (3006, B2, grupp 2)
65. Jag tycker att det är jättespännande. Jag vill inte att andra människor kan se till exempel mina sms:a. (3007, B2, grupp 2) (??)
66. Jag tycker att det är inte så roligt att sociala mediernas sätt att samla information om mig. (3008, B2, grupp 2)
67. ...är det inte riktig att samla information. (3022, B2, grupp 2)

Exemplen visar att det finns sammanlagt 5 andra adjektiv som informanterna använder i sina texter i denna kontext. Det är intressant att endast en informant använder ett negativt adjektiv (*oärlig*) i stället för ordkombinationen *inte* + antingen *bra* eller *okej*. Resultaten visar således att båda grupperna använder få negativa adjektiv vilket kan indikera att båda grupperna har ett bristande ordförråd vad gäller negativa adjektiv i kontexter där de uttrycker sina åsikter.

I exempel 65 är det något oklart vad informanten har velat uttrycka med adjektivet *jättespännande*. Det är ändå klart att hen uttrycker sin åsikt om de sociala mediernas informationssamling varför detta exempel presenteras här. Såsom i grupp 1 (se exempel 59) använder också en informant i grupp 2 adjektivet *riktig* (exempel 67) men igen är adjektivet olämpligt i denna kontext. Andra adjektiv, dvs. *oärlig*, *trevlig* och *rolig*, är lämpliga med tanke på kontext. Detta betyder att de informanter som har använt dessa adjektiv känner till betydelser av dessa ord och kan använda dem i rätt kontext. Resultaten indikerar att grupp 2 oftare använder lämpliga adjektiv än grupp 1 när de uttrycker sitt ogillande emot informationssamlingen.

8 Sammanfattning och diskussion

Syftet med denna studie har varit att undersöka lexikal komplexitet i texter skrivna av finska gymnasister på olika färdighetsnivåer. Som material har jag använt sammanlagt 15 skrivna texter som har insamlats inom projektet *Helsekvenser i tre morfologiskt olika inlärarespråk*. Dessa texter har bedömts enligt DIALUKI-skalan som baserar sig på CEFR. Jag har indelat texterna i två grupper enligt CEFR-nivåerna: grupp 1 som består av texterna på nivåerna A2 samt B1 och grupp 2 som består av texterna på nivåerna B2 samt C1. Jag har analyserat lexikal variation, lexikal täthet samt hur informanterna använder adjektiv i sina texter och jämfört den lexikala komplexiteten mellan de två grupperna.

Vad gäller den lexikala variationen kan det konstateras att grupp 2 har lexikalt mer varierande texter än grupp 1. Informanterna i grupp 2 har sannolikt ett bredare ordförråd än informanterna i grupp 1. Detta är inte överraskande eftersom informanterna i grupp 2 ligger på högre färdighetsnivåer. Mina resultat ligger i linje med tidigare forskning (se Agebjörn 2017 och Haikala 2012). Det finns dock variation inom båda grupperna vad gäller ordförrådets bredd. När jag jämförde mina resultat av den lexikala variationen med Haikalas (2012) framkom att den lexikala variationen är betydligt lägre hos informanterna i min undersökning än hos informanterna i Haikalas studie. Informanterna i denna undersökning verkar sålunda ha ett relativt begränsat ordförråd.

Angående den lexikala tätheten kom det fram att texterna skrivna av informanterna på lägre färdighetsnivåer är mer täta än texterna skrivna av informanterna på högre färdighetsnivåer. Detta är oväntat eftersom den lexikala tätheten, som Bulté och Housen (2012: 28) noterar, är ett tecken på lexikal komplexitet. Laurén (2003: 143) påpekar dock att den lexikala tätheten också kan ses som ett inlärarespråkligt drag så att tätheten blir hög om språkinläraren upprepar samma innehållsord. Detta gäller sannolikt även mina informanter på lägre färdighetsnivåer. Som redan ovan framkommit är den lexikala variationen lägre hos informanterna i grupp 1 än i grupp 2. Resultaten visar vidare att informanterna i grupp 1 upprepar mera samma innehållsord än grupp 2. Därför är tätheten högre hos grupp 1. Mina resultat ligger således i linje med Lauréns (2003) resultat vad gäller den lexikala tätheten.

När jag jämförde den lexikala tätheten i denna undersökning med Haikalas (2012) resultat framkom att den lexikala tätheten är relativt hög hos båda grupperna i mitt material och högre än i Haikalas studie. Anledningen till detta är sannolikt detsamma som redan framkommit. Den lexikala variationen är betydligt lägre hos båda grupperna i min undersökning än hos informanterna i Haikalas (2012) studie. Informanterna i min undersökning upprepar således sannolikt mer samma innehållsord än informanterna i Haikalas varför tätheten är högre i min studie.

I denna undersökning har jag även analyserat hur informanterna använder adjektiv i sina texter. Informanterna använder adjektiv i lika hög grad som infödda talare av svenska. Adjektivbruket är mer varierande i grupp 1 än i grupp 2. Även om informanterna i grupp 2 använder fler olika innehållsord än grupp 1 är användningen av olika adjektiv betydligt lägre i grupp 2 än i grupp 1. Detta är ett överraskande resultat eftersom andraspråksinlärare på högre färdighetsnivåer, som påpekats av Bergman & Abrahamsson (2004: 608–611), använder fler adjektiv än inlärare på lägre färdighetsnivåer.

Resultaten visar vidare att de mest frekventa adjektiven i de två grupperna totalt är adjektiven *social*, *bra* och *okej*. Andra frekventa adjektiv är *intresserad*, *god*, *viktig* och *dålig* i grupp 1 och *olik*, *gammal*, *riktig*, *ny*, *intresserad*, *kul*, *normal* och *rolig* i grupp 2. När dessa adjektiv jämfördes med Alléns (1971) lista över de 50 mest frekventa adjektiven i svenska språket framkom att informanterna på högre färdighetsnivåer använder fler adjektiv som hör till de vanligaste adjektiven än informanterna på lägre färdighetsnivåer. Detta är oväntat eftersom lexikal sofistisering enligt Bulté och Housen (2012: 28) är ett tecken på lexikal komplexitet och sålunda även på högre färdighetsnivå. Anledningen till detta kan dock vara att grupp 2 har ett större antal mest frekventa adjektiv än grupp 1. De mest frekventa adjektiven i båda grupperna är ändå relativt vanliga fast de 50 mest frekventa adjektiven enligt Allén (1971) inte innehåller alla dessa adjektiv. Detta kan bero på att materialet i Alléns (1971: XIII) frekvensordbok består av tidningar. Därför är de adjektiv som informanterna använder i denna undersökning och de adjektiv som förekommer i Alléns lista inte helt jämförbara.

Vad gäller användningen av adjektiven *social*, *bra* och *okej* i texterna visar resultaten att informanterna i båda grupperna upprepar adjektivet *social* mest av dessa tre adjektiv. Adjektivet *social* är frekvent i texterna eftersom uppgiftens tema anknyter till de sociala medierna varför resultatet inte är överraskande. Informanterna i de två grupperna använder adjektivet *okej* i lika hög grad. Adjektivet *bra* förekommer mer hos grupp 1 än hos grupp 2 och skribenterna i grupp 1 upprepar också adjektivet ofta i sin text.

Vidare visar resultaten att informanterna i båda grupperna kan stava och böja dessa tre adjektiv rätt. I Nations (2001) modell av djup produktiv ordkunskap finns det tre delar varav ordets form är den första. Utifrån resultaten kan det konstateras att informanterna i båda grupperna har tillägnat sig denna del av djup ordkunskap vad gäller åtminstone dessa tre adjektiv. Angående adjektivet *social* kan man dock inte dra den slutsatsen att informanterna själva kan stava och böja adjektivet eftersom adjektivet förekom i instruktionen för uppgiften och informanterna kan tänkas ha utnyttjat denna information.

Adjektiven *bra* och *okej* används mest när informanterna uttrycker om de accepterar eller ogillar de sociala mediernas informationssamling eller riktad reklam. Adjektiven är lämpliga med tanke på denna kontext. Detta kan tolkas så att informanterna känner till betydelsen av dessa två adjektiv. Informanterna i båda grupperna använder dock adjektiven *bra* och *okej* i relativt enkla ordkombinationer i denna kontext och dessa ordkombinationer upprepas många gånger i en och samma text. Därför kan dessa ordkombinationer vara exempel på helsekvenser (Sundman 2011) som lärts in som helhet. Vidare ger resultaten vid handen att informanterna i båda grupperna använder ett relativt enkelt sätt att uttrycka sina åsikter. Informanterna i grupp 2 använder dock förstärkningsprefixet *jätte* med adjektivet *bra* och adverbet *hel* med adjektivet *okej* i större utsträckning än grupp 1 vilket indikerar att användningen av dessa adjektiv är mer avancerad i grupp 2 än i grupp 1.

Vid analys av övriga adjektiv som informanterna använder när de uttrycker sina åsikter om de sociala mediernas informationssamling framkom att grupp 1 använder mer olämpliga adjektiv än grupp 2. Detta kan innebära att informanterna i grupp 1 inte känner till vilka betydelser dessa adjektiv kan ha och i vilka kontexter adjektiven kan användas. Exempelvis känner informanterna i grupp 1 inte till de betydelskillnader som adjektivet

bra och adjektivet *god* bär. Informanterna betraktar dessa två adjektiv som synonyma ord. Informanterna på högre färdighetsnivåer har således tillägnat sig en djupare ordkunskap än informanterna på lägre färdighetsnivåer vad gäller dessa adjektiv. Resultaten visar vidare att informanterna i båda grupperna dock har ett relativt begränsat ordförråd vad gäller negativa adjektiv i kontexter där informanterna uttrycker sina åsikter.

Som framgått är adjektivbruket mer varierande i grupp 1 än i grupp 2. Grupp 1 har ändå svårigheter att välja lämpliga adjektiv i kontexter där de uttrycker sina åsikter om de sociala mediernas informationssamling. Agebjörn (2017: 47) konstaterar att andraspråksinlärare behöver kunna använda ord i sitt basordförråd flytande och inte bara skapa ett brett ordförråd. Resultaten i denna undersökning visar att detsamma även gäller ordets betydelse. Det räcker inte att språksinlärare bara har ett brett ordförråd men inte känner till de betydelser som dessa ord bär och i vilka kontexter orden kan användas.

På grund av det ovannämnda borde språksinlärare fästa avseende vid adjektivens olika betydelser. Å andra sidan är det dock också viktigt att skapa ett brett ordförråd. Av resultaten i denna undersökning framgick att informanterna på högre färdighetsnivåer har ett bredare ordförråd än informanterna på lägre färdighetsnivåer. Exempelvis kan lärare i undervisningen uppmuntra elever att läsa mera olika slags texter. Som påpekas av Enström (2016: 25–26) påträffar inläraren på detta sätt ord i olika kontexter vilket hjälper hen att tillägna sig ordets olika betydelser. Läsning av olika slags texter hjälper eleven också att allmänt öka sitt ordförråd. Adjektiv kan dock vara svårare att tillägna sig från kontexter eftersom adjektiv inte är lika kontextberoende som t.ex. substantiv eller verb. Därför borde läraren fästa mera uppmärksamhet vid inläring av adjektiv också i klassrummet.

Sammanfattningsvis kan det konstateras att det finns ett samband mellan den lexikala komplexiteten och CEFR-nivåer vad gäller den lexikala variationen och användningen av lämpliga adjektiv med tanke på kontext. Detta kan tolkas så att även om DIALUKI-skalan inte innehåller direkta instruktioner om hur brett eller djupt ordförråd inläraren ska ha verkar dessa två aspekter ha betydelse när texterna bedöms. Det finns även ett samband mellan den lexikala tätheten och CEFR-nivåer och mellan adjektivvariation och CEFR-nivåer men i dessa fall var sambandet inte sådant som skulle ha förväntats.

I denna undersökning har jag studerat endast de mest frekventa adjektiven. I framtiden skulle man kunna undersöka mer ingående alla adjektiv i informanternas texter, dvs. hur djup ordkunskap informanterna har om alla adjektiv som de använder i sina texter. Lexikal sofistisering kunde också mätas mer omfattande genom att ta med alla adjektiv eller även alla ord i informanternas texter. Det skulle också vara intressant att undersöka om det finns ett samband mellan ordförråd som dessa informanter använder och ordförrådet som förekommer i läroböckerna.

Litteratur

- Agustín Llach, M.P., 2011: *Lexical errors and accuracy in foreign language writing*. Bristol: Multilingual matters.
- Allén, Sture 1971: *Nusvensk frekvensordbok baserad på tidningstext. 2, Lemman*. Stockholm: Almqvist & Wiksell.
- Agebjörn, Anders, 2017: Ordförråd och språkfärdighet – Lexikal komplexitet i (bedömning av) svenska som andraspråk. I: Sköldbberg, E., Andréasson, M., Adamsson Eryd, H., Lindahl, F., Lindström, S., Prentice, J. & Sandberg M. (red.), *Svenskans beskrivning 35*. Göteborg: Institutionen för svenska språket, Göteborg universitet. S. 37–52.
- Bergman, Pirkko & Abrahamsson, Tua, 2004: Bedömning av språkfärdigheten hos andraspråkselever. I: Hyltenstam K. & Lindberg I. (red.), *Svenska som andraspråk – i forskning, undervisning och samhälle*. Lund: Studentlitteratur. S. 597–625.
- Bulté, Bram & Housen, Alex, 2012: Defining and operationalising L2 complexity. I: Housen, A., Kuiken, F. & Vedder, I. (red.), *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam: John Benjamins Publishing Company. S. 21–46.
- Buss, Martina, 2002: *Verb i språkbudselevens lexikon: en sociolingvistisk studie i andraspråket*. Vaasan yliopiston julkaisuja. Acta Wasaensia Nr 105. Vasa: Vasa universitet.
- CEFR = Gemensam europeisk referensram för språk: Lärande undervisning och bedömning. Skolverket. Hämtad från: <https://www.skolverket.se/publikationer?id=2144>. Hämtad den 31 mars 2020.
- De Clercq, Bastine, 2015: The development of lexical complexity in second language acquisition. I: *EUROSLA Yearbook: Annual Conference of the European Second Language Association 15*. S. 69–94.
- Enström, Ingegerd, 2013: Ordförråd och ordinlärning: med särskilt fokus på avancerade inlärare. I: Hyltenstam K. & Lindberg I. (red.), *Svenska som andraspråk – i forskning, undervisning och samhälle*. Andra upplagan. Lund: Studentlitteratur. S. 169–195.
- Enström, Ingegerd, 2016: *Ordens värld: Svenska ord – struktur och inlärning*. Andra upplagan. Stockholm: Hallgren & Fallgren.

- Flyman Mattsson, Anna & Håkansson, Gisela, 2010: *Bedömning av svenska som andraspråk: en analysmodell baserad på grammatiska utvecklingsstadier*. Lund: Studentlitteratur.
- González-Fernández, Beatriz & Schmitt, Norbert, 2017: Vocabulary acquisition. I: Loewen, S. & Sato, M. (red.), *The Routledge Handbook of Instructed Second Language Acquisition*. New York: Routledge. S. 280–298.
- Haikala, Anne, 2012: Ordförrådet i gymnasist uppsatser – Hur skriver finskspråkiga abiturienter svenska i studentexamen? I: Björklund, S., Lönnroth, H. & Pilke, N. (red.), *Svenskan i Finland 13*. Vasa: Vasa universitet. S. 61–70.
- Helsekvenser i tre morfologiskt olika inlärspråk. Läroämnet nordiska språk. Hämtad från: <https://blogit.utu.fi/nordiskaforskning/helsekvenser-i-tre-morfologiskt-olika-inlararsprak/>. Hämtad den 30 mars 2020.
- Henriksen, Birgit, 1999: Three dimensions of vocabulary development. I: *Studies in second language acquisition 21* (2). S. 303–317.
- Housen, Alex, Kuiken, Folkert & Vedder, Ineke, 2012: Complexity, accuracy and fluency: definitions, measurement and research. I: Housen, A., Kuiken, F. & Vedder, I. (red.), *Dimensions of L2 Performance and Proficiency: Complexity, Accuracy and Fluency in SLA*. Amsterdam: John Benjamins Publishing Company. S. 1–20.
- Hultman, Tor G. & Westman, Margareta, 1977: *Gymnasistsvenska*. Lund: LiberLäromedel.
- Laurén, Ulla, 2003: *Språkbadslevers berättande. Studier i berättelsestruktur samt syntaktisk och lexikal kompetens i språkbadslevers muntliga och skriftligaproduktion*. Vaasan yliopiston julkaisuja. Tutkimuksia 247. Språkvetenskap 41. Vasa: Vasa universitet.
- LIX räknare. Hämtad från: <https://www.lix.se>. Hämtad den 23 maj 2019.
- Meara, Paul, 1996: The dimensions of lexical competence. I: Brown G., Malmkjær K. & Williams J. (red.), *Performance and Competence in Second Language Acquisition*. Cambridge: Cambridge University Press. S. 35–53.
- Nation, I.S.P., 2001: *Learning vocabulary in another language*. Cambridge: Cambridge University Press.
- Niitemaa, Marja-Leena, 2015: Kuinka vieraan kielen sanoja opitaan ja opetetaan. I: Pietilä P. & Lintunen P. (red.), *Kuinka kieltä opitaan: Opas vieraan kielen opettajalle ja opiskelijalle*. Helsinki: Gaudeamus.

- Puro, Tarja, 1999: Sanastollinen tieto ja suomen kielen oppikirjojen sanasto. I: *Virittäjä* 1. S. 2–26.
- Read, John, 2000: *Assessing Vocabulary*. Cambridge: Cambridge University Press.
- Schmitt, Norbert, 2010: *Researching Vocabulary: A Vocabulary Research Manual*. Basingstoke: Palgrave Macmillan.
- SAS = Hultman, Tor G., 2003: *Svenska akademiens språklära*. Stockholm: Svenska akademien.
- SAOB = Svenska Akademiens ordbok, utgiven av Svenska Akademien. Lund 1893–. Hämtad från: www.saob.se. Hämtad den 30 mars 2020.
- Sundman, Marketta 2011: Helsekvenser i finska grundskoleelevers skrivande. I: Edlund, A.C., Mellenius I. (red.), *Svenskans beskrivning 31*. Umeå: Institutionen för språkstudier, Umeå universitet. S. 327–336.
- Viberg, Åke, 1988: Ordförråd och ordinläring. Några utgångspunkter för en studie av hur ordförrådet lärs in. I: Hyltenstam K. & Lindberg I. (utg.), *Första symposiet om svenska som andraspråk. Volym I: Föredrag om språk och interaktion*. Stockholm: Centrum för tvåspråkighetsforskning, Stockholms universitet. S. 215–233.
- Viberg, Åke, 2004: Lexikal utveckling i ett andraspråk. I: Hyltenstam K. & Lindberg I. (red.), *Svenska som andraspråk – i forskning, undervisning och samhälle*. S. 197–220.

Bilagor

Bilaga 1: Bedömningskriterierna

DIALUKI-arviointiasteikko (perustuu EVK:n tasokuvauksiin)

	YLEIS- KUNVAUS	KIRJALLI- NEN VUOROVAI- KUTUS	KIRJEENVAIHTO & KIRJELAPUT, VIESTIT, LOMAKKEET	LUOVA KIRJOITTAMINEN & TEEMAN KEHITTELY & TEKSTIN SIDOSTEISUUS
A1	Pystyy kirjoittamaan yksinkertaisia, irrallisia ilmauksia ja lauseita.	Pystyy kysymään kirjallisesti toista koskevia yksinkertaisia henkilötietoja ja vastaamaan niitä koskeviin kysymyksiin.	Pystyy kirjoittamaan lyhyen, yksinkertaisen postikortin. Pystyy kirjoittamaan numeroita ja päivityksiä sekä oman nimen, kansallisuuden, osoitteen, iän, syntymäajan tai maahansaapumispäivän yms. esimerkiksi hotellin majoituskorttiin.	Pystyy kirjoittamaan yksinkertaisia ilmauksia ja lauseita itsestään ja kuvitelluista ihmisistä sekä siitä, missä he elävät ja mitä tekevät. Pystyy liittämään sanoja tai sanaryhmiä toisiinsa käyttäen yleisimpiä sidesanoja 'ja' tai 'sitten'.
A2	Pystyy yhdistämään yksinkertaisia ilmauksia ja lauseita tavanomaisilla sidesanoilla, kuten sanoilla 'ja', 'mutta', 'koska'.	Pystyy kirjoittamaan lyhyitä ja yksinkertaisia, vakiintuneista kaavamaisista ilmauksista koostuvia viestilappusia, jotka liittyvät välittömiin tarpeisiin.	Pystyy kirjoittamaan hyvin yksinkertaisia henkilökohtaisia kirjeitä, esimerkiksi kiitos- tai anteeksipyyntökirjeen. Pystyy kirjaamaan lyhyen, yksinkertaisen viestin, kun voi pyytää sen toistoa tai sen ilmaisemista toisin sanoin. Pystyy kirjoittamaan lyhyitä, yksinkertaisia kirjelappuja ja viestejä, jotka liittyvät välittömiin tarpeisiin.	Pystyy kirjoittamaan yksinkertaisia ilmauksia ja lauseita omasta perheestään, elinoloistaan, koulutustaustastaan sekä nykyisestä tai viimeisimmästä työstään. Pystyy kirjoittamaan lyhyitä, yksinkertaisia, mielikuvitukseen perustuvia tarinoita ja yksinkertaisia runoja ihmisistä ja heidän elämästään. Pystyy kirjoittamaan toisiinsa liitettyjen erillisten lauseiden avulla oman elinympäristönsä arkipäivään liittyvistä asioista, esimerkiksi ihmisistä, paikoista tai työhön tai opiskeluun liittyvistä kokemuksista. Pystyy kirjoittamaan hyvin lyhyitä, yksinkertaisia kuvauksia tapahtumista, menneistä toiminnoista ja henkilökohtaisista kokemuksista. Pystyy kertomaan tarinan tai kuvaamaan jotain yksinkertaisena asialuettelona.

				<p>Pystyy yhdistämään yksinkertaisia muutaman sanan muodostamia kokonaisuuksia tavanomaisimmilla sidosanoilla, kuten sanoilla 'ja', 'mutta' ja 'koska'.</p> <p>Pystyy tavallisimpia sidosanoja käyttäen liittämään yhteen yksinkertaisia lauseita kertomukseksi tai esittämään ne listana.</p>
--	--	--	--	--

	YLEIS- KUVAUS	KIRJALLI- NEN VUOROVAL- KUTUS	KIRJEENVAIHTO & KIRJELAPUT, VIESTIT, LOMAKKEET	LUOVA KIRJOITTAMINEN & TEEMAN KEHITTELY & TEKSTIN SIDOSTEISUUS
B1	<p>Pystyy kirjoittamaan yksinkertaisia, yhtenäisiä tekstejä tavallisista, itseään kiinnostavista aiheista yhdistämällä lyhyempiä, irrallisia, yksinkertaisia ilmauksia yhtenäiseksi tuotokseksi.</p>	<p>Pystyy välittämään tietoa ja ajatuksia abstrakteista ja konkreeteista aiheista, tarkistamaan tietoa sekä kysymään ongelmista ja selittämään niitä melko tarkasti.</p> <p>Pystyy henkilökohtaisissa kirjeissä ja viesteissä kysymään ja välittämään yksinkertaista, välittömän kokemuksen piiriin liittyvää tietoa ja korostamaan siinä tärkeinä pitämiään asioita.</p>	<p>Pystyy kirjoittamaan henkilökohtaisia kirjeitä, joissa kuvaa verrattain yksityiskohtaisesti kokemuksia, tunteita ja tapahtumia.</p> <p>Pystyy kirjoittamaan henkilökohtaisia kirjeitä, kertomaan niissä uutisia ja esittämään ajatuksia abstrakteista ja kulttuuriaiheista, kuten esimerkiksi musiikista ja elokuvista.</p> <p>Pystyy kirjoittamaan kirjalappusia, jotka sisältävät yksinkertaista, välittömään kokemukseen liittyvää tietoa ystäville, palveluhenkilökunnalle, opettajille ja muille hänen arkielämänsä liittyville henkilöille, sekä välittämään niissä tärkeinä pitämiään seikkoja ymmärrettävällä tavalla.</p> <p>Pystyy kirjaamaan viestejä, jotka sisältävät tiedusteluja tai selittävät ongelmia.</p>	<p>Pystyy kirjoittamaan yksinkertaisia, yksityiskohtaisia kuvauksia monista itseään kiinnostavista aiheista. Pystyy kirjoittamaan selostuksia kokemuksista ja kuvaamaan tunteita ja reaktioita yksinkertaisessa, yhtenäisessä tekstissä. Pystyy kirjoittamaan kuvauksia tapahtumista, esimerkiksi joko todellisesta tai kuvitellusta äskettäin tehdystä retkestä. Pystyy kertomaan tarinan.</p> <p>Pystyy esittämään suhteellisen sujuvasti yksinkertaisen kertomuksen tai kuvauksen asialuettelona.</p> <p>Pystyy yhdistämään joukon lyhyempiä, irrallisia, yksinkertaisia ilmauksia yhtenäiseksi tekstiksi.</p>

B2	Pystyy kirjoittamaan selkeää, yksityiskoh- taista tekstiä erilaisista itseään kiinnostavista aiheista. Pystyy tekemään yhteenvetoja sekä arvioimaan useista lähteistä poimimaan tietoja ja perusteluja.	Pystyy kirjoituksessaan ilmaisemaan tehokkaasti tietoja ja näkemyksiä ja suhteuttamaan ne muiden tietoihin ja näkemyksiin.	Pystyy kirjoittamaan kirjeitä, jotka välittävät erivahvuisia tunteita, korostavat tapahtumien ja kokemusten henkilökohtaista merkitystä ja kommentoivat kirjeen lähettäjän uutisia ja ajatuksia.	Pystyy kirjoittamaan selkeitä, yksityiskohtaisia kuvauksia hyvinkin erilaisista itseään kiinnostavista aiheista. Pystyy kirjoittamaan arvostelun elokuvasta, kirjasta tai näytelmästä. Pystyy kirjoittamaan selkeitä, yksityiskohtaisia kuvauksia todellisista tai kuvitelluista tapahtumista ja kokemuksista. Pystyy osoittamaan ajatusten väliset yhteydet selkeässä, yhtenäisessä tekstissä ja seuraamaan kyseisen tyylin vakiintuneita käytänteitä. Pystyy tuottamaan selkeän kuvauksen tai kertomuksen, jossa kehittää keskeisiä seikkoja ja esittää niitä tukevia yksityiskohtia ja esimerkkejä. Pystyy käyttämään joitakin kielellisiä keinoja rakentaakseen ilmauksista selkeän ja yhtenäisen tekstin. Pitkissä tuotoksissa voi ilmetä jonkin verran puutteita sidosteisuudessa. Pystyy käyttämään tehokkaasti useita sidosanoja osoittaakseen selvästi ajatusten välisiä suhteita.
-----------	---	--	--	---

	YLEIS- KUVAUS	KIRJALLI- NEN VUOROVAI- KUTUS	KIRJEENVAIHTO & KIRJELAPUT, VIESTIT, LOMAKKEET	LUOVA KIRJOITTAMINEN & TEEMAN KEHITTELY & TEKSTIN SIDOSTEISUUS
C1	Pystyy laatimaan selkeitä, hyvin jäsenneltyjä tekstejä monipolvisis- takin aiheista ja korostamaan tärkeitä asioita. Pystyy jossain määrin tukemaan näkemyksiään lisäseikoilla, perusteluilla ja osuvilla esimerkeillä	Pystyy ilmaisemaan itseään selvästi ja täsmällisesti ja ottamaan vastaanottajan huomioon joustavasti ja tehokkaasti	Pystyy ilmaisemaan itseään selkeästi ja täsmällisesti henkilökohtaisessa kirjeenvaihdossaan ja käyttämään kieltä joustavasti ja tehokkaasti myös emotionaalisiin, epäsuoria viittauksia sisältäviin ja leikillisiin tarkoituksiin.	Pystyy kirjoittamaan selkeitä, yksityiskohtaisia, hyvin jäsenneltyjä ja kehiteltyjä kuvauksia ja tekstejä kuvitelluista aiheista varman persoonallisella ja luontevalla tyylillä, joka sopii ajatellulle kohdeyleisölle. Pystyy esittämään yksityiskohtaisia kuvauksia ja kertomuksia, liittämään niihin alateemoja, kehittelemään keskeisiä seikkoja ja lopettamaan tekstin sopivalla tavalla. Pystyy tuottamaan selkeää ja loogisesti etenevää tekstiä. Osoittaa hallitsevansa tekstin

	sekä päättämään kirjoituksensa sopivaan lopetukseen.			jäsentämistavat ja sidosteisuudesta huolehtimisen.
C2	Pystyy laatimaan selkeitä, hyvin jäsenyneitä tekstejä monipolvisista aiheista ja tuottamaan asianmukaisen ja tehokkaan johdonmukaisen rakenteen, joka auttaa lukijaa löytämään tekstistä keskeiset seikat.	Kuten C1	Kuten C1	<p>Pystyy kirjoittamaan selkeitä, sujuvia ja mukaansatempaavia tarinoita ja kokemusten kuvauksia tyyllillä, joka sopii valittuun tekstilajiin.</p> <p>Pystyy tuottamaan sisällöltään ja muodoltaan johdonmukaista tekstiä käyttäen monipuolisesti ja tarkoituksenmukaisesti sopivia tekstin jäsentämistapoja ja sidosteisuutta parantavia keinoja.</p>

Asteikkojen lähde: EVK s. 96, 97, 121, 122, 175, 176

Bilaga 2: Instruktionen för uppgiften

Sinulla on n. 30 minuuttia aikaa kirjoittaa. Kirjoita 150–250 sanaa seuraavasta aiheesta. Otsikoi itse.

Pohjola-Nordens Ungdomsförbund vill veta vad nordiska ungdomar tycker om de sociala mediernas sätt att samla information om deras användare och att t.ex. anpassa reklam enligt denna information. Delta i diskussionen och skriv ett inlägg på deras webbplats där du berättar vad du själv tycker om detta. Kom ihåg att motivera dina åsikter. Sätt rubriken själv.

Bilaga 3: Adjektiv i informanternas texter enligt de två grupperna

Grupp 1		Grupp 2	
social	lätt	social	sjuk
bra	sjuk	okej	stor
okej	stor	bra	anonym
intresserad	allmän	olik	dum
god	effektiv	gammal	frivillig
viktig	enkel	riktig	full
dålig	etisk	ny	*olagvis
olik	fin	intresserad	omöjlig
rätt	liten	kul	oärlig
ny	lång	normal	sann
fel	möjlig	rolig	självklar
gammal	personlig	dålig	spännande
riktig	rädd	privat	svår
kul	tråkig	billig	trevlig
billig	privat	intressant	typisk
intressant		lätt	vanlig

Bilaga 4: De 50 mest frekventa adjektiven enligt Allén (1971)

1. stor
2. ny
3. svensk
4. gammal
5. olik
6. hög
7. god
8. lång
9. ung
10. svår
11. amerikansk
12. liten
13. viktig
14. politisk
15. möjlig
16. modern
17. klar
18. stark
19. vanlig
20. vit
21. fransk
22. ekonomisk
23. fri
24. tysk
25. tidig
26. allmän
27. brittisk
28. intressant
29. internationell
30. kort
31. fin
32. nära
33. bra
34. vacker
35. lätt
36. låg
37. enkel
38. full
39. nödvändig
40. röd
41. riktig
42. engelsk
43. svart
44. enskild
45. aktuell
46. naturlig
47. öppen
48. allvarlig
49. hård
50. känd

Lyhennelmä

Tässä pro gradu -tutkielmassa tarkastelen sanaston kompleksisuutta suomalaisten ruotsin kielen oppijoiden kirjoittamissa aineissa Eurooppalaisen viitekehyksen (CEFR) eri taitotasolla. Sanaston kompleksisuutta tarkastellaan leksikaalisen variaation sekä leksikaalisen tiheyden näkökulmasta. Leksikaalisella variaatiolla tarkoitetaan sitä, kuinka monipuolista sanastoa teksteissä käytetään ja leksikaalisella tiheydellä sitä, kuinka paljon oppijat käyttävät sisältösanoja teksteissään. Lisäksi tutkimuksessa pyritään selvittämään, miten oppijat käyttävät adjektiiveja teksteissään. Tarkoituksena on tutkia, vaihteleeko leksikaalinen variaatio, leksikaalinen tiheys sekä adjektiivien käyttö eri taitotasolla. Tavoitteena on vastata seuraaviin kysymyksiin:

- Kuinka monipuolista sanastoa oppijat käyttävät teksteissään eri taitotasolla?
- Kuinka leksikaalisesti tiheitä tekstit ovat eri taitotasolla?
- Mitä adjektiiveja oppijat käyttävät eri taitotasolla?
- Miten oppijat käyttävät adjektiiveja eri taitotasolla?

Tarkastelen eri adjektiivien esiintyvyyttä sekä kuinka tavallisia käytetyimmät adjektiivit ovat. Analysoin myös, miten oppijat tavuttavat ja taivuttavat adjektiiveja sekä missä kieliopillisissa funktioissa nämä adjektiivit esiintyvät. Tämän lisäksi selvitän, missä konteksteissa adjektiivit esiintyvät ja sopivatko nämä adjektiivit käytettäväksi kyseisissä konteksteissa.

Aineisto ja tutkimusmenetelmät

Tutkimusmateriaalini koostuu 15:sta suomalaisen lukiolaisen kirjoittamasta aineesta. Materiaali sisältyy Turun yliopiston tutkimusprojektiin *Helsekvenser i tre morfologiskt olika inlära språk* jota rahoittaa *Svenska kulturfonden*. Tehtävänantona oli kirjoittaa ruotsiksi 150–250 sanan argumentoiva teksti, jossa lukiolaiset ilmaisivat omia mielipiteitään tavasta, jolla sosiaalinen media kerää tietoa käyttäjistään. Yhteensä 16 lukiolaista kirjoitti aineen. Yksi teksteistä oli kuitenkin hyvin lyhyt (n. 50 sanaa), jonka vuoksi päätin jättää tämän tekstin pois analyysistä. Aineet on kirjoitettu Scriptlog-näppäilyntallennusohjelmalla, jolla voidaan tarkastella kirjoitusprosessia. Tässä

tutkielmassa käytän kuitenkin vain valmiita tekstejä, sillä tarkoituksena ei ole tutkia kirjoitusprosessia.

Aineet on arvioitu ns. DIALUKI-skaalan mukaan, joka pohjautuu Eurooppalaiseen viitekehukseen (CEFR). Tekstit sijoittuvat taitotasolle A2 (2 tekstiä), B1 (7 tekstiä), B2 (5 tekstiä) ja C1 (1 teksti). Koska taitotasolla A2 on vain kaksi tekstiä ja tasolla C1 yksi teksti, jaan aineet kahteen ryhmään: ryhmään 1, joka koostuu teksteistä taitotasolla A2 ja B1 (9 tekstiä) sekä ryhmään 2, joka koostuu teksteistä taitotasolla B2 ja C1 (6 tekstiä). Tulokset analysoidaan näiden kahden ryhmän perusteella.

Käytän tutkimuksessani sekä kvantitatiivisia että kvalitatiivisia menetelmiä. Kvantitatiivisessa analyysissä lasken juoksevat sanat sekä lekseemit lukiolaisten teksteissä. Tässä käytän apuna LIX-laskuria. Lekseemien laskemisessa käytän samaa periaatetta, jota Haikala (2012) on käyttänyt omassa tutkimuksessaan, eli lasken sanan eri taivutusmuodot yhdeksi ja samaksi lekseemiksi. Tämän lisäksi luokittelen lekseemit sanaluokkien mukaan. Leksikaalisen variaation laskemiseen käytän ns. OVIX-arvoa, jota myös Haikala (2012) on käyttänyt omassa tutkimuksessaan. Arvon laskemiseksi käytetään kaavaa $\ln(N)/\ln(2-\ln(V)/\ln(N))$, jossa N kuvaa juoksevia sanoja ja V lekseemejä. Leksikaalisen tiheyden lasken jakamalla sisältösanojen määrän juoksevien sanojen määrällä. Käytän tästä lyhennettä SAVKVOT, jota myös Haikala (2012) on käyttänyt omassa tutkimuksessaan. Vertailen yllä mainittujen ryhmien OVIX- sekä SAVKVOT-arvojen keskiarvoja keskenään.

Leksikaalisen variaation sekä leksikaalisen tiheyden lisäksi lasken kaikki teksteissä esiintyvät adjektiivit ja selvitän, missä määrin lukiolaiset käyttävät eri adjektiiveja. Tarkastelen molemmissa ryhmissä adjektiivien määrän sekä juoksevien sanojen määrän suhdetta sekä eri adjektiivien määrän sekä kaikkien adjektiivien määrän suhdetta. Selvitän myös, mitkä adjektiivit ovat käytetyimpiä molemmissa ryhmissä ja vertailen käytetyimpiä adjektiiveja Allénin (1971) laatiman frekvenssilistan kanssa, johon on kerätty ruotsin kielen käytetyimmät adjektiivit. Otan vertailuun mukaan ensimmäiset 50 frekvenssilistan adjektiivia. Tämän avulla tarkastelen, kuinka tavallisia adjektiiveja lukiolaiset käyttävät teksteissään. Selvitän myös, kuinka monta kertaa käytetyimmät adjektiivit esiintyvät molempien ryhmien teksteissä. Lopuksi tutkielman kvalitatiivisessa

osassa tarkastelen lähemmin teksteissä esiintyviä käytetyimpiä adjektiiveja käyttäen apuna Nationin (2001) mallia, jossa määritellään, mitä osa-alueita sanan syvälliseen osaamiseen kuuluu. Vertailen tuloksia molempien ryhmien välillä.

Tutkimuksen teoreettinen tausta sekä aiempi tutkimus

Tutkielmani teoriaosassa esittelen ensin, miten käsite *sana* on määritelty kirjallisuudessa. Niitemaa (2015) toteaa, että sana voidaan esimerkiksi määritellä sen mukaan, miten sanat lasketaan tekstissä. Tekstistä voidaan laskea esimerkiksi *juoksevat sanat* sekä *lekseemit*. (Niitemaa 2015: 140.) Lisäksi esittelen lyhyesti ruotsin kielessä esiintyvät sanaluokat sekä miten sanaluokkia voidaan luokitella. Sanaluokkien sanat voidaan jaotella esimerkiksi *sisältösanoihin* sekä *kielioppisanoihin* (Niitemaa 2015: 140). Tarkastelen myös lähemmin adjektiiveja sanaluokkana, eli mitä merkityksiä adjektiiveilla voidaan ilmaista sekä mitä kieliopillisia funktioita adjektiiveilla on. Lisäksi esittelen, miten *sanasto* käsitteenä on määritelty kirjallisuudessa. Tähän liittyy läheisesti ns. *mentaalileksikko*, joka tarkoittaa mielessä sijaitsevaa yksilöllistä sanavarastoa, jonka sanat muodostavat laajan verkoston (Niitemaa 2015: 145).

Kolmannessa kappaleessa käsittelen sanojen osaamiseen sekä oppimiseen liittyviä osa-alueita. Esittelen ensin *leksikaalisen kompetenssin* käsitteen, jota on usein kuvailtu *sanaston laajuuden* sekä *sanaston syvyyden* avulla. Sanaston syvyyden määrittelyn yhteydessä esittelen Nationin (2001) mallin, jossa määritellään mitä osa-alueita sanan syvälliseen osaamiseen kuuluu. Lisäksi esittelen *leksikaalisen kompleksisuuden* käsitteen, joka liittyy läheisesti sanaston laajuuden ja syvyyden käsitteisiin. Käsite kompleksisuus on yksi kielen osaamisen osa-alue ns. CAF-teoriassa (kompleksisuus, tarkkuus ja sujuvuus) (Housen, Kuiken & Vedder 2012: 1).

Leksikaalista kompleksisuutta voidaan tarkastella *leksikaalisen variaation*, *leksikaalisen tiheyden* sekä *leksikaalisen sofistikoituneisuuden* näkökulmasta (Bulté & Housen 2012: 28). Leksikaalinen variaatio kuvaa sitä, kuinka monipuolista sanastoa tekstissä käytetään (Haikala 2012: 63). Leksikaalinen tiheys kuvastaa taas sitä, kuinka paljon sisältösanoja teksti sisältää (Bulté & Housen 2012: 31). Leksikaalisella sofistikoituneisuudella tarkoitetaan sitä, miten paljon tekstissä esiintyy vähemmän tavallisia sanoja (Read 2000: 200). Näistä leksikaalinen variaatio sekä leksikaalinen tiheys kuvaavat sanaston laajuutta

ja leksikaalinen sofistikoituneisuus sanaston syvyyttä (Bulté & Housen 2012: 25). Leksikaalisen kompleksisuuden käsitteen määrittelyn jälkeen tarkastelen sitä, miten sanoja opitaan. Esittelen esimerkiksi, missä järjestyksessä eri sanaluokkien sanoja opitaan ja miten oppija voi oppia sanoja tietoisesti (ru. *intentionell inlärning*) tai tiedostamatta (ru. *incidentiell inlärning*).

Teoriaosassa esittelen myös Eurooppalaisen viitekehyksen sekä DIALUKI-skaalan, joka perustuu viitekehykseen. Koska tutkielmani aiheena on tarkastella kielenoppijoiden sanastoa taitotasolla A2, B1, B2 ja C1, esittelen mitä näiden DIALUKI-skaalan taitotasojen arviointikriteereissä sanotaan sanaston osaamisesta. DIALUKI-skaala ei mainitse tarkkoja listoja siitä, mitä sanoja oppijoiden tulee osata tietyllä taitotasolla, mutta skaala kuvailee kuitenkin tiettyjä teemoja, joista oppijan tulee osata kertoa.

Teoriaosuuden lopuksi esittelen tutkimukseni kannalta keskeistä aiempaa tutkimusta. Agebjörn (2017) on tarkastellut leksikaalista kompleksisuutta aikuisten ruotsin kielen oppijoiden suullisissa tuotoksissa, jotka on arvioitu Eurooppalaisen viitekehyksen taitotason B2 mukaan. Laurénin (2003) pitkittäistutkimuksessa on tarkasteltu ruotsin kielikylypöppiläiden kirjallisia sekä suullisia kertomuksia, niiden rakennetta sekä leksikaalisia ja syntaktisia piirteitä. Lisäksi näitä kertomuksia on verrattu äidinkielenään ruotsia puhuvien oppilaiden tuotoksiin sekä ruotsia perinteisessä opetuksessa opiskelleiden suomenkielisten oppilaiden tuotoksiin. Haikala (2012) on tarkastellut sanaston laatua suomalaisten abiturienttien aineissa, jotka he ovat kirjoittaneet ruotsin kielen ylioppilaskirjoituksissa. Haikala on jaotellut aineet kolmeen taitotasoryhmään Ylioppilastutkintolautakunnan arvioinnin perusteella. Haikala (2012), Agebjörn (2017) sekä Laurén (2003) ovat kaikki tutkineet sanastoa mittaamalla leksikaalisen variaation sekä leksikaalisen tiheyden oppijoiden tuotoksissa.

Haikalan (2012) tutkimuksessa kävi ilmi, että leksikaalinen variaatio on suurempi niiden abiturienttien aineissa, jotka ovat saaneet paremman arvosanan. Leksikaalinen tiheys ei kuitenkaan eroa juurikaan ryhmien välillä. Agebjörnin (2017) tutkimus osoitti, että vain sanaston laajuus korreloi merkittävästi arvioinnin kanssa. Lisäksi arviointi on ollut positiivisempi niiden oppijoiden tuotoksissa, joissa esiintyy paljon tavallisia sanoja. Laurénin (2003) tutkimuksessa kävi ilmi, että leksikaalinen variaatio on suurin ruotsia

äidinkielenään puhuvien oppilaiden aineissa. Leksikaalinen variaatio on aluksi suurempi kielikylpyoppilaiden teksteissä kuin ruotsia perinteisessä opetuksessa opiskelleiden oppilaiden teksteissä, mutta vuoden kuluttua ero on kuitenkin tasoittunut. Leksikaalinen tiheys on suurempi kielikylpyoppilaiden ja ruotsia perinteisessä opetuksessa opiskelleiden oppilaiden aineissa kuin ruotsia äidinkielenä puhuvien oppilaiden teksteissä. Leksikaalinen tiheys on aluksi suurempi perinteistä koulua käyneillä oppilailla kuin kielikylpyoppilailla, mutta vuoden kuluttua ero on tasoittunut.

Tärkeimmät tutkimustulokset

Tuloksista käy ilmi, että ylemmillä taitotasoilla B2 ja C1 leksikaalinen variaatio on suurempi kuin alemmilla taitotasoilla A2 ja B1. Tämä tarkoittaa todennäköisesti sitä, että ylemmillä taitotasoilla olevilla lukiolaisilla on laajempi sanavarasto kuin alemmilla taitotasoilla olevilla lukiolaisilla. Tulokseni ovat yhteneväisiä aiempien tutkimusten kanssa. Vertailu Haikalan (2012) tutkimuksen kanssa osoittaa, että minun tutkimuksessani olevien lukiolaisten leksikaalinen variaatio on huomattavasti pienempi kuin Haikalan tutkimuksessa olevien lukiolaisten leksikaalinen variaatio. Tutkimukseni lukiolaisilla vaikuttaa siis olevan suhteellisen suppea sanavarasto.

Leksikaalinen tiheys on suurempi alemmilla taitotasoilla kuin ylemmillä taitotasoilla. Tämä on yllättävää, sillä leksikaalista tiheyttä pidetään merkinä leksikaalisesta kompleksisuudesta. Laurén (2003: 143) toteaa kuitenkin, että korkea leksikaalinen tiheys voi olla merkki oppijankielelle ominaisesta piirteestä niin, että toistamalla samoja sisältösanoja voidaan saavuttaa korkeampi leksikaalinen tiheys. Tämän vuoksi olen leksikaalisen variaation lisäksi laskenut, kuinka paljon lukiolaiset käyttävät samoja sisältösanoja teksteissään. Tulokset osoittavat, että alemmilla taitotasoilla toistetaan enemmän samoja sisältösanoja kuin ylemmillä taitotasoilla. Tämä on siis mahdollinen selitys sille, että leksikaalinen tiheys on suurempi alemmilla taitotasoilla. Vertailtuani tuloksiani Haikalan (2012) tutkimuksen kanssa käy ilmi, että teksteillä minun tutkimuksessani on suhteellisen suuri leksikaalinen tiheys. Koska leksikaalinen variaatio on kuitenkin huomattavasti pienempi minun materiaalissani kuin Haikalan (2012) tutkimuksessa, voi tässäkin mahdollisena syynä pitää sitä, että lukiolaiset minun tutkimuksessani toistavat enemmän samoja sisältösanoja kuin lukiolaiset Haikalan tutkimuksessa.

Analyysistani käy myös ilmi, että lukiolaiset käyttävät yhtä paljon adjektiiveja kuin ruotsia äidinkielenään puhuvat. Adjektiiveja käytetään kuitenkin monipuolisemmin alemmilla taitotasoilla kuin ylemmillä taitotasoilla. Kolme käytetyintä adjektiiviä lukiolaisten teksteissä ovat *social*, *bra* ja *okej*. Näistä kolmesta adjektiivista adjektiivia *social* toistetaan eniten. Tämä ei ole yllättävää, sillä aineiden teemana on sosiaalinen media. Lisäksi adjektiivit *social*, *bra* ja *okej* on tavutettu sekä taivutettu oikein sekä alemmilla että ylemmillä taitotasoilla. Adjektiiveja *bra* ja *okej* käytetään eniten sellaisissa tilanteissa, joissa lukiolaiset ilmaisevat mielipiteitään siitä, onko sosiaalisen median tiedonkeruu tai kohdennettu mainonta hyväksyttävää. Adjektiivit sopivat käytettäväksi tässä kontekstissa.

Sekä alemmilla että ylemmillä taitotasoilla käytetään adjektiiveja *bra* ja *okej* suhteellisen yksinkertaisissa sanakombinaatioissa, joita toistetaan usein. Tämä osoittaa, että lukiolaiset käyttävät suhteellisen yksinkertaisia tapoja ilmaista mielipiteitään. Nämä sanakombinaatiot voivat olla myös esimerkki ns. sanastokeskittymistä (Sundman 2011), jotka opitaan kokonaisuutena. Ylemmillä taitotasoilla käytetään kuitenkin useammin vahvistussanoja *jätte* sekä *hel* adjektiivien *bra* ja *okej* kanssa, mikä osoittaa, että kyseisten adjektiivien käyttö on monipuolisempaa ja edistyneempää ylemmillä taitotasoilla kuin alemmilla taitotasoilla.

Muita usein käytettyjä adjektiiveja alemmilla taitotasoilla ovat *intresserad*, *god*, *viktig* ja *dålig* sekä ylemmillä taitotasoilla *olik*, *gammal*, *riktig*, *ny*, *intresserad*, *kul*, *normal* ja *rolig*. Mielenkiintoista on, että alemmilla taitotasoilla olevat lukiolaiset käyttävät enemmän tavallisia adjektiiveja kuin ylemmillä taitotasoilla olevat lukiolaiset. Tämä voi johtua siitä, että vertailussa käytetyn Allénin (1971) frekvenssilistan sanat on koottu sanomalehdistä. Tulokset osoittavat kuitenkin, että niin alemmilla kuin ylemmillä taitotasoilla eniten käytetyt adjektiivit ovat suhteellisen tavallisia.

Analyysissani olen tarkastellut myös muita adjektiiveja, joita käytetään niissä tilanteissa, joissa lukiolaiset ilmaisevat mielipiteitään tavasta, jolla sosiaalinen media kerää tietoa käyttäjistään. Tuloksista käy ilmi, että alemmilla taitotasoilla käytetään enemmän kontekstiin sopimattomia adjektiiveja kuin ylemmillä taitotasoilla. Alemmilla taitotasoilla olevat lukiolaiset eivät siis mahdollisesti tiedä, mitä merkityksiä kyseisillä

adjektiiveilla voidaan ilmaista. Tästä esimerkkinä on adjektiivin *god* käyttö mahdollisesti synonyymina adjektiiville *bra*. Ylemmillä taitotasolla adjektiivia *god* ei esiinny. Tulokset osoittavat myös, että lukiolaiset niin ylemmillä kuin alemmilla taitotasolla käyttävät vähäisissä määrin negatiivisia adjektiiveja yllä mainitussa kontekstissa.

Loppupäätelmät

Tutkimustulokset osoittavat, että Eurooppalaisen viitekehyksen taitotasojen sekä leksikaalisen kompleksisuuden välillä on yhteys, kun on kyse leksikaalisesta variaatiosta sekä sopivien adjektiivien käytöstä. Vaikka DIALUKI-skaalan kriteereissä ei suoraan mainita kuinka laaja tai syvä sanavarasto kielenoppijalla tulee olla eri taitotasolla, vaikuttaa näillä kahdella aspektilla olevan merkitystä, kun tekstejä arvioidaan. Myös leksikaalisella tiheydellä, adjektiivien sofistikoituneisuudella sekä adjektiivivariaatiolla on yhteys taitotasoihin, mutta näissä tapauksissa yhteys ei ole sellainen, kuin voisi odottaa.

Tulokset liittyen leksikaaliseen variaatioon osoittavat, että oppijan on tärkeää kasvattaa laaja sanavarasto. Tuloksista käy kuitenkin ilmi, että myös sanaston syvyydellä on merkitystä. Vaikka adjektiivien käyttö on monipuolisempaa alemmilla taitotasolla kuin ylemmillä taitotasolla, on alemmilla taitotasolla enemmän vaikeuksia valita sopivia adjektiiveja tilanteissa, joissa he ilmaisevat mielipiteitään tavasta, jolla sosiaalinen media kerää tietoa käyttäjistään. Kieltenoppijoiden tulisi siis kiinnittää huomiota myös siihen, että he oppivat mitä eri merkityksiä adjektiiveilla voi olla. Esimerkiksi koulussa oppilaan sanaston laajuutta sekä syvyyttä voidaan kehittää niin, että oppilaita rohkaistaan lukemaan enemmän erilaisia tekstejä. Toisaalta adjektiivien oppiminen lukemalla voi olla vaikeampaa kuin esimerkiksi substantiivien tai verbien oppiminen, sillä adjektiivit eivät ole yhtä kontekstisidonnaisia. Tämän vuoksi opettajan tulisi kiinnittää huomiota adjektiivien opettamiseen myös luokkahuoneessa.