

MISS 66

**Meddelanden från Institutionen för svenska språket
vid Göteborgs universitet**

Elisabeth Ohlsson

**Att synliggöra vokabulär
En kvantitativ studie
av gymnasieelevers skrivuppgifter på svenska
inom CLISS-projektets ram**

Januari 2018

MISS är en oregelbundet utkommande serie, som i enkel form möjliggör spridning av institutionens skriftliga produktion. Innehållet skall kunna vara skiftande: preliminära versioner av vetenskapliga artiklar, utredningar av allmänt intresse, recensioner, forskningsrapporter, licentiatuppsatser m.m. Bidragen är på ett naturligt sätt inriktade på ämnet nordiska språk inklusive svenska, men området får tas i vid mening. Varje enskild författare ansvarar för sitt bidrag. Synpunkter och förfrågningar mottas med tacksamhet.

ISSN 1102-4518

© Institutionen för svenska språket och författaren

Produktion:
Institutionen för svenska språket
Göteborgs universitet
Box 200
SE-405 30 GÖTEBORG

Reprocentralen, Humanisten
Göteborg

Abstract

Visualizing Vocabulary, an investigation of student assignments in CLIL and non-CLIL context

This licentiate thesis describes the investigation of students' texts written in L1, by CLIL and non-CLIL students participating in the CLISS-project during their three-year Upper Secondary School period, 2011-2014. The thesis focuses on the academic language proficiency in Swedish, where the practices of certain linguistic features characterizing academic texts are investigated.

The two groups of students involved in the study, a total number of 240, are students who have Swedish as the main language of instruction, the non-CLIL-group, and students who have English as the main language of instruction, the CLIL-group. Both groups follow the Swedish National Curriculum for Upper Secondary School (*Lgy11*, 2011) and, in learning Swedish, both groups have had mandatory Swedish teaching.

In this study, 520 texts were examined, written at four different times during the students' upper secondary education, the aim being to identify and study students' productive vocabulary in written assignments. Quantitative measures, such as text length, word length, words longer than six letters and word variation index were examined (see Magnusson & Johansson Kokkinakis 2009, Vagle 2005). Corpus linguistic methods were also applied to identify the usage of nominalizations, verbs in the passive tense and academic words, which are features of academic prose (Halliday & Martin 1993, Jansson *et al* 2012, Schleppegrell 2004). In addition, lexical profiling (Nation & Anthony 2016), identifying some of the most frequent words in Swedish, was carried out to visualize the general vocabulary and to explore to what degree the students' vocabulary contained the most frequent and common words or if, and to what extent, they used low-frequency words. The results are reported as means of the two student groups, CLIL and non-CLIL and also gender. The results were furthermore analyzed by using SPSS, version 22. The various parameters were used to explore the vocabulary use and to see if there were any differences or similarities in the vocabulary use between CLIL and non-CLIL students and male and female students.

The findings show that the results varied both within and between both groups. The boys in the CLIL group showed the highest means for several of the measured aspects but strong individuals were found in all four groups, i.e. males and females and CLIL and non-CLIL. The results also showed minor usage of features of academic prose. This indicates that explicit education concerning academic writing and the awareness of "the language of schooling" in L1, would be favorable for all students, regardless of language of instruction.

KEY WORDS: Swedish academic language proficiency, students' texts, word categories, nominalizations, verbs in passive tense, quantitative measures, CLIL and non-CLIL

Förord



I walk the maze of moments
But everywhere I turn to
Begins a new beginning
But never finds a finish
I walk to the horizon
And there I find another
It all seems so surprising
And then I find that I know

A Maze – Amazing!

Att vara doktorand och skriva en Licentiatuppsats har för mig varit som att ta sig igenom en labyrint; en fantastisk upplevelse med intressanta erfarenheter som har tagit sin tid, bara att få in tabeller och diagram i formatmallen samt hålla ordning på alla siffror gjorde dock att jag nästan gick vilse... Det har inneburit många återvändsgränder, många oväntade upptäckter, flera oväntade men angenäma möten och så till slut, en känsla av att hitta en lösning och att hitta ut!

Metaforen kommer av att jag har passerat en labyrint på min väg till och från skolan, där jag har arbetat under större delen av min doktorandtid. Som körsångare har jag ofta en melodislinga i huvudet och många gånger har det därför varit Enyas, se ovan.

A **maze** is designed to confuse you and challenge you to find your way to the exit or to the centre and then back out again. It is designed to test your problem-solving skills, your memory, and your tolerance for frustration!

Foto: privat/EO

Låttexten är hämtad från Enyas *Anywhere Is* © Sony/ATV Music Publishing LLC. Låtskrivare: Eithne Ní Bhraonáin/Nicky Ryan/Roma Ryan

Det är dags att sjunga tacksamhetens lov.



Jag framför min tacksamhet i kronologisk ordning från uppsatsidé till avslut, och börjar med professor Per. Om inte Du hade svarat och om inte Du hade övertalat mig att våga skriva en forskningsansökan (på en fredag) så hade detta inte hänt! Tack även professor Elisabeth Engdahl, då doktorandansvarig vid Institutionen för svenska språket, GU, som ringde och bjöd in mig till ett licentiatseminarium (Evas!), tack också för kloka och hoppfulla samtal. Det var så det hela började. Sedan krävdes ytterligare ett par ansökningar... så tack till Klas Eriksson, administrativt ansvarig för forskarskolan CUL, som alltid har varit behjälplig och svarat så vänligt på alla mina många frågor. Tack även till Jesper Boesen som var föreståndare för CUL när jag antogs, för bra möten och diskussioner när vi träffats.

Nu till CLISS-gänget! Tänk att bli antagen i ett forskningsprojekt med sju erfarna, duktiga seniora forskare och fyra glada doktorandkamrater, det är inte alla doktorander förunnat minsann. Så TACK; Liss-Kerstin Sylvén, projektledare, Britt-Marie Apelgren, Maria Lim Falk, Per Holmberg, Inger Lindberg, Sölve Ohlander, Tore Otterup, Eva Olsson, Helena Reierstam, Ylva Sandberg och BethAnne Yoximer Paulsrud. Så mycket roligt vi har haft, så lärorikt det har varit och så många goda stunder jag har fått uppleva tillsammans med er på olika platser världen runt. Tusen tack!

Mina handledare; utan er...

Sofie Johansson och Britt-Marie Apelgren, ni har varit fantastiska! Jag är *så* tacksam för allt stöd ni har gett, all tid och kunskap ni har bistått med, allt tålamod och all peppning och uppmuntran som jag fått! Mitt hjärta är fyllt av tacksamhet och jag vet att utan er hade jag inte kommit ut ur labyrinten. Ni finns mellan raderna!

Min doktorandkamrat Eva får en egen liten rad som ett extra tack för alla fina samtal och råd, på tåget, i mail, i telefonen och över en fika eller en lunch. Du har betytt mycket!

Ett stort tack till alla lärarna på skolorna vi har besökt samt alla elever som har deltagit i CLISS-projektet. Jag har blivit oerhört väl bemött både i klassrum och på personalrum. Trevliga samtal och trevliga möten, hela tiden, och mötena har varit många under materialinsamlingens tre år.

Tack alla roliga CUL-kamrater, ingen nämnd, ingen glömd. Tack all administrativ personal på Institutionen för svenska språket i Göteborg som oavsett vad, har funnits där med hjälp och råd. Ett särskilt tack till Sven Lindström på LT-gatan för textsättning av den slutliga uppsatsen med orubblig tro att "allting går att ordna" och tack till Tim och Anna Maria Hipkiss för engelsk språkhandledning.

Tack alla andra vänliga som jag har mött på LTS, på CUL-kurser och andra doktorandkurser, på IPS och på IDPP. Att vara distansdoktorand på halvtid har inte alltid varit enkelt, därav att jag är skyldig så många ett stort tack.

Tack alla fantastiska vänner, både fjärran och när, för ert oändliga tålamod, samtalsstöd, uppmuntran och vänskap. Så många ni är och så tacksam jag är!

Likaså tack till tidigare kollegor samt nuvarande och inte att förglömma, mina tålmodiga experimentglada elever. Gott att ni alla finns!

Jag är också mycket tacksam för de stipendier jag har erhållit under min doktorandtid, från Stiftelsen Paul och Marie Berghaus donationsfond, från Adlersbertska Stipendiestiftelsen samt från Filosofiska fakulteternas gemensamma donationsnämnd ur Stipendiefonden Viktor Rydbergs minne samt från LTS, Lärande, text och språk. Dessa stipendier har möjliggjort deltagande vid flera konferenser vilket har varit oerhört värdefullt och lärorikt, såväl att få delta som att få presentera min forskning.

Gällande brister och fel i uppsatsen så är dessa i de fall de finns uteslutande mina egna.

Slutligen min familj förstås; min kära härliga mamma Karin, min fantastiske datorfixande man Ecke, mina enastående underbara barn Kajsa, Gustav och Louise – vad vore jag utan er, ni är värdefullast av allt!

Skara, hösten 2017
Elisabeth Ohlsson

Innehåll

1. Inledning.....	1
1.1 Syfte och frågeställningar.....	4
1.2 Uppsatsens disposition och innehåll.....	5
1.3 Begreppsanvändning.....	5
2. Bakgrund.....	9
2.1 CLISS-projektet.....	9
2.1.1 Etiska överväganden.....	11
2.2 Språkintegrerad undervisning.....	12
2.3 Ämnesplaner och kunskapskrav i <i>Lgy11</i>	14
2.3.1 Argumenterande och utredande text.....	15
2.4 Begreppet <i>ord</i>	16
3. Material.....	17
3.1 Informanter.....	17
3.2 Materialinsamling i CLISS-projektet.....	19
3.3 Urval.....	20
4. Teoretisk bakgrund.....	21
4.1 Ett sociokulturellt perspektiv.....	21
4.2 Ett systemisk-funktionellt perspektiv.....	23
5. Tidigare forskning.....	27
5.1 Vardagsspråk och skolspråk.....	27
5.2 Ämnesspråk.....	30
5.2.1 Nominaliseringar.....	35

5.2.2	Verb i passiv form.....	37
5.2.3	Akademiskt språk, akademiska ord	37
5.3	Kvantitativa mätmetoder för textanalys	41
5.3.1	Kvantitativa mätmetoder och bedömningsaspekter	44
5.4	Korpuslingvistiska metoder för textanalys.....	48
6.	Metod.....	51
6.1	Kvantitativa ytstrukturella mätningar.....	53
6.2	Kvantitativa ämnesrelaterade mätningar	53
6.2.1	Elevtextkorpora.....	54
6.2.2	Ordlistor.....	54
6.2.3	Utveckling av specifika ordlistor.....	55
6.2.4	Akademiska ord	56
6.3	Lexikala profiler.....	57
6.4	Statistiska analyser.....	60
7.	Resultat	63
7.1	Resultatöversikt	64
7.1.1	Elevtexter.....	64
7.1.2	Skrivuppgifter och korpusar.....	65
7.2	Kvantitativa ytstrukturella mätningar.....	67
7.2.1	Textlängd	67
7.2.2	Ordmedellängd	75
7.2.3	Andel långord.....	79
7.2.4	Ordvariation.....	83
7.3	Kvantitativa ämnesrelaterade mätningar	86
7.3.1	Nominaliseringar	86
7.3.2	Verb i passiv form.....	89
7.3.3	Akademiska ord	92
7.4	Lexikala profiler.....	95
7.4.1	Täckningsgrad.....	96
8.	Diskussion.....	101
8.1	Metoddiskussion	102
8.2	Resultatdiskussion, jämförelse med tidigare studier	104
8.3	Resultatdiskussion, samvariation	115
8.4	Pedagogiska implikationer	119
8.5	Förslag till vidare forskning utifrån material och metod	122
8.6	Slutsatser och avslutande ord.....	124
9.	Referenser.....	127
10.	Bilagor.....	137
	Bilaga 1. Materialinsamlingen i CLISS-projektet	137
	Bilaga 2. Materialinsamling i CLISS-projektet, tidplan	138
	Bilaga 3. Argumenterande text, skrivuppgift 2.....	139
	Bilaga 4. Utredande text, skrivuppgift 3.....	140

Bilaga 5. Textlängd, individuella resultat.....	141
Bilaga 6. Textlängd, (F) och (P).....	142
Bilaga 7. Exempel på nominaliseringar i elevtexterna.....	143
Bilaga 8. Användningen av nominaliseringar.....	144
Bilaga 9. Exempel på verb i passiv form i elevtexterna.....	145
Bilaga 10. Användningen av verb i passiv form.....	146
Bilaga 11. Exempel på akademiska ord i elevtexterna.....	147
Bilaga 12. Användningen av akademiska ord.....	148
Bilaga 13. Resultat för lexikala frekvensprofiler.....	149
Bilaga 14. Förkortningar som används i uppsatsen.....	150

1. Inledning

Går det att ”se” ett ordförråd? Vilka typer av ord använder elever när de skriver, och hur kan deras skriftliga vokabulär se ut? I denna licentiatuppsats undersöks 520 elevtexter skrivna på svenska, insamlade under tre år inom *CLISS*-projektet, 2011-2014 (*Content and Language Integrated Learning in Swedish Schools*, se vidare i avsnitt 2.1 samt Sylvén & Ohlander, 2014). Undersökningen avser att besvara frågeställningarna ovan genom att försöka synliggöra gymnasieelevers skriftspråkliga val över tid, gällande deras svenska vokabulär.

För många elever är det självklart att man lär sig nya ord när man lär sig ytterligare språk som engelska, tyska, franska eller spanska, som är de språk elever vanligtvis möter som sina andra eller tredje språk i svensk skola. Att man behöver fylla på sitt svenska ordförråd är inte lika självklart som att ha glosläxor i engelska eller spanska. Det är heller inte självklart för alla elever att använda synonymer som gör att de varierar och förändrar sitt språk, inte heller att man arbetar med att gå från ett *receptivt* ordförråd till ett *produktivt* (Enström, 2013, Lindberg, 2007, Nation, 2013). Att eleverna medvetet arbetar med att i sina texter, faktiskt skriva de där nya orden som de har lärt sig och förstår, det kan ta sin tid. Skriftspråkskompetens utvecklas inte av sig själv utan kräver mycket tid och arbete, av såväl elev som av lärare (Knoch et al. 2014, Palmér & Östlund-Stjärnegårdh 2005).

Språk är nyckeln till all skolframgång och i skolan används ett alltmer specialiserat språk som skiljer sig från vardagsspråket (Lindberg 2007:15). Skolverket pekar i flera sammanhang på vikten av språkutveckling i skolan, ett språkutvecklande arbetssätt hos alla lärare, oavsett vilket ämne man undervisar i. Dessutom betonas värdet av att elever på alla stadier skall nå en hög språklig kompetens och behärska språket i olika verksamheter och ämnen genom exempelvis ett

gott ordförråd (se till exempel Skolverket 2012:25). I *Skolverkets lägesbedömning 2015* kan man läsa om betydelsen av att eleverna skall ”ges möjlighet att tillägna sig nya begrepp och erövra ett rikt språk” (ibid., 2015:41). I *Skolverkets lägesbedömning 2017* är motsvarande skrivelse att elever skall ”få goda möjligheter att utveckla kunskaper och värden” (ibid., 2017b:10). Förändringen kan ses kopplad till att svenska elevers läsförståelse har förbättrats något i de internationella mätningarna som gjorts. Dock brister likvärdigheten i svensk skola då resultaten ligger på en lägre nivå än när mätningarna började genomföras samt att elevernas socioekonomiska bakgrund i hög grad påverkar deras resultat (Skolverket 2017b:8).

Forskning visar att ordförrådet ”anses vara den enskilt viktigaste faktorn för att man framgångsrikt skall kunna tillägna sig kunskaper i skolans ämnesundervisning” (Lindberg 2006:66). I den omarbetade versionen av Skolverkets *Läsluft* (2017a) har i modulen ”Från vardagsspråk till ämnesspråk” en egen del tillkommit som heter ”Ordförrådet – en framgångsfaktor” (Lindberg, Skolverket 2017a, del 5). Enström (2013a, 2013b) beskriver det som att; ”Ju fler ord en individ förfogar över, desto bättre är han eller hon rustad för språkligt krävande situationer” (ibid., 2013a:169). Det är naturligtvis inte bara i svensk forskning som betydelsen av ett rikt ordförråd lyfts fram; ”Learners with big vocabularies are more proficient in a wide range of language skills than learners with smaller vocabularies.” (Meara 1996:37).

Engelska är lätt, men skriva på svenska är svårt, var titeln på min forskningsansökan till CLISS-projektet sommaren 2011. Jag antogs sedan på halvtid som licentiat i humaniora med inriktning mot utbildningsvetenskap associerad till CUL-forskarskolan, Centrum för utbildningsvetenskap och lärarforskning.¹ Detta gör att uppsatsen ingår i en kontext där skolutveckling och lärande skall ”vila på vetenskaplig grund, beprövad erfarenhet och evidens”.²

Som undervisande lärare i svenska och engelska på högstadiet har jag funderat mycket över elevers inställning till skrivande och skriftspråkliga krav. Forskningsansökans titel kom från mina egna elever, framför allt från pojkar och framför allt sedan IB-programmet tillkommit på vår gymnasieskola.³ Hur får man elever att utveckla sitt ordförråd på svenska och därmed skriva ”bra” texter och vad menas med en ”bra” text? Vad gör lärare för att elever skall utveckla sitt ordförråd och därmed sitt skrivande och vad gör lärare för att arbeta med skrivandet i svenska när engelskan för en del elever tycks vara ”lättare” och mer intressant? Dessa frågor har förmodligen flera lärare som läser och bedömer elevtexter, formativt och summativt, ställt sig, oavsett ämne. Likaså vem som bär ansvaret för att arbeta med elevernas språk- och skrivutveckling; lärarna i

¹ Se vidare <http://cul.gu.se/CUL+forskarskola>

² Skollagen 1 kap 5 §

³ International Baccalaureate, mer i avsnitt 2.2

svenska eller i svenska som andraspråk eller ämneslärarna eller alla tillsammans, och i så fall, hur (Berge 2002, Kouns 2010).

En väg till skolframgång är att alla elever ges möjligheten att *förstå* texter; att de kan skapa egna texter genom att förstå hur olika texter konstrueras, att de kan se textmönster som de sedan kan använda, att elever ges en textrepertoar. Att dessutom arbeta med det ämnesrelaterade språk som förekommer i ämnes-texter och som kan skapa inbäddade svårigheter då texten kan ”osynliggöra” deltagare och budskap genom grammatiska konstruktioner är lika viktigt. Skillnaden mellan en vardaglig vokabulär och den som benämns som skolspråk och ämnesspråklig blir angelägen att tydliggöra och införliva som en vital kompetens hos alla elever (se exempelvis Schleppegrell 2004, Lindberg 2006, Lindberg & Johansson Kokkinakis 2007, Skolverket 2008).

I föreliggande studie undersöks den svenska vokabulären i 520 texter som elever i de åtta deltagande klasserna i CLISS-projektet (se avsnitt 2.1) har skrivit vid fyra tillfällen under sin treåriga gymnasietid. Eleverna var den första årsgruppen som följde en ny läroplan för gymnasieskolan, *Läroplan, examensmål och gymnasiegemensamma mål för gymnasieskola 2011* (Skolverket 2011a, se avsnitt 2.3).

Genom att kombinera mer traditionella textlingvistiska metoder med korpuslingvistiska metoder samt att undersöka elevtexterna med lexikal profilering, har olika aspekter av elevernas vokabulär identifierats, synliggjorts och studerats. Då elevtexterna som undersöks är skrivna på svenska i en svensk skolkontext görs jämförelser av resultaten i huvudsak med tidigare svensk forskning.

Valet av textlingvistiska mätmetoder har gjorts med hänsyftning till att längre ord kan sägas vara mer innehållsrika än kortare ord och kräver mer kunskaper liksom förmågan att kunna skriva varierat och en längre text (Vagle 2005). Liknande underökningar har gjorts i tidigare nordiska studier vilket möjliggör jämförelser av dessa mått över tid (se exempelvis Ciolek Laerum 2009, Hultman & Westman 1977, Nyström 2000, Olevard 1999, Persson 2016, Ribeck 2015, Vagle 2005, Östlund-Stjärnegårdh 2002).

Korpuslingvistiska metoder har använts för att identifiera och studera de deltagande elevernas skriftspråkliga val genom att skapa ordlistor av alla texter där ordkategorier, som utifrån en systemisk-funktionell språksyn kännetecknar ett ämnesrelaterat, vetenskapligt och akademiskt språk, har valts ut. Ordkategorierna som undersöks i studien är nominaliseringar, verb i passiv form och akademiska ord (se Biber 1998, Christie & Beverly 2008, Halliday 1993, Johansson Kokkinakis & Frändberg 2013, Lindberg 2006, 2007, Magnusson 2008, 2011, Nation 2013, Nygård Larsson 2011, Persson 2016, Ribeck 2015, Schleppegrell 2004).

Lexikala profiler har använts för att studera spridning och förekomst av några av de mest frekventa orden i det svenska språket och därmed synliggöra elevernas skriftspråkliga val i relation till vokabulär (för jämförelse av motsvarande

undersökningar av de engelska uppsatserna i CLISS-projektet, se Olsson 2016, Sylvén & Ohlander 2014).

Dessutom undersöks om elevernas svenska skriftspråkliga vokabulär och de resultat studien visar avseende de undersökta skrivuppgifterna, korrelerar med elevernas programval, CLIL/icke-CLIL (se avsnitt 2.1 och 2.2), genus samt kursbetyg i svenska. Till de statistiska analyserna har SPSS, version 22, använts (*Statistical Package for the Social Sciences*).⁴

Studiens avgränsningar medför att den inte gör anspråk på att förklara de skriftspråkliga val som synliggörs i elevtexterna då undervisningssituationer, lärares planeringar eller explicita undervisning inte ingår i studien. Studien fokuserar inte heller på skrivprocesser, elevernas språkliga bakgrunder eller deras kognitiva förutsättningar och, eller förmågor.

Licentiatuppsatsen kan ses utgöra ett bidrag inom textforskning med fokus på gymnasieelevers produktiva, skriftliga färdigheter. Min förhoppning är att studien kan tillföra den didaktiska praktiken ett kunskapsfält där kvantitativa forskningsmetoder inte ofta används men här bidrar till att synliggöra en skolrelaterad, skriftspråklig, svensk vokabulär hos elever i en flerspråkig gymnasiekontext samt en jämförande svensk sådan, över tid. Det handlar om att försöka bidra till en förståelse för ”det rika språket och det goda ordförrådet”, hur det kan se ut, hur det kan utvecklas och vilka bakomliggande kunskaper som kan förmedlas, såväl till elever som till lärare och lärarutbildare för att på så vis också nå en ökad likvärdighet i skolan.

1.1 Syfte och frågeställningar

Syftet med licentiatuppsatsen är att definiera och identifiera, samt undersöka och beskriva, de deltagande gymnasieelevernas skriftliga svenska vokabulär i de inom CLISS-projektet insamlade skrivuppgifterna under tre år.

Licentiatuppsatsens frågeställningar är:

- Hur framträder en vokabulär i skrift som är skolrelaterad, ämnesrelaterad och akademisk, i gymnasieelevers texter på svenska?
- Finns det samband i den vokabulär som framträder som kan relateras till programval, CLIL respektive icke-CLIL, och genus i respektive grupp, och hur ser de i så fall ut?

⁴ <http://www-01.ibm.com/software/se/analytics/spss/>

1.2 Uppsatsens disposition och innehåll

Efter inledning, syfte och frågeställningar samt uppsatsens disposition kommer ett avsnitt om begreppsanvändning.

Kapitel 2 består av bakgrunden där CLISS-projektet presenteras och de etiska överväganden som har gjorts redovisas. Därefter följer en kortare beskrivning av språkinTEGRERAD undervisning i Sverige och internationellt, utifrån benämningarna *SPRINT* och *CLIL*. Vidare redogörs för gällande läroplan samt ämnes- och kursplaner liksom de texttyper som ingår i undersökningen och definitioner av ordbegreppet.

Kapitel 3 inleds med en redogörelse av informanterna, sedan behandlas materialet. Här beskrivs först hela CLISS-projektets materialinsamling därpå de fyra skrivuppgifterna vilka utgör analysmaterialet för denna licentiatuppsats samt urvalet av dessa.

Kapitel 4 beskriver studiens teoretiska bakgrund och i kapitel 5 ges en forskningsöversikt av tidigare forskning för att placera studien i ett forskningsfält.

I kapitel 6, redogörs för studiens metoder och de tillvägagångssätt som har använts för att försöka besvara uppsatsens frågeställningar om hur gymnasieelevers skriftliga vokabulär kan synliggöras och hur den framträder över tid.

Kapitel 7 omfattar resultaten av studien där en översikt ges inledningsvis och därefter presenteras de delundersökningar som ingår i studien var för sig.

Slutligen förs i kapitel 8, en metoddiskussion. Därefter diskuteras resultaten som har framkommit och jämförelser görs med resultat från tidigare forskning. Här ges även förslag till vidare forskning utifrån såväl metod som utifrån material liksom pedagogiska implikationer som studien ger.

Sist återfinns referenser, bilagor, tabell- och figurförteckning samt en lista med förklaringar till förkortningar som används i uppsatsen. Begrepp skrivs i uppsatsen initialt med kursiv stil därefter med vanlig stil.

1.3 Begreppsanvändning

I licentiatuppsatsen används begreppen *CLIL* och *icke-CLIL* om elevgrupperna som har deltagit i projektet. Benämningarna förklaras mer ingående i avsnitt 2.2 men avser programvalen där elevgruppen *CLIL* (*Content and Language Integrated Learning*) läser alla ämnen utom språk (svenska, engelska och eventuellt tillvalda språk) på engelska och *icke-CLIL*-eleverna läser sina program på svenska.

Begreppet *skrivuppgift* används om de fyra uppgifterna som har utformats inom projektet. Skrivtillfällena för dessa fyra uppgifter har varit under gymnasieelevernas terminer 1, 2, 4 och 5. Med *elevtexter/texter* menas de insamlade texterna som eleverna har skrivit utifrån de fyra skrivuppgifterna. Begreppet *uppsats* används med hänvisning till de nationella proven i svenska och engelska då det

också kan ses som en egen etablerad genre eller texttyp (Berge 1988, Nyström 2000:5).

Genom olika analysmetoder undersöks *vokabulären/ordförrådet*, som framträder i elevernas texter. Vokabulären delas in i *ordkategorier*, olika grupper av ord, här; *nominaliseringar*, *verb i passiv form/passiver/passiva verb* och *akademiska ord*, vars förekomst och spridning undersöks på individnivå, på gruppnivå utifrån programval CLIL och icke-CLIL (se avsnitt 2.2) och utifrån genus; *flickor* och *pojkar*. Gymnasieelever kan förstås ses som unga kvinnor och unga män, vilket i undersökningar avseende engelska anges som *Females* (F) respektive *Males* (M), (se Olsson 2016, Sylvén & Ohlander 2014 m.fl.). Här följer studien uppsatser som skrivits på svenska där förkortningarna (F) och (P) används för grupperna flickor respektive pojkar. Även förkortningarna FC (=flickor CLIL), PC (=pojkar CLIL), FnC (= flickor non/icke-CLIL) och PnC (= pojkar non/icke-CLIL) används för de olika grupperna

Förklaringar och avgränsningar för *ordkategorierna* (nominaliseringar, verb i passiv form, akademiska ord) redogörs i respektive avsnitt i kapitel 5.

Då det är elevers skriftliga vokabulär som undersöks i de insamlade texterna är det den *produktiva* vokabulären som fokuseras, det vill säga att man kan skriva och använda ord på ett korrekt sätt (se Gustafsson & Håkansson 2017, Nation 2013:9 ff.). Inom projektet har tester genomförs för att undersöka de deltagande elevernas *receptiva* ordförråd, att kunna känna igen ords form och betydelse (se Sylvén & Ohlander 2014, Lindberg & Johansson u.u.). I föreliggande studie görs alltså inga jämförelser mellan elevernas produktiva och receptiva kunskaper vilket görs i vissa andra studier (se vidare Enström, 2013:175–176, Lindberg, 2007:41–43, Nation, 2013:48 ff., Qian & Schedl 2004, Olsson, 2016:110).

Olika begrepp för *ord* förekommer i forskningsfältet, till exempel *textord*, vilket innebär bokstavstecken tillsammans med en avgränsning före och efter av mellanslag eller skiljetecken. Andra begrepp för textord är *graford*, *löpord* och *token* (eng.). Om samma ord förekommer flera gånger används begreppet *lexikonord*, *unika ord*, *typord* eller *type* (eng.), för ordets unika förekomst. Med *lemma* menas ett ord i grundform som skiljer sig åt i böjningsform och eller ordklass (Hultman, 2013:30). Beroende på vilken tidigare studie som refereras till kan dessa olika ordbegrepp variera. I föreliggande studie används begreppen *graford* och *lexikonord* i de två korpusar som har skapats (se tabell 10 i avsnitt 7.1.2).

Begreppet *textuella måttenheter* omfattar i den här studien de kvantifierbara enheter som finns i elevtexterna som undersöks: textlängd, ordmedellängd, andel långord och ordvariation, *OVIX*. Med *högrekventa ord* menas de allmänna ord har en hög *frekvens*, det vill säga förekomst, i det svenska språket (här utifrån en jämförande korpus, *Bloggmix* 2012) Med *lägrekventa ord* menas motsatsen, dessa ord förekommer mer sällan. Dessa ord synliggörs i de *lexikala profiler* som redovisas i avsnitt 7.4.

Begreppen *vardagspråk*, *skolspråk* och *ämnesspråk* diskuteras och definieras utifrån tidigare forskning i egna avsnitt (se kapitel 5).

I studier av exempelvis läromedeltext eller provtexter som i de internationella mätningarna PISA och TIMMS⁵, görs skillnad på textanalys på *ytnivå* respektive på *djupare* nivå. Där avses ytnivå omfatta undersökningar med textuella måttenheter som textlängd, ordlängd, meningslängd, läsbarhetsindex och ordvariation på grund av att ingen lingvistisk eller språkberoende grammatisk analys görs (Heimann Mühlenbock 2013, Persson 2016, Ribeck 2015).

Undersökningar av språket på djupare nivå kan innefatta *syntaktiska analyser* av till exempel ordklasser, verbens tempusformer, satsindelning och grammatisk satsanalys. Analysen kan även innefatta förhållanden däremellan liksom aspekter på språkliga register (se Skolverket 2008) eller själva skrivprocessen (se Andersson Varga 2014, Randahl 2014). För det behövs mer kunskaper *om* språket och det språkliga innehållet, det vill säga en språkberoende analys. Detta görs alltså inte i föreliggande studie. Undersökningar på ytnivå förutsätter inte samma kunskaper i och om språket som undersöks (Persson 2016:18, Ribeck 2015).

I uppsatsen används begreppet *kvantitativ textanalys* gällande de textuella måttenheterna, de *textlingvistiska variablerna*, som undersöks och analyseras. Begreppet *kvantitativ ämnesrelaterad textanalys* används om ordkategorierna, de *grammatiska variablerna*, som undersöks och analyseras (se kapitel 6).

Med uttrycket *skolrelaterad* avses här att texterna är skrivna i en skolkontext där det krävs, såväl en ändamålsenlig vokabulär av ord och fraser, liksom kunskaper om grundläggande grammatiska konstruktioner samt kunskaper om hur man uttrycker sig i texter förknippade med olika skolämnen (Lindberg 2006:89). Studiens undersökningar av en skolrelaterad vokabulär utgår från de 520 texterna, skrivna på svenska av studiens gymnasieelever, insamlade i det longitudinella projektet, i en skolkontext, där dessa krav förutsätts finnas.

Begreppet *validitet* används i uppsatsen med betydelsen tillförlitlighet och *reliabilitet* med betydelsen överförbarhet (se Cohen et al. 2011, Erickson 2016).

I uppsatsen refereras ibland tidigare forskning med direktcit. Då dessa är på engelska skrivs de på originalspråket utan översättning. Vid hänvisning till begrepp på engelska har i vissa fall dessa översatts med kommentar inom parentes.

⁵ PISA = *Programme for International Student Assessment*, en internationell kunskapsstudie av 15-åringars kunskaper i matematik, naturvetenskap och läsförståelse, genomförs vart tredje år.

<https://www.skolverket.se/statistik-och-utvardering/internationella-studier/pisa>

TIMMS = *Trends in International Mathematics and Science Study*, en internationell studie som undersöker fjärde- och åttondeklassares kunskaper i matematik och naturvetenskap, genomförs vart fjärde år.

<https://www.skolverket.se/statistik-och-utvardering/internationella-studier/timss>

2. Bakgrund

I detta kapitel beskrivs först projektet som uppsatsens studie utgår ifrån, därefter tydliggörs begreppet *språkintegrerad undervisning* (avsnitt 2.2). Sedan följer en beskrivning av rådande ämnes- och kursplaner i läroplanen *Lgy 11*, liksom av de texttyper som elevtexterna är skrivna utifrån (avsnitt 2.3, 2.3.1). Slutligen följer definitioner av begreppet *ord* då studien fokuserar på elevers produktiva ordförord i de insamlade texterna (avsnitt 2.4).

2.1 CLISS-projektet

Licentiatuppsatsen beskriver en studie som ingår i ett projekt kallat *CLISS*. CLISS står för *Content and Language Integrated Learning in Swedish Schools* och är ett longitudinellt projekt finansierat av Vetenskapsrådet 2011 – 2014.⁶ CLISS-projektets huvudfokus ligger på hur gymnasieelevers produktiva och receptiva färdigheter i skolrelaterat, akademiskt skriftspråk på svenska och engelska utvecklas i en flerspråkig skolkontext (se Sylvén & Ohlander, 2014).

CLISS-projektets övergripande forskningsfrågor är:

- Hur utvecklas produktiva och receptiva färdigheter i skolrelaterat, akademiskt skriftspråk på svenska och engelska?
- Hur väl behärskas ämnesspecifika begrepp och termer i historia respektive biologi på svenska och engelska?

⁶ <http://ips.gu.se/forskning/forskningsprojekt/cliss>

Vetenskapsrådet, projektnummer: 721-2010-5376

- I vad mån finns det skillnader mellan elever med svenska som modersmål och elever med annat modersmål?
- I vad mån finns det genusskillnader?

Projektet har designats och ansökts om av sju seniora forskare från Göteborgs och Stockholms Universitet. Fem doktorander har varit knutna till projektet och flera olika forskningsperspektiv av språk- och ämnesintegrerad undervisning i svensk gymnasieskola har därmed belysts i olika studier av projektdeltagarna (Apelgren, 2014, Apelgren u.u. 2018, Lim Falk, 2015, Lim Falk & Holmberg, 2016, Lindberg & Johansson u.u., Olsson, 2015, Olsson, 2016, Reijerstam, 2015, Sylvén, 2013, Sylvén, 2015, Sylvén & Ohlander, 2014a, Sylvén & Ohlander, 2014b, Sylvén & Olsson, 2015, Sylvén & Thompson, 2015, Yoxsimer Paulsrud, 2014).

I CLISS- projektet har åtta gymnasieklasser på tre olika skolor i Sverige deltagit, totalt har cirka 240 elever ingått i projektet. Deltagarsiffran har varierat något över tid då elever har bytt klasser och gått över till program som inte har ingått i studien eller tagit studieuppehåll samt att några elever har tillkommit. Avvikelserna gäller dock ett fåtal elever. Fem av de åtta deltagande klasserna har varit CLIL-klasser (se vidare i avsnitt 2.2) som har läst sina gymnasieprogram på engelska. Tre av klasserna har läst sina gymnasieprogram på svenska och är så kallade kontrollklasser. Mer om informanterna i avsnitt 3.1. Eleverna i studien tillhör den första elevkullen som följt den nya svenska läroplanen *Lgy11* (*Läroplan, examensmål och gymnasiegemensamma ämnen för gymnasieskola 2011*, Skolverket 2011a, fortsättningsvis refererad till som *Lgy11*) under hela sin gymnasietid. Det har varit en ny läroplan att möta och implementera även för de undervisande lärarna.

Test- och skrivuppgifter samt enkäter som ingår i CLISS-projektet, har konstruerats av projektets seniora forskare. Vid konstruktionen av de olika testerna och skrivuppgifterna har den nya läroplanen beaktats och skrivuppgifterna, på såväl svenska som engelska, följer de nya kunskapskraven i respektive ämne. Uppgifterna utgår även från kursplanerna för historia, samhällskunskap, biologi och naturkunskap i *Lgy11* då dessa ämnen är gymnasiegemensamma och ingår i forskningsprojektets övergripande frågor med fokus på hur väl eleverna behärskar ämnesspecifika begrepp i dessa ämnen då de skriver text på svenska respektive på engelska.

I såväl uppgifternas utformning som i insamlingen av all data har aspekter på validitet och reliabilitet tagits hänsyn till genom utformningen av tester och skrivuppgifternas konstruktion samt tillvägagångssättet vid genomförandet av dessa för att säkerställa likvärdighet (Crooks et al. 1996, Cohen et al. 2000, Erickson 2010, Erickson 2016).

I materialet som finns att tillgå och som har analyserats på olika sätt av projektgruppen kan nämnas bakgrunds- och motivationsenkäter, allmän läsförståelse och ordkunskap på svenska och engelska, genomförda termin 1 och 5, akademisk läsförståelse och ordkunskap på svenska och engelska, genomförda termin 3 och 6, skriv-

uppgifter av argumenterande och utredande texttyper på svenska och engelska, genomförda termin 1, 2, 4 och 5 (se avsnitt 3.2 samt bilaga 1 och 2).

2.1.1 Etiska överväganden

CLISS-projektet hade redan vid starten 2011 ansökt om Vetenskapsrådets tillstånd och prövats för de fyra allmänna huvudkraven på forskning; informationskrav, samtyckeskrav, konfidentialitetskrav och nyttjandekrav.⁷

Eftersom eleverna har fyllt 15 år och projektstudien inte har innefattat något som kan hänföras till etisk känslig karaktär har elevernas vårdnadshavare informerats, men samtycke av dessa har inte krävts utan eleverna har själva fattat beslutet om att delta i studien eller inte. Gymnasieeleverna som har deltagit i CLISS-projektet har alla fått fylla i en medgivandeblankett om sitt deltagande i projektet med information om Vetenskapsrådets etiska regler samt information om att de när som helst under projektets gång själva har kunnat välja att inte delta.⁸

Några få i varje klass valde redan från början att inte delta men har haft möjlighet att ändra sig under sin gymnasietid då projektet har pågått under deras tre gymnasieår. Även de elever som har tillkommit genom att byta klass eller har kommit tillbaka från studieuppehåll, har fått fylla i en blankett för ett medgivande om att delta i projektet eller om de har valt att avstå.

De olika insamlingsuppgifterna i projektet har inte ingått i skolornas ämneskurser och därmed inte varit betygsgrundande vilket gör att det inte har förelegat någon form av beroendeställning hos någon part. Däremot har möjligheten funnits för både lärare och elever att via projektledaren få tillgång till de skrivna texterna.

Skolorna som deltagit har haft lite olika förhållningssätt till sitt deltagande. Två av skolorna har på sin hemsida offentligt berättat att projektet har pågått på just deras skola och där gett lite information om vad CLISS-projektet har inneburit. Den tredje skolan har valt att inte ge motsvarande information offentligt.

Projektmedlemmarna har besökt skolorna regelbundet för materialinsamling men även minst en gång under projekttiden genomfört ett gemensamt besök och haft utbyte av information med skolans deltagande personal. Kontakten med alla tre skolorna har varit mycket god och även efter att projektet har avslutats och eleverna har slutat sina gymnasieprogram finns fortlöpande kontakt mellan projektgruppen och representanter för skolorna.

⁷ Vetenskapsrådets etiska regler <http://www.codex.vr.se/texts/HSFR.pdf>.

⁸ CLISS-projektet hade innan starten 2011 bedömts och godkänts av Vetenskapsrådets etiska kommitté.

Under projektets gång har projektgruppen ansvarat för insamlingen av materialet som finns i en databas med kodnyckel som endast projektmedlemmarna har tillgång till (se vidare avsnitt 3.2). Eleverna har själva inte vetat om sin kod utan de har vid varje uppgiftstillfälle skrivit sina namn på uppgifterna som har skrivits för hand alternativt på dator. Efter varje insamlingstillfälle har en projektmedlem kodat allt insamlat material och lagt in materialet i CLISS-databasen. Vi har kommit överens om vem som åkt till vilken skola och när. Vid enstaka tillfällen har någon av skolornas lärare genomfört insamlingen av material om ingen i projektgruppen har haft möjlighet att åka. Då har det funnits kommunikation mellan projektmedlemmar och ansvarig lärare för att säkerställa inramningen för uppgiftsgenomförandet och insamlingen av material. Att projektets deltagare har ansvarat för materialinsamlingen har varit ett sätt att garantera validitet och reliabilitet genom att vi haft gemensamma instruktioner och förhållningssätt vid insamling av data (Crooks et al. 1996). Insamlingen av respektive uppgift har skett vid ett tillfälle på respektive skola. Vi har inte gett möjlighet för frånvarande elever att genomföra de olika testen och skrivuppgifterna vid andra tillfällen med hänvisning till projektets validitet. Detta har i viss mån påverkat materialomfånget (se avsnitt 7.1.1).

För allt material ansvarar projektledare professor Liss Kerstin Sylvén, Institutionen för pedagogik och specialpedagogik vid Göteborgs universitet. Mot denna bakgrund kan de fyra allmänna huvudkraven på forskning anses vara uppfyllda.

2.2 Språkintegrerad undervisning

I Sverige är språk- och ämnesintegrerad undervisning (förkortat *SPINT* eller *SPRINT*) relativt utbrett både på kommunala skolor och i friskolor. År 2012 fanns det *SPRINT*-undervisning på omkring 26 % av alla svenska gymnasieskolor (Yoxsimer Paulsrud, 2014:62). En del skolor ser det som ett sätt att locka elever, andra som ett sätt att vara en del av ett globalt och internationellt sammanhang. I internationell kontext är benämningen *CLIL*, *Content and Language Integrated Learning*. *CLIL*-skolor, följer landets styrdokument för undervisning, det vill säga i Sverige den svenska läroplanen, när denna uppsats skrivs, *Lgy11*, med tillhörande kursplaner och kunskapskrav (se avsnitt 2.3). Däremot kan den exakta definitionen av begreppet skilja sig åt mellan länder liksom praxis för hur utbildningen erbjuds och vad den innehåller (se vidare Yoxsimer Paulsrud 2014). För *CLISS*-projektet var det viktigt att det fanns såväl språkklasser, vilka läste sina gymnasieprogram på engelska, här fortsättningsvis benämnda som *CLIL*-klasser, liksom deltagande kontrollklasser, vilka läste sin gymnasieutbildning på svenska.

Begreppet *CLIL* formulerades av David Marsh vid Universitetet i Jyväskylä i Finland (1994): "CLIL refers to situations where subjects, or parts of subjects,

are taught through a foreign language with dual-focused aims, namely the learning of content and the simultaneous learning of a foreign language."

Oftast erbjuds gymnasieprogram på engelska, det vill säga att elever kan läsa en programinriktning med ämnesundervisningen på engelska förutom undervisningen i modersmålet, i denna studie, i ämnet svenska där undervisningen sker på svenska (för en del elever svenska som andraspråk/SvA). Det finns dock skolor i Sverige som erbjuder undervisning på exempelvis tyska, franska eller finska. Det förekommer att skolor har en så kallad internationell inriktning där hela skolan från förskoleklass till årskurs nio och – eller även gymnasiet, erbjuds skolgång med språk- och ämnesintegrerad undervisning. Den ursprungliga tanken med denna form av undervisning är att främja inlärandet av ett annat språk än elevernas modersmål. Man kombinerar på så vis ämnes- och språkstudier. (Nixon 2001, Lim Falk 2008⁹, Sylvén 2004).

I Sverige finns även gymnasieskolor med så kallade IB-program, *International Baccalaureate*. I vissa av dessa skolor erbjuds språkutbildning från tidiga åldrar på grundskolenivå (PYP – Primary Years Programme och MYP – Middle Years Programme).¹⁰ IB har till skillnad från CLIL gemensamma internationella styrdokument, läroplaner och kursplaner, som gäller för alla IB-skolor världen över, där CLIL/SPRINT alltså följer en av landet, där utbildningen sker, fastställd läroplan.

Inom Europarådets samarbete har man på olika sätt sökt finna former för att öka europamedborgarnas språkkompetenser. Ett förslag som framfördes 2002 var att "Content and Language Integrated Learning – CLIL, can provide an answer".¹¹ Trots att Europakommissionen förordar att Europas medborgare bör kunna två språk utöver modersmålet och rekommenderar flerspråkighet, dominerar engelskan i utbildningssystemen vilket har lett till att CLIL av vissa benämns *CEIL – Content and English Integrated Learning* (Dalton-Puffer, 2011:183). Få utbildningssystem i Europa har integrerat CLIL i styrdokument, läroplaner eller andra policydokument. Undantag härvidlag är för närvarande Spanien och Nederländerna där flera utvecklingsprojekt har genomförts liksom stöttning för lärare och införandet av nationella kvalitetsaspekter (Dalton-Puffer 2011, Yoxsimer Paulsrud 2014). Inget liknande finns ännu i Sverige trots att nästan en fjärdedel av svenska gymnasieskolor erbjuder den här typen av språkintegrerad undervisning. Hur detta kan påverka undervisningen och därmed likvärdigheten har undersökts i en studie där CLIL-undervisning jämförs i fyra europeiska länder (Sylvén 2013).

⁹ Lim Falk använder förkortningen SPRINT = Språkintegrerad undervisning, i sin avhandling medan begreppet CLIL = Content and Language Integrated Learning används i denna.

¹⁰ <http://www.ibo.org/>

¹¹ http://www.edu.xunta.es/centros/cpicruce/system/files/clilbroch_en.pdf.

2.3 Ämnesplaner och kunskapskrav i *Lgy11*

Textinstruktionerna som eleverna har fått vid de fyra skrivtillfällena för de skrivuppgifter på svenska som ingår i studien, utgår från kunskapskraven för svenskämnet i *Lgy11* (Skolverket 2011a). De har även en utgångspunkt i inriktning mot de samhälls- respektive naturorienterade ämnesplanerna för att belysa hur eleverna i projektet behärskas ämnesspecifika begrepp och termer i historia, samhällskunskap, biologi och naturkunskap. Skrivuppgifterna har formulerats som *utredande* och *argumenterande* text då dessa texttyper explicit finns med i kunskapskraven för svenskämnet på gymnasiet (Skolverket, *Lgy11*:169, här se bilaga 3 och 4). Samtliga elever som har deltagit i studien har läst kurserna Svenska 1, 2 och 3 alternativt Svenska som andraspråk 1, 2 och 3. Denna studie omfattar alla deltagande elevers skrivuppgifter och skiljer inte på elevernas kursval.

I den läroplan som gällde före 2011, *Läroplan för de frivilliga skolformerna*, förkortad *Lpf 94*, formulerades i kursen Svenska 1 inte vilka texttyper eleverna skulle behärska. I ”*Mål att uppnå*” gällde då att eleverna skulle ”kunna använda skrivandet som ett medel för att utveckla tänkande och lärande”. För betyget *Godkänd/G* i *Lpf 94* gällande skrivande stod i kriterierna: ”eleven formulerar sig sammanhängande i tal och skrift i både formella och informella sammanhang med hänsyn till mottagaren”, för betyget *Väl godkänd/VG*; ”Eleven skriver i huvudsak korrekt såväl berättande som utredande texter där tankegången klart framgår” och för betyget *Mycket väl godkänd/MVG*; ”Eleven använder i såväl tal som skrift stilmedel som är effektfulla och slående och anpassar sitt budskap så att det både berör och förstås av mottagarna”.¹²

I *Lgy11* är ämnet svenska fortfarande gymnasiegemensamt, det vill säga att samma ämnesplan och kursplaner gäller för alla programval. Inledningen formulerar ämnets *syfte*, sedan följer ett *centralt innehåll* för varje delkurs (för svenskan finns Svenska 1, 2 och 3 samt tillvalskurser i Litteratur, Retorik och Skrivande). Därefter följer *kunskapskrav* för de olika betygen E, C och A (för betygen D och B gäller att kunskapskraven för det nästkommande betyget uppfylls till övervägande del).¹³

I det centrala innehållet för delkursen Svenska 1 står att ämnet, gällande skrivande, skall innehålla: ”Argumentationsteknik och skriftlig framställning av argumenterande text” (Skolverket 2011a:162). För delkursen Svenska 2 står

¹² Se Skolverkets hemsida för tidigare läroplaner

<https://www.skolverket.se/laroplaner-amen-och-kurser/gymnasieutbildning/gymnasieskola/kursplaner-fore-2011>

¹³ Observera att betyget F finns men innebär att eleven enligt läroplanen, *Lgy11*, ej har nått kunskapskraven. I den tidigare läroplanen *Lpf94* fanns betyget IG = icke godkänd som på grundskolan motsvarades av EUM = ej uppnått målen. Min anmärkning.

motsvarande: ” Skriftlig framställning av utredande och argumenterande texter. Normer och stildrag som hör till dessa texttyper” (Skolverket 2011a:169).

2.3.1 Argumenterande och utredande text

I svenskt skolskrivande har den berättande texten, *narrativen*, dominerat, men i de nya läroplanerna för både grundskolan, *Lgr11* och för gymnasieskolan, *Lgy11*, har kraven på explicita texttyper tillkommit (Skolverket, 2011b, 2011a).

Såväl den *argumenterande texten* som den *utredande texten* kan sägas tillhöra den analytiska genren med den engelska benämningen *Exposition* (se vidare Schleppegrell 2004:85, Axelsson et al. 2006:9). Utgångspunkten finns i den funktionella grammatikens genrebegrepp där en viss genre följer en viss struktur och ett speciellt socialt syfte (Gibbons 2009, Schleppegrell 2004, Sellgren 2005, m.fl.).

Nyström Höög (2010) för ett resonemang om den *diskursiva texten* och skriver att det skulle kunna vara ett överordnat texttypsbegrepp där den argumenterande texten liksom den utredande skulle kunna ses som undergrupper (2010:13–14). Ask (2005) beskriver att den *diskursiva kompetensen* innefattar kunskaper och färdigheter i det specifika som karaktäriserar akademisk text (ibid., s. 27).

Den utredande texttypen som återfinns i kunskapskraven för delkursen Svenska 2 (se avsnitt 2.3) ställer krav på skribentens förmåga att vara objektiv, analytisk och strukturerad och kan ses som en grundläggande kompetens för vuxenlivets texter (Ask 2005, Nyström Höög 2010, Norberg Brorsson 2007).

Studier om hur elever på olika nivåer behärskar och utvecklar sin förmåga att skriva argumenterande text visar att det är en texttyp som inte alltid följer de typiska drag som hör till (Nyström 2003, Nyström Höög 2010, Östlund-Stjärnegårdh 2002).

Vid Högsolan i Jönköping har undersökningar avseende argumenterande text gjorts av Lindberg och Sandblom (2014). De utgår bland annat från Toulmins (2003) modell för hur argumentation bör vara uppbyggd med genreanpassning och mottagaranpassning. Modellen har testats med elever från högstadiet, gymnasiet och studenter på högskolan samt med andraspråkslever. Resultaten som presenteras visar ”att en tydlig korrelation mellan olika språkliga egenskaper och argumentationens kvalitet är svår att fastställa” (Lindberg & Sandblom 2014:93).

I internationella studier är Toulmins (2003) argumentationsmodell ofta använd och refererad till där man genomför interventionsstudier för att se resultat och göra jämförelser hur elever och studenter klarar av att skriva argumenterande text med, respektive utan, förkunskaper och stöd om argumentationsteknik (se exempelvis Lee 2008, Prior 2005, Voss 1991, Wolfe et al. 2009, Wolfe 2012). Då föreliggande studie fokuserar på vokabulär och ordkategorier görs

ingen analys av texttyperna i elevtexterna eller ordkategoriernas eventuella kopplingar till texttyp.

2.4 Begreppet *ord*

Det vardagliga begreppet *ord* är enligt *Svenska Akademiens språklära* (Hultman 2013) ”mycket svårt att finna en vetenskaplig hållbar definition på” (ibid., 2013:28). Tre kännetecken finns: 1) man kan göra pauser emellan ord 2) de kan oftast bytas ut med ett annat ord med liknande betydelse 3) det kan oftast stå som ensamt yttrande, exempelvis som svar på en fråga. I skrift markeras gränsen mellan ord med mellanslag i svenskan, i andra språk kan andra regler förekomma (ibid., 2013:28–29).

Ord, såsom de förekommer i en löpande text (hur många gånger), benämns *graford*, *textord* eller *löpord*. Den engelska benämningen är *token*. För att ange ett ords unika förekomst i en text används begreppen *lexikonord*, *unika ord* eller *typord*. Här är den engelska benämningen *type*. Begreppet *hapax legomenon* används för att beteckna ord som endast förekommer en enda gång i en text. Hultman & Westman (1977) och Nyström (2000) kallar dessa *engångsord*, vilket är benämningen som också används i föreliggande uppsats.

Uppslagsformen av ett ord är dess *grundform*, denna oböjda form kallas även *lemma* (Hultman, 2013:30). Lemman skiljer sig åt antingen genom olika ordklass eller genom böjningsform. Vid språkliga undersökningar kan ord undersökas i lexikala enheter som består av en avgränsad betydelse, *lexem*. Genom detta kan man se om ord överensstämmer eller inte i fråga om: uttal, ordklass, böjning, etymologi och huvudbetydelse vilket då görs i jämförelser av *lexempar* (Svensén 2004:118). För att särskilja ords betydelser i korpuslingvistisk textanalys krävs *disambiguering* (analys av flertydigheter) på betydelsenivå. Se vidare Gardner (2007). I denna undersökning utförs inte kvalitativ analys av elevtexterna på betydelsenivå.

En mer detaljerad beskrivning av vad som ingår i att kunna ett ord ges av Nation (2013: 44 ff.) där han diskuterar tre huvudaspekter; *form*, *betydelse* och *användning* med ytterligare tre underaspekter till vardera av dessa. Förståelse eller kännedom om ords eventuellt olika betydelse ingår i olika språktest, exempelvis i det internationella språktestet *TOEFL (Test of English as a Foreign Language)*.

Oakhill et al. (2015) använder begreppen *bredd*, *djup* och *access* för motsvarande aspekter av att kunna ett ord. Komplexiteten av detta ingår även i Gustafsson och Håkansson (2017) studie av ordförståelse i högskoleprovet där i huvudsak den receptiva ordförståelsen mäts (ibid., 2017:19 ff.).

3. Material

I detta kapitel presenteras inledningsvis i avsnitt 3.1 informanterna i projektet, därefter görs i avsnitt 3.2 en genomgång av hela materialinsamlingen inom CLISS-projektet för att visa det större sammanhanget och i avsnitt 3.3 beskrivs sedan licentiatuppsatsens materialurval.

3.1 Informanter

CLISS-projektet startade höstterminen 2011 på tre gymnasieskolor med sammanlagt åtta klasser, fem CLIL-klasser och tre kontrollklasser. Skolorna i studien är belägna på tre olika orter i Sverige, från norra Sverige till södra Sverige. Skola 1/A, ligger i en storstad och erbjuder tre gymnasieprogram, vilka samtliga är språk- och ämnesintegrerade undervisningsklasser. Skolan omfattar även årskurs 1 – 9. Skola 2/B, ligger i en mellanstor stad (ca 45 000 invånare) och erbjuder 18 nationella gymnasieprogram varav några är språk- och ämnesintegrerade. Skola 3/C, ligger också i en mellanstor stad (ca 32 000 invånare) och är en gymnasieskola som erbjuder 20 nationella gymnasieprogram varav några är språk- och ämnesintegrerade. Skola B och C är regionala gymnasieskolor och har således stora upptagningsområden medan skola A är en innerstadsskola. De tre skolorna ligger i central stadsmiljö och eleverna kommer både från närliggande områden men de kommer också resande utifrån omgivande landsbygd eller förorter och tar sig till skolorna med allmänna kommunikationsmedel. Projektskolorna är alla inrymda i stora, gamla, vackra skolbyggnader som har moderniserats så att gammalt möter nytt i olika delar av husen. Kontakten med de deltagande skolorna har varit mycket god och alla projektdeltagarna har besökt alla skolorna

och alla klasserna men i olika omfattning beroende på tid, forskningsval och geografiska avstånd.

Undervisningen beträffande CLIL-klasserna, skiljer sig åt mellan de tre skolorna med avseende på i vilken omfattning undervisningen sker på engelska och i vilken omfattning undervisningen är på svenska eller hur språken används parallellt. Det har undersökts närmare av andra i projektgruppen (se Yoxsimer Paulsrud 2014, Sandberg u.u.).

På skola A har två CLIL-klasser deltagit, den ena med naturvetenskaplig inriktning och den andra med samhällsvetenskaplig inriktning. På skola B har två klasser deltagit, en CLIL-klass med naturvetenskaplig inriktning samt en kontrollklass även den med naturvetenskaplig inriktning. På skola C har fyra klasser deltagit, varav två CLIL-klasser den ena med samhällsvetenskaplig inriktning, den andra med ekonomisk inriktning, dessutom har där funnits två kontrollklasser, båda med ekonomisk inriktning. Alla elever läser således studieförberedande program. I tabell 1 nedan har gymnasieprogrammen förkortats till EK = ekonomiprogrammet, NA = naturvetenskapsprogrammet och SA = samhällsvetenskapsprogrammet. Dessa förkortningar används fortsättningsvis i uppsatsen.

TABELL 1: *Deltagande skolor A, B och C samt elevantal och fördelning av flickor och pojkar utifrån tillbandabärlna klasslistor.*

Skola	1-A	F/P	2-B	F/P	3-C	F/P
Program/antal elever						
EK CLIL					19	15/4
SA CLIL	37	26/11			28	20/8
NA CLIL	30	19/11	33	23/10		
NA icke- CLIL			33	21/12		
EK icke-CLIL					A-31	16/15
					B-26	13/13

Tabellen ovan visar att klasstorlekarna har varierat från 19 till 37 elever i de olika klasserna. Antalet deltagande elever har också varierat då vissa elever har bytt program under sin gymnasietid, till eller från den klass de har tillhört, några elever har gjort studieuppehåll och varit med en del av projekttiden men inte hela. Elevantalet som redovisas i tabell 1 är de som har funnits med på klasslistor vid tillfällena då skrivuppgifterna i denna studie har samlats in.

Totalt har 240 elever deltagit i projektet. Av dessa är 71 % unga kvinnor och unga män 29 %. I studien tillhör 68 % av de deltagande eleverna CLIL – klasser och 32 % icke-CLIL. Av eleverna har 85 % svenska som sitt modersmål, L1. Av de deltagande eleverna har 15 % ett annat modersmål än svenska där språk som ara-

biska, kurdiska och albanska representerar de största språkgrupperna men även andra språk förekommer. I bakgrundsenkäten visar resultaten att av de som deltagit att 83 av CLIL-eleverna är födda i Sverige (41,5 %), 98 i något annat land (49 %). Av icke-CLIL-eleverna är 17 födda i Sverige (8,5 %) och 2 elever (1 %) i annat land. På frågan om varför de valt att läsa sin gymnasieutbildning på engelska svarar en majoritet av CLIL-eleverna att de vill bli bättre på engelska, de vill kunna använda engelskan i högre utbildning och att de gillar engelska (statistik hämtad från bakgrundsenkäten, sammanställd och analyserad av Apelgren, 2014).

I analysen har resultaten kopplats till de olika programvalen; profilerna CLIL- och icke-CLIL, till genus med indelningen unga kvinnor (F) och unga män (P) samt med indelning av genus i CLIL- och icke-CLIL-grupperna. I uppsatsen redovisas resultaten för CLIL och icke-CLIL samt för (F) och (P) i respektive grupp där även förkortningarna PC, FC, PnC och FnC används.

3.2 Materialinsamling i CLISS-projektet

Materialinsamlingen i CLISS-projektet har varit omfattande och pågått regelbundet under hela den period som eleverna har gått sin gymnasieutbildning åren 2011-2014. Insamlingen har genomförts utifrån den ursprungliga projektutformningen (se bilaga 1).

Under termin 1 fick de deltagande eleverna fylla i en omfattande bakgrunds-enkät med frågor om deras läs- och skrivvanor både på svenska och engelska, varför de valt sitt gymnasieprogram, deras boende och fritidssysselsättningar, men också frågor om föräldrarnas utbildningsnivå och annat. Eleverna som deltog i projektet fick även fylla i en motivationsenkät som de sedan fick fylla i en gång till under sin sista termin på gymnasiet (se avsnitt 2.1.1 och Sylvén & Thompson 2015). Under termin 1 genomfördes även läsförståelse – och ordkunskapstest samt de första skrivuppgifterna på svenska och engelska (*Nuclear Power* och *Hälsa och välbefinnande*).

Vid skrivtillfällena har eleverna fått skrivuppgifterna utdelade i pappersform, därefter har den ansvarige projektmedlemmen läst/gett en instruktion som funnits med och så har eleverna haft 90 minuter till förfogande för att genomföra skrivuppgifterna på datorer eller för hand. På två av skolorna har eleverna haft tillgång till egna datorer och på den tredje skolan har man fått boka datorer innan insamlingstillfället. Efteråt har textinstruktionerna samlats in av ansvarig lärare och eleverna har fått en webbadress att skicka sina dokument till och i de fall då eleverna har skrivit för hand har allt material samlats in och därpå lämnats till projektledaren för att föras över till digital form i ursprungligt skick med eventuella stav- och skrivfel. Texterna har sedan förts över till CLISS-projektets databas av en medlem i projektgruppen som har kodat allt med elevernas individuella koder och lagt dem i mappar som återfinns i projektets databas (se 2.1.1).

3.3 Urval

Licentiatuppsatsens material består av 520 elevskrivna texter från fyra skrivuppgifter vilka har skrivits terminerna 1, 3, 4 och 5 under elevernas tre gymnasieår om totalt sex terminer. Textinstruktionerna som har inlett varje skrivuppgift utgår från kunskapskraven för svenskämnet i *Lgy11* (Skolverket 2011a), samt har utgångspunkt i en inriktning mot samhälls- och naturorienterade ämnesplaner (se avsnitt 2.3). Skrivuppgifterna har rubrikerna *Hälsa och välbefinnande*, *Bilfri stad*, *Folkhälsoproblem* och *Helt delad föräldrapeng* (tabell 2, 8 och 10, bilaga 3 och 4). Till sin utformning har skrivuppgifterna till viss del liknat uppsatserna som ingår i de nationella proven i svenska och engelska med en inledning, en uppgift och så tillhörande information av något slag som eleverna har kunnat inspireras av och har kunnat använda i sina texter. Denna form av igenkänningsfaktor har varit betydelsefull då det har inneburit att eleverna har vetat vad som förväntats av dem och använt den tid som har avsatts till ett fokuserat skrivande. Efter varje skrivuppgift har alla elever getts möjlighet att också kommentera skrivuppgiften, om den var lätt eller svår, hur tidsramarna var för uppgiften och så vidare. En del elever har använt möjligheten och elevernas kommentarer finns i materialinsamlingen men de redovisas inte närmare här.

Elevtexterna som analyseras i denna studie har inte ingått i elevernas ordinarie undervisning men de har utförts på ordinarie lektionstid. CLISS – projektet är inte en interventionsstudie där projektdeltagare varit delaktiga i den ordinarie undervisningen och implementerat instruktioner i förväg för att se vilken effekt det skulle ha på hur eleverna skrivit textuppgifterna. I projektgruppen vet vi alltså inte i vilken omfattning eleverna i studien har erhållit undervisning om former för argumenterande och utredande text, vilka ingår i kunskapskraven för gymnasiet kurser i svenska (se avsnitt 2.3). Vi vet heller inte om, och i så fall i vilken omfattning, man i klassrumskontexten har undervisat om ord och/eller ämnesspecifika begrepp som kännetecknar en mer vetenskapligt skriven text, dock finns skrivningar om detta i läroplanen och kursplanerna (se *Lgy11*). Elever och lärare har haft möjlig tillgång till skrivuppgifterna i efterhand som ett gemensamt arbetsunderlag. Möjligheten har även getts för lärare att använda elevtexterna som bedömningsunderlag, vilket några har gjort, dock inte alla.

TABELL 2: *Skrivuppgifterna på engelska och svenska (fet stil) i CLISS-projektet*

	Termin 1 (ht 2011)	Termin 2 (ht 2012)	Termin 3 (vt 2013)	Termin 5 (ht 2013)
Argumenterande text	<i>Nuclear Power (NO)</i>	<i>Bilfri stad (NO)</i>	<i>Violence/non-Violence (SO)</i>	<i>Helt delad föräldrapeng (SO)</i>
Utredande text	<i>Hälsa och välbefinnande (SO)</i>	<i>Gender and Equality (SO)</i>	<i>Folkhälsoproblem (NO)</i>	<i>Biodiversity (NO)</i>

4. Teoretisk bakgrund

I kapitlet följer en presentation av studiens teoretiska ramverk. Undersökningen av de 520 elevtexterna och elevernas skolrelaterade skriftspråkliga val utgår från ett sociokulturellt perspektiv då materialet som studeras har tillkommit genom mediering i en skolkontext och skrivuppgifterna utgår från rådande läroplan, ämnesplaner och kunskapskrav. För att identifiera ämnesspråket och därmed en akademisk vokabulär genom valda ordkategorier, används en systemisk-funktionell språksyn, vilken beskriver både ett språkssystem och språkets sociala och kognitiva funktion.

4.1 Ett sociokulturellt perspektiv

I ett sociokulturellt perspektiv blir språk och språkutveckling till i mötet mellan människor, i kommunikativa processer eller i mötet mellan människa och upplevelse. Elevtexterna som har undersökts har skrivits i en skolkontext vilken sannolikt har haft betydelse för elevernas språkliga val och som därmed kan kopplas till det språkliga samspel de har ingått i men som inte är den enda kontext som kan ha påverkat elevernas vokabulär (Vygotskij 1934/1999, Säljö 2015).

Det är den sociala interaktionen som bidrar till hur barn utvecklar sitt tänkande, språk och beteende. Vygotskij (1999) pekar på kulturens betydelse, vilket medför att barn kan utvecklas olika beroende på i vilka sociala sammanhang de ingår. Lärande föregår utveckling, och lärande är en förutsättning för kulturell utveckling. Ett bärande begrepp inom sociokulturen är *appropriering*, tillägnelse, att göra till sitt, som enligt Säljö (2015:95), är metaforen för lärande, där Säljö skiljer på *lärande* och *undervisning*. Lärande är osynligt och kan pågå när som helst

och var som helst, medan undervisning kan studeras i organiserade former där den följer styrdokument och planeringar (Säljö 2015:22 ff.).

Vygotskij (1999) menar att mognadsmetaforen; *Zone of Proximal Development* (ZPD), *den proximala utvecklingszonen*, att ett begreppsstadium måste nås innan utveckling kan ske. Han betonar samtalets betydelse för utveckling av tänkandet och att kunskaper genereras i sociala aktiviteter och underhålls kollektivt över tid genom processer där man med stöttning av andra, *scaffolding*, vidareutvecklar kunskaper (Wood, Bruner & Ross, 1976). Likaså talas om barnets *inre språk* som ett egocentriskt språk som sedan genomgår en utvecklingskedja och tillsammans med intellektuell mognad formar ett *yttre språk*. Ordets betydelse hör samman med det språkliga tänkandet och det meningsfulla ordet är en *enhet* av ord och tanke Därmed är ordet både ett språkligt och ett intellektuellt fenomen (Vygotskij 1999:394). Skolan en plats där man ges möjlighet att lära *vetenskapliga (akademiska)* begrepp som skiljer sig från de *spontana (vardagliga)* begreppen som man lär sig hemma (Säljö 2015:102). I sina kartläggningar av lärprocesser undersökte Vygotskij även skriftspråket och ställde frågan varför skriftspråket ter sig så mycket svårare och så mycket mindre utvecklat än talspråket att det till och med kan skilja upp till åtta år i ”lingvistisk ålder” (Vygotskij 1999:315 ff.). Jämförelser görs mellan det *inre språket* och det *yttre*, där skriftspråket beskrivs kräva en hög abstraktionsnivå hos individen då både ljud och samtalspartner saknas. Förutsättningarna för skriftspråket kräver mer av individen både när det gäller tanke och motivation vilket gör att ”skriftspråket är mycket mer medvetet än talspråket” (ibid. s. 320).

Ett annat begrepp inom sociokulturen är *mediering*, ett uttryck för det som sker mellan verktyget och individen. Barnet förs in i en förtolkad social och kulturell omgivning där språkbruket formar tänkandet. Säljö (2000) skriver: ”Ord och språkliga utsagor medierar således omvärlden för oss och gör att den framstår som meningsfull” (2000:82, 2015:94). För all sorts kommunikation människor emellan kan språket ses som den viktigaste resursen eftersom kunskap är språklig och medieras i ett socialt samspel mellan lärare och elever (Nygård Larsson 2011:40). Halliday (1993) uttrycker sambanden mellan lärande och språk som ”learning language, learning through language, learning about language” (1993:113).

Även Cummins (1979, 1980) visar i sin forskning om språkutveckling på de språkliga krav som olika kontexter kräver, där han använder begreppen *BICS – Basic Interpersonal Communicative Skills* och *CALP – Cognitive/Academic Language Proficiency*. Det förstnämnda representerar det grundläggande språkbruket, det vardagliga som sker mellan människor i vardagliga sammanhang. Det senare begreppet CALP definieras som “those aspects of language proficiency which are closely related to the development of literacy skills in L1 and L2”, alltså ett mer komplext och krävande språkbruk där färdigheter lärs och utvecklas (Cummins, 1980:177).

Språket som en social konstruktion ställer särskilt stora krav i skolsammanhang vilket något som även Schleppegrell (2004) visar på i sin forskning. Hon menar att: “the language of schooling, presents information and interpretation in new ways, using vocabulary, grammar, and text structures that students can learn to recognize when they read and to adopt when they write” (Schleppegrell et al. 2008:176). Detta fokuseras och undersöks i föreliggande studie.

4.2 Ett systemisk-funktionellt perspektiv

För att arbeta med språkets betydelse i text kan man använda den systemisk-funktionella grammatiken, SFG, som en språkmodell där betydelse och funktion är utgångspunkten (Halliday & Matthiessen 2014). I *An Introduction to Functional Grammar (IFG4)* uttrycks detta som: “Language is, in the first instance a resource for making meaning; so text is a process of making meaning in context.” (ibid. 2014:3). En tidig användning av SFG var i Australien när man kunde se hur andraspråksanvändare, L2, inte lika självklart som förstaspråksanvändare, L1, hade tillgång till framförallt skriftspråkets olika koder och mönster, vilket visade sig ha avgörande betydelse vid språkinläring i skolan.¹⁴ Genom SFG, och *Genrepedagogiken*, som kan ses som dess pedagogiska tillämpning, synliggjordes språkets funktioner och därmed kunde elever ges ett språkligt stöd, *scaffolding*, för att visa hur språk ”byggs” och kan ha olika funktioner i olika sammanhang (Gibbons 2009, Halliday 1993, Martin & Rose 2008, Schleppegrell 2004, Wood et al. 1976).

Benämningen *systemisk – funktionell lingvistik, SFL*, används om ett vidare fält där man även kan inkludera andra språk teorier som *appraisalteorin*, *genre teorin* och *multimodal analys*, därmed kan SFL sägas ha en bred ansats (af Geijerstam 2006, Holmberg, Karlsson & Nord 2011). Det systemisk-funktionella perspektivet innebär en språksyn som utgår ifrån språkets betydelse och funktion och inte främst form, vilket gör att den kan ses ha likheter med ett sociokulturellt perspektiv (af Geijerstam 2006, Holmberg & Karlsson 2006, Lim Falk 2008, Nygård Larsson 2011, Persson 2016, Schleppegrell 2004).

I denna studie av elevers skriftspråkliga val kan det systemisk-funktionella perspektivet ses som ett ramverk där ord har en språklig funktion som beror på sammanhangen de används i och kan därmed motivera ordkategorier som typiska för språkliga drag (Persson 2016). Holmberg och Karlsson (2006) menar att den funktionella grammatiken förutsätter att språket kan skapa betydelse i kommunikationen mellan människor (2006:18). I och med det kan språket ses som en meningsskapande resurs, *register*, som hänger samman med kontexten

¹⁴ L1=Language 1; förstaspråk, modersmål, L2=Language 2, andraspråk.

där språket återfinns, exempelvis i matematik eller i något annat skolämne (Schlepperegell 2004).

De ordkategorier som utifrån denna språksyn ses representera det vetenskapliga språket och som undersökts i studien är *nominaliseringar, verb i passiv form och akademiska ord* (Frändberg 2012, Johansson Kokkinakis & Frändberg 2013, Nygård Larsson 2011). Nominaliseringar är verb och adjektiv som görs om till substantiv (ex. att *representera* blir en *representation*). Nominaliseringar, är enligt SFG, exempel på *ideationell grammatisk metafor* (Halliday & Matthiessen 2014:729 ff., Holmberg & Karlsson 2006:162 ff., Magnusson 2011:2, 43). Begreppet grammatiska metaforer är, enligt Magnusson (2011), en samlingsterm inom SFL för hur språket kan ”packas ihop” så att deltagare och processer förs samman för ett mer specialiserat språkbruk vilket kan uppfattas som abstrakt och byråkratiskt (mer i kapitel 5.2.1). Användandet av grammatiska metaforer ses som ett sent steg i språkutveckling och hör samman med skriftspråklig utveckling (Halliday & Matthiessen 2014, Magnusson 2011:3). Även Vygotskij pekar på att ”de vetenskapliga begreppen är en produkt av skolutbildningen” och något som sker sent i barns utveckling (Vygotskij 1999:384).

Den passiva formen av verb, vars förekomster också undersöks i elevtexterna, innebär att pronomen i handlingen saknas (istället för; *jag* undersöker, står *undersöks*). Detta medför att texter blir mer komplexa, mer distanserade och därmed kan upplevas svårlästa (se vidare 5.2.2). Den systemisk-funktionella grammatiken erbjuder verktyg för att ”packa upp” den här typen av texter och på så vis skapa förståelse, både receptivt och produktivt (Schlepperegell 2004:45).

I af Geijerstams avhandling *Att skriva i naturorienterade ämnen i skolan* (2006) står:

Vi skapar mening i språk genom att välja bland tillgängliga grammatiska resurser. Språket kan betraktas som en meningspotential där vi uttrycker oss på olika sätt beroende på vad vi vill uppnå. De språkliga val som görs är dock villkorade av sammanhanget där språket används och även av den sociokulturella positionen hos individen. [...] Genom de lingvistiska val som görs skapas också sociala kon-
texter. (ibid. 2006:24).

Med detta som en bakgrund analyseras i föreliggande studie elevernas skriftspråkliga val genom vokabulären i deras skrivuppgifter/texter utan att analysera skolkontexten som texterna är skrivna i. Texterna analyseras heller inte genom de tre metafunktionerna som finns inom SFL; *interpersonell, ideationell* och *textuell* eller de *skikt* om kan ses samspela i dessa då texterna inte undersöks som helheter (Hipkiss 2014, Holmberg & Karlsson, 2006). Undersökningen av vokabulären i texterna görs utifrån de teoretiska ramar som har presenterats ovan och därmed utgår studien från den språksyn dessa representerar; att elevernas skriftspråkliga val delvis har formats i och av, den sociokulturella kontext som eleverna har befunnits i och där uppgifterna genomförts men att det givet också finns

fler faktorer som kan ha påverkat elevernas språk och deras skriftspråkliga val. Dessutom undersöks texterna utifrån en systemisk-funktionell språksyn genom val av ordkategorier som enligt denna kännetecknar ett mer specialiserat vetenskapligt, akademiskt sätt att skriva.

5. Tidigare forskning

Kapitlet ger en forskningsöversikt som avser placera uppsatsens frågeställningar om att synliggöra en skriftspråklig vokabulär i forskningsfält. Först beskrivs forskning som utgår från begreppen *vardagsspråk* och *skolspråk*, därefter följer *ämnesspråk*. För undersökning av ordkategorier relateras till forskning där korpuslingvistiska metoder har använts. Sedan följer tidigare forskning som har använt kvantitativa undersökningsmetoder, där textuella måttenheter har analyserats. Här redovisas även tidigare undersökningar av elevskrivna texter där korrelation med bedömning har studerats. Slutligen redogörs för undersökningen av lexikala profiler som har gjorts för att synliggöra spridning och förekomst av *hög- och lågfrekventa ord*.

5.1 Vardagsspråk och skolspråk

I skolan möter elever ett skolspråk som skiljer sig från det vardagliga språket man talar med familj och kamrater (Lindberg 2007, Magnusson 2008, Schleppegrell 2004). Schleppegrell skriver att lärande primärt är en språkprocess som är olika tillgänglig för olika barn beroende på bakgrundsfaktorer och refererar till Christie (1985) som menar att man kan kalla skolspråket för ”den dolda läroplanen” (Schleppegrell 2004:2, [... the hidden curriculum], min översättning). Edling (2006) beskriver det som en av de viktigaste saker en formell utbildning kan erbjuda, när man går från vardagskunskap till specialkunskap och därmed från ett vardagsspråk till ett specialiserat språk. ”I denna övergång omtolkas världen” (ibid. 2006:186).

Eftersom skolspråket till stor del är skriftspråksbaserat byggs det upp med en annan komplexitet och struktur än det vardagliga språket. Lindberg (2007) menar att skolspråket inte är någons modermål utan ett språk som måste läras som ett andraspråk. Oavsett modersmål skall alla elever ges ett språkligt stöd som ger dem möjlighet att lyckas i sitt skolspråk (Bailey 2007).

För att kunna utveckla skolans specialiserade språk och utveckla elevers språkliga repertoar behövs medveten och systematisk undervisning med explicit vägledning och tydliga språkliga modeller (Lindberg 2009:19). Det innebär att lärare måste arbeta språkstödande och tydliggöra språkliga strukturer som finns i olika texttyper vilka hör samman med olika skolämnen (Gibbons 2009 m.fl.). Halliday och Martin (1993) pekar på att det vetenskapliga språket som elever möter i skolan till viss del kännetecknas av tekniska termer, men de skriver:

“Of course, technical terms are an essential part of the scientific language [...]. But they are not the whole story. The distinctive quality of scientific language lies in the lexicogrammar (the ‘wording’) as a whole” (1993:4).

Holmberg och Karlsson (2006) skildrar det som att det är i *lexikogrammatiken*, som är det skikt, utifrån en systemisk-funktionell språksyn, där betydelsen ”för-språkligas” (ibid. 2006:30). Det innebär att det är grammatiken tillsammans med orden som ger material för att uttrycka betydelser. Avståndet mellan betydelse och lexikogrammatik är litet; ”processbetydelser uttrycks med verb, uppmaningar med imperativ och frågor med just frågegrammatik” (2006:162).

Genom att, som denna studie avser att göra, undersöka hur elevernas aktiva ordförråd framställs och ser ut över tid, kan frågor som rör elevers produktiva språkliga repertoar besvaras. Enligt Vygotskij (1999) är ”ordet ett tecken” och förmågan att gå från det vardagliga språket till en generalisering i form av begrepp är något barn inte når förrän de är i puberteten (ibid. s. 239).

Kraven på ett omfångrikt ordförråd som innehåller både allmänna ord men även specialicerade ord ökar med åren. Vardagsspråket kan jämföras med språket elever talat på raster och på sin fritid och skolspråket med det språk som förekommer såväl muntligt som skriftligt i mer kognitivt krävande uppgifter i skolans alla ämnen (Olofsson & Sjöqvist 2013:700). Hur detta sedan kommuniceras och implementeras i skolan är något som skiljer sig från klassrum till klassrum beroende på hur information och dialog förmedlas genom olika menings-skapande resurser (Hippkiss 2014).

Tabell 3 visar hur det vardagliga språket utgår ifrån personens vardagskunskaper, egna erfarenheter och byggs bit för bit, ofta i dialog med någon. Det vardagliga språket är mer oreflekterat och kan förhandlas fram till skillnad mot skolspråket som oftast är skriftspråksbaserat och uppdelat i discipliner, det vill säga här; skolämnen.

TABELL 3. *Vardagskunskap och skolkunskap*

Vardagskunskap	Skolkunskap
byggs upp sakta, gradvis och bit för bit på ett fragmentariskt sätt	byggs upp snabbt och systematiskt i en logisk sekvens
byggs upp i en takt som framför allt bestäms framför allt av barnet	byggs upp i en takt som bestäms av läraren, kursplanen
byggs upp på ett oreflekterat sätt	byggs upp på ett reflekterat sätt
förhandlas fram i talspråk	konstitueras i skriftspråk
är partikularistisk	är universalistisk
är baserad på personliga och ömsesidiga erfarenheter	är distanserad från personliga och ömsesidiga erfarenheter
har konkret och icke-teknisk betydelse	har abstrakt och teknisk betydelse
utgör en helhet	är uppdelad i discipliner

(Lindberg 2006:77, Magnusson 2008, från Painter 1999:353)

I det talade språket är vi ofta närvarande som en del av kontexten och kan förstå hur, vad och vem som menas genom användandet av olika pronomen. En skriven text måste vara självbärande, det vill säga förmedla syfte och tydligt uttrycka en mängd information som är kontextualiserat i tal men inte i skrift (Magnusson 2008:12, Schleppegrell 2004). Skolspråket som ofta utgår från exempelvis lärobokstext, eller avser att inleda en skrivuppgift för elever, kan därmed sägas vara både dekontextualiserat och kontextualiserat. Tabell 4 exemplifierar detta.

TABELL 4. *Karaktäristiska drag för skolspråk*

Drag	Karaktäristik av draget
Explicit	Exakt referens, mer elaborerat ordförråd
Distanserat	Läsaren ger inte någon interaktiv hjälp som visar att budskapet har gått fram. Skribenten måste sätta sig in i läsarens situation och inte ta förkunskaper för givna.
Komplex och strukturerat	Komplex struktur, sammanhållet.

(Magnusson 2008:13, fritt översatt från Snow et al. 1989, Schleppegrell 2004:9)

5.2 Ämnesspråk

I svensk skolkontext gällande ungdomsskolan används benämningarna *skolspråk* och *ämnesspråk* ofta parallellt, och Skolverket framhåller ett språkutvecklande arbetssätt för att stärka elevers prestationer. om ämnesspråk står bland annat:

Ett välutvecklat ämnesspråk gör det möjligt för eleverna att förstå hur grundläggande principer i ämnet är organiserade. Ämnesspråket gör det också möjligt för eleverna att kunna gå in i ämnesspecifika diskussioner och föra relevanta resonemang eller att läsa och skriva de texttyper som dominerar i det aktuella ämnet.[...]. Forskning visar till exempel att en ämnesspecifik och språkutvecklande undervisning är kontextrik och innehåller språklig stöttning och mycket interaktion. Sådan undervisning är gynnsam för alla skolans elever, men nödvändig för de flerspråkiga eleverna. (Skolverket 2014)

Skolspråket (se 5.1) uppvisar lexikala aspekter där det ”specialiserade språket blir en semantisk resurs” (Lindberg, 2007:15) som ger tillgång till det mer akademiska språkbruk, som förekommer i exempelvis SO- och NO- texter. Enligt Skolverket är *ämnesspråket* en förutsättning för att utveckla kunskaper i skolans alla ämnen när det handlar om att ”uttrycka, tolka, förstå och använda begrepp, fakta och tankegångar i tal och skrift” (ibid., 2016a). Det krävs en stor språklig medvetenhet hos ämneslärare som skall undervisa i ämnen med höga språkliga krav. Det naturvetenskapliga språket kännetecknas exempelvis av att det är specialiserat och informationstätt med abstrakta, vetenskapliga begrepp och nominaliseringar samt med en objektivitet i texten som gör deltagaren osynlig (af Geijerstam 2006, Nygård Larsson 2011).

Texter om historia är ofta innehållstäta med många abstrakta begrepp (Schleppegrell & de Oliveira, 2006). Av en läsare krävs en implicit kunskap om tid och rum, det vill säga en abstrakt förståelse för att kunna följa med i texten och förstå när och var händelser ägt rum. Händelser kan ha flera orsaksamband som kan vara osynliga i texten men finnas inbäddade i generaliseringar eller långa nominaliseringar som bli mycket svåra att följa och förstå. Likaså kan avancerade tidsfraser användas, som förväntas ingå i elevers vardagsspråk men som inte är självklara för alla elever (ibid., 2006).

I olika studier görs avgränsningar och definitioner av ordkategorier för att kunna mäta ords förekomst och spridning i ett textmaterial. Detta görs genom antaganden om vissa typer av delvokabulärer som representerar exempelvis ett ämnesrelaterat språk men teoretiskt är gränserna mellan dessa inte skarpa (Lindberg & Johansson Kokkinakis 2007). Gemensamt är att det visas hur ämnen är språkligt konstruerade och bygger upp språkliga mönster som kan utgöra stora svårigheter för elever som inte har erövrat ett skolspråk som ger dem tillgång till förståelse synnerhet de med annat modersmål och de som inte får med sig skolspråket hemifrån (se Lindberg 2007, Skolverket 2008, 2012).

En tidig ordförrädsstudie visar att det finns ordkategorier som är så vanliga, att man förväntar sig att alla elever har en förståelse för dem, trots att de inte tillhör kategorin ämnesrelaterade eller ämnesspecifika ord/*fackord*. Det visar sig att just de orden, de *icke-fackliga*, kan vara nog så svåra, speciellt för andraspråkslever. Exempel på några sådana är: gnida, kvist, hastighet, blandning (se Golden & Hvenekilde 1983, Golden 2006). Svårigheten kan dels komma av att orden inte uppfattas som svåra utan ”vanliga” för en lärare utifrån ett förstaspråksperspektiv så att explicit undervisning inte ges, dels att orden kan ha olika betydelse i olika sammanhang.

Macken-Horarick (1996) använder begreppet *domän* (eng. *domain*) för att beskriva hur språk används beroende av kontexten; i ”vardagsdomänen” (min översättning av *everyday*) används *funktionell litteracitet* i form av samtal och kommentarer, i den ”specialicerade” domänen (*specialized*) används språk för att konstruera, exempel här är skolspråket och i den ”reflexiva/reflekterande” domänen (*reflexive*) används språket för att beskriva lärande (ibid., 1996:241). Domänbegreppen kopplas till den systemisk-funktionella grammatiken, SFG, och till situationen, den kontext, där språket används.

Genom att identifiera ordkategorier i texter avsedda för olika åldersgrupper kan man få hjälp både med att veta vilka ord som är mest frekventa och som kan utgöra svårigheter om de inte ingår i elevernas vokabulär och att man på så vis får underlag för att kunna göra test av olika slag. I en studie av Nation (2013) identifieras fyra kategorier av ord; *högfrekventa ord*, *akademiska ord*, *tekniska termer* och *lågfrekventa ord* (min översättning).

I *OrdiL-rapporten* (Lindberg & Johansson Kokkinakis 2007, förkortning av projektet; *Ord i läromedel*) som är en kartläggning av ordförrådet i svenska läromedel i olika ämnen i grundskolans senare år har korpuslingvistiska metoder tillämpats. Genom att använda läromedelstexter i SO- och NO-ämnena samt i matematik (16 allmänt använda läroböcker för årskurs 8) samlades omkring 1 miljon graford som har gjorts tillgängliga och sökbara i digital form. *OrdiL-rapporten* har bidragit med viktiga kunskaper om de lexikala krav som läromedel kan ställa och visa på ord som oberoende av ämne kan betraktas som frekventa och allmänt skolrelevanta (ibid., 2007:10).

Vardagliga ord avser ord som är vanliga både i tal och i skrift men även i mer formella sammanhang. I *OrdiL* (2007) görs en skillnad mellan *allmänspråk* och *fackspråk*, dock menar man där att det inte finns några tydliga och distinkta gränser. De skillnader som finns kan vara både i grammatiken men även i ordförrådet där den största delen, även i fackspråk, utgörs av allmänspråkliga ord (2007:62). I *OrdiL-rapporten* skiljer man först på *ämnesneutrala ord* och *ämnesrelaterade ord* med två undergrupper till vardera kategorin, se nedan (*OrdiL* 2007:86).

TABELL 5. *Sammanfattning av ordkategorier i OrdL-rapporten (2007, s. 86-87).*

Ämnesneutrala	Exempel
allmänspråkliga, frekventa ord	ha, vara, komma, människa, exempel
allmänna ofta abstrakta ord	utbredning, resurser, bilda, avta, föremål
Ämnesrelaterade ord	Exempel
allmänspråkliga, ämnestypiska ord	klimat, sträcka, blandning, strålning
fackord, facktermer, ofta unika för ett visst ämne	decimalform, kromosom, reformationen,

Lim Falk (2008) använder benämningarna *ämnesrelaterat* och *ämnesanknutet* parallellt i sin studie där hon undersöker i vilken omfattning CLIL-elever använder ett ämnesrelaterat språkbruk (ibid., 2008:53). I sina undersökningar analyserar hon både text och kontext i en CLIL-klass och en kontrollklass på gymnasiet för att se elevernas bruk av ett ämnesrelaterat/ämnesanknutet språk i kemi, fysik, historia och svenska. Begreppen *ämnesspecifikt/ämnestypiskt* används om språkknutet till kemidiskursen, fysikdiskursen och så vidare (ibid.).

Resultaten i Lim Falks studie visar tydligt på språkanvändningens betydelse för språkutvecklingen (ibid., 2008:277). Trots att CLIL-eleverna i allmänhet är mer studiemotiverade än kontrollklassens elever (se Sylvén & Thompson 2013) och trots att CLIL-eleverna oftast har högre betyg i de flesta ämnena än eleverna i kontrollklassen (se även Apelgren u.u. 2018) så använder kontrollklassens elever i Lim Falks studie (2008), i större utsträckning ett ämnesrelaterat språk i de ämnen som ingår i studien och visar en större säkerhet i användningen av ett ämnesanknutet språk. Dock är förhållandena omvända i svenskämnet. Slutsatsen som dras är att undervisningsspråk och språkträning har en avgörande betydelse (ibid., 2008:280).

I analyser av läromedelstexter i svenska, SO och NO för skolår 5 och 8 samt år 2 på gymnasiet använde sig Edling (2006) av kategorierna *konkreta* och *abstrakta* begrepp där hon kombinerar begreppsparen till en analys av abstraktion i tre kategorier: *specifika*, *generell* och *abstrakt* (ibid. 2006:188 ff.). Resultaten visar att abstraktions- och generaliseringsgraden i texterna ökade i alla ämnen utifrån skolåren samt att texterna i svenska i allmänhet tillhörde en vardaglig domän och de naturvetenskapliga texterna en mer specialiserad domän även om det fanns drag av båda i de olika grupperna. Eleverna i studien visade mindre textrörlighet i texter med högre grad av abstraktion. Edling menar att för att kunna hantera de textkrav samhället ställer så är lösningen inte att ta bort den sortens texter utan i stället skapa aktiviteter runt texterna som stärker mötet mellan elev och text (ibid., s. 190).

En annan studie av elevers skrivande i de naturvetenskapliga ämnena är den som af Geijerstam (2006) genomfört i samma projekt, som Edlings (2006) stu-

die.¹⁵ Här undersöks undervisningspraktiker, elevaktiviteter och texter i ett antal klassrum i årskurs 5 och 8 för att kunna beskriva skrivförlopp, texter och textförståelse. Resultaten i af Geijerstams studie visar att de naturorienterade ämnena till stor del är muntligt baserade, och att de elevskrivna texterna i stort sett enbart är skrivna med någon förlaga samt att förståelse uttrycks muntligt och inte kopplas till skrivande och text (ibid., 2006:160 ff.).

Man kan också ifrågasätta till vilken grad de texter eleverna skriver fungerar som verk samma redskap för lärande. Då elevernas textbaserade rörlighet är mycket låg verkar texterna inte fungera som grund för lärande. (af Geijerstam 2006:165)

I en undersökning av en läromedelstext i kemi samt elevsvar från en kemiuppgift i TIMMS (Trends in International Mathematics and Science Study),¹⁶ används kategorierna: *tekniska termer, nominaliseringar, passiva verb, vardagliga och snedfördelade ord, vardagliga och ämnesspecifika ord, namn/formler och akademiska ord* (se Frändberg 2012, Johansson Kokkinakis & Frändberg 2013). Studien innehåller även en korrelation med hur elevsvaren har bedömts där elevsvaren har delats in i tre grupper utifrån resultat (låg-, medel-, högpresterande). Resultaten av hela studien visar att högpresterande elever använder sig av fler tekniska termer, fler akademiska ord, fler passiver och vardagliga ord med ämnesspecifik betydelse (exempel; koppar, massa) samt skriver längre texter (2013:65). Detta överensstämmer med den tidigare, mer omfattande studie av Frändberg som bygger på samma material (2012).

På senare år har studier med fokus på flerspråkiga elevers förutsättningar presenterats. I avhandlingen *Biologiämnets texter*, av Nygård Larsson (2011) undersöks text, språk och lärande i gymnasiets kurs Biologi A. Materialet i studien består av kursläroboken, lärarens texter samt elevtexter (114 prov, 49 skrivuppgifter från en NA-klass i årskurs 1, Gy). Resultaten av textanalys av elevernas texter visar att de flerspråkiga eleverna inte når högre betygskrav än godkänt, sannolikt för att uppgifterna är för krävande gällande skol- och ämnesrelaterad språklig nivå (ibid. 2011:303 ff.).¹⁷ Resultaten visar även att mer självständigt skrivande av ämnesrelevanta texter utgör en utmaning för många flerspråkiga elever.

Utifrån studiens resultat resonerar Nygård Larsson (2011) hur fler elever skall få möjlighet till att nå högre betyg då mer avancerade textaktiviteter i sig är en svårighet, framför allt för flerspråkiga elever (ibid., 2011: 316) I Skolverkets *PM: Betyg och studieresultat i gymnasieskolan 2013/2014* står: ”Elever med svensk bakgrund når gymnasieexamen på tre år i större utsträckning (69 procent) än elever med ut-

¹⁵ Projektet *Elevers möte med skolans textvärldar* 1999-2003, projektledare professor Caroline Liberg.

¹⁶ I TIMSS deltar länder från hela världen och studien genomförs vart fjärde år, man undersöker elevers kunskaper i matematik och naturvetenskap i årskurs 4 och årskurs 8.

¹⁷ Studien gjord före *Gy11* med dåvarande betygsskala IG, G, VG, MVG, se tidigare fotnotsref.

ländsk bakgrund (43 procent)”¹⁸ I *Skolverkets lägesbedömning 2015* står ”elever med utländsk bakgrund har fortsatt sämre resultat än elever med svensk bakgrund”¹⁹.

Nygård Larssons undersökning (2011) visar att en språk- och textmedveten undervisning är en förutsättning för flerspråkiga elevers språkutveckling och lärande men att en sådan också gynnar alla elever eftersom ”ämne och språk kan ses som integrerade och därmed konstituerande varandra” (ibid., 2011:320).

Vilka möjligheter en ämneslärare i kemi har att från ett språkdidaktiskt perspektiv utveckla en språklig repertoar för andraspråkselever, genom sina undervisningsstrategier diskuteras i en studie av Kouns (2010). Studien genomförs i en klass med en lärare och Kouns undersöker om de allmäntdidaktiska strategierna som läraren tillämpar och ser lärarens egna engagemang i eleverna, i ämnet och i höga förväntningar på eleverna som en av de mest kraftfulla resurserna, vilket gör att studiens resultat i sin helhet inte kan sägas vara generaliserbart. En språkdidaktisk diskussion förs som pekar på ”att lärandets språkliga dimension är ett ämne som endast ganska nyligen fått uppmärksamhet i samhället och i skolorna i Sverige. Den språkliga mångfalden i många svenska klassrum har dock satt fingret på en öm punkt.” (ibid., 2010:158 ff.). Kouns (2014) undersöker vidare i en interventionsstudie hur fysiklärare kan utveckla och pröva språk-inriktade undervisningsaktiviteter för att ge eleverna möjlighet till samtal och skrivande med ämnesinnehåll. Här berörs relationen mellan språk och ämnesundervisning och förändringar i hur språk, ämne och undervisning kan samspela och begreppen ”talking science och talking science education” används (ibid., 2014:259, Lemke 1990). Resultaten av interventionsstudien visar såväl svårigheter men också möjligheter där kontextualisering, stöttning och språklig interaktion förutsätts för att få ämne och språk att samverka och samspela (Kouns 2014:253 ff.).

Elevers skrivande i olika ämnen studeras och rapporteras i *Strategiska skribenter. Skrivprocesser i fysik och svenska* (Randahl 2014). Där följer Randahl tre elevers skrivande i både i skolan och hemma och undersöker om, och i så fall på vilket sätt, en viss ämneskultur kan påverka eller ”tvinga fram” olika skrivprocesser (ibid., 2014:37). Resultaten visar att när eleverna skriver en laborationsrapport börjar de att skriva direkt medan det går en ganska lång tid, som eleverna beskriver som en ”inspirationsfas”, innan de tar sig an att skriva en text i svenskan kopplad till realismen (ibid., 2014:202 ff.).

I en språkdidaktisk studie av *Klassrummets semiotiska resurser* (Hipkiss 2014) undersöks hur ämnesspråket i hem- och konsumentkunskap, kemi och biologi

¹⁸ s.37 Skolverket (2014) *PM: Betyg och studieresultat i gymnasieskolan 2013/2014*. Elever med svensk bakgrund är elever födda i Sverige och med minst en förälder född i Sverige.

¹⁹ s.10 Skolverket (2015) *Betyg och studieresultat i gymnasieskolan år 2016*, Skolverket (2016a) *PISA 2015. 15-åringars kunskaper i naturvetenskap, läsförståelse och matematik* och Skolverket (2016d) *TIMSS 2015. Svenska grundskoleelevers kunskaper i matematik och naturvetenskap i ett internationellt perspektiv*.

realiseras i undervisningen samt i samtal mellan elever och lärare i årskurs 7-9. Undersökningen visar att den muntliga kommunikationen i klassrummen ofta sker på vardagsspråk och ämnesspråket företrädesvis synliggörs i läromedlen. Eleverna får receptiva kunskaper men tränar inte i tillräckligt hög grad på att använda det ämnesspecifika språket produktivt och därmed ”saknas samtal, diskussioner, påpekanden om skillnader mellan vardagligt och ämnesspecifikt språkbruk” (ibid., 2014:241).

Det naturvetenskapliga språket i läromedel för högstadiet och gymnasiet undersöks ur flera aspekter i avhandlingen *Steg för steg* (Ribeck 2015). Studien är genomförd med korpuslingvistiska metoder och statistiska beräkningar där flera mått på läsbarhet och texters komplexitet har undersökts utifrån olika mått på syntaktisk komplexitet (meningslängd, nominalfrastäthet, nominalfraslängd, parsingsdjup, ibid., 2015:242 ff.). Resultaten visar en avsevärd ökning av komplexitet i läromedelstexterna mellan stadierna gällande både måtten för komplexitet och ordförråd, framförallt i de naturvetenskapliga ämnestexterna. Studien innehåller inte analys av, eller jämförelser med, elevtexter.

Ordtypsmodellen som utvecklades i *OrdL*-projektet (2007) har i en vidareutvecklad form använts i forskningsprojekt som till exempel Kanebrant *et al.* (2015). De mer finfördelade ordtypskategorierna har kunnat beläggas genom evidensbaserad formativ bedömning av elevers ordkunskap.

I en annan studie undersöks språkliga drag i de naturvetenskapliga uppgifterna i TIMMS och vilken betydelse dessa kan ha för elevers möjligheter att lösa uppgifterna (Persson 2016). Resultaten visar på skillnader mellan de olika ämnens språkkonstruktioner vilket betyder att de naturvetenskapliga språken inte kan betraktas som ett enhetligt språk utan som olika språk med vissa gemensamma språkliga drag (ibid., 2016:85). Vidare visar studien att användningen av passiv form innebär svårigheter för flera elever, liksom att vokabulären har en stor betydelse för elevers förståelse av provuppgifterna.

5.2.1 Nominaliseringar

I elevtexter kan *nominaliseringar* ses som ett steg mot ett mer specialiserat språkbruk och är därför en av de ordkategorier som undersökts i denna studie. Nominaliseringen är, i enlighet med *Svenska Akademiens Grammatik*, *nomen actionis* ”substantiv som prototypiskt är avlett av verb som i sin grundbetydelse anger verbets aktion, t.ex. rengöra: rengöring, reducera: reduktion, veta: vetande” (SAG 1:201, Teleman et al. 2010).

I *Svensk universitetsgrammatik för nybörjare* beskriver Josefsson (2001) nominaliseringar som substantiv som har gjorts om genom tillägg av suffix till en verbrot, exempelvis tilläggen -ing/-ning, -ande/-ende, -het, -skap, -dom (ibid., 2001:56–57).

Lindberg (2007) redogör för hur text kan distanseras från konkret verklighet till att bli mer abstrakt och koncentrerad när handlingar, processer och egenskaper görs om till substantiv genom nominaliseringar (ibid., 2007:78).

Till skillnad mot det vardagliga språket som talas i familjen och i vardagliga sammanhang packas det vetenskapliga språket ihop och reduceras ju mer specialiserat språkbruk som används. Det vetenskapliga språkets framväxt kan kopplas till behovet av att skapa tydliga tekniska taxonomier och en ökad användning av nominaliseringar var beroende av varandra då tekniska termer kan skapas genom nominaliseringar (af Geijerstam 2006:75).

Begreppet *grammatiska metaforer* är, enligt den systemisk-funktionella grammatiken, ett sådant sätt att packa ihop deltagare och processer (Halliday & Matthiessen 2014). De talar om ”inbäddade satser” (min översättning). Genom att konstruera substantiv av processer skapas även tekniska termer som kan vara nominaliseringar vilket inte skapar en tydlig kongruens (grammatisk överensstämmelse). Det är ”en samlingsterm för icke-typiska sätt att uttrycka betydelse, exempelvis nominaliseringen *inlämning* kan uttryckas med verbet *lämna in*” (se Magnusson 2011:2). För andraspråkselever kan den implicita kopplingen i en grammatisk metafor vara mycket svår att förstå (Halliday & Martin 1993, Magnusson 2011).

Med den systemisk-funktionella lingvistik som teoretiskt ramverk undersöks i avhandlingen *Skolspråk i utveckling* (Magnusson 2011) flera typer av grammatiska metaforer (förkortat som GM) i nationella provtexter skrivna av gymnasieelever. Magnusson skiljer mellan *nominaliserad egenskap, process, fas, modalitet, omständighet (tid), kausalitet, adjektiv* och *verb* (ibid., 2011:97). I materialet är nominalisering av process vanligast, eleverna i svenska B på gymnasiet skriver fler GM-täta texter än eleverna i år 9 och tendensen i studien visar att det är stora skillnader mellan en- och flerspråkiga elever där de förstnämnda använder GM i en högre utsträckning och skillnaderna ökar med ålder (ibid., 2011:225 ff.).

Nygård Larsson (2011) undersöker i sin avhandling av texter i biologiämnet *ideationella grammatiska metaforer* som rör nominaliseringar och participkonstruktioner ”där processer uttrycks genom ett substantiv och inte genom ett verb” (ibid., 2011:62). Även Holmberg och Karlsson (2006) diskuterar den ideationella metaforen – systemet där deltagare och process packas ihop och hur grammatiska strukturer på så vis samverkar (ibid., 2006:163).

I en undersökning av elevsvaren i svenska TIMMS 2007 för åk 8 (Frändberg 2012, Johansson Kokkinakis & Frändberg 2013) granskades nominaliseringar som ett lingvistisk mått utöver tekniska termer och passiva verb. De hänvisar även de till det naturvetenskapliga specialiserade språket med en högre grad av abstraktion och skriver att nominaliseringar ”är en användbar mekanism för definitioner och klassificeringar i naturvetenskap” (ibid., 2013:57). I båda studierna ingår analys av en läromedelstext i kemi samt elevsvar från en kemiuppgift i

TIMMS. Andelen nominaliseringar som återfinns i läromedelstexten var högre än den som återfanns i elevtexterna (ibid.).

5.2.2 Verb i passiv form

Verb i passiv form är ett annat steg i ett specialiserat språkbruk som kännetecknar ett mer ämnesspecifikt språk och särskilt i den naturvetenskapliga diskursen är det vanligt att aktören avlägsnas och på så sätt ställs utanför de processer som beskrivs. I en analys av läromedelstexter pekar Hägerfelth (2004) på hur aktörer kan göras passiva genom att händelsen ställs i fokus för läsaren som på så vis positioneras medvetet eller omedvetet genom att inte med säkerhet veta vem som står bakom det som utförs i satsen. Denna förskjutning av processen, händelsen från deltagarna genom användningen av verb i passiv form är ett fenomen som är släkt med den ideationella metaforen (Holmberg & Karlsson 2006).

I undersökningen av elevsvar i svenska TIMMS 2007 för åk 8 (Frändberg 2012, Johansson Kokkinakis & Frändberg 2013) granskades de passiva verben som ett lingvistiskt mått utöver tekniska termer och nominaliseringar. De passiva verben beskrivs som ett kongruent sätt att använda dem på då ”fysikala fenomen i omvärlden förklaras med fysikaliska lagar och kausala samband” (2013:57). Detta görs då det inte finns någon agent bakom händelser i den naturliga fysikaliska världen. I undersökningarna visas att såväl nominaliseringar som passiva verb förekom i så låg grad i elevernas svar att det inte gick att säkerställa statistisk signifikans. I deras undersökning av läromedelstext var 4,1 % av graforden passiva verb, i elevtexterna var resultatet 1,8 % (ibid.).

De flesta så kallade *transitiva verb*, det vill säga verb där händelsen har en aktant som utför något för någon, kan konstrueras med s-form med passiv funktion (SAG 2:506, Telman et al. 2010). Lyngfelt (2007, 2010) beskriver hur man skiljer på två huvudtyper av s-verb-konstruktioner i svenskan, *passiv* och *deponens*, där de senare, *deponens*, inte har passiv funktion utan kan indelas i fyra grupper, exempel ges i parentes; *absoluta* (knuffas, luras), *reciproka* (kivas, samsas), *mediala* (förenas, möts*) och *intransitiva* (yvas, vistas), (ibid., 2007:111 ff.).

5.2.3 Akademiskt språk, akademiska ord

Utgångspunkten för CLISS-projektet har varit att fokusera på skolans ”akademiska” skriftspråk som på många sätt skiljer sig från det vardagliga och informella språket. Det kännetecknas av att det är ett skriftspråksrelaterat kontextoberoende och tankemässigt mer krävande språk än det informella samtalspråk som barn och ungdomar utvecklar i vardagslivet (se exempelvis Schleppegrell 2004, Johansson Kokkinakis & Lindberg 2007).

Tillgången av, och förståelse för ett akademiskt språk, det vill säga användningen av ord och uttryck som är mer frekventa i akademiska texter än i exempelvis tidningstext eller skönlitterär text, är något som betonas och diskuteras. Schleppegrell beskriver det som att:

It is important for students to develop academic register options in different disciplines because particular grammatical choices are functional for construing the kinds of knowledge typical of a discipline. (Schleppegrell, 2004:137)

Det akademiska ordförrådet är en resurs som elever och studenter, oavsett modersmål, behöver ha tillgång till och som kan vålla problem för dem som inte har det (Sköldberg & Johansson Kokkinakis 2011). Detta är något som liknas vid en portvakt, en ”Gatekeeper”; om man har tillgång till det akademiska språket får man komma in, annars riskerar man att bli stående utanför.

”The ability to use the specialized language of the academic sphere is like a gatekeeper: it opens doors for those who have it, and fastens them shut for those who do not.” Stanford Teacher Education Program. (Bailey 2007: ix)

Söker man vägledning om *akademiskt språk* och/eller språkbruk på Skolverkets hemsida finns begreppet med i olika kursförordningar för gymnasiet, angående behörighetskrav för lärare, som en hänvisning till Olssons avhandling inom CLISS-projektet (2016) samt en hänvisning till området Skolutveckling där en text nämner boken *Språkinriktad undervisning* (Hajer & Meestringa 2014). Det är, utifrån min sökning (våren 2017), den enda referensen, förutom Olssons (2016) där Skolverket på sin hemsida berör begreppet akademiskt språk gällande alla elevers språkutveckling.

Hajer och Meestringa (2014) beskriver begreppet som centralt i det didaktiska arbetet för att utveckla elevernas språk; ”Undervisningen ska utveckla elevers kunskaper i akademiskt språkbruk eftersom många elever inte har möjlighet att möta ett sådant språk utanför skolan” (ibid., 2014:37). De använder begreppet *Akademiskt språk* som en underrubrik till begreppet *Skolspråk* (2014:8) vilket är något som även Lindberg (2006) gör med hänvisning till den anglosaxiska forskningen där man skiljer på en mer social språkfärdighet som bygger på det informella vardagliga språket och en akademisk språkfärdighet som förutsätter ett mer formellt ”akademiskt” språk (ibid., 2006:87).

I en kunskapsöversikt *Att läsa och förstå* (Skolverket 2016c) hänvisas till en studie där förståelsen för det akademiska språket kopplas till att utveckla läsförståelse (se Ucceli et al. 2015). Innebörden av ett akademiskt språk har mer handlat om att kunna ämnesspecifika ord än att få språkliga verktyg som hjälper elever att erövra det som i studien benämns *Core Academic Language Skills* (ibid., 337-356). Dessa, förkortade som *CALS*, innefattar att känna igen ett akademiskt

register genom sex olika förmågor som tränas vid läsning av akademiska texter (ibid., s. 349).

Begreppen ”*Academic Language*” och ”*Academic Vocabulary*” har internationellt använts i forskningssammanhang där böcker och kurser handlar om *English for Academic Purposes, EAP* (Bailey 2007, Coxhead 2002, Hyland & Tse 2007, Nation 2013, Paquot 2010, Schleppegrell 2004).

Definitionen av begreppet *Academic Language* liksom avgränsningen av *Academic Words* diskuteras bland forskare utifrån de olika typer av akademiska ordlistor för olika språk och ändamål som har skapats och som används runt om i världen. Hyland och Tse (2007) menar att en högre förekomst av vissa ord i akademiska texter än i andra texter är bekräftat liksom specifika språkliga drag men att det därmed inte går att skapa en allmän akademisk ordlista. Vidare menar de att frekvensen och fördelningen är en av de stora svårigheterna med den här sortens listor;” [...] a major difficulty of such lists, and not just the AWL (*The Academic Word List*), is the assumption that a single inventory can represent the vocabulary of academic discourse and so be valuable to all students irrespective their field of study” (ibid., 2007:238). Deras undersökningar pekar på att täckningsgraden kan vara väldigt ojämn beroende på ämnesdisciplin samt att olika akademiska ämnesfält använder och skapar olika vokabulär för att passa olika diskurser (ibid., 2007:240). De diskuterar därför om det över huvud taget kan sägas finnas en akademisk vokabulär, ”By considering context, context, and use, academic vocabulary becomes a chimera” (Hyland & Tse, 2007:250) och framhåller i och med detta vikten av att ordlistor och konkordanser hämtas från de språkliga register inom vilka studenterna förväntas läsa och skriva (ibid. 2007:251).

Utifrån Hyland och Tses forskning och ställningstaganden förordar Granger och Paquot ”a happy medium approach” (2009:8, Paquot 2010:26) i sitt framtagande av en ordlista för produktiva syften där orden kan förekomma bland de 2000 vanligaste orden i engelska men även förekomma och vara frekventa i akademiska texter. Paquot (2010) definierar begreppet: ”[A]cademic vocabulary would be more usefully defined as a set of options to refer to those activities that characterize academic work, organize scientific discourse and build the rhetoric of academic texts” (ibid., 2010:28).

Definitionen av begreppet akademiska ord eller akademisk vokabulär menar Paquot (2010) bör baseras på språkforskning som visar att det är ord som är användbara vid beskrivningar av aktiviteter som karaktäriserar akademiskt arbete, det vill säga, forskning, analys och utvärdering.

Andra begrepp man möter i engelskspråkiga forskningsartiklar om akademiska ord/vokabulär är *sub-technical vocabulary*, *semi-technical vocabulary*, *specialized non-technical lexis* (Nation 2013, Hyland & Tse 2007). Dessa ses representera det språkbruk som används i högre grad i det akademiska registret än i andra texter. Enligt Nation (2013) görs ofta en tredelad indelning av vokabulären i akad-

emiska texter i: ”general service or basic vocabulary, sub-technical vocabulary and technical vocabulary [...] although this distinction ignores low-frequency vocabulary with no sub-technical features” (Nation, 2013:289).

När behovet av språklig kompetens ökar som det gjort i vårt alltmer globaliserade och teknikbaserade samhälle, kan diskussionerna om begreppsdefiniering och täckningsgrad tyckas vara något som handlar mer om positionering. Att olika akademiska fält har olika diskurser och därmed olika språkliga register förefaller rimligt och de specialiserade vetenskapliga ord som tillhör respektive fält möter inlärare oftast via skriven text där texten ingår i undervisningssekvenser vilket leder till en receptiv ordinlärning som sedan när studenterna själva förväntas skriva akademiska texter i bästa fall har resulterat i en produktiv inlärning. Problemet är de ord som utgör stommen av texten och får den att hänga samman där *kollokationer* (två eller flera ord som hör samman i satser) och *konnektorer* (bindeord) kan vara sådana ord som studenter förväntas kunna men som i det akademiska språkbruket är andra än i ett vardagligt språkbruk liksom *homografer* (ord som stavas likadant men som har olika betydelse) och *polysema ord* (flertydiga, relaterade ord). Nation (2013) menar att dessa ord inte alltid är så väl kända för inläraren trots att de inte tillhör den specialiserade vokabulären som anses svår utan är ord som ligger mellan de vanligaste och de specialiserade.

Olsson (2016), som inom CLISS-projektet har undersökts skrivuppgifterna som gymnasieeleverna har skrivit på engelska, skiljer på *domänspecifika* och *allmänakademiska vokabulär* med hänvisning till Bauman & Graves artikel *What is Academic Vocabulary?* (2010). Olsson analyserar den allmänakademiska vokabulären som ”ord som förekommer i många discipliner men mer sällan i icke-akademisk kontext” (ibid., 2016:100).

I svensk kontext har *En svensk akademisk ordlista* (Jansson et al. 2012, Sköldb- berg & Johansson Kokkinakis 2012), utarbetats vid Institutionen för svenska språket vid Göteborgs universitet som en elektronisk och fritt tillgänglig resurs (fortsättningsvis förkortad som *ESAO*).²⁰ Den baseras på *SveAk*-korpuserna, en *Svensk akademisk korpus* (Borin et al. 2012) som består av 25 miljoner ord från avhandlingar och vetenskapliga artiklar inom humaniora och samhällsvetenskap. Till skillnad från de ovan nämnda engelska korpusarna baseras *ESAO* (ibid. 2012) på lemman istället för ordfamiljer. Den tänkta målgruppen är ”universitetsstudenter med olika språklig bakgrund” (Jansson et al. 2012). Med hänvisning till de senast rapporterade PISA-resultaten för svenska grundskoleelever (Skolverket 2016c) kan målgruppen komma att utökas till att inbegripa alla högskole- och universitetsstudenter i Sverige då olika metoder för att utveckla elevers receptiva och produktiva språkliga färdigheter eftersöks. För närvarande finns 652 huvudord i *ESAO* (ibid. 2012). Dessa är ordnade enligt ”ett fallande

²⁰ <https://spraakbanken.gu.se/ao/>

akademiskt index med följande ord i topp: *dock, studie, beskriva, social, enligt, innebära, samt, form, betydelse, fall*’ (se Ribbeck 2015:27).

Inom CLISS-projektet har Lim Falk och Holmberg (2016) undersökt 260 elevtexter skrivna på svenska från skrivuppgift 1 och 3 (delar av materialet i denna studie). De har gjort en kvantitativ täckningsundersökning av andelen akademiska ord utifrån *ESAO* (ibid., 2012) samt en kvalitativ studie av 36 elevskrivna texter utifrån ett SFL-perspektiv (*Systemisk-funktionell lingvistik* se Holmberg, Karlsson & Nord 2011). I täckningsanalysen ingår samtliga 652 huvudord från listan (för jämförelse se avsnitt 7.3.3). I den kvalitativa analysen av de 36 texterna ges exempel på vilka av orden som har använts mest och i vilka sammanhang. Ett annat resultat i deras undersökning är att användningsgraden av akademiska ord har ökat i alla grupperna från den första skrivuppgiften som skrevs termin 1 till den tredje från termin 4 (Lim Falk & Holmberg 2016).

Med allt högre krav på skriftspråklig kompetens är det viktigt att studera elevers skriftliga produktion i sådana texttyper som är relevanta i olika utbildningssammanhang (se Andersson Varga 2014, Randahl 2014, m.fl.).

5.3 Kvantitativa mätmetoder för textanalys

Textanalys av normativa texter har en lång tradition. En av de tidiga och mest kända läsbarhetsformlerna är Irving Lorges från 1939.²¹ Genom att räkna och mäta olika textuella enheter har texter analyserats, läsbarhet diskuterats och med datorers hjälp har ytterligare mått på textenheter i allt större material kunnat mätas och jämföras (Biber et al. 1998, Laufer & Nation 1995, Melin & Lange 2000, Nation 2001 m.fl.).

I boken *Läsbarhet* (1968) presenterade Björnsson ett sätt att mäta språklig tillgänglighet i brukstexter som han kallade *läsbarhetsindex*, LIX.²² Det räknades ut genom att det genomsnittliga antalet ord per mening i en text adderas med procentalet för långa ord (ord med fler än sex tecken). Med mätmetoden kunde texter bedömas vara allt från mycket lättlästa till mycket svår byråkratsvenska beroende på det LIX-tal som texterna fick.

Kritik mot LIX har framförts, bland annat av Westman (1969), som menade att en tillförlitlig metod för liknande beräkningar” måste föregås av omfattande undersökningar av ords och ordformers och satsstrukturers frekvenser, kort sagt skilda språkmönster” (ibid., 1969:8, se vidare 6.1).

I avhandlingen, *I see what you mean* (2013) visar Heimann Mühlenbock hur nya språkteknologiska metoder och elektroniska språkresurser kan bidra till att mer

²¹ Den bestod av tre komponenter: genomsnittlig meningslängd, andel prepositionsfraser och andelen svåra ord.

²² <http://www.lix.se/>

exakta läsbarhetsanalyser kan genomföras. Resultaten är en språkmodell för svenska, kallad SVIT, som i studien visar högre noggrannhet än LIX gällande delar av beräkningarna som exempelvis textkomplexitet (ibid., 2013:159 ff.).

Inom svensk kvantitativt orienterad textlingvistik är Hultmans och Westmans stora undersökning *Gymnasistsvenska* en av de mer omfattande studier som har genomförts (1977). I studien har man kartlagt ordförråd, ordklassfördelning, syntax och grafiska förhållanden i texter från gymnasieelevers centrala prov i årskurs 3 (151 uppsatser) samt vuxentexter, som samhällsinformerande broschyrer (25 st.), allmänt informativa tidningsartiklar (28 st.) och utdrag ur läroböcker (20 texter). Dessa innehöll tillsammans 88 757 graford (1977:52–53). Undersökningen utgick från två stora projekt på 70-talet, *Skrivsyntax* och *Talsyntax*. Alla ordräkningar utfördes med hjälp av dator.

Syftet med studien var att kunna göra jämförelser mellan de olika textmaterialen för att se om gymnasieungdomarna ”behärskade den slitstarka bruksprosan” (ibid. 1977:11). Dessutom ville man se hur textlängd korrelerade med betyg samt med kön. För att kunna göra jämförelser mellan andelen olika ord i förhållande till textlängd i olika textmaterial, skapades måttet ordvariation, *OVIX*.²³ För att en undersökning av *OVIX* skall fungera behöver texterna vara längre än 200 ord då måttet räknas fram med en logaritmisk formel (Hultman 1994:60).

I en forskningsöversikt av *Skrivteori och skrivforskning* (2006, reviderad 2010), placerar Blåsjö undersökningen av Hultman & Westman i ett eget kapitel, *Svensk elevspråksforskning*, tillsammans med Larssons (1984) *Skrivförmåga*. Blåsjö menar att deras undersökningar kan ses som ”sprungna ur ett paradigmskift mellan den grammatisk-stilistiska forskningen med en mer kognitivistisk syn på väg mot en mer social språksyn” (Blåsjö 2010:28).

Vidare nämner Blåsjö arbetsgruppen för nationella prov vid Uppsala universitet som över tid har fortsatt med kvantitativa undersökningar av elevtexter men som även studerar fler aspekter av elevers skrivande (se exempelvis Nyström 2000, Nyström Höög 2010, Olevard 1999, Östlund-Stjärnegårdh 2002).

Nyströms (2000) data utgår från de nationella proven på gymnasiet där antalet insamlade provtexter var 936 men där en intensivgrupp bestående av 42 elever har intervjuats i studien. Texterna är skrivna som olika texttyper, delar av resultaten kommer att refereras till gällande textlängd och betyg (se avsnitt 5.3.1). I Östlund-Stjärnegårdhs studie (2002) har 60 gymnasieelevtexter hämtats från den nationella provgruppens arkiv där urvalet är statistiskt säkerställt genom Skolverkets urvalssystem (2002:35 ff.). Studiens material omfattar 31 124 graford.

²³ *OVIX* räknas ut med formeln $\log(\text{tokens})/\log(2-(\log(\text{types})/\log(\text{tokens})))$

OVIX skapat av Tor G Hultman, anger kvoten mellan olika ord och det totala antalet ord i

En text med kompensation för textens längd. $V=N(2-Nk)$. N är antalet löpord (även graford, antalet ord som finns i en text) och V är antalet lexikonord (orden som finns i en text). k är en konstant och $OVIX=1/k$.

Flera studier finns i form av *FUMS*-rapporter (*Forskning och utbildning i modern svenska*). Här tillkommer också forskning om bedömning av elevtexter då den nationella provgruppen i Uppsala utformar bedömningsstöd till lärare som underlag vid genomförandet av de nationella proven i svenska i både grundskolan och gymnasieskolan. (Palmér 2013, Palmér & Östlund-Stjärnegårdh 2005).

En fortsättning i den kvantitativt undersökande texttraditionen är en uppsats på avancerad nivå även den från Institutionen i nordiska språk i Uppsala (Bratt 2014). Här undersöks 24 elevtexter från de nationella proven i Svenska B på gymnasiet och jämförelser görs av olika textuella variabler mellan provuppsatser från 1996 (7 st.), 2003 (7 st.) och 2013 (10 st.). Motiveringen för att ta med en studentuppsats här är att det som refereras till är de textuella mätningarna som möjliggör jämförelser över tid.

I tre andra studier som refereras till består elevtextmaterialet av nationella provtexter från årskurs 9 (Ciolek Laerum 2009, Nordenfors 2011, Olevard 1999). Både Ciolek Laerum och Olevard har hämtat sitt material från de insamlade nationella provtexterna för årskurs 9 som varje år skickas in till nationella provgruppen i Uppsala utifrån elevers födelsedatum.²⁴ Olevards undersökning omfattar 60 elevtexter där 15 är skrivna av flickor och 15 av pojkar från provet år 1987 och lika många från provet 1996 (1999:2). Totalt består de 60 undersökta texterna av 34 285 graford. Den andra studien av nationella prov i svenska för årskurs 9 består av ett urval av 40 elevtexter som omfattar totalt 27 650 graford (Ciolek Laerum 2009). I den tredje studien som refereras till här ingår ett longitudinellt textmaterial bestående av olika typer av texter insamlade från 31 elever mellan deras skolår 5 till 9, däribland nationella provuppsatser (Nordenfors 2011:53).

Det norska KAL-projektet (Evensen 2003), *Kvalitetssikring av læringsutbyttet i norsk skriftlig*, har 3368 elevtexter från 179 gymnasieavgångsklasser undersökts och analyserats av en stor forskargrupp. Texterna omfattar 2,1 miljoner graford.²⁵ Insamlingen av texter har gjorts under två olika tidsperioder, 1998 och 1999 samt 2000 och 2001 för att se förändring över tid samt för att möjliggöra jämförelser av examenstexter som har skrivits under två olika läroplaner (Evensen 2003:3-4).

I Lim Falks (2008) granskning av en CLIL-klass textproduktion görs även jämförelser med material insamlat i en kontrollklass liksom i denna studie. I elevtextundersökningen ingår olika textmaterial från kemi, fysik, historia, religion och svenska, även jämförelser med de nationella provtexterna i svenska görs. Antalet elevtexter som ingår i Lim Falks studie är 44 stycken varav 25 är skrivna av CLIL-elever och 19 av elever i kontrollklassen (ibid., 2008:229).

²⁴ Slumpvisa datum som anges i Lärarens informationshäfte till de Nationella proven i de olika provämnena.

²⁵ <http://www.forskningsradet.no/csstorage/vedlegg/evensen.pdf>

I en tvåårig studie i Storbritannien undersöktes vad som benämns som *linguistic development* (språklig utveckling) hos elever i åldrarna 12-15 år (Myhill 2008). Fem olika aspekter som omfattade totalt 31 variabler analyserades. En sammanfattning av resultaten var att se skrivutveckling som att gå från något till något annat, med en ökad medvetenhet om hur ord, fraser och meningar kan formas och förändras beroende på uppgifter, krav och behov hos mottagare (Myhill 2009:407).

5.3.1 Kvantitativa mätmetoder och bedömningsaspekter

Vid kvantitativa undersökningar görs ofta korrelationsanalyser mellan olika textuella mått och hur elevskrivna texter sedan bedöms för att se om måtten har betydelse för bedömning. Vid bedömningen av elevuppsatserna i *Gymnasist-svenska* (Hultman & Westman 1977) visade sig interbedömarreliabiliteten vara låg, dock var överensstämmelsen om uppsatserna med betygen 1 och 5 hög, däremot rådde diskrepans om mellanbetygen.²⁶ Resultaten visade att flickorna skrev klart längre uppsatser än pojkarna, hög korrelation fanns mellan textlängd och betyg upp till och med betyget 4, sedan konstaterar forskarna att skillnaden mellan de högsta betygen snarare är beroende av kvalitet än av textlängd. (ibid., 1977:53). Korrelationen mellan betyg och olika ord var högre än den mellan antalet ord och betyg; 0,61 mot 0,57 det vill säga, undersökningen visade att större ordvariation, *OVIX*, hängde ihop med högre betyg men att det snarare, enligt forskarnas tolkning, var ”variationer i informationsrikedom än synonymvariationer som ordvariationsmättet mäter” (ibid., 1977:60).

Andra studier där sambandet mellan textlängd och betyg har undersökts bekräftar en positiv korrelation mellan antalet ord i en text och hur den bedöms (se exempelvis Larsson 1984).

Josephson, Melin och Oliv (1990) analyserade tio elevtexter från åk 1 till 9 skrivna av både flickor och pojkar med olika förutsättningar som skribenter. De gjorde både strukturella och funktionella analyser med avsikten att ge lärare redskap och utgångspunkter för att arbeta med elevers skrivande och skriftspråk. Däremot fanns ingen korrelation till betyg med. I diskussionen om ett varierat ordförråd nämner de även mätmetoden *OVIX* som en variabel och de menar ”att vara precis är att skriva varierat” (ibid., 1990:45).

I Nyströms avhandling *Gymnasisters skrivande. En studie av genre, textstruktur och sammanhang* (2000) där undersökningsmaterialet består av flera olika skrivuppgifter från gymnasieelever, är ett resultat i studien att textlängden för de olika texttyperna har ett samband med betyg men att textlängden är relaterad till den

²⁶ Betygssystemet 1 (lägst) – 5 (högst) enligt dåvarande läroplan *Lgy 70*.

texttyp som valts samt att även ordvariationen påverkas av typen av text. Diskussionen förs av Nyström vilket begrepp, *texttyp* eller *genre*, som bäst beskriver skoltexter i svenskämnet där det diskursiva och narrativa skrivandet dominerar (2000:234). Olikheter mellan textlängd i förhållande till skrivuppgiften är något som även Vagle (2005) lyfter fram i sina kvantitativa analyser av texterna i KAL-projektet (2005:327ff.). Även Nordenfors (2011) visar detsamma i sin longitudinella studie.

Ett annat resultat som Nyström (2000) redovisar angående textlängd är att den påverkas av skrivverktyget, de datorskrivna texterna är betydligt längre än de handskrivna (ibid. 2000:175). Nyström pekar även på att skrivforskning flyttat fokus från textens enheter till sambandet mellan text och skrivande där skrivprocessen, den didaktiska skrivforskningen i Norge samt genrepedagogiken, haft stor betydelse för förändringen (se exempelvis Berge 1988, Dysthe 1996).

Att använda textlängd och betyg som en av flera analysmodeller görs även av Östlund-Stjärnegårdh i hennes avhandling *Godkänd i svenska? Bedömning och analys av gymnasieelevers texter* (2002). Östlund-Stjärnegårdh gör jämförelser med Hultmans och Westmans undersökning (1977) och påvisar en utveckling över tid där eleverna med stigande ålder verkar skriva allt längre texter. Elever med lägre betyg skrev tidigare kortare texter men i Östlund-Stjärnegårdhs (2002) undersökning är även deras texter längre. Orsaker därtill menar hon kan vara dels att provsituationerna har förändrats över tid liksom att elevkategorierna har förändrats i och med att gymnasiet erbjuder fler programval. Granskningen bekräftar också att högre betyg och längre texter hör ihop (ibid., 2002:75 ff.).

Inom *SUF*-projektet (*Språk och språkbruk i flerspråkiga storstadsmiljöer*, 2001-2006) har Johansson Kokkinakis och Magnusson (2008) genomfört en studie där de jämför elever med svenska som modersmål och elever med flerspråkig bakgrund, något som inte har gjorts med kvantitativ textlingvistik tidigare. Deras material är 200 elevuppsatser skrivna som nationella prov på gymnasiet och de har undersökt *lexikal densitet*, *nominalkvot*, *ordvariation* och *ordlängd*. Resultaten visar att alla mått korrelerar starkt med betyg utom lexikal densitet samt att måtten korrelerar med varandra och ordlängd är det mått som har den starkaste korrelationen. Studien visar också att elever som skriver på sitt förstaspråk, i det här fallet svenska, skriver uppsatser med högst nominalkvot och lexikal täthet, ett resultat som även har framkommit i andra studier (ibid., 2008:54).

I *FUMS*-rapportserie (*Forskning och utbildning i Modern Svenska*, vid Uppsala universitet) har två studier som behandlar elevers uppsatser och lärares bedömning av dessa, genomförts. Olevard (1999) utförde en kvantitativ undersökning av nationella provuppsatser för årskurs 9 från åren 1987 och 1996. För att göra studien longitudinell gjordes en liknande undersökning av årskurs 9:s nationella provuppsatser från 2006 av Ciolek Laerum (2009). Urvalet till studien består av 40 uppsatser varav 20 har skrivits av pojkar och 20 av flickor. Texterna har skrivits in i *Word* av externa bedömare med de skrivfel och andra normbrott som

fanns i de handskrivna originalen för att kvantitativa beräkningar skulle kunna göras. Olika aspekter av uppsatserna har analyserats. Resultaten överensstämmer med Östlund-Stjärnegårdhs undersökning (2002) på det sättet att uppsatser som har skrivits har blivit längre över tid och i att textlängd och betyg korrelerar. ”MVG-texterna innehåller i genomsnitt tre gånger fler ord än texter som bedömts EUM” (ibid. 2002:33).²⁷

I ovanstående rapportserie har Nyström Höög (2010) gjort en senare studie där hon med flera textlingvistiska metoder undersöker texter från de nationella proven i svenska för årskurserna 5 och 9. Resultaten visar att de äldre eleverna skriver längre texter med längre meningar och mer varierad textstruktur medan de yngre eleverna skriver mer tydligt argumenterande. Detta kan kanske förklaras utifrån instruktioner och formuleringar i undervisningen men att båda grupperna skulle gynnas av en mer explicit skrivundervisning (ibid., s. 55 ff.).

Vid institutionen för nordiska språk i Uppsala har Bratt (2014) genomfört en studie av elevuppsatser med högsta betyg från de nationella proven i svenska, på gymnasiet. Urvalet är gjort med sju års mellanrum och är tänkt att representera de tre olika kursplanerna som har varit gällande under perioden från 1996 till 2014.²⁸ Materialet består av 24 slumpmässigt valda uppsatser, sju från vardera 1996 och 2003 samt tio från 2013. Materialet är undersökt med nio olika variabler varav denna studie gör jämförelser med fyra av dem; *textlängd*, *ordlängd*, *andel långa ord* och *ordvariationsindex*. Jämförelsen som görs av bedömning av uppsatserna över tid och i olika betygssystem visar att det finns förändringar i de drag som över tid kännetecknar textkvalitet. Trots att flera av variablerna visar en nedgång över tid är andelen elever som får högsta betyg konstant (ibid., 2014: 30).

Andra studier som har använt kvantitativa mätningar är Lim Falks (2008), som nämnts tidigare, och Nordenfors (2011). Lim Falk undersökte om elever i CLIL-klasser på gymnasiet ”tar ett ämnesanknutet språkbruk i bruk på svenska, när, till vad och i vilken utsträckning och på vilket sätt (ibid., 2008:22). I analysen av elevtexter från en CLIL-klass och en kontrollklass undersökte Lim Falk bland annat hur textlängd korrelerade med betyg (ibid., 2008:228 ff.). Resultaten visade att textlängd korrelerade med betyg i kontrollklassen men att CLIL-eleverna över lag skrev längre texter. För betyget MVG innehöll deras texter andra aspekter än textlängden som förklarar ett högre betyg.²⁹

Nordenfors (2011) använder begreppen *textlängd*, *ordmängd/ordmedellängd*, *ordvariation (OVIX)* för att undersöka elevtexter, totalt 266 texter av nio olika texttyper, skrivna av 31 elever i åk 5-9 mellan åren 2003 och 2007, för att se vilka föränd-

²⁷ På grundskolan motsvarades gymnasiebetyget IG av omdömet EUM=ej uppnått målen.

²⁸ *Lpf94*, de reviderade kursplanerna 2000 och *Lgy11*.

²⁹ Betyg enligt *Lpf94*, IG= icke godkänd, G=godkänd, VG= väl godkänd, MVG=mycket väl godkänd.

ringsmönster som uppvisas i texterna. Genus har inte ingått som en variabel (ibid. 2011:10). Resultaten visar att textlängd och *OVIX* ökar under grundskoleåren för eleverna i studien men avtar efter årskurs 8. Olika texttyper ger varierande resultatvärden, vilket överensstämmer med Nyströms (2000) resultat liksom Vagles (2005). Korrelationen med betyg överensstämmer inte med tidigare studier hos Nordenfors (2011) vilket leder till slutsatsen att textmått i studien är problematiska som mått på skrivutveckling liksom att texttyp är en variabel man måste ta hänsyn till (2011:247–248). I Nordenfors studie undersöks även *röstbegreppet*.

Internationella studier där kvantitativa lingvistiska metoder har använts blir svåra att jämföra med då det i engelskan ofta är vanligt med två ord där svenskan för samma ord till ett ord, så jämförelser gällande exempelvis ordmedellängd blir därmed inte tillförlitliga (se Lim Falk 2008:173 ff.).

Huruvida ordräkningar kan och bör relateras till textkvalitet och bedömning kan givetvis diskuteras. Hultman och Westman (1977) menar att trots det triviala med ordräkning (som i deras undersökning av elevtexter är gjord på grafordsnivå) i jämförelse med andra mer sofistikerade analyser, så fann de i sin studie ”att även ganska enkel ordstatistik kan ha mycket att berätta om språk och innehåll” (ibid., 1977:46).

De kvantitativa textstudier av elevtexter som har genomförts av flera forskare vid Uppsala universitet där de nationella proven i svenska görs, visar också på att kvantitativa resultat kan hänga samman med textkvalitet (Bratt 2014, Ciolek Laerum 2009, Nyström 2000, Nyström Höög 2010, Olevard 1999, Östlund-Stjärnegårdh 2002).

Även i norsk textforskning har kvantitativa mätningar av elevtexter gjorts. I det stora KAL-projektet, 2005, där 3364 norska examenstexter har analyserats undersöker Vagle i en delstudie *produktivitet/textlängd (TL)* och *lexikaliskt-semantiskt innehåll genom genomsnittlig ordlängd (GOL)*. Vagle (2005) skriver att textlängden i flera internationella studier visar samvariation med bedömning av text även om produktivitet inte ensamt är avgörande för textkvalitet. Dessutom visas även samvariation med bakgrundsfaktorer som ålder, kön och social bakgrund. Ordmedellängd, visar sig trots sin enkelhet vara ”et godt mål på lexikalisk specifikasjon og variasjon” eftersom det finns ”en klar sammenheng mellom: (1) ords lengde, (2) deres frekvens og (3) deras semantiske tyngde”. Längre ord har, enligt Vagle, mer innehåll än kortare ord (ibid., 2005:305).

I ett nytt stort norskt projekt, *Normprosjektet* (se exempelvis Solheim & Matre 2014, Matre & Solheim 2016, Berge & Skar 2015), har skrivforskare samverkat med lärare för att utveckla nationella förväntningsnormer för skrivkompetens. Projektet pågick mellan 2012 och 2014 och målet har varit att alla lärare skall få kompetens i skrivutbildning. De normer man gemensamt har arbetat fram för elever i årskurs 5 och årskurs 8 kan delas in i två huvudgrupper: 1) *Functional Competencies* och 2) *Coding Competencies* (Matre & Solheim 2016). Till den första gruppen hör normer gällande kommunikation, innehåll, textstruktur och språk-

bruk (där vokabulär ingår). Till den andra gruppen hör stavning och morfologi, interpunktion och hur man använder skrivandet i utförande (ibid., 2016:190). Projektet har innehållit en interventionsstudie med 500 lärare, 3670 elever på 20 skolor där man såg signifikanta positiva resultat vid skrivmomenten, särskilt för de yngre eleverna vid tidiga insatser.

Beträffande bedömning av vokabulär finns följande bedömningsstöd för lärare på mellanstadiet “use a relevant and varied vocabulary, including terms relevant to school subjects” och för högstadiet “use a relevant, varied and precise vocabulary, including discipline-specific terms” (ibid., 2016:202).

5.4 Korpuslingvistiska metoder för textanalys

Lingvistiska analyser över tid möjliggörs av *korpuslingvistiska metoder* där man kan studera och hitta empiriska belägg i stora textmängder, *korpusar*. (Biber et al. 1998, Gardner 2007, Nation 2013, McEnery et al. 2006). Detta görs framför allt med hjälp av datorer och med olika datorprogram (se Nation & Anthony 2016).

En korpus utgörs av en stor samling språklig data som kan vara muntlig, det vill säga inspelad och transkriberad, eller skriftlig, i form av text. Textkorpusar kan bestå av text från tidningar, böcker och ibland omfatta en författares hela författarskap. De kan även bestå av texter från sociala medier, det vill säga av texter som skrivs på nätet, exempelvis i form av bloggar (Språkbanken är en digital resurs vid Göteborgs universitet där texter och lexikon från olika textgener finns fritt tillgängliga för studier och nedladdning).³⁰ En korpus kan variera i storlek. En större korpus gör det möjligt att generalisera runt språkliga fenomen medan en mindre korpus medför att man kan analysera enskilda språkliga mönster och strukturer. För att genomföra grammatisk och syntaktisk analys krävs bearbetning av textkorpusar. Efter sådan bearbetning kan samförekomster av ord i uttryck och i grammatiska konstruktioner enklare beskrivas (se Olsson 2011, Ribbeck 2015). Det här sättet att märka upp texter på metalingvistisk nivå kallas *taggning* eller *annotering*. I en korpus kan man göra lexikala beräkningar, det vill säga man tittar på ordkomponenter som vokaler, konsonanter, morfem, eller textord, textordtyper och grundformer – *lemman* och *lemgram* (ords samtliga böjningsformer; en katt-en-er-erna).

Korpuslingvistik används bland annat när man gör ordlistor. Flera sådana ordlistor har skapats för det som benämnts ‘*Academic Words*’ eller ‘*Academic Vocabulary*’ där grundmaterialet är texter från olika vetenskapliga discipliner som sammanställs till en korpus. Ett tidigt exempel på en ordlista är *A General Service List of English Words* (GSL, West 1953) som innehåller de 2000 vanligaste engelska orden som förekom i en textkorpus på cirka 5 miljoner ord. Efterföljande

³⁰ <http://spraakbanken.gu.se/>

ordlistor baseras på fler texter från fler ämnesområden till exempel *The Academic Word List (AWL)*, Coxhead 2002). En ännu senare efterföljare är *Academic Vocabulary Lists (AVL)*, Gardner & Davies 2013). Den baseras på *COCA, Corpus of Contemporary American English*, en korpus med 120 miljoner ord från akademiska texter skrivna fram till 2012 och hämtade från nio olika discipliner. Till skillnad mot de tidigare ordlistorna är *AVL* indelad i lemman i stället för i ordfamiljer.

Ett användningsområde som följer av att kartlägga förekomst och frekvens av ord i stora korpusar är att skapa ordtest av olika slag för att göra kvantitativa bedömningar av ordförråd hos elever eller studenter i diagnostiskt syfte eller som intagningsprov till olika utbildningar. (Nation 2013, Laufer & Nation 1995). Den svenska *OrdiL*-korpusen (Lindberg & Johansson Kokkinakis 2007, se avsnitt 5.2), har utöver huvudsyftet som var att kartlägga ordförrådet i läromedels-texter, också använts för att skapa olika typer av ordtest för att mäta ordförråd hos språkanvändare.

Inom *OrdiL*-projektet utvecklades även en *datorbaserad lexikalisk profil* (Johansson Kokkinakis 2007). En sådan innebär att man kan beskriva hur orden i en text ”representerar olika frekvensområden i förhållande till en utvald frekvensordnad ordlista” (ibid., 2007:174, Laufer & Nation 1995, Laufer 2005, Meara 2005). En frekvensordnad ordlista kan till exempel skapas med hjälp av en balanserad textkorpus där normativa texter som representerar svenskt skriftspråk ingår. Man kan vid lexikal profilering även använda andra lexikala mått som *lexikalisk originalitet*, *lexikalisk densitet* och *nominalkvot* för att göra mer detaljerade undersökningar av texter. Genom att göra detta kan man ”mäta” ordförrådets storlek, bredd och djup hos personer för att skapa förståelse för omfång och tillgänglighet hos språkanvändare (Johansson Kokkinakis 2007:175).

Några av de studier som med korpuslingvistiska metoder undersöker elevspråk och gör jämförelser med läromedelspråk och/eller provspråk har gjorts av Nygård Larsson (2011), Frändberg (2012, 2013) och Persson (2016). Studier där fokus ligger på läromedelspråket i läromedel för olika stadier har gjorts av exempelvis Edling (2006) som undersökt kvalitativa egenskaper av ord och texter och Ribbeck (2015) som undersökt kvantitativa aspekter i läromedeltexter för högstadiet och gymnasiet.

6. Metod

För att kunna besvara forskningsfrågorna om hur en vokabulär i skrift som är skolrelaterad, ämnesrelaterad och akademisk, framträder i gymnasieelevers texter på svenska samt om det finns samband i den vokabulär som framträder som kan relateras till faktorerna: CLIL/icke-CLIL, genus samt kursbetyg i ämnet svenska, och hur dessa i så fall ser ut, har nedanstående metoder använts.

I denna studie definierades och karakteriserades ordförrådet inledningsvis utifrån traditionella textlingvistiska mått som har brukats i tidigare studier (Hultman & Westman 1977 m.fl.). Därefter användes korpuslingvistiska metoder för att definiera och identifiera specifika lingvistiska variabler som enligt en systemisk-funktionell språksyn kännetecknar ett mer vetenskapligt språkbruk (Halliday 1993, Holmberg & Karlsson 2006, Schleppegrell 2004 m.fl.). Lexikala profiler har använts för att synliggöra ord med hög- respektive låg förekomst och statistiska analyser för att studera resultaten som har framkommit har utförts med SPSS.

I en studie som denna, av elevskrivna texter, blir vissa av de textlingvistiska måtten inte reliabla, exempelvis meningslängd och läsbarhetsindex (LIX, se avsnitt 6.1). Detta då elever, till skillnad mot läromedelsproducenter och provkonstruktörer, helt enkelt inte alltid sätter ut några skiljetecken, vilket innebär att vissa meningar kan bli orimligt långa och därmed förvanska resultat.

De ämnesrelaterade orden visade sig vara de svåraste att avgränsa utifrån en validitetsaspekt eftersom inga jämförande texter som läromedel, nationella provtexter eller provkonstruktioner har kunnat utnyttjas vid analysen och avgränsningen av ord. Här har det alltså inte funnits fastställda korpusar med annat textmaterial att tillgå för att på så vis se relationen mellan elevernas ordförråd och den vokabulär som realiserats i exempelvis läromedelstext, vilket har gjorts i

andra undersökningar (se Edling 2006, Frändberg 2012, Johansson Kokkinakis & Frändberg 2013, Lindberg & Johansson Kokkinakis 2007, Persson 2016).

De ord som här kan betraktas som ämnesrelaterade ord utöver ordkategorierna nominaliseringar, passiva verb och akademiska ord, är vissa av de ord som vid den lexikala profileringen faller utanför de 5000 av de mest frekventa orden i en frekvensordnad ordlista. Framför allt gäller detta ord med endast en förekomst, vilket synliggörs i figur 1, 2, 20 och 21. Dessa ord kan ses höra samman med textrubrikerna för de fyra skrivuppgifterna; *Hälsa och västånd* (SO), *Bilfri stad* (NO), *Folkhälsoproblem* (NO) och *Helt delad föräldrapeng* (SO). Då en valid avgränsning inte kan göras så redovisas de inte som en egen ordkategori, däremot finns de representerade bland lågfrekventa ord i redovisningen av täckningsgrad i lexikala profiler, se avsnitt 7.4.1. De elevskrivna texterna har alla undersökts genom statistiska analyser men även genom manuell analys, se vidare i resultatdelen, kapitel 7.

Nedan tydliggörs de analysmetoder som har används i studien.

TABELL 6. *Översikt av analysmetoder och analys typer*

Kvantitativ textanalys genom ytstrukturella mätningar av:	textlängd ordmedellängd andel långord ordvariationsindex - OVIX
Kvantitativ ämnesrelaterad textanalys genom mätningar av:	nominaliseringar verb i passiv form akademiska ord
Lexikala frekvensprofiler genom undersökningar av:	lågfrekventa ord högfrekventa ord

I kapitlet beskrivs inledningsvis i avsnitt 6.1 hur de kvantitativa ytstrukturella mätningarna av textlängd, ordmedellängd, andelen långord och ordvariationsindex har gjorts. Därefter beskrivs i avsnitt 6.2 hur de kvantitativa ämnesrelaterade mätningarna har genomförts för att kunna beräkna andelen nominaliseringar, andelen verb i passiv form och andelen akademiska ord i elevtexterna. Där presenteras även hur elevtextkorpusarna, de frekvensbaserade ordlistorna och de ämnesspecifika ordlistorna har skapats för att kunna analysera elevtexterna. I avsnitt 6.3 beskrivs arbetet med lexikala frekvensprofiler och avslutningsvis i 6.4 beskrivs hur materialet har studerats med statistiska analyser.

I studien analyseras 520 elevtexter med sammanlagt 276 898 graford (tabell 10).

6.1 Kvantitativa ytstrukturella mätningar

De lingvistiska variablerna för en skolrelaterad skriftspråklig vokabulär har identifierats med deskriptiva undersökningsmetoder. De 520 elevtexterna har undersökts med kvantitativa textlingvistiska mätningar av *antalet ord per text = textlängd*, *ordmedellängd (antalet tecken delat med antalet ord)*, *andelen långord/långordsfrekvens (ord med >6 tecken av totala antalet ord)* samt *ordvariation, OVIX* (formel för beräkning se avsnitt 5.3). Till detta har i huvudsak datorprogrammet *LIX-räknaren* använts.³¹ Den baseras på Björnssons (1968) uträkningar för läsbarhetsindex av brukstexter i första hand och är en fri programvara som är enkelt sökbar på Internet. Programvaran LIX användes för att få reda på hur långa texter eleverna skrev, hur många meningar deras texter innehöll, antalet graford/tokens i texterna, andelen långord (mer än 6 tecken) samt texternas ordvariationsindex. Uträkningen av *ordmedellängd* fick göras separat genom att antalet tecken delats med antalet ord för varje elevtext. Alla resultat har förts in i Excelark och senare i statistikprogrammet, SPSS. Medelvärden för grupper av elever, såväl för klasserna som för CLIL/icke-CLIL samt för individer, flickor (F) och pojkar (P) i respektive grupp har också noterats.

Måttet LIX, läsbarhetsindex (som är ett av flera mått som rapporteras i programvaran med samma namn) för att med beräkningar kunna se olika texters svårighetsgrad bör alltså *inte* användas på elevtexter. Om elever undviker att sätta ut kommatecken kan detta medföra LIX-tal som pekar på att elevernas texter når tal som anger ”mycket svår byråkratsvenska” trots att texterna består av fem rader men endast en eller två meningar. Det är något som har lett till att just LIX-talen har kritiserats (se exempelvis Westman 1969, Lundberg & Reichenberg 2008). I några av elevtexterna förekommer detta, att elever skriver extremt långa meningar utan skiljetecken, så LIX-talet används inte i studien, dels på grund av brist på reliabilitet, dels på grund av brist på jämförande material i form av läromedelstext eller standardiserat provmaterial (gällande LIX se även Heimann Mühlenbock 2013, Ribeck 2015). I stället följer jag tidigare undersökningar, se avsnitt 5.3, och refererar och diskuterar studiens resultat i förhållande till dessa i avsnitt 8.2.

6.2 Kvantitativa ämnesrelaterade mätningar

De grammatiska variablerna för en ämnesrelaterad vokabulär har identifierats med hjälp av korpuslingvistiska metoder (Biber 1998, McEnery et al. 2006). De grammatiska variablerna i studien innefattar *nominaliseringar*, *passiva verb* och *akademiska ord*. De sistnämnda utgår ifrån en *Svensk akademisk ordlista, ESAO* (Jans-

³¹ <http://www.lix.se>

son et al. 2012, Sköldberg & Johansson Kokkinakis 2012). Resultaten av identifieringen av de grammatiska variablerna har analyserats med statistiska metoder.

6.2.1 Elevtextkorpusar

Skrivuppgifterna som undersökts ligger i mappar som Wordfiler i CLISS – projektets databas och är benämnda med uppgiftsnummer; text 1, text 2, text 3 och text 4. I varje textmapp har texterna dessutom ett elevnummer så att vi i projektet genom att använda våra kodnycklar har kunnat följa varje deltagande elev och se vilka uppgifter de har deltagit i. Texterna ligger i en följd för varje deltagande klass så att klasserna lätt har kunnat avgränsas som grupper vid olika analyser.

För att kunna analysera elevtexterna skrivna på svenska som ingår i denna studie, skapades först två korpusar av alla elevtexterna som textfiler. Elevtexter tillhörande skrivuppgift 1 och 4 med SO – inriktning (*Hälsa och välbefinnande* och *Helt delad föräldrapeng*) fördes samman i en gemensam textfil och alla elevtexter tillhörande skrivuppgift 2 och 3 med NO – inriktning (*Bilfri stad* och *Folkhälsoproblem*) samlades i en annan textfil som två olika elevtextkorpusar, fortfarande med elevnumren kvar för att kunna se vem som skrivit vilken text. Fortsättningsvis kallas dessa för Korpus 1 (skrivuppgift 1 och 4 – SO) och Korpus 2 (skrivuppgift 2 och 3 – NO). I båda korpusarna ingår 260 elevtexter vilket gör att korpusarna kan sägas vara jämförbara storleksmässigt (se avsnitt 7.1.2).

6.2.2 Ordlistor

De specifika ordlistorna som har skapats utgår från korpusarna som satts samman av alla elevtexter från de fyra skrivuppgifterna. Den första korpusen, Korpus 1, innefattar elevtexterna om *Hälsa och välbefinnande* och *Helt delad föräldrapeng*. Korpus 2 består av elevtexterna *Bilfri stad* och *Folkhälsoproblem*.

Nästa steg i undersökningen var att alla elevtexterna i de två korpusarna, Korpus 1 och Korpus 2 fördes samman till varsin stor ordlista i programmet *AntConc 3.2.4w* (Nation & Anthony 2016).³² Genom att använda funktionen *Word List* så listades alla graford i en lista efter förekomst och frekvens där det synliggjordes hur många förekomster varje ord hade i hela korpusen. Denna fördes sedan över till ett *Excelark*.

För att kunna utveckla de specifika ordlistorna som senare har använts för att undersöka täckningsgraden av de olika ordkategorierna, har alla graford förs över i Excelark där de kommit att listas efter förekomst med antalet förekomster angivet framför varje ord. På så vis har lexikonord och engångsord synliggjorts.

³² <http://www.laurenceanthony.net/software.html>

6.2.3 Utveckling av specifika ordlistor

För att skapa specifika ordlistor har alla ord i respektive korpus granskats manuellt och kategoriserats enligt de ordkategorier som kunnat utskiljas i elevtexterna och som kan definierats som kännetecknande för ett skolrelaterat akademiskt språkbruk. De grammatiska variabler som har karaktäriseras i elevtextmaterialet är *nominaliseringar*, *verb i passiv form* och *akademiska ord*. Utgångspunkten var ordkategorier som i tidigare forskning identifierats känneteckna ett vetenskapligt och akademiskt språk (se kapitel 5). Genom att granska alla orden okulärt flera gånger har de kategorier som ingår i studien urskiljts och därefter förts in i egna listor i Excelark.

Då texterna i denna studie inte analyseras i sin helhet har avgränsningen av nominaliseringar, för att skapa en analysordlista, gjorts manuellt enligt *nomen actionis* (SAG 1 s. 201; SAG 3 s. 31 ff., Teleman et al. 2010). För att kunna avgränsa nominaliseringar i elevtexterna har detta beaktats med ett ”omskrivningstest” samt att kriterierna form, betydelse, funktion och etymologi har använts (exempel på nominaliseringar i elevtexter se bilaga 7).

Verb i deponensform (se Lyngfelt 2007, 2010) har inte tagits med i den analysordlista med verb i passiv form som använts i studien för att identifiera passiv användning i elevtexterna. Här har SAG:s lista över deponensformer varit vägledande och dessa har, i de fall de har förekommit i elevtexterna, tagits bort från analysordlistan (exempel på verb i passiv form från elevtexter se bilaga 9). Exempel på några sådana som har tagits bort från elevtexterna är (här i infinitiv): *fattas*, *finnas*, *frodas*, *hoppas*, *lyckas* och *synas* (SAG 2, Verb § 48, s. 554 – 555, Teleman et al. 2010).

Det visade sig i denna studie att i korpusarna av elevtexter använde eleverna inga namn (utöver geografiska namn och personnamn) eller formler, inte heller några tekniska termer och korpusarna var för små för att delas in i två undergrupper av vardagliga ord. Det som skiljer denna studie från andra där ordkategorier som är relaterade till ämne har undersökts är att där har kontrolltexter som lärobokstexter och/eller provtexter ingått, vilket har inneburit att det har funnits en form av kontrolltext att mäta emot. Exempel på sådana studier är Lindberg & Johansson Kokkinakis 2007, Lim Falk 2008 och Johansson Kokkinakis & Frändberg 2013. Detta är något som måste finnas för att avgränsningarna för ordkategorierna skall kunna anses vara valida, antingen en kontrollvokabulär eller tydliga grammatiska avgränsningar vilket här har använts för ordlistorna för nominaliseringar, verb i passiv form och akademiska ord.

Efter att de olika ordkategorierna hade identifierats och markerats i respektive tabell i Excelarket sammanfördes respektive lista till ett Worddokument som sedan sparades som oformaterad text (*.txt) i en textfil med annan kodning, *Unicode UTF8* samt infogning av specifik radbrytning för att skilja orden åt. Därefter har dessa ordlistor sparats både i Word och som textfiler för vidare arbete.

För att kunna studera förekomst och distribution av ordkategorierna i varje elevtext har alla 520 texter analyserats mot varje ordlistalista i programmet *Ant-WordProfiler* (Nation & Anthony 2016).

Alla elevtexter har redovisats med täckningsgraden för respektive ordkategori och med ett procenttal som har förts in i ett Excelark där alla elever funnits med sina kodnummer. På så vis har varje deltagande elev fått resultat för hur många procent av uppsatsens graford som har utgjorts av nominaliseringar, passiver, akademiska ord och hög- respektive lågfrekventa ord, (avgränsningarna för dessa kategorier se avsnitt 7.3).

6.2.4 Akademiska ord

Undersökningen av akademiska ord har i denna studie utgått från *En svensk akademisk ordlista*, fortsättningsvis förkortad som *ESAO* (Jansson et al. 2012, Sköldbberg & Johansson Kokkinakis 2012), som har utarbetats vid Institutionen för svenska språket vid Göteborgs universitet. Den baseras på *Sve.Ak-korpusen*, en *Svensk akademisk korpus* (Borin et al. 2012), som består av 26 miljoner ord från avhandlingar och vetenskapliga artiklar inom humaniora och samhällsvetenskap som publicerats 1997–2012. Korpusen är fritt tillgänglig via korpusinfrastrukturen, *Korp* i Språkbanken. *ESAO* (ibid., 2012) baseras på lemman istället för ordfamiljer som traditionellt sett har använts i internationella akademiska ordlistor. Detta innebär att den omfattar ordklass, böjning, betydelser, redigerat språkprov och engelsk översättning för de första 100 orden, som är ordnade efter så kallat ”akademiskt index, ett mått som speglar ordens frekvens, utbredning och spridning i korpusen” (Ribeck 2015:27). Följande ord listas som högst rankade: dock, studie, beskriva, social, enligt, innebära, samt, form, betydelse, fall (ibid.).

För närvarande finns 652 huvudord i ordlistan som är jämnt fördelade mellan ämnena i korpusen och är mer frekventa i den akademiska korpusen än i en referenskorpus, *LäSBarT* (se Heimann Mühlenbock 2009). *LäSBarT* -korpusen innehåller lättlästa skönlitterära texter, nyhetstexter och samhällsinformation. Den akademiska ordlistan (*ESAO*, 2012) innehåller inte heller ord som återfinns bland de 1000 vanligast förekommande orden i *LäSBarT* -korpusen.³³

I föreliggande studie gjordes inledningsvis en manuell analys av andelen akademiska ord från de båda Korpusordlistorna 1 och 2 mot *ESAO* (ibid., 2012). Därefter fördes dessa över till en gemensam lista som akademiska ord och undersöktes mot ordlistorna med högfrekventa ord som har använts i studien för de lexikala profilerna. De ord i *ESAO* (ibid.) som också återfinns bland de 5000 av de mest frekventa i svenska språket har tagits bort så att samma ord inte re-

³³ <https://spraakbanken.gu.se/swe/resurs/lasbart>

dovisas vid flera tillfällen vid olika undersökningar. Detta är ett beslut som gör att denna studie av texterna och avgränsningen för akademiska ord skiljer sig från undersökningen av elevtexter av skrivuppgift 1 och 3 som gjorts av Lim Falk och Holmberg (2016) där alla ord som förekommer i *ESAO* (ibid., 2012) har undersökts. I studien av Johansson och Frändberg (2013) har ord kunnat tillhöra flera ordkategorier och därmed ingå i andel i procent som redovisas vilka då har kunnat överstiga 100 % men så är inte fallet här.

I och med detta blir förekomsten av akademiska ord som redovisas i denna studie som andel i procent av andelen graford i texterna, lägre än i Lim Falk och Holmbergs (2016) undersökning. Då flera av orden i *ESAO* (ibid., 2012) är *allmänakademiska/allmänspråkliga* och därmed inte förekommer enbart i akademiska texter så återfinns dessa i granskningen med lexikala profiler. Antalet akademiska ord som inte finns bland de 5000 av de vanligast förekommande orden i svenska språket och vars förekomst finns i *ESAO* (ibid.) är totalt 363 ord av de 652 som här representerar den akademiska vokabulären, några av dem redovisas i bilaga 9.

6.3 Lexikala profiler

Elevtexterna har även analyserats genom en lexikalisk profilering där täckningsgraden för olika ords förekomst visas med procenttal. Denna form av täckningsanalys bygger på korpusar där språkliga företeelser representeras och visar distributionen av ordanvändning (se avsnitt 7.4).

Undersökningen med lexikalisk frekvensprofil visar de 1000 – 5000 av de vanligast förekommande orden på svenska som baseras på en bloggkorpus som innehåller 80 miljoner ord (*Bloggmix* 2012).

Begreppet *token coverage* anger andelen av texten som täcks av de 5000 mest frekventa orden i svenskan utifrån denna lista. I frekvensband 1 visas förekomst och distribution för 1000 av de mest frekventa orden i korpusen, för frekvensband 2 förekomst och distribution av de ord som återfinns bland de 2000 av de mest frekventa orden och så vidare upp till de 5000 av de mest frekventa orden i baserat på den frekvensordnade ordlistan. Den lexikala frekvensprofileringen genomfördes genom användning av programvaran *AntiWordProfiler* (Nation & Anthony 2016)³⁴ samt den frekvensordnade ordlistan och därmed skapas frekvensband i *Level List*. Därefter analyserades en elevtext i taget (som textfil) i *Users File*. Genom analysen har varje elevtext fått en lexikalisk profil med alla frekvensbanden angivna som procent av textens helhet. Alla resultat har förts in i Excelark.

³⁴ <http://www.laurenceanthony.net/software.html>

Figur 1 och 2 visar två elevtexter där frekvensbandsundersökning gjorts och färgerna i texterna representerar olika förekomst och distribution för frekvensband 1-5 med olika procentsatser. Varje färg representerar ett band med 1000 ord som omfång. I det första bandet, *Level 1* finns ord som: *och, jag, det, att, är, på, i*. I det andra bandet, *Level 2* finns ord som: *skapa, chans, död, packa, känsla*. I det tredje bandet, *Level 3* finns ord som: *grön, skiva, politiken, ämnet*. I det fjärde bandet, *Level 4* finns ord som: *jord, väst, sidorna, artisten*. I det femte bandet, *Level 5* finns ord som: *evigheter, släktingarna, klassiker, försäljning, helger*. Ord­längden, andelen långord synliggörs här genom att orden blir längre och därmed inte lika vanligt förekommande i språket ju högre frekvensbanden blir. De mest frekventa orden är en- eller tvåstaviga.



FIGUR 1. *Lexikal profil för icke-CLIL-elev, skrivuppgift 4*

I elevtexten som visas i figur 1 illustreras att 95,9 % av täckningsgraden återfinns i *Level 1-5*. De övriga 4,1 % markeras svarta i texten och procentsiffran visas på *Level 0* som även benämns *off-list*. Dessa är alltså ord vilka inte finns i den frekvensbaserade ordlistan. Dessa kan vara ord som är geografiska namn, ord som är felstavade eller ord som representerar någon av de ordkategorier som undersöks vidare i denna studie. Här bör påpekas att några av orden, exempelvis *Sverige* och *delad*, vilka är svartmarkerade i både figur 1 och 2, kan tyckas vara fre-

kventa i vårt språk men är det inte utifrån denna korpus (*Bloggmix* 2012), dock synliggörs orden i *Level 0*. I figur 2 finns fler ord som tydligare kan sägas vara lågfrekventa.

Elevtexterna i figur 1 och 2 är valda för att visa hur skillnaderna kan se ut gällande de lexikala profilerna. Båda texterna är från samma skrivuppgift, nummer 4, med samma rubriksättning men sedan skiljer sig vokabuläranvändningen synligt åt.



FIGUR 2. Lexikal profil för CLIL-elev, skrivuppgift 4

I figur 2 har eleven använt 77,7 % av de 5000 mest frekventa orden i den frekvensbaserade ordlistan. 62,1 % av orden återfinns bland de 1000 vanligaste orden, 5,5 % av de 2000 vanligast förekommande, 4,4 % av ord som finns bland de 3000 mest frekventa, 3,2 % av orden i det fjärde frekvensbandet och 2,5 % av ord bland de 5000 mest förekommande i detta frekvensband. Bland *off-listorden*, de ord som är svartmarkerade i elevtexten i figur 2, finns ord som: *basala*, *filantropiska*, *påtvingade*, *patriarkala*, *tudelad*, *bevåg* med flera. Dessa utgör 22,3 % av alla ord i texten vilket är en betydande del.

Genom att göra denna form av analys och sedan göra jämförelser med de vokabulärlistor som har gjorts speciellt för denna studie kan antaganden om ele-

vernas allmänspråkliga kompetens gällande skriftspråk göras. I avsnitt 8.3 diskuteras resultaten som syns i figurerna vidare.

6.4 Statistiska analyser

Resultaten av de olika kvantitativa och korpuslingvistiska mätningarna har alla analyserats statistiskt. De resultat som har registrerats är medelvärden, median, max- och minimivärden. För att göra jämförelser på gruppnivå mellan CLIL och icke-CLIL har *Independent Samples T-test* använts med *Levene's Test for Equality of Variances*. För att jämföra och analysera resultaten mellan de fyra undergrupperna CLIL (F) och (P) och icke-CLIL (F) och (P) har *Oneway ANOVA* genomförts med efterföljande *Tukey HSD post hoc-test*. Detta för att se signifikanta skillnader i prestation mellan alla grupper.

För att synliggöra resultaten av de statistiska analyserna används *stapeldiagram* för jämförelse av de två grupperna CLIL och icke-CLIL och *lådagram* (eng. *box-plots*) för att illustrera de fyra gruppernas resultatfördelning. De visar de fyra olika grupperna med olika färger: pojkar CLIL – mörklila, flickor CLIL – ljusare lila, pojkar icke-CLIL – mörk grön och flickor icke-CLIL – ljusare grön, gällande spridning och frekvens i form av max- och minimieresultat, medianvärde samt eventuella *uteliggare* (eng. *outliers*.) Uteliggare är värden som avviker från övriga gruppen och därför hamnar utanför en låda. Även standardavvikelse (*standard deviation*, SD) redovisas.

Lådagrammen, där själva ”lådan”, som i figurerna markerats med olika färger för de olika grupperna, omfattar 50 % av värdena. I den syns även medianen för respektive grupp i varje uppgift samt övre och undre kvartilen det vill säga det 75 % värdet och det 25 % värdet. Dessutom visas maxvärde respektive minimivärdena som även kallas variationsvidd (eng. *range*). Dessa är de tvärgående yttre strecken i varje lådagram. I vissa fall syns stjärnor eller prickar som ligger utanför dessa, de kallas uteliggare (eng. *outliers*) och betraktas i SPSS som extremvärden (Djurfeldt et al. 2010:63). Standardavvikelsen (SD) är måttet på hur mycket värden avviker från medelvärdet. Ett högt värde på SD innebär att en grupp har stor spridning i sina resultat och kan betraktas som heterogen medan ett lägre värde är ett tecken på en mer homogen grupp.

För att illustrera allt detta i de fyra grupperna synliggörs resultaten i de fyra skrivuppgifterna i lådagram, där CLIL (P) representeras av den första boxen från vänster sett, följt av CLIL (F) därefter icke-CLIL (P) och så icke-CLIL (F) längst åt höger. Förkortningarna PC (pojkar CLIL), FC (flickor CLIL), PnC (pojkar icke-/non-CLIL) och FnC (flickor icke-/non-CLIL) används också i resultatredovisningen.

Här bör nämnas att elever som syns som uteliggaren med ett nummer vid en prick eller en stjärna i lådagrammen, *inte* har samma nummer där som i CLISS-

databasen. Detta då SPSS-numren hänger samman med sorteringen av klasserna i variabelinformationen i programvaran och blir därmed unika elevnummer i den här studien vilket ger ytterligare validitet.

För att besvara uppsatsens om hur en svensk vokabulär i skrift framträder i gymnasieelevers texter har olika analysmetoder av variabler som har beskrivits i kapitel 6 utförts. Vokabulären som har studerats är skolrelaterad, ämnesrelaterad och akademisk. Avsikten har också varit att synliggöra om det finns samband i den vokabulär som framträder som kan relateras till faktorerna CLIL./icke-CLIL och/eller genus. Dessa frågor besvaras i resultat- och diskussionskapitlen.

7. Resultat

Resultatkapitlet inleds med en sammanställning av studiens kvantitativa data där relevantaste redovisas klassvis och med fördelningen av flickor respektive pojkar i de deltagande klasserna, CLIL och icke-CLIL (se avsnitt 7.1). Därefter visar tabell 8 skrivuppgifterna på svenska som ingår i studien, med dess rubriker samt antalet texter (N) från eleverna som har skrivit de olika uppgifterna vilka undersöks i studien.

Därpå följer resultaten av elevernas skrivuppgifter som avser besvara studiens frågeställningar. Till att börja med besvaras frågan om hur en skolrelaterad skriftspråklig vokabulär framträder utifrån kvantitativa ytstrukturella mätningar av *textlängd*, *ordmedellängd*, *andel långord* och *ordvariation* (se avsnitt 7.2). Sedan redovisas hur elevernas ämnesrelaterade vokabulär har urskiljts med korpuslingvistiska metoder, samt resultaten för de olika kvantitativa ämnesrelaterade mätningarna av *nominaliseringar* och *verb i passiv form* (avsnitt 7.3.1 och 7.3.2). Vidare redovisas hur de *akademiska orden* har identifierats, även de undersökta med korpuslingvistiska metoder, samt resultaten för denna del (avsnitt 7.3.3). Varje del inleds med resultat där medelvärdena för grupperna CLIL och icke-CLIL jämförs med varandra vilket synliggörs med *stapeldiagram*, sedan följer resultaten för CLIL/icke-CLIL flickor respektive pojkar som redovisas med tabeller och i vissa fall även med *lådagram* (se avsnitt 6.4). Vissa individuella resultat lyfts fram för att illustrera spridning i grupperna.

Till varje delresultat redovisas även de statistiska analyser som har gjorts. Resultaten och jämförelser med tidigare studier diskuteras i kapitel 8.

7.1 Resultatöversikt

I studien undersöks 520 texter. Av studiens elever har 62 % utgjorts av CLIL-elever och 38 % av kontrollelever. Elevantalet i klasserna på de tre deltagande skolorna (skola 1/A, 2/B, 3/C) har varierat mellan 19 elever i den minsta klassen och 37 elever i den största. Fördelningen mellan flickor och pojkar skiljer sig mest i CLIL-klasserna där det har varit en övervägande del flickor medan det i två av kontrollklasserna har varit en nästan helt jämn fördelning. Den tredje kontrollklassen följer fördelningen av fler flickor än pojkar (se tabell 7).

TABELL 7. *Deltagande skolor samt elevantal och fördelning av flickor och pojkar utifrån tillhandahållna klasslistor.*

Skola	1-A	F/P	2-B	F/P	3-C	F/P
Program, klass/antal elever						
EK CLIL					19	15/4
SA CLIL	37	26/11			28	20/8
NA CLIL	30	19/11	33	23/10		
NA icke- CLIL			33	21/12		
EK icke-CLIL					A-31	16/15
					B-26	13/13

7.1.1 Elevtexter

Tabell 8 visar antalet deltagande elever totalt vid de olika skrivuppgifterna samt hur fördelningen ser ut mellan CLIL-elever och icke-CLIL elever samt mellan könen (F) och (P). I CLIL-klasserna har fler flickor än pojkar deltagit. I kontrollklasserna är det motsatta förhållanden, vilket återspeglar fördelningen mellan flickor och pojkar i CLIL-klasserna respektive kontrollklasserna.

TABELL 8. *Antalet deltagande elever vid de fyra skrivuppgifterna samt fördelning av CLIL, icke-CLIL, flickor och pojkar.*

	Skrivuppg. 1	Skrivuppg. 2	Skrivuppg. 3	Skrivuppg. 4
Deltagare totalt	N=135	N=134	N=126	N=125
	57 %	56,5 %	53,2 %	52,7 %
Varav flickor/pojkar	88/47	88/46	88/38	81/44
CLIL/icke-CLIL	80/55	84/50	75/51	67/58
Varav flickor/pojkar				
CLIL	57/23	63/21	58/17	54/13
Icke-CLIL	31/24	25/25	30/21	27/31

Sammanlagt ingår skrivuppgifter från 189 elever i studien varav 153 uppgifter är skrivna av flickor och 84 av pojkar där några elever har skrivit flera skrivuppgifter. Deltagarantalet har legat mellan 52,7 % och 57 %, med det högsta deltagandet vid första skrivuppgiften (termin 1) och det lägsta antalet deltagare vid det sista skrivtillfället (termin 5), se tabell 8 och avsnitt 2.1.1.

Tabell 9 visar antalet elever som har skrivit alla fyra skrivuppgifterna. Av de 189 eleverna har 63 elever skrivit alla fyra skrivuppgifterna på svenska, 43 av dem är flickor och 20 är pojkar fördelade över alla de deltagande klasserna. 43 elever av de 63 som har skrivit alla fyra skrivuppgifterna går i CLIL-klasser och 20 elever går i någon av kontrollklasserna. De som har skrivit alla skrivuppgifterna på svenska i projektet utgör 33 % av det totala deltagarantalet. I tabell 9 syns fördelningen av de elever som har skrivit alla skrivuppgifterna.

TABELL 9. *Antalet elever som har skrivit alla fyra skrivuppgifterna*

Antal elever som har skrivit alla fyra texter	63	CLIL/ icke-CLIL
Flickor/Pojkar	43/20	40/23
Flickor/Pojkar		
CLIL	32/8	
icke-CLIL	11/12	

7.1.2 Skrivuppgifter och korpusar

Av skrivuppgifterna med nummer 1 – *Hälsa och välbefinnande* och 4 – *Helt delad föräldrapeng*, båda med en SO-inriktning, har en delkorpus, *Korpus 1* gjorts, där alla elevtexter i form av ökad text har förts samman till ett dokument. Samma sak har gjorts med alla elevtexter som har skrivits till uppgifterna 2 – *Bilfri stad* och 3 – *Folkhälsoproblem* med NO-inriktning, dessa utgör då *Korpus 2*. Detta synliggörs i tabell 10 där de två korpusarna beskrivs utifrån antalet texter som ingår liksom antalet graford och lexikonord. Skrivuppgift 2 och 3 finns som bilaga 3 och 4.

TABELL 10. *Statistiskt underlag för studien med korpusar, graford och lexikonord*

Skrivuppgift (se bilagor)	1. Hälsa och välbefinnande (SO)	4. Helt delad föräldrapeng (SO)	2. Bilfri stad (NO)	3. Folkhälsoproblem (NO)
Korpusar	Korpus 1 – SO 260 texter		Korpus 2 – NO 260 texter	
Graford (antal ord)	129 315		147 583	
Lexikonord (antal unika förekomster)	7 509		10 396	

Korpus 1, som utgörs av skrivuppgifterna 1 och 4 med en SO-inriktning består av 260 elevskrivna texter som har undersökts på olika sätt.

I Korpus 1 finns 129 315 olika ord, graford, flera av dessa återkommer frekvent. De tre vanligast förekommande orden är: *att, det* samt *och* med 5082, 4048 respektive 3668 gånger som de förekommer i texterna.

Korpus 2 utgörs av skrivuppgifterna 2 och 3 med NO-inriktning och består även den av 260 elevskrivna texter. I denna korpus finns 147 583 olika ord, graford. De vanligast förekommande är här: *att, och* samt *är* som förekommer 5771 gånger respektive 4750 och 3509. Ordet *att* är alltså det vanligast förekommande ordet i båda korpusarna.

Korpus 2 är större än Korpus 1 trots att båda korpusarna innehåller samma antal elevtexter, 260 stycken i båda. Orsaken till detta är troligtvis ämnesrelaterad. De längsta texterna skrevs i skrivuppgift 3 – *Folkhälsoproblem* där texterna också visar sig i högre grad vara personliga och mer berättande. Även skrivuppgift 2 – *Bilfri stad*, visar liknande uttryck där ämnet också har engagerat eleverna på ett mer personligt plan och då skrev de längre texter.

Det finns naturligtvis även ord som återfinns endast en gång i korpusarna (*engångsord*). I studien finns det 3831 olika ord (av 129 315) med endast en förekomst i Korpus 1. Några exempel är: *acceptabel, adopterat, basala, befästa, diverse, övertid* och *överväldigande*. När man arbetar med en korpus så kan man sortera och filtrera och få graforden i alfabetisk ordning och med antalet förekomster så att man sedan lätt kan leta och hitta i materialet.

I Korpus 2 finns följande exempel på ord med endast en förekomst: *acceleration, administrativa, alzheimer, bantningsmedel, och klyftpotatis* för att nämna några. I denna korpus, som alltså omfattar fler elevskrivna texter, är det därmed fler graford som förekommer endast en gång i korpusen (5638 av totalt 147 583 graford).

Av de ord som förekommer enbart vid ett tillfälle i någon av de elevskrivna uppgifterna finns även namn, både geografiska men också personnamn, dessutom ett antal felstavade ord som eleverna, trots att de har skrivit på dator med rättstavningsprogram, inte har rättat till. Inte sällan är det bokstavsväxlingar som kan kopplas till felslag på tangentbordet, liksom att en bokstav har utelämnats i tvåbokstavsord, men felstavningar förekommer också. Några exempel på felstavningar från Korpus 1 är: *aggerar, aktuväll, anlednig, anorlunda, arbetslös, ekonmi, förelidren, medlämsländerna, hämska, igenkligen, jämnställd, orakerna, osraker, probelmen, repsektea, rättivst*. Felstavningar är dock få med tanke på materialets storlek. I Korpus 1 återfinns 238 felstavade ord av engångsorden, vilket motsvarar 6 % av de 3831 orden som endast förekommer en gång men dessa utgör 0,002 % av antalet graford, det vill säga av alla orden i Korpus 1.

Flertalet engångsord kan ses höra samman med skrivuppgifternas ämnen. Exempel på sådana är: *artificiell, arvs massa, barnadödlighet, barnafödandet, befolkningsmängd, biologiskt, bromsmedicin, farföräldrar, friskola, fysiken, geografisk, geologisk, hand-sprit, humanitära, högteknologiutbildade, immun, koldioxidutsläppen, lattepappa, miljöfak-*

torer, segregering, teknologisk, tendensen, vårdnadshavare med flera. Engångsorden markeras som svarta i frekvensbandsundersökningen tillsammans med ord som utifrån den frekvensordnade ordlistan är att betrakta som *off-list-ord* eller *lägfrekventa ord*. Orden har vid manuell granskning av Excelarken och frekvensbilderna undersökts om de är ämnesrelaterade eller den typ av felstavade graford som getts exempel på ovan. Då andelen felstavade graford är relativt lågt liksom andelen namn, såväl personnamn som geografiska namn, kan ett antagande göras att en övervägande del av de ord som synliggörs i *Level 0* är ämnesrelaterade. Dock kan ett valitt resultat av ämnesrelaterade ord inte fastställas i föreliggande studie eftersom ingen jämförande korpus av läromedel eller provfrågor med ämnesspecifika och ämnesrelaterade ord finns (se figur 1, 2, 20, 21, 23 och 24).

En viss sammanblandning av engelsk och svensk stavning går att se i graford som *action, corrupt, guerillor, iraq, money, notes, overcomes, peace, pointerade, supreme, survival*. Det finns även några få exempel på sammanskrivning av graford som *påtal, pågrundav, tillsdessatt*. Även dessa svartmarkeras i de lexikala profilerna och har granskats manuellt.

7.2 Kvantitativa ytstrukturella mätningar

I avsnitt 7.2 presenteras de kvantitativa mätningarna av de textuella måtten *textlängd, ordmedellängd, andel långord* och *ordvariation/OVIX* under egna rubriker. En ansats är att presentera dessa analyser med en likartad struktur för att jämförelser mellan de olika måtten skall kunna göras. Inledningsvis ges en överblick, därefter görs gruppjämförelser av medelvärden för grupperna CLIL och icke-CLIL samt för de fyra undergrupperna PC, FC, PnC och FnC. I vissa aspekter lyfts individuella resultat, för att beskriva spridning och ge exempel på det lägsta respektive högsta värdena samt även för att belysa enskilda extremvärden som synliggörs i lådagrammen. I avsnitt 8.3 diskuteras även samvariationer av individuella resultat som återfinns i flera variabler som höga respektive låga värden.

7.2.1 Textlängd

Textlängd är en variabel som har valts för att synliggöra storleken av ordförråd vilket är av vikt för att kunna skriva längre textstycken.

Den kvantitativa delundersökningen av texterna visar att textlängden varierar stort. Av de elevskrivna texterna som har undersökts är det endast en elevtext som inte har används då den innehöll 17 graford som bara var ”trams” och har tagits bort från studien, annars har alla elever skrivit texter som utgår ifrån de givna textinstruktionerna, även om ett par texter är väldigt korta (se bilaga 5 med de kortaste respektive längsta elevtexterna i respektive klass angivna). I

materialet består den längsta texten av 1552 graford och den kortaste av 42 graford. Textlängden för de fyra skrivuppgifterna i studien varierar mellan 89 och 1552 graford i CLIL-klasserna med en genomsnittlig textlängd på 589 graford. De tre kontrollklasserna i den här studien har skrivit texter mellan 42 och 1081 graford, med ett genomsnitt på 436 graford.

För en av CLIL-klasserna finns skrivuppgiften som ingick i de nationella proven i *Svenska 1* tillgänglig för oss. Elevuppsatserna är offentlig handling men vi har också fått tillstånd av rektor att använda dem som ett jämförande material. I den CLIL-klass, där vi har tillgång till deras nationella provtexter skrivna i kursen Svenska 1, så är antalet graford i genomsnitt för hela klassen 443 att jämföra med den genomsnittliga textlängden för de fyra texterna samma klass skrev i CLISS-projektet. För skrivuppgift 1 är antalet graford för just denna klass i genomsnitt – 672 graford, för skrivuppgift 2 – 569 graford, för skrivuppgift 3 – 597 graford och för skrivuppgift 4 – 551 graford. Det vill säga; för samtliga skrivuppgifter som är skrivna inom projektets ram så är den genomsnittliga textlängden av CLISS-texterna *längre* än för den nationella provtexten som motsvarighet *trots* att den är skriven som ett prov och därmed är ett betygsgrundande bedömningsunderlag och *trots* att skrivtiden för projektets skrivuppgifter har varit kortare än skrivtiden vid det nationella provet. Detta resultat kan sägas ha betydelse för studiens validitet i undersökningen av de elevskrivna texterna då det kan ses visa elevernas engagemang i de inom projektet insamlade uppgifterna.

De nationella proven på gymnasiet har 180 minuter som tidsram för uppsatsdelen i svenskämnet, på grundskolan är motsvarande tid 160 minuter. För skrivuppgifterna i CLISS-projektet hade ungdomarna 120 minuter som tidsram vid de två första tillfallet men det ansågs vara alltför lång tid då det visade sig vara svårgenomförbart eftersom fler lektioner än en kom att tas i anspråk och flera lärare blev berörda, så för skrivuppgifterna 3 och 4 var tidsramen 90 minuter. Tilläggas kan att flera elever dessutom var klara efter 60 minuter, färre använde hela tiden.

Tabell 11, synliggörs den genomsnittliga textlängden vid de fyra skrivtillfällena under terminerna 1, 2, 3 och 5 för CLIL-klasserna och kontrollklasserna samt för flickorna (F) och för pojkar (P).

TABELL 11. *Genomsnittlig textlängd för samtliga skrivuppgifter/CLIL-icke-CLIL/F-P.*

Text=> antal ord	1 SO	F/P	2 NO	F/P	3 NO	F/P	4SO	F/P
Genomsnittlig textlängd CLIL	573	590/531	579	592/540	692	726/578	512	528/450
icke-CLIL	441	437/445	456	460/451	491	486/499	357	351/363

Tabell 11 visar att CLIL-flickorna i genomsnitt har skrivit längre texter än CLIL-pojkarna vid alla skrivtillfällen samt att CLIL-eleverna har skrivit längre texter än kontrollgrupperna, icke-CLIL, vid samtliga fyra skrivtillfällen. För kontrollklasserna är resultaten omvända, här skriver pojkarna i genomsnitt längre texter än flickorna i tre av fyra skrivuppgifter. I skrivuppgift 2 är flickornas genomsnittliga textlängd något längre än pojkarnas, 460 graford att jämföra med 451 graford.

Texterna blir inte längre över tid utan textlängden verkar hänga samman med ämnet. De längsta texterna för båda grupperna, CLIL och icke-CLIL, är i skrivuppgift tre (termin 3), då ämnet var *Folkhälsoproblem*, vilket skapade engagemang. Den längsta texten som skrevs innehöll 1552 graford, vilket är den längsta texten av alla. Den kortaste texten vid det tredje skrivtillfället är 148 graford vilket är längre än de kortaste texterna som skrevs vid de andra tre skrivtillfällena, så ämnet verkar ha haft betydelse för elevernas skrivande. Den längsta texten är skriven av en flicka i en CLIL-klass och den kortaste på 42 graford av en flicka icke-CLIL-klass (se bilaga 5 med antal graford för de kortaste respektive längsta texterna i samtliga skrivuppgifter).

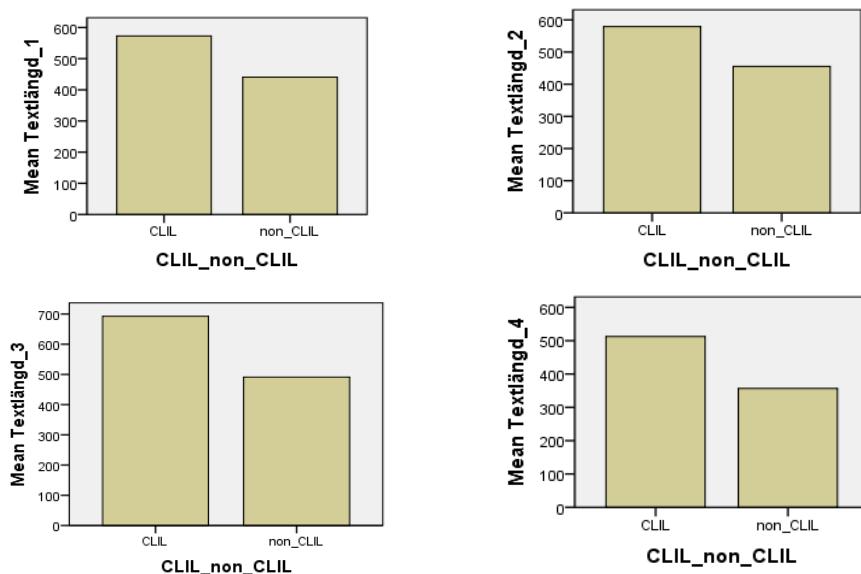
Tabell 13 visar antalet deltagare vid skrivuppgift 1 i grupperna CLIL/icke-CLIL respektive (P) och (F) med medelvärden och standardavvikelse. Den gemensamma genomsnittliga skrivlängden är längst för gruppen flickor CLIL med ett totalt medelvärde på 590 graford per text och kortast för flickor icke-CLIL med en genomsnittlig skrivlängd på 437 ord. Standarsavvikelsen är högst för pojkar icke-CLIL följt av gruppen pojkar CLIL. I lådagrammen som följer (figur 3-6) blir detta mer visuellt synligt vilket beskrivs i texter i anslutning till varje figur.

CLIL-gruppens medelvärden visar genom en analys med *Independent Samples T-test* stark statistisk signifikans i jämförelse med icke-CLIL gruppens, gällande textlängd i alla fyra skrivuppgifterna vilket tabell 12 visar. Detta innebär att CLIL-eleverna, både pojkar och flickor skriver längre texter i denna studie.

TABELL 12. Signifikansvärden för textmedellängd i samtliga skrivuppgifter för grupperna CLIL och icke-CLIL

Skrivuppgift	1	2	3	4
Mellan grupper	.000	.001	.000	.000
CLIL/icke-CLIL				

I figur 3, synliggörs medelvärdet för textlängd i grupperna CLIL och icke-CLIL i de fyra skrivuppgifterna genom stapeldiagram. I stapeldiagrammet för skrivuppgift 3 visar y-axeln ett högre värde än i de övriga tre figurerna då texterna där var längre.



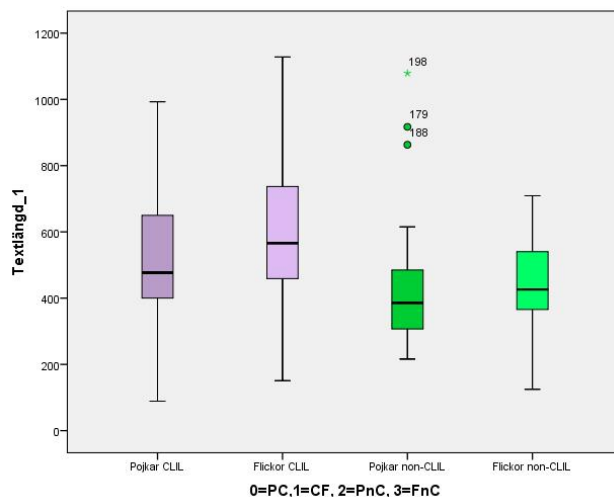
FIGUR 3. Resultat för textlängd i samtliga skrivuppgift 1-4, grupperna CLIL och icke-CLIL

I bilaga 5 visas en tabell med tal som är understruken. De markerar att samma elev i respektive klass har skrivit kortast vid två eller fler tillfällen liksom att elever som har skrivit de längsta texterna i vissa fall gjort så vid mer än ett tillfälle. Analysen gjordes vid manuell granskning av alla elevresultaten som i Excelarken har färgmarkerats för högsta respektive lägsta resultat i samtliga undersökta variabler.

Man kan alltså utläsa att *samma* elev har skrivit den kortaste texten vid två tillfällen. Detta gäller i flera klasser. Dessa elever återfinns både i CLIL-klasser och i en kontrollklass. I de båda NA CLIL-klasserna skriver två elever, i respektive klass, de kortaste texterna vid *tre* tillfällen vardera men i en av de andra klasserna är det olika elever som skriver kortast vid varje skrivuppgift. En elev i EK B icke-CLIL har skrivit den längsta texten i klassen vid *tre* tillfällen, för skrivuppgift 1, 2 och 4 och för skrivuppgift 3 skrev eleven den näst längsta texten i klassen med 1041 graford, det vill säga endast 40 färre graford än den som skrev längst. Ytterligare en CLIL-elev skriver vid två tillfällen den längsta texten i klassen. Tre andra CLIL-elever skriver vid samtliga fyra skrivtillfällen de näst längsta texterna i sina respektive klasser.

Figur 4 visar att medellängden för CLIL-flickornas texter i den första skrivuppgiften är den längsta av de fyra grupperna med 592 graford. Spridningen är ganska stor i båda CLIL-grupperna men den största spridningen finns hos pojkar icke-CLIL med tre uteliggare och den minsta spridningen hos flickor icke-CLIL.

En av uteliggarna (n=198) syns även i lådagrammen för skrivuppgift 3 och 4. Detta diskuteras i kapitel 8.



FIGUR 4. Resultat och spridning för skrivuppgift 1 - textlängd för grupp och kön

Ett *Independent Samples T-Test* (för jämförelse av två grupper) visar ett statistiskt signifikant värde mellan grupperna CLIL och icke-CLIL (p -värde = .000). För att analysera de fyra undergrupperna användes *ANOVA Tukey HSD post hoc* som visar signifikanta värden mellan flickor CLIL och pojkar icke-CLIL (p = .015) samt mellan flickor CLIL och flickor icke-CLIL (p = .004). Mellan pojkar och flickor CLIL är det däremot ingen signifikans (p = .623). Medianvärdet för respektive grupp är: PC – 477 graford, PnC – 385 graford, FnC – 426 graford och flickor CLIL, som skriver de längsta texterna och har ett medianvärde som är högst med 566 graford.

TABELL 13. Textlängd i skrivuppgift 1 med deltagarantal och standardavvikelse i de fyra undergrupperna

Kön	Grupp	Medel	N	SD
Pojkar	CLIL	531	23	217
Flickor	CLIL	590	57	204
Pojkar	icke-CLIL	445	24	223
Flickor	icke-CLIL	437	31	131

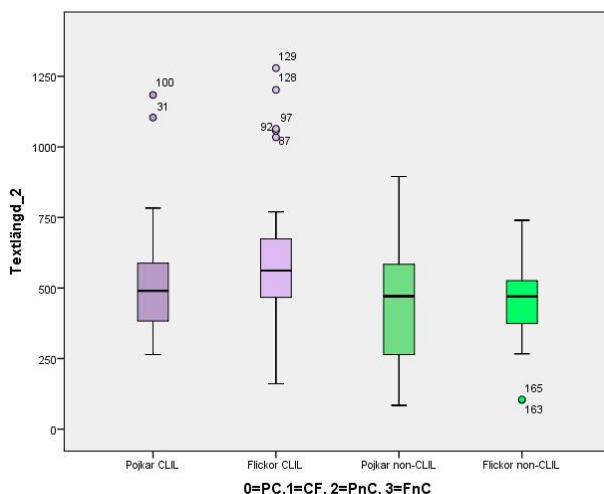
För skrivuppgift 2, är samtliga gruppers textmedellängd något längre än i skrivuppgift 1 och standardavvikelsen är här högst för pojkar CLIL, vilket syns i tabell 14. Liksom i text 1 är spridningen minst i gruppen flickor icke-CLIL.

TABELL 14. *Textlängd i skrivuppgift 2 med deltagarantal och standardavvikelse i de fyra undergrupperna*

Kön	Grupp	Medel	N	SD
Pojkar	CLIL	540	21	240
Flickor	CLIL	592	63	206
Pojkar	icke-CLIL	451	25	211
Flickor	icke-CLIL	460	25	155

Även här visar *Independent Samples T-Test* statistisk signifikans mellan grupperna CLIL och icke-CLIL (p -värde = .001). *ANOVA Tukey Post Hoc* visar signifikanta värden mellan flickor CLIL och pojkar icke-CLIL (p = .022) samt mellan flickor CLIL och flickor icke-CLIL (p = .036). Mellan CLIL-grupperna (F) och (P) är det inte heller här signifikans (p = .742). Inte heller mellan icke-CLIL-grupperna (p = .999).

Medianvärdet i tre av grupperna; pojkar CLIL, pojkar och flickor icke-CLIL är ganska nära varandra (PC – 490 graford, PnC – 471 graford, FnC – 470 graford) medan flickor CLIL även här skriver de längsta texterna och har ett medianvärde som är högst (562 graford). Tre av grupperna, pojkar och flickor CLIL samt flickor icke-CLIL, har också flera uteliggare där en CLIL-flicka återkommer även i skrivuppgift 4 ($n=97$).



FIGUR 5. *Resultat och spridning för skrivuppgift 2 för grupp och kön*

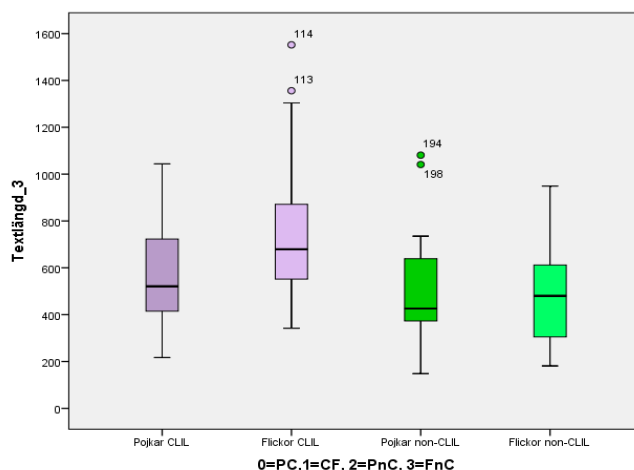
För skrivuppgift 3 är samtliga gruppers textmedellängd längre än i någon annan skrivuppgift. Det är den skrivuppgift som har de längsta texterna, även de kortaste texterna är längre, vilket möjligen kan kopplas till ämnet *Folkhälsoproblem* vilket verkar ha inspirerat ungdomarna mest. Dessa texter är också mer personliga. Standardavvikelsen i skrivuppgift 3 är högst för flickor CLIL, detta syns i tabell 15 nedan och även i lädagrammet i figur 6. Liksom i både text 1 och 2 är spridningen minst i gruppen flickor icke-CLIL.

Independent Samples T-Test visar ett statistiskt signifikant värde (p -värde = .000) mellan grupperna CLIL och icke-CLIL, och *ANOVA Tukey Post Hoc*-testet visar signifikanta värden mellan flickor CLIL och pojkar icke-CLIL (p = .003) samt mellan flickor CLIL och flickor icke-CLIL (p = .000). Mellan CLIL-grupperna är det däremot ingen signifikans (p = .158). Det är heller inga signifikanta resultat mellan pojkar CLIL och de två icke-CLIL-grupperna (PnC) och (FnC) (p = .746, p = .594).

TABELL 15. Textlängd i skrivuppgift 3 med deltagarantal och standardavvikelse i de fyra undergrupperna

Kön	Grupp	Medel	N	SD
Pojkar	CLIL	578	17	247
Flickor	CLIL	726	58	263
Pojkar	icke-CLIL	498	21	241
Flickor	icke-CLIL	486	30	202

Medianvärdet för textlängden i skrivuppgift 3 är; pojkar CLIL - 521, pojkar icke-CLIL - 426, flickor icke-CLIL - 480 och flickor CLIL skriver även här de längsta texterna och har ett medianvärde på 681. Figur 6 synliggör detta.



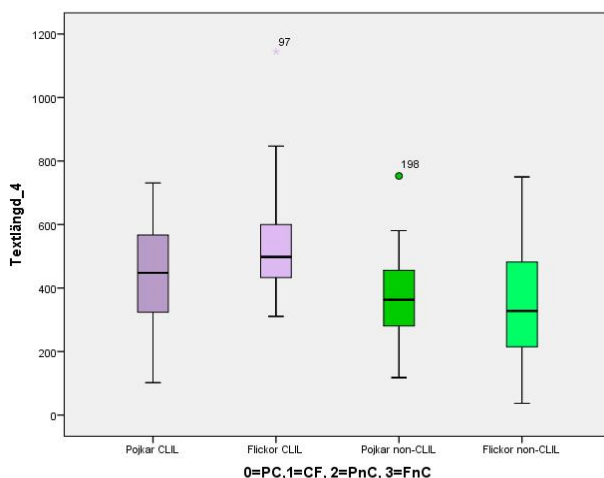
FIGUR 6. Resultat och spridning för skrivuppgift 3 för grupp och kön

I den sista skrivuppgiften 4, är medellängden den kortaste för alla fyra grupperna. Endast en text, skriven av en CLIL-flicka är längre än 1000 graford, att jämföra med tidigare skrivuppgifter där det varit fler (uppgift 1: 2 elever, uppgift 2: 4 elever, uppgift 3: 6 elever). Spridningen är störst för pojkar CLIL med en standardavvikelse på 186 tätt följda av flickor icke-CLIL med 185. Dock är spridningen i alla grupper lägre i denna uppgift vilket syns i tabellen nedan liksom att även deltagarantalet är lägre för alla grupper förutom pojkar icke-CLIL.

TABELL 16. *Textlängd i skrivuppgift 3 med deltagarantal och standardavvikelse i de fyra undergrupperna*

Kön	Grupp	Medel	N	SD
Pojkar	CLIL	450	13	186
Flickor	CLIL	528	54	149
Pojkar	icke-CLIL	362	31	142
Flickor	icke-CLIL	351	27	185

Independent Samples T- Test visar ett statistiskt signifikant värde mellan grupperna (p -värde = .000) och *ANOVA Tukey Post Hoc* testet visar signifikanta värden mellan flickor CLIL och pojkar icke-CLIL och flickor icke-CLIL ($p = .000$). Mellan CLIL-grupperna är det däremot ingen signifikans ($p = .404$). Det är heller inga signifikanta resultat mellan pojkar CLIL och båda icke-CLIL-grupperna ($p = .339$, $p = .256$).



FIGUR 7. *Resultat och spridning för skrivuppgift 4 för grupp och kön*

Medianvärdet är här; pojkar CLIL – 448 graford, flickor CLIL – 498 graford, pojkar icke-CLIL – 363 graford, flickor icke-CLIL – 328 graford vilket är det lägsta genomsnittliga medianvärdet för alla grupper i samtliga fyra skrivuppgifter. Observera uteliggarna ($n=97$), FC och ($n=198$) PnC och jämför med figur 4, 5 och 6.

Sammanfattningsvis gällande textlängd kan sägas att CLIL-gruppernas resultat för (P) och (F) visar statistisk signifikans i jämförelse med båda icke-CLIL grupperna (PnC) och (FnC) i alla fyra skrivuppgifterna. Detta innebär att CLIL-eleverna, både pojkar och flickor skriver längre texter på svenska än kontrollgruppen genom hela studien. *ANOVA Tukey Post Hoc* visar även statistisk signifikans för CLIL-flickorna gentemot de båda icke-CLIL grupperna i samtliga skrivuppgifter. Flickor CLIL, skriver alltså de längsta texterna sett till värdena för textlängd.

I skrivuppgift 3 är medelvärdena för textlängd högst, eleverna skriver längst texter i den skrivuppgiften och kortast i skrivuppgift 4. I skrivuppgift 4 är spridningen (SD) minst men texterna är också kortare. I skrivuppgift 2 finns flest uteliggare, det vill säga elever som har avvikande värden det som i SPSS kallas *extremvärden* (Djurfeldt *et al* 2010:63, se även 6.4). I lådagrammen synliggörs även att samma elever vid upprepade tillfällen har värden som är så kallade extremvärden, såväl höga som låga. Detta diskuteras vidare i kapitel 8.

7.2.2 Ordmedellängd

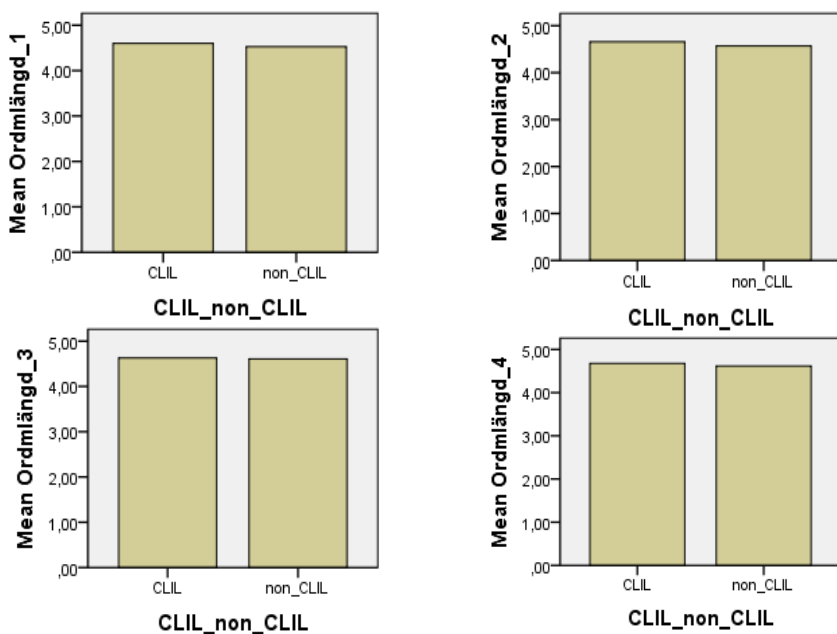
Ordmedellängd (antalet tecken delat med antalet ord) är ett mått som har använts i flera studier och som räknas ut genom att man räknar på antalet tecken i en text och delar med antalet ord. Ordmedellängd är ett tecken på variation i en text genom att blanda korta och långa ord. De vanligaste orden i språket är en- eller tvåstaviga vilket synliggörs i ordlistorna som utgör grunden för de två korpusarna som har skapats av alla 520 (260+260) elevtexter. I Korpus 1 är de 30 första graforden en- och/eller tvåstaviga med tillsammans 51 428 förekomster i de 260 texterna. På 30:e plats kommer ett trestavigt ord; *Sverige* med 577 förekomster, därefter fortsätter de en- och tvåstaviga orden med två undantag ända till graford nummer 100. Från plats 2109 till och med plats 7509 återfinns flerstaviga ord vilka har en till tre förekomster i texterna. Flerstaviga ord är alltså mer lågfrekventa. Korpus 2 följer samma mönster. Bland de 250 först graforden finns tolv trestaviga ord, ett med fem stavelser övriga är en- eller tvåstaviga ord. Även här har de längre orden få förekomster, det vill säga är lågfrekventa.

I tabell 17 återfinns medelvärdena för ordmedellängd i de två undersökningsgrupperna CLIL och icke-CLIL i samtliga fyra skrivuppgifter tillsammans med standardavvikelsen och deltagarantal för respektive grupp.

TABELL 17. Ordmedellängd för CLIL och icke-CLIL i de fyra olika skrivuppgifterna

Skrivuppg.	Grupp	N	Medel	SD
1	CLIL	80	4,60	,032
	icke-CLIL	55	4,53	,038
2	CLIL	84	4,66	0,35
	icke-CLIL	50	4,57	,049
3	CLIL	75	4,63	,033
	icke-CLIL	51	4,61	,053
4	CLIL	67	4,68	,039
	icke-CLIL	57	4,61	,041

Stapeldiagrammen i figur 8 visar att det är väldigt små skillnader mellan grupperna CLIL och icke-CLIL med ordmedellängden angiven på y-axeln.



FIGUR 8. Resultat för ordmedellängd i skrivuppgift 1-4 för respektive grupp CLIL och icke-CLIL

Det högsta medelvärdet återfinns för CLIL-gruppen i text 4, *Helt delad föräldrapeng*, där flera ord innehåller förledet *föräldra-* vilket ger många, långa ord. Några exempel är: *föräldraledighet – en, föräldrapenning – en, föräldrapeng – en*, samt andra långa ord som *försäkringskassan, gemensam, mammorna – s, lenadsstandard, barnadödlighet, sjukvården, utnyttja* med flera.

I tabell 18 redogörs för den genomsnittliga ordmedellängden vid de fyra olika skrivtillfällena för de fyra grupperna; pojkar CLIL, flickor CLIL, pojkar icke-CLIL och flickor icke-CLIL.

I samtliga skrivuppgifter; *Hälsa och välbefinnande, Bilfri stad, Folkhälsoproblem* och *Helt delad föräldrapeng* har pojkar CLIL den högsta genomsnittliga ordmedellängden. Gruppen flickor icke-CLIL har i alla fyra skrivuppgifter den kortaste ordmedellängden. Variationen mellan grupperna flickor CLIL och pojkar icke-CLIL visar att grupperna har en högre ordmedellängd vid två tillfällen vardera. Den genomsnittliga ordmedellängden i samtliga grupper PC, FC, PnC och FnC, varierar mellan 4,46 och 4,79. Tittar man på minimum- och maximumresultat för individer är det lägsta 3,48 i skrivuppgift 4, för en pojke CLIL samt det högsta 6,29 för en flicka icke-CLIL i skrivuppgift 3 och det näst högsta 6,13 för en flicka CLIL i skrivuppgift 2. Det högsta individuella värdet i alla fyra grupper för en flicka icke-CLIL kan ställas mot den genomsnittliga ordmedellängden för gruppen flickor icke-CLIL som ligger lägst av de fyra grupperna i tre av de fyra skrivuppgifterna.

TABELL 18. *Genomsnittlig ordmedellängd för de olika grupperna i de fyra skrivuppgifterna*

Grupp	Ordmedellängd			
	1	2	3	4
Pojkar CLIL	4,67	4,75	4,79	4,73
Flickor CLIL	4,57	4,63	4,58	4,67
Pojkar icke-C	4,61	4,60	4,67	4,61
Flickor icke-C	4,46	4,53	4,57	4,62

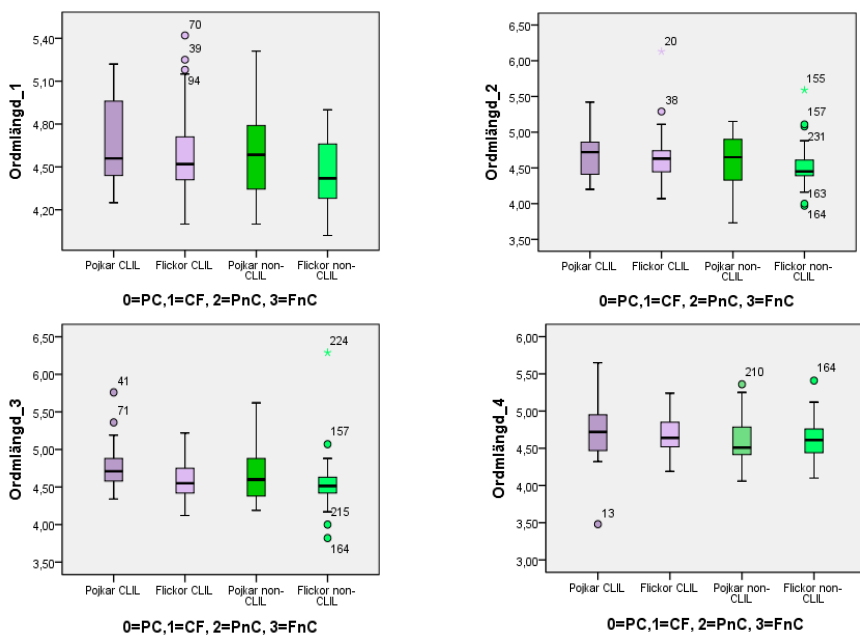
Pojkarna har en genomsnittlig högre ordmedellängd än flickorna, något som syns i tabellen ovan för både CLIL-gruppen och kontrollgruppen vid de fyra skrivtillfällena under termin 1, 2, 3 och 5.

Vid ett *Independent Samples T-test*, för att se statistisk signifikans mellan grupperna CLIL och icke-CLIL, visas ingen signifikans i någon av uppgifterna vilket synliggörs i tabell 19. Gruppernas resultat ligger nära varandra med ytterst små skillnader.

TABELL 19. *Signifikansvärden för ordmedellängd i samtliga skrivuppgifter för grupperna CLIL och icke-CLIL*

Skrivuppgift	1	2	3	4
Mellan grupper CLIL/icke CLIL	.073	.178	.082	.628

I figur 9 synliggörs spridningen mellan de fyra undergrupperna i de fyra skrivuppgifterna.



FIGUR 9. Resultat och spridning för ordmedellängd i skrivuppgift 1-4 för respektive grupp och kön

Inte i någon skrivuppgift visar *T-testet* ett statistiskt signifikant värde mellan grupperna CLIL och icke-CLIL. Inte heller i *ANOVA Tukey Post Hoc*-testet visas signifikanta värden mellan undergrupperna flickor och pojkar respektive flickor och pojkar icke-CLIL. Flickor CLIL har den högsta standardavvikelsen av grupperna i skrivuppgift 1 ($SD = .3244$) med tre uteliggare. Medianvärdena i skrivuppgift 1 är: pojkar CLIL – 4,56, flickor CLIL – 4,52, pojkar icke-CLIL – 4,59, flickor icke-CLIL – 4,42.

I skrivuppgift 2 har pojkar-CLIL den högsta standardavvikelsen ($SD = .3643$). Flickor icke-CLIL har flera uteliggare. Medianvärdena i skrivuppgift 2 är: pojkar CLIL – 4,72, flickor CLIL – 4,63, pojkar icke-CLIL – 4,65, flickor icke-CLIL – 4,45. I skrivuppgift 3 har flickor icke-CLIL den högsta standardavvikelsen ($SD = .4043$) återigen med flera uteliggare. Medianvärdena i skrivuppgift 3 är: pojkar CLIL – 4,71, flickor CLIL – 4,55, pojkar icke-CLIL – 4,60, flickor icke-CLIL – 4,52. I skrivuppgift 4 har pojkar CLIL den högsta standardavvikelsen ($SD = .5509$). Medianvärdena i skrivuppgift 2 är: pojkar CLIL – 4,72, flickor CLIL – 4,64, pojkar icke-CLIL – 4,51, flickor icke-CLIL – 4,61.

Sammanfattningsvis kan sägas att gruppernas resultat inte visar statistisk signifikans gällande ordmedellängd i någon av de fyra skrivuppgifterna. Båda pojkgруппerna har i samtliga skrivuppgifter ett högre medelvärde för ordmedellängd än de två flickgrupperna. Detta kan ställas i relation till att flickor CLIL skrev

längre texter än båda pojkgrupperna i samtliga skrivuppgifter. Gruppen flickor icke-CLIL har i tre av fyra skrivuppgifter den lägsta ordmedellängden.

Gruppernas resultat gällande ordmedellängd skiljer sig ganska lite åt. Däremot är det större skillnader mellan individer vilket synliggörs i lådagrammen i figur 9. I skrivuppgift 1 är spridningen mellan de högsta och lägsta resultaten stor. I samtliga skrivuppgifter finns det individer som hamnar utanför och har så kallade *extremvärden* (se avsnitt 6.4). En sådan elev, en flicka icke-CLIL, nummer 164, syns som en uteliggare i skrivuppgift 2, 3 och 4. Här gick jag tillbaka in i texterna för att se vad som orsakade detta då eleven i skrivuppgift 2 och 3 har låga resultat men i skrivuppgift 4 ett högt resultat. Samma elev syns även i figur 11 gällande långordsfrekvens med ett lågt resultat i skrivuppgift 3 (extremvärde) och ett högt resultat i skrivuppgift 4 (extremvärde). Detta visades vid en manuell analys av texten hänga samman med elevens ordval där flera graford var långa ord som upprepades flera gånger vilket påverkar denna variabel.

7.2.3 Andel långord

Resultaten av andelen *långord* (ord med fler än sex tecken angivet som procentandel av den totala andelen ord i elevtexterna) för de 520 skrivuppgifterna redovisas som medelvärden för de olika grupperna CLIL, icke CLIL samt för pojkar och flickor i respektive grupp. Dessutom kommenteras vissa individuella resultat. Ser man till de åtta klassernas medelvärden så ligger de mellan 15,4 % för en kontrollklass som har det lägsta medelresultatet och 20,9 % för en CLIL-klass som har det högsta medelvärdet för andel långord.

I tabell 20 återfinns medelvärdena för elevgrupperna CLIL och icke-CLIL för de olika skrivuppgifterna samt även deltagarantal vid varje skrivtillfälle samt standardavvikelsen. Resultaten varierar mellan 16,00 % för icke-CLIL-gruppen i skrivuppgift 4 som den lägsta andelen långord till 18,42 % för CLIL-gruppen i den tredje skrivuppgiften.

TABELL 20. Medelvärden av andelen långord mellan CLIL/icke-CLIL samt standardavvikelse för de fyra olika skrivuppgifterna

Skrivuppg.	Grupp	N	Medel	SD
1	CLIL	80	18,20	3,30
	icke-CLIL	55	17,22	3,64
2	CLIL	84	17,23	4,02
	icke-CLIL	50	16,48	4,71
3	CLIL	75	18,42	4,22
	icke-CLIL	51	17,28	4,33
4	CLIL	67	17,30	4,12
	icke-CLIL	57	16,00	4,08

De individuella resultaten är varierande; från en elev, en icke-CLIL-pojke, med den allra lägsta andelen långord med 4,06 % följd av en flicka icke-CLIL, med 4,19 %, båda i skrivuppgift 4. Till det högsta värdet för en CLIL-pojke med 31,05 % i skrivuppgift 3.

Av de 520 texterna har 18 texter en andel av långord under 11 % (mellan 6,15 och 11 %). Av de 18 texterna finns 8 texter i en av kontrollklasserna, inga alls i två av CLIL-klasserna och de övriga 10 med en andel av långord under 11 % fördelade på de övriga klasserna, både CLIL och icke-CLIL. Gruppen av texter med väldigt låg andel långord utgör 3,5 % av materialet. Fyra skrivuppgifter har över 30 % långord vilket utgör knappt 1 % av materialet och 99 skrivuppgifter har en andel långord mellan 20 % och 30 %. Den gruppen utgör 19 % av alla skrivna uppgifter.

Värt att notera är att för den CLIL-klass där vi har nationella provtexter att jämföra med så är medelvärdet för andelen långord i de texterna 15 %. Av texterna har 20 % av eleverna en andel långord under 10 %, det vill säga att eleverna har skrivit *färre* långa ord i sina nationella provtexter i jämförelse med de fyra skrivuppgifterna som eleverna har skrivit inom CLISS-projektet.

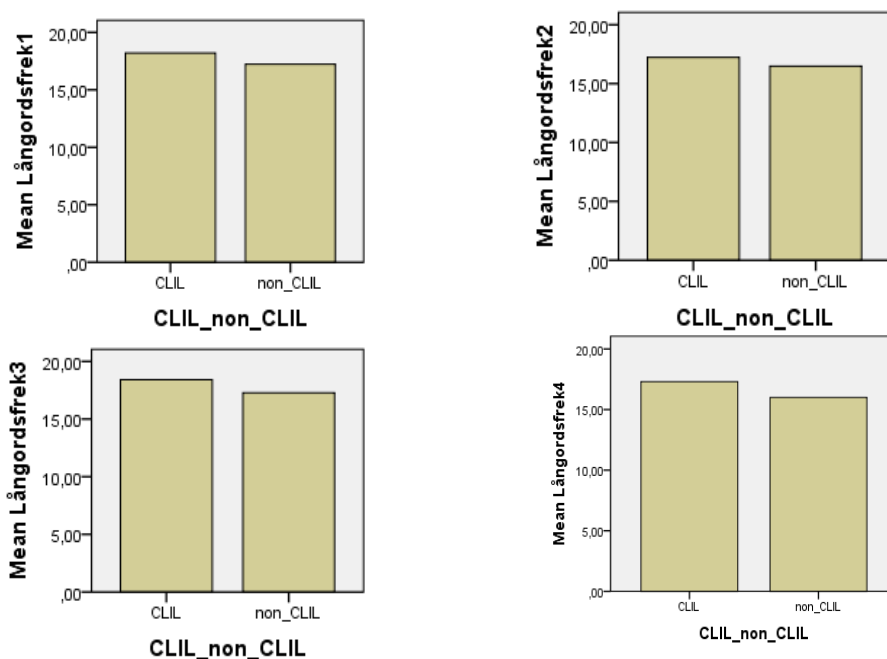
Elevtexterna som har högst andel långord är också relativt långa och de som har en låg andel långord är kortare men ligger ändå över 200 ord. Några texter är mellan 500 och 1000 ord men har en andel långord runt 10 %. Detta kan ställas i relation till den lexikala profileringen och diskuteras ytterligare i kapitel 8.

Vid ett *T-test* visas statistisk signifikans mellan grupperna CLIL och icke-CLIL i skrivuppgift 2 och 3. I tabell 21 syns *p*-värde mellan grupperna i de fyra skrivuppgifterna.

TABELL 21. *Signifikansvärden för andelen långord i samtliga skrivuppgifter för grupperna CLIL och icke-CLIL*

Andel långord	Skrivuppg. 1	Skrivuppg. 2	Skrivuppg. 3	Skrivuppg. 4
CLIL/icke-CLIL	$p = .070$	$p = .018$	$p = .016$	$p = .098$

I stapeldiagrammen i figur 10 visas medelvärdena för andel långord i procent för de båda grupperna CLIL och icke-CLIL.

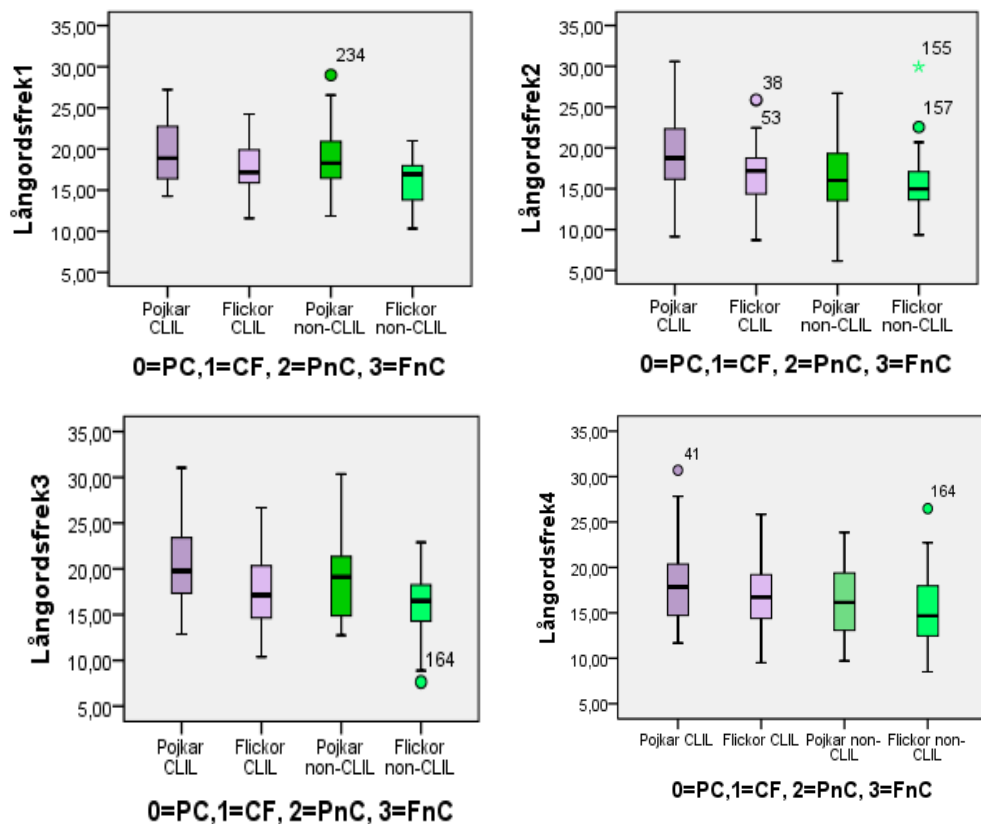


FIGUR 10. Resultat för procentuell andelen långord i skrivuppgift 1-4 för respektive grupp

Vid en analys av de fyra undergrupperna CLIL (F) och (P) respektive icke-CLIL (F) och (P) synliggörs att pojkar CLIL har det högsta medelvärdet av andelen långord vid samtliga skrivtillfällen, därefter följer pojkar icke-CLIL i tre skrivuppgifter av fyra, sedan flickor CLIL och det lägsta medelvärdet i samtliga skrivuppgifter har flickor icke-CLIL. Dessa resultat redovisas i tabell 22 samt synliggörs i lådagram för de fyra skrivtillfällena i figur 11.

TABELL 22. Genomsnittlig procentuell andel långord för de olika grupperna i de fyra skrivuppgifterna

Grupp	Andel långord 1	Andel långord 2	Andel långord 3	Andel långord 4
Pojkar CLIL	19,26	19,02	20,56	19,00
Flickor CLIL	17,78	16,64	17,79	16,89
Pojkar icke-C	18,73	17,12	18,98	13,03
Flickor icke-C	16,06	15,84	16,10	11,47



FIGUR 11. Lådagram för procentuell andelen långord i skrivuppgift 1-4 för respektive grupp

I figur 11 synliggörs, som tidigare nämnts, att pojkar CLIL vid samtliga skrivtillfällen har det högsta medelvärdet av andelen långord. Detta innebär att pojkar CLIL i högre grad använder ord som innehåller mer än sex tecken. Lådagrammen synliggör även att flickor icke-CLIL har de lägsta medelvärdena. Det finns färre uteliggare i jämförelse med tidigare rapporterade resultat. Texterna uppvisar alltså inte i samma omfattning extremvärden inom grupperna gällande användningen av långa ord som tidigare rapporterade textuella måttenheter.

Den största spridningen finns hos pojkar CLIL i skrivuppgift 4 med en standardavvikelse på ($SD = 5,66$), den minsta spridningen återfinns i gruppen flickor icke-CLIL i skrivuppgift 1 ($SD = 2,90$). Samma grupp, flickor icke-CLIL, har även de lägsta medianvärdena i samtliga skrivuppgifter, i uppgift 4 är värdet 11,47 %, vilket är gruppens lägsta att jämföra med medianvärdena för pojkar CLIL i samma uppgift som är 19,00 %.

Kontrollgruppens flickor skriver därmed färre långa ord i jämförelse med CLIL-pojkarna vilket även p -värdena bekräftar: i skrivuppgift 1 (p -värde = .003), i skrivuppgift 2 (p = .073), i skrivuppgift 3 (p -värde = .003) och i skrivuppgift 4 (p -värde = .000).

Sammanfattningsvis kan sägas att det finns ett signifikant resultat i skrivuppgift 4 gällande andelen långord där både flickor CLIL och pojkar CLIL använder längre ord i en högre omfattning än kontrollklasserna. I övriga tre skrivuppgifter finns ingen statistisk signifikans mellan grupperna CLIL och icke-CLIL däremot mellan undergrupperna vilket har redovisats ovan samt i skrivuppgift 4 där även signifikans visas mellan flickor CLIL och de två kontrollgrupperna pojkar och flickor icke-CLIL: FC – PnC (p = .006), FC – FnC (p = .000).

Rresultaten kan alltså även här, liksom för ordmedellängd, kopplas till text 4, *Helt delad föräldrapeng*, där flera ord innehåller förledet föräldra- vilket ger många långa ord, samt även andra långa ord som kan kopplas till ämnet.

7.2.4 Ordvariation

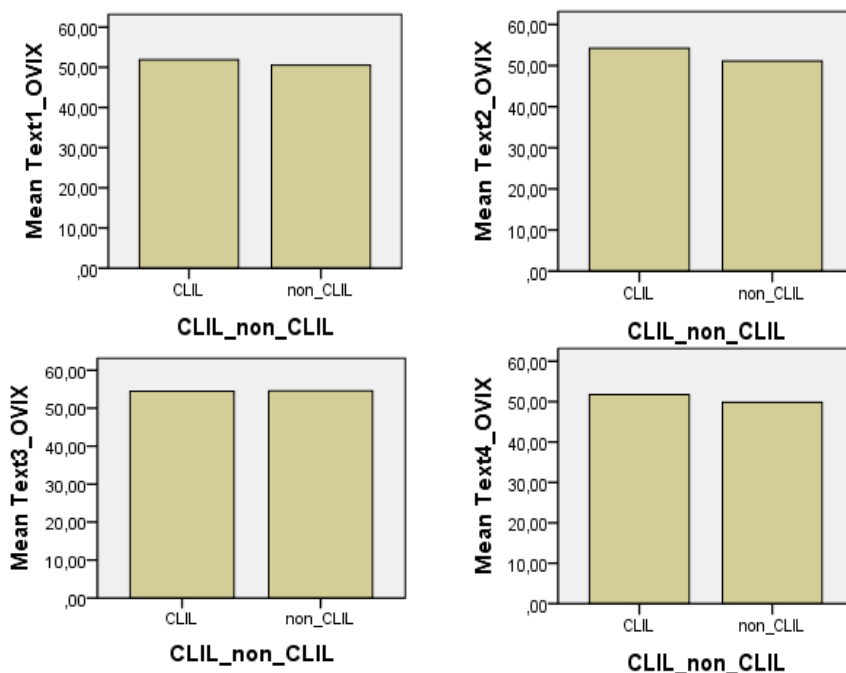
För att möjliggöra jämförelser av texter med olika längd togs måttet *OVIX* – ordvariationsindex, fram av Hultman inom projektet Skrivsyntax och användes sedan i *Gymnasistsvenska* (1977). Det ”visar antalet olika ord (lexord) i förhållande till det totala antalet ord (löpord) med kompensation för olika textlängd” (Östlund-Stjärnegårdh 2002:89). Texter som är kortare än 200 ord har inte tagits med i undersökningen av *OVIX* då de siffrorna, enligt tidigare studier, blir missvisande på grund av den logaritmiska skalan (se avsnitt 5.3). Resultaten i denna studie visar att skrivuppgifterna för texter som är längre än 200 ord har *OVIX*-värden mellan 32,35 (FnC) och 74,35 (PC). Av de 520 texterna som undersöks är 17 texter kortare än 200 ord och har inte räknats med i analysen av *OVIX*. Av de 17 är två skrivna av CLIL-elever, resten av icke-CLIL-elever och elva av dessa 17 är skrivna i uppgift 4 som har de kortaste texterna över lag.

De längsta texterna skrevs vid skrivuppgift 3 *En bilfri stad*, dessa visar också de höga medelvärdena för *OVIX*. Även vid en sammanslagning av de olika resultaten för att få aritmetiska medelvärden för hela grupper, i det här fallet för varje klass, så framträder mönstret att den klass som har skrivit de längsta texterna i genomsnitt också har det högsta *OVIX*-värdet samt att CLIL-klasserna skriver längre texter med högre ordvariation. Tabell 23 visar signifikansvärden för *OVIX* i samtliga skrivuppgifter.

TABELL 23. Signifikansvärden för *OVIX* i samtliga skrivuppgifter för grupperna CLIL och icke-CLIL

OVIX	Skrivuppg. 1	Skrivuppg. 2	Skrivuppg. 3	Skrivuppg. 4
CLIL/icke-CLIL	$p = .263$	$p = .009$	$p = .930$	$p = .130$

Vid undersökning med *SPSS Independent T-test* mellan CLIL och icke-CLIL gruppen visas statistisk signifikans i skrivuppgift 2 ($p = .009$), men inte i övriga texter, detta redovisas i tabell 23. I figur 12 synliggörs resultaten för de båda grupperna CLIL och icke-CLIL i stapeldiagram där y-axeln anger *OVIX*-värde.



FIGUR 12. Stapeldiagram för ordvariation/*OVIX* i skrivuppgift 1-4 för respektive grupp

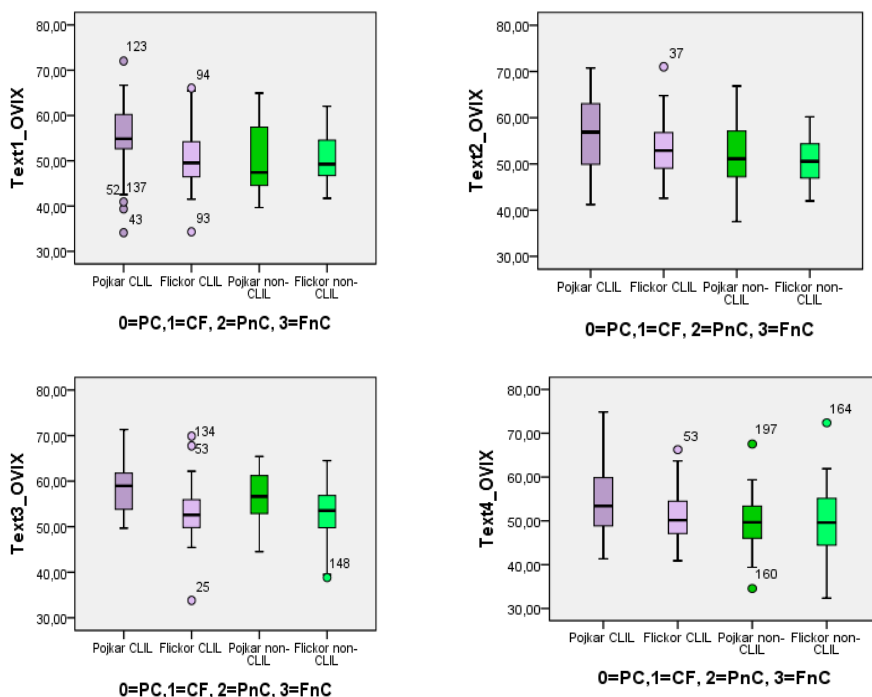
I tabell 24 syns att det är en variation mellan de fyra undergrupperna, dock har pojkar CLIL det högsta medelvärdet för *OVIX* i samtliga skrivuppgifter och flickor icke-CLIL det lägsta i samtliga utom den sista. De två andra grupperna, flickor CLIL och pojkar icke-CLIL, har en variation sinsemellan av medelvärden i de olika skrivuppgifterna.

TABELL 24. Medelvärde av *OVIX* för *PC*, *FC*, *PnC* och *FnC* för samtliga skrivuppgifter

Grupp	<i>OVIX</i>	<i>OVIX</i>	<i>OVIX</i>	<i>OVIX</i>
Skrivuppg.	1	2	3	4
Pojkar CLIL	54,57	56,41	58,60	55,12
Flickor CLIL	50,80	53,45	53,24	51,00
Pojkar icke-CLIL	50,16	51,86	56,50	49,61
Flickor icke-CLIL	50,77	50,35	53,19	50,12

Resultaten för ordvariation i de fyra undergrupperna överensstämmer därmed med resultaten för ordmedellängd och längordsfrekvens.

För att synliggöra spridning, max- och minivärde samt medianvärde för de fyra grupperna används lådagram i figur 13.



FIGUR 13. Lådagram för ordvariation/OVIX i skrivuppgift 1-4 för respektive grupp

I samtliga fyra skrivuppgifter har PC ett maxvärde över 70, dock uppvisar en elev (PC, $n=43$) i skrivuppgift 1 ett minivärde på 34,12 vilket är det näst lägsta av alla elever vid alla skrivtillfällen. En elev i gruppen FnC ($n=25$) har det lägsta värdet i skrivuppgift 4 med 32,35. En elev, en CLIL-flicka ($n=53$), syns som uteliggare i både skrivuppgift 3 och 4. Hon är den enda elev som återfinns utanför lådagrammen fler än en gång i denna variabel. De högsta medianvärdena finns i samtliga grupper i skrivuppgift 3, där de längsta texterna skrevs. Där är OVIX – medelvärdet för samtliga grupper över 50.

ANOVA Post Hoc Testet visar en signifikans mellan PC och FnC i skrivuppgift 2, ($p = .010$), samt mellan PC och båda flickgrupperna FC och FnC i skrivuppgift 3 ($p = .007$, $p = .017$) men ingen signifikans återfinns mellan någon av grupperna i skrivuppgift 1 och 4.

Medelvärdena i de skrivuppgifter som har undersökts i denna studie är lägre än i tidigare studier. Det sammanlagda medelvärdet för *OVIX* är här 53 (avrundat från 52,86). Diskussion angående detta förs i 8.3 och 8.4.

7.3 Kvantitativa ämnesrelaterade mätningar

Den ämnesrelaterade vokabulären omfattar i den här uppsatsen ordkategorier som utifrån en systemisk-funktionell språksyn kännetecknar det vetenskapliga språket (Halliday 1993, Schleppegrell 2004 m.fl.). Ordkategorierna som har identifierats, undersökts och analyserats är *nominaliseringar* och *verb i passiv form*. De *akademiska orden* har urskiljts på delvis annat sätt och rapporteras i avsnitt 7.3.3.

För att identifiera en ämnesrelaterad vokabulär och därigenom få resultat för förekomst och spridning av nominaliseringar och verb i passiv form har korpuslingvistiska metoder tillämpats och två olika korpusar har skapats, Korpus 1 och Korpus 2 (se avsnitt 6.2.1 samt 7.1.2). Alla elevtexter har undersökts i webbaserade program och resultaten finns både för individer samt som medelvärden för klasser, CLIL- och icke CLIL och för genus (P) och (F).

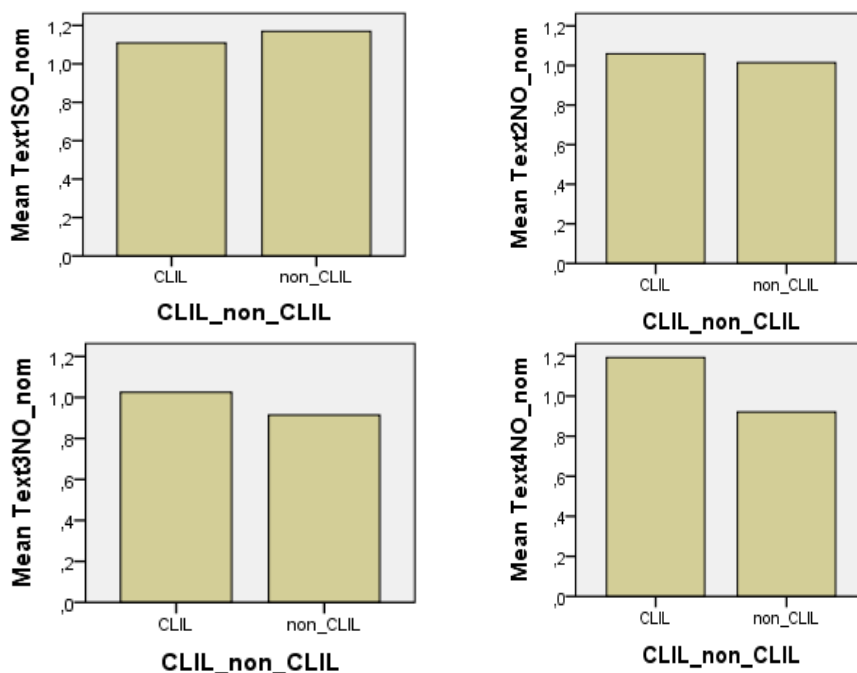
7.3.1 Nominaliseringar

Användningsgraden av nominaliseringar, det vill säga att verbprocesser har gjorts om till substantiv (se 5.2.1), är över lag väldigt låg. I flera elevtexter återfinns inga nominaliseringar alls. Detta diskuteras i 8.2. Användningsgraden som redovisas är här uttryckt i form av procentandel av den totala andelen ord i texterna. Dock är det ingen elev som har 0 % i alla skrivuppgifter som de har deltagit i. Endast en elev som bara har skrivit en text, hade 0 % i den, övriga kan ha 0 % i någon text men har använt en eller ett par nominaliseringar i övriga texter. Den högsta användningsgraden är 3,6 % som återfinns i tre texter, två av dem är skrivna av samma CLIL-elev i skrivuppgift 3 och 4, den tredje är av en icke-CLIL-elev i skrivuppgift 4.

Med *T-test* visas ingen tydlig signifikans mellan grupperna CLIL och icke CLIL förutom i skrivuppgift 4 där *p*-värdet är nära. I skrivuppgift 1 är signifikansen mellan grupperna ($p = .641$), i skrivuppgift 2 ($p = .728$), i skrivuppgift 3 är signifikansen ($p = .426$) och i skrivuppgift 4 är värdet mellan CLIL och icke-CLIL ($p = .059$).

TABELL 25. Signifikansvärden av nominaliseringar i samtliga skrivuppgifter, 1-4

Skrivuppgift	1	2	3	4
Mellan grupper CLIL/icke-CLIL	.641	.728	.426	.059



FIGUR 14. Stapeldiagram för procentuell andel nominaliseringar i skrivuppgift 1-4 för respektive grupp

Stapeldiagrammen visar visserligen synliga skillnader mellan grupperna CLIL och icke-CLIL men y-axelns gradering, från 0 till 1,2 som är högst, visar 0,2 för varje gradering vilket i procent blir väldigt lite, så de små skillnaderna synliggörs men de är ändå små och ger inte säkerställd statistisk signifikans. I uppgift 4 är skillnaden mellan grupperna störst vilket också visades i p -värdet.

Nedan redovisas genomsnittet för de fyra undergrupperna i studien vid samtliga skrivtillfällen. Det är inte några stora skillnader mellan gruppernas användning. I den fjärde skrivuppgiften har pojkar CLIL det högsta medelvärdet på 1,692 %.

I tabell 26-29, synliggörs medeltalen för nominaliseringar för grupperna CLIL (P) och (F) samt icke-CLIL (P) och (F), vid de fyra olika skrivtillfällen. I skrivuppgift 1 har alla grupperna ett medeltal strax över 1 %, i skrivuppgift 2 och 3 har både flickor CLIL och flickor icke-CLIL ett medel som ligger under 1 % och i skrivuppgift 4 är det de båda icke-CLIL grupperna som har ett medel under 1 %. Standardavvikelsen är lägst vid tre tillfällen för gruppen flickor CLIL, vilket innebär att spridningen inom den gruppen är minst. De har även det högsta antalet deltagare vid samtliga skrivtillfällen.

TABELL 26. *Procentuell andel av nominaliseringar i skrivuppgift 1*

Kön	Grupp	Medel	N	SD
Pojkar	CLIL	1,048	23	,6788
Flickor	CLIL	1,133	57	,7806
Pojkar	icke-CLIL	1,154	24	,7241
Flickor	icke-CLIL	1,133	31	,7355

TABELL 27. *Procentuell andel av nominaliseringar i skrivuppgift 2*

Kön	Grupp	Medel	N	SD
Pojkar	CLIL	1,271	21	,9419
Flickor	CLIL	0,989	63	,5597
Pojkar	icke-CLIL	1,032	25	,8484
Flickor	icke-CLIL	0,996	25	,7850

TABELL 28. *Procentuell andel av nominaliseringar i skrivuppgift 3*

Kön	Grupp	Medel	N	SD
Pojkar	CLIL	1,406	17	1,0703
Flickor	CLIL	0,914	58	,6309
Pojkar	icke-CLIL	1,043	21	,8066
Flickor	icke-CLIL	0,823	30	,7384

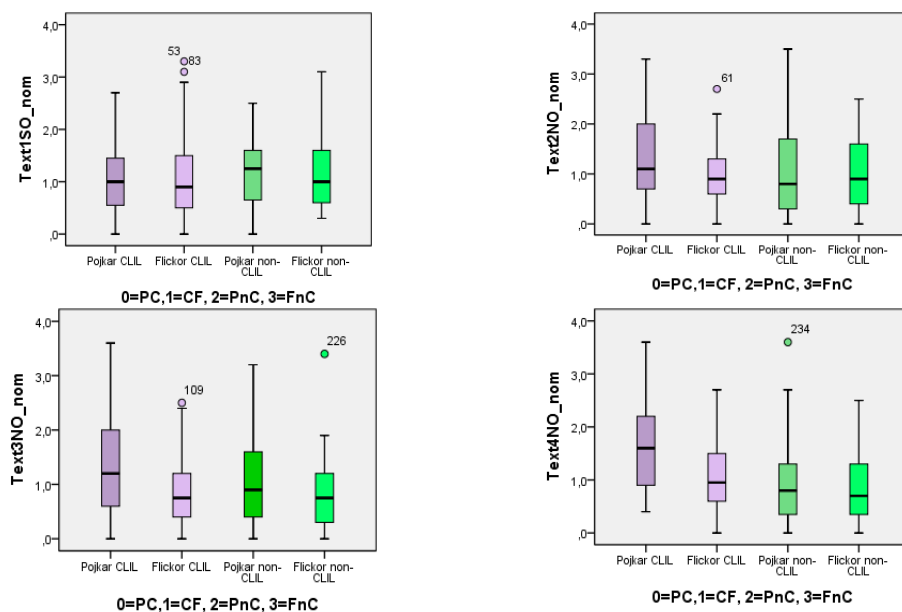
TABELL 29. *Procentuell andel av nominaliseringar i skrivuppgift 4*

Kön	Grupp	Medel	N	SD
Pojkar	CLIL	1,692	13	1,1071
Flickor	CLIL	1,072	54	,6888
Pojkar	icke-CLIL	0,955	31	,8512
Flickor	icke-CLIL	0,881	27	,6895

I den fjärde skrivuppgiften visar *Tukey post hoc* statistisk signifikans mellan pojkar CLIL och pojkar icke-CLIL ($p = .026$), mellan pojkar CLIL och flickor icke-CLIL ($p = .014$) men däremot inte mellan de andra grupperna i någon av skrivuppgifterna. Exempel på nominaliseringar från elevtexterna återfinns i Bilaga 6.

I tabellerna ovan synliggöra något som skulle kunna tolkas som en progression i gruppen pojkar CLIL, vars medelvärde höjs från skrivuppgift 1 till skrivuppgift 4 utan att gå ned i skrivuppgift 2 och/eller skrivuppgift 3. För att ta reda på om så är fallet krävs en djupare granskning av texterna.

Genom att använda lådagran för att synliggöra förekomst och fördelning inom gruppen av användningsgraden av nominaliseringar i de fyra skrivuppgifterna framträder även något annat.



FIGUR 15. Lådagram för procentuell andel nominaliseringar i skrivuppgift 1-4 för respektive grupp

Medelvärdena för användningsgraden av den ämnesrelaterade vokabulären som undersöks i denna uppsats, här representerad av ordkategorin *nominaliseringar*, är låg. I samtliga undergrupper finns flera elever som inte använder någon men gällande nominaliseringar synliggörs i lådagrammen ovan att i skrivuppgift 1 använder alla flickor i icke-CLIL-gruppen någon nominalisering i sina texter vilket gör att gruppens lägsta värde inte är 0 utan 0,3 %. Detsamma gäller för pojkar CLIL i skrivuppgift 4 där deras lägsta värde inte är 0 utan 0,4 % förekomst av nominaliseringar i texterna. Detta diskuteras i avsnitt 8.3. I bilaga 8 återfinns de fyra undergruppernas medelvärden samt max- och minivärden, standardavvikelse och medianvärden för användningsgraden av nominaliseringar i de fyra skrivuppgifterna.

7.3.2 Verb i passiv form

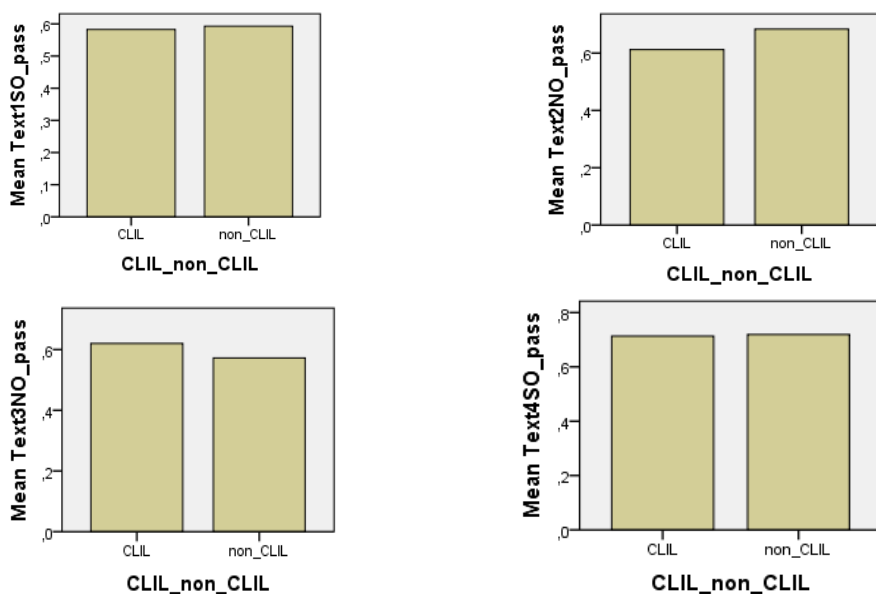
Användningsgraden som redovisas är uttryckt i form av procentandel av den totala andelen ord i texterna. Användningen av verb i passiv form är ännu lägre än användningen av nominaliseringar i elevernas texter, något som diskuteras vidare i kapitel 8. Ingen grupp har ett medelvärde över 1 vid något skrivtillfälle

och flera elever använder inga verb i passiv form alls i sina texter. Två elever, båda pojkar CLIL, har 0 % i båda texterna de har skrivit, annars är mönstret detsamma som för nominaliseringar, att elever har en blandning av eventuellt 0 procent verb i passiv form i en text men i andra texter finns det en eller ett par verb i passiv form vilket ger utslag. Några elever har maxvärden på 2,4 – 3,2 % vilket är lägre än maxvärdena av användningsgraden av nominaliseringar. Dessa elever finns i samtliga grupper. Dock bör påpekas att detta är förekomst i procent av hela texter vilket innebär att användningsgraden är låg.

Ett *Independent Samples T-test* visar ingen signifikans mellan grupperna CLIL och icke-CLIL i någon av de fyra skrivuppgifterna vilket sisas i tabell 30.

TABELL 30. Signifikansvärden av verb i passiv form i skrivuppgift 1-4 för grupperna CLIL/icke-CLIL

Mellan grupper CLIL/icke-CLIL	.898	.426	.602	.956
----------------------------------	------	------	------	------



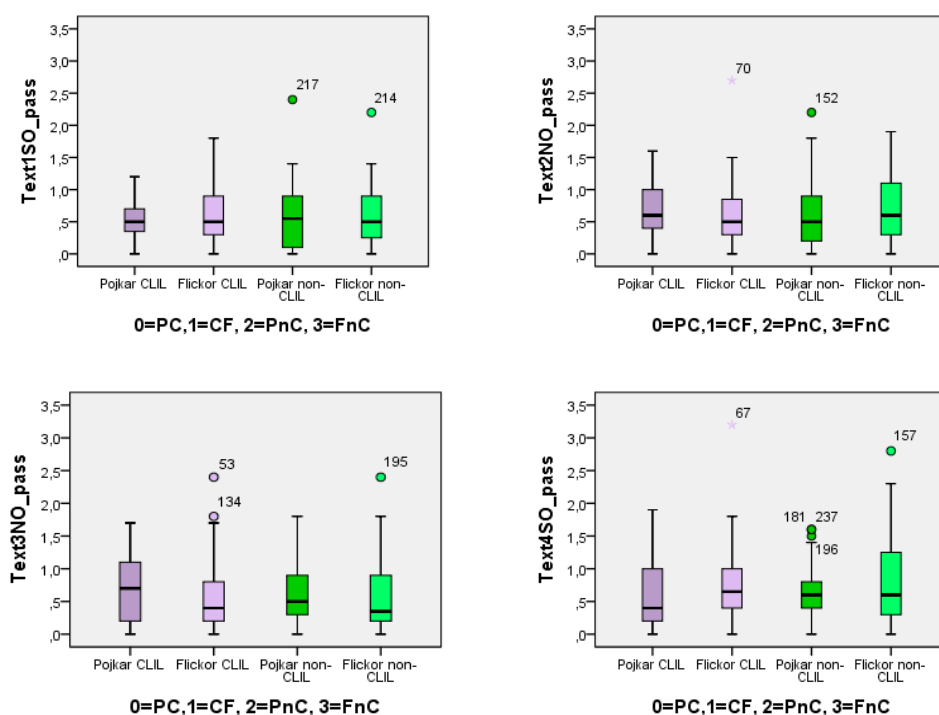
FIGUR 15. Stapeldiagram för användningen av verb i passiv form i skrivuppgift 1-4 för respektive grupp

Genom att visa resultaten i form av stapeldiagram i figur 15, synliggörs dels att medelvärdena mäts i tiondelar på y-axeln vilket anger en mycket låg användningsgrad, dels att CLIL-gruppen som kollektiv ligger lägre än icke-CLIL-

gruppen i skrivuppgifte 1, 2 och 4 vilket inte återfinns i någon av de andra mätningarna (se 8.2 och 8.3).

Observera y-axelns gradering med tiondelar vilket innebär att skillnaderna som syns dock är väldigt små, därav att ingen statistisk signifikans finns här.

Vid en statistisk analys av resultaten för de fyra undergrupperna finns inte heller där någon signifikans mellan PC, FC, PnC, FnC, gällande användningsgraden av variabeln. Elever med högre förekomst av verb i passiv form återfinns bland flickor CLIL, flickor icke-CLIL och pojkar icke-CLIL. Flickorna i de båda undergrupperna, FC och FnC, har en högre andel verb i passiv form än pojkar CLIL i alla skrivuppgifter utom i skrivuppgift 3. Detta synliggörs i lådagrammen i figur 16 samt i bilaga 10 där de fyra undergruppernas medelvärden samt max- och minivärden, standardavvikelse och medianvärden återfinns för de fyra skrivuppgifterna.



FIGUR 16. Lådagram för procentuell andel verb i passiv form i skrivuppgift 1-4 för respektive grupp

Såväl användningsgrad som spridningen för CLIL-pojkarna ligger vid mätningen av denna variabel lägre än vad gruppens resultat gör vid andra gruppjämförelser.

Medianvärdet och även medelvärdet för gruppen PC är lägst av de fyra grupperna i skrivuppgift 1 och 4, vilket inte återfinns i någon av de andra mätningarna. Gruppen flickor icke-CLIL har här det högsta medelvärdet i två av skrivuppgifterna, 2 och 4, vilket gruppen inte når i något annat variabelvärde. Detta diskuteras vidare i 8.2.

7.3.3 Akademiska ord

Inledningsvis gjordes en manuell analys av andelen akademiska ord från de båda korpusordlistorna 1 och 2 som jämfördes med *En svensk akademisk ordlista, ESAO* (Jansson et al. 2012, Sköldbberg & Johansson Kokkinakis 2012). Därefter fördes orden som matchade över till en gemensam lista som akademiska ord. Därpå kördes den ordlistan mot ordlistorna med generella ord som har använts i studien för de lexikala frekvensprofilerna. De ord i *ESAO* (ibid.) som också återfinns bland de 5000 av de mest frekventa i svenska språket har tagits bort så att samma ord inte redovisas vid flera tillfällen i olika analyser.

Antalet akademiska ord som inte ingår bland de 5000 mest förekommande orden i svenska språket och vars förekomst även finns i *ESAO* (ibid.) är totalt 363 (av 652 huvudord) ord som här representerar de akademiska, några av dem redovisas i Bilaga 9. Vissa ord finns i *ESAO* (ibid.) men återfinns inte i någon av elevtexterna exempelvis ordet *avhandling*.

Användningsgraden av akademiska ord är över lag väldigt låg. I flera elevtexter återfinns inga alls utifrån de analyser som har gjorts här. Användningsgraden som redovisas är uttryckt i form av procentandel av den totala andelen ord i texterna. Inte heller här finns någon elev som har 0 % i alla skrivuppgifter som de har deltagit i. Endast en elev, som bara har skrivit en text, hade 0 % i den, övriga deltagande elever kan ha 0 % i någon text men har använt en eller ett par akademiska ord i övriga texter. Två icke-CLIL-flickor har 0 % i tre av skrivuppgifterna och 0,2 % respektive 0,5 % i den fjärde. Den högsta användningsgraden är 3,8 % i skrivuppgift 3 och 3,5 % i skrivuppgift 2. Båda resultaten är av samma CLIL-elev som för övrigt är samma elev som hade den högsta användningsgraden av nominaliseringar samt som har höga resultat gällande övriga kategorier också.

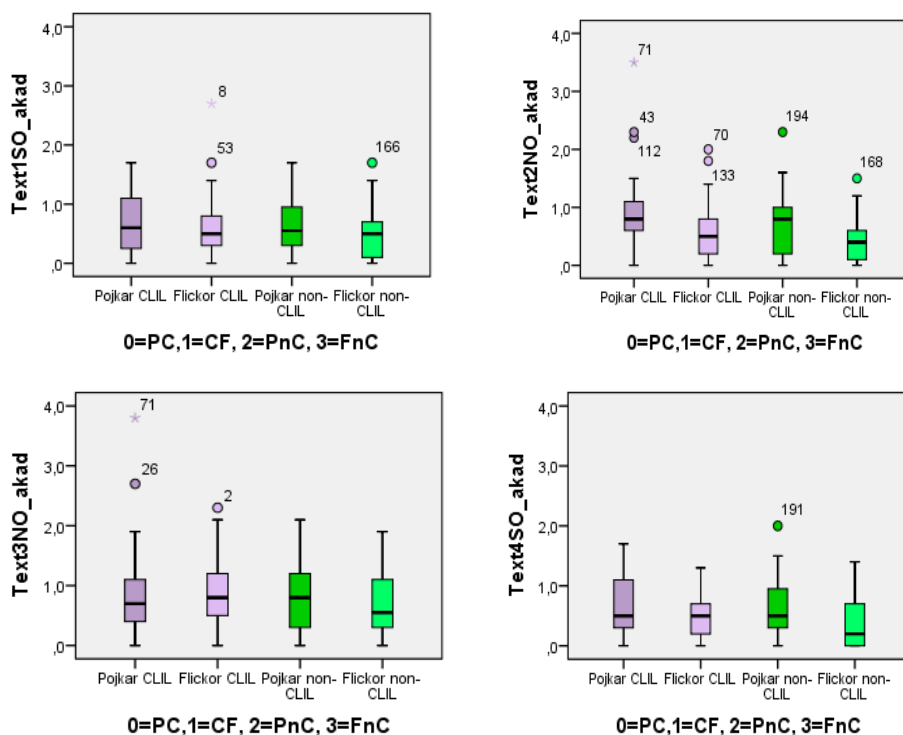
Vid *T-test* visas ingen signifikans mellan de två grupperna CLIL/icke-CLIL i någon av de fyra skrivuppgifterna, se tabell 31.

TABELL 31. *Signifikansanalys av akademiska ord i samtliga skrivuppgifter 1-4, p-värde*

Skrivuppgift	1	2	3	4
Mellan grupper CLIL/icke-CLIL	.288	.502	.121	.769

Vid en manuell analys av Excelarken där alla elevers resultat för de fyra skrivuppgifterna finns införda blir det tydligt genom färgmarkeringar av höga och låga resultat att det är samma elever som återkommande har de högsta resultaten (för jämförelse se Apelgren u.u. 2018).

Mellan de fyra undergrupperna, PC, FC, PnC och FnC finns signifikans i skrivuppgift 2, då mellan pojkar CLIL och flickor CLIL ($p = .002$) samt mellan pojkar CLIL och flickor icke-CLIL ($p = .002$). Detta synliggörs i lådagrammet för uppgiften nedan där spridningen är störst hos PC med flest extremvärden.



FIGUR 17. Lådagram för procentuell andel akademiska ord i skrivuppgift 1-4 för respektive grupp

TABELL 32. Medelvärden för akademiska ord i skrivuppgifter 1-4 CLIL-icke-CLIL/F-P.

Grupp	Medel uppg.	1	2	3	4
CLIL pojkar		0,670	1,048	1,041	0,662
CLIL flickor		0,604	0,552	0,871	0,509
icke-CLIL pojkar		0,629	0,768	0,871	0,606
icke-CLIL flickor		0,458	0,448	0,640	0,411

I både lådagrammen i figur 17, och i tabell 32, visas att medelvärdena för andelen akademiska ord är högst i gruppen PC i samtliga skrivuppgifter. Observera extremvärd för elev (n= 71) i uppgift 2 och 3. Dock är det inte stora skillnader då förekomsten anges i procent av texterna och det handlar om hundradelar vilket visas i tabellen. Medelvärdet är högst i skrivuppgift 2, *Bilfri stad* följt av skrivuppgift 3, *Folkhälsoproblem*. Gruppen PnC har de näst högsta värdena i skrivuppgift 1, 2 och 4 och i skrivuppgift 3 har de samma medelvärde som gruppen FC. Flickor icke-CLIL har de lägsta medelvärdena i samtliga fyra skrivuppgifter. Medelvärdena ökar inte över tid, från termin 1 då den första skrivuppgiften skrevs till termin 5 då den fjärde skrivuppgiften genomfördes.

Liksom för tidigare resultat är det skrivuppgift 3 som tycks ha engagerat eleverna, följt av skrivuppgift 2 där andelen akademiska ord är något högre även om den förefaller låg och inte ens når upp till 1 % av texternas ord.

I tabell 33, visas medianvärdena för samtliga fyra undergrupper vid de fyra skrivtillfällena. Här syns att medianvärdena är desamma för flera av grupperna, exempelvis i skrivuppgift 4.

TABELL 33. Medianvärden för akademiska ord i skrivuppgifter 1-4 CLIL-icke-CLIL/F-P.

Grupp	Median uppg.	1	2	3	4
CLIL pojkar		0,600	0,800	0,700	0,500
CLIL flickor		0,500	0,500	0,800	0,500
icke-CLIL pojkar		0,550	0,800	0,800	0,500
icke-CLIL flickor		0,500	0,400	0,550	0,200

Vid analys av medianvärdena i de fyra undergrupperna i skrivuppgift 1-4 synliggörs att grupperna ligger nära varandra i skrivuppgift 1, att båda pojkgrupperna har samma högre värde i skrivuppgift 2, att FC och PnC har samma högre värde i skrivuppgift 3 samt att alla grupperna har lägre medianvärde i skrivuppgift 4 där PnC har det lägsta av samtliga grupper i samtliga uppgifter. Detta illustreras även i figur 17. Flickor icke-CLIL ligger lägst i tre av fyra skrivuppgifter.

En jämförelse med medelvärden som redogjordes för i tabell 33 ovan visar att PC inte har de högsta medianvärdena i samtliga uppgifter där de hade de högsta medelvärdena, detta synliggörs genom den spridning som kan ses i lådagrammen i figur 17, där uteliggarna i skrivuppgift 2 och 3 samt den stora spridningen påverkat medelvärdena. I bilaga 12 återfinns de fyra undergruppernas medelvärden samt max- och minivärden, standardavvikelse och medianvärden gällande användningsgrad av akademiska ord i de fyra skrivuppgifterna.

7.4 Lexikala profiler

Alla 520 elevtexter har genomgått lexikal profilering i programvaran *AntWord-Profiler* (Nation & Anthony, 2016). Begreppet *token coverage* anger procentuell andel av texten som täcks av de 5000 orden utifrån den frekvensbaserade ordlistan (se figur 1 och 2, samt figur 20 och 21).

I bilderna som visar varje elevtext, färgas orden beroende på vilken frekvens, vilken *Level*, vilken frekvensnivå, de tillhör. Utöver de 5000 mest frekventa orden som färgläggs är en del ord fortfarande svarta och får ingen färg. Dessa ord representerar *Level 0/off-list-ord/Other* som kan vara felstavningar, siffror, namn ämnesrelaterade ord eller lågfrekventa ord. Vid manuell granskning av varje elevtextbild har det blivit synligt hur elevernas skriftliga vokabulär fördelas mellan högfrekventa ord och lågfrekventa ord. Även omfattningen av stavfel har synliggjorts, dock är dessa få (se 7.1.2). Namn är också ovanliga i elevtexterna men finns med i skrivuppgift 2 – *Bilfri stad* där flera elever beskrivit hur de vill att deras stad skall se ut om den blir bilfri och har då angett respektive stads namn samt torg- och vägnamn. De siffror som förekommer är också få. Detta sammantaget innebär att orden som synliggörs som svarta i *Level 0/off-list-ord/Other* har kunnat studeras och analyserats manuellt. Alla frekvensnivåer rapporteras med procentandel av antalet ord som texterna innehåller. Dessa resultat har förts in i Excelark samt i SPSS för vidare analys.

I avsnitt 7.4.1 redovisas analys av i vilken omfattning eleverna använder sig av de 5000 mest frekventa orden i det svenska språket genom *Token coverage* samt de lågfrekventa orden som synliggjorts i *Level 0*, det vill säga i vilken grad grupperna använder sig av de mest frekventa respektive de minst frekventa orden. I gruppjämförelsen av CLIL och icke-CLIL finns även en rapportering av *Level 1* med, detta då det vid en manuell granskning av resultaten visar sig att elever som i hög grad använder sig av de 5000 vanligast förekommande orden och därmed inte använder många lågfrekventa ord har en stor andel av orden i det första frekvensnivån, *Level 1*. Det är alltså inte så att elever som använder de vanligaste orden har en spridning fördelad mellan de olika frekvensnivåerna 1 – 5 utan de flesta orden återfinns för många elever bland de 1000 mest frekventa i nivå 1/*Level 1*.

Resultaten för alla elevers användningsgrad av alla frekvensband finns rapporterade och kan redovisas, dock avgränsas beskrivningen i denna uppsats med de två variablerna *Token coverage* och *Level 0*, då de kan sägas ge en uppfattning om eleverna produktiva skriftliga vokabulär. Observeras bör att här läses resultaten omvänt, det vill säga att ett lågt värde för de mest frekventa orden innebär ett högt värde för de minst frekventa (se figur 1, 2, 20 och 21). Om elever har ett större ordförråd har de tillgång till fler ord än de mest frekventa i svenska språket och har därmed ett lägre procenttal för *Token coverage* samt ett högre procenttal för *Level 0*, vilket synliggörs i denna analys.

7.4.1 Täckningsgrad

Inledningsvis beskrivs de statistiska analyserna för de två grupperna CLIL och icke-CLIL gällande den totala täckningsgraden för 5000 högfrekventa ord och *Level 0/Other* för övriga ord. Därefter beskrivs SPSS-analyser för de fyra undergrupperna PC, FC, PnC och FnC samt resultaten för medelvärden, max- och minivärden samt standardavvikelse. Slutligen redogörs även för en manuell analys på individnivå där mönster av höga respektive låga resultat återkommer liksom vid de tidigare rapporterade kvantitativa och korpuslingvistiska mätningarna.

När det gäller den totala täckningsgraden, det vill säga den genomsnittliga användningen av de 5000 av de mest frekventa orden visar ett *Independent Samples T-test* ingen statistisk signifikans mellan grupperna CLIL och icke-CLIL i någon skrivuppgifterna, se tabell 34, dock är det nära i skrivuppgift 1 där p -värdet är .059.

TABELL 34. *Signifikansanalys av täckningsgrad i samtliga skrivuppgifter 1-4*

Skrivuppgift	1	2	3	4
Mellan grupper CLIL/icke-CLIL	.059	.698	.809	.367

Om man analyserar täckningsgraden, användningen av *Level 0/Other*, det vill säga de lågfrekventa orden, liksom analysen av användningen av ord som hör till frekvensnivå 1, de 1000 vanligast förekommande orden blir resultaten mellan grupperna inte heller genomgående signifikanta, se tabell 35 och 36.

TABELL 35. *Signifikansanalys av Level 0/Other i samtliga skrivuppgifter 1-4*

Skrivuppgift	1	2	3	4
Mellan grupper CLIL/icke-CLIL	.056	.684	.945	.884

TABELL 36. *Signifikansanalys av Level 1 i samtliga skrivuppgifter 1-4*

Skrivuppgift	1	2	3	4
Mellan grupper CLIL/icke-CLIL	.021	.752	.323	.146

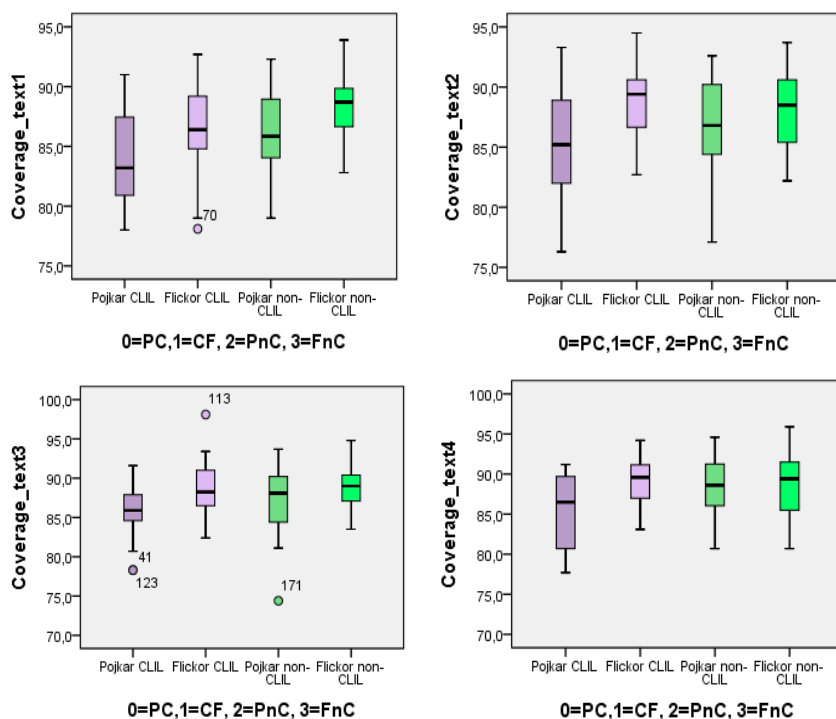
Det är i skrivuppgift 1, *Hälsa och välbefinnande*, som det finns en statistisk signifikans mellan CLIL och icke-CLIL-grupperna gällande användningsgraden av de 1000 mest frekventa orden i det svenska språket (Tabell 36, $p = .021$). Resultaten för

användningsgraden av ord utanför nivå 1-5, *Level 0/Other* ligger nära (Tabell 35, $p = .056$), och för samma skrivuppgift är *Token coverage* (Tabell 34, $p = .059$).

En analys av de fyra undergruppernas resultat ger däremot signifikanta värden gällande täckningsgraden för 5000 av de vanligast förekommande orden mellan de båda CLIL-grupperna och kontrollgrupperna.

För skrivuppgift 1 visar *ANOVA Tukey HSD post hoc* signifikanta värden mellan pojkar CLIL och flickor CLIL ($p = .006$) samt mellan pojkar CLIL och flickor icke-CLIL ($p = .000$). Mellan flickor CLIL och de båda icke-CLIL-grupperna är det däremot ingen signifikans, inte heller mellan de båda icke-CLIL-grupperna.

I skrivuppgift 2 är resultaten liknande; mellan PC och FC finns signifikans ($p = .000$), liksom mellan PC och FnC ($p = .011$), däremot inte mellan någon av de andra grupperna. Resultaten upprepas i skrivuppgift 3 med signifikans mellan PC och FC där det finns signifikans ($p = .005$), liksom mellan PC och FnC ($p = .006$) men inte heller här mellan någon av de andra grupperna. I skrivuppgift 4 finns ingen signifikans mellan någon av grupperna.



FIGUR 18. Lådagram, procentuell täckningsgrad i skrivuppgift 1-4 för respektive grupp

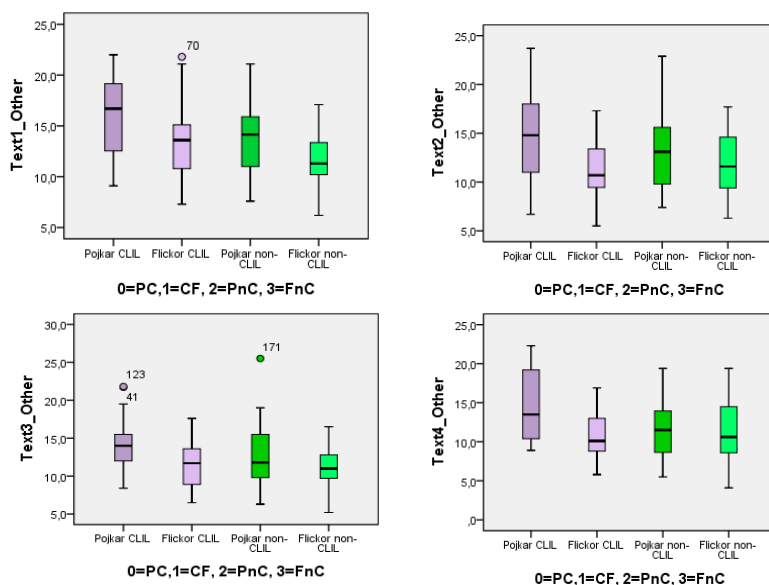
I lådagrammen i figur 18 bör påminnas att resultaten skall läsas ”omvänt”, det vill säga att Pojkar CLIL som har det lägsta medelvärdet i samtliga fyra figurer och dessutom två uteliggare i skrivuppgift 3 därmed använder den minsta andelen högfrekventa ord vilket i sin tur innebär att deras genomsnittliga värde för andelen lågfrekventa ord är högst.

Spridningen inom grupperna är högst i skrivuppgift 2, *Bilfri stad*, och minst i skrivuppgift 3, *Folkhälsoproblem*. En orsak till detta kan vara att i skrivuppgift 2 återfinns flera engångsord som: *bilanvändande, bilanvändandet, bilanvändningen, bilbarnstol, bilbranschen, bildebatten* och så vidare (35 graford som är bilrelaterade). Därefter fortsätter det med flera engångsord som är sammansatta och bussrelaterade, cykelrelaterade, elrelaterade, energirelaterade och så vidare.

Om man analyserar resultaten av *Level 0/Other* för de fyra undergrupperna PC, FC, PnC och FnC visas samma mönster; För skrivuppgift 1 visar *ANOVA Tukey HSD post hoc* signifikanta värden mellan PC och FC ($p = .006$) samt mellan PC och FnC ($p = .000$). Mellan FC och de båda icke-CLIL-grupperna är det ingen signifikans, inte heller mellan de båda icke-CLIL-grupperna. I skrivuppgift 2 är resultaten liknande; mellan PC och FC finns signifikans ($p = .000$), liksom mellan PC och FnC ($p = .011$), däremot inte mellan någon av de andra grupperna. Resultaten upprepas i skrivuppgift 3 där det återigen finns signifikans mellan PC och FC ($p = .004$), liksom mellan PC och FnC ($p = .004$) men inte heller här mellan någon av de andra grupperna. I skrivuppgift 4 finns signifikans mellan PC och FC ($p = .001$), mellan PC och FnC ($p = .023$), mellan PC och PnC ($p = .017$) men inte mellan övriga grupper; FC, PnC och FnC. I figur 19 nedan synliggörs detta i lådagram.

I lådagrammen i figur 19 läses resultaten som vanligt, det vill säga här anger de höga resultaten att eleverna i hög grad använder lågfrekventa ord. Det som syns är att båda pojkgrupperna gör så i högre utsträckning än flickgrupperna.

Vid manuell analys av vilka elever som använder mest vardagliga ord visar det sig att av alla fyra skrivuppgifter som genomförts i de åtta klasserna (=32) är det 29 flickor och 3 pojkar som använder lägst andel av de lågfrekventa orden. Det innebär alltså att de använder en hög andel högfrekventa, vardagliga ord. Samma analys synliggör att två pojkar CLIL har en *täckningsgrad* under 80 % i samtliga fyra skrivuppgifter, det vill säga de använder mer än 20 % av ord som förekommer utanför nivå 1-5 i samtliga texter de har skrivit inom CLISS-projektets ram. De använder alltså fler *off-list-ord* än övriga elever.



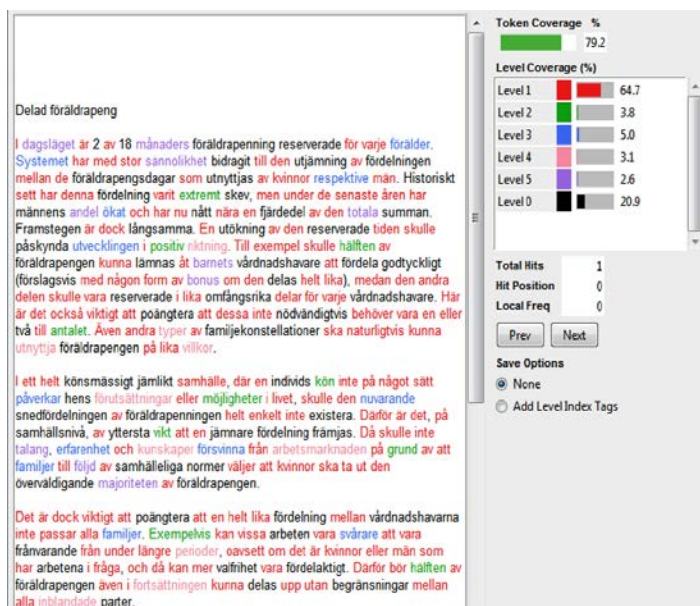
FIGUR 19. Lådagram för andel off-list ord i skrivuppgift 1-4 för respektive grupp

För samtliga 520 undersökta texter finns lexikala frekvensbilder i form av figur 1 och 2 (se 6.3). Genom färgmarkeringarna synliggörs frekvens och spridning av förekommande vokabulär. I figur 20 och 21 synliggörs två andra elevtexter än i tidigare figurer (1 och 2). Båda är, liksom de tidigare figurerna från en icke-CLIL-elev och en CLIL-elev.



FIGUR 20. Lexikal profil för icke-CLIL-elev, skrivuppgift 4.

Figur 20 visar texten för skrivuppgift 4, *Helt delad föräldrapeng*, skriven av en icke-CLIL-elev.



FIGUR 21. *Lexikal profil för CLIL-elev, skrivuppgift 4.*

Figur 21 synliggör en lexikal profil från en CLIL-elevs texter från skrivuppgift 4 illustrerar ännu en gång hur vokabulären kan visualiseras och hur olika den ser ut beroende på elevernas användning av och tillgång till vokabulär. Det är framför allt användningsgraden av ord utanför de 5000 vanligaste och ord högfrekventa, vardagliga ord som synliggörs. Detta diskuteras i 8.4.

I bilaga 13 finns en jämförelse mellan grupperna CLIL och icke CLIL i samtliga skrivuppgifter för medelvärdena av andelen ords förekomst i *Level 1* och *Level 0/Other*. Här visas även minimivärden och maxvärden i grupperna vid de olika skrivtillfällena 1-4.

8. Diskussion

I denna del diskuteras resultaten som har framkommit i studien i relation till syfte och frågeställningar samt till tidigare forskning. I avsnitt 8.1 diskuteras inledningsvis metodval, därefter redovisas i avsnitt 8.2 en variabel i taget där resultat jämförs med tidigare forskning. Resultaten diskuteras senare även sammanfattningsvis. I 8.3 diskuteras om, och i så fall hur, studiens resultat samvarierar. Kapitlet avslutas med pedagogiska implikationer i avsnitt 8.4 och förslag till vidare forskning i avsnitt 8.5.

Syftet att identifiera och definiera, samt undersöka och beskriva, de deltagande gymnasieelevernas skriftliga svenska vokabulär i de av CLISS-projektet insamlade skrivuppgifterna under tre år, har gjorts genom kvantitativa ytstrukturella mätningar samt kvantitativa ämnesrelaterade mätningar.

Undersökningar av *textlängd*, *ordmedellängd*, *andel långord*, *ordvariationsindex* – *OVI*X har möjliggjort textanalys av elevtexterna över tid och därmed även möjliggjort jämförelser med tidigare studier, i huvudsak svenska sådana. Även de korpuslingvistiska mätningarna av *nominaliseringar*, *verb i passiv form* och *akademiska ord* för att se vokabulären ur ett systemiskt-funktionellt perspektiv, redovisas över tid och även där möjliggjort jämförelser med tidigare studier. De olika variablerna har identifierats utifrån tidigare forskning där ett sociokulturellt perspektiv och ett systemiskt-funktionellt perspektiv samverkar som teoretiskt ramverk för studien.

Genom manuella och statistiska undersökningar av de 520 elevtexterna har variablerna beskrivits och analyserats i egna avsnitt samt illustrerats genom olika diagram och tabeller.

8.1 Metoddiskussion

När elevtexter studerats är det ofta ur perspektiv som innefattar mer av kvalitativ forskning där textanalyser görs på djupnivå av olika aspekter eller av deltagarnas skrivprocess eller av annat som undersökts. Det är en av orsakerna till valet att här pröva kvantitativa metoder. Genom att dessutom göra statistiska analyser har resultaten kunnat påvisa statistisk signifikans eller inte. De olika variablerna kan var och en motiveras och tillsammans anser jag att de skapar en förståelse för elevers produktiva ordförråd som tar sig uttryck i de 520 texter som analyserats i studien.

Granskningen av elevtexter med kvantitativa textuella måttenheter har en lång tradition och därmed är jämförelser över tid möjliga. Hur långt bakåt i tiden man bör gå för att göra den typen av jämförelser kan diskuteras men då flera av de tidigare studierna som refereras till har gjort sina jämförelser med *Gymnasist-svenskans* undersökning från 1977 (Hultman & Westman) har jag valt att göra detsamma. Detta innebär en tidsperiod på 40 år där gymnasieelevers skrivande synliggörs med kvantitativa ytstrukturella mätningar. De didaktiska frågorna Vad? Hur? och Varför? kan därmed diskuteras gällande förändringar som visas ha skett över tid.

Att ha med jämförande studier där nationella prov från årskurs 9 samt texter av yngre elever också ingår motiveras dels genom att kvantitativa mått kan vara den del som lärare ser i elevtexter, såväl textlängd som variablerna ordlängd och andelen långord. Dessa har synliggjorts och lyfts fram i studien som en del av en vokabulär. Dels att eleverna lärare möter kommer från ett annat stadium alternativt skall vidare till nästa. Då är det som lärare viktigt att ha kännedom om, och få förståelse för, elevers kunskapslägen och utvecklingsmöjligheter.

Till detta har den lexikala profileringen, där fördelningen mellan hög- och lågfrekventa ord, synliggjorts på ett sätt som kan med dator tekniska hjälpmedel skapa en medvetenhet om hur dessa påverkar text samt ger förståelse för vokabulär vilket kan gynna både lärare och elever.

Att dessutom synliggöra och undersöka aspekter som kännetecknar det vetenskapliga och akademiska språket utifrån en systemisk-funktionell språksyn tillför en insikt om hur nominaliseringar, verb i passiv form samt att vissa ord har en högre användningsgrad i akademiska texter (Halliday 1993, Martin & Rose 2008, Schleppegrell 2004). Hur dessa ord identifieras och används är en värdefull kunskap som kan innebära en utvecklingspotential för lärare i det vardagliga skolarbetet där mediering och appropriering av skolspråket pågår (Säljö 2000, 2015).

Beslutet att inte undersöka huruvida alla 652 ord som återfinns i *ESAO* (Jansson et al. 2012, Sköldberg & Johansson Kokkinakis 2012) kunde ses i elevtexterna då 292 av dem återfinns i frekvensbanden för 5000 av de mest frekventa orden i svenska språket, innebär att resultaten rapporteras som akademiska ord

respektive högre frekventa ord men har analyserats som olika kategorier. Det innebär därmed att något annat synliggörs gällande användningsgraden av akademiska ord än den motsvarande studien av delar av materialet (Lim Falk & Holmberg 2016).

En annan aspekt utifrån redogörelsen ovan är om *alla* mätningar i studien behöver vara med och synliggöras som har gjorts i text, stapeldiagram och lådagram. Min motivering är att genom denna inkludering synliggörs komplexiteten i ordförrådsbegreppet gällande en produktiv vokabulär. Min förhoppning är att metoden och resultaten kan bidra till en ökad insikt för vilka olika språkliga drag som finns och därmed kan förstås och ingå i språkliga läroprocesser. Vokabulär består inte enbart av receptiva och produktiva ord, utan det har stor betydelse *vilka* ord man lär och att man får en förståelse för i vilka sammanhang de kan användas liksom *hur* de används. Här samverkar en sociokulturell språksyn med en systemisk-funktionell sådan i hög grad, där skolan kan bidra med kunskaper om det vetenskapliga, akademiska sättet att skriva och hur betydelse och form samspelar, liksom krav och förväntningar på det diskursiva skrivandet (Andersson Varga 2014, Ask 2005). Till detta kan läggas innebörden av att *kunna* ett ord som beskrivs mer ingående av Nation (2013), Oakhill *et al* (2015) och Gustafsson och Håkansson (2017), men som inte undersökts i denna studie.

Att ha med genus som en variabel i studien hänger delvis samman med att det finns med som en av de övergripande frågeställningarna i CLISS-projektet men även att det, som nämndes inledningsvis, har funnits med som en frågeställning i min egen praktik.

De kvantitativa mätningarna som har genomförts i studien är gjorda i ett forskningssyfte där resultaten visar mönster och tendenser som kan föranleda slutsatser och ge kunskaper. Kvantitativa mätningar bör inte användas som enskilda bedömningsunderlag utan samverka med flera aspekter för formativ och summativ bedömning (se Erickson 2016). Däremot kan studiens resultat synliggöra och skapa förståelse för komplexitet i en produktiv vokabulär. Vid textanalys som omfattar bedömning finns ytterligare aspekter som bör lyftas fram (se exempelvis *Normprojektet*, Berge & Skar 2015). Dock är förståelsen för en vokabulärs sammansättning och hur en sådan kan synliggöras och därmed bidra till kunskaper om vikten av att utveckla och förändra beroende på användningsområde vara värdefull då vokabulär anses som en viktig komponent i framgångsrikt skrivande (se Skolverkets *Läslyft*, Lindberg 2017, *Normprojektet*, Matre & Solheim 2015).

I föreliggande licentiatuppsats görs inga regressionsanalyser, det vill säga jämförelser av, och i så fall i vilken omfattning, specifika elever har utvecklat sina resultat från den första skrivuppgiften till den fjärde. Detta skulle dock vara möjligt att göra då det finns 63 elever som har skrivit alla fyra skrivuppgifterna. En manuell analys av dessa resultat visar dock inte någon progression på individnivå men en utförligare analys kan göras genom exempelvis en *Growth Curve*

Model. Jag har heller inte gjort korrelationsanalyser i SPSS vilka skulle kunna genomföras för att se eventuella samband mellan olika variabler. Detta utvecklas i avsnittet om förslag till vidare forskning.

Deltagarantalet kan synas lågt då det vid de fyra skrivtillfällena har varit mellan 50 % och 60 % av det totala elevantalet som har deltagit. Detta beroende på att skrivuppgifterna har genomförts vid endast ett tillfälle och därefter har insamlingen räknats som avslutad. I ordinarie undervisning har lärare möjlighet att få frånvarande elever att skriva på en annan tid, en annan dag och så vidare men den möjligheten har inte funnits inom CLISS-projektets insamlingsram. Materialet kan ändå sägas vara tillräckligt stort för att kunna visa resultat över tid och genom de olika analysmetoderna säkerställa vissa definitiva resultat.

Av informanterna är det en klass av åtta som har motiverats lite extra av lärare och rektor då intresset för CLISS-projektet minskade något mitt i den treåriga projektiden. Övriga klasser har hela perioden tagit sig an alla olika uppgifter med en ärlig och uppriktig inställning. Endast en text har räknats bort av de insamlade texterna och inte tagits med, då den enbart bestod av upprepade tramsord. Övriga texter uppvisar alla, vid manuell genomläsning, ett innehåll kopplat till respektive uppgift, även de kortaste texterna. Att vi dessutom har haft tillgång till en klass nationella prov i svenska där samtliga kvantitativa och korpuslingvistiska mätningar av variabler som ingår i denna studie, trots olika tidsramar för skrivtid, visar lägre värden gällande alla analystyper, bidrar med validitet till studiens resultat.

Mängden data som har analyserats har varit omfattande och då alla resultat har bearbetats både manuellt i Excel och därefter genom SPSS har felaktigheter emellanåt upptäckts i den manuella hanteringen vilket har lett till ett otal omräkningar. Detta kan givetvis ses som ett led i uppsatsens reliabilitet men även att det kan finnas fel som har missats. Genom materialets storlek torde dock eventuella felräkningar inte påverka de tendenser och mönster som uppmärksammas.

8.2 Resultatdiskussion, jämförelse med tidigare studier

För att besvara studiens första fråga om hur en skolrelaterad skriftspråklig vokabulär framträder har kvantitativa textlingvistiska mått som *textlängd*, *ordmedellängd*, *långordsfrekvens* och *ordvariation*, *OVIX* identifierats och analyserats. Studiens andra fråga om, och i så fall hur, faktorerna programval, genus och betyg kan samvariera med resultaten besvaras och diskuteras parallellt. I undersökningarna som refereras till har resultaten prövats statistiskt i Vagles studie (2005) samt i Ribecks studie (2015). I de övriga refererade studierna redovisas medelvärden utan statistisk analys.

Analysen av dessa kvantitativa ytstrukturella mätningar visar att gällande **textlängd** så är den genomsnittliga textlängden för CLIL-eleverna längre än för

kontrollklasserna med medelvärden på 588 graford och 436 graford för respektive grupp. Statistisk signifikans visas i samtliga skrivuppgifter. CLIL-eleverna skriver alltså längre texter än CLIL-eleverna genomgående i studien. Resultaten följer därmed en tidigare studie vid jämförelse av CLIL och icke CLIL (Lim Falk 2008). Dock var texterna i den studien längre med genomsnittliga textlängder på 762 graford för CLIL-eleverna respektive 644 ord för kontrollgruppens elever. Detta kan bero på dels en längre skrivtid då det var texter som skrevs vid ett nationellt prov samt att provsituationer påverkar elever att skriva längre texter då texterna ingick som ett bedömningsunderlag vid betygssättning vilket skrivuppgifterna inte har gjort i denna studie. En annan faktor kan vara att det är flera år mellan studierna och liksom i andra mätningar (ex. PISA) kan resultaten här följa den trend som varit.

Flickornas medelvärde för texter är i denna studie längre än pojkarnas när det gäller CLIL-gruppen. I icke-CLIL-gruppen skriver flickorna i genomsnitt längre än pojkarna vid ett tillfälle, och pojkarna längre vid tre av de fyra skrivtillfällena. Ser man på skrivuppgifterna som ett kollektiv (se Bilaga 6) så skriver flickorna genomgående längre texter vilket till viss del överensstämmer med tidigare studier (Ciolek Laerum 2009, Hultman & Westman 1977, Lim Falk 2008, Olevard 1999, Vagle 2005, Östlund-Stjärnegårdh 2002). I denna studie är det CLIL-flickornas resultat som påverkar detta. I skrivuppgift 3 och 4, där skrivuppgift 3 är den som har de högsta medelvärdena för antalet graford och skrivuppgift 4 den som har det lägsta medelvärdet för antalet graford finns det statistisk signifikans för att flickgruppen som kollektiv skriver längst (skrivuppgift 3 => p -värde = .032, skrivuppgift 4 => p -värde = .015).

Trots att skrivtiden skiljer mellan de olika studierna som refereras till så överensstämmer resultaten gällande CLIL-elever som skriver längre än icke-CLIL-elever och att flickor till övervägande del skriver längre än pojkar. De statistiska analyserna av textlängd ger statistiskt signifikanta resultat både gällande skillnaderna mellan grupperna CLIL och icke-CLIL samt mellan undergrupperna PC, FC, PnC och FnC. Dock visas ingen statistisk signifikans mellan CLIL-flickorna och CLIL-pojkar, flickornas medelvärden är högre men det är inte signifikant i någon av skrivuppgifterna.

Medelvärdet för textlängd av uppsatserna i *Gymnasistsvenska* (Hultman & Westman 1977:53) var totalt 588 graford vilket är nästan exakt detsamma som medelvärdet för CLIL-elevernas texter i denna studie som är 589 graford. För flickor i Hultmans & Westmans (ibid.) undersökning var medelvärdet 620 graford och 565 för pojkar. Det högsta medelvärdet här för CLIL-flickorna är 726 graford för skrivuppgift 3 samt 578 graford för CLIL-pojkar vid samma skrivtillfälle. Texterna har i denna studie huvudsak skrivits på dator vilket de inte gjorde i Hultman & Westmans studie (1977).

I två tidigare studier gjorda av Olevard (1999) och Ciolek Laerum (2006), med nationella provtexter skrivna av årskurs 9, visar resultaten av antalet graford i

snitt (för både flickor och pojkar) ha ökat i omfång vid undersökningarna som är gjorda från åren 1987, 1996 (Olevarud 1999) samt från år 2006 (Ciolek Laerum 2009).

TABELL 37. Resultat från Olevaruds (1999) och Ciolek Laerums (2009) studier

	1987	1996	2006
Flickor	517 ord	805 ord	741 ord
Pojkar	446	517	642
Totalt	482	661	688

TABELL 38. Resultat från Bratts (2014) studie

	1996	2003	2013
Antal ord i genomsnitt:	794	851	583

Ciolek Laerums studie (2009) visar dock att längden för flickornas texter har minskat i omfång mellan åren 1996 och 2006 medan pojkarnas har ökat under samma tid. Förklaringar som ges är att det kanske kan bero på att textlängderna börjar närma sig en maxlängd för vad som hinns med att skriva under angiven skrivtid. En annan förklaring kan vara att flickornas texter har blivit mer kärnfulla genom mer fördjupade resonemang men ytterligare undersökningar behöver göras enligt Ciolek Laerum (2009:17). I Bratts undersökning (2014) ingår ett mindre, slumpvist urval av nationella provtexter från 1996, 2003 och 2013 skrivna av gymnasieelever i årskurs 2 och 3. Där är den genomsnittliga textlängden något högre för texterna från 1996 vid jämförelsen i tabell 38, ytterligare en ökning sker för texterna från 2003 men sedan är det en kraftig nedgång i omfånget för texterna från 2013. Antalet graford i genomsnitt för texterna från 2013 är 583 graford att jämföra med *Gymnasistsvenskans* 588 graford och för denna studie 589. Någon genusjämförelse finns inte i Bratts undersökning då antalet texter är få (24 st.). I de tidiga studierna är texterna skrivna för hand medan de i de senare är skrivna på dator, vilket kan påverka omfånget (Nyström 2000).

Nordenfors (2011) undersökning som också är longitudinell med elevtexter från årskurs 5 till årskurs 9 visar att textlängden ökar med åldern (2011:97).

Motsvarande resultat, med längre texter högre upp i åldrarna syns inte i denna studie där den genomsnittliga textlängden inte uppvisar någon ökning över tid, under de deltagande elevernas tre gymnasieår. Istället verkar textlängden höra samman med den skrivuppgift som har engagerat eleverna mest vilket är nummer 3, *Folkhälsoproblem*, som skrevs under gymnasieelevernas termin 3. Tilläggas bör att svårighetsgraden på de olika skrivuppgifterna, såväl på svenska som på engelska, inte har ökat under projektets gång då ett syfte i projektet var att kunna undersöka progression hos eleverna över tid genom en konstant upp-

giftsnivå. Här skiljer sig skrivuppgifterna från de nationella proven som ökar i svårighetsgrad ju högre upp i skolsystemet de prövas.

En jämförelse med skrivuppgifterna på engelska inom CLISS-projektet gällande textlängd är möjlig då dessa har undersökts av Olsson (2015), dock bör påpekas att ord kan skrivas olika och därmed bli svårjämförbara då det till exempel i engelskan ofta är vanligt med två ord där svenskan för samma ord till ett ord (se Lim Falk 2008:173 ff.). I samtliga skrivuppgifter på engelska skriver CLIL-eleverna längre texter precis som de gör på svenska och statistisk signifikans uppvisas i alla dessa. Även det resultatet överensstämmer med denna studie. I tabell 39 nedan redovisas textmedellängden för de båda grupperna för respektive skrivuppgift på svenska och engelska. De engelska resultaten är hämtade från Olsson (2015:72).

TABELL 39. *Jämförelse av textlängd i skrivuppgift 1-4 skrivna på svenska och engelska (jämför Olssons resultat 2015:72).*

Text=> antal ord	1 Sv	1 Eng	2 Sv	2 Eng	3 Sv	3 Eng	4 Sv	4 Eng
Genomsnittlig textlängd CLIL	573	570	579	587	692	605	512	525
icke-CLIL	441	364	456	348	491	395	357	379

För CLIL-gruppen är den genomsnittliga textlängden i stort sett densamma. I två av skrivuppgifterna, uppgift 1 och 4, skriver CLIL-eleverna marginellt längre texter på svenska och i uppgift 2 och 3 kortare på svenska än på engelska. Skillnaden är störst i skrivuppgift 3 där ämnet *Folkhälsoproblem* engagerade alla elever så att de längsta texterna återfinns där för samtliga grupper. Icke-CLIL-elevernans genomsnittliga textlängd är i skrivuppgift 1-3 på engelska nästan 100 graford kortare än motsvarande skrivuppgifter på svenska. I uppgift 4 är dock medellängden något längre på engelska än på svenska. Dessa resultat av genomsnittlig textlängd visar att CLIL-eleverna kan anses ha tillgång till fler ord när det gäller att skriva på engelska än icke-CLIL-eleverna men spridningen inom grupperna rapporteras inte på individnivå här så där kan variation finnas.

Gällande **ordmedellängd** har pojkar CLIL den längsta genomsnittliga ordmedellängden i samtliga skrivuppgifter. Gruppen flickor icke-CLIL har i alla fyra skrivuppgifter den kortaste ordmedellängden. Variationen mellan grupperna flickor CLIL och pojkar icke-CLIL visar att de har en längre ordlängd än den andra gruppen vid två tillfällen vardera. Den genomsnittliga ordlängden för grupperna CLIL och icke-CLIL varierar mellan 4,53 och 4,68 i de fyra skrivupp-

gifterna, vilket är högre siffror än i KAL-projektet (2005) men där överensstämmelse råder gällande att pojkar har en högre ordmedellängd vid utredande text än flickor, om än liten. I KAL-materialet som helhet (3365 texter) är den genomsnittliga ordlängden 4,02 bokstäver per ord (Vagle 2005:380). På individnivå varierar ordmedellängden i denna studie mellan 3,48 för en pojke CLIL i skrivuppgift 4, och 6,29 för en flicka icke-CLIL i skrivuppgift 3. Spridningen inom grupperna kan alltså vara större än medelvärden, något som även visas i Vagles studie (2005), då gällande pojkarna, vars resultat återfanns bland både de starkaste och de svagaste beroende på texttyp (2005:360).

I Östlund-Stjärnegårds studie (2002) redovisas hur den genomsnittliga ordmedellängden för gymnasieeleverna korrelerar med betyg; IG – 4,49, G – 4,44 VG – 4,90 vilket kan ses stämma relativt väl överens med variationen i denna studie, från de lägsta till de högsta resultaten. En jämförelse görs även med Hultman & Westmans studie (1977) där medeltalet i hela materialet var 5,07. Könsskillnaderna i båda nämnda studier bedöms vara relativt liten, däremot skiljer sig ordmedellängden mellan texttyper, vilket visas i fler studier (Vagle 2005: 381, Östlund-Stjärnegård 2002:86). I Magnusson & Johansson Kokkinakis studie (2008) är medelvärdet för ordmedellängd 4,31 vilket är lägre än här, där CLIL-elevernars totala medelvärde är 4,64 och icke-CLIL-elevernars är 4,58. Skillnaderna är små men kan möjligtvis kopplas till att i KAL-materialet (Vagle 2005) ingår olika typer av gymnasieklasser, liksom i Magnusson & Johansson Kokkinakis studie (2008). I denna studie ingår dels enbart högskoleförberedande klasser, dels CLIL-klasserna vilka som grupp har högre meritvärden redan vid antagningen samt ofta är väl studiemotiverade (se Apelgren 2017, Sylvén & Thompson 2015).

I Ribecks (2015) undersökning av läromedel i SO och NO för högstadiet och gymnasiet ökar ordmedellängden mellan stadierna i samtliga ämnens läromedel. Den genomsnittliga ordmedellängden är där 5,30 bokstäver (ibid., 2015:91). Detta att jämföra med berättande text i samma material som ligger strax under 4,40 och *Sve-Ak-korpusen* (Borin et al. 2012) som ligger på 5,60. Här blir skillnaden mellan läromedel och elevtexter tydlig. Ordmedellängden nämns i flera av de refererade studierna ovan som en skriftspråklig kompetens samt som en variabel som i hög grad samvarierar med betyg och därmed kan betraktas som en kvalitetskompetens som hör samman med ”de två aspekterna variation och specifikation” (Hultman & Westman 1977:76). Detta diskuteras vidare i avsnitt 8.4.

Vid undersökningen av **andelen långord** följer denna studie det som beskrivs som en nedåtgående trend som Östlund-Stjärnegårdh fann i sin studie (2002) av nationella provtexter från 1997. I hennes material var andelen långord i genomsnitt 16 % att jämföra med en undersökning i samma studie av centrala provtexter från 1984 till 1996 som redovisar först en andel långord på 25 % som sedan gått ner till 19 % (2002:87). I Bratts studie (2014) som också jämför nat-

ionella provtexter över tid; 7 stycken från 1996, 7 från 2003 och 10 stycken från år 2013, visar resultaten även där en nedgående trend gällande andelen långord. År 1996 var medelvärdet 23,1 %, år 2003 var medelvärdet 21 % och 2013 var medelvärdet 20,1 %.

Resultaten av andelen långord i denna studie för de 520 elevtexterna är att klasserna ligger på medelvärden för långord mellan 15,3 % för en CLIL-klass (uppgift 2), tätt följd av en icke-CLIL-klass med 15,5 % (uppgift 4), som är de lägsta resultaten samt 20,9 % (uppgift 3) för en CLIL-klass som har den högsta andelen långord. För enskilda individer är variationen ännu större där den som har lägst har 6,15 (PnC) och den som har högst har 31,05 % (PC).

Fyra femtedelar av texterna här har medelvärden strax över det på 16 % som provtexterna hade i Östlund-Stjärnegårdhs studie (2002) att jämföra med medeltal på 20 % och däröver i andra studier (Bratt 2014, Hultman & Westman 1977). Eleverna i denna studie verkar därmed skriva kortare ord, det vill säga fler ord som inte består av sex bokstäver eller mer i jämförelse med tidigare studier som refereras till.

Det innebär att en femtedel av materialet i studien ligger över snittet i jämförelse med en tidigare studie (Östlund-Stjärnegårdh 2002) men att flertalet, det vill säga fyra femtedelar, 80 %, ligger nära resultatet på 16 % som framkom där, med ett genomsnitt på 17,9 % för CLIL-klasserna och 16,9 % för kontrollklasserna och således bryter denna studies resultat inte den nedåtgående trenden som nämns i den jämförande studien gällande användningen av långord i texter skrivna av gymnasieelever (ibid., 2002: 87-88). De längre orden har få förekomster, det vill säga är lågfrekventa, vilket överensstämmer med tidigare forskning (Nation 2013 m.fl.).

Nordenfors (2011) studie visar att resultaten av ordmedellängd och långordsfrekvens starkt hänger samman med texttyp. I textkollektivet som har undersökts i hans studie får sagan 12,7 % i andel långord och skräcknovellen 15 % medan ett faktaarbete når 30,3 %. Vid en närstudie av faktaarbetet visar det sig att skribenten har upprepat vissa ord 10-19 gånger. Nordenfors för resonemanget om måtten för ordmedellängd och andel långord kan anses vara pålitliga och verkligen säga något om textkvalitet. Resultaten i studien visar dock, förutom sambandet med texttyp, att textmåttsvärde ökar med ålder och att måtten generellt är högre i de texter som har fått högre betyg (2011:105–111).

I denna studie syns inte den ökningen av textmåttsvärden över tid utan hänger här samman med skrivuppgifternas inriktning där eleverna verkar ha engagerats mer av att skriva om *Folkhälsoproblem* (uppgift 3) än om *Helt delad föräldrapeng* (uppgift 4). Om det i sin tur bör relateras till ämnen som kanske ligger mer nära dem, eller om det handlar om en viss trötthet där plikt känslan bidrog till att de ändå skrev relevanta texter även i den sista omgången går inte fullt ut att besvara men svårighetsgraden i skrivuppgifterna ökade inte mellan uppgift 3 och 4. Svårighetsgraden på de olika skrivuppgifterna, såväl på svenska som på

engelska, har varit likvärdig under projektets gång då ett syfte i projektet var att kunna undersöka progression hos eleverna över tid genom en konstant uppgiftsnivå. Här skiljer sig skrivuppgifterna från de nationella proven som ökar i svårighetsgrad ju högre upp i skolsystemet de provas. Dock gäller resultaten, oavsett om det är argumenterande eller utredande text, det vill säga det *diskursiva skrivandet*. Andelen långord är trots en genomsnittlig kortare textlängd för båda grupperna CLIL och icke-CLIL i uppgift 4, längre där för CLIL-gruppen och obetydligt kortare för icke-CLIL-gruppen jämfört med i skrivuppgift 2; *Bilfri stad*. Andelen långord kan därmed tolkas korrelera med ämnen där långa ord finns i elevernas produktiva ordförråd som i *Helt delad föräldrapeng* där många sammansatt ord skapas utifrån förledet *föräldra-* (se avsnitt 7.2.3).

Den genomsnittliga ordmedellängden är också högst för CLIL-gruppen i skrivuppgift 4 och har samma medelvärde för icke-CLIL-gruppen i uppgift 3 och 4. Detta kanske kan tolkas som en progression i CLIL-gruppen då uppgifterna är skrivna i årskurs 2 respektive 3.

Hultman & Westman (1977) för en diskussion om det är bra eller inte bra med långa ord i texter och får i sin undersökning fram att ordlängden är positivt kopplad till betyg. De menar att gymnasisterna i deras undersökning ”i verkligheten behöver lära sig att skriva så att deras ordlängd ökar” utan att de för den skull skall siva byråkratiskt och krångligt (1977:78). Detta diskuteras vidare i avsnitt 8.4.

Gällande medelvärdena för **ordvariation** är samtliga klassers **OVIX-värden** i föreliggande studie betydligt lägre än de som framkommit i tidigare studier, förutom en, vilka redovisas nedan. Resultaten i denna studie visar att skrivuppgifterna har **OVIX-värden** mellan 32 och 72 (texter kortare än 200 ord har inte räknats med, se Hultman 1994:60, Nordenfors 2011:92). I denna studie har 17 texter tagits bort från denna del i analysen av **OVIX**. Medelvärdet för alla klasser är här 53 där medelvärdena i de tidigare studier som refereras till har medelvärden på 60 och över för motsvarande åldersgrupp (se tabell 40). De längsta texterna som har skrivits i uppgift 3 uppvisar också de högsta värdena för ordvariation. Det är pojkar CLIL som genomgående har de högsta medelvärdena för **OVIX** och flickor icke-CLIL de lägsta i tre av de fyra skrivuppgifterna.

I Östlund-Stjärnegårds undersökning varierar **OVIX**-värdet i enskilda texter mellan 43 och 71 (2002:90). Då bör man betänka att de texterna är skrivna som nationella prov med längre tidsförhållanden och en annan form av bedömningsaspekt än skrivuppgifterna som undersöks i denna studie. I *Gymnasistsvenska* (Hultman & Westman 1977) har gymnasisternas uppsatser ett genomsnittligt **OVIX**-värde på 64,8 och som jämförelse med bruksprosa hade tidningstexter 67,6, läroböcker 71, broschyrer 61 och debatttexter 69,3 (ibid., s. 56). De har också funnit en korrelation i gymnasistmaterialet mellan större ordvariation och betyg.

Även Josephson, Melin & Oliv (1990) ger förslag på olika analysredskap för att kartlägga, analysera och utveckla elevers skriftspråk och har undersökt tio texter skrivna av elever från årskurs 1 till 9 där de bland annat mäter ordvaria-

tionen med *OVIX*. De anser att just förmågan att kunna variera sitt ordförråd är så viktigt och att ett stort och varierat ordförråd är ett tecken på mer än förmågan att kunna beskriva saker väl. De menar att det knappast är realistiskt för enskilda lärare att genomföra den sortens beräkningar men att vetskapen om att metoden finns är viktigt. Normala *OVIX*-värden för stora elevgrupper anger Josephson et al. för årskurserna 4, 6 och 9 på grundskolan och för gymnasiet år 3 enligt följande (1990:45):

TABELL 40. *Ovix*-värden i Josephson, Melin & Olivs studie (1990:45)

Årskurs	4	6	9	Gy3
OVIX	50	55	60	67

Medelvärdet för *OVIX* i denna studie som är 53 motsvarar utifrån studien ovan årskurs 5, med en spridning i materialet från det lägsta värdet 32,35 till det högsta som är 72,38. Detta är något som diskuteras vidare i 8.4.

Resultaten för ordvariation i studien av Magnusson & Johansson Kokkinakis (2008) där de undersöker användningsgraden av fyra kvantitativa mått; *lexikal tätthet*, *nominalkvot*, *ordvariation* och *ordlängd*, är medelvärdet för *OVIX* 56,6 vilket alltså är högre än här. Gällande jämförelsen av ordlängden var resultaten från deras studie lägre än här, vilket möjligen kan förklaras av att materialet inte enbart består av nationella prov från studieförberedande klasser, men *OVIX*-resultatet motsäger detta antagande. Det är tio år mellan studierna vilket föranleder en undran om vad annat som kan påverka resultaten.

I Nordenfors (2011) longitudinella undersökning har eleverna *OVIX*-värden mellan 50 och 60, på olika texttyper från de nationella proven i åk 5 till de nationella proven i åk 9 (2011:111). Hans studie ligger således närmare Josephson, Melin & Olivs studie.

Sammanfattningsvis kan sägas att medelvärdena de **kvantitativa ytstrukturella mätningarna** som har undersökts i denna studie; *textlängd*, *andel långord* och *OVIX* visar en nedgång i jämförelse med flertalet tidigare refererade studier medan analystypen *ordmedellängd* ligger jämförelsevis lika med tidigare studier av elevskrivna texter. Hur detta bör tolkas kan givetvis diskuteras och undersökas vidare. Det är inom studien statistiskt säkerställt att CLIL-eleverna skriver längre texter än icke-CLIL-eleverna samt att gruppen pojkar CLIL använder längre ord och uppvisar en större ordvariation. Flickor icke-CLIL är den grupp som sammantaget har de lägsta resultaten.

De kvantitativa ytstrukturella måtten ökar inte med ålder, det vill säga skillnaderna uppvisas inte över de terminer som texterna har skrivits utan följer tendensen att de olika variablerna verkar hänga samman med texttyp eller personligt intresse. Undantaget är andelen nominaliseringar som ökar något i varje skrivuppgift för gruppen pojkar CLIL. Detta trots att skrivuppgifternas svårig-

hetsgrad är avsedd att vara densamma i alla uppgifter för att inom projektet kunna mäta progression.

Eleverna har i studien skrivit de längsta texterna i det ämne som verkar ha engagerat dem mest och där de har haft tillgång till längre, ofta sammansatta ord som hänger samman med skrivuppgifternas ämnen. Individuella variationerna återfinns i såväl CLIL som i icke-CLIL-klasser genom hela materialet.

Användningsgraden för de kvantitativa ämnesrelaterade mätningarna av *nominaliseringar*, *verb i passiv form* och *akademiska ord* är för samtliga grupper i studien låg i samtliga skrivuppgifter. Användningsgraden rapporteras som procentuell andel ord i respektive text och liksom ovan har undersökningar gjorts av medelvärden för grupperna CLIL och icke-CLIL samt för flickor och pojkar i respektive grupp.

För **nominaliseringar** varierar de individuella resultaten mellan 0 %, det vill säga inga nominaliseringar har återfunnits i elevtexten alls, till 3,6 % som finns i tre texter, två av dem är skrivna av samma CLIL-elev i skrivuppgift 3 och 4, den tredje är av en icke-CLIL-elev i skrivuppgift 4. Mönstret att samma elever upprepar höga respektive låga resultat återkommer i studien vid de olika mätningarna.

För nominaliseringar finns dock ett annat mönster som inte synliggörs i analysen av mätningarna av *verb i passiv form* och *akademiska ord*, nämligen att i två skrivuppgifter, skrivuppgift 1 och i skrivuppgift 4, använder alla elever i två av grupperna; flickor icke-CLIL och pojkar CLIL någon nominalisering i sina texter vilket gör att gruppens lägsta värde inte är 0 procent.

Medelvärdena för användningsgraden av nominaliseringar i samtliga skrivuppgifter är för gruppen CLIL: 2,38 % och för gruppen icke-CLIL 2,0 %. Det finns ingen statistisk signifikans för grupperna i någon uppgift. Ser man däremot till undergrupperna PC, FC, PnC och FnC så finns signifikans för pojkar CLIL i skrivuppgift 4 gentemot båda kontrollgrupperna PnC och FnC. I lådagrammen synliggörs spridningen inom varje undergrupp som är så stor att variationen av resultaten påverkas av detta.

I Johansson Kokkinakis & Frändbergs (2013) granskning av ordkategorierna *tekniska termer*, *nominaliseringar*, *passiva verb*, *vardagliga och snedfördelade ord*, *vardagliga och ämnesspecifika ord*, *namn/formler* och *akademiska ord* så undersöktes dessa i läromedelstext samt i elevsvar (årskurs 8) i TIMMS 2007. Gällande nominaliseringar återfanns 1,9 % i läromedelstext och 0,9 % i elevsvaren. Eleverna var indelade i tre undergrupper (låg- medel- och högpresterande) där variationen var från 0,6 till 1,1 % (ibid., 2013:63). Alla värden är alltså lägre än i denna studie. Om det beror på att eleverna var yngre än i föreliggande studie eller att eleverna kommer från mer homogena klasser på grundskolan än här, där samtliga elever går studieförberedande klasser, går inte att säkerställa.

Användningsgraden av **verb i passiv form** är som beskrivits låg. Ingen grupp har ett medelvärde över 1 % vid något skrivtillfälle och flera elever använder inga verb i passiv form alls. Maxvärden här ligger på 2,4 – 3,2 % vilket är lägre

än motsvarande värden av den procentuella andelen nominaliseringar. Ingen signifikans finns mellan grupperna CLIL och icke-CLIL. Medelvärdena för grupperna är CLIL - 0,63 % och för icke-CLIL - 0,64, icke-CLIL gruppen ligger alltså lite högre men det är obetydligt. Elever med högre förekomst av verb i passiv form återfinns bland flickor CLIL, flickor icke-CLIL och pojkar icke-CLIL. Flickorna i de båda undergrupperna har i skrivuppgifterna 1 och 4 en högre andel verb i passiv form än pojkar CLIL vilka som grupp inte uppvisat de lägsta medelvärdena i någon annan variabelmätning. Gruppen flickor icke-CLIL har de högsta medelvärdena i två av skrivuppgifterna, något som heller inte återfinns för gruppen i någon annan variabelmätning.

I undersökningen av elevsvaren i TIMMS (se ovan Johansson Kokkinakis & Frändbergs 2013) är medelvärdena för den totala användningen av verb i passiv form i lärobokstexten 4,1 % av alla ord och i elevsvaren 1,8 % vilket är dubbelt så mycket som här. Fördelningen mellan elevgrupperna som undersöktes var från 0,9 till 2,6 % (ibid., 2013:63). Resultaten för dessa värden är alltså de motsatta som för nominaliseringar. Hur detta bör tolkas kan möjligen ha att göra med att det i TIMMS-provens frågeställningar återfinns verb i passiv form och så följer eleverna frågan vid omskrivning till ett svar. Detta är något som Nygård Larsson (2011) resonerar om i sin studie av elevtexter i biologiämnet; hur når man en undervisning där elever använder ämnesrelevanta ord i självständiga formuleringar? Frågan har relevans för alla skolans ämnen.

Resultaten, inte minst CLIL-pojkarnas, pekar på vad som kan tolkas som en okunskap, en omedvetenhet om hur den passiva formen av verb kännetecknar det vetenskapliga språket och därmed är en resurs i den här typen av skoltext som är argumenterande och utredande (Halliday & Matthiessen 2014, Holmberg & Karlsson 2006, Schleppegrell 2004).

De **akademiska orden** i denna studie har haft sitt utsprung i *En svensk akademisk ordlista*, ESAO (Jansson et al. 2012, Sköldberg & Johansson Kokkinakis 2012). Vissa av orden har avlägsnats då de också visat sig ingå i en grupp av ord som tillhör de 5000 av de vanligast förekommande orden i svenska språket. Av de ursprungliga 652 akademiska orden återstod därefter 363. Andelen som återfinns i elevtexterna uppvisar en motsvarande låg användningsgrad som nominaliseringar och passiver. Flera elever har 0 % förekomst och den högsta användningsgraden är 3,8 % i skrivuppgift 3 och 3,5 % i skrivuppgift 2. Båda resultaten är av samma CLIL-elev som för övrigt är samma elev som hade den högsta användningsgraden av nominaliseringar samt som har höga resultat gällande övriga kategorier också. Statistisk signifikans återfinns inte mellan grupperna CLIL och icke-CLIL men gruppen pojkar CLIL har en stor spridning och flera uteliggare vilket påverkar en signifikans mellan PC och båda flickgrupperna i två av skrivuppgifterna.

I TIMMS-undersökningen (Johansson Kokkinakis & Frändbergs 2013) studeras även ordkategorin akademiska ord. I läromedelstexten som används som

referensmaterial återfinns 9,2 % akademiska ord. I elevsvaren återfinns 1,6 % som ett medelvärde och vid uppdelningen i de tre undergrupperna låg- medel- och högpresterande är resultaten; 0,6 %, 1,0 % och 2,5 % (ibid., 2013:63). Detta kan jämföras med nominaliseringar som liksom de akademiska orden används i lägre grad i TIMMS-studien (2013) men där andelen verb i passiv form förekom i högre grad än i denna studie.

I Lim Falk & Holmbergs studie (2016) undersöks 260 elevtexter skrivna inom CLISS-projektet. Det är samma material som här men motsvarar hälften, då det omfattar skrivuppgifterna 1, *Hälsa och välbefinnande* och 3, *Folkhälsoproblem*, här undersöks även skrivuppgifterna 2 och 4. I Lim Falk & Holmbergs studie (ibid.) görs även en kvalitativ analys av 36 elevtexter utifrån ett SFL-perspektiv. Här refereras till den kvantitativa undersökningen där de 260 texterna har analyserats utifrån de 655 ord som ingår i *En svensk akademisk ordlista, ESAO* (Jansson et al. 2012, Sköldbberg & Johansson Kokkinakis 2012). Resultatet visar då att procentuell andel akademiska ord varierar mellan 1 % och 15 % i elevernas texter. Medelvärdena för CLIL-gruppen är 5,03 % i skrivuppgift 1 och 5,87 % för skrivuppgift 3. Motsvarande medel för icke-CLIL-gruppen är 4,6 % och 5,3 (Lim Falk & Holmberg, 2016:6). Deras studie visar därmed en progression i användningsgraden av akademiska ord hos båda grupperna, CLIL och icke-CLIL, mellan termin 1 och termin 3. I den refererade studien görs en annan indelning av grupperna än här nämligen med avseende på skoltillhörighet (skola A, B och C, se tabell 1, avsnitt 3.1). Hos en av CLIL-grupperna, från skola A, syns inte denna progression då de har det högsta värdet vid första skrivtillfället och sedan ligger kvar på samma medelvärde vid skrivuppgift 3. Av de elever som syns som uteliggare i Figur 17 går endast en av dem på den skolan som rapporteras i Lim Falk & Holmbergs studie (2016), övriga extremvärden som visas här representerar elever i de övriga klasserna.

Användningen av *ESAO* (ibid., 2012) ger alltså olika resultat beroende på om man tar med de allmänspråkliga orden, som finns representerade bland de 5000 av de vanligaste orden i svenska språket, eller inte.

Gällande de **lexikala profilerna** så visas inga signifikanta resultat mellan grupperna CLIL och icke-CLIL gällande den totala täckningsgraden eller användningsgraden av de lågfrekventa orden i *Level 0*, dock är det nära signifikant i skrivuppgift 1. Däremot ges signifikans gällande *Level 1* i skrivuppgift 1. Det vill säga i den första uppsatsen som skrevs under termin 1 använde CLIL-eleverna i mindre utsträckning de mest frekventa orden och i högre utsträckning andra ord. Eftersom det inte finns signifikans i övriga skrivuppgifter mellan grupperna CLIL och icke-CLIL kan en slutsats vara att eleverna i samtliga klasser utvecklar ordförrådet i avseende hur de använder hög- respektive lågfrekventa ord under sin gymnasietid. Detta skulle kunna undersökas med en regressionsanalys för att säkerställa reliabilitet.

Vid de statistiska analyserna på gruppnivå för de fyra undergrupperna, pojkar CLIL, flickor CLIL, pojkar icke-CLIL och flickor icke-CLIL visas statistisk signifikans för gruppen pojkar CLIL gentemot de båda flickgrupperna i samtliga

skrivuppgifter gällande användningen av lågfrekventa ord, de som syns i *Level 0*. Signifikans visas även gentemot gruppen pojkar icke-CLIL i skrivuppgift 4. Även gällande *Token Coverage* visas signifikans för pojkar CLIL gentemot de båda flickgrupperna i skrivuppgift 1-3, däremot inte i skrivuppgift 4.

Det som synliggörs är att de båda pojkgrupperna i högre grad använder lågfrekventa ord och att flickor i högre grad återfinns bland de elever som nästan enbart använder orden som återfinns bland de 5000 av de vanligaste svenska orden. Detta överensstämmer med att gruppen pojkar CLIL har höga medelvärden gällande ordmedellängd, långordsfrekvens och *OVIX*. Denna typ av ord återfinns i högre grad i *Level 0* som lågfrekventa ord. Två CLIL-pojkar använder i samtliga skrivuppgifter många lågfrekventa ord, har en hög andel långord och skriver dessutom inte de längsta texterna (mer i 8.3).

8.3 Resultatdiskussion, samvariation

Studiens frågeställning om det finns samband mellan den vokabulär som framträder som kan relateras till programvalen CLIL respektive icke-CLIL, samt genus i respektive grupp besvaras nedan.

Till de olika mätningarna har tre Excelark skapats, ett för LIX-räknarens beräkningar av de kvantitativa ytstrukturella måtten *antalet ord per text = textlängd, ordmedellängd (antalet tecken och delar med antalet ord), andelen långord/långordsfrekvens (ord med >6 tecken av totala antalet ord)* samt *ordvariationsindex, OVIX*. Ett Excelark för de kvantitativa ämnesrelaterade mätningarna av *nominaliseringar, verb i passiv form och akademiska ord* som identifierades genom korpuslingvistiska metoder där programvaran *AntConc* (Nation & Anthony 2016) sedan användes för att studera förekomsten av de lingvistiska variablerna i varje elevtext. Ett tredje Excelark skapades för de lexikala frekvensprofilerna där programvaran *AntWordProfiler* (Nation & Anthony 2016) har använts för att ta fram procentuell förekomst av ord från de olika frekvensnivåerna som finns i en text. Initialfärgmarkerades höga och låga resultat i Excelarken. När sedan alla resultat hade matats in i SPSS för statistiska analyser så har de olika graferna samt signifikansprövning av resultaten tillfört ytterligare validitet.

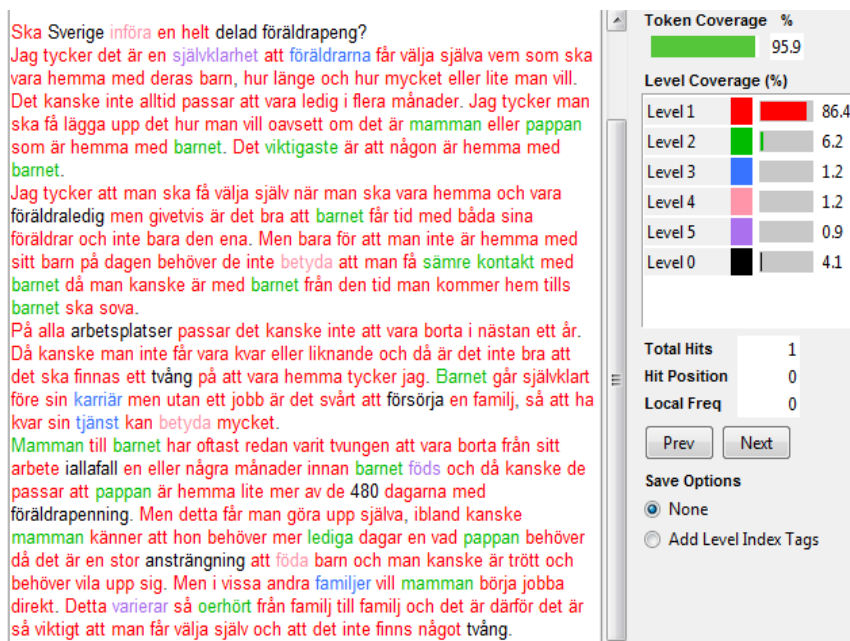
Vid manuell analys synliggörs hur samma elever har skrivit de längsta texterna respektive de kortaste. Eftersom det är en longitudinell studie har det förvånat mig att en progression inte synliggjorts trots att skrivuppgifterna konstruerats för att hålla samma svårighetsgrad studien igenom. Variationen återfinns i mittengruppen men elever vars resultat är högre eller lägre än mittengruppens, förefaller skriva som de "alltid" har gjort. De elever som inte skriver så långa texter ökar inte sin textlängd under sin gymnasietid och de ordrika eleverna som flödesskriver och skriver långa texter redan från start fortsätter så, sin gymnasietid igenom. Dock bör här påpekas att djupare textanalyser skulle kunna visa om den diskursiva text-

strukturen följs (Nyström 2000), så att vissa kortare texter kan betraktas som akademiska och därmed uppvisa progression inom det diskursiva skrivandet.

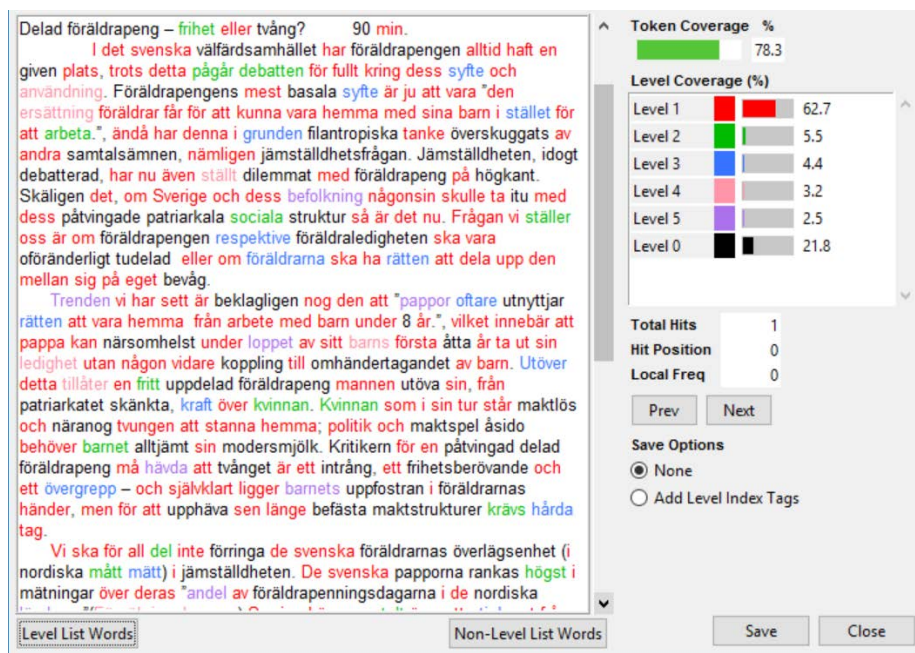
Eleverna som skriver långa texter använder däremot inte alltid de längsta orden. Flickor CLIL skriver de längsta texterna i studien men det är pojkar CLIL som har de högsta resultaten för ordmedellängd, långordsfrekvens och ordvariation. Gällande ordmedellängd och långordsfrekvens så har grupperna pojkar CLIL och pojkar icke-CLIL högre variabelvärden i skrivuppgifterna 1, 2, 3 än båda flickgrupperna. Pojkar CLIL ligger högst även i skrivuppgift 4 då icke-CLIL pojkar som grupp har lägst eller näst lägsta värden.

I de lexikala profilerna synliggörs också att de längsta texterna inte alltid innehåller den högsta andelen långord eller har höga *OVIX*-värden. Detta innebär att några av texterna som har höga värden gällande användningsgrad av lågfrekventa ord också har en hög procentuell användningsgrad för ordmedellängd långord, och *OVIX*. Dock är dessa inte alltid bland de längsta texterna. Här kan diskussionen om förståelsen av det diskursiva skrivandet utvecklas. Det synliggörs alltså att några av eleverna, från både CLIL och icke-CLIL grupperna, skriver mer koncentrerat men ändå med ordvariation och komplexitet gällande vokabulär.

I figur 1, 2, 20 och 21 visas lexikala profiler för fyra elever, två CLIL och två icke-CLIL. Nedan återkommer två, tidigare figur 1 och 2, nu figur 23 och 24 samt samtliga värden som alla mätningar av de olika variablerna visar i respektive elevtext (förkortningar används i tabellerna), därefter diskuteras dessa.



FIGUR 23. Lexikal profil av text av icke-CLIL-elev, skrivuppgift 4 (figur 1)



FIGUR 24. Lexikal profil av text av CLIL-elev, skrivuppgift 4 (figur 2)

TABELL 37. Variabelmätningar för lexikal profil för icke-CLIL-elev i figur 23

Variabel	Textlängd	Ordm.l.	Långord	OVIX	Nom.	Passiv	Akadem	T.Cov.	Level 0
Figur 1	329	4,27	8,51	41,82	0,3	0,3	0	95,9	4,1

TABELL 38. Variabelmätningar för lexikal profil för CLIL-elev i figur 24

Variabel	Textlängd	Ordm.l.	Långord	OVIX	Nom.	Passiv	Akadem	T.Cov.	Level 0
Figur 2	521	5,54	27,83	74,85	2,7	1,9	0,6	77,7	22,3

De två elevtexter som visas som figur 1 – 23 och figur 2 – 24, tillhör de som har högst respektive lägst variabelvärden, därav de stora skillnaderna men samvariationen mellan variablerna är synligt, mer eller mindre, i alla elevtexter som har undersökts.

I tabell 37 och 38 visas mätvärdena för de olika variablerna. Gällande textlängd och ordmedellängd är skillnaden inte så stor men när man sedan jämför andelen långord i procent och ordvariationen, *OVIX*-värde, så är det markanta skillnader. Dessa samvarierar tydligt med andelen låg- och högfrekventa ord, vilket syns både i procentuell andel av *Token Coverage* respektive *Level 0* men också i färgerna som följer med respektive frekvensnivå. Samvariation finns

även med den lexikala profileringen och användningsgraden av nominaliseringar, verb i passiv form och akademiska ord.

Dock visar resultaten i denna studie att medelvärdena för *OVIX* är lägre än i tidigare refererade studier, med en spridning i materialet från det lägsta värdet 32,35 till det högsta som är 72,38. Resultatet implicerar att alla elever inte har tillgång till en vokabulär som gör att de kan skriva precist genom att skriva varierat (Josephson *et al* 1990). Läser man textutdragen för figur 1 – 23 och figur 2 – 24 märks detta av innehållet som är mer upprepande i figur 1 – 23 och utförligare och mer varierat i figur 2 – 24.

Flickor icke-CLIL har de lägsta medelvärdena gällande ordmedellängd, långordsfrekvens och ordvariation i samtliga skrivuppgifter förutom i skrivuppgift 4 då gruppen pojkar icke-CLIL har lägre medelvärden. Med lådagram synliggörs också den spridning som finns inom de olika grupperna. Resultaten visar att pojkgруппerna i flera mätningar har de högsta medelvärdena och flickor CLIL och pojkar icke-CLIL ofta har resultat som är nära varandra samt att flickor icke-CLIL har lägre resultat som grupp i alla mätningar, dock finns det extremvärden/uteliggare. Dessa extremvärden är flest i gruppen flickor CLIL med sammanlagt 38 uteliggare i de olika mätningarna vilket innebär en stor spridning inom gruppen i de flesta variabelmätningar. Av uteliggarna representerar 34 höga värden och fyra låga. Motsvarande antal för de övriga grupperna är 19 uteliggare i pojkar CLIL, varav 15 höga värden och fyra låga, 20 uteliggare i gruppen pojkar icke-CLIL, varav 19 höga och ett lågt värde och 23 extremvärden/uteliggare i gruppen flickor icke-CLIL, varav 16 höga och åtta låga. De båda pojkgруппerna är i dessa mätningar mer homogena med färre uteliggare än de båda flickgruppen och flest uteliggare med låga värden finns bland flickor icke-CLIL.

Här är jämförelsen med rapporter och mätningar från andra studier av intresse där flickor över lag presterar bättre än pojkar, vilket inte stämmer med alla variabelmätningar i den här studien, särskilt avseende gruppen flickor icke-CLIL (se exempelvis Skolverket 2012, 2017c).

Den variabel där pojkar CLIL ligger lägst, trots att de oftare ligger högst, är vid användningen av verb i passiv form. En tolkning kan vara att det är ett sätt att skriva som inte har uppmärksammats av gruppen vilket också kan förbrylla då de läromedelstexter eleverna möter under sin skoltid liksom provuppgifter, särskilt i de naturvetenskapliga ämnena, ofta använder passiv-konstruktion (se Frändberg 2012, Johansson Kokkinakis & Frändberg 2013, Persson 2016, Ribeck 2015).

8.4 Pedagogiska implikationer

Om språket är så viktigt att det betraktas som en framgångsfaktor i skolan, och ordförrådet den enskilt viktigaste faktorn som rustar individen för språkligt krävande uppgifter (Enström 2013, Lindberg 2006, Meara 1996, Skolverket 2015, Uccelli et al. 2015 m.fl.) så kan den här studien bidra till att synliggöra komplexiteten i ett ordförråd. Det är många aspekter som ingår i en vokabulär, även om undersökningarna här fokuserar på den produktiva och inte den receptiva delen. Studien visar även vad som händer i elevtexter insamlade under tre år hos gymnasieelever i åtta högskoleförberedande klasser varav fem klasser är språk- och ämnesintegrerade och tre har läst sina gymnasieprogram helt på svenska (förutom eventuella språktillval).

De pedagogiska implikationer som studien ger är flera; komplexiteten av hur ett ordförråd kan se ut behöver synliggöras i all undervisning liksom hur texttyper är sammansatta av olika vokabulär där ordens betydelse och funktion kan variera (Gibbons 2009, Halliday 1993, Halliday & Matthiessen 2014, Martin & Rose 2008, Schleppegrell 2004). Även komplexiteten av att *kunna* ord behövs lyftas fram och tydliggöras (Gustafsson & Håkansson 2017, Nation 2013, Oakhill et al. 2015). Att förenkla skolspråket i tron att man då överbrygger språkliga svårigheter genom att sänka abstraktionsgraden lämnar istället elever sämre rustade för att möta högre utbildningskrav, vilket inte på något sätt ökar likvärdigheten i svensk skola, snarare tvärtom (Edling 2006, Macken-Horarick 1996).

Kunskaperna finns genom forskning som presenterats inom forskarvärlden men som även görs tillgänglig för alla genom exempelvis Skolverkets olika publikationer (Skolverket 2008, 2012, 2014, 2016a, 2016b, 2016c, 2016d, 2017a, 2017b, 2017c). Det är en fråga om tid som måste ges så att lärare får tillgång till den forskning som finns samt ges möjligheter att omsätta den tillgängliga forskningen empiriskt så att skola och undervisning vilar på ”en vetenskaplig grund, beprövad erfarenhet och evidens” (Skollagen 1 kap 5 §).

I *Lgy11* står för ämnet svenska i det *Centrala innehållet* i kursplanen (som anger vilket obligatoriskt innehåll som ska behandlas i undervisningen) att eleverna skall möta ”vetenskapliga texter” och ”texternas syften, innehåll, uppbyggnad och språkliga drag.” (Skolverket 2011a: *Lgy11* s. 176-178). På samma ställe i läroplanen finns även *kunskapskraven* som innefattar hur språket skall bedömas; ”Språket är träffsäkert, klart och varierat och innehåller goda formuleringar. Viktiga generella drag som rör disposition, språk och stil i texter av vetenskaplig karaktär. Språket är varierat och innehåller goda formuleringar över lag välformulerat” (ibid., s.176-178).

Föreliggande studie visar att de deltagande eleverna i mycket liten utsträckning använder sig av de variabler som kännetecknar ett mer vetenskapligt sätt att skriva. Som undervisande lärare tänker jag att elever tidigt, redan på grundskolan, behöver få bekanta sig med detta. I det *Centrala innehållet* för svenskämnet

för årskurs 7-9 framgår det att ”vetenskapliga texter” skall ingå som en del av ”Berättande texter och sakprosatexter” (Skolverket 2011b). Behovet av detta är något som har visats och förts fram även i tidigare studier (se Andersson Varga 2014, Ask 2005, Norberg Brorsson 2007, Nyström Höög 2010).

Man behöver tidigt läsa läromedelstexter tillsammans i klassrummet med ordens betydelse såväl som funktion i fokus, så att passiva verb uppmärksammas, liksom nominaliseringar och akademiska ord (Edling 2006, Frändberg 2012, Hipkiss 2014, Johansson Kokkinakis & Frändberg 2013, Persson 2016, Ribeck 2015). Det är inte ovanligt att lärare läser och skriver och elever lyssnar och tittar på. Detta räcker dock inte. För att ord skall ingå i den produktiva vokabulären behöver de skrivas, sägas, användas och tränas, *tillsammans*.

I skrivande stund redogörs i media för Svenska Akademiens satsning på Intensivutbildning i svenska för nyanlända elever där ett av momenten för skolframgång är ”en riktad satsning på utveckling av ordförrådet”³⁵. Inom utbildningssatsningen används beprövade metoder för ordinläring hämtade från Försvarets Tolkskola och kända minnestekniker. Dessa metoder är enkla och kostar inte mer än några ark papper och tid men visar att glosor och nyordsinläring kan och bör tränas oavsett språk.

I min studie synliggörs också att för att kunna skriva ”träffsäkert, klart och varierat med goda formuleringar” (Skolverket 2011b, s.176-178) så behöver elever, och lärare, få en förståelse för hög- och lågfrekventa ord, liksom för måttet *OVIX*, som visar hur variationsrikt språket kan vara. Här synliggörs ingen tydlig progression i elevtexterna över tid gällande de textlingvistiska mätningar av olika variabler som har gjorts. Detta trots att texternas svårighetsgrad har varit densamma över tid i syfte att just kunna mäta detta. Elever som skriver långa texter, som skriver variationsrikt och i högre grad använder längre och mer lågfrekventa ord, gör så i de insamlade skrivuppgifterna under sina tre gymnasieår. De som skriver kortare texter med fler högfrekventa, vardagliga ord, med lägre ordvariationsindex fortsätter att göra detta sin gymnasietid igenom.

De kvantitativa mätningarna visar att elever som har låga värden studien igenom återfinns i såväl CLIL som i icke-CLIL-gruppen. Här är det alltså inte programvalet som tycks ha påverkat. Även det är ett resultat som överensstämmer med Apelgrens (u.u. 2017) rapport till Skolverket *Höga betyg in, höga betyg ut* (ibid.). I studien refereras även att ordmedellängd och andel långord samvarierar med läromedelstext vilket torde påverka elevers skrivande då denna typ av texter innehåller sådana kännetecken (se Ribeck 2015).

Båda resultaten ovan bekräftar det som Andersson Varga (2014) fann i sin undersökning av skrivundervisningen i gymnasieskolan. I studien redogörs för lärares förväntningar på elevers prestationer som kan kopplas till programval

³⁵ <http://intensivsvenska.se/category/ordfarrad/> SVT 2017-10-22

liksom att skrivandet förväntas utvecklas av sig självt (ibid., 239-242). Lärares förväntningar på elever spelar roll vilket även en explicit skrivundervisning gör.

Vid jämförelser med tidigare studier (Bratt 2014, Ciolek Laerum 2009, Hultman & Westman 1977, Josephson et al. 1990, Lim Falk 2008, Nordenfors 2011, Olevard 1999, Vagle 2005, Östlund-Stjärnegårdh 2002) visar variabelmätningarna här att medelvärdena för textlängd samt användningsgraden av långord och *OVIX* har fortsatt att minska över tid.

Myhill (2009) ser skribenter som *designers*; det är en process där man blir medveten om vad som förväntas, både läraren i sin skrivundervisning och eleven i sitt skrivande (ibid., s. 412-413).

”The pedagogical implications of this are that teaching writing with linguistic development in mind should not be about grammatical rules, error correction or formulaic approaches to effectiveness: rather, enabling teaching with focus on option and possibilities available to them and encouraging discussion about the links between linguistic choices, rhetorical impact and the communication of meaning.” (Myhill 2009:413)

CLISS-projektets huvudfokus har legat på att undersöka hur gymnasieelevers produktiva och receptiva färdigheter i skolrelaterat, akademiskt skriftspråk på svenska och engelska utvecklas i en flerspråkig skolkontext. I föreliggande studie bekräftas det som har framkommit i andra studier i projektet att CLIL-elevernas resultat i flera avseenden är högre än icke-CLIL-elevernas (se Apelgren, 2014, Apelgren u.u. 2018, Lim Falk, 2015, Lim Falk & Holmberg, 2016, Olsson, 2015, Olsson, 2016, Reijerstam, 2015, Sylvén, 2013, Sylvén, 2015, Sylvén & Ohlander, 2014a, Sylvén & Ohlander, 2014b, Sylvén & Olsson, 2015, Sylvén & Thompson, 2015, Yoxsimer Paulsrud, 2014). I föreliggande kvantitativa studie är det dock inte alla resultat som visar statistisk signifikans, vilket innebär att flera elever i icke-CLIL-gruppen också uppvisar höga resultat.

När det gäller kvantitativa undersökningar har en diskrepans gällande studieresultat synliggjorts mellan flickors och pojkars resultat till exempel i de internationella mätningar där Sverige deltar. Detsamma gäller vid olika jämförelser av skolprestationer mellan könen, där flickor presterar bättre än pojkar.³⁶

I föreliggande studie visas att pojkarna vid mätningar av variabelerna har högre värden än flickorna. Gällande betyg visar undersökningen av de deltagande klasserna att flickorna som grupp (CLIL och icke-CLIL) har högre betygspoäng än pojkarna i samtliga kurser i svenska, Svenska 1, 2 och 3, (se Apelgren u.u. 2018). Däremot har gruppen CLIL högre betyg i samtliga kurser i svenskämnet än gruppen icke-CLIL (ibid.). CLIL-gruppen visar också som grupp en högre motivationsgrad vid sin gymnasiestart (Sylvén & Thompson, 2015).

³⁶ Skolverket, pressmeddelande 2017-11-21 angående PISA-resultat.

Avslutningsvis bör sägas att samtliga skolor har visat en alltigenom positiv inställning under projektiden. Alla klasser som deltog i CLISS-projektet var högskoleförberedande och lärarna som jag har mött har varit ambitiösa och mycket intresserade av projektets inriktning och fokus. Detta till trots implicerar resultaten i föreliggande studie att ytterligare uppmärksamhet kan och bör ägnas åt skolspråkets komplexitet.

8.5 Förslag till vidare forskning utifrån material och metod

Det insamlade materialet inom CLISS-projektet är, som har beskrivits i avsnitt 3.2, omfattande (se även bilaga 1). Utifrån denna studie är det möjligt att gå vidare med korrelationsanalyser där exempelvis jämförelser görs av elevernas produktiva ordförråd och deras receptiva ordförråd i svenska, vilket finns i de ordförståelsetest som har genomförts inom projektet. Genom att granska resultaten i ordtesten och jämföra med ordlistorna som har skapats i denna studie kan ordkategorierna undersökas närmare.

Även en analys av det akademiska språket kan göras genom ordtesten via skrivuppgifternas resultat gentemot de 652 orden i *ESAO* (Jansson et al. 2012, Sköldberg & Johansson Kokkinakis 2012). Detta då jag beslutade om en annan typ av jämförelse av dessa utifrån *ESAO* (ibid., 2012) jämfört med Lim Falk & Holmberg (2016). Här skulle en fördjupad analys av skillnaderna i undersökningsmetod kunna vara av intresse för att se i vilka frekvensband de orden återfinns som jag fann bland de högfrekventa orden vid undersökningen av lexikala profiler. Vilka ord är det och hur har de använts av eleverna?

Något annat att gå vidare med är andelen *nominaliseringar* och *verb i passiv form*. Samma frågor som ställdes ovan kan även ställas här; vilka är de, hur används de, och återfinns de i skrivuppgifternas inspirationstexter? En aspekt är hur det kommer sig att just nominaliseringar finns i alla elevtexter hos gruppen flickor i icke-CLIL i skrivuppgift 3. Likaså för pojkar CLIL i skrivuppgift 4 där deras lägsta värde inte är 0 % utan 0,4 %. Vad är det för nominaliseringar i just de texterna? Användningsgrad och spridningen för CLIL-pojkar ligger vid mätningen av denna aspekt lägre än gruppen gör vid andra gruppjämförelser. Även medianvärdet för gruppen pojkar CLIL är lägst av de fyra grupperna i två av skrivuppgifterna, något som inte återfinns i någon av de andra mätningarna. Vad går att se vid en djupare textanalys beträffande detta?

Motsvarande kvantitativa analys kan göras av de ämnesrelaterade orden som synliggjorts genom de lexikala profilerna; vilka är de, hur används de, återfinns de i skrivuppgifternas inspirationstexter? Kan man se om individer återanvänder vissa ämnesrelaterade och/eller andra lågfrekventa ord i flera texter? Kan det i så fall kopplas till programval, kön och/eller betyg?

Undersökningen av de lexikala profilerna kan utvecklas och studeras i relation till annat material som återfinns i insamlingsdatabasen som visar elevernas receptiva ordförråd för att se om förekomster av ord återfinns såväl receptivt som produktivt. De lexikala profilerna synliggör vokabulären ur en aspekt som många sannolikt inte tänker på i det vardagliga skrivandet. Med teknikens hjälp kan olika typer av textanalyser göras och på så vis bidra till en ökad kännedom om ords förekomst och frekvens för att därmed kunna utveckla skrivkompetens (jfr Andersson Varga 2014, Ask 2005, Randahl 2015).

I en vidare studie skulle en djupare analys av de kvantitativa ytstrukturella måttenheterna *textlängd*, *långord*, *ordmedellängd* och *OVTX* kunna innebära att man i varje elevtext granskar hur många gånger ord har upprepats för att på så vis se om detta har påverkat resultaten där ordvariationen kanske mer hänger samman med uppräknings än variation (jfr Nordenfors 2011).

Man kan även i en vidare studie genomföra fler statistiska korrelationsanalyser av materialet. Här finns flera undersökningsmöjligheter; korrelerar variabler med varandra och ger de i så fall statistisk signifikans? Betygen är en sådan variabel som inte har undersökts i denna studie men som skulle kunna korrelera med samtliga variabler för att se om mönster tydliggörs eller inte.

En annan väg är att analysera texterna från de 63 eleverna som har skrivit samtliga skrivuppgifter på svenska under de tre projektåren (se tabell 9). Detta genom att göra jämförelser inom gruppen av de resultat som har framkommit, dels genom att undersöka om gruppen som sådan eventuellt uppvisar mönster genom att ha skrivit samtliga skrivuppgifter. Man kan även göra andra typer av statistiska analyser av exempelvis regressionsanalyser alternativt att göra mer kvalitativa textanalyser.

Mitt försök att synliggöra elevers ordförråd genom de olika mätningarna som har genomförts har visat att tendenser av olika aspekter är möjliga att upptäcka.

I en pilotstudie av nationella provtexter från elever i årskurs 6 användes de ytstrukturella mätmetoderna och visade dels skillnader mellan skolor med olika socioekonomiska förhållanden, dels ordvariationen eller snarare bristen på ordvariation. Metoden visade sig vara användbar och kan förstås i en enkel omfattning genomföras av enskilda lärare men med kunskaper om att det är samband som synliggörs, *inte* bedömningsunderlag.

Genom att göra en interventionsstudie på en av skolorna som deltagit i projektet där fyra klasser; två CLIL och två icke-CLIL i så fall skulle ingå, kan det möjligtvis gå att se effekter av en explicit språkstödande undervisning. En sådan intervention skulle då innehålla undervisning i två grupper; en CLIL, en icke-CLIL om språkets funktion i olika typer av texter, här det akademiska och vetenskapliga sättet att skriva där nominaliseringar, verb i passiv form samt specifika ord har en högre användningsgrad i den typen av text. I de två kontrollgrupperna skulle motsvarande undervisning inte ske. Därefter skulle de fyra klasserna få skriva två av skrivuppgifterna som ingår i denna studie. Resultaten

undersöks och analyseras sedan med samma metoder som har använts i föreliggande studie för att se hur utfallet blir gällande testgrupp och kontrollgrupp, CLIL – icke-CLIL, genus och betyg.

En aspekt som inte är med i denna studie är kopplingen mellan pojkars resultat och deras kunskaper i engelska. Korrelerar en hög användningsgrad av nominaliseringar, verb i passiv form samt akademiska ord och en hög andel av lågfrekventa ord med elevers kunskaper i engelska där fritidsengelskan i tidigare studier har visat sig ha en påverkan deras skrivande (se Olsson 2011, 2016)?

Vid en interventionsstudie skulle även mer kvalitativa undersökningsmetoder kunna tillföras som att intervjua deltagarna om hur de ser sin vokabulär och hur de arbetar med sitt receptiva och produktiva ordförråd. En frågeställning är vilka strategier de har och hur de kan se sin egen utveckling över tid i förhållande till lärande och undervisning (Säljö 2015).

8.6 Slutsatser och avslutande ord

Går det att ”se” ett ordförråd? Jag skulle vilja svara ”Ja!” på den frågan. De olika sätt som beskrivs gällande undersökning och analys av de 520 elevtexterna visar ett ordförråds komplexitet samt hur mycket det faktiskt går att lära sig och lära ut. Detta är något som ger mig nya infallsvinklar, både som lärare och lärarutbildare.

De, enligt mig, viktigaste slutsatserna är dels att elever oavsett programval behöver få explicit undervisning om ”vetenskapliga texters syften, innehåll, uppbyggnad och språkliga drag” (Skolverket 2011a: *Lgy11* s. 176-178). Det räcker inte med skrivningen att eleverna skall ”få möta” (ibid.) denna typ av texter. Det är en alldeles för vag beskrivning i ett kunskapskrav som skall verka som stöd för såväl elever som för lärare. Lika tydligt framstår det att elever måste få träna på att skriva texter ”av vetenskaplig karaktär” så att de ges möjlighet att nå kunskapskraven om ett kunna skriva ”träffsäkert, klart och varierat med goda formuleringar och ha med generella drag som rör disposition, språk och stil” (ibid., s.176-178). När man skall träna texttyper behövs explicit stöttning; *scaffolding* (Gibbons 2009, Schleppegrell 2004, Wood, Bruner & Ross, 1976), något som i svensk skola inte alltid betyder att man *skriver tillsammans*, vilket är gynnsamt för alla elever (Hajer & Meestringa 2014).

Den allmänt låga användningsgraden av ordkategorierna nominaliseringar, verb i passiv form och akademiska ord, vilka kännetecknar ett vetenskapligt sätt att skriva (Halliday 1993, Schleppegrell 2004 m.fl.), visar att det är något som eleverna i denna studie under sin gymnasietid inte nyttjar i någon större omfattning. Detta trots att eleverna, med stor sannolikhet, möter den typen av ordkategorier i läromedel och i provkonstruktioner (Edling 2006, Frändberg 2012, Johansson Kokkinakis & Frändbergs 2013, Persson 2016, Ribeck 2015). Även

här krävs tydlig undervisning där den receptiva förståelsen och det produktiva skrivandet tränas.

Ytterligare en slutsats i föreliggande studie är hur ords frekvens, som synliggjorts i de lexikala profilerna där användningsgraden av hög- respektive lågfrekventa, vardagliga ord visas, åskådliggör att ordvariation och förståelse för den innehållsliga sidan av ord spelar roll. För att kunna skriva som förväntas i kunskapskraven ovan behövs tillgång till en rik vokabulär. Förståelsen för hur språket kan varieras och hur oändligt många möjligheter som finns måste explicitgöras i undervisningen av alla lärare, oavsett vilket ämne man undervisar i.

Behovet av mera systematisk skrivträning är inget nytt utan har varit en slutsats som framkommit i flera av studierna som refererats till i denna uppsats. Nyström Höög (2010) uttrycker att: ”det krävs mera systematisk skrivträning som explicit formulerar vilka språkliga egenskaper som krävs av olika slags texter och som ger eleverna handfasta instruktioner om vilka moment som måste genomföras för att olika slags texter skall kunna fylla sina syften” (2010:9). En fråga blir då varför detta inte syns i elevtexterna flera år och en läroplan senare.

Efter att ha synliggjort vokabulären i studien med kvantitativa ytstrukturella mätningar, kvantitativa ämnesrelaterade mätningar och lexikal profilering framträder en bild av hur oerhört olika resurser elever har när det kommer till deras produktiva skrivande. Det behöver inte vara så. Om elever ges tidig stöttning i att utveckla vokabulär genom att få tillgång till många ord i en kontextrik undervisning, där även det svenska ordförrådet uppmärksammas och tränas kontinuerligt, liksom om elever erbjuds en klart uttalad skrivutvecklande undervisning, så kommer detta att samverka med såväl läsande som skrivande och kunna ge synbara resultat.

Likaså ger det som synliggjorts i denna uppsats implikationer för det likvärdiga lärandet och den likvärdiga skolan. Det handlar om en förståelse för ”det rika språket och det goda ordförrådet”, hur det kan se ut, hur det kan utvecklas och vilka bakomliggande kunskaper som kan förmedlas, såväl till elever som till lärare och lärarutbildare, för att på så vis också nå en ökad likvärdighet i skolan där språket är allas.

9. Referenser

- Andersson Varga, Pernilla (2014). *Skrivundervisning i gymnasieskolan. Svenskämnets roll i den sociala reproduktionen*. Diss. Gothenburg Studies in Educational Sciences 359. Göteborgs universitet.
- Apelgren, Britt-Marie (2014). *CLISS Students' Background Factors*. Brisbane: AILA.
- Apelgren, Britt-Marie. *The CLISS Student – some Background Factors of the Participating Students in the CLISS Project*. I Sylvén, L.K. (red.), (bokmanus).
- Apelgren, Britt-Marie (u.u. 2018). *Höga betyg in, höga betyg ut En studie om betygsutveckling hos elever med engelska som undervisningsspråk*. Rapport, Skolverket
- Ask, Sofia (2005). *Tillgång till framgång. Lärare och studenter om studieövergången till högre utbildning*. Licentiatuppsats. Växjö: Institutionen för humaniora.
- Axelsson, Monica, Olofsson, Mikael, Philipsson, Anders, Rosander, Carin & Sellgren Mariana (2006). *Ämne och språk – språkliga dimensioner i ämnesundervisningen*. Rapport. http://www.andrasprak.su.se/polopoly_fs/1.84033.1333713181!/menu/standard/file/Amneochsprak.pdf
- Bailey, Alison L. (2007). *The Language Demands of School. Putting Academic English to the Test*. New Haven & London: Yale University Press.
- Berge, Kjell-Lars (1988). *Skolestilen som genre. Med påtvungen penn*. Lillehammer: LNU og J.W. Cappelens Forlag.
- Berge, Kjell-Lars (2002). Hidden norms in Assessment of Students' Exam Essays in Norwegian Upper Secondary Schools. *Written Communication*. October 2002 vol. 19 no. 4, (s. 458-492).
- Berge, Kjell-Lars & Skar, Gustaf (2015). *Har Normprosjektet bidratt til å styrke elevenes skriveferdigheter?* (Normprosjektet: Resultater pr. 9.03.15 – Notat til Ludvigsen utvalget.) Oslo: ILN, UiO.

- Biber, Douglas, & Conrad, Susan & Reppen, Randi (1998). *Corpus Linguistics. Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Björnsson, Carl Hugo (1968). *Läsbarhet*. Stockholm: Liber.
- Bloggmix* (2012). <https://spraakbanken.gu.se/swe/resurs/bloggmixodat>
- Blåsjö, Mona (2006/2010). *Skrivteori och skrivforskning. En forskningsöversikt*. MINS, Meddelanden från Institutionen för nordiska språk vid Stockholms universitet, nr. 56.
- Borin, Lars, Forsberg Markus, Roxendal Johan (2012). *Korp – the Corpus Infrastructure of Språkbanken*. I: Proceedings of LREC 2012. Istanbul: ELRA, s. 474–478.
- Bratt, Elin (2014). *Toppbetyg i svenska – En studie av kopplingen mellan textkvalitet och högsta betyg i svenska över tid*. C-uppsats, Institutionen för nordiska språk, Uppsala universitet: Uppsala.
- Christie, Frances (1985). *Language Education*. Deakin University: Deakin University Press.
- Christie, Frances & Beverly, Derewianka (eds). (2008). *School Discourse*. London: Continuum.
- Ciolek Laerum, Beatrice (2009). *Elever skriver och lärare bedömer – en studie av elevtexter i åk 9*. (Svenska i utveckling nr 25. FUMS Rapport nr 226.) Uppsala: Uppsala Universitet.
- Cohen, Louis, Manion, Lawrence, Morrison, Keith. (2011). *Research Methods in Education*. 7th ed. London & New York: Routledge Applied Linguistics.
- Coxhead, April (2002). *The Academic Wordlist: A Corpus-based Word List for Academic Purposes. Language and Computers, Teaching and Learning by Doing Corpus Analysis*. Proceedings of the Fourth International Conference on Teaching and Language Corpora, Graz. Ketteman B. & Marko, G. (Eds.) (s. 73-89).
- Crooks, Terry J., Kane, Michael T. & Cohen, Allan S. (1996). Threats to the Valid Use of Assessments. *Assessment in Education: Principles, Policy & Practice*, 3 (3), (s. 265-285).
- Cummins, John (1979) Cognitive academic language proficiency, linguistic independence, the optimum age question and some other matters. *Working Papers on Bilingualism, No.19* (s. 121-129).
- Cummins, John (1980). The Cross-Lingual Dimensions of Language Proficiency: Implications for Bilingual Education and the Optimal Age Issue *TESOL Quarterly, Vol. 14*, No. 2 June 1980. (s. 175-187)
- Dalton-Puffer, Christiane (2011). Content-and-Language Integrated Learning: From Practice to Principles? *Annual Review of Applied Linguistics*. (2011) 31, (s. 182–204).
- Dysthe, Olga (1996). *Det flerstämmiga klassrummet*. Lund: Studentlitteratur.
- Edling, Agnes (2006). *Abstraction and authority in textbooks. The textual paths towards specialized language*. ACTA UNIVERSITATIS UPSALIENSIS Uppsala universitet.
- Enström, Ingegerd (2013a). Ordförråd och ordinlärning – med särskilt fokus på avancerade inlärare. Kenneth Hyltenstam & Inger Lindberg (red.) *Svenska som andraspråk-i forskning, undervisning och sambälle*. Lund: Studentlitteratur (s.169-195).
- Enström, Ingegerd (2013b). *Att undervisa om ord. En studie av lärares behov, förutsättningar och kunskapsutveckling*. Forskningsrapporter från Institutionen för svenska språket, Göteborgs universitet. ISSN 1401-5919.

- Erickson, Gudrun (2010). Good Practice in Language Testing and Assessment – A Matter of Responsibility and Respect. In Kao, T. and Lin Y. (Eds.), *A New Look at Teaching and Testing: English as Subject and Vehicle* (s. 237-258). Taipei, Taiwan: Bookman Books Ltd.
- Erickson, Gudrun (2016). Likt, olik, gemensamt – om god bedömning i språk. *Lingua*, nr 4, (s. 7-12).
- Europarådet (2002). *European Language Policy and CLIL*. Selection of EU-funded Projects. http://www.edu.xunta.es/centros/cpicruce/system/files/clilbroch_en.pdf.
- Evensen, Lars Sigfred (2003). *Sammendragsrapport, juni 2003. Kvalitetssikring av læringsutbyttet i norsk skriving. KAL-prosjektet*. Institutt for språk- og kommunikasjonsstudier, NTNU. <http://prosjekt.hihm.no/r97-kal>
- Frändberg, Birgitta (2012). *Elevers språk i naturvetenskapliga förklaringar: Analys av skriftliga svar från åk. 8 i TIMMS 2007*. Licentiatuppsats. Göteborg: Chalmers University of Technology.
- Gardner, Dee (2007). Validating the Construct of *Word* in Applied Corpus-based Vocabulary Research: A Critical Survey. *Applied Linguistics* 28/2: pp. 241-265. Oxford University Press: Oxford.
- Gardner, Dee & Davies, Mark (2013). A New Academic Wordlist. *Applied Linguistics* 2014: 35/3: (s. 305-327).
- af Geijerstam, Åsa. (2006). *Att skriva i naturorienterade ämnen i skolan*. ACTA UNIVERSITATIS UPSALIENSIS. Uppsala Universitet.
- Gibbons, Pauline (2009). *Lyft språket. Lyft tänkandet. Språk och Lärande*. Hallgren & Fallgren: Stockholm.
- Golden, Anne (2006). Å gripe poenget. Förståelse av metaforiska uttryck fra lærebøker i samfunnskunnskap hos minoritets elever i ungdomsskolen. *Norsk Lingvistisk Tidsskrift*, 24(2), (s. 281-286).
- Golden, Anne & Hvenekilde, Anne (1983). *Rapport fra prosjektet Læbokspråk*. Sentret for språkpedagogikk, Universitetet i Oslo.
- Granger, Sylviane & Paquot, Magali (2009). *In search of a General Academic vocabulary: A Corpus-Driven Study Université Catholique de Louvain Centre for English Corpus Linguistic*, http://www.uclouvain.be/cps/ucl/doc/adri/documents/In_search_of_a_general_academic_english.pdf
- Hajer, Maaike & Meestringa, Theun (2010/2014). *Språkinriktad undervisning*. Stockholm: Hallgren & Fallgren Studieförlag AB.
- Halliday, Michael A. K. (1993). Towards a Language-Based Theory of Learning. *Linguistics and Education* 5, (s. 93-116).
- Halliday, Michael A. K. & Martin, James Robert (1993). *Writing Science. Literacy and Discursive Power*. London/Washington D.C.: The Falmer Press.
- Halliday, Michael A.K. & Matthiessen, Christian M.I.M. (2014). *An Introduction to Functional Grammar*. Fourth Edition. London: Arnold.
- Heimann Mühlenbock, Katarina (2013). *I See What You Mean*. Diss. Göteborg: University of Gothenburg.

- Hipkiss, Anna Maria (2014). *Klassrummets semiotiska resurser. En språkdidaktisk studie av skollämnena Hem- och konsumentkunskap, Biologi och Kemi*. Umeå: Umeå universitet.
- Holmberg, Per & Karlsson, Anna-Malin (2006). *Grammatik med betydelse. En introduktion till funktionell grammatik*. Stockholm: Hallgren & Fallgren Studieförlag AB.
- Holmberg, Per, Karlsson, Anna-Malin & Nord, Andreas (red.), (2011). *Funktionell textanalys*. Stockholm: Norstedts.
- Hultman, Tor G. (1994). Hur gick det med OVIX? I: *Språkebruk, grammatik och språkförändring. En festskrift till Ulf Teleman 13.1 1994*. Lund (s. 55–64).
- Hultman, Tor G. (2003/2013). *Svenska Akademiens Språklära*. Stockholm: Norstedts.
- Hultman, Tor G. & Westman, Margareta (1977). *Gymnasistsvenska*. Svenskläraryöreningen 167. Lund: Liber Läromedel.
- Hyland, Ken & Tse, Polly (2007). Is There an "Academic Vocabulary"? *TESOL Quarterly*, Volume 41, Issue 2 (s. 235-253).
- Hägerfelth, Gun (2004). *Språkpraktiker I naturkunskap i två mångkulturella gymnasieklassrum*. Malmö: Malmö Studies in Educational Science, Diss. Malmö Högskola.
- Jansson, Håkan, Johansson Kokkinakis, Sofie, Ribbeck, Judy & Emma Sköldberg (2012). *A Swedish Academic Word List: Methods and Data*. I: Proceedings of the XV Euralex International Congress, Oslo, augusti 2012, s. 955-960.
- Johansson Kokkinakis Sofie (2007). Språkteknologiskt arbete. I: Lindberg, Inger, & Johansson Kokkinakis, Sofie (red.). *OrdL. En korpusbaserad kartläggning för grundskolans senare år*. ROSA-rapport 8. Institutet för svenska som andraspråk, Göteborgs universitet (s. 101-134).
- Johansson Kokkinakis Sofie & Frändberg, Birgitta (2013). Högstadielära användning av naturvetenskapligt språkbruk i kemiämnet i TIMMS. *Utbildning och Demokrati*. Örebro universitet (s. 53-68).
- Josefsson, Gunlög (2001). *Svensk universitetsgrammatik för nybörjare*. Lund: Studentlitteratur.
- Josephson, Olle, Oliv, Tomas & Melin, Lars (1990). *Elevtext: Analyser av skoluppsatser från åk 1 till åk 9*. Lund: Studentlitteratur.
- Kanebrant, Erik, Mühlenbock, Katarina Heimann, Johansson Kokkinakis, Sofie, Jönsson, Arne, Liberg, Caroline, af Geijerstam, Åsa & Falkenjack, Johan (2015). TMaster – A Tool for Assessing Students' Reading Abilities. In *7th International Conference on Computer Supported Education (CSEDU 2015), Lisbon, Portugal, 2015*. (s. 220-227).
- Knoch, Ute, Rouhsad, Amir, Storch, Neomy (2014). Does the Writing of Undergraduate ESL Students Develop after one Year of Study in an English-Medium University? *Assessing Writing 21, 2014*. (s. 1-17).
- Kouns, Maria (2010). *Inga IG i kemi A! En språkdidaktisk studie av en kemilärares undervisningsstrategier i en gymnasieklass med elever med svenska som andraspråk*. Licentiatuppsats, Malmö Högskola.
- Kouns, Maria (2014). *Beskriv med ord: fysiklärare utvecklar språkinriktad undervisning på gymnasiet*. Malmö: Malmö Studies in Educational Science, diss. Malmö Högskola.
- Larsson, Kent (1984). *Skrinförmåga. Studier i svenskt elevspråk*. Uppsala: Uppsala universitet.

- Laufer, Batia & Nation, Paul (1995) *Vocabulary Size and Use: Lexical Richness in L2 Written Production*, American Association of Applied Linguistics.
- Laufer Batia (2005). Lexical Frequency Profiles: From Monte Carlo to the Real World a Response to Meara. *Applied Linguistics*, American Association of Applied Linguistics.
- Lee, Sook Hee (2008). An Integrative Framework for the Analyses of Argumentative/Persuasive Essays from an Interpersonal Perspective. *Text & Talk* 28-2 (2008), (s. 239-270).
- Lemke, Jay (1990). *Talking science: language, learning, and values*. Norwood, N.J.: Ablex.
- Lim Falk, Maria (2008). *Svenska i engelskspråkig skolmiljö Ämnesrelaterat språkbruk i två gymnasieklasser*. ACTA UNIVERSITATIS STOCKHOLMIENSIS: Stockholms universitet.
- Lim Falk, Maria (2015). English and Swedish in CLIL student texts. In Dalton-Puffer, C. & Nikula, T. (red.), *Language Learning Journal*, Special issue: CLIL, Vol. 43, nr 3, s. 304-318. London: Routledge.
- Lim Falk, Maria & Holmberg, Per (2016). Paths to Academic Writing in a Globalized World: A Longitudinal Study of Content and Language Integrated Learning in Upper Secondary School in Sweden. In: *Recherches en écritures: regards pluriels / [ed] Sylvie Plane, Charles Bazerman, Fabienne Rondelli, Christiane Donahue, Arthur N. Applebee, Catherine Boré, Paula Carlino, Martine Marquilló Larruy, Paul Rogers, David Russell*, Crem: Université de Lorraine, 2016, (s. 187-211).
- Lindberg, Inger (2006). Med andra ord i bagaget. I: Bjar, Louise. (red.). *Det hänger på språket*. Lund: Studentlitteratur.
- Lindberg, Inger, & Johansson Kokkinakis, Sofie (red.) (2007). *OrdIL. En korpusbaserad kartläggning för grundskolans senare år*. ROSA-rapport 8. Institutet för svenska som andraspråk, Göteborgs universitet.
- Lindberg, Inger (2007). Forskning om läromedelsspråk och ordförrädsutveckling. I: Lindberg, Inger, & Johansson Kokkinakis, Sofie (red.). *OrdIL. En korpusbaserad kartläggning för grundskolans senare år*. ROSA-rapport 8. Institutet för svenska som andraspråk, Göteborgs universitet. (s. 13-60).
- Lindberg, Inger (2009). I det nya mångspråkiga Sverige. *Utbildning & Demokrati* 2009, vol. 18, NR 2, (s. 9-37).
- Lindberg, Inger (2017). *Ordförrådet – en framgångsfaktor*. <https://larportalen.skolverket.se>
- Lindberg, Inger & Johansson, Sofie (u.u.), The Development of Swedish Receptive Vocabulary in CLIL. A Multilingual Perspective. *Multilingual Matters*.
- Lindberg, Ylva & Sandblom Elisabeth (2014). En makro- och microanalys av elevtexter. I *Mångfaldens möjligheter. Litteratur- och språkdiradaktik i Norden*. (red.) Andersson, P., Holmberg, Per, Lyngfelt, Anna, Nordenstam, Anna, Widhe, Olle. Nationella nätverket för svenska med didaktisk inriktning (SMDI) och Nordiska nätverket för modersmålsdidaktik (NNMF) och författarna. (s. 73-97).
- Lundberg, Ingvar & Reichenberg, Monica (2008). *Vad är lättläst?* Härnösand: Specialpedagogiska skolmyndigheten.

- Lyngfelt, Benjamin (2007). Mellan polerna. Reflexiv- och deponenskonstruktioner i svenskan. I: *Språk och stil*, NF 17. (s. 86-134).
- Lyngfelt, Benjamin (2010). En akademiledamot pensioneras inte - om konstruktioner med s-verb i svenskan. I: *Bo65. Festskrift till Bo Ralph*. Göteborg: Meijerbergs. (s. 178-192).
- McEnery, Tony, Xiao, Richard Tono, Yukio (2006). *Corpus-based Language Studies*. London & New York: Routledge Applied Linguistics.
- Macken-Horarick, Mary (1996). Literacy and Learning across the Curriculum: Towards a Model of Register for Secondary School Teachers. In Hasan, Ruqaiya, Williams, Geoff. 1996. *Literacy in Society*. Addison Wesley Longman: London.
- Magnusson, Ulrika (2011). *Skolspråk i utveckling. En- och flerspråkiga elevers bruk av grammatiska metaforer i senare skolor*. Diss. Göteborgsstudier i nordisk språkvetenskap 17: Göteborgs universitet.
- Magnusson, Ulrika. & Johansson Kokkinakis, Sofie (2008). *Quantitative Measures on Students' Texts*. ASLAS: skriftserie nr. 22/2008, s. 43-56.
- Martin, James Robert, & Rose, David (2008). *Genre Relations. Mapping Culture*. Equinox Textbooks and Surveys in Linguistics.
- Matre Synnøve & Solheim Randi (2016). Opening Dialogic Spaces: Teachers' Metatalk on Writing Assessment. *International Journal of Educational Research*. Volume 80, (s. 188-203).
- Meara, Paul (1996). The Dimensions of Lexical Competence. I: Brown, G., Malmkjær, K., Williams, J. (eds.) *Performance and Competence in Second Language Acquisition*. Cambridge: Cambridge University Press, (s. 35-52).
- Meara, Paul, (2005) Lexical frequency profiles: A Monte Carlo analysis, *Applied Linguistics*, American Association of Applied Linguistics.
- Melin, Lars, Lange, Sven (2000). *Att analysera text. Stilanalys med exempel*. Lund: Studentlitteratur
- Mühlenbock, Katarina (2009). Readable, legible or Plain Words. Presentation of an Easy-to-Read Swedish Corpus. I: Multilingualism, Proceedings of the 23rd Scandinavian Conference of Linguistics (Studia Linguistica Upsaliensia 8). Uppsala: Acta Universitatis Upsaliensis, (s. 325-327).
- Myhill, Debra (2008). 'Towards a Linguistic Model'. *Language and Education*, 22 (5), (s. 271-288).
- Myhill, Debra (2009). Becoming a Designer: Trajectories of Linguistic Development. I Beard, Roger, Myhill, Debra, Riley, Jeni and Nystrand, Martin (eds.) *The SAGE Handbook of Writing Development*. London: SAGE Publications Ltd. (s. 401-414).
- Nation, Paul (2001/2013). *Learning Vocabulary in Another Language*. Cambridge: Cambridge University Press.
- Nation, Paul & Anthony, L. (2016). Measuring vocabulary size. In *Handbook of Research in Second Language Teaching and Learning*, Volume III, E. Hinkel (Ed.) New York: Routledge.

- Nixon, John (2001). Kvalitet i SPRINT. Föreliggande broschyr är en kort sammanfattning av en längre undersökning av John Nixon; *Quality in SPRINT: Towards quality assessment and assurance in content and language integrated education - A field study report.* <http://www.skolverket.se/publikationer?id=854>
- Norberg Brorsson, Birgitta (2007). *Man liksom bara skriver. Skrivande och skrivkontexter i grundskolans år 7 och 8.* Diss. Örebro universitet.
- Nordenfors, Mikael (2011). *Skriftspråksutveckling under högstadiet.* Diss. Göteborgsstudier i nordisk språkvetenskap 16: Göteborgs universitet.
- Nyström, Catharina (2000). *Gymnasisters skrivande. En studie av genre, textstruktur och sammanhang.* Institutionen för nordiska språk, Uppsala universitet, Diss: Uppsala.
- Nyström, Catharina (2003). *Argumentera! En presentation av gymnasisters texter i databasen ARGUS.* (Svenska i utveckling nr 19. FUMS Rapport nr 209.) Uppsala.
- Nyström Höög, Catharina (2010). Mot ökad diskursivitet? Skrivutveckling speglad i provtexter från årskurs 5 och årskurs 9. *Svenska i utveckling nr 27.* Uppsala universitet. Uppsala. <http://uu.diva-portal.org/smash/get/diva2:380369/FULLTEXT01.pdf>
- Nygård Larsson, Pia (2011). *Biologiämnets texter: text språk och lärande i en språkligt heterogen gymnasieklass.* Diss. Malmö: Lärarytbildningen, Malmö Högskola.
- Oakhill, Jane, Cain, Kate & Elbro, Carsten (2015). *Understanding and Teaching Reading Comprehension.* London & New York: Routledge.
- Olevard, Helena (1999). *"Tonårsliv". En pilotstudie av 60 elevtexter från standardproven för skolor 9 åren 1987 och 1996.* (Svenska i utveckling nr 11. FUMS Rapport nr 209.) Uppsala.
- Olofsson, Mikael & Sjöqvist, Lena (2013). Bedömning i svenska som andraspråk I: Kenneth Hyltenstam & Inger Lindberg (red.) *Svenska som andraspråk-i forskning, undervisning och sambälle.* Lund: Studentlitteratur.
- Olsson, Eva (2011). *"Everything I Read on the Internet is in English". On the Impact of Extramural English on Swedish 16-year-old pupils' Writing Proficiency.* Licentiatuppsats, ROSA 15, Göteborgs universitet.
- Olsson, Eva (2015) Progress in English Academic Vocabulary Use in Writing among CLIL and non-CLIL Students in Sweden. *Moderna Språk.* 109 (2), (s. 51-74).
- Olsson, Eva (2016). *On the Impact of Extramural English and CLIL on Productive Vocabulary.* Gothenburg Studies in Educational Sciences. Diss. University of Gothenburg: Göteborg.
- Olsson, Eva & Sylvé, Liss-Kerstin (2015) Extramural English and Academic Vocabulary. A Longitudinal Study of CLIL and non-CLIL Students in Sweden. *Apples - Journal of Applied Language Studies.* 9 (2), (s. 77-103).
- Palmer, Anne & Östlund-Stjärnegårdh, Eva (2005). *Bedömning av elevtext.* Stockholm: Natur och Kultur.
- Paquot, Magali (2010). *Academic Vocabulary in Learner Writing: From Extraction to Analysis.* London & New-York: Continuum.

- Persson, Tomas (2016). *De naturvetenskapliga ämnesspråken. De naturvetenskapliga uppgifterna i och elevers resultat från TIMSS 2011 år 8*. Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Educational Sciences 9, Uppsala universitet.
- Prior, Paul (2005). Toward the Ethnography of Argumentation: A Response to Richard Andrews' 'Model of Argumentation in Educational Discourse'. *Text* 25(1), (s. 129-144).
- Qian, David D. & Schedl Mary (2004). Evaluation of an in-depth vocabulary knowledge measure for assessing reading performance. *Language Testing* 2004, 21 (s. 28-52).
- Randahl, Ann-Kristin (2014). *Strategiska skribenter. Skrivprocesser I fysik och svenska*. Studier från Örebro i svenska språket 9. Södertörn Doctoral Dissertations 91. Örebro universitet.
- Reijerstam, Helena (2015). *Assessing Content or Language? A Comparative Study of the Assessment Practices in three Swedish Upper Secondary CLIL Schools*. Licentiate Thesis in Subject Matter Education at the Department of Education and Special Education, Faculty of Education, University of Gothenburg.
- Ribeck, Judy (2015). *Steg för steg. Naturvetenskapligt ämnesspråk som räknas*. PhD. Diss., Data Linguistica. No. 28, Institutionen för svenska språket, Göteborgs universitet.
- Ribeck, Judy, Jansson, Håkan, & Emma Sköldberg (2014). *Från aspekt till övergripande – en ordlista över svensk akademisk vokabulär*. I: Nordiska studier i lexicografi 12. Rapport från Konferensen om lexicografi i Norden, Oslo, augusti 2013.
- Sandberg, Ylva (u.u.). *Teaching and Learning Content through two languages: The Teacher Perspective*. I Sylvén, L.K. (red.), (bokmanus)
- Schleppegrell, Mary J. (2004). *The Language of Schooling. A Functional Linguistics Perspective*. Mahawa, New Jersey & London: Lawrence Erlbaum Associates, Publishers.
- Schleppegrell, Mary & de Oliveira, Luciana (2006). An Integrated Language and Content Approach for History Teachers. *Journal for English for Academic Purpose* 5. (s. 254-268).
- Schleppegrell, Mary J., Greer, Stacy & Taylor, Sarah (2008). Literacy in History: Language and Meaning. *Journal of Language and Literacy*. Vol. 31, no. 2, 2008, (s. 174-187).
- Sellgren, Mariana (2005). Ämnesundervisning för flerspråkiga elever – integrering av språk och kunskap. I: Axelsson, M., Rosander, C. & Sellgren, M., *Stärkta trådar – flerspråkiga barn och elever utvecklar språk, litteracitet och kunskap*. Stockholm: Språkforskningsinstitutet i Rinkeby.
- Skolverket (2008). *Språk i ämnet*. Rapport, Magnusson, Ulrika. Stockholm: Fritzes.
- Skolverket (2011a). *Läroplan, examensmål och gymnasiegemensamma mål för gymnasieskola 2011, Lgy11*. <http://www.skolverket.se/publikationer?id=2705>
- Skolverket, (2011b). *Läroplan för grundskolan, förskoleklassen och fritidshemmet 2011, Lgr11*. <https://www.skolverket.se/publikationer?id=2575>
- Skolverket (2012). *Få syn på språket*. <http://www.skolverket.se/publikationer?id=2843>
- Skolverket (2014). *Vad innebär en språkutvecklande undervisning?*
<http://www.skolverket.se/skolutveckling/lorande/nt/grundskoleutbildning/amnesovergripande-sprak/2,8730>

- Skolverket (2015) *Betyg och studieresultat i gymnasieskolan år 2016*.
<https://www.skolverket.se/publikationer?id=3733>
- Skolverket (2016a). *Skolverkets lägesbedömning 2015*.
<http://www.skolverket.se/publikationer?id=3432>
- Skolverket (2016b). *Språkutvecklande arbetsätt*.
<https://www.skolverket.se/laroplaner-amnen-och-kurser/grundskoleutbildning/diskutera-och-utveckla/sprakutvecklande-arbetsatt>
- Skolverket (2016c). *Att läsa och förstå. Läsförståelse av vad och för vad?*
<https://www.skolverket.se/publikationer?id=3694>
- Skolverket (2016d). *PISA 2015. 15-åringars kunskaper i naturvetenskap, läsförståelse och matematik*. <https://www.skolverket.se/publikationer?id=3725>
- Skolverket (2017a). *Ordförrådet – en framgångsfaktor*. <https://larportalen.skolverket.se>
- Skolverket (2017b). *Skolverkets lägesbedömning 2017*.
<https://www.skolverket.se/publikationer?id=3786>
- Skolverket (2017c). *PM. Slutbetyg i grundskolan, våren 2017*. Dokumentdatum 2017-09-28, Dnr. 2017:01232 <https://www.skolverket.se/publikationer?id=3831>
- Sköldberg, Emma & Johansson Kokkinakis, Sofie (2011). *A och O om akademiska ord*.
<http://spraakdata.gu.se/carola/akademiska-ordlistor/AO.pdf>
- Solheim, Randi & Matre, Synnøve (2014). Forventninger om skrivekompetanse. Perspektiver på skrivning, skriveopplæring og vurdering i «Normprosjektet». *Viden om Læsning* 2014 (15), (s. 76-89).
- Språkbanken, Göteborgs universitet. <http://spraakbanken.gu.se/>
- Svensén, Bo (2004). *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag.
- Sylvén, Liss-Kerstin (2004). *Teaching in English or English Teaching? On the Effects of Content and Language Integrated Learning on Swedish Learners' Incidental Vocabulary Acquisition*. Diss. Göteborg: Department of English, Göteborgs universitet.
- Sylvén, Liss-Kerstin (2013) CLIL in Sweden — Why does it Not Work? A Metaperspective on CLIL across Contexts in Europe. *International Journal of Bilingual Education and Bilingualism*, 16 (3), (s. 301-320).
- Sylvén, Liss-Kerstin & Ohlander, Sölve (2014) The CLISS Project: Receptive Vocabulary in CLIL versus non-CLIL Groups, *Moderna Språk*, 108 (2), (s. 80-114).
- Sylvén, Liss-Kerstin & Thompson, Amy S. (2015) Language Learning and CLIL: Is there a Connection? *Journal of Immersion and Content-Based Language Education*, Volume 3, Number 1, 2015, (s. 28-50).
- Säljö, Roger (2000). *Lärande i praktiken – Ett sociokulturellt perspektiv*. Stockholm: Norstedts.
- Säljö, Roger (2015). *Lärande – en introduktion till perspektiv och metaforer*. Malmö: Gleerups.
- Teleman, Ulf (1999/2010). *Svenska Akademiens Grammatik*. Svenska Akademien, Stockholm: Norstedts.
- Toulmin, Stephen (1958). *The Uses of Argument*. New York, NY: Cambridge University Press.

- Uccelli, Paula, Galloway Philips, Emily, Barr, Christoffer, Menese, Alejandra & Dobbs, Christina. (2015). Beyond Vocabulary: Exploring Cross-Disciplinary Academic Language Proficiency and Its Association With Reading Comprehension. *Reading Research Quarterly* 50(3) 337–356
- Vagle, Wenche (2005). Tekstlengde + ordlengdesnitt = Kvalitet? Hva kvantitative kriterier forteller om avgangselevernes skriveprestasjoner. (s. 303-386). I: Berge, Kjell Lars, Evensen, Lars Sigfred, Hertzberg, Frøydis, Vagle, Wenche (red.). *Ungdommers skrivekompetanse. Bind II: Norskesamen som tekst*. Oslo: Universitetsforlaget.
- Vetenskapsrådets etiska regler. <http://www.codex.vr.se/texts/HSFR.pdf>
- Voss, James, & Means, Mary (1991). Learning to Reason via Instruction in Argumentation. *Learning and Instructions*, 1, (s. 337-350).
- Vygotskij, Lev (1934/1999). *Tänkande och språk*. Göteborg: Daidalos.
- Westman, Margareta (1969). *Språkvård/Språkets myller*. Stockholm: Norstedts.
http://fungerandemedier.se/sites/default/files/bifogade_filer/L%C3%A4sbarhet.pdf
- Wolfe, Christopher R. (2012). Individual Differences in the “Myside Bias” in Reasoning and Written Argumentation. *Written Communication* 29, (s. 477- 501).
- Wolfe, Christopher R., Britt, M. Anne, & Butler, Judie A. (2009). Argumentation Schema and the Myside Bias in Written Argumentation. *Written Communication*, 26, (s. 183-209).
- Wood, David, Bruner, Jerome S., Ross, Gail (1976). The Role of Tutoring in Problem Solving. *Child Psychol. Psychiat.*, Vol. 17, (s. 89-100). Pergamon Press: Great Britain
- Yoxsimer Paulsrud, Beth Anne (2014). *English Medium Instruction in Sweden: Perspectives and Practices in two Upper Secondary Schools*. Academic Dissertation for doctoral Degree. Stockholm: US-AB.
- Östlund-Stjärnegårdh, Eva (2002). *Godkänd i svenska? Bedömning och analys av gymnasieelevers texter*. Institutionen för nordiska språk, Uppsala universitet, Diss.

10. Bilagor

Bilaga 1. Materialinsamlingen i CLISS-projektet

Prov	Term 1	Term 2	Term 3	Term 4	Term 5	Term 6	Total
Bakgrunds- och motivationsenkäter	X						150
Allmän läsförståelse och ordkunskap	Svenska Engelska				Svenska Engelska		600
Akademisk läsförståelse och ordkunskap			Svenska Engelska			Svenska Engelska	600
Argumenterande skrivning	Svenska Engelska			Svenska Engelska			600
Utredande skrivning		Svenska Engelska			Svenska Engelska		600
No & So- prov						X	150
Totalt per klass	5	2	2	2	4	2	

Bilaga 2. Materialinsamling – exempel på genomförande

Typ av data	Tidsåtgång	Genomförs på:	Sept. – okt. 2013	Nov. – dec. 2013	Jan. – feb. 2014	Mars – april 2014
Skrivuppgift svenska	90 min.	dator	x			
Skrivuppgift engelska	90 min.	dator	x			
Läsförståelse svenska	40 min.	papper	x			
Läsförståelse engelska	40 min.	papper		x		
Vokabulärtest svenska	40 min.	papper		x		
Vokabulärtest engelska	40 min.	papper		x		
Motivationsenkät	60 min.	dator			x	
Språkdagbok		dator			x	
Synonymer och kollokationer svenska	40 min.	papper			x	
Synonymer och kollokationer engelska	40 min.	papper				x
Bakgrundsenkät	30 min.	dator				x
Klassrumsobservationer			x	x	x	x
Lärintervjuer			x	x	x	x
Elevintervjuer			x	x	x	x

Bilaga 3. Argumenterande text, skrivuppgift 2



För eller emot en bilfri stad

När vi vill övertala och övertyga andra i en viktig fråga, kan det vara en god idé att göra det genom att skriva.

Din uppgift är att skriva en text på svenska där du formulerar och motiverar din egen ståndpunkt i denna fråga: Är du för eller emot att göra din stad till en bilfri stad? Formulera din ståndpunkt klart och effektivt. Försök övertyga läsaren om att din ståndpunkt är den bästa med starka argument och träffande exempel. Ta gärna hänsyn till vad du tror att läsaren kan ha för motargument.

På nästa sida finns ett bildcollage som kanske kan inspirera dig att utveckla din argumentation.

Du har 120 minuter på dig att skriva din text. Tänk på att du behöver lite tid för att läsa igenom vad du skrivit.

(+tillhörande bildcollage)

Bilaga 4. Utredande text, skrivuppgift 3



Folkhälsoproblem

Att skriva kan vara ett sätt att reda ut svåra samband, resonera kring olika alternativ och dra slutsatser.

En del tecken tyder på att Sveriges befolkning mår bättre nu än tidigare. Vi får t.ex. allt längre genomsnittlig livslängd. Men på två olika områden har befolkningen under de senaste 25 åren börjat må sämre. Det gäller för det första övervikt och för det andra psykisk ohälsa, dvs. problem som exempelvis ångest, oro och sömnsvårigheter.

Din uppgift är att skriva en text där du resonerar kring varför dessa två folkhälsoproblem – övervikt och psykisk ohälsa – ökar i Sverige.

På det bifogade faktabladet (se texten nedan) finns några statistiska uppgifter som kan vara användbara för ditt resonemang.

Du har 90 minuter på dig att skriva din text. Tänk på att du behöver lite tid för att läsa igenom vad du skrivit.

Psykisk ohälsa

Sedan början av 1990-talet har statistiken visat att det psykiska välbefinnandet i vårt land har försämrats. Man har bland annat hur många som uppgett att de lider av ångest, oro, sömnbesvär, återkommande huvudvärk med mera.

Nu ser det ut som om det kanske är på väg att vända till det bättre igen. För åren 2004-2005 uppgav 24 procent av kvinnorna och 17 procent av männen att de hade dessa besvär. Det är en förbättring med två respektive tre procent när det gäller den psykiska hälsan jämfört med siffrorna från början av 2000-talet.

Utvecklingen är positiv i alla åldersgrupper utom unga kvinnor och män mellan sexton till tjugofyra år där den andel personer som säger sig uppleva lätt eller svår ångslan, ångest eller oro fortsätter att öka.

Övervikt och fetma

Sedan slutet av 1980-talet har det pågått en kraftig ökning av antalet överviktiga personer i Sverige. År 1989 led 37 procent av männen i Sverige av fetma eller övervikt, år 2000 var samma siffra över 50 procent. För kvinnor var siffrorna 28 respektive 36 procent. Vid den nya mätningen gjord 2005 så har ökningen avstannat. För männen är det fortfarande en ökning, men nu endast någon enstaka procent. För kvinnorna en liten minskning.

(från webbplatsen för stiftelsen Statshälsan: www.statshalsan.se)

Bilaga 5. Textlängd, individuella resultat

Understrukna resultat = samma elevs högsta respektive lägsta resultat vid flera tillfällen. Resultat inom parentes = samma elev med näst högsta resultat.

Text=> Antal graf- ord individ	1 <i>Hälsa och väl- stånd</i> (SO)		2 Folkhälsop- roblem (NO)		3 <i>Bilfri stad</i> (NO)		4 Helt delad föräldrapeng (SO)	
Program	variation, text- längd kortast/längst		variation textlängd kortast/längst		variation textlängd kortast/längst		variation textlängd kortast/längst	
EK icke-CLIL A								
B	125	593	84	506	148	607	42	556
	280	<u>1079</u>	264	<u>895</u>	207	1081 (1041)	132	<u>753</u>
NA icke-CLIL	233	709	366	858 (526)	225	<u>730</u>	159	750 (612)
EK CLIL	297	<u>784</u>	329	719	<u>217</u>	774	<u>102</u>	<u>665</u>
SA CLIL								
1	358	<u>993</u>	294	<u>1104</u>	303	1260	324	535
2	151	760 (609)	<u>161</u>	1184 (1062)	415	1552 (1304)	<u>311</u>	<u>1145</u>
NA CLIL								
1	89	1128 (768)	<u>280</u>	<u>783</u>	<u>323</u>	<u>1044</u>	<u>283</u>	847 (731)
2	<u>319</u>	<u>893</u>	264	<u>1279</u>	<u>379</u>	907	<u>395</u>	695

I tabellen ovan synliggörs att samma elever skriver antingen de kortaste *eller* de längsta texterna vid flera tillfällen. Exempelvis en elev i NA CLIL 2 som i skrivuppgift 1, 3 och 4 skriver den kortaste texten i sin klass vid de tre skrivtillfällena och vid det fjärde skrivtillfället i skrivuppgift 2 skrev elev den tredje kortaste texten. I NA CLIL 1, SA CLIL 2, NA icke-CLIL och EK icke-CLIL B skriver i respektive klass, samma elever de längsta texterna vid ett eller två tillfällen samt och vid upprepade tillfällen även de näst längsta texterna som visas genom understrukna tal i parentes.

Bilaga 6. Textlängd, (F) och (P)

Statistiska värden; medelvärde, standardavvikelse, SD av medelvärdet av populationen

0=P, 1=F		N	Mean	Std. Deviation	Std. Error Mean
Textlängd_1	pojkar	47	487,36	221,798	32,353
	flickor	88	536,06	195,050	20,792
Textlängd_2	pojkar	46	491,76	226,914	33,457
	flickor	88	554,59	201,235	21,452
Textlängd_3	pojkar	38	534,32	243,487	39,499
	flickor	88	644,08	268,088	28,578
Textlängd_4	pojkar	44	387,98	159,544	24,052
	flickor	81	468,67	181,500	20,167

Bilaga 7. Exempel på nominaliseringar i elevtexterna

acceleration
acceptans
analys
andning
anpassning
anspänningar
ansträngning
anställning
användning
användningen
argumentationer
assistans
avbetalningar
avgasutsläpp
avgångar
avkopplingen
avlämning
avslappning
avslutning
backning
bantning
begränsningar
behandling
behandlingen
behovet
belastningar
belastningen
belöning
Beroende
beroendet
beröm
besiktning
Beslut
betalning
betoning
bevarandet
Bidraget
bidraget
bilberoende
bilförbrukningen
bilförbud

Bilaga 8. Användningen av nominaliseringar

Medelvärden, min- och max samt SD och median för samtliga fyra undergrupper i de fyra skrivuppgifterna.

0=PC,1=CF, 2=PnC, 3=FnC		Text1SO_nom	Text2NO_nom	Text3NO_nom	Text4SO_nom
Pojkar	Mean	1,048	1,271	1,406	1,692
CLIL	N	23	21	17	13
	Std. Deviation	,6788	,9419	1,0703	1,1071
	Std. Error of Mean	,1415	,2055	,2596	,3071
	Minimum	,0	,0	,0	,4
	Maximum	2,7	3,3	3,6	3,6
	Flickor	Mean	1,133	,989	,914
CLIL	N	57	63	58	54
	Std. Deviation	,7806	,5597	,6309	,6888
	Std. Error of Mean	,1034	,0705	,0828	,0937
	Minimum	,0	,0	,0	,0
	Maximum	3,3	2,7	2,5	2,7
	Pojkar	Mean	1,154	1,032	1,043
non-CLIL	N	24	25	21	31
	Std. Deviation	,7241	,8484	,8066	,8512
	Std. Error of Mean	,1478	,1697	,1760	,1529
	Minimum	,0	,0	,0	,0
	Maximum	2,5	3,5	3,2	3,6
	Flickor	Mean	1,181	,996	,823
non-CLIL	N	31	25	30	27
	Std. Deviation	,7282	,7850	,7384	,6895
	Std. Error of Mean	,1308	,1570	,1348	,1327
	Minimum	,3	,0	,0	,0
	Maximum	3,1	2,5	3,4	2,5
	Total	Mean	1,133	1,043	,980
N		135	134	126	125
Std. Deviation		,7355	,7284	,7694	,8050
Std. Error of Mean		,0633	,0629	,0685	,0720
Minimum		,0	,0	,0	,0
Maximum		3,3	3,5	3,6	3,6

Bilaga 9. Exempel på verb i passiv form i elevtexterna

fraktas
framställs
fylls
färdas
föds
förbjuds
förbrukas
förbättras
förebyggas
förgiftas
förklaras
förminskas
förpestras
förs
förstärks
förstörs
försvåras
försämras
förvandlas
förväntas
förvärras
förändras
globaliseras
glöms
gynnas
hängs
höjs
hörs
idealiseras
ignoreras
införs
intalas
introduceras
kallas
klassas
kompenseras

Bilaga 10. Användningen av verb i passiv form

Medelvärden, min- och max samt SD och median för samtliga fyra undergrupper i de fyra skrivuppgifterna.

0=PC,1=CF, 2=PnC, 3=FnC		Text1SO_pass	Text2NO_pass	Text3NO_pass	Text4SO_pass
Pojkar CLIL	Mean	,530	,690	,706	,631
	N	23	21	17	13
	Std. Deviation	,3253	,4867	,5018	,6019
	Std. Error of Mean	,0678	,1062	,1217	,1669
	Minimum	,0	,0	,0	,0
	Maximum	1,2	1,6	1,7	1,9
Flickor CLIL	Mean	,604	,586	,595	,733
	N	57	63	58	54
	Std. Deviation	,4057	,4714	,4968	,5460
	Std. Error of Mean	,0537	,0594	,0652	,0743
	Minimum	,0	,0	,0	,0
	Maximum	1,8	2,7	2,4	3,2
Pojkar non- CLIL	Mean	,600	,656	,638	,639
	N	24	25	21	31
	Std. Deviation	,5883	,5701	,4213	,4318
	Std. Error of Mean	,1201	,1140	,0919	,0776
	Minimum	,0	,0	,0	,0
	Maximum	2,4	2,2	1,8	1,6
Flickor non- CLIL	Mean	,587	,712	,527	,811
	N	31	25	30	27
	Std. Deviation	,5136	,5480	,5601	,7035
	Std. Error of Mean	,0923	,1096	,1023	,1354
	Minimum	,0	,0	,0	,0
	Maximum	2,2	1,9	2,4	2,8
Total	Mean	,587	,639	,601	,716
	N	135	134	126	125
	Std. Deviation	,4531	,5049	,4990	,5619
	Std. Error of Mean	,0390	,0436	,0445	,0503
	Minimum	,0	,0	,0	,0
	Maximum	2,4	2,7	2,4	3,2

Bilaga 11. Exempel på akademiska ord i elevtexterna

accepterad
accepterade
accepterat
agerat
allmänna
allternativet
alternativa
alternativen
analyserade
anledning*
angivet
angivit
angående
anledningarna
anpassad
anpassade
anpassar
anpassat
antingen
antya
användningen
argument
argumenten
argumentera
argumenterar
argumentet
aspekt
aspekter
avsedd
avseende
avseenden
avsikt
avsikter
baserad
baserar
baseras
baserat
bedöma,
bedömer

Bilaga 12. Användningen av akademiska ord

Medelvärden, min- och max samt SD och median för samtliga fyra undergrupper i de fyra skrivuppgifterna.

0=PC,1=CF, 2=PnC, 3=FnC		Text1SO_akad	Text2NO_akad	Text3NO_akad	Text4SO_akad
Pojkar CLIL	Mean	,670	1,048	1,041	,662
	N	23	21	17	13
	Std. Deviation	,5112	,8023	1,0112	,5606
	Std. Error of Mean	,1066	,1751	,2453	,1555
	Minimum	,0	,0	,0	,0
	Maximum	1,7	3,5	3,8	1,7
Flickor CLIL	Mean	,604	,552	,871	,509
	N	57	63	58	54
	Std. Deviation	,4964	,4384	,4924	,3315
	Std. Error of Mean	,0658	,0552	,0647	,0451
	Minimum	,0	,0	,0	,0
	Maximum	2,7	2,0	2,3	1,3
Pojkar non-CLIL	Mean	,629	,768	,871	,606
	N	24	25	21	31
	Std. Deviation	,5000	,6135	,6341	,5066
	Std. Error of Mean	,1021	,1227	,1384	,0910
	Minimum	,0	,0	,0	,0
	Maximum	1,7	2,3	2,1	2,0
Flickor non-CLIL	Mean	,458	,448	,640	,411
	N	31	25	30	27
	Std. Deviation	,4081	,4204	,5062	,4750
	Std. Error of Mean	,0733	,0841	,0924	,0914
	Minimum	,0	,0	,0	,0
	Maximum	1,7	1,5	1,9	1,4
Total	Mean	,586	,651	,839	,528
	N	135	134	126	125
	Std. Deviation	,4812	,5709	,6172	,4397
	Std. Error of Mean	,0414	,0493	,0550	,0393
	Minimum	,0	,0	,0	,0
	Maximum	2,7	3,5	3,8	2,0

Bilaga 13. Resultat för lexikala frekvensprofiler

Användningsgrad av *Level 1* och *off-list*-ord i samtliga skrivuppgifter för CLIL och icke-CLIL

Grupp: CLIL/ icke-CLIL	Text1 Level 1	Text 1 Off-list	Text2 Level 1	Text 2 Off-list	Text3 Level 1	Text 3 Off-list	Text4 Level 1	Text 4 Off-list
CLIL N	80		84		75		67	
Medel	72,39	14,03	75,28	12,20	75,10	12,14	75,81	11,41
Stand. dev.	4,56	3,72	4,72	3,77	4,62	3,28	4,7	3,6
Minimum	62,0	7,3	61,8	5,5	63,1	6,5	62,1	5,8
Maximum	81,2	22,0	86,5	23,7	83,4	21,8	84,9	22,3
icke-CLIL N	55		50		51		58	
Medel	74,18	12,84	73,62	12,57	75,95	11,9	77,08	11,37
Stand. dev.	4,10	3,20	10,96	3,69	4,62	3,49	4,96	3,61
Minimum	65,9	6,2	6,0	6,3	56,1	5,2	65,7	4,1
Maximum	82,0	21,1	83,9	22,9	85,3	25,5	86,9	22,3
T-value	-2,33	1,93	1,21	-,56	-,99	,381	-1,46	,146
p-value	,021*	,056	,228	,576	,323	,702	,146	,884

Bilaga 14. Förkortningar som används i uppsatsen

- ANOVA** = *Analyses of Variances*, en samling statistiska metoder för hypotesprövning
- CEIL** = *Content and English Integrated Learning*, att läsa alla skolämnen på engelska förutom språkämnen
- CLIL** = *Content and Language Integrated Learning*, att läsa alla skolämnen på ett annat målspråk än modersmålet
- CLISS** = *Content and Language Integrated Learning in Swedish Schools*, projektet som studien utgår ifrån
- EK** = ekonomiprogrammet
- ESAO** = *En svensk akademisk ordlista*
- FC** = flickor CLIL
- FnC** = flickor icke (non)-CLIL
- IB** = International Baccalaureate, internationell utbildning med gemensamma styrdokument
- L1** = Language 1; förstaspråk, modersmål
- L2** = Language 2; andraspråk
- Lgy11** = *Läroplan, examensmål och gymnasiegemensamma mål för gymnasieskola 2011*
- LIX** = läsbarhetsindex, ett sätt att mäta språklig tillgänglighet
- Lp94** = *Läroplan för de frivilliga skolformerna 1994*
- N** = Populationsstorleken
- NA** = naturvetenskapsprogrammet
- NO** = naturorienterade skolämnen; biologi, kemi, fysik (teknik)
- nC** = non CLIL => icke-CLIL
- OrdiL** = förkortningen på projektet *Ord i läromedel*
- OVIX** = ordvariationsindex
- p-värde** = probability – sannolikhet. Ett *p-värde* beräknas i en statistisk analys av sannolikheten för att ett resultat skulle bero på tillfälligheter
- PC** = pojkar CLIL
- PnC** = pojkar icke (non) CLIL
- PISA** = *Programme for International Student Assessment*, en internationell kunskapsstudie av 15-åringars kunskaper i matematik, naturvetenskap och läsförståelse, genomförs vart tredje år.
- SA** = samhällsvetenskapsprogrammet
- SAG** = *Svenska Akademiens Grammatik*
- SD** = Standardavvikelse är ett statistiskt mått på hur mycket de olika värdena för en population avviker från medelvärdet
- SFG** = systemisk – funktionell grammatik
- SFL** = systemisk – funktionell lingvistik, en språk teori
- SO** = samhällsorienterade ämnen; geografi, historia, religion, samhällskunskap
- SPINT** = språk- och ämnesintegrerad undervisning

SPRINT = språk- och ämnesintegrerad undervisning

SPSS = *Statistical Package for the Social Sciences*, ett datorprogram för statistisk analys

TIMMS = *Trends in International Mathematics and Science Study*, en internationell studie som undersöker fjärde- och åttondeklassares kunskaper i matematik och naturvetenskap.

TOEFL-testet = världens mest kända språktest i engelska som används av över 9 000 universitet världen över. *TOEFL* står för Test av engelska som främmande språk