

Nr 38

En fråga om ord?

**Automatisk och manuell analys av gymnasieelevers
språkanvändning i nationella provet i kursen svenska 3**

Annika Hillbom

2022

Abstract

Annika Hillbom: *En fråga om ord? Automatisk och manuell analys av gymnasieelevers språkanvändning i nationella provet i kursen svenska 3*. Svenska i utveckling nr 38. Institutionen för nordiska språk, Uppsala universitet.

At the upper secondary school level, academic writing is trained during the courses *Swedish 3* and *Swedish as a Second Language 3*. The students' ability to write an academic text is evaluated in a national test, as part of the courses. In this integrated test, the students have to use content from source texts, i.e. texts collected from newspapers and popular science books. The aim of the present study is to examine, analyse and discuss the students' language from different perspectives. A total of 90 student texts are analysed. For the purpose of comparison, the 14 source texts of the test are also analysed, together with 13 published scientific articles. The methods used are an automatic analysis of ten linguistic traits (Heimann Mühlenbock 2013), and a manual analysis of four categories of words and expressions (developed from Järborg 2007) in a smaller sample of the student texts.

The results show that the student texts are closer in many linguistic aspects to the published articles than they are to the source texts. The latter have much shorter sentences and a higher OVIX-value than the other texts. The published articles have the highest value of mean word length, noun/pronoun ratio and nominal ratio. The main differences between the student texts and the published scientific articles are that the students tend to use longer sentences with more subordinate clauses, less semantic precision and concentration at the level of words, expressions and phrases. Higher grades correlate with longer texts, longer words and higher word variation, as well as with a higher proportion of nouns, subject words, academic words and expressions, and with more references to the source texts. Stylistic errors seem to have a clearly negative effect on the grading of the texts.

The main conclusions drawn are the following: 1. At least some of the students seem to be able to differentiate between the non-academic language of the source texts and the academic language which they are expected to use in order to succeed with the task. 2. Automatic text analysis might provide a valuable complement to the teachers' qualitative assessment for the grading of student texts. 3. The national tests in *Swedish 3* and *Swedish as a Second Language 3* seem to test the students' linguistic and structural skills of academic writing, including how to summarize and reference the texts of others. However, in order for the students to develop scientific thinking, reading and writing at a deeper level, these abilities need to be trained over time and in many subjects during upper secondary school.

Keywords: academic writing, national tests, Swedish, upper secondary school, automatic linguistic analysis, word categories.

Kph, Trycksaksbolaget, Uppsala 2022.

Innehåll

1 Inledning.....	8
2 Teori och tidigare forskning	10
2.1 Trender inom skrivforskningen	10
2.2 Forskning om elevtext, ordförråd och ordanvändning	13
2.2.1 Språkutveckling och ordinlärning	13
2.2.2 Elevspråksstudier	14
2.2.2.1 1970-tal: <i>Gymnasistsvenska</i>	14
2.2.2.2 1980- och 1990-tal: språk, kön och socialgrupp	16
2.2.2.3 2000-tal: genrer och textbindning	18
2.2.3 Ordkunskapstester och passivt ordförråd	21
2.2.4 Mätning av texters språkliga komplexitet	22
2.2.5 Läromedelsspråk	23
2.3 Mer om skrivande i skolan	24
2.3.1 Genre, texttyp och diskurs	24
2.3.2 Skrivsituationer och skrivkontexter	27
2.3.3 Bedömning, textnormer och svenskämneskonceptioner	29
3 Syfte och frågeställningar.....	32
4 Material och metoder	34
4.1 Det nationella provet i kurs 3	34
4.2 Studiens källmaterial: elevtexter och jämförelsematerial	35
4.2.1 Elevtexter	36
4.2.1.1 Uppgifter och betyg	36
4.2.1.2 Övriga urvalsvariabler	37
4.2.1.3 Mer om NP år 2014 och 2015	38
4.2.1.4 Provets bedömningsmatris	39
4.2.2 Jämförelsematerial	40
4.3 Undersökningsmetoder	41
4.3.1 Automatisk analys med SVIT-modellen	41
4.3.1.1 Analysvariabler	42
4.3.1.2 Rensning och uppmärkning av materialet	47
4.3.1.3 Statistisk analys.....	48
4.3.2 Manuell analys: ordkategorier	49
5 Resultat från automatisk analys	54
5.1 Medelvärden i elevtexter, källtexter och vetenskapliga artiklar	54
5.2 Textlängd och meningslängd	55
5.3 Ordlängd, ordvariation och lemmavariation	59
5.4 Läsbarhetsindex.....	62
5.5 Bisatser och träddjup	63
5.6 Substantiv/pronomen-kvot och nominalkvot	66
6 Resultat från manuell analys	70
6.1 Ämnesord	70

6.2 Centrala begrepp	76
6.3 Akademiska ord och uttryck	80
6.3.1 Allmänvetenskapliga ord och uttryck	81
6.3.2 Referatmarkörer	81
6.3.3 Övriga akademiska textbindningsmarkörer	82
6.3.4 Akademiska ord och uttryck: sammanfattande kommentarer...	84
6.4 Stilbrott.....	84
6.5 Övriga språkliga iakttagelser i elevtexterna	85
6.6 Förhållandet till källtexterna	87
7 Sammanfattning och diskussion	90
8 Slutsatser	96
Referenser.....	100
Bilaga 1. Resultattabeller från automatisk analys.....	106
Bilaga 2. Tablåer över analyserade texter	118

Tabellförteckning

Tabell 1. Antal texter som ingår i studien.	35
Tabell 2. Medelvärden (SVIT) för alla textgrupper.	54
Tabell 3. Ämnesord i elevtexter med betyg A, uppgiften <i>Särskrivning och stavfel</i>	71
Tabell 4. Ämnesord i elevtexter med betyg E, uppgiften <i>Särskrivning och stavfel</i>	71
Tabell 5. Antal förekomster av de centrala begreppen i elevtexter med betyg A och E, uppgiften <i>Särskrivning och stavfel</i>	79
Tabell 6. Genomsnittligt antal referatmarkörer i elevtexter med betyg A och E, uppgiften <i>Särskrivning och stavfel</i>	82
Tabell 7. Medelvärden (SVIT) i elevtexter med betyg A, C och E på uppgiften <i>Särskrivning och stavfel</i> och <i>Skrivregler och sms</i>	106
Tabell 8. SVIT-värden i elevtexter, uppgiften <i>Särskrivning och stavfel</i> , betyg A.	107
Tabell 9. SVIT-värden i elevtexter, uppgiften <i>Särskrivning och stavfel</i> , betyg C.	108
Tabell 10. SVIT-värden i elevtexter, uppgiften <i>Särskrivning och stavfel</i> , betyg E.	109
Tabell 11. SVIT-värden i elevtexter, uppgiften <i>Skrivregler och sms</i> , betyg A.	110
Tabell 12. SVIT-värden i elevtexter, uppgiften <i>Skrivregler och sms</i> , betyg C.	112
Tabell 13. SVIT-värden i elevtexter, uppgiften <i>Skrivregler och sms</i> , betyg E.	114
Tabell 14. SVIT-värden i texthäftet.	116
Tabell 15. SVIT-värden i <i>Språk och stil</i> -artiklarna.	117

Tablåförteckning

Tablå 1. Exempel på ämnesord i kategori 2a i elevtexter, uppgiften <i>Särskrivning och stavfel</i>	72
Tablå 2. Exempel på ämnesord i kategori 2b i elevtexter, uppgiften <i>Särskrivning och stavfel</i>	74
Tablå 3. Förekomst av de centrala begreppen <i>särskrivning</i> , <i>felstavning</i> och <i>rättstavning</i> i elevtexter med betyg A och E: termer och böjningsformer.	78
Tablå 4. Användningen av de olika källtexterna i elevtexter med betyg A och E, uppgiften <i>Särskrivning och stavfel</i>	87
Tablå 5. Förteckning över elevtexter som analyseras i studien.	118
Tablå 6. Förteckning över källtexter som analyseras i studien (provets texthäfte vårterminen 2014).	120
Tablå 7. Förteckning över <i>Språk och stil</i> -artiklar som analyseras i studien.	121

Förord

Rapportserien Svenska i utveckling ges ut vid Institutionen för nordiska språk, Uppsala universitet. I serien redovisas främst studier som utförs med de nationella proven i svenska som underlag, men även andra arbeten med anknytning till skola och utbildning. Rapporten *En fråga om ord?* är en forskningsstudie av elevtexter från ett av de första nationella proven enligt den läroplan som infördes år 2011, Gyll.

Det aktuella provet är ett så kallat kursprov, vilket innebär att provet ska spegla ämnesplanens mål för undervisningen, det centrala innehållet och kunskapskraven i den aktuella kursen. Skrivuppgiften leder till skrivande av ett pm, en utredande text om ett svenskämnesanknutet ämne. Den efterfrågade texten ska representera akademiskt skrivande, vilket i ämnesplanen benämns texter av vetenskaplig karaktär. Att läsa och använda källor utgör en väsentlig del av uppgiften, vilket innebär att detta är vad man i bedömningsforskning kallar ett integrerat skrivprov där fler förmågor än just skrivförmåga provas.

Rapportens författare, Annika Hillbom, är universitetslektor vid Institutionen för pedagogik, didaktik och utbildningsstudier och har en bakgrund inom språkvetenskap med inriktning på svenska. I sin analys arbetar hon med tre korpusar, varav elevtextkorpuserna är huvudmaterial. Elevtexterna jämförs med dels de källtexter som ingår i provmaterialet och dels en korpus med artiklar från tidskriften *Språk och stil*. Genom en automatisk, kvantitativ analys av språkliga drag kan hon visa hur elevtexterna språkligt förhåller sig till de två jämförelsekorpusarna. Språket i elevtexterna ligger på många sätt närmare den vetenskapliga korpuserna än källtextkorpuserna, vilket är ett intressant resultat. Dessutom genomförs en manuell kvalitativ analys av ämnesord och akademiska ord i elevtexterna. Analysen visar bland annat att eleverna visserligen hämtar sitt huvudsakliga ordförråd från källtexterna, men när det kommer till vetenskapliga ord och uttryck bygger de på ett ordförråd som måste ha inhämtats via undervisning och annan läsning.

Rapporten vänder sig till läsare som är intresserade av skrivbedömning och skrivundervisning, till exempel forskare, provkonstruktörer och blivande eller verksamma lärare.

Uppsala i maj 2022

Anne Palmér
universitetslektor och docent,
vetenskaplig ledare för Nationella prov i svenska
och svenska som andraspråk

1 Inledning

Det akademiska skrivandet är en viktig del av alla högskolestudier. För många studenter innebär detta skrivande en av de största utmaningarna under studietiden. I skrivuppgifter såsom hemtentamina, pm och i längre uppsatser är förmågan att kunna leverera en akademisk text en förutsättning för att uppgiften ska bli godkänd. Högskolelärare vittnar ur sitt perspektiv om studenters svårigheter att skriva akademiska texter, och på många utbildningar ägnas mer och mer tid åt att stötta studenterna i detta skrivande.

Varför upplever då många studenter att det är extra svårt att skriva akademiskt, till skillnad från skrivande i andra sammanhang, såsom mer personligt, kreativt eller enklare faktabaserat skrivande? Det akademiska skrivandet är i många avseenden mycket krävande. Det förutsätter att man sätter sig in i ett ämne ur olika perspektiv, att man förstår och kan använda sig av akademiska texter som man läser, att man tänker igenom och planerar sin egen text och att man visar förmåga till både analys och precision. Själva texten ska vara välstrukturerad med god textbindning. Språket ska vara korrekt, precist och formellt. Referenshanteringen ska vara tydlig, lätt att följa och överensstämna med anvisningarna för det referenssystem som ska användas. Ämnesterminologin ska vara korrekt och konsekvent använd, och studenten måste även behärska akademiska ord och uttryck som används i samband med referenshantering och i textens resonemang.

Att lära sig att skriva akademiskt tar tid och kräver övning. Redan på gymnasiet ingår akademiskt skrivande i undervisningen i olika sammanhang, bland annat med syftet att förbereda eleverna inför högskolestudier. Det akademiska skrivandet är en central del i kurserna svenska 3 och svenska som andraspråk 3, och testas där i en skrivuppgift i ett nationellt prov. I denna skrivuppgift, ett så kallat pm, prövas elevernas förmåga att utifrån en given frågeställning referera, diskutera och dra slutsatser om ett antal texter. I bedömningen utvärderas olika aspekter av det akademiska skrivandet: innehåll och källhantering, disposition och sammanhang, samt språk och stil.

Ämnet för denna studie är gymnasieelevers akademiska skrivande i det nationella provet i kursen svenska 3. Termen *vetenskapligt skrivande* används i studien om produktion av ny kunskap, medan *akademiskt skrivande* används mer generellt om allt skrivande på högskolan, inklusive texter som pm, hemtentamina och kurslitteratur. I det nationella provet på kursen svenska 3 används formuleringen ”skrivande av vetenskaplig karaktär”, vilken kan sägas motsvara beskrivningen av akademiskt skrivande.

Ur högskolans perspektiv är det angeläget att känna till studenternas förkunskaper inom akademiskt skrivande, för att kunna erbjuda adekvat stöd. Ur gymnasieskolans perspektiv är det viktigt att få kunskap om vad eleverna inför högre studier behöver träna på inom akademiskt skrivande, för att kunna erbjuda en relevant undervisning på området. I studien jämförs därför gymnasieelevers texter med autentiska vetenskapliga artiklar, liksom med de källtexter som eleverna

utgår ifrån när de skriver sina nationella prov-texter. Studien syftar även till att öka kunskapen om det nationella provet i kursen svenska 3 i sig, eftersom provet i nuvarande utformning endast har funnits sedan 2014. Sådan kunskap kan till exempel vara till nytta i samband med uppgiftskonstruktioner och utformning av bedömningsanvisningar.

Studien har ett språkligt fokus, särskilt på ordförrådet – akademiska ord och uttryck, ämnesord med mera. Detta är ett område där många högskolelärare har vittnat om brister i studenttexter (se till exempel Blücker 2006), samtidigt som ordförrådet är avgörande för att bära upp en texts innehåll. Syntaxen ingår dock också i studien i avsikt att placera in ordförrådet i ett mer fullständigt språkligt sammanhang.

Det finns många studier av gymnasieelevers texter (se vidare avsnitt 2.2.2). Mycket av det som analyseras i denna studie har studerats tidigare, i olika sammanhang. Elevers och studenters språkförmåga och kunskap om olika genrer är dock något som förändras över tid och påverkas av sådant som nya läs- och skrivvanor i samhället, nya läroplaner och nya krav inom skola och universitet samt förändrad språklig och breddad socioekonomisk bakgrund hos elever och studenter. Det är därför viktigt med kontinuerliga studier av elevers och studenters skrivförmåga, vilket alltså föreliggande studie är tänkt att bidra till.¹

¹ Jag vill tacka Katarina Heimann Mühlenbock för utförandet av de automatiska analyserna med den SVIT-modell för automatiserad textanalys som hon har utvecklat. Jag vill också tacka Anne Palmér för stöd och hjälp under hela mitt arbete med studien. Anne Palmér och Saga Bendegard gav värdefulla kommentarer på mitt manus inför publicering. Tack också till Catrin Isaksson för ett gediget redigeringsarbete inför tryckningen av rapporten. Ytterligare forskare, samt provkonstruktörerna av det nationella provet i svenska 3, har bidragit med kloka synpunkter när studien har presenterats vid seminarier och på konferenser – tack!

2 Teori och tidigare forskning

I kapitlet beskrivs utvecklingslinjer inom framför allt den svenska skrivforskningen från 1970-tal fram till idag, med särskilt fokus på skolskrivande och elevtext. Teoretiska begrepp av relevans för studien presenteras. Forskningsstudier som denna studie i olika avseenden bygger vidare på refereras.

2.1 Trender inom skrivforskningen

Mona Blåsjö (2010) beskriver i en översikt skrivteori och skrivforskning från 1970-talet och framåt. Fokus ligger på nordisk forskning, men Blåsjö gör även internationella utblickar. Hon ger följande sammanfattande beskrivning av skrivforskningens inriktningar och utveckling under tidsperioden:

På ett mycket generellt plan kan man säga att skrivforskningen gått från individcentrerad elevforskning och psykologisk forskning till social forskning (även om vuxna), först textcentrerad och därefter mer etnografisk och kontextuellt inriktad. Intressefokus har alltså vidgats från individens skrivprocess till den sociala dialogen. Med det har även metoderna ändrats från experiment till etnografiska studier av verklig interaktion; textanalyser har förekommit parallellt. [...] Som parallellt löpande kan man också se språkforskningens allmänna breddning från grammatisk-strukturalistisk via textlingvistisk-kognitivistisk och social-kontextuell till kritisk [...] (Blåsjö 2010:8–9)

Blåsjö beskriver den tidiga svenska elevspråksforskningen (bland annat Hultman & Westman 1977; Larsson 1980, 1984) som mycket språkligt inriktad, med utgångspunkt i en kvantitativ grammatisk-stilistisk tradition. Fokus låg på språkliga enheter i den färdiga texten, vilka gärna relaterades till variabler som kön och socialgrupp. Samtidigt urskiljer Blåsjö i denna forskning ansatser till en mer socialt inriktad och genreinriktad språksyn. Blåsjö nämner också att man inom arbetsgruppen för nationella prov vid Uppsala universitet har gjort och gör studier av elevspråk som delvis kan ses som uppföljningar av de nämnda studierna från 1970- till 1980-tal, vilket alltså även föreliggande studie i viss mening kan ses som (Blåsjö 2010:26-28).

På senare tid har den norska skolskrivningsforskningen växt sig stark. Blåsjö beskriver denna som, i jämförelse med den svenska, ”mindre text- och språkcentrerad, mer social/pedagogisk och förankrad i den sociokulturella forskningens rötter hos Bachtin, Vygotskij m.fl.” (Blåsjö 2010:40). I denna forskning återfinns tankegångar som Olga Dysthes uppfattning att dialog i undervisningen är viktig för att elevers skrivande och lärande ska utvecklas, och hennes uppfattning att elevens samspel med andra texter och yttranden, liksom skrivandet av den egna texten, bidrar till utvecklingen av tankar och kunskaper hos eleven. Skrivandet ses som ett medel för att utveckla kunskaper och tankar, och inte endast som en produkt i sig (Blåsjö 2010:41–42).

Ytterligare ett viktigt perspektiv som Blåsjö (2010:74) berör, apropå den socialt inriktade skrivforskningen, är huruvida ”skrivutveckling [ses] som en följd av socialisering in i existerande kulturer som inte ifrågasätts eller som ett mer aktivt deltagande i nya kulturer som därmed utmanas och utvecklas”. Blåsjö ser dock inte någon direkt motsättning mellan dessa perspektiv, utan menar att det snarare är en fråga om var man lägger fokus.

Catharina Nyström (2000) urskiljer två utvecklingslinjer inom den svenska elevspråksforskningen under 1970- till 1990-tal. För det första börjar man i allt högre grad intressera sig för det skrivande som leder fram till en text, och för sambandet mellan text och skrivande; det vill säga fokus ligger inte längre uteslutande vid den färdiga produkten. För det andra börjar man intressera sig för större språkliga enheter än satser och meningar.

Att man under 1970- till 1990-tal började se på skrivandet på nya sätt hade mycket med introduceringen av den processororienterade skrivpedagogiken att göra. Inte minst inom den norska skrivforskningen, vilken ofta har didaktiska inslag, började sidor av skolskrivandet såsom responsarbete, loggskrivande och skrivande som kognitivt redskap uppmärksammas under tidsperioden (Nyström 2000:23–24). I Blåsjös (2010) forskningsöversikt över skrivteori och skrivforskning ges många ytterligare exempel på inriktningar och områden inom skrivforskningen, inom både skola och yrkesliv. Bland annat berörs skrivandet med digitala verktyg, främst dator. Det har funnits en farhåga att datorskrivandet skulle kunna leda till sämre texter, med fler muntliga drag etc. Detta har dock inte kunnat bekräftas av forskningen; snarare beror den skiftande kvaliteten på att det idag är lättare att publicera texter och att fler gör det (Blåsjö 2010:34–35). En studie som behandlar detta ämne är *Elevernas skriftspråkande i skolan och nya medier. Slutrapport från projektet ”Att lära sig skriva i IT-samhället”* av Sylvana Sofkova Hashemi och Ylva Hård af Segerstad (2007). Ytterligare en nyare inriktning inom skrivforskningen är studier av multimodalitet: samverkan mellan text och bild med mera. Charlotte Engblom analyserar till exempel yngre barns multimodala texthantering vid läsande och skrivande på dator i skolan och i hemmet (Engblom 2012), samt yngre barns användning av bilder från internet i egna texter (Engblom 2016). Studierna belyser bland annat den ökande multimodala kompetensen hos barn och unga samt pekar på behovet av att multimodala perspektiv på textbehandling fångas upp i undervisningen i skolan.

På 2000-talet har ytterligare nya metoder börjat användas inom skrivforskningen. Bland annat används automatisk analys, vilket också görs i den här studien. Victoria Johansson (2009) undersöker tal- och skrivutveckling under skoltiden hos tio-, tretton- och sjuttonåringar samt på universitetet hos vuxna studerande. Både talade och skrivna texter analyseras med hjälp av dator teknik och kvantitativa metoder. De talade texterna videofilmas och de skrivna texterna spelas in med så kallad tangentbordsloggning, vilket bland annat gör det möjligt att se hur lång tid skribenterna ägnar åt olika moment i texterna och vilka pausmönster de har. Ett flertal mått analyseras (antal ord och satser, fraser, lexikal diversitet och

densitet med mera). Resultaten indikerar bland annat att det sker ett stort utvecklingssprång mellan tretton och sjutton år, samt att sjuttonåringarna behöver längre tid på sig för att skriva texter som är ungefär lika långa som de vuxnas, vilket tyder på att uppgiften är mer krävande för sjuttonåringarna än de vuxna.

Megyesi, Näsman och Palmér (2016) har iordningställt *Uppsala elevtext-korpus*, som vid projektets start bestod av 2500 elevtexter från de nationella proven i svenska och svenska som andraspråk. I den annoterade korpusen beskrivs ordens ordklass, morfologiska drag och syntaktiska struktur. Det är möjligt att göra sökningar i korpusen utifrån variablerna kön, geografiskt område, ålder och betyg. Megyesi, Näsman och Palmér (2016) har också utvecklat det automatiska analysverktyget Swegram, som är fritt tillgängligt på nätet och kan användas för analys av till exempel ordklassfördelning, ordlängd, ordvariation och nominalkvot i elevtexter och andra texter. I Palmér (2018) beskrivs en pilotstudie där hon har analyserat så kallade benchmarktexter från de nationella proven i svenska och svenska som andraspråk, kurs 1 och kurs 3 på gymnasiet. Benchmarktexter är elevtexter som har bedömts av flera expertlärare i samband med utprovning av nya prov och som används som exempel på hur bedömningskriterierna kan tillämpas. Palmér gör en automatisk analys genom Swegram av textlängd, ordlängd, ordvariationsindex (OVIX) och nominalkvot. Syftet är att undersöka om medelvärden, medianvärden eller totalvärden korrelerar med betyg, vilket skulle kunna innebära möjligheter till att använda automatisk analys åtminstone inledningsvis vid bedömning av elevtexter i de nationella proven. Den automatiska analysen skulle kunna användas för en första grovsortering av texternas betygsnivåer, varefter läraren skulle kunna fokusera på de textegenskaper som inte fångas av automatisk analys, såsom innehållets relevans, originalitet och källhantering (Palmér 2018).

Resultatet av Palmérs studie visar på statistiskt signifikanta korrelationer mellan de analyserade måtten och betyg i nationella provet för kurs 1. I provet för kurs 3 (samma prov som analyseras i denna studie, dock andra elevtexter) korrelerar ordlängd och nominalkvot med betyg, men inte textlängd och OVIX. Palmér drar bland annat slutsatserna att automatisk analys bör designas i förhållande till de textnormer som råder i ett specifikt skrivprov och att den automatiska analysen måste vara mer mångfasetterad än i pilotstudien (Palmér 2018).

Pilán och Volodina (2018) undersöker vilka mått som vid automatisk analys bäst ringar in vilken språknivå inlärare av andraspråk befinner sig på i läsförståelse och egen textproduktion. De identifierar 14 olika mått som tillsammans verkar kunna ge en adekvat bild av inlärarens nivå. Både morfo-syntaktiska och lexikala mått ingår. En del av dessa mått är relevanta även för denna studie, medan andra är inriktade på att mäta språkdrag som kan visa vilken nivå inläraren befinner sig på.

Automatisk analys skulle kunna tjäna som komplement vid bedömning av elevtexter, bidra till en mer likvärdig bedömning och spara tid för lärare. För engelskspråkiga texter används redan en del automatiska bedömningsverktyg

(Palmér 2018). För bedömning av svenska elevtexter skulle automatiska verktyg kunna utvecklas ännu mer. Denna studie kan ses som ett led i en sådan utveckling.

2.2 Forskning om elevtext, ordförråd och ordanvändning

I avsnittet fokuseras elevtextstudier av relevans för den aktuella studien, med särskilt fokus på ordförråd. Även ordinlärning, aktivt och passivt ordförråd samt elevers textuella förebilder behandlas.

2.2.1 Språkutveckling och ordinlärning

Teleman (1989:7–8) beskriver tre olika perspektiv på barns och ungdomars språkutveckling: det strukturella/lingvistiska (hur barnet lär sig ord, grammatik och textmönster), det funktionella/psykologiska (fokus på språkbruket; de kommunikativa och kognitiva processerna) och det pedagogiska/sociologiska (vad som påverkar elevens språkutveckling, till exempel i undervisningen). Inom ramen för det strukturella/lingvistiska perspektivet konstaterar Teleman att elever lär sig de flesta ord utanför skolan, och att merparten av de generella grammatiska reglerna har lärts in innan skolan börjar. När det gäller ord inkluderar inlärningen bland annat hur orden uttalas, stavas, böjs, vad de betyder, hur ord kombineras med varandra och vad kombinationerna betyder (Teleman 1989:8).

Åke Pettersson (1989) pekar särskilt på de långa ordens betydelse i elevers språkutveckling. När elever lär sig att använda fler långa ord blir det möjligt för dem att skriva enklare om komplicerade saker, vilket enligt Pettersson tyder på att skrivförmågan har utvecklats. Pettersson tänker sig tre stadier i utvecklingen av skrivförmågan: 1) eleven skriver enkelt om näraliggande ting, 2) eleven skriver komplicerat om mer generella och invecklade ämnen, och 3) eleven förmår skriva enkelt om svåra ämnen. Pettersson menar att språkutvecklingen har ett nära samband med mental, eller kognitiv, utveckling.

Språkutveckling innebär bland annat att ordförrådet växer och att möjligheterna till syntaktisk variation ökar. Här ingår också att kunskapen om hur språket används inom olika genrer och i olika kontexter ökar. På ordnivå innebär språkutveckling även att kunskapen om ords betydelsenyanser och stilistiska värden ökar. Språkutveckling på ordnivå innefattar också lexikaliseringsprocessen, det vill säga att gå från att situationer beskrivs med längre formuleringar till att behärska lexikaliserade, självbärande begrepp (se Larsson 1984:278–280). Den innefattar även att gå från passiv ordförståelse till att aktivt kunna använda orden när man skriver. Här kan till exempel lexikal osäkerhet, i form av sådant som särskrivningar, stavfel och böjningsfel, indikera att begreppet inte är helt lexikaliserat hos språkbrukaren. Ordet kanske finns, men är ofullständigt lagrat, i språkbrukarens mentala lexikon (Aitchison 2012). Å andra sidan kan kanske (många) korrekta böjningar indikera att begreppet är lexikaliserat hos språkbrukaren.

Katharina Hallencreutz (2003) analyserar särskrivningar, bindestrecksfel och sammanskrivningar i mellanstadieelevers texter. De olika fenomenen diskuteras i förhållande till ordklass, antal stavelser, betoning, skolans undervisning och engelskans inflytande (vilket Hallencreutz menar förefaller mindre än förväntat). Hallencreutz påpekar att även om skolelever med svenska som modersmål skulle ha svårt att avgränsa ordbegreppet (Larsson 1984), så har de dock i alla fall klart för sig var ord- och morfemgränserna går; detta visar mellanrummen och bindestrecken i elevtexterna.

2.2.2 Elevspråksstudier

I avsnittet tecknas en övergripande bild av elevspråksforskningen från 1970-tal till 2000-tal, med fördjupade redogörelser för studier som är särskilt relevanta för denna studie.

2.2.2.1 1970-tal: Gymnasistsvenska

Gymnasistsvenska av Tor G. Hultman och Margareta Westman (1977) utgör ett pionjärarbete inom den svenska elevspråksforskningen. Här introduceras kvantitativa mått av olika slag, vilka genomgående kombineras med kvalitativa analyser. Bokens syften är dels att kartlägga gymnasisters språkbruk, och dels att, utifrån kartläggningen, bidra med skrivråd till elever och svensklärare. Studien är både metodmässigt och resultatmässigt relevant för denna studie, varför en detaljerad redogörelse är motiverad.

Materialet i *Gymnasistsvenska* utgörs av 151 elevuppsatser från centralprovet i svenska i årskurs 3. Provet ägde rum våren 1970. Elevernas uppgift är att skriva en debattartikel på temat "Familjen och äktenskapet än en gång". Som utgångspunkt har de ett antal så kallade stimulanstexter om detta ämne. Som jämförelsematerial i studien används ett antal bruksprosatexter av professionella skribenter: broschyrer, tidningstext, läromedel och debatttexter. Genomgående i *Gymnasistsvenska* fokuseras betygs- och genusskillnader. Hur skriver femmorna jämfört med ettorna? Hur skriver flickorna jämfört med pojkarna? Och hur skriver gymnasister jämfört med vuxna, professionella skribenter?

Ungefär halva *Gymnasistsvenska* ägnas åt ordförrådet i elevtexterna. I kapitlet "Orden" behandlas bland annat ordvariation, vanliga och ovanliga ord, ordlängd och andel ord som lånats från stimulansmaterialet. *Nusvensk frekvensordbok* (Allén 1970–1975) används i analysen av vanliga och ovanliga ord. Resultaten visar bland annat att gymnasisterna använder fler vanliga och mindre precisa ord än de professionella skribenterna, samt att de håller sig mycket nära ordförrådet i stimulansmaterialet. Runt 80 procent av orden i elevtexterna återfinns i stimulans-texterna. Längre elevtexter får generellt högre betyg än kortare, upp till en viss nivå. För att få det högsta betyget, 5, verkar det krävas koncentration och precision på ordnivå, vilket snarare innebär ett mer ekonomiskt skrivsätt.

Eftersom några av de mått som används i *Gymnasistsvenska* även används i denna studie återges resultaten för dessa i elevtexterna. Gällande antal löpord får Hultman och Westman fram ett totalt genomsnitt på 588 ord per elevtext (flickor 620 ord, pojkar 565 ord, betyg fem 698 ord, betyg fyra 728 ord, betyg ett 357 ord). Det genomsnittliga OVIX-värdet för alla elevtexter är 64,8 (betyg fem 67,6 och betyg ett 58,7). Gällande ordlängd är genomsnittet 5,07 bokstäver per ord (betyg fem i genomsnitt 5,52, betyg ett i genomsnitt 4,83).

I nästa kapitel, "Ordklasserna", analyseras ordklassfördelningen i elevtexterna. Analysen är i första läget maskinell (dock med manuell ordklassstaggning) och kvantitativ, men alla resultat kontrolleras också för hand, vilket Hultman och Westman menar är nödvändigt för att siffrorna ska kunna förstås rätt. Det visar sig finnas många exempel på att enstaka ord och uttryck snedvrider frekvenser. Inom vissa underkategorier visar de procentuella andelarna ett starkt samband med betyg. Detta gäller till exempel förekomsten av verbalsubstantiv (till exempel *posthämtning* istället för *hämta posten*), där andelen ökar med högre betyg, och förekomsten av personbetecknande pronomen, där andelen är högre i uppsatser med låga betyg. Inom många andra underkategorier hittas inga direkta samband med betyg. Sammantaget ligger femmornas ordklassfördelning närmast värdena för bruksprosan, medan ettornas ordklassfördelning ligger närmast talspråkets. I de övergripande slutsatserna pekar Hultman och Westman på att ordklassfördelningen speglar å ena sidan precisionen i elevtexterna, å andra sidan skillnaden mellan pratighet och koncentration. Ettorna använder oftare ord och ordklasser som i sig själva har oprecisa betydelser. Det är inte alltid som kontexten räcker till för att kompensera för detta. I uppsatser med högre betyg förekommer fler "ord med rikare semantisk specifikation" (Hultman & Westman 1977:178). Det finns också en tendens hos vissa skribenter, särskilt flickor med betyg tre, att skriva på ett alltför pratigt sätt, med många betydelsefattiga ord.

Resultaten av syntaxundersökningen bekräftar resultaten av ordundersökningarna. Det visar sig att långa meningar och många bisatser *inte* har något samband med höga betyg. Istället är det förmågan till exakthet och koncentration, främst genom nominalfraser, som korrelerar med höga betyg.

Avslutningsvis berörs i *Gymnasistsvenska* grafiska förhållanden och språkliga fel i elevtexterna. Vad gäller grafiska medel som styckeindelning, växling i meningslängd och skiljetecken visar det sig att gymnasisterna som grupp inte har lärt sig att utnyttja dessa. Här är skillnaderna mot jämförelsematerialet stora för alla betygsgrupper. Bland de språkliga felen behandlas sju huvudtyper, varav lexikonfel, semantiska fel, stilistiska fel och funktionella fel har relevans för denna studie. En del fel på dessa områden har att göra med att eleven blandar ihop närliggande ord och uttryck (till exempel *uråldrad* istället för *uråldrig* eller *föråldrad* och *i en viss mån* istället för *i viss mån*). Andra fel har att göra med att eleven inte behärskar vissa ordbetydelser (till exempel *utlämnad* kontra *ute-lämnad*). De semantiska felen kan också sträcka sig över en mening eller formulering. Hultman och Westman lyfter fram de semantiska felen som särskilt

allvarliga, eftersom de riskerar att påverka förståelsen av elevens resonemang. Det kan också vara så att de semantiska felen visar att eleven inte har förstått eller förmår resonera om det som hen skriver om.

Stilistiska fel definieras av Hultman och Westman som avvikelser från en jämn stilnivå. Hultman och Westman betraktar denna feltyp som mindre allvarlig, och menar att det snarare är oroande att eleverna uppvisar så liten språklig kreativitet. De pekar också på att det i sammanhanget är problematiskt att stimulanstexterna är skrivna med olika stilar.

Som funktionella fel nämns fel som har att göra med att eleverna har svårt att återge tankar i stimulanstexterna, samt att citera och hänvisa till dessa på ett korrekt sätt. Här lyfter Hultman och Westman fram att provmaterialet kan vara för omfattande, att stimulanstexterna är för svåra samt att de styr elevernas språk och tankar på ett olyckligt sätt. Kanske verkar provets utformning rentav rakt emot skolans uttalade mål att fostra eleverna till kritiskt och självständigt tänkande (Hultman & Westman 1977:259, 81–82). Eleverna förefaller bland annat behöva mer träning i att läsa och analysera texter kritiskt och de behöver få ett större aktivt och passivt ordförråd.

Slutligen påpekar Hultman och Westman att studien visar på stora språkliga skillnader mellan å ena sidan professionella skribenter och gymnasister, å andra sidan mellan olika grupper av gymnasister. Författarna nämner att elevernas skilda sociala förhållanden kan inverka på deras språk – en faktor som dock inte kunnat studeras i *Gymnasistsvenska*. Författarna uttrycker också farhågor att skolan i sig inte skapar lika förutsättningar för alla elever att utveckla sitt språk. Elevuppsatserna med låga betyg förefaller vara skrivna av elever som varken har tilltro till sitt eget språk eller lust att använda det. Hultman och Westman menar att skolan måste skapa rikliga tillfällen för eleverna att skriva sådant som upplevs som meningsfullt, liksom rikliga tillfällen att diskutera språk och innehåll.

2.2.2.2 1980- och 1990-tal: språk, kön och socialgrupp

Under 1980–1990-tal fortsätter den språkligt inriktade elevtextforskningen med fokus på variabler som kön och socialgrupp. Tor G. Hultman (1980) beskriver, utifrån ett antal studier av könsskillnader i elevspråk, några skillnader mellan flickors och pojkars språk. Flickorna skriver längre uppsatser med färre fel än pojkarna. Pojkarna använder en mer nominal stil än flickorna. Respektive grupps ordförråd speglar enligt Hultman skillnader i intressen och förhållningssätt i tillvaron: flickorna är till exempel mer inriktade på relationer, hus och hem, medan pojkarna är mer inriktade på skolarbetet och livet efter skolan. Flickornas språkutveckling ligger före pojkarnas i de tidiga skolåren, men pojkarna går om flickorna under gymnasietiden, vilket bland annat visas av att pojkarna då använder fler ovanliga och längre ord än flickorna. Hultman påpekar dock att könsskillnaderna i språkbruket överlag är sekundära i förhållande till ålder, betyg och socialgrupp. Vidare påpekar han att tolkningen av data är beroende av hur

man ser på sambandet mellan grammatik och tanke. Är grammatiska skillnader kopplade till tankemässiga skillnader eller kan de betraktas som funktionellt likvärdiga stilar?

Birgitta Garne och Siv Strömquist (1980) ifrågasätter att det skulle vara svårare för flickor än för pojkar att få det högsta betyget på sina uppsatser. Utifrån en ny studie av centralprov i årskurs 3 på gymnasiet, skrivna vårterminen 1970 (samma prov som Hultman och Westman använder i *Gymnasistsvenska*, dock andra uppsatser) visar de att fler flickor än pojkar får det högsta betyget på sina uppsatser. Resultaten bekräftar visserligen att pojkarna som grupp har en mer nominal stil än flickorna, men detta verkar inte leda till högre betyg i skolan. Garne och Strömquist ifrågasätter därmed att den nominala stilen skulle vara särskilt uppskattad av svensklärare. Istället hittar de ett annat område där flickorna har ett försprång framför pojkarna, nämligen textbindningen. God textbindning förefaller, enligt Garne och Strömquist, ha ett samband med höga betyg.

Temat *Skola, språk och kön* återkommer året därpå i en antologi med denna titel (Larsson 1981). Pettersson, Franzén och Rehman analyserar här förekomsten av så kallade emotiva ord (känslord) i elevtexter, till exempel *bra*, *viktig(-)*, *triv-*, *intress-* (positiva emotiva ord) och *svår*, *fel*, *problem(-)* och *dålig* (negativa emotiva ord). På gymnasiet förefaller förekomsten av emotiva ord vara positiva för betyget. Flickor använder något fler emotiva ord än pojkar. Pojkar använder något fler negativa emotiva ord än flickor. Könsskillnaderna är små.

I sin avhandling berör Larsson (1984) bland annat elevernas ordkunskap i förhållande till deras skrivande. Elevernas resultat på två olika ordkunskapsprov – ett matchningsprov och ett öppet synonymprov – ställs mot de betyg de fått på sina skrivprestationer. Larsson konstaterar bland annat att poängen på ordproven stiger med elevernas ålder och betyg, att det finns ett samband mellan socialgrupp och poäng på ordkunskapsproven, och att sambandet mellan kön och resultat på ordkunskapsproven i studien visar sig vara svagt. Naturligtvis finns det individuella skillnader. Larsson betonar att situation och användning har en avgörande inverkan i individens inläring av ordbetydelser. Skrivandet kräver högre ordbehärskning än läsandet:

[...] bland de skrivstarkare är det rätt många som också presterar bra på ordproven och bland de lässtarkare rätt många som också presterar sämre på ordproven. [...] Det synes alltså föreligga ett visst stöd för tanken att skrivandet i högre grad än läsandet är förbundet med självbärande eller operativt-kreativa ordenheter. (Larsson 1984:281)

Larsson (1984) utvecklar en teoretisk modell för analys av elevers skrivförmåga. Denna testas empiriskt på elever i årskurs 7, 9 och i årskurs 2 på gymnasiet. Ålder, social klass och kön visar sig ha starka samband med skrivförmåga. Texttypen är avgörande för hur väl eleverna lyckas i sitt skrivande. Språkutveckling ses av Larsson som utveckling av kognitiv förmåga. En avgörande fas i utvecklingen av skrivförmågan sker när individen försöker hitta språkliga uttrycksmedel för ett

mer abstrakt tänkande. De individuella skillnaderna kan vara stora, vilket återspeglas på både ord- och textnivå.

Ytterligare en studie av gymnasieelevers ordförråd görs av Björn Melander (1997). Melander studerar tre olika ordanalysmått i elevtexter skrivna i samband med utprovning av uppgifter till nationella provet i svenska 1995. De tre måtten är antal löpord, ordens medellängd i antal bokstäver och OVIX. Hultman och Westman (1977) visade att dessa tre mått hade ett starkt samband med betyg: längre uppsatser, längre ord och högre ordvariation visade sig ha ett samband med högre betyg (se också avsnitt 2.2.2.1). Melander ställer nu frågorna om typen av uppgift (PM, baksidestext och så vidare), samt skribentens kön och inställning till uppgiften (rolig, svår etc.), påverkar de undersökta måtten. Eleverna har i samband med utprovningen fått besvara en enkät med frågor om hur de upplever uppgiften.

Resultaten visar tydligt att typen av uppgift påverkar måtten. Skribentens kön förefaller inte ha någon större inverkan. Skribentens inställning förefaller inte heller ha någon större inverkan, även om resultaten här är något svårtolkade. Det visar sig bland annat att elever med högre ordmåttsvärden är mer benägna att besvara frågorna i enkäten än elever med lägre ordmåttsvärden.

Melander jämför även vilka ordmåttsvärden elever på teoretiska respektive yrkesinriktade program har i genomsnitt (alla uppgifter sammanslagna) och får fram följande värden (Melander 1997:131):

- Antal löpord: teoretiska program 262, yrkesinriktade program 217.
- Ordens medellängd: teoretiska program 4,81 bokstäver, yrkesinriktade program 4,59.
- OVIX: teoretiska program 67,8, yrkesinriktade program 59,9.

I förhållande till *Gymnasistsvenska* är elevtexterna avsevärt kortare i Melanders studie, vilket troligen till stor del beror på typ av uppgift och genre. Ordens medellängd är lägre i Melanders studie. OVIX för teoretiska program är ungefär samma som för betyg fem i Hultmans och Westmans studie, och OVIX för yrkesinriktade program ligger på samma nivå som betyg ett.

2.2.2.3 2000-tal: genrer och textbindning

Catharina Nyström (2000) undersöker genre, textstruktur och sammanhang i gymnasisters texter. Texterna samlades in 1996–1997 och består dels av texter skrivna i skolan, inklusive inom det nationella provet (NP) i svenska, dels av texter som eleverna skrivit på fritiden. Eleverna går på både teoretiska och yrkesförberedande program. Nyström strävar efter att inte ha ett normativt perspektiv i studien, och därför ingår inte till exempel betyg som en variabel.

Resultaten visar att eleverna visserligen – totalt sett – skriver texter inom många olika genrer, men att ett fåtal genrer dominerar i skolan, nämligen den

utredande uppsats, faktaredovisning, berättelse och bokrecension. Nyström beskriver de första två som rena skolgenrer som inte är särskilt gångbara utanför skolkontexten. Skrivandet i det nationella provet, vid denna tid inriktat på tidningsgenrer, skönlitterärt skrivande, litterär analys med mera, skiljer sig från det övriga skolskrivandet genom betoningen av kommunikativa situationer.

Nyström intervjuar också en del av eleverna som lämnat in texter, utifrån frågor om den egna skrivprocessen, provsituationen, beskrivning av provtext och självvald text, kriterier för textkvalitet och skrivundervisningens mål. Elevernas synpunkter används som tolkningsram vid genreklassificeringen och vid textanalysen.

Tyngdpunkten i Nyströms avhandling ligger på kvalitativa textanalyser, men Nyström genomför också en del kvantitativa undersökningar av bland annat antal löpord och OVIX. Resultaten jämförs med *Gymnasistsvenska* (Hultman & Westman 1977). De visar att eleverna i Nyströms studie (NP-texterna) skriver lite kortare texter och att OVIX är lite lägre än i *Gymnasistsvenska*. Elever på yrkesförberedande program har som grupp lägre värden än elever på studieförberedande program. Skillnaden i textlängd mellan flickor och pojkar har minskat, jämfört med i *Gymnasistsvenska*. OVIX varierar med genre: värdet för alla elever på utredande uppsats är 58, på faktaredovisning 66 och på debattartikel 62 (Nyström 2000:178). Faktaredovisningen är den genre som har högst OVIX. Detta är elevtexter som innehåller många avskrifter och lån från andra texter, och där sammanhanget mellan textens olika delar ibland brister.

Nyström närstuderar två elevtexter med höga OVIX-värden. Det visar sig att båda elevtexterna är ganska innehållsrika och ofta byter teman. I den första texten handlar det om växling mellan delaspekter av temat, medan det i den andra mer handlar om växling mellan olika exempel på ett problemområde. Ingen av elevtexterna har någon särskilt markerad satskonnektion; de lexikaliska sambanden blir istället viktiga för att binda samman texten. Dessa är av vissa typer:

De lexikaliska sammanhangsstrategierna i texten är däremot inte bindningar substantiv–pronomen utan de vilar på andra mekanismer: delidentitet, hyponymirelationer och inferens. Det är dessa mönster som ger utslag i de höga OVIX-värdena. (Nyström 2000:192)

Nyström jämför dessa elevtexter med höga OVIX-värden med en elevtext med lågt OVIX-värde, och konstaterar att den sistnämnda visar ett annat textbindningsmönster, med mer frekvent och tydligare satskonnektion samt ett mer omfattande pronomenbruk. Nyström konstaterar att ordvariation kan ha ett samband med en texts sammanhangsstruktur. Hon gör också reflektionen att det verkar finnas en optimal OVIX-nivå: ”Ett alltför högt värde tyder på en alltför snabb ämnesprogression, medan ett alltför lågt kan tyda på innehållslig tunnhet” (Nyström 2000:192).

Även ordlängd ingår i Nyströms studie. Hon konstaterar att den genomsnittliga ordlängden i hela undersökningsmaterialet är 4,7 bokstäver per ord. Den genre

som har högst genomsnittlig ordlängd är faktaredovisningarna med 5,26 bokstäver per ord, och den kortaste genomsnittliga ordlängden finns i provmaterialets öppna brev (4,29 bokstäver/ord) (Nyström 2000:179).

Eva Östlund-Stjärnegårdh (2002) undersöker var gränsen går mellan betygen godkänt och icke godkänt i nationella provet i svenska på gymnasiet. Materialet utgörs av uppsatser skrivna inom ramen för Svenska B-kursen 1997, vilka tillhör olika uppgiftstyper: skönlitterära, utredande och argumenterande uppgifter samt litterär analys. Uppsatserna har dels betygsatts av ordinarie lärare, dels av ett antal ytterligare lärare. Vid omdömet visar det sig att ett antal uppsatser som blivit godkända av den ordinarie läraren nu bedöms som icke godkända.

Östlund-Stjärnegårdh gör både kvantitativa och kvalitativa analyser av de 60 uppsatser som ingår i studien. I den kvantitativa analysen ingår bland annat ett antal ordanalysmått, såsom antal ord per text, per mening och per makrosyntagm, ordlängd, OVIX, andel så kallade banalord samt skriftspråklig eller talspråklig variant av vanliga ord. Dessutom analyseras textbindning, styckeindelning, meningsbyggnad, ordföljd, förekomsten av bisatser, innehåll, hur eleven följt instruktionen och hur texten fungerar i sitt sammanhang (så kallad helhetsbedömning). Östlund-Stjärnegårdh redogör också för resultaten av en enkätundersökning med lärare, där dessa får gradera hur viktiga olika faktorer är i avgörandet mellan IG och G. De viktigaste faktorerna enligt lärarna är helhetsbedömning, funktion i den angivna situationen, relevant innehåll, hur eleven följt instruktionen, anpassning till situationen, röd tråd och meningsbyggnad. Faktorer som stavning och textomfång anses mindre viktiga i sammanhanget (Östlund-Stjärnegårdh 2002:51–52).

Resultaten visar sammantaget att adekvat innehåll, och sättet att organisera innehållet, väger tyngre i lärarens bedömning än språkliga detaljer. När det gäller ordförrådsundersökningen bekräftas visserligen bilden sedan tidigare studier att längre elevtexter generellt får högre betyg, men det visar sig också att en del riktigt långa elevtexter – liksom mycket korta elevtexter – får IG. Jämfört med tidigare studier, till exempel *Gymnasistsvenska*, förefaller meningar och ord ha blivit kortare men elevtexterna längre. Undersökningarna av OVIX, banalord med mera visar enligt Östlund-Stjärnegårdh inga tydliga resultat i förhållande till gränsen mellan G och IG. Hon sammanfattar resultaten av ordförrådsundersökningarna enligt följande:

Ordförrådsundersökningarna visar små och ibland svårtolkade skillnader mellan betygen IG och G. Egentligen är det bara användningen av *dom* som är tydligt betygsskiljande. Också för kriteriet *Ordval* bekräftas lärarenkätens resultat att detta inte är ett avgörande kriterium för bedömningen av gränsen mellan Icke godkänd och Godkänd. (Östlund-Stjärnegårdh 2002:108)

2.2.3 Ordkunskapstester och passivt ordförråd

I de tidigare avsnitten har framför allt språkbrukares aktiva ordförråd fokuserats. Men det finns också forskning som är inriktad på språkbrukares passiva ordförråd. Här finns ofta ett samhällsperspektiv. Förstår medborgarna vad myndigheterna vill säga? Ordförståelsen blir en fråga om demokratiska rättigheter.

En tidig studie på detta område som ofta nämns är *Språkklyftan: Hur 700 ord förstås och missförstås* av Nils Frick och Sten Malmström (1976). Författarna undersöker hur deltagare i dåvarande Arbetsmarknadsstyrelsens kurser förstår ord från det sociala, politiska och ekonomiska livet. Informanterna får se ordet i en mening, till exempel *[h]an har inga abstinensbesvär*. De kryssar i ett av alternativen ”vet säkert vad ordet betyder”, ”tror jag vet vad det betyder” eller ”vet inte vad det betyder”. Därefter ska de försöka ange vad ordet betyder genom att ersätta det med ett eller flera ord med liknande betydelse (Frick & Malmström 1976:9, 13). Resultaten visar, enligt Frick och Malmström (1976:21), att en mycket stor del av informanterna inte tillfredsställande kan förklara vad många av orden i studien betyder, trots att de själva anser sig veta vad orden betyder. Det finns många exempel på feltolkningar och oklar förståelse, det vill säga att svaret bedöms som *delvis rätt* (Frick & Malmström 1976:15).

Frick och Malmström gör en gruppering av informanternas feltolkningar av olika ord och urskiljer följande typer:

- feltolkningar styrda av ljudlikhet, till exempel *egalt* förväxlas med *legalt*, *arvode* med *arv*
- förväxling med ord som har motsatt betydelse, till exempel *arbetsoförmögen* förväxlas med *kan arbeta*
- annan förväxling med ord som berör samma område, till exempel *amortering* förväxlas med *skuld*, *bronkit* med *halsont*
- svårigheter att hålla isär olika betydelser av ett och samma ord, till exempel *deklarera åsikter* förväxlas med *deklarera inkomster*
- svårigheter vid ordsammansättningar, till exempel *affärsbank* förväxlas med *bank i affären* (Frick & Malmström 1976:16–20).

I artikeln *Språkklyftan: 30 år senare* (2011) redogör Anna Gustafsson och David Håkansson för en mindre, uppföljande studie till Frick och Malmström (1976). Informanterna utgörs av ett fyrtiotal deltagare i arbetsmarknadsåtgärder. Testet utformades på samma sätt som i Frick och Malmström. I artikeln redogörs för informanternas förståelse av 14 ord som också ingick i Fricks och Malmströms studie. Resultaten visar att förståelsen av de aktuella orden förefaller ha minskat med tiden, åtminstone hos de yngre informanterna i studien. Gustafsson och Håkansson nämner att faktorer som informanternas utbildningsnivå, ålder och motivation förefaller inverka på resultatet. De redogör även för en studie beställd av Skatteverket 2005 och drar, till skillnad från vad som framkom i den studien, slutsatsen att språkklyftan i samhället snarast förefaller ha ökat.

Peter Cassirer (1977) menar att själva testkonstruktionen vid ordförståelsetest har avgörande betydelse för resultatet. Samma ord är olika svåra för olika grupper av språkbrukare. Det är skillnad på språkbrukares passiva och aktiva ordförståelse, vilket kan påverka resultatet om ett multiple choice-test ställs mot ett där deltagarna själva ska förklara orden. Det är inte heller helt okomplicerat att använda frekvensordböcker som underlag för bedömningen av vilka ord som kan anses vara svåra eller ovanliga – frekvensordböcker utgår ju från producerat språk och mäter inte sådant som hur ofta vi tänker på vissa saker, eller hur närvarande de är i våra liv. Även Christina Stage (1981) lyfter fram att provkonstruktionen kan inverka på resultaten på ordförståelsetest. Hon undersöker könsskillnader i ordkunskap i högskoleprovet. Stage konstaterar att resultaten ibland påverkas av könsens skilda erfarenheter och intressen, vilket bör vägas in vid provkonstruktionen.

I en studie från 2017 analyserar Anna Gustafsson och David Håkansson, med utgångspunkt i högskoleprovets orddel, hur förståelsen av olika ord har förändrats under perioden 2000–2011 i Sverige. I studien ingår också en jämförelse mellan olika gruppers ordförståelse baserad på kön, ålder och utbildningsnivå. Det visar sig bland annat att skillnaderna i ordförståelse relaterat till ålder och utbildningsnivå ökar under perioden.

2.2.4 Mätning av texters språkliga komplexitet

Hur lättläst är en text? Från 1920-talet och in på 2000-talet har olika sätt att mäta texters läsbarhet utvecklats. Det mest kända svenska måttet är läsbarhetsindex, LIX, som utvecklades av Carl-Hugo Björnsson (1968). LIX grundar sig på genomsnittlig meningslängd och andel långa ord. Som långa ord räknas ord med fler än 6 bokstäver. LIX utvecklades i en skolkontext (läroböcker, elevtexter), och det finns fastställda intervallvärden för olika genrer såsom skönlitteratur och tidningstext. LIX har kritiserats för att vara ett ytligt och osäkert mått, särskilt beträffande den genomsnittliga meningslängden. Kritiken till trots används LIX fortfarande.

Katarina Heimann Mühlenbock (2013:28) presenterar en översikt över 16 olika läsbarhetsmått, från 1921 till 1968. Alla utom ett (LIX) rör engelskan. Totalt 13 olika faktorer förekommer i läsbarhetsmått. Merparten av dessa rör ordförrådet: ordlängd i antal bokstäver eller antal stavelser, jämförelser med ordlistor, så kallade svåra ord, ordvariation, med mera. Därutöver förekommer några mått som rör fraser, meningar, kontext och idéer. De olika läsbarhetsmått grundar sig på olika kombinationer av 1 till 7 av dessa faktorer. Björnssons LIX-mått, som grundar sig på 2 faktorer, är därmed ett av de mått som inkluderar ett fåtal faktorer.

Heimann Mühlenbock utarbetar ett eget läsbarhetsmått, SVIT, vilket står för *Sentence structure, Vocabulary load, Idea density, and human and personal intresT*, eller, på svenska, *Syntax, Vokabulär, Intressegrad och (idé)Täthet*. Syftet är att skapa en modell som bättre än till exempel LIX mäter texters svårighetsgrad.

Heimann Mühlenbock är primärt intresserad av att kunna mäta så kallade lättlästa texter, men det mått hon utvecklar kan, liksom andra läsbarhetsmått, användas på alla typer av texter. I studien ingår korpusar med olika typer av material. I SVIT ingår drygt 20 olika variabler, inom fem områden (se vidare metoddelen, avsnitt 4.3.1.1).

2.2.5 Läromedelsspråk

Läromedel är alltså en av de genrer som har använts vid utvecklingen av läsbarhetsmått ovan (avsnitt 2.2.4). Lärobokstexter av olika slag, och i olika ämnen, är något som möter eleverna under hela skoltiden, och det är förstås en typ av texter som kan påverka elevernas eget skrivande på olika sätt. Forskning om läromedel, framför allt språkligt inriktad sådan, har därför relevans för föreliggande studie. I antologin *Läroboksspråk* (Strömquist 1999) behandlas språket i läroböcker av några olika språkvetare. I sin inledning konstaterar Siv Strömquist att det oftast är faktaurval och presentationsmodeller som fokuseras i diskussionen av läroböcker. Hon fortsätter: ”Detta är olyckligt av flera anledningar. Läroboken är inte bara en källa till kunskap och ett redskap för inläring och utveckling av tanke- och kunskapsstrukturer. Den ger också sina läsare mönster och förebilder av betydelse för den ständigt pågående språkutvecklingen” (Strömquist 1999:7). I sin inledning till volymen sammanfattar Strömquist de olika artiklarna. I studien av Melander, en översikt över svensk språkforskning om läroböcker, framkommer att det finns två olika perspektiv i denna forskning: dels ett deskriptivt, där syntax, ordförråd, stilistiska grepp och textuella egenskaper fokuserats, dels ett funktionellt perspektiv, där frågor om språkets lämplighet för avsedd åldersgrupp, om texterna främjar elevernas språkutveckling och dylikt stått i centrum. Forskningen – både den tidigare och studierna i den aktuella volymen – pekar på att läroboksspråket utmärks av en koncentrerad, nominal stil med ganska dålig textbindning. Sandqvists studie av språket i läroböcker i historia visar på hög informationstäthet och därmed rik ordvariation. Ordvariationen är kopplad till att innehållet är faktaspäckat, med tätt skiftande ämnen. Den beror inte på att språket i sig är varierat och rikt på synonymer. Texterna ger ett alltför komprimerat intryck (Strömquist 1999:8–11).

Ribeck (2015:231) analyserar ”de naturvetenskapliga ämnesspråken” i svenska läroböcker för årskurs 7–9 och gymnasiet, med mått som meningslängd, ordlängd, LIX, OVIX, typiskt ordförråd, nominalfraser, meningsstrukturer med mera. Som jämförelsematerial används språket i läroböcker i matematik och samhällsvetenskapliga ämnen, samt berättande och akademiska texter (de sistnämnda i form av korpusen SveAk i Språkbanken; avhandlingar och tidskriftsartiklar inom humaniora och samhällsvetenskap). Resultaten visar bland annat att de naturvetenskapliga ämnesspråken innehåller många ämnesspecifika termer, många fler än i samhällsvetenskapliga läroböcker. Resultaten visar också att läroboksspråket överlag är mycket enklare än akademiskt språk. Det akademiska

språket innehåller bland annat fler och längre nominalfraser, och många meningar som inleds med långa, utbyggda subjekt. Lärobokstexterna förefaller dock vara mer informationspackade än de akademiska texterna, i form av en hög andel ämnesord.

2.3 Mer om skrivande i skolan

Skrivande i skolan, liksom i alla andra sammanhang, sker i en kontext och utifrån olika typer av förutsättningar, vilka på olika sätt inverkar på textproduktionen. I detta avsnitt behandlas olika typer av förutsättningar som inverkar på textproduktion och textbedömning i skolan, med utgångspunkt i vissa centrala begrepp. Många av dessa hör till den nyare skrivforskningen, och avsnittet ger därmed också en överblick över denna nyare forskning.

2.3.1 Genre, texttyp och diskurs

Något som har diskuterats mycket inom skrivforskningen är begreppet *genre*. Vad menar man med genre? Är genren främst språkligt konstituerad, eller bör mer sociala kriterier användas för att definiera genrebegreppet?

Ledin (2001:26–30) definierar genre enligt följande:

- En genre kopplar texter till en återkommande social process där människor samhandlar genom texter.
- En genre innefattar prototypföreställningar om textutformningen.
- En genre är normalt namngiven och på så sätt språkligt och socialt kodifierad.
- En genre är en tradition som tas i bruk i en situation, varför den förändras över tid.

Gibbons (2016:116) ger, i svensk översättning, en liknande definition: ”Här använder vi ordet genre [...] med avseende på hur vissa saker hanteras genom språket i ett samhälle eller i en kultur.” Enligt Gibbons förekommer en genre i en specifik kultur och har ett specifikt socialt syfte, en specifik övergripande struktur och specifika språkliga särdrag. Det är inte endast skrivna texter som tillhör genrer. Gibbons exemplifierar genrer med tv-nyheter, vigselakt, lekprogram, lektion, skämt, telefonsamtal med en vän, rapport, instruktion och skriftlig argumentation.

Vid sidan av genrebegreppet, där alltså den sociala kommunikationssituationen betonas, förekommer också begreppet *texttyp* (Hellspong & Ledin 1997:24; Östlund-Stjärnegårdh & Palmér 2015:15). Texttyper kategoriseras främst utifrån språkliga grunder. Exempel på texttyper är berättande, beskrivande, instruerande, förklarande/utredande och argumenterande. En text kan innehålla en enda texttyp, men den kan också innehålla flera. I vissa sammanhang används termen *basgenre* ungefär motsvarande det som ovan kallas för texttyp. Vanliga sådana basgenrer i

skolan är berättande, beskrivande, förklarande, utredande, instruerande, argumenterande och diskuterande (Gibbons 2016:123–125). I den aktuella studien är genren PM, och huvudtexttypen utredande.

Per Holmberg, inspirerad av den systemisk-funktionella lingvistik (2006), visar i en studie att elevtexter i gymnasiet ofta består av flera *textaktiviteter*, ett begrepp som ligger nära texttyp och syftar på grundläggande, lingvistiskt specificerade framställningsformer. Ett exempel är genren recension, som kan innehålla både berättande, beskrivande och argumenterande textaktiviteter. Holmberg (2009) använder också termen basgenre för textaktivitet och ansluter sig då till den terminologi som används av Gibbons.

I Australien och på Nya Zeeland har den så kallade genrepdagogen fått stort genomslag. Den har spridits till många andra länder, bland annat Sverige. Gibbons (2016) beskriver den i en lärobok som undervisning om både innehåll och form, i funktionella sammanhang. Undervisningen sker i fyra steg, i en cykel för undervisning och lärande. Steg 1 (*bygga upp fältet*) innebär att man skaffar sig faktakunskaper om det ämne som texten ska handla om. I steg 2 (*modellera eller dekonstruera genren*) visar läraren hur texter inom den aktuella genren är uppbyggda. I steg tre (*gemensam konstruktion*) skriver eleverna tillsammans med läraren en text om det valda ämnet inom den aktuella genren. I steg 4 (*individuell konstruktion*) skriver slutligen eleverna egna texter. Det är vanligt med olika typer av stöd i skrivandet (till exempel skrivmallar, självskattningsscheman), respons och formativ bedömning (Gibbons 2016:127–151).

Kjell Lars Berge, Lars Sigfred Evensen och Ragnar Thygesen (2016) presenterar den sociokulturella och funktionella pedagogiska modellen för skrivande i skolan som under lång tid utvecklats inom den norska skrivforskningen, det så kallade skrivhjulet. Skrivandet förstås i denna modell som en aktivitet; när vi skriver yttrar vi oss genom olika typer av handlingar. Vi har ett syfte med skrivandet, och vi använder oss av olika semiotiska verktyg när vi skriver.

Berge m.fl. (2016) ser skrivandet i skolan som ett verktyg för lärande, reflektion, kunskapsutveckling och interaktion. Interaktionen kan ske med andra, men den kan också innebära att en individ använder skrivandet som ett sätt att kommunicera med sig själv. Berge m.fl. (2016) menar vidare att skrivandet sker i två kontexter, å ena sidan situationskontexten (skolan), å andra sidan i den vidare kulturella kontexten. Särskilt yngre elever har dock begränsad kunskap om den kulturella kontexten, varför ett betoning av denna riskerar att leda till att elevernas texter ses som misslyckade, istället för som ett led i skrivutvecklingen.

I skrivhjulet (Berge m.fl. 2016:9) ingår sex olika syften med skrivandet vilka i typfallet korresponderar med varsin skrivhandling. Dessa är, i min översättning, följande:

1. Syfte: att skapa textvärldar (fantasi, teorier). Handling: att föreställa sig (berätta, skapa, teoretisera).

2. Syfte: att övertala. Handling: att övertyga, uttrycka åsikter, argumentera, diskutera, rekommendera.
3. Syfte: att utbyta information via dialog. Handling: att interagera, samarbeta.
4. Syfte: att forma sin identitet, självutvärdering och metakommunikation. Handling: att reflektera över sina erfarenheter, tankar, känslor och sitt arbete.
5. Syfte: att organisera och lagra kunskap. Handling: att beskriva, organisera och strukturera innehåll.
6. Syfte: att utveckla kunskap. Handling: att utforska, undersöka, jämföra, analysera, diskutera, tolka, förklara, resonera.

Skrivandet i nationella provet i kurs 3 beskrivs som utredande, och eleverna ska skriva texter av så kallad vetenskaplig karaktär (se avsnitt 4.1). Det finns även ett element av att ta ställning. Sammantaget kan skrivandet därmed ses som en kombination av punkt 2, 5 och 6 ovan.

Berge m.fl. (2016) betonar att de ser skrivandet som situationellt, och att något genrebegrepp eller genrepedagogik därmed inte ingår i modellen. De avfärdar genreskrivande med motiveringen att genrer ändå tenderar att bli på låtsas i skolan. Fokus ligger inte heller på den språkliga eller textuella formen. Berge m.fl. placerar bedömningsfokus på om texten kan betraktas som en relevant representation av den aktuella skrivhandlingen och dess syfte.

Ett annat centralt begrepp inom skrivforskningen är *diskurs*. Anne Palmér (2013) analyserar vilka skrivdidaktiska diskurser som kan sägas ingå i de tidigare nationella proven på gymnasiet (Lpf 94, Svenska B) respektive de nuvarande (Gyl1, svenska 1 och svenska 3). Palmér utgår från Ivaničs definition av skrivdidaktisk diskurs enligt vilken begreppet omfattar ”värderingar, uppfattningar och praktiker (handlingsmönster) som leder till ett visst agerande, vissa beslut, val och bortval, liksom till användning av särskilda språkliga uttryck och begrepp” (Palmér 2013:7). Ivanič urskiljer sex olika skrivdidaktiska diskurser, vilka i Palmérs översättning kallas för färdighetsdiskurs, kreativitetsdiskurs, processdiskurs, genrediskurs, diskurs för sociala praktiker samt sociopolitisk diskurs. Palmér analyserar kursplaner, provmaterial och forskningsdokumentation kring proven. Hon drar slutsatsen att alla tre skrivproven (Svenska B, svenska 1, svenska 3) kan betecknas som skrivdidaktiska hybrider. De innehåller inslag av flera diskurser. Alla tre innehåller spår av processdiskursen. Genrediskursen har blivit starkare i de senare proven. Starkast är dock den funktionella synen på skrivande, vilken enligt Palmér (2013:58) kan ses som ”en variant i utkanten av diskursen om sociala praktiker”. Tyngdpunkten i proven ligger på ett skrivande som är funktionellt och kommunikativt inriktat. Kommunikationssituationerna i proven är visserligen fiktiva, men det är fortfarande funktionellt, samhällsinriktat och kommunikativt skrivande som utgör inriktningen. Denna inriktning i proven överensstämmer med kursplanernas formuleringar kring kommunikation och skrivande, i både Lpf 94 och Gyl1. Det antal genrer som förekommer bland

skrivuppgifterna har dock smalnats av. Svenska 3-provet innehåller endast en genre: pm på en universitetskurs.

Palmér nämner också att hon har uteslutit färdighetsdiskursen ur analysen av provens diskurser. Visserligen ingår bedömning av till exempel skriftspråksnormer i proven, men inte på det dekontextualiserade sätt som vanligen avses inom färdighetsdiskursen.

I läroboken *Bedömning av elevtext* (2015) presenterar och diskuterar Palmér och Östlund-Stjärnegårdh den bedömningsmodell som används i de nationella skrivproven. Här lyfter de fram det kommunikativa skrivandet (skrivande riktat till andra) såsom åtskilt från det kognitiva skrivandet (att skriva för att tänka och lära). Som exempel på kognitivt skrivande ges att föra anteckningar, skriva loggbok och att göra ett första utkast inför kommunikativa texter. Det kognitiva skrivandet resulterar enligt Palmér och Östlund-Stjärnegårdh inte i färdiga texter.

2.3.2 Skrivsituationer och skrivkontexter

Det är viktigt att betona att läsande och skrivande sker i ett sammanhang, vilket begreppet *literacy* (sv. litteracitet) handlar om. Liberg och Säljö (2014) definierar litteracitet som bruk av skrift och skriftspråkliga uttrycksformer (jfr skriftspråklighet, läs- och skrivförmåga). Literacypraktiker är verksamheter där skriftspråk (även bilder etc.), läsande och skrivande på olika sätt ingår i meningsskapande. Praktikerna förändras över tid och ser olika ut i olika sammanhang, till exempel i skolans olika ämnen. Liberg och Säljö lyfter fram att i mötet mellan texten och deltagarna inverkar många faktorer såsom å ena sidan deltagarnas erfarenhet av liknande texter, deras intressen, ämneskunskaper, språkliga och metakognitiva kompetens och å andra sidan textens medium och modalitet, innehåll och kulturella bas, deltagaraktivitet, genrestruktur, språk med mera (Liberg & Säljö 2014:357–378).

Något som också diskuteras inom skrivforskning med fokus på skolan är hur fiktiva skrivsituationer, vilket det oftast är fråga om i skolan, inverkar på elevernas texter. Kjell Lars Berge (1988) talar om ”påtvungen penn”, Ulf Teleman (1991) om ”låtsasgenrer”. Östlund-Stjärnegårdh (2006) menar å andra sidan att elever måste få öva sig på att skriva inom olika genrer i skolan, och att det så kallade låtsasskrivandet inte utgör något större problem – det är inte märkligare än att man måste övningsköra innan man kan köra bil på egen hand.

Man kan notera en viss motsättning här. Å ena sidan har det under senare decennier i den svenska skolan betraktats som positivt att låta eleverna föreställa sig tydliga kommunikativa situationer när de skriver olika texter, såsom när de i nationella provet i kurs 3 ombeds föreställa sig att de precis har börjat på universitetet och ska skriva ett PM med kurskamrater och lärare som mottagare. Den kommunikativa situationen används här som ett sätt att betona skrivandets syfte och tydliggöra vilka krav som uppgiften ställer eleven inför (Palmér 2013:59–60).

Å andra sidan blir skrivandet kanske än mer fiktivt av sådana påhittade skriv-situationer, jämfört med om eleverna helt enkelt hade ombetts att skriva en utredande text i största allmänhet om det aktuella ämnet.

Melander (1997) lyfter fram att skrivuppgifter i skolan, inklusive i nationella provet, förutsätter en betydande rollspelsförmåga hos eleven. Eleven ska dels kunna skriva en text med fiktiv mottagare, fiktivt syfte och fiktivt sammanhang, samtidigt som eleven samtidigt har en verklig mottagare (läraren) etc. i skolan. Melander konstaterar att elever är olika bra på detta rollspel och att elevens texterfarenheter utanför skolan, vilka ofta har ett samband med socioekonomiska faktorer, inverkar. Donald Broady (1981, 2007) talar om skolans dolda läroplan, det vill säga om de implicita förväntningar på elever som inte direkt uttrycks i styrdokument eller i klassrummet, men som elever ändå bedöms utifrån. Broady menar att elever från medelklassen generellt har lättare att avkoda dessa förväntningar än elever ur arbetarklassen; de sistnämnda missgynnas därmed av förekomsten av en dold läroplan i skolan.

Hammarbäck (1989) följer en elev på tvåårigt yrkesförberedande programs skrivande under gymnasietiden. Hammarbäck lyfter fram att elevens skrivförmåga dels påverkas av elevens egen syn på denna, dels av den så kallade göra-kultur som många av dessa elever är vana vid hemifrån och inom sina blivande yrken. Göra-kulturen ställs emot skolans och medelklassens tala-kultur. Enligt göra-kulturen är det bland annat logiskt att så fort som möjligt försöka avsluta en text. Hammarbäck återger en episod där eleverna ombeds hitta på en fortsättning på berättelsen och en verkstads elev nöjer sig med att skriva: "Och så åkte alla hem" (Hammarbäck 1989:147). Den elev som Hammarbäck följer gör under de två gymnasieåren stora framsteg i sitt skrivande, men har svårigheter att skriva kontextoberoende, det vill säga att kunna se sin text med en okänd läsares ögon. Eleven har också svårt att hitta en övergripande, välstrukturerad disposition för sin text.

I vilken mån kan elever lära sig att skriva akademiska texter på gymnasiet? Sofia Ask (2005) ger i en studie av studieövergången från gymnasiet till högskolan uttryck för följande uppfattning:

Akademiskt skrivande som kunskapsproduktion kan knappast tränas på gymnasieskolan. Däremot kan skrivande i form av korrekt, offentligt språk med god textbindning och fungerande disposition vara fruktbart. Övning i att vara saklig och att använda egna ord i förhållande till en källa är också viktigt, likaså att träna upp ett kritiskt-analytiskt tänkande. (Ask 2005:111)

Asks uppfattning kan sägas överensstämma med den inriktning som det aktuella nationella provet i kursen svenska 3 har. Det är i detta prov inte fråga om kunskapsproduktion, utan det handlar mer om en övning i den akademiska texttypen.

Anna-Malin Karlsson (1997) jämför i en studie 50 autentiska insändare till *Kamratposten* med 50 fiktiva insändare skrivna av elever i skolan. Samtliga är svar på en insändare publicerad i *Kamratposten*. Resultaten visar att skoltexterna

är mycket mer implicita och kontextberoende än de autentiska insändarna, det vill säga att de med Larssons (1980, 1981, 1984, 1989) terminologi skulle kunna kallas för satellittexter, medan de autentiska insändarna i högre grad fungerar som egentexter. I sin analys tittar Karlsson på situationsmarkörer, refererande element och makrostrukturer i texterna.

Blåsjö (2004) studerar studenters akademiska skrivande i två olika akademiska kontexter: historia och ekonomi. Blåsjö drar slutsatsen att kunskapssynen och synen på hur resonemang ska föras skiljer sig väsentligt åt mellan ämnena, vilket också får konsekvenser för vilka typer av akademiska texter studenterna förväntas skriva. Ask (2007) undersöker lärarstudenters akademiska skriftspråkskompetens i början och slutet av utbildningen. De studenter som befinner sig i slutet av utbildningen har lärt sig att använda ytspråkliga drag som karaktäriserar akademiska texter, men har fortfarande svårigheter att förstå det akademiska skrivandet på djupet (kritiskt-analytiskt tänkande, att förhålla sig till tidigare forskning). Både Blåsjö (2004) och Ask (2007) pekar på att nya studenter möts av implicita förväntningar som bland annat rör det akademiska skrivandet. Förväntningarna är självklara för lärarna, men inte för studenterna. Även Lea och Street (1998) lyfter fram samma bild, ur ett engelskt perspektiv.

2.3.3 Bedömning, textnormer och svenskämneskonceptioner

I studien *Säker stil eller rätt i sak?* analyserar Maria Ohlsson (2001) en skrivuppgift från nationella provet i svenska på gymnasiet där eleverna ska analysera resultatet av en fiktiv enkätundersökning om trivsel i ett bostadsområde, samt föreslå hur problemen i området ska lösas. Ohlssons analys av elevtexterna visar att eleverna överlag är bra på att förstå och analysera enkätundersökningen i sig, och att de även för resonemang om statistisk reliabilitet (det vill säga är källkritiska), men att de är olika bra på att skriva själva texten. Det är alltså inte primärt innehållet som skiljer mellan olika betyg, utan själva textproduktionen. Delbetygen på områdena *textens funktion* och *textens språk och stil* visar hög korrelation med sammanfattningsbetyget, medan delbetyget för *redovisning av enkäten* visar betydligt lägre korrelation. Ohlsson drar slutsatsen ”att det i termer av betyg är viktigare för eleven att producera en text med en god behandling av språk och stil än att producera en text med en genomtänkt analys av enkäten” (Ohlsson 2001:34–35). Det visar sig till och med att elever som fått höga betyg kommit undan med tillägg av påhittade enkätuppgifter; Ohlssons tolkning är att uppgifterna förmodligen lagts till av eleverna för att göra texten fylligare och skapa mer flyt.

Eva Östlund-Stjärnegårdh (2006) analyserar i studien *Att förmedla egna och andras tankar* hur gymnasieelever klarar källhanteringen i nationella provet, och hur detta återspeglas i betyget. Elevtexterna är skrivna år 2000, och provet ingick i Svenska B-kursen. Uppgiften gick ut på att argumentera för barnböcker och barns läsning. Resultaten visar att elevtexter som saknar källhänvisningar får låga

betyg (IG och G), även om innehållet har hämtats från stimulanstexterna. I andra änden av betygsskalan (VG och MVG) visar det sig att dessa elever gör många källhänvisningar, men att deras texter också innehåller fler så kallade stölder från stimulanstexterna (källanvändning utan källhänvisning) än de elevtexter som fått lägre betyg. Östlund-Stjärnegårdh gör tolkningen att stölderna uppfattas som godtagbara av lärarna, om eleven också har med källhänvisningar i sin text och därmed har visat sig behärska hantverket. Stölderna bidrar till att skapa en rik och fyllig text. Majoriteten av eleverna använder sig av fakta och åsikter från endast två stimulanstexter. Det är mycket sällsynt att de uppvisar ett kritiskt förhållningssätt gentemot stimulanstexterna – till skillnad från i Ohlssons (2001) studie (se ovan).

Eric Borgström (2014) ställer i sin avhandling *Skrivbedömning: Uppgifter, texter och bedömningsanvisningar i svenskämnets nationella prov* frågor om hur provuppgifters utformning påverkar elevtexternas så kallade globala struktur, och hur bedömningsanvisningar inverkar på konstituerandet av textkvalitet, samt hur definitionen av skrivförmåga påverkar lärares bedömning av elevernas texter. Avhandlingen innehåller fyra separata studier, varav tre rör de nationella proven på gymnasiet (Svenska B-kursen). Bland annat diskuterar Borgström det faktum att elevtexternas form, genom uppgiftsformuleringar och bedömningsmatriser, är hårt styrd, och formulerar följande fundering: ”[m]en om det i begreppet skrivande ingår att självständigt fatta beslut om hur en situation bör hanteras genom texten, och vilka textstrukturella strategier som gör sig bäst givet syfte och målgrupp, borde uppgifter och matriser skapa ett sådant utrymme för eleverna [...]” (Borgström 2014:80). Teoretiskt bygger Borgström bland annat på Berges (1988) syn på provskrivande som en typ av social handling medierad genom text. Skrivandet av provtexten anses bestå av tre handlingar med varsitt mål: 1) en strategisk handling med målsättningen att leverera en bedömningsprodukt, 2) en rituell handling, med målsättningen att vara en del i en gemenskap, och 3) en kommunikationshandling med målsättningen att skapa interpersonell betydelse (Borgström 2014:84–85).

Margareta Westman (1980) diskuterar redan i *Elevsvenska skolans textnormer*. Hon säger följande:

Den textform som betonas på bekostnad av andra i skolan är och har väl rätt länge varit den utredande, generaliserande uppsatsen om ett neutralt eller lagom kontroversiellt ämne. [...] Att just den utredande uppsatsens texttyp har ansetts som särskilt värdefull i skolans skrivträning kan hänga ihop med att den är bedömningsbar just som form. Eftersom vi i skolan har velat eller känt oss tvungna att bedöma språkförmågan isolerat, alltså bortsett från vad texten handlar om, så har skolan utvecklat en egen texttyp, vars betydelse bärs av själva språkbehandlingen, av syftet att visa att man kan behandla språket. (Westman 1980:109)

Westman önskar att bedömningen av utredande uppsatser skulle utgå från både innehåll och form. Hon menar dock att det är en texttyp som egentligen inte passar barn och ungdomar särskilt väl, eftersom den förutsätter mognad och erfarenhet.

Eleverna skulle förmodligen lyckas bättre inom texttyper som tjänar på fantasi, lekfullhet med mera. Westman påpekar också att man som skribent måste lära sig att behärska olika textmönster innan man kan börja utnyttja och kanske förändra formen. Enligt en tidigare studie av Sandell (1972), som Westman refererar, skriver elever av eller citerar ganska mycket ur de så kallade stimulanstexter som hör till centrala provet i svenska på gymnasiet. Här menar alltså Westman att dessa avskrifter skulle kunna ses som ett naturligt led i att eleverna håller på att lära sig en ny texttyp.

Åke Pettersson (1997) lyfter fram att textkvalitet och skrivförmåga inte alltid korrelerar. Pettersson menar att skolans skrivundervisning bör syfta till att utveckla elevernas skrivförmåga snarare än till att eleverna ska skriva så bra texter som möjligt. Skrivförmågan är dock enligt Pettersson svårdefinierad. Den kanske visar sig först när eleven har lämnat skolan, och då kan andra influenser ha tillkommit. Den kanske inte kan utläsas direkt ur texternas kvalitet, eftersom dessa kan ha tillkommit under olika förutsättningar. Och, slutligen, skrivförmåga är något oerhört komplext:

Den [elevernas skrivförmåga] omfattar en kaotisk komplexitet med låt oss säga hundratal komponenter som ingen kan överblicka. Och språkdragen samspelar: Ofta måste nämligen i skrivutvecklingen förtjänster eller framsteg på ett område betalas med brister eller reträtter på ett annat; så är det med språklig korrekthet och originalitet. Av sådana och andra skäl skiljer sig åsikterna om god skrivförmåga åt ganska avsevärt bland gymnasiets svensklärare [...]. (Pettersson 1997:160)

Vad som ska ingå i ett ämne och hur innehållet ska hanteras är inte självklart. Inom svenskämnet har olika ämnessyner, eller ämneskonceptioner, avlöst varandra. Bergöö (2005) beskriver hur svenskämnet under olika tidsperioder har betraktats som ett färdighetsämne, som ett kulturarvsämne eller litteraturhistoriskt bildningsämne, som ett erfarenhetspedagogiskt ämne och som ett demokratiämne. I min egen studie (Bergström 2007), med titeln *Två olika ämnen? Svenska språket på gymnasiet och på högskolan*, som bygger på enkätundersökningar med gymnasieelever, gymnasielärare, studenter och högskolelärare, visar resultaten bland annat att svenskämnet på gymnasiet – förutom att det också inkluderar litteratur – är mycket kommunikativt inriktat. Elevens språkutveckling och kommande förmåga att använda skrivandet i olika samhälleliga situationer står i fokus. På högskolan fyller istället kunskaper om svenska språket och språkanalytiska inslag ett stort utrymme i ämnet. Det är rimligt att tänka sig att sådana skilda synsätt även inverkar på bedömningen av gymnasieelevers respektive studenter språkbehandling i akademiska texter i ämnet svenska. Studien visar också att både elever och studenter i stor utsträckning antar och håller med om den bild av ämnet som presenteras för dem i undervisningen.

3 Syfte och frågeställningar

Kursen svenska 3 på gymnasiet syftar till att förbereda eleverna inför det akademiska skrivandet på högskolan. Kursen är obligatorisk på högskoleförberedande program och valbar på yrkesförberedande program. En central del av kursens examination utgörs av det nationella prov som ingår i kursen. I provet ingår en muntlig och en skriftlig uppgift. Den skriftliga uppgiften innebär att eleven ska skriva ett så kallat PM med frågeställning, diskussion och slutsats. Ämnet anknyter till antingen språk eller litteratur, med ett specifikt tema. Eleven ska hänvisa till så kallade källtexter som behandlar temat. Den text som eleven skriver ska vara av så kallad vetenskaplig karaktär.

Det akademiska skrivandet är en stor och viktig del av högskolestudier. I allt från mindre skrivuppgifter till avslutande examensarbeten behöver studenterna kunna bemästra detta skrivande för att klara av sina studier. Ur högskolans perspektiv är det därför angeläget att känna till så mycket som möjligt om nya studenters förkunskaper i akademiskt skrivande, för att kunna erbjuda en adekvat undervisning på detta område. Kursen svenska 3 infördes med läroplanen Gy11. Det nationella prov som ingår i kursen genomfördes första gången vårterminen 2014. Forskning på elevtexter från detta prov är än så länge begränsad. Det är angeläget med studier som ur olika infallsvinklar undersöker elevernas prestationer på provet. Sådan kunskap behövs i samband provkonstruktionen, till exempel i utformningen av lärarinstruktioner och bedömningsanvisningar.

Ordförrådet är en central del när det kommer till sakprosatexter, inklusive vetenskapliga texter. Centrala begrepp, övriga ämnesord samt akademiska ord och uttryck bär upp och strukturerar innehållet i vetenskapliga texter. Hur de används avgör till stor del den vetenskapliga textens tydlighet och kvalitet. Många synpunkter om studenters eventuellt bristande förmåga till akademiskt skrivande har handlat om brister i just ordanvändningen (se till exempel Blücker 2006).

Studien syftar till att undersöka, analysera och diskutera gymnasieelevers ordanvändning i det nationella prov som ingår i kursen svenska 3, samt att sätta in denna ordanvändning i ett vidare språkligt sammanhang med syntaktiska variabler och olika jämförelsematerial.

Studien utgår från följande frågeställningar:

1. Vad utmärker på en övergripande nivå elevernas ordanvändning och syntax, och hur förhåller sig dessa till texterna i provets texthäfte respektive till autentiska vetenskapliga artiklar? Finns det uppgifts- eller betygsrelaterade skillnader mellan elevtexterna med avseende på ordanvändning och syntax?
2. Hur ser förekomsten av ämnesord samt akademiska ord och uttryck ut i elevtexter med betyg A respektive E? Hur förhåller sig, på dessa områden, elevtexterna till texterna i texthäftet samt till autentiska vetenskapliga artiklar?

Frågeställning 1 besvaras genom en automatisk analys med den så kallade SVIT-modellen (Heimann Mühlenbock 2013), vilken både inkluderar mått som förekommer i tidigare elevspråksforskning, såsom ordlängd och ordvariationsindex, såväl som nyare mått, såsom traddjup och lemmavariationsindex. Frågeställning 2 besvaras genom en manuell analys av ord och uttryck i olika kategorier, vilka bygger på Järborgs (2007) indelning, vidareutvecklad för att passa den aktuella studiens inriktning. Metoderna presenteras ingående i kapitel 4.

Som framgår av frågeställningarna har studien ett tydligt språkligt fokus, och följer därmed den språkligt inriktade elevspråksforskningstradition som sedan några decennier tillbaka är etablerad vid Institutionen för nordiska språk vid Uppsala universitet. Den ingår också i den forskning om automatiska analysers användbarhet vid bedömning som växt fram under senare tid. Studiens resultat diskuteras också ur ett vidare perspektiv, utifrån faktorer som normer, skrivsituation, genrer, bedömning samt förhållandet mellan språk och tänkande.

4 Material och metoder

I kapitlet presenteras det nationella prov i kursen svenska 3 på gymnasiet som studien behandlar. Det primära källmaterialet (elevtexter) beskrivs, liksom de jämförelsematerial som används inom studien. Därefter presenteras och kommenteras de metoder för automatisk analys och de manuella ordanalysmetoder som har använts på materialet.

4.1 Det nationella provet i kurs 3

I den svenska grund- och gymnasieskolan ges ett antal nationella prov i utvalda ämnen. Avsikten med de nationella proven är att de ska tjäna som ett bedömningsstöd för läraren samt hjälpa till att upprätthålla en nationell likvärdighet mellan skolor över hela landet. Tanken är inte att proven ensamma ska avgöra elevens betyg, i de årskurser där betyg ges. I ämnena svenska och svenska som andraspråk ges nationella prov på grundskolan i årskurs 3, 6 och 9, och på gymnasiet i kurs 1 och i kurs 3. Dessa gymnasiekurser infördes med läroplanen Gy11. De föregicks av kurserna Svenska A, B och C i Lpf 94, varav endast Svenska B inkluderade ett nationellt prov.

De nationella proven i svenska och svenska som andraspråk har delvis olika inriktningar beroende på i vilken årskurs eller kurs de ges. Det prov som står i fokus i den aktuella studien ges inom kurserna svenska 3 och svenska som andraspråk 3 och har en inriktning mot akademiskt skrivande, eller ”skrivande av vetenskaplig karaktär”, vilket är den formulering som används av Skolverket och i provsammanhanget. Det akademiska skrivandet är en viktig del av kurs 3, förstås med tanken att kursen ska förbereda för högskolestudier. Kursen är obligatorisk på högskoleförberedande program och valbar på yrkesförberedande program.

Varje nationellt prov har ett specifikt tema. Temat hämtas från svenskämnet och är antingen språkligt eller litterärt inriktat. Det prov som används i studien har ett språkligt tema om skriftspråkliga normer, med rubriken *Hur skriver man?* Det gavs vårterminen 2014 och var det första nationella prov som skrevs av en hel årskurs elever inom kurserna svenska 3 och svenska som andraspråk 3. Det skiljer sig med sin inriktning mot akademiskt skrivande en del från det nationella prov som tidigare gavs inom kurserna Svenska B och Svenska som andraspråk B på gymnasiet. (För en mer utförlig presentation, analys av och jämförelse mellan dessa prov, se Palmér 2013.) Att ett nytt nationellt prov har införts gör det förstås extra angeläget med forskning på detta prov, vilket alltså denna studie är tänkt att bidra med.

Det nationella provet inom kurs 3 förbereds och genomförs under några veckors tid, huvudsakligen inom ramen för den ordinarie undervisningen. Läraren förbereder eleverna och de får ett häfte med texter som behandlar provets aktuella tema som de kan läsa igenom och arbeta med på olika sätt. Provet består sedan av

två delprov: delprov A, som innebär att eleven ska skriva ett så kallat pm baserat på texter ur texthäftet, samt delprov B, vilket innebär att eleven ska hålla ett muntligt anförande med anknytning till temat inför klassen. Delprov A, vilket är det prov som undersöks i denna studie, har en skrivtid som är begränsad till 4 timmar och genomförs vid ett gemensamt skrivtillfälle i skolan. Vid skrivtillfället får eleven välja en uppgift att skriva om, med anknytning till temat. I provmaterialet ingår också ett häfte med lärarinformation och ett häfte med bedömningsanvisningar. Läraren bedömer delproven utifrån så kallade bedömningsmatriser. Bedömningsmatrisen för delprov A innehåller aspekterna *Innehåll och källhantering*, *Disposition och sammanhang*, *Språk och stil* samt *Helhetsbedömning*. Varje aspekt innehåller olika punkter som talar om vad läraren ska fokusera på i sin bedömning. Läraren har även tillgång till redan bedömda elevtexter, med utförliga kommentarer från provkonstruktörerna. Dessa texter har skrivits av elever och bedömts av verksamma lärare under de utprovningar som görs inför varje nytt prov. Provet är detsamma för elever som läser svenska respektive svenska som andraspråk, men punkterna i bedömningsmatrisen skiljer sig åt i viss mån mellan dessa ämnen.

Provet's uppgiftsinstruktioner är relativt omfattande. I studien sammanfattar jag dem och återger för studien relevanta delar. Citat ur elevtexterna ges för att exemplifiera resultat, men fullständiga elevtexter återges inte av utrymmes- och integritetsskäl. Sekretess råder på allt nyare provmaterial. För det aktuella provet hävdades sekretessen 1 juli 2020.

4.2 Studiens källmaterial: elevtexter och jämförelsematerial

I avsnittet presenteras studiens källmaterial närmare, dels det primära källmaterialet, elevtexterna från nationella provet i kursen svenska 3 vårterminen 2014 (avsnitt 4.2.1), dels det sekundära källmaterial som har använts som jämförelsematerial till elevtexterna, och som utgörs av de texter som eleverna refererar i provet samt autentiska språkvetenskapliga artiklar (avsnitt 4.2.2). I tabell 1 ges en översikt över de texter som analyseras i studien.

Tabell 1. Antal texter som ingår i studien.

	Elevtexter, uppgiften <i>Skrivregler och sms</i>	Elevtexter, uppgiften <i>Särskrivning och stavfel</i>	Käll- texter	Språkvetenskapliga artiklar
Antal texter	60	30	14	13

4.2.1 Elevtexter

I de följande avsnitten beskrivs studiens primära källmaterial, elevtexterna, med fokus på urvalet av texter utifrån olika variabler. Redan kända fakta om betygsutfall med mera när provet gavs vårterminerna 2014 och 2015 presenteras. Avslutningsvis beskrivs provets bedömningsmatris, med fokus på för studien relevanta punkter.

4.2.1.1 Uppgifter och betyg

Studiens primära källmaterial är elevtexter från kursprovet vårterminen 2014, delprov A, kursen svenska 3. Varje gång ett nationellt prov i svenska och svenska som andraspråk ges skickas ett urval elevtexter in till Institutionen för nordiska språk vid Uppsala universitet, där de arkiveras. Ur detta arkiv har sedan elevtexter digitaliserats och kommit att ingå i Uppsala elevtextkorpus (Megyesi, Näsman & Palmér 2016), som är Gruppen för nationella prov i svenska och svenska som andraspråks särskilda databank för forskning om elevtexter. Elevtexterna skickas in i pappersform och har sedan överförts till digital form genom skanning. Handskrivna elevtexter har transkriberats och därefter lagts in i korpusen.

Det är alltså ur Uppsala elevtextkorpus som studiens material har hämtats. Jag har valt ut de två uppgifter som flest elever valde att skriva om vårterminen 2014. Valet grundar sig framför allt på att tillgången på elevtexter är större för dessa uppgifter. Det kan också vara intressant att titta närmare på dessa uppgifter just för att många elever ville skriva om dem, kanske för att de upplevdes som intressanta, aktuella eller lätta. Samtliga elevtexter som analyseras var från början skrivna på dator.

De två uppgifterna, med antal elevtexter med olika betyg, är

- *Skrivregler och sms*, totalt 60 elevtexter (betyg A: 20 stycken, betyg C: 20 stycken, betyg E: 20 stycken)
- *Särskrivning och stavfel*, totalt 30 elevtexter (betyg A: 10 stycken, betyg C: 10 stycken, betyg E: 10 stycken).

Antalet elevtexter som analyseras har styrts av texttillgången i Uppsala elevtextkorpus. Jag valde att analysera så många elevtexter som möjligt upp till närmaste tiotal. Det fanns alltså fler elevtexter att tillgå för uppgiften *Skrivregler och sms*, som var den uppgift som flest elever valde att skriva om, än för den näst populäraste uppgiften *Särskrivning och stavfel*. Jag ville analysera samma antal elevtexter i olika betygssteg, för att underlätta den kvantitativa jämförelsen mellan textgrupperna.

Betygsskalan går från A till F, där F är underkänt. Jag har valt att analysera elevtexter med betygen A, C och E eftersom det är dessa betyg som det finns bedömningskriterier för i matrisen, vilket möjliggör att resultaten även kan diskuteras i förhållande till dessa kriterier. Värt att notera är att vårterminen 2014

ombads den betygsättande läraren inte att redogöra för betygen för de olika aspekterna vid inskickningen av elevtexter. Betyget är alltså ett sammanfattningsbetyg (så kallad helhetsbedömning) på hela elevtexten, vilket innebär att det inte går att veta exakt hur varje aspekt har bedömts. Enligt bedömningsmodellen har dock en elevtext som bedömts med A, C eller E som helhetsbetyg minst detta betyg på de tre aspekterna i bedömningsmodellen.

Bedömning är ingen helt exakt vetenskap; en annan lärare hade kanske kommit fram till ett annat betyg på samma elevtext. Ett möjligt sätt att komma ifrån detta problem är att låta ett antal bedömare göra nya bedömningar av samma elevtexter, och att utgå från ett genomsnitt i dessa nya bedömningar. Så gör till exempel Hultman och Westman (1977) i *Gymnasistsvenska*, men de påpekar ändå att de inte helt har kunnat undgå problematiken med bedömningsvariation kopplad till bedömare. Inom ramen för den här studien har det inte varit möjligt att låta göra nya bedömningar av elevtexterna. Min förhoppning är att det relativt stora antalet elevtexter ändå ska göra det meningsfullt att resonera om skillnader mellan betyg, åtminstone i form av tendenser.

4.2.1.2 Övriga urvalsvariabler

Vårterminen 2014 skickades 979 elevtexter in, för elever födda ett visst datum varje månad under hela året. 56 procent av dessa elevtexter är skrivna av kvinnor, 44 procent av män. Endast 4 procent av de inskickade elevtexterna är skrivna av elever som läser svenska som andraspråk. 83 procent av elevtexterna är skrivna av studenter på högskoleförberedande program. 62 procent av de inskickade elevtexterna är skrivna på dator, och resten för hand (Broman & Dalberg 2014).

Jag har valt att i denna studie undersöka elevtexter skrivna på dator, primärt inom ämnet svenska. Jag har valt bort variabler som kön och program. Valet grundar sig dels på tillgången på elevtexter – antalet elevtexter skulle helt enkelt, som framgår av procentsatserna i föregående stycke, bli för litet i textgrupperna om ytterligare variabler inkluderades för att tillförlitliga slutsatser skulle kunna dras – dels på studiens fokus. Studiens fokus ligger på hur väl gymnasieelever som grupp bemästrar det ”skrivande av vetenskaplig karaktär” som tränas under kursen svenska 3 och testas i det nationella prov som undersöks, samt på eventuella skillnader kopplade till betyg i hur väl de klarar av detta skrivande. Betygsstegen utgör en ingång till att analysera och diskutera dessa nivåer. Naturligtvis vore det intressant att inkludera ytterligare variabler, inte minst en jämförelse mellan elevtexter skrivna inom svenska respektive svenska som andraspråk, men på grund av texttillgången och studiens fokus och omfattning görs inte detta. Materialet är huvudsakligen hämtat från ämnet svenska och med några få undantag från ämnet svenska som andraspråk. Eftersom provuppgiften är densamma för båda svenskämnen och bedömningskriterierna mycket lika, görs ingen skillnad mellan elevtexterna beroende på svenskämne. Däremot hämtas samtliga textexempel ur elevtexter klassificerade med svenska.

4.2.1.3 Mer om NP år 2014 och 2015

Elevernas uppgift under delprov A är att skriva ett så kallat pm med frågeställning, referat av källtexter, diskussion och slutsats(er). Eleven ombeds föreställa sig att hen precis har börjat läsa på universitetet. Det pm som hen skriver ska kunna ge kurskamraterna information och idéer om provets tema, och det ska bedömas av universitetsläraren. Provet utgår alltså från en fiktiv skrivsituation (se avsnitt 2.3.2).

Under det aktuella provet kunde eleverna välja mellan fyra olika uppgifter, varav elevtexter från de två uppgifter som flest elever valde att skriva om ingår i studien. Uppgiften *Skrivregler och sms*, som hela 55 procent av eleverna valde, innebar att eleven skulle skriva om varför skrivregler inte verkar följas på samma sätt i sms som i andra texter. Denna uppgift hade också högst andel underkända, inskickade lösningar, och den var mer populär bland männen än kvinnorna. Uppgiften *Särskrivning och stavfel* var näst populärast bland eleverna – knappt 30 procent valde den, varav något fler kvinnor än män. Här skulle eleven diskutera i vilken mån olika typer av brott mot skrivnormer stör läsningen av en text. Samtliga uppgifter i delprov A innebär att eleven först ska presentera en frågeställning (det vill säga den frågeställning som ges i uppgiftsinstruktionen), därefter referera vad som sägs om det aktuella ämnet i några av texterna i det texthäfte som de har fått inför provet, och slutligen dra slutsatser eller presentera en jämförelse som svar på frågeställningen, baserat på de egna referaten. Det anges i instruktionen att ett lämpligt omfång för elevens text är 600–800 ord.

Som tidigare nämnts inkluderas inte variabler som kön och program i studien. Vissa saker är dock redan kända. Broman och Dalberg (2014) redogör i en resultatrapport för provet för elevernas prestationer inom olika grupper. Eleverna har generellt lyckats bättre på det muntliga provet, delprov B. Här fick 17 procent av eleverna det högsta betyget, A, och endast 5 procent underkänt, F. På det skriftliga delprovet, delprov A, fick endast 8 procent A, och hela 27 procent F. Kvinnorna har som grupp fått högre betyg än männen, på båda delproven. Fler kvinnor än män når också de högsta betygen i delprov A.

Eleverna som läser svenska som andraspråk har som grupp fått lägre betyg än de som läser svenska. Broman och Dalberg påpekar dock att antalet inskickade elevtexter inom svenska som andraspråk är för litet för att säkra slutsatser ska kunna dras.

Betygsskillnaderna mellan datorskrivna och handskrivna elevtexter är små, och det förefaller inte finnas något samband mellan skrivmedium och betyg. I studien har datorskrivna elevtexter valts för att underlätta den automatiska analysen.

Broman och Dalberg konstaterar att baserat på betygsutfallet i delprov A når en ganska stor grupp elever *inte* upp till kunskapskraven för vetenskapligt skrivande. I lärarenkäter framgår att vanliga orsaker till att elever får underkänt på delprov A är att de inte har följt uppgiftsinstruktionen samt brister i referat-tekniken. Lärare ansåg också att instruktionen till uppgiften *Skrivregler och sms* var något otydlig. Broman och Dalberg påpekar slutligen att det faktum att provet

fortfarande var nytt för många av lärarna kan ha påverkat utfallet i negativ riktning (Broman & Dalberg 2014).

I resultatrapporten för motsvarande prov vårterminen 2015, denna gång med litterärt tema, bekräftas bilden från 2014 års prov. Eleverna presterar bättre på det muntliga provet än det skriftliga, kvinnor presterar bättre än män, och elever som läser svenska presterar generellt bättre än elever som läser svenska som andraspråk. I denna resultatrapport jämförs också elever på högskoleförberedande och yrkesförberedande program, och det visar sig att eleverna på yrkesförberedande program generellt får lägre betyg. På delprov A har hela 40 procent av eleverna på yrkesförberedande program blivit underkända. Den totala andelen underkända elever på delprov A är densamma som på 2014 års prov, 27 procent (Broman & Dalberg 2015).

Från vårterminen 2015 redovisar lärarna även betygen för de olika aspekterna som ingår i bedömningen av delprov A. Andelen F är högst för aspekten Innehåll och källhantering (25 procent), vilket överensstämmer med den bild som ges i lärarenkäterna – elevernas svårigheter handlar ofta om svårigheter med att följa instruktionen, att hålla sig till ämnet och svårigheter med källhanteringen. För aspekterna Disposition och sammanhang samt Språk och stil är andelen F densamma, med 12 procent F för båda (Broman & Dalberg 2015).

Lärarna förefaller överlag nöjda med provet, enligt sammanställningen av den lärarenkät som har gjorts av Gruppen för nationella prov i svenska och svenska som andraspråk (2014). Kritik finns förstås också. Bland annat påpekar en del lärare att temat om skrivnormer var oinspirerande för eleverna, och att den tänkta provsituationen inte fullt ut skulle motsvara en autentisk skrivsituation på högskolan.

Som tidigare nämnts var det vårterminen 2014 första gången som en hel årskurs elever skrev det nya kursprovet. Provet modifieras kontinuerligt baserat på de erfarenheter som görs. I väsentliga delar är provet fortfarande (2021) utformat på samma sätt, men detaljer kan ha ändrats. Till exempel har vissa formuleringar i bedömningsmatrisen modifierats.

4.2.1.4 Provets bedömningsmatris

Som tidigare nämnts bedömer lärarna elevernas texter i delprov A utifrån en så kallad bedömningsmatris vilken inkluderar tre olika aspekter samt helhetsbedömning. Inom de olika aspekterna är det vissa bedömningskriterier som särskilt berör min studie. Dessa är följande, med betygsnivån inom parentes:

Innehåll och källhantering

Texten utmärks av objektiv saklighet (A). Texten utmärks i huvudsak av objektiv saklighet (C). Texten utmärks delvis av objektiv saklighet (E).

Källhänvisningarna är väl fungerande med till exempel referatmarkeringar (A–C). Eleven tillämpar regler för citat- och referatteknik (E).

Disposition och sammanhang

Texten är väldisponerad genom väl fungerande inledning, avslutning och textbindning [...] (A–C). Texten har en tydlig disposition med inledning och avslutning samt fungerande textbindning (E).

Språk och stil

Språket är välformulerat med precision i ordval och uttrycksformer (A). Språket är klart och bär upp elevens resonemang (C). Språket följer i huvudsak skriftspråkets normer för språkriktighet. Språket är varierat och innehåller goda formuleringar (E).

Stilen är väl anpassad till den vetenskapliga texttypen (A). Stilen är anpassad till den vetenskapliga texttypen (C). Stilen är till viss del anpassad till den vetenskapliga texttypen (E).

Ovanstående bedömningskriterier har i viss mån påverkat urvalet av ordanalyskategorier (se avsnitt 4.3.2).

4.2.2 Jämförelsematerial

I likhet med Hultman och Westman (1977) har jag valt att ställa elevtexterna mot jämförelsematerial. Dessa utgörs av

- provets texthäfte, totalt 14 texter
- vetenskapliga artiklar från tidskriften *Språk och stil*, åren 2014–2015, totalt 13 artiklar.

Min tanke med jämförelsematerialet är att jag dels vill närma mig frågan varifrån eleverna hämtar sina språkliga förebilder, dels vill jämföra elevernas texter av ”vetenskaplig karaktär” med autentiska vetenskapliga texter. Texterna i provets texthäfte är de texter som eleverna ska utgå från och använda sig av när de skriver sina egna texter. Är det dessa texter eleverna också inspireras av rent språkligt, eller har de andra förebilder? Hur lika eller olika autentiska språkvetenskapliga texter är elevernas texter? Vari består likheterna och skillnaderna?

Texterna i provets texthäfte, hädanefter kallade källtexterna, är 14 till antalet. De utgörs huvudsakligen av populärvetenskapliga artiklar från tidskrifter och dagstidningar och utdrag ur handböcker och uppslagsverk (se tablå 6 i bilaga 2). De är skrivna av forskare (språkvetare) eller vetenskapsjournalister. Omfånget varierar ganska mycket, från omkring 300 till 1500 ord (se tabell 14 i bilaga 1). Alla källtexter är inte relevanta för alla uppgifter. Ofta är omkring 4 källtexter av direkt relevans för en viss uppgift. Det är inget krav att eleverna ska använda sig av samtliga relevanta källtexter; de får göra ett eget urval. Dock ska minst två källtexter användas.

De autentiska vetenskapliga texter som jag också använder som jämförelsematerial utgörs av samtliga 13 forskningsartiklar i *Språk och stil* 2014 och 2015 (se tablå 7 i bilaga 2). Artiklarna är omfattande och varierar mellan cirka 3000 och 9000 ord (se tabell 15 i bilaga 1), efter att tabeller och liknande rensats bort. Recensioner och så kallade smärre bidrag har uteslutits vid analysen av *Språk och*

stil-materialet; endast forskningsartiklar ingår. Jag har valt *Språk och stil* eftersom det är en väl etablerad, svensk språkvetenskaplig tidskrift med artiklar skrivna av verksamma språkforskare. Tidskriften ges ut vid samma institution som nationella provet i svenska görs vid; dock ingår medlemmar också från andra lärosäten i både styrelsen för det sällskap som ger ut tidskriften och i tidskriftsredaktionen. Artiklar valdes framför till exempel avhandlingar eftersom textmängden är mer begränsad, vilket var viktigt med tanke på att texterna behöver gås igenom noga och rensas på allt som inte utgör löpande text inför den automatiska analysen (se vidare avsnitt 4.3.1.2). *Peer-review*-granskade artiklar som det är fråga om här är dessutom vanligen noggrant genomgångna och bearbetade med avseende på sådant som precision och stilnivå, och kan därför antas väl spegla den språkliga normen i vetenskaplig text, i detta fall inom språkvetenskap. Detta gör dem till ett lämpligt jämförelseobjekt i förhållande till elevtexterna i studien.

Naturligtvis är det så att eleverna inte förväntas skriva på samma sätt som etablerade forskare. Likväl kan det vara intressant att jämföra deras texter ”av vetenskaplig karaktär” med riktiga vetenskapliga texter, för att analysera eventuella språkliga skillnader består. Sådan kunskap kan sedan utnyttjas i provmaterialet (anvisningar, bedömning) och i undervisningssituationen. Det är också viktigt att jämföra elevernas texter med vetenskaplig text inom samma ämnesområde (språkvetenskap), eftersom vetenskapligt skrivande till viss del är ämnes-specifikt.

4.3 Undersökningsmetoder

I studien används både automatiska metoder (avsnitt 4.3.1) och manuell analys av ord och uttryck (avsnitt 4.3.2). De kvantitativt inriktade automatiska analyserna syftar till att ge en bred bild av elevtexterna i sig själva och i förhållande till jämförelsematerialet (frågeställning 1). Här ingår variabler med fokus på ordförrådet, men också syntaktiska variabler för att sätta in ordanalyserna i ett bredare språkligt sammanhang. De manuella analyserna, vilka görs på ett mindre antal elevtexter, rör endast ord och uttryck, för att ge en fördjupad bild av elevernas ordanvändning, med fokus på akademiskt skrivande (frågeställning 2).

4.3.1 Automatisk analys med SVIT-modellen

De automatiska analyserna har gjorts med hjälp av den så kallade SVIT-modellen (Heimann Mühlenbock 2013). I avsnittet presenteras modellen övergripande. De analysvariabler som har använts i studien presenteras mer utförligt. Avslutningsvis beskrivs hur de analyserade texterna har förberetts för analyserna, samt vilka statistiska mått som har framtagits i den automatiska analysen.

4.3.1.1 Analysvariabler

SVIT-modellen inkluderar ett tjugotal mått relaterade till ord och syntax. Dels ingår mått som brukar användas vid automatisk textanalys, såsom antal löpord, läsbarhetsindex och ordvariationsindex, vilket möjliggör jämförelser med tidigare studier, dels ingår relativt nya automatiska mått som utgör nya analysmöjligheter, såsom lemmavariationsindex och träddjup.

Totalt har 22 olika variabler analyserats med SVIT-modellen. Jag väljer att redovisa resultaten för 10 av dessa i resultatdelen. Jag har valt ut variabler som är relevanta för ordanvändningen, vilket är studiens primära fokus, samt några variabler som rör grafiska meningar och syntax, för att kunna sätta in ordförrådet i ett vidare språkligt sammanhang, och för att underlätta jämförelser med källtexterna och *Språk och stil*-artiklarna. De variabler som inte redovisas i studien är dels några ytterligare variabler som rör syntaxen, dels några ordförrådsvariabler som inte är relevanta för studien. Variabeln *ord med mer än 13 bokstäver* är i och för sig relevant i studien. Här visade det sig dock att de kvantitativa skillnaderna mellan elevtexterna, med något undantag, var mycket små, varför jag inte bedömde det som relevant att löpande inkludera denna variabel i resultatredovisningen (se också avsnitt 5.3).

De SVIT-mått som redovisas och analyseras i studien är följande:

- *words* = antal löpord
- *mean sentence length in words* (MSL) = antal ord per mening i genomsnitt
- *mean word length in characters* (MWLC) = antal bokstäver per ord i genomsnitt
- ordvariationsindex (OVIX)
- läsbarhetsindex (LIX)
- lemmavariationsindex (LVIX)
- *subordinate classes* (UA) = andel bisatser
- *parse tree depth* (PT) = träddjup
- *noun/pronoun-ratio* (NoPr) = substantiv/pronomen-kvot
- *nominal ratio* (NR) = nominalkvot.

Den stora fördelen med automatisk analys är förstås att den kan ge övergripande, jämförbara värden för stora material. Samtidigt finns ett antal felkällor. Analysen av ordrelaterade mått är överlag säkrare än syntaktiska analyser, eftersom de senare är mer komplexa att göra med automatiska verktyg. Vidare innehåller elevtexter fler språkliga fel än publicerade texter av professionella skribenter, vilket kan påverka analysen på olika sätt. Den automatiska analysen bör kompletteras med manuell analys, vilket alltså görs i studien (se avsnitt 4.3.2).

Nedan presenteras de automatiska mått som ingår i studien mer utförligt, tillsammans med motiveringar, begreppsdefinitioner och resonemang om felkällor.

Löpord: Denna variabel talar om hur många ord en text innehåller. Det är alltså ett mått på textlängden. Ett löpord består av bokstäver och avgränsas av mellanrum

före och efter ordet. Icke-verbala tecken, såsom siffror och skiljetecken, räknas inte som löpord. I elevtexterna innebär särskrivningar en felkälla: ett särskrivet ord räknas som två löpord istället för ett, vilket alltså i viss mån kan dra upp värdet för textlängden. Som felkälla bör den dock vara försumbar: det är trots allt inte så många ord som naturligt går att särskriva i en text. Oftast handlar det om vissa sammansatta substantiv. Textlängden, mätt i löpord, är en variabel som i tidigare forskning har visat sig korrelera med betyg (se avsnitt 2.2.2.1–2.2.2.3). Åtminstone upp till en viss nivå får längre elevtexter högre betyg. I Palmér (2018), en analys av så kallade benchmark-texter från nationella provet i svenska 3, visar sig dock detta samband inte gälla. Eftersom textlängd i många studier har visat sig ha ett samband med betyg har jag likväl valt att inkludera den i den aktuella studien. Variabeln är alltså framför allt av intresse vid jämförelsen av elevtexter som har fått olika betyg.

Genomsnittlig meningslängd, eller antal ord per mening: Detta är ett mått som talar om hur många löpord meningarna i en text innehåller i genomsnitt. Variabeln säger alltså ingenting om variation i meningslängd inom en text. Vidare är det viktigt att påpeka att det är grafiska meningar som den automatiska analysen utgår från. En grafisk mening börjar med stor bokstav och slutar med punkt, frågetecken eller utropstecken. Satsradningar, det vill säga att skribenten använder komma-tecken mellan huvudsatser som skulle kunna utgöra egna grafiska meningar (eller bindas ihop med ett bindeord, kolon eller semikolon), innebär alltså att den genomsnittliga meningslängden ökar. Inom skönlitteratur används ofta satsradningar som ett stildrag. De är mer accepterade inom populärvetenskaplig text än i vetenskaplig, där de helst bör undvikas, eftersom de ofta innebär en bristande innehållslig precision. Elever som satsradar mycket kan alltså enligt den automatiska analysen i denna variabel förefalla skriva längre meningar än andra elever. Sambandet mellan genomsnittlig meningslängd och textkvalitet är förstås komplext. Upp till en viss nivå kan säkert längre meningar kopplas till ett mer avancerat och nyanserat språkbruk. Men inte minst i vetenskaplig text är det också ett kvalitetstecken att kunna skriva koncentrerat med bland annat nominaliseringar, vilket kan innebära kortare meningar. I studien är variabeln av särskilt intresse i jämförelsen mellan elevtexterna, källtexterna och de vetenskapliga artiklarna.

Genomsnittlig ordlängd, eller antal bokstäver per ord: I denna variabel analyseras det genomsnittliga antalet bokstäver per löpord i texterna. Måttet har i tidigare forskning visat sig ha ett tydligt samband med betyg, och även med genre/texttyp. Elever som får högre betyg har ofta en högre genomsnittlig ordlängd, och vetenskaplig text innehåller många långa ord. Många vardagliga ord i språket är korta, medan mer specialiserade ord, dit vetenskapliga termer hör, är längre. Variabeln är därför av intresse både vid jämförelsen mellan elevtexter med olika betyg, och vid jämförelsen mellan elevtexterna och jämförelsematerialet. Felkällorna bör vara försumbara; enstaka särskrivna eller felstavade ord i elevtexterna kan knappast inverka på helhetsbilden.

I SVIT-modellen ingår också en variabel för antalet stavelser per ord. Eftersom detta sätt att mäta ordlängden på visar ungefär samma sak som mätningen av antalet bokstäver per ord, väljer jag att endast redovisa antal bokstäver per ord. Valet grundar sig på att det är detta sätt att mäta ordlängd som har varit, och fortfarande är, vanligast inom svensk elevspråksforskning, vilket möjliggör jämförelser med tidigare forskning. Jag har också, som tidigare nämnts, valt att inte kontinuerligt redovisa resultaten av variabeln extra långa ord, det vill säga ord som består av 13 bokstäver eller mer. Den ingår i SVIT-modellen, men visade sig ge mycket litet utslag vid analyserna av texterna i studien (se vidare avsnitt 5.3).

Ordvariationsindex (OVIX) är ett från början svenskt mått som togs fram i samband med de projekt som bland annat resulterade i *Gymnasistsvenska* (Hultman & Westman 1977). Ett högre OVIX-värde innebär fler unika ord i en text, det vill säga en högre *ordvariation*. Dock är det unika ord, eller löpord, som analyseras. Böjningsformer av samma ord räknas alltså som olika ord. En elev som ibland stavar fel, och ibland sär skriver vissa ord, skulle också i princip kunna uppnå ett högre OVIX-värde än en elev som är konsekvent i sin stavning och sammansättning; denna felkälla bör dock vara marginell.

OVIX räknas ut med en formel vari ingår viss kompensation för textlängden – i en längre text kommer ett antal ord, inte minst språkets vanligaste ord, att upprepas många gånger, vilket felaktigt kan ge intrycket att längre texter har en lägre ordvariation än kortare. OVIX är en vidareutveckling av måttet *type token ratio* (TTR). TTR innebär att antalet unika (löp)ord i en text (*type*) delas med det totala antalet löpord (*tokens*). Variabeln ingår även den i SVIT-modellen, men redovisas inte eftersom OVIX generellt är ett säkrare mått, och dessutom det mått som brukar användas inom elevspråksforskning. Dock kan inte heller OVIX-formeln fullt ut kompensera för textlängden när det kommer till riktigt långa texter. Variabeln är av stort intresse i studien med tanke på dess fokus på elevers ordanvändning. Hur varierat är elevernas ordförråd? Har ordvariationen samband med betyg? Hur förhåller sig elevernas ordvariation till källtexternas och de vetenskapliga artiklarnas ordvariation? Dock finns det alltså ett antal reservationer att göra beträffande vad OVIX verkligen mäter – främst att böjningsformer räknas som unika ord, och att textlängden i viss mån kan dra ner OVIX-värdet.

Läsbarhetsindex (LIX) är även det ett svenskt mått, utvecklat av Björnson 1968. Det grundar sig på genomsnittlig meningslängd (antal ord per mening) och andelen långa ord i en text. Som långa ord räknas ord med fler än 6 bokstäver (Lix räknare u.å.). Ju lägre LIX-värde, desto mer lättläst anses en text vara. Måttet har kritiserats för att vara alltför enkelt; det säger till exempel ingenting om syntaktisk komplexitet. Det har dock använts, och används fortfarande, som ett av många sätt att mäta hur språkligt avancerade texter är. Måttet ingår i SVIT-modellen och jag har valt att redovisa det i studien som ett mått på eventuella skillnader mellan olika textgrupper, och också för att möjliggöra jämförelser med tidigare forskning där LIX ingår.

Lemmavariationsindex (LVIX) är ett nyare automatiskt mått som kan ses som en vidareutveckling av OVIX. LVIX räknas ut med samma formel som OVIX. Skillnaden är att i LVIX slås alla böjningsformer av ett ord först samman, det vill säga ett ords alla lemmaformer räknas tillsammans som ett enda ord, medan de i OVIX räknas som olika ord. LVIX är därför ett bättre mått på verklig lexikal variation i form av olika ordbetydelser, eller lexem, än OVIX. I vetenskaplig text föredras på ordnivå precision och stringens framför stilistisk variation, vilket innebär att det oftare är bättre att upprepa samma ämnesord (begrepp, termer) än att använda synonymer. Eftersom LVIX är ett relativt nytt automatiskt mått finns inte så många jämförbara studier på elevtext att tillgå. Dock kan det vara intressant att i studien ställa OVIX- och LVIX-måtten mot varandra. Om en text har ett klart lägre LVIX-värde än OVIX-värde skulle det kunna tyda på att skribenten använder många böjningsformer av samma ord, vilket alltså kan indikera en större ämnesmässig precision och en mer vetenskaplig stil. Om LVIX- och OVIX-värdena ligger nära varandra innebär det istället att fler ord är unika i texten (det vill säga tillhör olika lemman), vilket alltså på motsvarande sätt skulle kunna tyda på en större ämnesmässig splittring, eller åtminstone på att skribenten inte använder en konsekvent terminologi i behandlingen av ämnet. Felkällorna vid beräkning av LVIX i elevtext är desamma som för övriga ordmått; felstavningar och särskrivningar kan leda till ett högre värde om de inte är konsekventa, eftersom antalet unika lemmaformer då blir högre. I förhållande till det totala antalet ord i den analyserade elevtexten bör dessa felkällor vara försumbara.

Andel bisatser (UA) är ett syntaktiskt mått som har medtagits i analysen som ett led i att sätta in ordanalyserna i ett större språkligt sammanhang. En högre andel bisatser i förhållande till andelen huvudsatser förknippas ofta med ett mer avancerat språkbruk. Bisatsen ingår syntaktiskt i huvudsatsen, där den utgör en satsdel eller bestämning i fraser. Ett mer koncentrerat språkbruk genom bland annat nominaliseringar och mer utvecklade fraser, vilket är typiskt för vetenskaplig text (Ribeck 2015) skulle möjligen kunna leda till färre bisatser som utgör primära satsdelar. I sakprosatexter, till exempel i läroböcker, är andelen huvudsatser generellt hög, eftersom sådan text ofta består av relativt korta meningar med en hög andel ämnesord (Strömquist 1999; Ribeck 2015). I skönlitteratur kan vidare en hög andel huvudsatser utgöra ett medvetet stilistiskt val. Andelen bisatser är således ett mått som i och för sig kan säga något om en texts språkliga komplexitet, men det är samtidigt ett mått som är relaterat till genre. Vidare säger inte den automatiska analysen något om vilka olika typer av bisatser som förekommer i texten, vilket förstås är relevant inte minst i vetenskaplig text, där bisatserna kan beskriva olika typer av förutsättningar, orsakssamband med mera.

Måttet *träddjup* (PT) är liksom andel bisatser ett syntaktiskt mått. Det utgår från en i grunden generativ syntaktisk analys av förhållandet mellan olika syntaktiska enheter i satsen (satsdelar, fraser, bisatser), i figurform illustrerade som så kallade trädstrukturer. Träddjupet mäter avståndet från det så kallade

trädetts rot till dess yttersta grenar. Ett högre värde tyder på en större syntaktisk komplexitet inom satsen, vilket ofta förknippas med en skriftspråklig och innehållsligt mer nyanserad framställning – jämför till exempel de påhittade exemplen *språket innehåller olika regler med det skrivna språket innehåller avsevärt fler och mer precisa regler än det talade, i vilka inkluderas regler som rör sådant som stavning, ordföljd och kongruens*. I det senare exemplet är trädjupet avsevärt större än i det första, på grund av längre och mer komplexa nominalfraser. I studien kan en jämförelse av textgruppers och individuella texters genomsnittliga trädjup bidra till analysen av texternas språkliga stil.

Det bör påpekas att automatisk analys av syntaktiska drag av naturliga skäl har fler felkällor än automatisk analys av enstaka ord. Syntax innebär ofta komplexa strukturer, där det inte alltid är självklart hur ord, fraser och satser förhåller sig till varandra. Den automatiska analysen utgår dessutom, som tidigare nämnts, från grafiska meningar, vilka inte alltid sammanfaller med syntaktiska meningar. Särskilt i analysen av elevtexterna, där meningsbyggnaden inte alltid är korrekt, får man räkna med att analysen av andel bisatser har brister. Analysen av trädjupet har utan tvekan brister i alla textgrupper, men särskilt i elevtexterna. I studien används de syntaktiska måtten för allmänna jämförelser mellan textgrupperna, för att sätta in de mer ordrelaterade måtten i ett vidare språkligt sammanhang. Siffervärdena måste dock betraktas som ungefärliga.

Substantiv/pronomen-kvot (NoPr) innebär att antalet substantiv i en text delas med antalet pronomen. Pronomen är kontextberoende och används bland annat för att syfta tillbaka på substantiv. En hög andel pronomen i en text anses tyda på en mer talspråklig stil. Omvänt indikerar en hög andel substantiv en mer skriftspråklig stil (Liberg, af Geijerstam & Wiksten Folkeryd 2010). Eftersom ämnesord ofta utgörs av substantiv kan en hög andel substantiv också tyda på ett rikare ämnesinnehåll i en text. Ett högt NoPr-värde tyder alltså på en mer skriftspråklig stil än ett lågt, och förmodligen också på ett rikare eller mer nyanserat ämnesinnehåll, även om den faktiska förekomsten av substantiv i en viss text förstås behöver analyseras för att en sådan slutsats ska kunna dras med säkerhet. Variabeln är av intresse i jämförelsen av de olika textgrupperna (elevtexterna, källtexterna, de vetenskapliga artiklarna), men också i jämförelsen av elevtexter med olika betyg. Har till exempel E-texterna lägre NoPr-värden än A-texterna?

Det sista automatiska mått som inkluderas i analysen, nominalkvot (NR) är liksom substantiv/pronomen-kvoten ett mått som baseras på vissa ordklassers förekomst i text. I en så kallad enkel nominalkvot delas antalet substantiv med antalet verb. I en så kallad utökad nominalkvot, som är det som analyseras i denna studie, delas antalet substantiv, prepositioner och particip med antalet verb, pronomen och adverb. I båda fallen är avsikten att analysera om texten har en mer nominal eller verbal stil. Ett högre värde tyder på en mer nominal stil, vilket förknippas med en mer skriftspråklig stil. Formell text, som vetenskaplig text hör till, utmärks bland annat av nominaliseringar istället för verbala uttryck (till exempel *rättstavning* istället för *stava rätt*). Måttet är, liksom

substantiv/pronomen-kvoten, av intresse i jämförelsen mellan de olika textgrupper som ingår i studien, liksom i jämförelsen mellan individuella texter. Båda måtten grundar sig på en automatisk ordklassanalys, eller ordklasstaggning, av orden i texterna, vilken utgår från både ordlistor, jämförelsematerial och ordens positioner i förhållande till omgivande ord i texten. Automatisk analys av ordklassstillhörighet har, till skillnad från syntaktisk analys, en ganska liten felmarginal. Enstaka ord kan förstås analyseras felaktigt, till exempel om de utgör homonymer, såsom substantivet *man* och pronomenet *man*.

4.3.1.2 Rensning och uppmärkning av materialet

Materialet har rensats inför den automatiska analysen, med målsättningen att behålla rubriker och vanlig text, det vill säga skribentens eget språk. Det är främst *Språk och stil*-artiklarna som har behövt rensas. Nedan beskrivs vad som har tagits bort. Flertalet punkter gäller alltså *Språk och stil*-artiklarna.

Följande har tagits bort inför den automatiska analysen:

- författarens namn, mejladress, universitet etc.
- all text i sidhuvud och sidfot (sidnummer, artikelns namn och dylikt)
- abstrakt på engelska
- tabeller, tablåer, figurer och transkriptioner av tal
- fristående språkliga exempel, till exempel numrerade meningar eller ett antal ord som räknas upp fristående från texten
- alla citat på engelska – och övriga andra språk än svenska – både som blockcitat och i den löpande texten²
- blockcitat på svenska³
- alla fotnoter⁴
- referenslistor
- illustrationer och faktarutor.

Språkliga exempel på svenska i den löpande texten, liksom citat på svenska i den löpande texten, har behållits. Dessa är oftast inte så långa och är språkligt integrerade i författarens text. I elevtexterna förekommer även så kallade dolda citat, det

² I den löpande texten ingår de visserligen i författarens framställning, men de påverkar den kvantitativa ordanalysen av till exempel OVIX och LVIX på ett icke-representativt sätt, och har därför uteslutits. Skiljetecken, vanligen punkt, har lagts till om borttagningen av citatet har lett till en oavslutad mening. Även ändring av skiljetecken från kolon till punkt har gjorts i sådana fall.

³ Dessa är vanligen längre än citat i den löpande texten, och de är inte lika språkligt integrerade i texten som dessa. De skulle kunna påverka de kvantitativa ordanalysmåtten och har därför borttagits.

⁴ Vissa skulle kunna behållas, men det skulle ta omotiverat lång tid att göra ett sådant urval. Fotnotssiffran har tagits bort ur texten.

vill säga att eleven mer eller mindre har skrivit av formuleringar ur källtexterna utan att markera dessa som citat. Dessa har av naturliga skäl inte kunnat uteslutas.

Före den automatiska analysen har även en uppmärkning av materialet gjorts, vilken främst har inneburit att referenser inom parentes har markerats i texterna för att undvika att de ska påverka mätningarna av vissa mått på fel sätt.

Att rensa och märka upp ett så pass omfattande material som ingår i studien (90 elevtexter, 14 källtexter, 13 vetenskapliga artiklar) är en tidskrävande process som kräver stor noggrannhet och innebär överväganden av olika slag. På enstaka punkter hade man kanske kunnat göra andra val. Enstaka förekomster av olika slag som borde ha rensats bort kan vidare ha missats. I dess helhet bör dock materialet vara så väl förberett inför den automatiska analysen att värdena för de olika måtten är rättvisande och sinsemellan jämförbara.

4.3.1.3 Statistisk analys

I studien används de deskriptiva statistiska måtten *medelvärde*, *maxvärde* och *minimivärde* där så är relevant. I samband med den automatiska analysen har också värden för standardavvikelse och varians beräknats inom olika textgrupper.

Standardavvikelsen är ett så kallat spridningsmått och anger hur mycket olika textvärden inom olika mått avviker från textgruppens medelvärde. En låg standardavvikelse innebär att texternas värden ligger samlade nära medelvärdet. En hög standardavvikelse innebär att individuella texters värden ligger långt över eller under medelvärdet, det vill säga att spridningen mellan texterna är stor.

Standardavvikelsen räknas ut ur det närbesläktade måttet *varians*. Måtten ger ungefär samma information. De räknas ut med samma ekvation och standardavvikelsen räknas ut ur variansen. Värdena för standardavvikelse och varians anges i resultattabellerna i bilaga 1 (tabell 8–15). Standardavvikelsen är det mått som är mest brukligt att använda i forskningsstudier. Därför har jag valt att kommentera standardavvikelsen i presentationen av resultaten av den automatiska analysen (kapitel 5), men inte variansen. Att både standardavvikelse och varians redovisas i tabellerna möjliggör för forskare som önskar jämföra värden för standardavvikelse och varians i liknande automatiska textanalyser med värdena i denna studie att göra så.

Det bör påpekas att statistiska värden kan vara svårtolkade. Vad bör till exempel uppfattas som en liten respektive stor spridning mellan olika texter inom ett visst mått? Finns jämförbara värden för olika genrer att tillgå utanför studien? Med dessa reservationer gjorda så är det förstås så att statistiska värden är viktiga i tolkningen av kvantitativa data, som vid automatiska analyser av språkliga drag i stora textmaterial.

4.3.2 Manuell analys: ordkategorier

Den automatiska analysen i studien kompletteras med en manuell analys av ord och uttryck inom olika kategorier (ämnesord, akademiska ord och uttryck) i en begränsad del av elevtexterna, i avsikt att få fram en fördjupad bild på det område som utgör studiens fokus, det vill säga elevernas ordanvändning. I delstudien analyseras de tio A-texter och de tio E-texter som hör till uppgiften *Särskrivning och stavfel*. Eftersom analyserna är tidskrävande är det inte möjligt att i delstudien analysera samtliga 90 elevtexter som ingår i den automatiska analysen på detta sätt. Delstudien har därför begränsats till en av de uppgifter som ingår i materialet. Elevtexterna med betygen A och E har valts ut eftersom det är rimligt att anta att det går att hitta störst skillnader i användningen av ord och uttryck mellan dessa betygssteg. Det är vidare rimligt att anta att de olika sätt som elever använder ord och uttryck på i dessa totalt 20 elevtexter är någorlunda representativa för elevtextsmaterialet i dess helhet. Att analysera ytterligare elevtexter, inklusive elevtexter som bedömts med betyg C, skulle troligen inte förändra den övergripande resultatbilden i någon större omfattning.

I delstudien analyseras endast elevtexter, eftersom dessa utgör studiens fokus och primära källmaterial. Ord och uttryck i källtexterna och i *Språk och stil*-artiklarna, vilka utgör jämförelsematerial i den automatiska analysen, analyseras inte i delstudien. Jämförande exempel ur dessa material ges dock när så är relevant.

Det finns ett antal tidigare studier av ord och uttryck i olika sorters elevtext och i läromedel. Ribbeck (2015) inkluderar *typiskt ordförråd* som en kategori i analysen av läromedel i olika ämnen i årskurs 7–9 samt på gymnasiet (se avsnitt 2.2.5). af Geijerstam (2006) undersöker elevers skrivande i naturorienterade ämnen i årskurs 5 och 8, bland annat med fokus på det naturvetenskapliga språket.

I en lärobok från 2010 beskriver Liberg, af Geijerstam och Wiksten Folkeryd en uppdelning mellan så kallat vardagsspråk och skolspråk. Skolspråket utmärks bland annat av att det innehåller specialord (ämnesord, termer) och är kontextoberoende, medan vardagsspråket utmärks av att det är konkret och kontextberoende. Vardagsspråket är oplanerat och har ofta en verbal stil, medan skolspråket är planerat och har en mer nominal stil. (Liberg, af Geijerstam & Wiksten Folkeryd 2010:30–31). Beskrivningen rör primärt mellanstadiet. Det som gemensamt kallas för skolspråk skiljer sig vidare åt mellan olika ämnen och inom de olika genrer som förekommer i skolan. Indelningen ger dock en första, övergripande beskrivning av vad som utmärker skriftspråk i skolan i förhållande till tal, vilken är relevant även för den aktuella studien.

Inom projektet *OrdiL – en korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år* (Lindberg & Johansson Kokkinakis 2007) analyseras ordförrådet i läromedel inom NO, SO och matematik, med fokus på

vilka utmaningar flerspråkiga elever ställs inför i förståelsen av ord och fraser. Järborg definierar inom projektet fyra olika vokabulärkategorier:

- a) **Allmänspråkliga, frekventa ord:** förekommer i både tal och skrift – till exempel *vara, komma, människa, exempel*.
- b) **Allmänna, ofta abstrakta, skriftspråkliga ord:** förekommer i sakprosa – till exempel *utbredning, motsvara, föremål*.
- c) **Allmänspråkliga, ämnestypiska ord:** till exempel *klimat, strålning, muskel, företag*.
- d) **Fackord och facktermer, ofta unika för ett visst ämne:** till exempel *decimalform, kopplingsschema, kromosom*. (Järborg 2007:86–87)

Även om vokabulärkategorierna ovan presenteras i samband med en studie av läromedel inom NO, SO och matematik avsedda för grundskolans senare år, är de så pass allmänna att de fungerar som analyskategorier för alla typer av ämnesspråk, och för läroböcker såväl som för vetenskaplig text. Att lära sig nya ämnesspråk är också en utmaning för alla elever, oavsett modersmål, även om flerspråkiga elever förstås ställs inför extra utmaningar, bland annat i form av polysema och homonyma ord, fasta uttryck av olika slag och metaforer.

Den övergripande utgångspunkten i utvecklingen av ordanalyskategorier i studien är dess fokus på elevernas ord och uttryck i den aktuella texttypen. Denna utmärks av ämnesord och akademiska ord och uttryck. Järborgs vokabulärkategorier, med fokus på kategori b, c och d, har tjänat som en utgångspunkt i utvecklingen av relevanta analyskategorier. Hänsyn har även tagits till de kriterier som på olika sätt rör aspekten språk och stil i provets bedömningsmatris (se avsnitt 4.2.1.4). Dessa kriterier innebär att texten ska ha en saklig ton, ha fungerande källhänvisningar och fungerande textbindning, precision i ordval och uttrycksformer (för betyg A), ha en vetenskaplig stil och vara språkligt korrekt.

Nedan presenteras studiens ordanalyskategorier, tillsammans med underkategorier och ordexempel:

1. **Centrala begrepp och termer** (jfr OrdiL kategori d): begrepp och termer som är centrala i den aktuella uppgiften, till exempel *särskrivning, stavfel*.
2. **Ämnesord** (jfr OrdiL kategori c–d): alla ord och uttryck som hör till de ämnen som texten behandlar, inklusive orden i kategori 1, inom två olika underkategorier:
 - a) språk och språkliga normer, till exempel *det svenska skriftspråket, stavningsregler, felböjningar*
 - b) läs- och skrivprocess, texter, till exempel *den som läser, skolelevs uppsatser*.
3. **Akademiska ord och uttryck** (jfr delvis OrdiL kategori b): alla ord och uttryck som visar att texten hör till den akademiska texttypen, inom tre olika underkategorier:
 - a) allmänvetenskapliga ord och uttryck, till exempel *frågeställning, resultat, distinktion*

- b) referatmarkörer, till exempel *Borg belyser..., enligt Borg..., i artikeln står det...*
 - c) övriga akademiska textbindningsmarkörer, till exempel *dock, för att sammanfatta*.
4. **Stilbrott:** ord, uttryck, formuleringar och syntaktiska drag som inte är förenliga med en akademisk stil, till exempel *kolla, medans, jag tycker...*, satsradningar, meningsfragment.

Analyserna av ord, uttryck och formuleringar inom ovanstående kategorier har gjorts i en text i taget i ett separat dokument, med utrymme för de olika kategorierna och underkategorierna. Även övriga iakttagelser rörande ord och uttryck som inte direkt ryms i analyskategorierna, men som kan vara av intresse för studien i dess helhet, har noterats. En sådan övrig iakttagelse kan till exempel gälla att en elev ofta, utöver direkt akademiska ord och uttryck, använder en formell stil, eller växlar mellan en formell och informell stil. I resultatdelen presenteras en sammanfattning av analysen, med fokus på sådant som typiska exempel, andelar ord och uttryck i olika kategorier samt jämförelser mellan A- och E-texter.

Som framgår av exemplen i de olika kategorierna ovan har enstaka ord, fraser och längre formuleringar inkluderats i analyskategorierna baserat på kategori-tillhörigheten. Ibland kan samma ord, uttryck eller formulering ingå i flera kategorier. Detta gäller till exempel en formulering som "[...] om man alltid skriver som designern till en snygg reklamskylt [...]" (A2⁵, *Särskrivning och stavfel*), vilken dels ingår i kategori 2b (ämnesord kopplade till läs- och skrivprocess), dels i kategori 4 (stilbrott i form av populärvetenskaplig stil).

I vetenskaplig text bärs mycket av analysen upp av vissa centrala begrepp. Att centrala begrepp är väldefinierade och används på ett konsekvent sätt i en analys, vilket inkluderar att de benämns på samma sätt, med samma term, är en förutsättning för resonemangets vetenskapliga kvalitet. Detta är anledningen till att kategori 1 har medtagits i delstudien. Vilka begrepp som ska betraktas som centrala i elevernas texter kan förstås diskuteras. En rimlig utgångspunkt kan vara de ämnesrelaterade begrepp som nämns i uppgiftsinstruktionen. I den aktuella uppgiften nämns *särskrivning, stavfel, skribent/skribenter, [brott mot] skrivnormer, skrivnormerna, stavningsreglerna, skriva isär [sammansatta ord], stava rätt och läsningen*. Av dessa kan *särskrivning* och *stavfel* betraktas som särskilt centrala, eftersom de även ingår i uppgiftens rubrik. Att analysera samtliga begrepp som nämns i uppgiftsinstruktionen i alla 20 elevtexter skulle vara alltför tidskrävande. Analysen av centrala begrepp har därför begränsats till *särskrivning* och *stavfel*, samt *rättstavning* (förekomster, böjningsformer och synonymer). Det kan vara av intresse att undersöka om de elever som har fått betyget A använder dessa begrepp

⁵ I hänvisningar till elevtexter anger bokstaven textens betyg och siffran vilken text som avses. En fullständig förteckning över analyserade texter, med beteckningar, återfinns i bilaga 2.

mer konsekvent än de elever som har fått betyget E. I bedömningsmatrisen nämns ”precision i ordval och uttrycksformer” (se avsnitt 4.2.1.4) som ett kriterium på A-nivå. En sådan precision kan till exempel handla om en konsekvent användning av centrala begrepp.

Kategori 2, ämnesord, har delats in i två underkategorier, baserat på de ämnesbegrepp som nämns i uppgiftsinstruktionen. Dessa pekar ganska tydligt på två olika områden, eller teman, i uppgiften: ett som har ett språkligt fokus (språk, språkliga normer) och ett annat tema som handlar om att läsa och skriva olika typer av texter. Ämnesord i uppgiftsrubriken och i direkta citat har ej inkluderats i kategorin. Boktitlar (till exempel SAOL), personnamn, och språkliga exempel (vanligen från källtexterna) har inte heller inkluderats.

Det bör påpekas att de ord som ingår i kategori 1, centrala begrepp och termer, också ingår i kategori 2, ämnesord. Kategorierna behandlas dock var för sig i resultat och analys, varför det i praktiken inte utgör ett problem att enstaka ord ingår i båda kategorierna.

Ibland överlappar också ord i kategori 2a och 2b, särskilt om de ingår i en längre formulering som anknyter till båda ämnesområdena. De kan därmed ibland ha placerats i båda kategorierna. Eftersom andelen ämnesord generellt sett är hög i texterna bör det inte inverka på resultatet i dess helhet att enstaka ord räknats två gånger.

Ämnesord, inklusive centrala begrepp och termer, bär upp en texts innehåll och utgör en fundamental och utmärkande del av sakprosatexter, inklusive vetenskapliga texter. De har en direkt koppling till textens innehåll och kvalitet och utgör således en central del av det ordförråd som är relevant att analysera inom delstudien.

Även kategori 3, akademiska ord och uttryck, har delats in i underkategorier. Dessa tre utgår ifrån tre typiska språkliga inslag i akademisk text: allmänvetenskapliga ord och uttryck (kategori 3a), referatmarkörer (kategori 3b) samt övriga akademiska textbindningsmarkörer (kategori 3c). Elevernas användning av dessa ord och uttryck kan relateras till hur väl de lyckas med texttypen. Kategorierna ingår även på olika sätt i bedömningsmatrisen (se avsnitt 4.2.1.4).

Den fjärde kategorin, stilbrott, har inkluderats som jämförelse med kategori 3. I stilbrottskategorin inkluderas ordval, uttryck, formuleringar och syntaktiska konstruktioner som bryter mot den akademiska texttypen. En hypotes är att sådana stilbrott leder till lägre betyg, och att de därmed skulle kunna vara fler i E-texterna än i A-texterna. Sambandet är av intresse i en övergripande diskussion av i vilken grad ämnesmässig precision kontra stilistisk medvetenhet inverkar på lärarens bedömning av texten. Stilen bedöms även av lärarna, enligt kriterier i bedömningsmatrisen.

Vad som ska betraktas som ett stilbrott i en akademisk text är förstås inte självklart. Hultmans och Westmans (1977) definition, att stilbrott innebär avvikelser från en jämn stilnivå (se avsnitt 2.2.2.1), har tjänat som utgångspunkt. Hultman

och Westman betraktar denna feltyp som mindre allvarlig och menar att det snarare är oroande att eleverna uppvisar så liten språklig kreativitet. De pekar också på att det i sammanhanget är problematiskt att stimulanstexterna är skrivna med olika stilar. I den akademiska texttyp som undersöks i den aktuella studien bör dock stilnivån ha större betydelse. Här är ju tanken bland annat att eleverna ska lära sig att använda en akademisk stil. Därför ägnas stilbrott större uppmärksamhet i denna studie. I resultatdelen (avsnitt 6.4) redogör jag för de olika typer av stilbrott som jag har urskilt inom kategori 4.

Att analysera förekomsten av ord, uttryck och formuleringar inom olika kategorier i elevtext är förstås inte helt enkelt. Elever är inte professionella skribenter och till skillnad från publicerade texter har dessa elevtexter inte granskats och redigerats före inlämning. Elevtexterna innehåller språkliga fel och oklara formuleringar. Det är ofta långt ifrån självklart hur en fras eller formulering ska avgränsas. Analyskategorierna är i sig tydliga, men naturligtvis förekommer gränsfall dem emellan, och, som nämnts, ord, uttryck och formuleringar som beroende på analytiskt perspektiv kan ingå i flera kategorier. Min förhoppning är dock att analyserna inom delstudien sammantaget bidrar med en tillräckligt tydlig bild av elevernas användning av ämnesord och akademiska uttryck, relaterat till betyg, vilken kan tjäna som ett värdefullt komplement till den automatiska analysen.

5 Resultat från automatisk analys

I kapitlet presenteras resultaten av den automatiska analysen av elevtexter och jämförelsematerial, det vill säga källtexterna och de vetenskapliga artiklarna från tidskriften *Språk och stil*. Fokus ligger på jämförelser mellan de tre textgrupperna, inklusive jämförelser mellan de båda uppgifter från nationella provet som ingår i studien, och på jämförelser av elevtexter med betygen A, C och E. Även individuella skillnader inom textgrupper behandlas. Ytterst syftar denna del av resultatanalysen till att karaktärisera elevernas akademiska skrivande i förhållande till källtexter och autentiska vetenskapliga artiklar, som ett led i att besvara studiens första frågeställning.

Samtliga resultat av den automatiska analysen går att hitta i avsnitt 5.1, tabell 2, och i bilaga 1, tabell 7–15 (medelvärdena har i samtliga tabeller avrundats till 2 decimaler). Hänvisningar till dessa tabeller görs när så är relevant.

5.1 Medelvärden i elevtexter, källtexter och vetenskapliga artiklar

I tabell 2 redovisas medelvärden för de tio olika automatiska mått som ingår i studien (se avsnitt 4.3.1.1). I tabellen har alla 30 elevtexter med samma betyg (A, C och E) slagits samman, vilket alltså innebär att både elevtexter som behandlar uppgiften *Särskrivning och stavfel* samt *Skrivregler och sms* ingår i samma betygsgrupp i tabellen. Elevtexter med olika betyg jämförs i tabellen med texterna i texthäftet, det vill säga med de källtexter som eleverna använder sig av när de skriver sin egen text, samt med autentiska språkvetenskapliga artiklar från tidskriften *Språk och stil*.

Tabell 2. Medelvärden (SVIT) för alla textgrupper.

Mått	Elevtexter betyg A N=30	Elevtexter betyg C N=30	Elevtexter betyg E N=30	Källtexter N=14	Vetenskapliga artiklar N=13
Löpord	764,27	703,37	656,77	918	6805
Antal ord/mening	20,33	19,71	21,69	14,9	19,6
Antal bokst./ord	5,11	4,94	4,83	5,03	5,66
OVIX	65,83	63,08	59,33	75,2	67,5
LIX	47,16	44,01	44,54	39	53
LVIX	55,09	52,87	50,6	62,9	58,1
Andel bisatser	0,92	0,91	1,03	0,51	0,56
Träddjup	8,33	8,17	8,59	6,66	7,72
Subst./Pron.-kvot	3,28	2,67	2,62	2,62	6,48
Nominalkvot	0,98	0,9	0,84	0,91	1,59

Utifrån tabell 2 ovan är det tydligt att det finns både likheter och stora skillnader mellan olika textgrupper. *Språk och stil*-artiklarna är till exempel avsevärt längre än elevtexterna och källtexterna och har i genomsnitt längst ord, liksom högst andel substantiv i förhållande till pronomen och högst nominalkvot. Elevtexterna ligger i flera variabler närmare *Språk och stil*-artiklarna än källtexterna; detta gäller till exempel meningslängd och ordvariation. Källtexterna har det högsta värdet för ordvariation och i genomsnitt avsevärt kortare meningar än övriga texter. Andelen bisatser är vidare avsevärt lägre i källtexterna och i de vetenskapliga artiklarna än i elevtexterna.

De genomsnittliga värden som presenteras i tabell 2 måste förstas tolkas med försiktighet. De ger en övergripande och möjligen trubbig bild av texternas språkliga landskap. Det kan finnas stora variationer mellan individuella texter inom textgrupperna (se vidare avsnitt 5.2–5.6). Värdena säger vidare ingenting om de faktiska språkliga uttrycken – vilka ordval, formuleringar och syntaktiska konstruktioner döljer sig bakom siffrorna? I elevtexterna kan också, såsom påtalats i metodkapitlet (se avsnitt 4.3.1.1), språkliga fel av olika slag påverka vissa värden. Med dessa reservationer gjorda så är det förstas så att analysen också har styrkor; den tillhandahåller en övergripande jämförelse av språkliga drag i ett omfattande textmaterial, vilken inte hade varit möjlig att genomföra manuellt.

I de följande avsnitten kommenteras värdena i tabell 2 mer djupgående, varvid även skillnader inom textgrupper berörs, inklusive skillnader mellan elevtexter med olika betyg och mellan elevtexter som utgår från uppgiften *Särskrivning och stavfel* respektive *Skrivregler och sms*. De tabeller som de senare värdena har hämtats ur återfinns i bilaga 1.

5.2 Textlängd och meningslängd

Hultman och Westman (1977, se även avsnitt 2.2.2.1) lyfter fram att längre elevtexter generellt får högre betyg än kortare upp till en viss nivå. För det högsta betyget verkar dock ett mer koncentrerat skrivsätt premieras, vilket bland annat innebär mer utbyggda nominalfraser framför långa meningar och många bisatser. Hultmans och Westmans studie behandlar bruksprosa. Det är rimligt att anta att slutsatsen i minst lika hög grad gäller akademisk text. I detta avsnitt presenteras resultaten beträffande textlängd och meningslängd, vilka alltså är intressanta att analysera i förhållande till betygen på elevtexterna och i förhållande till genre (elevtexter, källtexter och vetenskapliga artiklar).

Variabeln löpord talar om hur långa texterna är i genomsnitt, det vill säga hur många ord de innehåller. De vetenskapliga artiklarna i *Språk och stil* är, trots att de har rensats på sådant som abstrakt på engelska, tabeller och referenslistor, ändå avsevärt mer omfattande än både elevtexterna och källtexterna. I genomsnitt är de, i rensat skick, knappt 10 gånger längre än elevtexterna. Källtexterna, vilka dock i flera fall utgörs av utdrag ur mer omfattande texter, ligger i genomsnitt

avsevärt närmare elevtexterna i omfattning. Resultatet var förväntat och bekräftar vad man kan se med blotta ögat. Det speglar det förhållande att vetenskapliga artiklar ofta är resultatet av ett omfattande arbete och inkluderar detaljerade redogörelser för sådant som tidigare forskning, empiri och analys. De har ofta skrivits under lång tid och har bearbetats och redigerats i flera led. Skrivsituationen är därmed radikalt annorlunda än den skrivsituation som elevtexterna har tillkommit i. Eleverna har visserligen kunnat förbereda sig genom att läsa de källtexter som de sedan refererar till i den egna texten i förväg. Den egna texten är dock tillkommen under en begränsad skrivtid. Uppgiften meddelas på plats och det är inte möjligt att bearbeta texten efter att den har lämnats in.

Utifrån medelvärdena är elevtexter med högre betyg längre än elevtexter med lägre betyg. A-texterna är i genomsnitt 100 ord längre än E-texterna. Eleverna kunde inte skriva hur långt de ville; elevtexternas omfattning skulle enligt rekommendationen i uppgiftsinstruktionen ligga mellan 600 och 800 ord. Det innebär att dessa elevtexter är av liknande omfattning som elevtexterna (debattartiklar) i Hultmans och Westmans (1977) studie. Dessa var i genomsnitt runt 600 ord.

Resultatet bekräftar vad som är känt från tidigare forskning, det vill säga att det finns ett tydligt samband mellan textlängd och betyg, på så sätt att längre elevtexter generellt får högre betyg (Hultman & Westman 1977; Melander 1997; Nyström 2000; Östlundh-Stjärnegårdh 2002). Det är rimligt att anta att skillnaderna i textlängd mellan elevtexter med högre och lägre betyg hade varit ännu större om det inte hade funnits en rekommenderad begränsning av textlängden i uppgiftsinstruktionen.

Värdena för standardavvikelse (se tabell 8–15 i bilaga 1) visar på mycket stora skillnader i textlängd mellan *Språk och stil*-artiklarna, trots att de samtliga är långa. Dessa skillnader kan delvis förklaras av att artiklarna i varierande grad innehåller tabeller och andra bortrensade inslag. Det finns, som för alla vetenskapliga tidskrifter, ett högsta tillåtet sidantal, men vad som inkluderas i dessa sidor, det vill säga huruvida det är fråga om ren text eller andra delar, kan variera. Även källtexterna har en hög standardavvikelse för löpord, vilket illustrerar att även de varierar i omfattning. Den kortaste texten är på knappt 300 ord och den längsta på knappt 2000 ord.

Standardavvikelsen är avsevärt lägre i de 6 olika grupperna av elevtexter (betyg A, C och E för respektive uppgift, se tabell 8–13 i bilaga 1), vilket förstås speglar omständigheten att textomfånget var begränsat för eleverna. Dock är standardavvikelsen högre i elevtextgrupperna med betyg E, för båda uppgifterna, vilket alltså tyder på större variationer i textlängd för elevtexter som får betyg E. Medelvärdena för löpord är vidare något högre för uppgiften *Särskrivning och stavfel*, i alla tre betygsstegen, än för uppgiften *Skrivregler och sms*. Möjligen skulle denna omständighet kunna spegla ett förhållande att *Skrivregler och sms*, som var den uppgift som flest elever valde, lockade fler skrivovana elever än den kanske lite torrare uppgiften *Särskrivning och stavfel*.

Variabeln antal ord per mening talar om hur många ord meningarna i en text innehåller i genomsnitt. Resultaten är intressanta: elevtexternas medelvärden, och då särskilt betyg C, ligger avsevärt närmare *Språk och stil*-artiklarna än källtexterna. Källtexterna utmärker sig i denna variabel med ett avsevärt lägre medelvärde (i snitt knappt 15 ord per mening), än övriga textgrupper, där medelvärdet är runt 20 ord per mening.

Det finns förstås variationer inom textgrupperna. Några exempel kan ges för att visa på spridningen. Det lägsta värdet inom betyg A, uppgiften *Särskrivning och stavfel*, är knappt 16 ord per mening, medan det högsta är 25 ord per mening. Bland E-texterna, uppgiften *Skrivregler och sms*, är det lägsta värdet knappt 17 ord per mening, och det högst 34 ord per mening. Bland källtexterna är det lägsta värdet drygt 9 ord per mening, och det högsta drygt 21 ord per mening. *Språk och stil*-artiklarnas genomsnittliga meningslängd går från drygt 16 ord till drygt 28 ord. Standardavvikelsen visar på störst variationer i meningslängd inom elevtextgrupperna med betyg E.

Att källtexterna generellt innehåller kortare meningar än både elevtexterna och *Språk och stil*-artiklarna speglar att källtexterna i flertalet fall kan betecknas som populärvetenskapliga. De har hämtats från populärvetenskapliga tidskrifter, dagstidningar och handböcker. I förhållande till vetenskaplig text innehåller populärvetenskaplig text ofta ett lite enklare, eller mer kortfattat, språk, i syfte att förklara på ett lättfattligt sätt. Här kan en parallell dras till lärobokstext, där faktatätheten ofta är hög och textbindningen mindre utvecklad (Strömquist 1999; Ribeck 2015).

Ett exempel kan ges ur den källtext som har den i genomsnitt kortaste meningslängden:

Så hur ska språkvårdarna råda? Talspråksanpassning eller skriftspråksanpassning? Konsekvent dom-bruk vore enkelt för skribenterna men ett stort ingrepp. Alla texter tillkomna före reformen bleve patinerade. Men vi vill befria svensklärare och skribenter från de-dem-grubblrier. (Källtext 14)

Som framgår i citatet ovan är meningarna mycket korta. Den andra meningen innehåller endast 3 ord, och utgör i syntaktisk bemärkelse ingen fullständig sats; bland annat saknar den helt verb. Sådana ofullständiga satser accepteras i genrer som populärvetenskap och skönlitteratur, där de ofta används medvetet som stildrag, för att skapa rytm i texten. I akademisk text betraktas de dock som just ofullständiga satser, det vill säga som grammatiskt inkorrekta. Eftersom akademisk text ställer höga krav på precision och tydlighet förväntas språket vara korrekt, vilket bland annat innebär ett undvikande av ofullständiga meningar. Det bör dock påpekas att språket i citatet ovan inte i alla avseenden kan betraktas som enkelt; till exempel används konjunktiv (*vore, bleve*) och det ovanliga ordet *patinerade*. Detta är några exempel som visar på hur komplext det är att bedöma en texts svårighetsgrad.

Det har flera gånger påpekats att elevtexternas uppmätta värden kan vara påverkade av olika typer av språkliga fel, vilka sällan återfinns i redigerade texter

av professionella skribenter. Ett sådant område är just meningslängden. Satsradningar, det vill säga att skribenten fogar samman huvudsatser med kommatecken istället för att tydliggöra relationen mellan dem med punkt, bindeord eller semikolon, är vanligt förekommande i elevtexter, som i följande exempel:

Det visade sig att två stavfel skulle vara nog för att läsarna skulle tappa förtroendet för texten samt för att dem skulle läsa långsammare, förtroendet skulle minska ännu mer när det var upptill⁶ sex fel per sida och det gör att man får en negativ helhetsbedömning av texten. (E1, *Särskrivning och stavfel*)

Satsradningen i citatet ovan återfinns på rad två, där skribenten har valt att sätta kommatecken mellan orden *långsammare* och *förtroendet*, istället för att, vilket hade varit naturligtast i detta fall, låta *förtroendet* inleda en ny grafisk mening. Det framgår också av citatet att meningslängden även, vid sidan av satsradning, kan dras upp av många satser i samma grafiska mening, samt av en ordrik eller pratig ton. Som tidigare påpekats är elever inte professionella skribenter och alla behärskar inte fullt ut interpunktion. Deras kunskap om det ämne de skriver om är begränsad. Skrivtiden och möjligheten till textbearbetning är också begränsade. Långa och pratiga meningar är ofta ett resultat av att skribenten inte kan eller hinner tänka igenom alla orsakssammanhang och hitta de mest precisa och koncentrerade uttryckssätten.

Utifrån elevtexternas medelvärden (se tabell 2 i avsnitt 5.1 och tabell 7 i bilaga 1) finns det skäl att anta att elevtexter med lägre betyg innehåller fler satsradningar och pratigare meningar än elevtexter med högre betyg. Meningslängden är också i genomsnitt högre i uppgiften *Skrivregler och sms*, i betygen C och E, vilket skulle kunna bero på att, som tidigare nämnts, detta var den populäraste skrivuppgiften och att den kan ha lockat många ovana skribenter. Kanske bidrar också ett större personligt engagemang i ett ämne till ett mer ordrikt och obearbetat flöde i skrivandet. Det är dock viktigt att påpeka att pratiga formuleringar och satsradningar även återfinns i elevtexter med högre betyg, som i följande exempel:

Under sjuttonhundratalet började man massproducera text. Tidigare hade var och en stavat lite som de ville, man stavade ord som man tyckte att de lät och detta kunde variera mellan dialekter och individer. (A1, *Särskrivning och stavfel*)

Satsradningen återfinns i den andra grafiska meningen, mellan orden *ville* och *man*.

Hur kan då ett mer koncentrerat skrivsätt, med kortare meningar, se ut? Ett exempel kan ges ur den *Språk och stil*-artikel som har den lägsta genomsnittliga meningslängden i studien:

Synestesi är ett område som har ådragit sig stort intresse inom språkforskningen. Synestesi innebär enkelt uttryckt att ett intryck från ett sinne används metaforiskt för att

⁶ Samtliga citat ur elevtexterna återges ordagrant, inklusive eventuella språkfel.

beskriva ett intryck från ett annat sinne, som i *varma färger* och *vassa ljud*. Språkforskningen på detta område har bland annat varit inriktad på att analysera vilka sinnen som beskriver vilka [...]. (SoS 12)

Koncentration på meningsnivå innebär ofta, som i citatet ovan, att innehållet i den grafiska meningen begränsas till en omständighet eller delaspekt av ämnet, vilket ofta föredras i akademisk text av tydlighetsskäl. Ytterligare ett exempel kan ges ur en elevtext:

Särskrivning är däremot inte det enda brottet mot skrivnormer som påverkar läsningen. Även stavfel kan ställa till problem. (A4, *Särskrivning och stavfel*)

Hade eleven i exemplet ovan istället valt att skriva samman de båda meningarna till en enda hade innehållet i den andra meningen inte framhävts lika tydligt.

5.3 Ordlängd, ordvariation och lemmavariation

Ordlängd, ordvariation och lemmavariation är tre mått som alla handlar om ordförrådet i en text. Ordlängdsmåttet anger hur långa orden i texten är i genomsnitt, det vill säga hur många bokstäver de innehåller. Ordvariationsmåttet och lemmavariationsmåttet mäter istället, lite förenklat, andelen unika ord i texten. Ordförrådet inom olika kategorier i elevtexterna behandlas mer utförligt i kapitel 6; där ges också exempel på ord och uttryck i elevtexterna.

Variablerna ordlängd och ordvariationsindex har i tidigare forskning visat sig ha ett starkt samband med både betyg och genre (se till exempel Hultman & Westman 1977; Melander 1997; Strömquist 1999; Nyström 2000; Ribeck 2015). Elever som får högre betyg använder i genomsnitt längre ord och uppvisar ett större ordförråd i sitt skrivande. Ordlängden är också i viss mån genreberoende; sakprosatexter innehåller fler ämnesord och specialtermer av olika slag, vilka vanligen är längre än mer vardagliga ord, som i sin tur är vanligare i mer personliga texter, och även i skönlitteratur. Vidare är specialtermerna förstås fler i en vetenskaplig text än i en populärvetenskaplig text om motsvarande ämne.

Som förväntat är den genomsnittliga ordlängden högst i de vetenskapliga artiklarna, med 5,66 bokstäver per ord (se tabell 2). Även elevtexterna följer det förväntade resultatet; den genomsnittliga ordlängden är högst i A-texterna (5,11), lite lägre i C-texterna (4,94) och lägst i E-texterna (4,83). Källtexterna, med ett genomsnitt på 5,03 bokstäver per ord, befinner sig mellan A- och C-texterna. Men det är alltså elevtexterna med betyg A som i genomsnitt ligger närmast de vetenskapliga artiklarna, sett till genomsnittlig ordlängd.

I Hultmans och Westmans studie (1977) var den genomsnittliga ordlängden i alla elevtexter 5,07. I elevtexter med betyg 5 var den i genomsnitt 5,52, och i elevtexter med betyg 1 var den 4,83. Detta innebär att betyg 1-texterna i *Gymnasistsvenska* hade exakt samma genomsnittliga ordlängd som betyg E-texterna i denna studie. Betyg A-texterna har dock en klart lägre genomsnittlig ordlängd än betyg 5-texterna i *Gymnasistsvenska* – detta trots att det i denna studie är akademiska texter som har analyserats, och i *Gymnasistsvenska* debattartiklar. Utifrån dessa siffror förefaller skillnaderna mellan olika grupper av elever ha minskat, och ordlängden förefaller generellt ha blivit lägre, med kortare ord. Betyg 5-texternas genomsnittliga ordlängd (Hultman & Westman 1977) ligger nära genomsnittet för de vetenskapliga artiklarna i denna studie.

I jämförelse med Melanders (1997) studie av bland annat ordlängd i nationella prov i svenska 1995, är ordlängden i denna studie högre. Melander fick fram medellängden 4,81 på teoretiska program, och 4,59 på yrkesinriktade. Nyström (2000) fick fram varierande genomsnittliga värden kopplat till genre i elevtexter på gymnasiet, från 4,29 (så kallade öppna brev i nationella provet) till 5,26 (fakta-redovisningar). Det går förstås inte att utesluta att ordlängden i elevtexter generellt sett kan öka eller minska över tid, men de skilda värdena i förhållande till dessa senare studier förefaller snarare vara relaterade till elevtexternas skilda genrer.

Den genomsnittliga ordlängden i elevtexterna är i alla betygsstegen något högre för uppgiften *Särskrivning och stavfel* än för *Skrivregler och sms* (se tabell 7). Skillnaden är minst för betyg A och något större för betyg C och E. Individuella variationer inom textgrupperna finns naturligtvis, men standardavvikelsen visar att de generellt sett är ganska små. Kanske är det helt enkelt inte möjligt att inom en viss genre variera den genomsnittliga ordlängden i någon större omfattning.

I SVIT-modellen ingår också variabeln extra långa ord, det vill säga ord på minst 13 bokstäver. Sådana ord är generellt sett sällsynta i text. Skillnaderna mellan och inom de olika textgrupperna i studien är alltför små för att det ska vara meningsfullt att redovisa siffrorna i detalj. Det kan dock nämnas att de vetenskapliga artiklarna i genomsnitt, som förväntat, innehåller en något högre andel (omkring 4 procent) extra långa ord än källtexterna och elevtexterna (omkring 2 procent). Individuella variationer finns, alltifrån någon enstaka elevtext som knappt innehåller något enda extra långt ord, till vetenskapliga artiklar som innehåller 6 procent extra långa ord.

I variabeln ordvariationsindex (OVIX) mäts andelen unika ord i en text. Måttet har använts sedan 1970-talet och det finns därför resultat från tidigare studier att jämföra med. Som nämntes i metodkapitlet har analysen bland annat svagheten att olika böjningsformer av samma ord räknas som unika ord. Vetenskaplig text bör, sett till användning av specialtermer, ha en högre ordvariation än en lärobokstext eller populärvetenskaplig text om motsvarande ämne. Samtidigt eftersträvas enhetlig terminologi i vetenskaplig text, det vill säga ett undvikande av synonymer, vilket i vissa avseenden kan leda till lägre ordvariation.

Bland textgrupperna i studien (se tabell 2) är det OVIX i elevernas A-texter (knappt 66) som ligger närmast de vetenskapliga artiklarna (67,5). Värdena för C- och E-texterna är lägre – runt 63 för C-texterna och runt 59 för E-texterna. Högre ordvariation förefaller alltså, vilket även tidigare studier har visat (se till exempel Hultman & Westman 1977), ha ett samband med högre betyg. Skillnaderna mellan de båda uppgifterna är små. Den textgrupp som avviker mest i denna variabel är källtexterna, med ett OVIX-medelvärde runt 75, vilket är klart högre än i övriga textgrupper. Enligt Hultman och Westman innebär ett OVIX-värde under 60 en låg lexikal variation, och ett OVIX över 70 en hög lexikal variation. Baserat på de siffrorna skulle alltså E-texternas ordvariation, med ett medelvärde på 59, kunna anses relativt låg, medan källtexterna kan anses ha en hög ordvariation.

Nyström (2000) konstaterar att OVIX varierar med genre. I hennes studie av elevtexter på gymnasiet har faktaredovisning högst OVIX (66). Dessa elevtexter innehåller dock många avskrifter och lån från andra texter, vilket även elevtexterna i denna studie gör. Nyström påpekar att ett högt OVIX kan spegla en alltför snabb ämnesprogression med otillräcklig textbindning, medan ett lågt OVIX kan indikera innehållslig tunnhet. Att källtexterna har högst OVIX i denna studie, och elevtexterna med betyg E lägst OVIX, är överensstämmande med denna bild.

I *Gymnasistsvenska* (1977) var det genomsnittliga OVIX-värdet i alla elevtexter (debattartiklar) 64,8. I betyg 5-texterna var genomsnittet 67,6 (ungefär detsamma som i de vetenskapliga artiklarna i denna studie), och i betyg 1-texterna var det 58,7 (ungefär samma som E-texterna i denna studie). I Melanders (1997) studie var det genomsnittliga OVIX-värdet knappt 68 på teoretiska program, och knappt 60 på yrkesförberedande.

Standardavvikelsen är för OVIX hög inom samtliga textgrupper (se tabell 8–15 i bilaga 1); de individuella skillnaderna i den ordvariation som OVIX mäter är alltså stora mellan texterna. Det är dock intressant att notera att standardavvikelsen är lägst mellan de vetenskapliga artiklarna; dessa befinner sig således, sett till ordvariation, närmare varandra än elevtexterna och källtexterna, vilket kan spegla en mer etablerad och konsekvent genomförd stilistisk norm i dessa texter. Den högsta skillnaden i standardavvikelse finns mellan källtexterna, där värdena sträcker sig från knappt 59 till knappt 92. Bland elevtexterna finns det texter med OVIX-värden runt 50, såväl som texter med värden runt 75.

I lemmavariationsindex, LVIX (se tabell 2), mäts, till skillnad från i OVIX, andelen unika ord i en text; här är böjningsformerna sammanslagna till ett gemensamt ord. Måttet är relativt nytt och det finns därför inte så många jämförbara studier. Som nämntes i metodkapitlet kan det framför allt vara intressant att jämföra OVIX och LVIX inom texter och textgrupper. Ett för samma text lågt LVIX och högt OVIX bör tyda på att färre unika ord används oftare, i fler böjningsformer, i texten, vilket kan vara förenligt med en vetenskaplig stil med precisa termer och färre synonymer. Samtidigt leder många olika termer, vilket ofta krävs för att på djupet behandla ett ämne, till högre ordvariation; i det fallet behöver inte näraliggande LVIX- och OVIX-värden alls tyda på bristande

precision i form av synonymer, eller en ämnesmässigt hoppig text. Men närliggande LVIX- och OVIX-värden skulle alltså kunna tyda på att en hög andel av orden i en text är unika, vilket skulle kunna innebära en ämnesmässigt spretig text. Det bör dock påpekas att merparten av orden i en text inte utgörs av ämnesord, vilket förstås också inverkar på analysen.

Det är svårt att dra några tydliga slutsatser utifrån LVIX-värdena i studien. De följer övergripande samma mönster som OVIX-värdena (se tabell 2), det vill säga källtexterna ligger högst, därefter kommer de vetenskapliga artiklarna, och sedan elevtexterna i fallande betygsordning. Differensen mellan OVIX och LVIX är faktiskt, utifrån medelvärdena, högst i källtextsgruppen (ungefär 13 enheter), och lägre i övriga textgrupper (ungefär 10 enheter). Individuella texter uppvisar även i denna variabel mönster som avviker från gruppens medelvärden. Det vore intressant att göra en kvalitativ näranalys av ordanvändningen i ett fåtal texter med olika förhållanden mellan LVIX- och OVIX-värdena, för att närmare undersöka vad förhållandena mellan värdena kan stå för i faktisk ordanvändning. Detta har dock inte varit möjligt att genomföra inom ramen för denna studie.

5.4 Läsbarhetsindex

Läsbarhetsindex (LIX) grundar sig på två av de mått som redan behandlats: genomsnittlig meningslängd och ordlängd i form av andel långa ord i texten. Med långa ord avses ord med fler än sex bokstäver. Måttet utvecklades 1968 och är ett vanligt förekommande sätt att mäta svårighetsgraden i en text – ju lägre LIX-värde desto mer lättläst text, och tvärtom. Som nämndes i metodkapitlet är dock måttet ganska kritiserat och vissa forskare avråder från att använda det på elevtext (Megyesi, Palmér & Näsman 2019). Måttet säger ingenting om sådant som syntaktisk komplexitet och typ av ordförråd. Det kan dock ge en första, övergripande bild av en texts språkliga svårighetsgrad, och fungera som ett av flera verktyg vid jämförelser mellan texter.

Lixskalan för olika genrer är följande:

- under 25 = barnböcker
- 25–30 = enkla texter
- 30–40 = normaltext, skönlitteratur
- 40–50 = sakinformation
- 50–60 = facktexter
- över 60 = svåra facktexter, forskning, avhandlingar (Wikipedia, 2020).

Som förväntat har de vetenskapliga artiklarna det högsta genomsnittliga LIX-värdet, 53 (se tabell 2). Därefter kommer elevtexterna med betyg A (drygt 47). Elevtexterna med betyg C och E ligger ungefär lika, runt 44. E-texterna har till och med ett något högre LIX-värde än C-texterna. Som tidigare nämnts kan många satsradningar dra upp meningslängden, vilket kan vara en förklaring till ett

synbart högre LIX-värde i E-texterna. Det lägsta LIX-värdet i studien har källtexterna, 39, vilket var förväntat med tanke på att den genomsnittliga meningslängden är avsevärt lägre i denna textgrupp än i de övriga.

I betygssteg A och E är det genomsnittliga värdet ungefär två enheter högre för uppgiften *Särskrivning och stavfel*, än för *Skrivregler och sms* (se tabell 7 i bilaga 1). I betygssteg C är medelvärdet dock högre för uppgiften *Skrivregler och sms* (ungefär en enhet). Individuella variationer förekommer inom alla textgrupper. Standardavvikelsen varierar och visar inte på tydliga spridningsmönster i de olika textgrupperna (se tabell 8–13 i bilaga 1).

Rent allmänt kan iakttagelsen göras att LIX-värdena i de olika textgrupperna i studien, i förhållande till LIX-skalan ovan, är relativt låga (se tabell 8–15 i bilaga 1). Det är endast en vetenskaplig artikel som med ett LIX-värde på drygt 63 enligt skalan motsvarar vetenskaplig text (se tabell 15 i bilaga 1). Merparten av de vetenskapliga artiklarna har värden som enligt skalan innebär facktext, elevtexterna hamnar överlag inom kategorin sakinformation och källtexterna inom normaltext/skönlitteratur. Kanske är det så att LIX-skalan grundar sig på ett äldre språkbruk. Kanske är det den aktuella vetenskapliga disciplinen och ämnena som inverkar; språkbruk handlar ju i grunden om människor, vilket kanske medför en lite ledigare stil än resonemang och analys inom till exempel de naturvetenskapliga ämnena.

5.5 Bisatser och träd djup

Andel bisatser och träd djup är två syntaktiska mått som har medtagits i analysen för att sätta in ordanalysmått i ett större språkligt sammanhang. Måtten beskrivs i metodkapitlet, där det också påpekas att automatisk analys av syntax innehåller större felkällor än automatisk analys av ordförråd. Till exempel grundar sig den automatiska analysen av syntaktiska variabler på grafiska meningar, vilka inte alltid behöver sammanfalla med syntaktiska meningar. Ofullständiga satser förekommer både i elevtexterna och i källtexterna. Satsradningar är vanligt förekommande i elevtexterna. Båda dessa företeelser, liksom olika typer av grammatiska meningsbyggnadsfel, kan snedvrider den automatiska analysen av syntaktiska variabler.

Sambandet mellan andel bisatser och språklig komplexitet i en text är inte självklart. Att lära sig att använda bisatser i sitt skrivande innebär ett steg framåt i elevens språkutveckling och en högre andel bisatser kan tyda på en större språklig komplexitet. För att man ska kunna avgöra hur varierad och precis bisatsanvändningen är krävs dock en manuell analys av bisatserna i en viss text. Andelen bisatser är också relaterad till genre. Bisatserna är ofta fler i till exempel skönlitterär text och färre i sakprosa, där mycket av innehållet förmedlas av ordvalen (Strömquist 1999; Ribeck 2015). I vetenskaplig text finns mycket av den språkliga komplexiteten på ordvals- och frasnivå (Ribeck 2015), samtidigt som bisatser på

primär satsdelsnivå också används för att uttrycka olika typer av innehållsliga samband. Bisatser ingår också som attribut i fraser.

Genom variabeln genomsnittlig meningslängd framgick det att elevtexterna och de vetenskapliga artiklarna har ungefär samma meningslängd, runt 20 ord per mening (se tabell 2). Källtexterna har i genomsnitt avsevärt kortare meningar, runt 15 ord. När det gäller andel bisatser är det istället de vetenskapliga artiklarna och källtexterna som ligger närmast varandra, med i genomsnitt drygt en halv bisats per huvudsats, medan elevtexterna har en avsevärt högre andel bisatser, med ungefär lika delar bisatser och huvudsatser. Det finns vissa skillnader mellan de båda uppgifterna, men mönstret är inte tydligt (se tabell 7 i bilaga 1). Individuella skillnader finns naturligtvis, som inom alla textgrupper. Standardavvikelsen visar dock att de är minst inom gruppen vetenskapliga artiklar (se tabell 15 i bilaga 1), vilket även i denna variabel förmodligen tyder på en tydligare och mer konsekvent genomförd stilistisk norm än i övriga textgrupper.

Hur kan det då komma sig att andelen bisatser är ungefär densamma i källtexterna och i de vetenskapliga artiklarna, när meningarna i de senare i genomsnitt är avsevärt längre? Och hur kan andelen bisatser skilja sig åt så mycket i elevtexterna och de vetenskapliga artiklarna, när den genomsnittliga meningslängden är ungefär densamma i dessa textgrupper? I de vetenskapliga artiklarna måste det vara något annat än bisatser som utgör satsdelar, det vill säga något annat än bisatser på primär satsdelsnivå, som fyller ut meningarna. Rimligen måste orsaken vara längre fraser i de vetenskapliga artiklarna, och då främst nominalfraser (jfr avsnitt 5.6, variabeln nominalkvot). Attribut i nominalfraser bidrar med precision och nyanseringar av olika slag, vilket är viktiga inslag i vetenskaplig text.

Hur kan det då se ut i de aktuella texterna? Bland elevtexterna med betyg A och E, uppgiften *Särskrivning och stavfel*, är det elevtexterna A3 och E8 som har den högsta andelen bisatser. Några exempel på bisatser (understrukna) ges nedan:

Martin Borg skriver i sin artikel att särskrivningar i det svenska språket har funnits sedan länge och att de har retat folk enda sedan början på 1900-talet. (A3, *Särskrivning och stavfel*)

När man läser texter menar Ulf Teleman (vad är de för fel på fel?) att de finns tre faktorer man tar hänsyn till dessa är: tid, irritation och motivation, och då vi fel stavar samt särskrivningar rubbas detta vilket ledder till att texten tar längre tid att läsa och en irritation på grund av dessa brister väcks. (E8, *Särskrivning och stavfel*)

I det första exemplet förekommer två att-bisatser. I det andra återfinns sex bisatser, med en större variation av bisats-inledare och fler olika typer av bisatser. Här hade det dock varit önskvärt att eleven delat upp texten i fler grafiska meningar; den långa meningen ger ett lite pratigt intryck och inkluderar allt för mycket information. Dessutom är bisatsen då vi fel stavar samt särskrivningar felkonstruerad; för att den skulle utgöra en språkligt korrekt bisats hade eleven behövt skriva *då vi felstavar samt särskriver*. Nästan samtliga bisatser ovan utgörs

av bisatser som fungerar som egna satsdelar på primär satsdelsnivå. En av bisatserna, [*som*] *man tar hänsyn till* (E8, *Särskrivning och stavfel*) utgör attribut till huvudordet *faktorer*.

I följande exempel ur en av de vetenskapliga artiklarna förekommer dels en bisats som ingår som attribut i en satsdel [1], dels en bisats som utgör en satsdel [2]:

Synestesi är ett område som har ådragit sig stort intresse inom språkforskningen. [1] Synestesi innebär enkelt uttryckt att ett intryck från ett sinne används metaforiskt för att beskriva ett intryck från ett annat sinne, som i *varma färger* och *vassa ljud*. [2] (SoS 12)

Slutligen ges ett exempel på populärvetenskaplig stil med många huvudsatser, ur en av källtexterna:

Särskrivning av sammansatta ord har blivit allt mer uppmärksammat, och många känner sig kallade att bekämpa, eller åtminstone påpeka, åkomman. Av hävd har den mest syns på handtextade skyltar och dito restaurangmenyer. Lunchvalet mellan ägg röra och kebab tallrik var där inte alltid så enkelt. (Källtext 2)

I citatet ovan (källtext 2) förekommer ingen bisats.

Hultman och Westman (1977) kommer fram till att långa meningar och en hög andel bisatser *inte* korrelerar med höga betyg. Det är istället exakthet och koncentration, främst genom nominalfraser, som har ett samband med höga betyg. Hultmans och Westmans resultat bekräftas i stora drag i denna studie; den genomsnittliga meningslängden och den genomsnittliga andelen bisatser är högst i E-texterna (se tabell 2). A- och C-texterna ligger dock nära varandra i dessa variabler. En mer fullständig analys av förekomsten av bisatser i undersökningsmaterialet borde förstås inkludera en analys av bisatstyper – förekommer till exempel fler typer av bisatser i A-texterna än i E-texterna? – samt en analys av bisatsernas funktioner: utgör de till exempel oftare attribut i fraser i de vetenskapliga artiklarna än i elevtexterna och i källtexterna? Utifrån den automatiska analysen i studien (se även nominalkvot) finns det skäl att misstänka att så är fallet, och denna omständighet pekar i så fall på ett tydligt språkligt drag i vetenskaplig text i förhållande till andra texttyper, det vill säga behovet av koncentration och precision på frasnivå istället för tillägg av många enheter på samma nivå.

Variabeln trädjup uppvisar ett liknande mönster som andelen bisatser. Värdet är lägst i källtexterna, lite högre i de vetenskapliga artiklarna, och högst i elevtexterna (se tabell 2). Standardavvikelsen visar att de individuella skillnaderna är klart mindre inom gruppen vetenskapliga artiklar än inom övriga textgrupper. Eftersom trädjupet mäter avståndet mellan de syntaktiska enheterna i den överordnade satsen, från trädets så kallade rot till dess yttersta grenar, är det logiskt att en högre andel bisatser medför ett större trädjup.

Som redan påpekats innehåller den automatiska analysen av syntaktiska variabler avsevärt större felkällor än automatisk analys av olika ordförrädsvariabler.

Felkällorna blir förstås än större om de analyserade texterna i sig innehåller syntaktiska fel av olika slag, vilket är fallet med framför allt elevtexterna. Siffrorna måste därför ses som ungefärliga, men kan ändå bidra till att sätta in resultaten av ordförrådsanalyserna i ett vidare språkligt sammanhang.

5.6 Substantiv/pronomen-kvot och nominalkvot

I variabeln substantiv/pronomen-kvot mäts andelen substantiv i förhållande till andelen pronomen. En hög andel substantiv anses tyda på en mer skriftspråklig text och, eftersom ämnesord ofta utgörs av substantiv, också på ett rikare ämnesinnehåll. Omvänt tyder en hög andel pronomen på en mindre formell, och kanske mer talspråklig, text. Även variabeln nominalkvot baseras på förhållandet mellan vissa ordklasser i en text, men här är det förhållandet mellan substantiv och verb som är i fokus. I studien är det en så kallad utökad nominalkvot som har analyserats, vilket innebär att substantiv, prepositioner och particip har delats med verb, pronomen och adverb. Även här anses ett högre värde tyda på en mer skriftspråklig, eller nominal, stil och ett lågt värde på en mindre formell, eller verbal, stil.

Samtliga texter i studien utgörs av sakprosatexter, dock inom olika genrer. De vetenskapliga artiklarna är autentiska vetenskapliga texter; i dem presenteras resultat av olika forskningsstudier och de har bearbetats i flera led. Författarna är experter inom de ämnesområden de skriver om. Elevtexterna är texter ”av vetenskaplig karaktär”, där syftet är att eleverna ska öva sig i akademiskt skrivande. Eleverna använder sig av, och hänvisar till, källtexter som handlar om de ämnen de skriver om, men eleverna har förstås inte själva bedrivit forskning, och de är inte heller experter på de ämnen de skriver om. I källtexterna, slutligen, är syftet att på ett mer intresseväckande och lätt samt sätt presentera fakta och exempel inom olika ämnesområden relaterade till skrivande och språkliga normer.

I de båda variablerna substantiv/pronomen-kvot och nominalkvot är värdena avsevärt högre i de vetenskapliga artiklarna än i både elevtexterna och källtexterna (se tabell 2). De vetenskapliga artiklarna innehåller alltså en avsevärt högre andel substantiv och har en avsevärt mer nominal stil än vad som är fallet i övriga textgrupper. Resultatet är förväntat, och tydligt. Det speglar sannolikt omständigheten att de vetenskapliga artiklarna har ett rikare ämnesinnehåll med mer omfattande terminologi än övriga texter. Det speglar också att de vetenskapliga artiklarna tydligt följer en språklig norm, med en så kallad nominal stil. Heimann Mühlenbock (2013) presenterar nominalkvotsvärden i några olika genrer (ursprungligen från Melin & Lange), såsom broschyrer, läroböcker, tidningar och tal. Genrerna är inte direkt jämförbara med genrerna i denna studie, men det kan nämnas att värdena för texthäftet och elevtexterna befinner sig på ungefär samma nivå som olika typer av tidningar och tidskrifter (0,84–0,98), medan de vetenskapliga artiklarna med värdet 1,59 ligger avsevärt högre än läroböcker och broschyrer, som ligger på 1,18 respektive 1,19 (Heimann Mühlenbock 2013:45–46).

Skillnaderna mellan övriga textgrupper är avsevärt mindre än mellan dessa och de vetenskapliga artiklarna. Elevtexterna med betyg A ligger högst, det vill säga närmast de vetenskapliga artiklarna, vilket förefaller bekräfta att de elever som har skrivit dessa texter av lärarna anses ha lyckats bäst med den vetenskapliga stilen. Resultatet är i linje med Hultmans och Westmans resultat (1977), där de konstaterar ett samband mellan en mer nominal stil (fler verbalsubstantiv, färre pronomen) och högre betyg på elevtexter. Det bör som tidigare påpekas att betyget den aktuella terminen var ett sammanslaget betyg för alla de bedömda aspekterna i elevtexterna. Ett A som helhetsbetyg innebär dock att alla aspekter har bedömts som A.

C- och E-texterna har lägre genomsnittliga värden än A-texterna (se tabell 2). E-texterna har något lägre värden än C-texterna, men skillnaderna är små. Källtexternas genomsnittliga värden ligger på samma nivå som elevtexterna i betyg C och E, vilket är intressant; utifrån dessa värden är det alltså inte så att källtexterna innehåller varken en högre andel substantiv eller en mer nominal stil än elevtexterna; snarare förefaller källtexterna i dessa variabler, och utifrån medelvärden, ha en lite mer vardaglig stil än åtminstone A-elevtexterna.

Det finns vidare vissa skillnader mellan de båda uppgifterna. Gällande substantiv/pronomen-kvoten är de genomsnittliga värdena genomgående högre för uppgiften *Särskrivning och stavfel*, än för *Skrivregler och sms* (se tabell 7 i bilaga 1). Det är svårt att med säkerhet tolka de skilda värdena mellan uppgifterna. Kanske speglar de, som tidigare nämnts, att *Skrivregler och sms* kanske inbjöd till en lite personligare stil, medan *Särskrivning och sms* kanske uppfattades som en lite torrare uppgift där ämnesinnehållet hamnade ännu mer i fokus. Kanske fanns det fler ämnesord att tillgå gällande uppgiften *Särskrivning och sms*, utifrån källtexterna. Gällande nominalkvoten är värdena ungefär desamma för båda uppgifterna i betyg A. I betyg C och E är värdena något högre för uppgiften *Särskrivning och sms*.

Värdena för standardavvikelsen visar faktiskt på större individuell variation mellan de olika källtexterna och mellan de olika vetenskapliga artiklarna, än inom de olika elevtextgrupperna (se tabell 8–15 i bilaga 1). Bland källtexterna är det särskilt källtext 10 som avviker uppåt, med värden som ligger på ungefär samma nivå som de genomsnittliga värdena för de vetenskapliga artiklarna. Denna text utgörs av ett utdrag ur en riksdagsmotion och är alltså inte en populärvetenskaplig text. De skilda värdena mellan de vetenskapliga artiklarna är mer svårtolkade, bland annat eftersom det handlar om omfattande texter, där språket också kan skilja sig åt mellan olika delar i texterna, såsom mellan tidigare forskning och diskussion. Värdena kan ha att göra med de ämnen som behandlas, eller med delvis olika stilar mellan författarna, eller vara kopplade till vilken omfattning olika delar har i de olika texterna.

Vilka språkliga uttryck kan då skillnader i de aktuella värdena ta sig? Nedan återges de två första meningarna ur det andra stycket⁷ i de två elevtexter (uppgiften *Särskrivning och stavfel*) som har de högsta respektive de lägsta värdena för substantiv/pronomen-kvot och nominalkvot. Flerordiga nominalfraser har markerats med understrykning och huvudorden är fetstilsmarkerade.

Mycket av **språkvården** som bedrivs idag bygger på **antaganden** om vad som stör en läsare, snarare än erkända **sanningar**. I den klassiska **skriften *Språkrätt*** från 1979 har professor **Ulf Teleman** vid Lunds universitet beskrivit **diskussionen** i fråga och utifrån den tagit fram de allvarligaste **konsekvenserna** som regelrätta fel kan frambringa för läsaren av en text. (A10, *Särskrivning och stavfel*)

I **artikeln** ”Vad är det för fel på ett fel?” (Lars Melin, *Forskning & framsteg* 2007:1) Skriver Lars Melin om hur vi reagerar på olika **skrivfel** och hur det påverkar läsningen. Enligt en **studie** som använts i **artikeln** påstås det att det tar **22 % längre tid** att läsa **ett ord** som är felstavat än ett rättstavat. (E3, *Särskrivning och stavfel*)

I exemplen ovan kan vi bland annat se att elevtext A10 innehåller fler ovanliga och skriftspråkliga substantiv än elevtext E3. I elevtext A10 förekommer, som huvudord i nominalfraser (varav vissa utgör rektion i prepositionsfraser), substantiven *språkvården*, *antaganden*, *sanningar*, *skriften Språkrätt*, *professor Ulf Teleman*, *diskussionen*, *konsekvenserna* och *läsaren*. I E3 används på motsvarande sätt substantiven *artikeln*, *Lars Melin*, *skrivfel*, *läsningen*, *studie*, *tid* och *ord*. Fraserna är avsevärt längre i A10 än i E3 (se understrykningar, med huvudordet fetstilsmarkerat). Den elev som skrivit E3 väljer också att uttrycka sig mer genom personliga pronomen (mening 1: *vi*, *det*) medan eleven som skrivit elevtext A10 istället håller en mer opersonlig och formell stil.

Hur kan den nominala stilen ta sig uttryck i vetenskaplig text? Nedan ges ett exempel ur den vetenskapliga artikel som har högst genomsnittlig nominalkvot och näst högst substantiv/pronomen-kvot:

Metodiskt görs i den här **undersökningen** inte **någon skillnad** mellan kvalitativa och kvantitativa litterära **kretslopp** eller **någon uppdelning** i specifika **typer av litteratur** eller litterära **genrer** när **översättningsverksamheten** granskas inom de skandinaviska **länderna**. Sådana **studier** är planerade i **framtiden**. Här tillämpas i stället en **analysmodell** som bygger på spatiala **relationer**, det vill säga en **centrum/periferi-modell** för hur **översättningar** migrerar i **flöden** [...] (SoS 10:71)

I exemplet ovan förekommer många och specifika substantiv, varav flera utgörs av långa ord. Flertalet substantiv ingår i prepositions- eller nominalfraser (understrukna) och specificeras ytterligare av olika typer av attributiva bestämningar, som till exempel i *kvalitativa och kvantitativa litterära kretslopp* (adjektiviska

⁷ Eftersom det första, inledande stycket i elevtexterna ofta utgörs av likartade formuleringar, ofta hämtade från uppgiftsinstruktionen, valdes det andra stycket i texterna; det är i detta stycke som eleverna mer tydligt börjar formulera sin egen text, med eget språkbruk, efter att i inledningen ha introducerat ämnet och frågeställningen.

attribut) och en *analysmodell* som bygger på *spatiala relationer* (attribut i form av relativbisats). Verben står ofta i passiv form (*görs, granskas, tillämpas*). Även detta är ett typiskt stildrag i vetenskaplig text, och ett stildrag som automatiskt medför färre personliga pronomen.

I exemplet ur SoS 10 finns också en tydlig textbindning mellan meningarna, med ord och uttryck som syftar tillbaka på sådant som nämnts i föregående meningar: *Sådana studier [...], Här tillämpas i stället [...]*. Även detta är ett typiskt stildrag i vetenskaplig text, där relationer mellan olika enheter och led i analyser måste tydliggöras och kopplas samman. I populärvetenskaplig text finns inte samma krav på explicit uttryckta samband, vilket bland annat kan ta sig uttryck i korta huvudsatser med relativt många substantiv, där det snarare är ämnet i sig som utgör textbindningen. Följande exempel kan ges:

Givetvis skulle skriftspråket aldrig förändras om alla människor följde samma norm slaviskt hela tiden. Att medvetet göra "fel" kan vara en viktig del av utvecklingen i ett språk; det krävs alltid ett visst mod för att få till förändringar. Former som "har" och "blir" var i början av 1900-talet radikala experiment hos unga skribenter, som så småningom fick fler och fler efterföljare. (Källtext 1:4)

Framställningen i exemplet ovan är utan tvekan ämnesmässigt sammanhängande och lätt att följa, även om meningarna inte relateras till varandra med olika textbindande ord och uttryck. En mindre skicklig skribent kan dock riskera en ämnesmässigt hoppig och osammanhängande text när tydliga textbindningsmarkörer utelämnas. Detta kan många gånger bli fallet i elevtexterna, när eleverna i en tidsmässigt begränsad skrivsituation utifrån en given frågeställning ska referera och diskutera ämnesinnehåll ur några olika textutdrag.

Strömquist (1999) och Ribeck (2015) pekar på att läroboksspråk ofta utmärks av en koncentrerad, nominal stil med hög informationstäthet och med ganska dålig textbindning. Ribeck lyfter fram att det akademiska språket är mer avancerat än läroboksspråket, med bland annat fler och längre nominalfraser och med många meningar som inleds med långa, utbyggda subjekt. Att lära sig att skriva inom en ny genre har naturligtvis delvis att göra med språkliga förebilder. Eleverna i studien har kanske snarare populärvetenskapliga texter och lärobokstexter som språkliga förebilder, än faktiska vetenskapliga texter. Vidare är en vetenskaplig text, såsom redan påpekats, nästan alltid ett resultat av en längre tankeprocess och av bearbetningar i flera led, vilket alltså innebär en radikalt annorlunda skrivsituation än den som eleverna har under det nationella provet under kursen svenska 3. De förutsättningar och omständigheter som råder under det aktuella provet är förmodligen helt rimliga i förhållande till elevernas förkunskapsnivå, läroplanen, provsituationen och skolkontexten. Samtidigt utgör de begränsningar i elevernas möjligheter att kunna skriva texter som i större omfattning påminner om autentiska vetenskapliga texter.

6 Resultat från manuell analys

I kapitlet redovisas och diskuteras resultaten av de fördjupande manuella ordanalyserna i totalt 20 elevtexter med betyg A och E, uppgiften *Särskrivning och stavfel*. Analyskategorierna utgörs av ämnesord, centrala begrepp, akademiska ord och uttryck samt stilbrott, med underkategorier (se avsnitt 4.3.2). I slutet av kapitlet behandlas förhållandet till källtexterna. Kapitlets innehåll svarar mot studiens andra frågeställning.

6.1 Ämnesord

Så kallade ämnesord utgör en fundamental del av ordförrådet i alla sakprosatexter. Det är ämnesorden som bär upp textens innehåll. Andelen ämnesord i en text kan säga något om i vilken utsträckning en skribent fokuserar på ämnet. Vilka ämnesord som används kan säga något om precision och stringens, vilket är viktiga drag i vetenskaplig text.

Ämnesorden har i studien, utifrån uppgiftsinstruktionen för uppgiften *Särskrivning och stavfel*, delats in i två kategorier: språk och språkliga normer (2a) och läs- och skrivprocess, texter (2b). I den första kategorin ligger fokus på språket i sig. I den andra kategorin ligger fokus på arbetet med att skriva och läsoplevelsen av olika typer av texter, i förhållande till särskrivningar och stavfel. Ord som ingår i direkta citat, boktitlar och så vidare har ej inkluderats; analysen fokuserar i möjligaste mån på elevernas egna formuleringar, även om mycket av ordförrådet förstås har hämtats från källtexterna.

Att avgränsa ämnesorden från övriga ord, särskilt om de ingår i fraser eller längre formuleringar, är förstås inte helt enkelt. Antalet ord i de olika kategorierna måste därför ses som något ungefärligt. Gränsen mellan kategori 2a och 2b är dessutom inte alltid självklar; det förekommer att ord och formuleringar har inkluderats i båda dessa kategorier. Med dessa reservationer gjorda (se också avsnitt 4.3.2) kan en uppskattning av andelen ämnesord ändå bidra till en övergripande bild av det ämnesrelaterade ordförrådet i elevtexterna.

Som framgår i tabell 3 och 4 (se nästa sida) är andelen ämnesord i genomsnitt högre i A-texterna (27 procent av det totala antalet ord) än i E-texterna (22 procent av det totala antalet ord). Resultatet förefaller bekräfta att en högre andel ämnesord har en koppling till högre betyg. Variationerna mellan olika elevtexter är dock stora, även inom samma betygssteg. Generellt är andelen ämnesord i kategori 2a (språk och språkliga normer) något högre än andelen ämnesord relaterade till att läsa och skriva olika typer av texter (kategori 2b), men även i detta avseende finns det variationer mellan olika elevtexter.

Tabell 3. Ämnesord i elevtexter med betyg A, uppgiften *Särskrivning och stavfel*.⁸

Elevtext	Totalt antal ord	Ämnesord: kategori 2a		Ämnesord: kategori 2b		Alla ämnesord	
		Antal	Procent	Antal	Procent	Antal	Procent
A1	724	142	20	83	11	225	31
A2	849	127	15	138	16	265	31
A3	926	132	14	112	12	244	26
A4	849	114	13	139	16	253	30
A5	712	96	13	136	19	232	33
A6	781	124	16	58	7	182	23
A7	830	117	14	55	7	172	21
A8	771	125	16	89	12	217	28
A9	694	125	18	41	6	166	24
A10	814	94	12	85	10	179	22
Min–Max			12–20		6–19		21–33
Medelvärde			15		12		27

Tabell 4. Ämnesord i elevtexter med betyg E, uppgiften *Särskrivning och stavfel*.⁹

Elevtext	Totalt antal ord	Ämnesord: kategori 2a		Ämnesord: kategori 2b		Alla ämnesord	
		Antal	Procent	Antal	Procent	Antal	Procent
E1	719	134	19	131	18	265	37
E2	412	57	14	36	9	93	23
E3	551	72	13	37	7	109	20
E4	602	68	11	73	12	141	23
E5	707	127	18	83	12	210	30
E6	885	122	14	102	12	224	25
E7	797	63	8	33	4	96	12
E8	575	59	10	52	9	111	19
E9	802	70	9	32	4	102	13
E10	672	121	18	27	4	148	22
Min–Max			8–19		4–18		12–37
Medelvärde			13		9		22

Vilka ämnesord kan då ingå i kategori 2a och 2b, i några olika elevtexter? I tablå 1 och tablå 2 nedan visas ämnesorden ur de två A-texter som innehåller högst andel ämnesord i respektive kategori, samt ämnesorden i den E-text som innehåller lägst andel ämnesord i båda kategorierna. Både enstaka ord, fraser och satser har noterats. De har placerats i alfabetisk ordning i tablåerna. I de fall ett ord eller uttryck

⁸ Antal ord utgörs av antalet ord i elevtexterna vid den automatiska analysen. Referenser inom parentes är ej medtagna (se även avsnitt 4.3.1.2). Avrundning till hela procent.

⁹ Se föregående fotnot.

förekommer flera gånger anges detta med en siffra inom parentes. Versal i början av ord och uttryck har ändrats till gemen.

Tablå 1. Exempel på ämnesord i kategori 2a i elevtexter, uppgiften *Särskrivning och stavfel*.

Kategori 2a. Språk och språkliga normer	
Elevtext A1	Elevtext E7
andelen särskrivningsfel	att stava rätt
betoningskillnader	den svenska språkjournalisten Martin Borg
betydelseskiljande	dessa smittsamma särskrivningar
dagens språkliga norm	dessa stavfel
den svenska stavningen	det felstavade ordet
det svenska skriftspråket	det svenska språket (2)
det svenska språket	detta språkfel
det vanligaste språkelet	detta särskrivande
dialekter	detta väldiga delande på sammansatta ord
distinktion	felstavning (2)
en betydelsefull distinktion	några futtiga stavfel
en norm (2)	ordet
en norm som bara ändrats i grunden en gång	språkforskare
en omfattande stavningsreform	stavar
en språklig standard	stavfel (8)
en standard	särskrivning (8), särskrivnen, särskrivningen, särskrivningar
ett gemensamt skriftspråk	
ett medvetet stilgrepp	
felstavningar, felböjningar eller felaktiga ändelser	
man stavade ord som man tyckte att de lät	
normen	
normen som utvecklades på tryckerierna	
ord kan ändra betydelse	
ordet	
sammanblandningar av ord	
sammanblandningar av ord som "chans" och "risk"	
språket (2)	
språket blir mindre precist	
språkliga fel	
stavat lite som de ville	
stavfel (7)	
stavfel och felböjningar	

stavningsregler och grammatik

svenskan (2)

särskrivning, särskrivningen (3), särskrivningar (3),
särskrivningarna

särskrivningen av ord som egentligen ska vara
sammansatta

tre olika språkfel

två stavfel

viktiga betydelskillnader

vissa ords stavning

Tablå 2. Exempel på ämnesord i kategori 2b i elevtexter, uppgiften *Särskrivning och stavfel*.

Kategori 2b. Läs- och skrivprocess	
Elevtext A5	Elevtext E7
att datorerna följer en rättstavningsfunktion som saknar fallenheten att veta när en sammansättning är läglig	budskapet (4)
att innebörden går förlorad	budskapet i en text (2)
att läsa	en text
att tappa förtroendet för en text	flytet i texten
budskapet	förståelsen för texten
böcker	informationen som skall läsas
börjar läsaren slarvläsa	läsare, läsaren (3)
de som har författat texten, har ett budskap och ett syfte med det som de skrivit	texten (2), texter
effekten av stavfel	verklig läsning
en seriös skribent	
en text (2)	
en text, där budskapet gått förlorat	
fokus dras bort från budskapet och kontentan ändras	
förtroendet för texten sjunker	
gå igenom din text en gång till	
hela texten	
irritation och minskad motivation att läsa texten	
leda till att budskapet ändras	
läsa det du skrivit	
läsningen (2)	
man skriver som det låter	
man tappar förtroendet för texten	
meddelandet	
oseriös skribent	
skribenten framstår som oseriös	
skriva samma fel	
skrivstilens begränsningar	
skyltar	
slarvläsa	
SMS	
texter	
tidningar	
vi läser varje dag	
våra kommunikationssätt	

I jämförelsen mellan ämnesorden i de A-texter som har de högsta andelarna inom respektive kategori och den E-text som har den lägsta andelen ämnesord i kategori 2a och bland de lägsta i kategori 2b, blir tre saker tydliga: A-texterna innehåller i förhållande till E-texten fler olika och mer varierade ämnesord, de innehåller fler längre nominalfraser med bestämningar till ett ämneshuvudord och de innehåller även fler längre formuleringar som inkluderar ämnesord. Inom kategori 2a, språk och språkliga normer, förekommer till exempel i texten A1, vid sidan av termer för *särskrivning* och *stavfel*, ord som *stilgrepp*, *stavningsreform*, *dialekter*, *felböjningar*, *distinktion*, *betoningsskillnader* och *betydelseskilnader*, medan ämnesorden i kategori 2a i elevtext E7 i mycket hög grad utgörs av just orden *särskrivning* och *stavfel*, med vissa variationer. Elevtexterna A1 och A5 innehåller också fler längre och mer varierade nominalfraser, till exempel *en betydelsefull distinktion*, *särskrivningen av ord som egentligen ska vara sammansatta* och *[e]n text där budskapet gått förlorat*, medan de nominalfraser i E7 som innehåller bestämningar till huvudord överlag är korta och inte så originella, till exempel *det svenska språket* och *budskapet i en text*. Tre av de lite längre eller mer speciella nominalfraser som förekommer i E7 innehåller dessutom stilbrott i form av en mer talspråklig eller populärvetenskaplig stil: *detta väldiga delande på sammansatta ord*, *några futtiga stavfel* och *dessa smittsamma särskrivningar*, där ordet *vældiga* kan anses talspråkligt och *futtiga* och *smittsamma* populärvetenskapliga. Det förekommer vidare ett antal längre formuleringar (satser) i A1 och A5 som innebär preciseringar av ämnesinnehåll, till exempel *ord kan ändra betydelse* (A1) och *att innebörden går förlorad* (A5). Sådana satser har inte hittats i E7.

Elevtexterna A1 och A5 har enligt den automatiska analysen högre värden för nominalkvot än E7, det vill säga de har en mer nominal stil än E7. Bland ämnesorden förekommer nominaliseringar som *felstavningar*, *felböjningar* (A1) och *skrivstilens begränsningar*, *effekten av stavfel*, *minskad motivation [att läsa texten]* (A5). Möjliga nominaliseringar i E7 förefaller avsevärt färre, och verkar mest utgöras av centrala termer (*särskrivning*, *felstavning*).

Sammantaget visar analysen av ämnesord att A-texterna generellt innehåller en högre andel ämnesord än E-texterna. De fördjupande exempel som behandlats indikerar vidare att A-texterna innehåller mer variation och precision i ämnesorden, i form av fler unika ämnesord och fler längre uttryck där ämnesord preciseras (nominalfraser, satser). Resultatet är i linje med Hultman och Westman (1977:178) som konstaterar att det i gymnasieuppsatser med högre betyg förekommer fler ”ord med rikare semantisk specifikation”.

Det är rimligt att anta att eleverna hämtar merparten av ämnesorden från uppgiftsinstruktionen och från källtexterna, eftersom det är just källtexterna som utgör elevernas fakta på området; de skriver om ett ämne som de själva saknar specialkunskap om, även om både särskrivningar och stavfel väl får sägas vara ganska vanligt förekommande begrepp i diskussioner av språkfel, inte minst i skolan. Att A-texterna överlag innehåller fler och mer varierade ämnesord än E-texterna

kan således tyda på att de elever som har skrivit A-texterna i högre utsträckning tagit till sig och utnyttjar källtexterna i sin egen text, än de elever som har skrivit E-texterna. Det har inte inom ramen för studien varit möjligt att i någon större utsträckning jämföra ämnesorden i elevtexterna och i källtexterna (se dock jämförelsen mellan centrala begrepp i elevtexter och källtexter i avsnitt 6.2).

Elevernas användning av källtexterna kommenteras vidare i avsnitt 6.6.

6.2 Centrala begrepp

I en vetenskaplig studie är de centrala begreppen inom det ämnesområde som behandlas särskilt viktiga. Att dessa är väldefinierade och används på ett precist och konsekvent sätt är fundamentalt för analysens kvalitet. Elevtexterna i det nationella provet i kursen svenska 3 innebär förstås endast en övning i vetenskapligt skrivande; det är texter av så kallad vetenskaplig karaktär som ska skrivas, och eleverna gör inga egna forskningsstudier utan återger och diskuterar några olika skribenters inlägg på området, utifrån en given frågeställning. Med reservation för dessa omständigheter kan det ändå vara intressant att se hur eleverna förhåller sig till vissa centrala begrepp i sin text. Är benämningarna precisa och konsekventa, eller förekommer växlingar? Finns det skillnader i dessa avseenden mellan A- och E-texterna?

Orden i kategori 1 ingår, som nämndes i metoddelen (avsnitt 4.3.2), även i ämnesorden (kategori 2a och 2b). Begrepp och termer är ju naturligtvis också ämnesord. I förhållande till övriga ämnesord är de dock mer specialiserade – jämför facktermer (OrdiL kategori D). Det finns ett antal ord i elevtexterna som skulle kunna betecknas som facktermer. Som nämndes i metoddelen (4.3.2) har två begrepp från början valts ut för denna del av analysen. Det är de två i uppgiften mest centrala begreppen, det vill säga de som nämns i rubriken till uppgiften och som betecknar de två språkliga företeelser som texten ska handla om: *särskrivning* och *stavfel*. Därutöver har analysen kompletterats med begreppet *rättstavning* (se nedan). Analysen är tidskrävande och det har därför varit nödvändigt med en tydlig avgränsning till ett fåtal begrepp, även om det hade varit intressant med en mer fullständig kartläggning av elevernas användning av facktermer.

I uppgiftsrubriken ingår alltså termerna *särskrivning* och *stavfel*. I uppgiftsinstruktionen nämns även *skriva isär [sammansatta ord]*, det vill säga *särskrivning* nämns både som nominalisering (företeelsen i sig) och som verbal process (*att skriva isär ord*). Det bör också påpekas att *särskrivning* rent språkligt både kan beteckna företeelsen i sig, och ett faktiskt fall av särskrivning (*en särskrivning, flera särskrivningar*). I uppgiftsrubriken förefaller den rimliga tolkningen vara att det är företeelsen i sig som avses genom singularformen. Detta innebär i sig en diskrepans i förhållande till rubrikens *stavfel*, som betecknar faktiska fall av felaktig stavning, och inte företeelsen i sig. En benämning motsvarande *särskrivning* hade kunnat vara *felstavning*, men det är alltså inte en

sådan som ges i rubriken, utan begreppen i rubriken befinner sig på olika abstraktionsnivå. Vidare kan noteras att singular- och pluralformerna av *stavfel* sammanfaller – *ett stavfel, flera stavfel*.

Analysen har kompletterats med begreppet *rättstavning*, som också nämns i uppgiftsinstruktionen (*stava rätt*). *Rättstavning* och *felstavning* utgör motsatta underbegrepp till huvudbegreppet *stavning*, och det kunde vara intressant att se om elever använde sig av båda dessa former, av bara den ena eller om de istället valde att beskriva dem med verben (*stava rätt, stava fel*). *Rättstavning* inkluderades också eftersom det kunde vara intressant att undersöka om eleverna i de analyserade elevtexterna själva valde att nominalisera ett begrepp som inte redan var nominaliserat i uppgiftsinstruktionen.

Analysen av de centrala begreppen i elevtexterna med betyg A och E utgår alltså från *särskrivning* och *stavfel*, samt *rättstavning* (i uppgiftsinstruktionen *stava rätt*). Alla alternativa benämningar som kan sägas beteckna dessa företeelser inkluderas. Analysen utgår från de ämnesord som urskilts i kategori 2a (språk och språkliga normer). Ordet *fel* har inte inkluderats, eftersom det kan syfta på många olika typer av språkliga fel. Inte heller i förhållande till *särskrivning* angränsande eller motsatta begrepp har inkluderats, såsom *sammansättningar* eller *hopskrivningar*. Dessa nämns inte heller i uppgiftsinstruktionen. Artiklar, räkneord och så vidare har inte noterats i sammanställningen; det kan alltså till exempel vara så att ett fall av *särskrivning* avser både företeelsen i sig och ett faktiskt fall av *särskrivning*. *Sammansättningar* med *särskrivning* och *stavfel* har inkluderats, liksom fraser eller uttryck som betecknar dessa företeelser.

I tablå 3 på nästa sida presenteras de olika termer, med böjningsformer, samt uttryck, som sammantaget används av eleverna för att beteckna de centrala begreppen *särskrivning*, *felstavning* och *rättstavning*.

Tablå 3. Förekomst av de centrala begreppen *särskrivning*, *felstavning* och *rättstavning* i elevtexter med betyg A och E: termer och böjningsformer.¹⁰

Begrepp	Termer, böjningsformer och uttryck i A-texter	Termer, böjningsformer och uttryck i E-texter
<i>Särskrivning</i>	särskrivning–särskrivningen– särskrivningar–särskrivningarna– särskrivningarnas särskrivningsfel särskrivnings fall särskrivningsproblem–särskrivnings- problemet särskrivningssjukan sär- eller hopskrivning skriva isär–skriver isär–skrivs isär– skrivits isär särskriva–särskriver–särskrivs– särskrivas–särskrivit särskrivna	särskrivning–särskrivningen– särskrivningar–särskrivningarna särskrivningen särskrivningar särskrivningsproblemet skriva isär särskriva–särskriver–särskrivs vilka ord som ska skrivas tillsam- mans och vilka inte, vilken ord skrivs tillsammans och vilken separerad särskriven–särskrivet–särskrivna detta särskrivande, detta väldiga delande på sammansatta ord
<i>Felstavning</i>	stavfel–stavfelen felstavning–felstavningar–felstav- ningens felaktiga stavningar stavnings- och slarvfel stavfelsproblemet rättstavningsfel stava fel–stavar fel stavat lite som de ville man stavade ord som man tyckte att de lät felstavade–felstavat stavats fel	stavfel–stavfels–stavfelet–stavfelen felstavning–felstavningar fel stavning stavproblem stavfelsproblem särstavningar stava fel felstavar fel stavar felstavat–felstavade valfritt stavande ord
<i>Rättstavning</i>	rättstavning rättstavningsregler stava rätt rättstavat–rättstavade korrekt stavat den visuella stavningen, den korrekta	rättstavningsregel–rättstavningsregler stava rätt–stavar rätt ska stavas rättstavat

¹⁰ Samtliga citat ur elevtexterna återges ordagrant, inklusive eventuella språkfel.

Det är förstås stora skillnader i hur ofta de olika termerna och deras olika böjningsformer används av eleverna. *Särskrivning*, i olika böjningsformer, och *stavfel* står för en hög andel av det totala antalet förekomster. Övriga substantiv, verb och particip förekommer mer sparsamt. De mer speciella uttrycken i tablå 3 förekommer enstaka gånger.

Det är stora skillnader i termanvändning mellan olika elevtexter. Vissa elever använder flera olika termer för de centrala begreppen, i olika böjningsformer, medan andra endast använder ett fåtal termer och kanske nästan inga böjningsformer av dessa. Två exempel kan ges, på elever som använder få respektive många olika termer och böjningsformer (antal förekomster inom parentes):

- elevtext A5: särskrivning (4), särskrivningen (1), särskrivningar (7), stavfel (10), felstavade (1), rättstavade (1)
- elevtext A3: särskrivning (3), särskrivningar (16), särskrivningarna (1), särskrivna (1), stavfel (2), felstavning (2), felstavningar (8), felstavat (3), felstavade (1), stava rätt (3), rättstavat (2).

I tabell 5 finns en sammanställning av det totala antalet förekomster av de centrala begreppen i A- och E-texterna, fördelat på substantiv, verb och adjektiv. *Felstavning* och *rättstavning*, vilka ofta behandlas tillsammans i elevtexterna, har slagits samman i översikten.

Tabell 5. Antal förekomster av de centrala begreppen i elevtexter med betyg A och E, uppgiften *Särskrivning och stavfel*.

Begrepp	Elevtexter betyg A		Elevtexter betyg E	
	Antal	Procent	Antal	Procent
<i>Särskrivning</i>				
varav substantiv	135	87	117	85
varav verb	14	9	14	10
varav particip	6	4	6	4
Totalt	155	100	137	100
<i>Felstavning och rättstavning</i>				
varav substantiv	112	73	83	72
varav verb	19	12	23	20
varav particip	23	15	10	9
Totalt	154	100	116	100

Sammantaget framstår skillnaderna i termanvändning som större mellan enskilda elevtexter (antal typer och antal förekomster) än mellan betygsgrupperna. Det förefaller svårt att baserat på denna begränsade undersökning dra slutsatser om termanvändning relaterat till betyg, även om mindre skillnader som kan relateras

till betyg möjligen är märkbara. Antalet förekomster av termer som betecknar de centrala begreppen i delstudien är något högre i A-texterna. Eftersom A-texterna generellt är något längre än E-texterna behöver detta dock inte innebära att termerna används mer frekvent i A-texterna, i förhållande till antalet löpord. Dock är den procentuella andelen totala ämnesord högre i A-texterna (se avsnitt 6.1), vilket inkluderar de centrala termerna, så det skulle kunna vara så att de centrala termerna används något oftare i A-texterna.

Språkliga fel och lite luddiga formuleringar som ersättning för precisa termer är något vanligare i E-texterna än i A-texterna (se tablå 2). Resultatet är i linje med betygskriterierna, där precision i ordval och uttrycksformer är ett kriterium för betyg A. Antalet elevtexter och antalet exempel är dock för litet för att säkra slutsatser ska kunna dras på detta område.

Gällande begreppen *felstavning* och *rättstavning* är den procentuella förekomsten av verb högre i E-texterna, medan förekomsten av particip är högre i A-texterna. Resultatet är i linje med tidigare forskning (till exempel Hultman och Westman 1977) som visar att en nominal stil dels utmärker akademisk text i allmänhet, dels visar att en mer nominal stil leder till högre betyg i elevtexter. Antalet förekomster är dock så lågt att det är svårt att dra några större slutsatser baserat på delstudien.

I uppgiftsinstruktionen nämns alltså *särskrivning*, *skriva isär*, *stavfel* och *stava rätt*. I de tre källtexter som är mest använda av eleverna (se avsnitt 6.6) förekommer *särskrivning* (företeelsen), *särskrivna ord* och *felstavning* (källtext 1, Lindström), *särskrivning* (om företeelsen och om enstaka exempel, i olika böjningsformer), *särskrivningssjukan*, *särskrivningsproblemet*, *särskrivningsfrekvensen*, *stavningsproblem*, *stavfel* och *rättstavningsfunktion* (källtext 2, Borg) och *stavfel*, *rättstavat* och *felstavat* (olika böjningsformer) (källtext 3, Melin).

Eleverna använder i hög grad de termer som förekommer i uppgiftsinstruktionen och i källtexterna, vilket är naturligt. Källtext 2 (Borg) är tydligt populärvetenskaplig, med uttryck som *särskrivningssjukan*. Även sådana uttryck förekommer alltså i elevtexterna (se tablå 2).

6.3 Akademiska ord och uttryck

Akademiska ord och uttryck är, vid sidan av ämnesorden, en fundamental del av det ordförråd som visar att en text tillhör den akademiska texttypen. De har i studien delats in i tre olika underkategorier: allmänvetenskapliga ord och uttryck (avsnitt 6.3.1), referatmarkörer (avsnitt 6.3.2) och övriga akademiska textbindningsmarkörer (avsnitt 6.3.3).

6.3.1 Allmänvetenskapliga ord och uttryck

Med allmänvetenskapliga ord och uttryck avses sådana ord och uttryck som syftar på den vetenskapliga analysen och produkten. Följande ord och uttryck förekommer i elevtexterna, i underkategorierna nedan:

- **delar av den vetenskapliga processen:** frågeställning, fråga, tes, hypotes, teori, metod, dra slutsatsen att, slutsats
- **empiri:** sammanställning, belägg, underlag, utredning, undersökning, exempel, resultera, resultat, experiment, försöksperson, studie
- **orsaker och orsakssamband:** faktor, orsak, konsekvens, effekt, förklaring, påstående, antagande, anledning
- **diskussion och analys:** diskutera, diskussion, debatt, jämförelse, konstatera, problematik, problem, kontentan, argument, synvinkel, jämföra, likhet, skillnad, distinktion
- **delar av den vetenskapliga texten:** inledning, avslutning.

Antalet allmänvetenskapliga ord och uttryck varierar i elevtexterna mellan något enstaka fall till något tiotal. Allmänvetenskapliga ord och uttryck är i genomsnitt något fler i A-texterna (cirka 6 per text) än i E-texterna (cirka 4 per text). Även med hänsyn till att A-texterna generellt är lite längre än E-texterna är andelen allmänvetenskapliga ord och uttryck något högre i A-texterna. Det handlar dock i alla elevtexter om enstaka ord och uttryck, men de är viktiga för att signalera att det är fråga om texter av så kallad vetenskaplig karaktär.

6.3.2 Referatmarkörer

Att hänvisa till andras forskning, bakgrundsfakta och så vidare är en fundamental del av vetenskaplig text. I de aktuella elevtexterna är dessutom en stor del av uppgiften att återge innehållet i några olika källtexter om det ämne som behandlas. Referenser består vanligen av en referatmarkör, det vill säga en introduktion av själva referensen, ofta genom användning av skribentens efternamn, tillsammans med ett verb som beskriver vad personen gör (till exempel *Lindström menar...*), och en slutparentes, där efternamn, årtal och ofta sidnummer anges (till exempel *Lindström 2008:73–75*). Eftersom de källtexter som eleverna hänvisar till vanligen är mycket korta finns det inte så stora behov av att ange sidnummer. De så kallade slutparenteserna är ganska få i elevtexterna. Jag har därför valt att fokusera på referatmarkörerna, vilka är fler och där möjligheten till språklig variation är avsevärt större. De är också på ett annat sätt del av själva texten, och en del av elevens eget språk.

Två olika typer av referatmarkörer har urskilts i studien: *med efternamn* (det vill säga efternamn och verb, alternativt *enligt* och efternamn) och *övriga* (ofta med själva källtexten som utgångspunkt, kan inkludera efternamn). Exempel ges nedan:

- **med efternamn:** Borg skriver/menar/påpekar/ger exemplet/presenterar/hänvisar till/belyser/sammanfattar/lägger fram/drog slutsatsen att...; enligt Borg...
- **övriga (ofta med artikeln som utgångspunkt):** i artikeln står det...; Martin Borgs budskap i artikeln...; jag håller med Lindström...

I tabell 6 sammanfattas det genomsnittliga antalet referatmarkörer i A-texterna och i E-texterna.

Tabell 6. Genomsnittligt antal referatmarkörer i elevtexter med betyg A och E, uppgiften *Särskrivning och stavfel*.

Elevtexter	Referatmarkörer med efternamn (medelvärde)	Unika referatmarkörer med efternamn (medelvärde)	Övriga referatmarkörer (medelvärde)	Samtliga referatmarkörer (medelvärde)
Betyg A	11,7	7,0	2,5	14,2
Betyg E	7,4	5,1	2,7	10,1

Sammantaget visar studien av referatmarkörer i elevtexterna att dessa är fler och mer varierade i A-texterna än i E-texterna. I den större referatmarkörskategorin (efternamn och verb, alternativt *enligt* och efternamn) använder de elever som skrivit A-texterna i genomsnitt fler unika uttryck än E-eleverna, vilket kan tyda på en större precision och en mer nyanserad framställning när olika aspekter ur källtexterna återges.

E-texterna innehåller vidare fler felaktigheter och klumpiga formuleringar gällande referatmarkörerna än A-texterna. De innehåller oftare dåtid (till exempel *har framfört*, *kom fram till*), felaktiga konstruktioner (till exempel *hävdar på*, *menar på*, *påpekar på*, *ställde sig själv en tes*), talspråk (till exempel *är inne på*), och möjligen oftare bristande precision (till exempel *I skriften påstår Teleman [...]*).

6.3.3 Övriga akademiska textbindningsmarkörer

Med *övriga akademiska textbindningsmarkörer* avses ord och uttryck som används för att föra den vetenskapliga analysen framåt i texten – relatera olika infallsvinklar till varandra, sammanfatta en diskussion och så vidare. Det handlar

om ord och uttryck som *dock*, *emellertid*, *därtill*, *däremot*, *vidare*, *även*, *som ovan nämnts* och *för att sammanfatta*. Eftersom redan vissa ord och uttryck i kategori 3a (allmänvetenskapliga ord och uttryck) kan innebära en viss textbindning (till exempel sammanfattning, avslutning) kallas kategorin för **övriga** akademiska textbindningsmarkörer. Till skillnad från ord och uttryck i kategori 3a har orden och uttrycken i kategori 3c primärt en textbindande funktion, medan det i kategori 3a är det vetenskapliga innehållet som primärt är i fokus.

Det bör påpekas att orden och uttrycken i denna kategori (3c) inte endast förekommer i akademisk text. Till skillnad från orden och uttrycken i kategori 3a och 3b är de mer allmänt förekommande i formellt skriftspråk. De är dock vanliga i akademisk text, där de bidrar till den akademiska stilen, varför det är relevant att inkludera dem i studien.

Det är inte helt lätt att göra en kvantitativ jämförelse mellan elevtexterna avseende förekomsten av övriga akademiska textbindningsmarkörer. Textbindningen kan sträcka sig över flera meningar, vilket gör det svårt att avgränsa antalet textbindningsmarkörer i en viss elevtext. Textbindning är ju dessutom ett i högsta grad kvalitativt språkdrag, som måste analyseras och bedömas i sitt sammanhang. En övergripande uppskattning visar dock att antalet fall av textbindning sammantaget är lite fler i A-texterna, men att de individuella variationerna mellan elevtexterna är stora både mellan A-texterna och mellan E-texterna. Några exempel på textbindande uttryck i en A-text och två E-texter kan ges (textbindningsmarkörerna i fetstil):

Särskrivning är **däremot** inte det enda brottet mot skrivnormer som påverkar läsningen. **Även** stavfel kan ställa till problem. [...] **sammanfattningsviss** tre faktorer [...] **Dock** finns det **även andra typer** av särskrivningar som leder till värre konsekvenser. De omedvetna reparationerna tar **dock** tid [...] **dock** kan vissa fel [...] Om man **jämför** [...] Det gör **däremot** [...] **Sammanfattningsvis** [...] (A4, *Särskrivning och stavfel*)

En **ganska liknande** teori presenteras i språkriktighetsboken [...] **Men** när det kommer till **andra** särskrivningar [...] (E3, *Särskrivning och stavfel*)

[...] **däribland** [...] **Det här** förklarar Lars Melin [...] **Trots** de förändringar SAOL genomförde [...] **I det här** fallet [...] **När** man pratar om stavfel [...] **En annan typ** av stavfel [...] Engstrand menar **även** [...] Särskrivningar är **tillika** vanligt förekommande [...] **I vissa fall** kan [...] **I första meningen** handlar det om [...] **Samtidigt** ger **andra meningen** [...] **Jämför** man [...] **På liknande vis** menar Martin Borg [...] **Sådana** fel gör att [...] Det kan **också** vara svårt att veta [...] **Här** är det oklart om [...] **Parallellt** menar Lindström [...] **Därför** är det minst lika viktigt [...] (E10, *Särskrivning och stavfel*)

Exemplen ovan, som kan anses någorlunda representativa för de undersökta elevtexterna, visar dels att det är ganska stora skillnader mellan enskilda elevtexter i hur mycket textbindning de innehåller, dels att de textbindande uttrycken kan vara mer eller mindre elegant och korrekt formulerade. E-texterna innehåller sammantaget många fler exempel på klumpiga och delvis inkorrekta textbindande uttryck

(till exempel [*e*]n ganska liknande teori), liksom fler exempel på populärvetenskapliga ordval (till exempel [*d*]et här förklarar Lars Melin, [*n*]är man pratar om stavfel; jämför avsnitt 6.4, om stilbrott).

6.3.4 Akademiska ord och uttryck: sammanfattande kommentarer

Analysen av akademiska ord och uttryck i de tio A-texterna och de tio E-texterna, uppgiften *Särskrivning och stavfel*, visar att eleverna på detta område har en egen språklig repertoar. Akademiska ord och uttryck används inte i någon större omfattning i källtexterna, som många gånger är populärvetenskapliga. Enstaka uttryck (*presentera frågeställningen, beskriv, jämför, jämförelsen*) ges i uppgiftsinstruktionen, men huvuddelen av de allmänvetenskapliga orden och uttrycken, de olika referatmarkörerna och de övriga akademiska textbindningsmarkörerna som förekommer i elevtexterna måste eleverna ha lärt sig i andra sammanhang. Delvis kan de ha lärts in i undervisning inför det aktuella provet, men de har säkert också dykt upp i andra undervisningssammanhang under kursen svenska 3 och i kurser i andra ämnen, liksom under gymnasiearbetet.

Resultaten visar vidare på tydliga skillnader mellan A- och E-texterna. A-texterna innehåller sammantaget något fler allmänvetenskapliga ord och uttryck, fler och mer varierade referatmarkörer, samt något fler övriga akademiska textbindningsmarkörer. I E-texterna finns det vidare fler exempel på att ord och uttryck inom de undersökta kategorierna är inkorrekt, klumpigt eller talspråkligt formulerade. Det bör dock påpekas att de individuella variationerna mellan elevtexterna är stora.

Analysen av akademiska ord och uttryck bekräftar tidigare forskning om nationella provet i svenska på gymnasiet. Ohlsson (2001) visar att bedömningen av elevtextens språk och stil har en hög korrelation med elevtextens betyg. Det är rimligt att anta att just förekomsten av akademiska ord och uttryck, som är viktiga för att markera att en text tillhör den vetenskapliga texttypen, och hur dessa används, spelar en stor roll för bedömningen av de aktuella elevtexterna. A-texterna innehåller fler och mer varierade akademiska ord och uttryck än E-texterna, och färre fel på området. Även Östlund-Stjärnegårdhs (2006) resultat som visar på ett samband mellan många källhänvisningar och höga betyg, bekräftas i denna studie. Garne och Strömquist (1980) ser vidare ett samband mellan god textbindning och höga betyg, vilket alltså även resultaten i denna studie indikerar, även om det är fråga om en tendens som skulle behöva undersökas mer djupgående, och i fler elevtexter.

6.4 Stilbrott

Med stilbrott avses ord, uttryck, formuleringar och meningsbyggnad som avviker från en akademisk stil. Stilbrott är inte samma sak som språkliga fel, utan det

handlar, med Hultmans och Westmans (1977) definition, om avvikelser från en jämn stilnivå. Kategorin har inkluderats utifrån hypotesen att stilbrott leder till lägre betyg, och därför skulle kunna vara fler i E-texterna än i A-texterna.

Ett antal olika typer av stilbrott har identifierats i elevtexterna. De sammanfattas och exemplifieras nedan:

- **talspråkliga ordval:** en *mängd* texter, *kolla* på stavfel
- **talspråkliga formuleringar:** Det här tycker vi är helt okej [...], Det blir mer jobb för läsaren [...]
- **utelämnat att:** Den frågeställningen jag kommer [att] ta reda på [...]
- **man = bristande precision:** När *man* pratar om stavfel [...]
- **eget tyckande:** Jag tycker [...]
- **satsradningar:** Stavar fel gör vi alla någon gång, det går inte att undvika.
- **ofullständiga meningar och meningsfragment:** Vilket innebär att man skriver isär sammansatta ord., Frivilligt eller ofrivilligt.
- **informell eller populärvetenskaplig ton:** [...] bör vi därför ta oss i kragen [...], Det är dags att Svenska språkademien gör som vårdcentralerna gjorde för några år sedan och massvaccinerar!

Det exakta antalet stilbrott av olika typer i elevtexterna har inte räknats – vilket är svårt att göra med exakthet eftersom olika typer av stilbrott ofta går in i varandra – men min uppskattning är att stilbroten är avsevärt fler i E-texterna än i A-texterna; i genomsnitt är de två till tre gånger fler. Även A-texterna innehåller dock stilbrott. Bland annat verkar eleverna påverkas av den populärvetenskapliga stilen i flera av källtexterna. Många elever uttrycker sig på ett levande och lite humoristiskt sätt när de skriver, vilket i sig kan innebära goda språkliga formuleringar och tyda på språkkänsla, men det är tyvärr inte förenligt med den akademiska texttypen. Hultman och Westman (1977) påpekar i *Gymnasistsvenska* att det med tanke på att eleverna förväntas ha en viss språklig stil i sina texter är problematiskt att de så kallade stimulanstexterna (motsvarigheten till källtexterna i denna studie) är skrivna med olika stilar. Samma förhållande råder alltså i det undersökta nationella provet i kursen svenska 3.

Det förekommer ganska stora variationer mellan olika elevtexter gällande stilnivån. Det finns elever som använder sig av ett mer formellt, akademiskt språkbruk, och det finns elever som har en mer vardaglig eller populärvetenskaplig stil. Det är vanligt att stilar blandas i elevtexterna, till exempel genom inslag av populärvetenskapliga formuleringar i en i övrigt mer formell elevtext.

6.5 Övriga språkliga iakttagelser i elevtexterna

Analys av rena språkfel ingår inte i studien. Naturligtvis förekommer många sådana i elevtexterna. Det allmänna intrycket är att språkfelen överlag är fler i E-texterna, även om variationerna mellan olika elevtexter även på detta område är stora.

Några lexikala och semantiska drag som det vore intressant att analysera närmare i en senare studie är till exempel det som skulle kunna kallas för lexikal osäkerhet, det vill säga att en elev har svårt hitta rätt ord eller uttryck (till exempel *valfritt stavande ord, särstavningar*), bristande semantisk precision (till exempel *handlingen av själva texten*), eller företeelsen att mer precisa termer ersätts med längre formuleringar (till exempel [...] *vilken ord skrivs tillsammans och vilken separerad [...]* (E-text), *Man stavade ord som man tyckte att de lät [...]* (A-text). Det förefaller vara så att dessa typer av fel samt förekomst av pratigare formuleringar istället för mer koncentrerade och precisa uttryck är vanligare i E-texterna, men detta skulle behöva undersökas mer djupgående och systematiskt.

Hultman och Westman (1977) beskriver olika typer av lexikala och semantiska fel i elevtexter. Frick och Malmström (1976) beskriver olika typer av feltolkningar som informanter gör i ett ordförståelsetest. Gustafsson och Håkansson (2011; 2017) gör en uppföljande studie av informanternas ordförståelse, samt av resultaten på orddelen i högskoleprovet, och visar att det fortsatt finns brister i många språkbrukares ordförståelse. De lexikala och semantiska problem som uppmärksammas i dessa studier är märkbara även i denna studie.

6.6 Förhållandet till källtexterna

I tablå 4 visas vilka källtexter eleverna som arbetade med uppgiften *Särskrivning och stavfel* i delstudien har använt sig av.

Tablå 4. Användningen av de olika källtexterna i elevtexter med betyg A och E, uppgiften *Särskrivning och stavfel*.¹¹

Elevtext	Källtext 1	Källtext 2	Källtext 3	Källtext 4	Källtext 5	Källtext 8	Källtext 11
Betyg A							
A1	×	×	×			×	
A2	×	×	×				
A3	×	×	×	×			
A4		×	×				
A5	×	×	×				
A6		×	×				
A7	×		×				
A8	×	×	×				
A9	×	×	×				
A10	×	×	×				
Antal elever som använder texten	8	9	10	1	1		
Betyg E							
E1	×	×	×				
E2		×				×	
E3		×	×				×
E4		×	×				
E5	×	×					
E6		×	×			×	
E7		×	×				
E8		×	×				
E9	×	×	×				
E10	×	×				×	×
Antal elever som använder texten	4	10	7	0	3	1	1

Som framgår i tablå ovan är det framför allt tre olika källtexter som eleverna hänvisar till. Två av dessa är särskilt populära: 19 av 20 elever hänvisar till texten av Martin Borg (källtext 2, "Hur en djup fryst kyckling lever", *Nationalencyklopedin*) och 17 av 20 elever till texten av Lars Melin (källtext 3, "Vad är det för fel på ett fel?", *Forskning och framsteg*). Även texten av Fredrik Lindström (källtext 1), ett utdrag ur boken *Världens dåligaste språk*, används av

¹¹ De källtexter som inte användes alls i de granskade elevtexterna är inte med i tablå.

många elever. Alla tre texterna är populärvetenskapliga och innehåller populärvetenskapliga formuleringar. Borgs text (källtext 2), som är den mest använda, innehåller särskilt många populärvetenskapliga inslag. Den inleds på följande sätt:

Den sprider sig. Snabbt. Smittsam är den också. Den florerar inte blott bland ostadiga kråkfötter. Den härjar mitt ibland oss. Den? Ja, just de. Sär skrivnings sjukan. (Källtext 2)

Det är en konst att kunna skriva bra populärvetenskapliga texter, både rent innehållsligt och språkligt. Borgs stil är levande och humoristisk och lockar till läsning. Uppenbarligen har den lockat många av eleverna, eftersom nästan samtliga i delstudien använder sig av denna källtext.

Även Lars-Gunnar Anderssons text (källtext 4, ”Därför blir vi så arga på språkfel”, *Språktidningen*) är populärvetenskaplig. Att den inte används så mycket av eleverna beror troligen på att den handlar mer allmänt om språkfel, till skillnad från Borgs text (källtext 2), som fokuserar mer direkt på ett av de två språkfel som är i fokus i uppgiften.

Den andra texten av Lars Melin (källtext 5) utgörs av ett utdrag ur boken *Människan och skriften*. Även texten av Olle Engstrand (källtext 8) är ett utdrag ur en bok, *Hur låter svenskan, ejengklien*. Både Melin och Engstrand är professorer, och även om texterna inte är vetenskapliga, utan snarare kurslitteratur, är de skrivna med en mer neutral och saklig stil än till exempel Borgs text (källtext 2). Till viss del beror detta förstås på formatet; en utpräglat populärvetenskaplig stil med speciella uttryck som är tänkta att locka till läsning är mer motiverade i kortare texter, som i artiklar, än i böcker, där de mer fängslande formuleringarna åtminstone brukar blandas med avsnitt i neutralare ton. Även utdraget ur *Språkriktighetsboken* (källtext 11) har en saklig stil. Det handlar om *de/dem/dom*, och är inte omedelbart relevant för den aktuella uppgiften.

I genomsnitt hänvisar de elever som har skrivit A-texterna till nästan tre källtexter (2,9), medan de elever som har skrivit E-texterna i genomsnitt hänvisar till lite färre källtexter (2,5). Detta skulle kunna vara ett av skälen till att A-texterna i genomsnitt innehåller en högre andel ämnesord än E-texterna, samt fler referatmarkörer.

När det gäller ämnesorden är eleverna mycket beroende av källtexterna. Delstudien om centrala begrepp och termer (avsnitt 6.2) visade att samtliga sådana termer hämtats ur källtexterna. Det är förstås helt naturligt att ämnesord hämtas ur de källtexter som innehåller fakta om ämnet. Samtidigt speglar detta förhållande också att eleverna på gott och ont förefaller ämnesmässigt utlämnade åt källtexterna. Till skillnad från vad som är fallet när forskare skriver vetenskapliga artiklar om ett ämne som de behärskar, försöker eleverna skriva en text av så kallad vetenskaplig karaktär om ämnen som de inte har specialkunskap om. Men det är i de flesta fall troligen svårt för eleverna att avgöra hur olika begrepp förhåller sig till varandra, vilken terminologi som är bäst att använda och att vara konsekvent.

Till skillnad från vad som är fallet med ämnesorden har eleverna ett eget förråd av akademiska ord och uttryck. Eftersom sådana förekommer sparsamt i källtexterna – som ju inte är vetenskapliga texter – är det helt nödvändigt att eleverna har egna ord och uttryck här, vilket de också har. Allmänvetenskapliga ord och uttryck och det som i studien kallas för övriga akademiska textbindningsmarkörer, förekommer i nästan alla elevtexter, i något högre grad i A-texterna än i E-texterna. Referatmarkörer förekommer i alla elevtexter, utom i någon enstaka där eleven använt sig av fotnotsreferenser, men de är mer frekventa och mer varierade i A-texterna än i E-texterna.

Även stilbrott förekommer i princip i samtliga elevtexter, men är överlag mycket vanligare i E-texterna, liksom språkfel. Stilbrott i form av populärvetenskaplig stil förefaller i många fall bero på att eleverna har inspirerats av källtexterna. Flera av dessa, och särskilt de som merparten av eleverna hänvisar till, är tydligt populärvetenskapliga i tonen. Som tidigare nämnts konstaterar även Hultman och Westman (1977) att elever influeras av källtexternas stil i det egna skrivandet.

Det bör påpekas att det förstås finns mycket mer att säga om hur elevtexterna förhåller sig till källtexterna. Har eleverna valt rimliga texter att referera till och sammanfattas källtexternas innehåll på ett relevant och rättvisande sätt? I vilken omfattning innehåller elevtexterna direkta och dolda citat? Sådana aspekter har uppmärksamrats i tidigare forskning, av bland andra Ohlsson (2001) och Östlund-Stjärnegårdh (2006), men behandlas inte i denna studie.

7 Sammanfattning och diskussion

Det akademiska skrivandet är en viktig del av alla högskolestudier. Som student förväntas man både kunna förstå, och själv producera, akademiska texter. På gymnasiet förbereds eleverna på olika sätt inför vidare studier. Det akademiska skrivandet är en viktig del av kursen svenska 3, där det bland annat testas i det nationella prov som ges under kursen. I detta prov ska eleverna skriva ett så kallat pm, där de utifrån en given frågeställning refererar vad några olika författare har skrivit om ämnet, diskuterar ämnet utifrån frågeställningen och drar slutsatser. Eleverna ombeds föreställa sig att de precis har börjat på universitetet och att de ska skriva detta pm som ett underlag inför en diskussion om ämnet med lärare och medstudenter. Till skillnad från vad som är fallet i verklig forskning är det inte fråga om produktion av ny kunskap, utan om att återge och diskutera andras inlägg i ämnet. Termen vetenskapligt skrivande används om produktion av ny kunskap, medan akademiskt skrivande kan användas mer generellt om allt skrivande på högskolan, och då inkluderas även sådant som pm, hemtentamina och kurslitteratur. I det nationella provet på kursen svenska 3 används formuleringen ”skrivande av vetenskaplig karaktär”, vilken kan sägas motsvara beskrivningen av akademiskt skrivande.

Ur högskolans perspektiv är det viktigt att känna till vilka kunskaper och färdigheter nya studenter har, till exempel med avseende på akademiskt skrivande. Ur provkonstruktionsperspektiv är det viktigt med kontinuerliga analyser av elevers prestationer på proven, för att kunna bedöma hur väl proven fungerar och för att kunna modifiera dem på olika sätt. Sådan kunskap är även viktig ur ett didaktiskt perspektiv, för undervisningen i svenska, och i andra ämnen, på gymnasiet. Vilka svårigheter har elever gällande akademiskt skrivande, och hur kan sådana svårigheter mötas i den undervisning som ges? Även ur ett samhällsperspektiv är det intressant att undersöka gymnasieelevers akademiska skrivande närmare. Under lång tid har unga vuxnas bristande skrivförmåga debatterats i media, ofta i samband med inträdet på högskolan. Nya studenter sägs ha dåligt ordförråd och bristande förmåga att kunna uttrycka sig begripligt, korrekt och precist.

Den aktuella studien syftar till att undersöka, analysera och diskutera gymnasieelevers ordanvändning i det nationella prov som ingår i kursen svenska 3, samt att sätta in denna ordanvändning i ett vidare språkligt sammanhang med syntaktiska variabler och olika jämförelsematerial. Det aktuella provet, som gavs vårterminen 2014, var då nytt, och det är av detta skäl extra viktigt med olika typer av studier på detta prov. Provet hade vårterminen 2014 temat *Hur skriver man?* Texterna i det texthäfte som eleverna fick inför provet handlade på olika sätt om skriftspråkliga normer.

Studien utgår från två frågeställningar:

1. Vad utmärker på en övergripande nivå elevernas ordanvändning och syntax, och hur förhåller sig dessa till texterna i provets texthäfte respektive

till autentiska vetenskapliga artiklar? Finns det uppgifts- eller betygsrelaterade skillnader mellan elevtexterna med avseende på ordanvändning och syntax?

2. Hur ser förekomsten av ämnesord samt akademiska ord och uttryck ut i elevtexter med betyg A respektive E? Hur förhåller sig, på dessa områden, elevtexterna till texterna i texthäftet samt till autentiska vetenskapliga artiklar?

Den första frågeställningen besvarades genom en automatisk analys av 90 elevtexter, fördelade på två olika uppgifter och jämnt fördelade mellan betyg A, C och E, samt genom motsvarande analyser på två olika jämförelsematerial: dels de 14 så kallade källtexter som eleverna skulle använda sig av när de skrev sina pm, dels 13 autentiska vetenskapliga artiklar från tidskriften *Språk och stil*, 2014–2015. I delstudien analyserades textlängd, meningslängd, ordlängd, ordvariation, lemmavariation, läsbarhetsindex, bisatser, traddjup, substantiv/pronomen-kvot samt nominalkvot.

Resultaten av den automatiska analysen visar att elevernas grafiska meningar i genomsnitt innehåller runt 20 ord, vilket även de grafiska meningarna i de vetenskapliga artiklarna gör. I källtexterna är det genomsnittliga antalet ord per grafisk mening lägre, runt 15 ord. Den genomsnittliga ordlängden i elevtexterna är 5,11 bokstäver per ord för betyg A, 4,94 för betyg C, och för 4,83 betyg E. Motsvarande siffror för de vetenskapliga artiklarna är 5,66 och för källtexterna 5,03. Bland elevtexterna är det i A-texterna som ordlängden ligger närmast de vetenskapliga artiklarna, även om skillnaden är relativt stor. Även beträffande ordvariationsindex befinner sig A-texterna med ett genomsnittligt värde på 65,83 närmast de vetenskapliga artiklarnas motsvarande värde på 67,5. C-texterna har ett OVIX på 63,08 och E-texterna 59,33. I källtexterna är det genomsnittliga OVIX avsevärt högre än i övriga textgrupper: 75,2. Andelen bisatser är lägre i de vetenskapliga artiklarna (0,56) och i källtexterna (0,51) än i elevtexterna, där bisatserna är ungefär lika många som huvudsatserna (A 0,92; C 0,91; E 1,03). Traddjupsvärdena uppvisar ett liknande mönster som bisatsvärdena. De vetenskapliga artiklarna innehåller en avsevärt högre andel substantiv i förhållande till andelen pronomen (substantiv/pronomen-kvot 6,48) än elevtexterna (A 3,28; C 2,67; E 2,62) och källtexterna (2,62). Värdena för nominalkvoten bekräftar bilden av en mer nominal och koncentrerad språklig stil i de vetenskapliga artiklarna i förhållande till elevtexterna och källtexterna. I de vetenskapliga artiklarna är nominalkvoten i genomsnitt 1,59, i A-texterna 0,98, i C-texterna 0,90, i E-texterna 0,84 och i källtexterna 0,91.

Resultaten visar sammantaget på både likheter och skillnader mellan de olika textgrupperna. Elevtexterna ligger i flera variabler närmare *Språk och stil*-artiklarna än källtexterna; detta gäller till exempel meningslängd och ordvariation. Källtexterna har det högsta värdet för ordvariation, och i genomsnitt avsevärt kortare meningar än övriga texter. Andelen bisatser är avsevärt lägre i källtexterna

och i de vetenskapliga artiklarna än i elevtexterna. *Språk och stil*-artiklarna är mycket längre än elevtexterna och källtexterna och har i genomsnitt längst ord, liksom högst andel substantiv i förhållande till pronomen och högst nominalkvot.

Elevtexter med högre betyg är generellt längre, innehåller längre ord, har högre ordvariation, har en högre andel substantiv och en högre nominalkvot, än elevtexter med lägre betyg. Dessa skillnader är även i viss mån märkbara i jämförelsen mellan de två uppgifter som analyseras (*Särskrivning och stavfel* respektive *Skrivregler och sms*) – värdena i de nämnda variablerna är i genomsnitt något högre för elevtexterna om *Särskrivning och stavfel*.

I jämförelsen mellan de tre textgrupperna elevtexter, källtexter och vetenskapliga artiklar är det källtexterna som framstår som språkligt avvikande i förhållande till de andra två. De innehåller generellt kortare meningar och har mycket höga värden för ordvariation. Värdena speglar att det är fråga om populärvetenskapliga texter som är skrivna för att på ett lättfattligt sätt presentera olika infallsvinklar på språk, ofta med många exempel och med en humoristisk ton. Elevtexterna ligger generellt närmare de vetenskapliga artiklarna i de olika variabler som har analyserats; eleverna förefaller i den bemärkelsen ha lyckats skriva texter av vetenskaplig karaktär. Detta skulle kunna tolkas som att eleverna i viss utsträckning klarar av att hantera stilskillnaden mellan källtexterna och den text av vetenskaplig karaktär som de själva förväntas skriva.

Automatiska analyser har felkällor och kan till exempel inte ge en fullständig bild av sådant som faktisk ordanvändning i texter. För att besvara studiens andra frågeställning utfördes en manuell analys av förekomsten av ämnesord och akademiska ord och uttryck i 10 A-texter och 10 E-texter, tillhörande uppgiften *Särskrivning och stavfel*. Analysen visar bland annat att andelen ämnesord är något högre i A-texterna än i E-texterna, och att ämnesorden i mycket hög utsträckning hämtas från källtexterna. När det gäller akademiska ord och uttryck har däremot eleverna en egen språklig repertoar. Förekomsten av allmänvetenskapliga ord och uttryck, referatmarkörer samt av övriga akademiska textbindningsmarkörer är något högre i A-texterna än i E-texterna. Dessa ord och uttryck är också generellt mer varierade, och innehåller färre språkliga fel, i A-texterna. E-texterna innehåller vidare avsevärt fler stilbrott än A-texterna, och fler språkliga fel generellt.

Studiens resultat bekräftar i många avseenden olika aspekter av vad som är känt från tidigare forskning på det nationella provet och dess föregångare (bland annat Hultman & Westman 1977; Garne & Strömquist 1980; Melander 1997; Nyström 2000; Ohlsson 2001; Östlund-Stjärnegårdh 2002, 2006), vilket också var förväntat, eftersom de nationella proven i svenska och själva provkontexten i många avseenden är densamma över tid. Elevtexterna i det analyserade provet tillhör visserligen en annan genre ("skrivande av vetenskaplig karaktär") än elevtexterna som analyserades av Hultman och Westman i *Gymnasistsvenska* 1977 (debattartiklar), men i båda fallen är det fråga om att sammanfatta vad som sägs

om någon aspekt av ett ämne i olika texter, och att föra en diskussion utifrån detta, på en begränsad skrivtid i skolan.

I jämförelsen mellan elevtexterna och de vetenskapliga artiklarna förefaller den stora skillnaden vara just en fråga om ord och uttryck: de vetenskapliga artiklarna innehåller generellt längre ord och har en mer nominal och koncentrerad språklig stil än elevtexterna. Detta är en stil som förutsätter både god ämneskunskap och mycket tid för textplanering och textbearbetning, vilket alltså innebär andra förutsättningar än de som råder under det nationella provet i kursen svenska 3. Att verkligen förstå och självständigt kunna använda avancerade ord och uttryck är krävande och tar tid, vilket på olika sätt har uppmärksammats i forskning om språkbrukares ordförståelse och ordanvändning (se till exempel Frick och Malmström 1976; Gustafsson & Håkansson 2011, 2017; Aitchison 2012). Många gånger kan fel på ordnivå spegla att begreppet kanske inte är helt lexikaliserat hos språkanvändaren.

Skolkontextens och skrivsituationens inverkan på elevers texter har berörts av bland annat Westman (1980), Berge (1988), Hammarbäck (1989), Teleman (1991), Karlsson (1997), Melander (1997), Ohlsson (2001) och Borgström (2014). I denna forskning uppmärksammas bland annat att skrivandet i skolan omges av många implicita förväntningar på elevernas texter, och att skrivandet av eleverna kan upplevas vara på låtsas, när de ombeds föreställa sig olika typer av tänkta situationer som texterna ska kunna fungera i. Blåsjö (2004), Ask (2007) och Liberg och Säljö (2014) pekar alla på att allt skrivande sker i en kontext. Kontexten ser olika ut i olika sammanhang och i olika ämnen. Att lyckas avkoda denna kontext är en förutsättning för att lyckas i sitt skrivande. Tyvärr förblir kontexten många gånger dold för elever och studenter, och säkerligen många gånger omedveten för de lärare som verkar inom den.

Både Ledin (2001) och Gibbons (2016) pekar på att en genre består av mycket mer än språkliga drag. Att skriva en text inom en viss genre innebär att inom en viss kultur åstadkomma en text med en specifik struktur och specifika språkliga drag, och med ett speciellt socialt syfte. Huruvida det egentligen fullt ut är möjligt att skriva verkliga genretexter inom de påhittade skrivsituationer som det i många fall måste bli fråga om i skolan ifrågasätts av till exempel Borgström (2014) och Berge m.fl. (2016). Berge m.fl. (2016) förespråkar en skrivundervisning där eleverna snarare får träna på att använda skrivandet för olika syften, utan att man eftersträvar att de ska skriva färdiga texter. Det utredande, eller vetenskapliga, skrivandet beskrivs enligt följande:

If we write to *explore*, the purpose is to write in such a way that evidence and assumptions about the same subject matter are critically examined. In this way, we may discuss and assess what the most valid interpretation of a phenomenon is. We often write exploring texts when we want to develop our knowledge. This is the dominant way of writing in science. It is supposed to be valued in schools by developing students' critical awareness and attitudes towards knowledge, placing different understandings against each other. (Berge m.fl. 2016:10–11)

I det utredande skrivande som beskrivs av Berge m.fl. i citatet ovan är det innehållet i elevernas texter som är i fokus. Det handlar om att utveckla elevernas förmåga att kunna tolka, analysera och vara kritiska till hur olika typer av kunskapsinnehåll presenteras. Borgström (2014) menar att formen i de nationella proven i svenska på gymnasiet är alltför styrd för att tillåta eleverna att fatta självständiga beslut om hur de ska hantera ett visst innehåll i sin text. Ask (2005) påpekar å andra sidan att verklig kunskapsproduktion, i den betydelse som avses inom vetenskaplig forskning, knappast är möjlig på gymnasiet, och att det med avseende på akademiskt skrivande är rimligt och relevant att som förberedelse inför studier på högskolan fokusera på sådant som disposition, språklig korrekthet, saklig ton och att kunna sammanfatta andras resonemang, vilket alltså är delar som prövas i det nationella prov som studien behandlar.

Vilka textuella förebilder har eleverna när de skriver sina texter av vetenskaplig karaktär? I studien har det blivit tydligt att de källtexter som eleverna använder sig av när de skriver sina egna texter i många avseenden skiljer sig åt språkligt och stilistiskt från vetenskaplig text. De utgörs av mer populärvetenskaplig text i form av tidningsartiklar, utdrag ur läroböcker och liknande. Rent innehållsligt är detta förmodligen nödvändigt; det skulle troligen vara alltför utmanande för eleverna att försöka förstå motsvarande innehåll i vetenskaplig text, inom ramen för provets omfattning. Samtidigt förefaller många elever influeras av källtexternas populärvetenskapliga stil, särskilt på lägre betygsnivåer. Kanske möter inte eleverna särskilt många exempel på verkliga vetenskapliga texter på gymnasiet och saknar därmed egen uppfattning om vad som utmärker sådana texter på ett djupare plan. Kanske tror många elever att ”skrivande av vetenskaplig karaktär” är liktydigt med användning av ytspråkliga drag som referatmarkörer och vissa andra akademiska ord och uttryck, medan de är omedvetna om sådant som vikten av tydliga begreppsdefinitioner och konsekvent ämnesordsterminologi.

I den aktuella studien jämförs elevtexterna mellan olika betyg, oberoende av variabler som kön, gymnasieprogram och socialgrupp. Elevtexterna är med några få undantag skrivna av elever som läser svenska, och inte svenska som andraspråk. Det finns forskning som pekar på en ökande skolsegregation i dagens Sverige (se till exempel Holmlund m.fl. 2014). Broady (2007) talar om en ökande tendens i den svenska skolan till att familjebakgrund och skolval får allt större betydelse för elevers skolframgång. Akademiskt skrivande är ett område där akademiska traditioner inom familjen förmodligen kan ha stor betydelse för hur väl elever lyckas. Sättet att tänka, implicita förväntningar med mera kan vara lättare att avkoda för elever vars vårdnadshavare har läst på universitet och som därmed har möjlighet att stötta sina barn i denna avkodning, än för elever vars vårdnadshavare inte har gjort det.

I varje ny tid brukar det sägas att dagens unga kan mindre än tidigare generationers unga. Larmrapporterna om svenska elevers och studenters i olika avseenden bristande skrivförmåga har varit många på senare tid. Säkerligen finns det förbättringsområden. Samtidigt är det viktigt att uppmärksamma att dagens unga också är

avsevärt kunnigare inom vissa områden än oss äldre. Det gäller till exempel sådant som rör digitalisering, internet och multimodalitet. Engblom (2012, 2016) pekar på behovet av att skolan fångar upp elevers ökande kunskap på sådana områden. När nu de nationella proven gradvis digitaliseras är detta kanske ett steg i en sådan riktning.

8 Slutsatser

Resultaten av den automatiska analysen i studien bekräftar på centrala punkter vad som är känt från tidigare forskning om elevtext (Hultman & Westman 1977; Melander 1997; Strömquist 1999; Nyström 2000; Ribeck 2015), nämligen att textlängd, ordlängd och ordvariation har ett samband med både betyg och genre. Längre elevtexter, fler långa ord och högre ordvariation leder generellt till högre betyg, men både ordlängd och ordvariation är också genreberoende. I jämförelse med Hultmans och Westmans studie av elevers debattartiklar (1977) är elevernas ordlängd i denna studie lägre, men i jämförelse med Melanders (1997) och Nyströms (2000) studier är den högre, åtminstone i jämförelse med de mer personliga eller litterära genreerna i de prov som Melander och Nyström undersöker. Ordvariationen (OVIX) i elevtexterna befinner sig i denna studie generellt på en liknande nivå som i Hultmans och Westmans (1977) studie, men skillnaderna mellan betygsstegen är större i Hultmans och Westmans studie. Elevtexter med betyg 5 hade där i genomsnitt högre OVIX-värde än elevtexter med betyg A i denna studie, och elevtexter med betyg 1 något lägre OVIX-värde än elevtexterna i denna studie. Även beträffande ordlängd framkom i Hultmans och Westmans studie större skillnader mellan betygsstegen än i denna studie. De språkliga skillnaderna mellan elevtexter med höga och låga betyg förefaller baserat på dessa värden ha minskat över tid. Hultman och Westman (1977) kommer också fram till att långa meningar och en hög andel bisatser inte korrelerar med höga betyg, vilket även denna studie visar. Den genomsnittliga meningslängden och andelen bisatser är högst i E-texterna.

I förhållande till Palmérs (2018) automatiska analys av benchmark-texter från det nationella provet i svenska och svenska som andraspråk i kurs 3 på gymnasiet (samma prov som analyseras i denna studie men andra elevtexter) är värdena för textlängd i betyg A och OVIX något högre i denna studie, medan värdena för textlängd i betyg C och E, ordlängd och nominalkvot är lägre. Till skillnad från Palmérs resultat som visar att endast ordlängd och nominalkvot korrelerar med betyg i det nationella provet i svenska och svenska som andraspråk, i kurs 3, korrelerar samtliga fyra mått, det vill säga även textlängd och OVIX, med betyg i denna studie. Det är svårt att säga vad detta kan bero på. En möjlig förklaring skulle kunna vara att de väl genomarbetade bedömningarna av flera expertlärare under provets konstruktionsprocess innebär att textlängd och ordvariation inte påverkar lärarnas bedömning på samma sätt som de kanske gör i skolan när bedömningstid och möjligheter till medbedömning är mer begränsade. En annan möjlighet är att skillnaderna är relaterade till de statistiska värden som har använts. Till exempel utgår Palmér i OVIX från medianvärden medan medelvärden används i denna studie. Ytterligare en möjlig förklaring skulle kunna vara skillnader i själva mätverktygen eller i annoteringen av texterna. Sådana skillnader bör dock inte kunna påverka förhållandet mellan värden relaterade till betyg inom samma mått.

Den manuella analysen av 10 elevtexter med betyg A och 10 elevtexter med betyg E visar att andelen ämnesord i genomsnitt är högre i A-texterna (27 procent) än i E-texterna (22 procent). Eleverna är gällande ämnesorden mycket beroende av källtexterna. Beträffande akademiska ord och uttryck har de i större utsträckning ett eget språkförråd, vilket också är nödvändigt med tanke på att källtexterna överlag saknar sådana uttryck eftersom de är mer populärvetenskapliga. Att elever genom akademiska ord och uttryck lyckas efterlikna en akademisk stil förefaller väga tungt i lärares bedömning. A-texterna innehåller något fler allmänvetenskapliga ord och uttryck (*frågeställning*, *slutsats* med flera), fler och mer varierade referatmarkörer, och något fler akademiska textbindningsmarkörer (*i motsats till*, *vidare* med flera). Att stilen inverkar på lärarens bedömning bekräftas också av att stilbrott är vanligare i E-texterna än i A-texterna.

En första slutsats som kan dras utifrån studiens resultat är att språket i elevtexterna generellt ligger närmare språket i verkliga vetenskapliga artiklar än språket i de källtexter som eleverna sammanfattar i sina texter, vilket skulle kunna tolkas som att eleverna förstår uppgiften och utifrån provsituationens ramar förmår hantera den. De språkliga skillnaderna mellan elevtexterna och de vetenskapliga artiklarna handlar i stor utsträckning om att de senare har en mer koncentrerad, nominal stil än de förra, med stort fokus på ämnesord och terminologi, och med fler utvecklade nominalfraser. De språkliga skillnader mellan källtexter och vetenskapliga artiklar som framkommer i studien påminner om Strömquists (1999) och Ribecks (2015) beskrivning av läroboksspråk som koncentrerat med nominal stil, hög informationstäthet och ganska dålig textbindning, samt Ribecks (2015) beskrivning av akademiskt språk som mer avancerat än läroboksspråket, med bland annat fler och längre nominalfraser och med fler långa, utbyggda subjekt som meningsinledare. Det bör dock påpekas att källtexterna i många fall utgörs av tidningstext snarare än lärobokstext.

En andra slutsats av studien är att automatisk analys skulle kunna användas som ett tidsbesparande komplement till, samt öka likvärdigheten i, lärares bedömningar av elevtexter. Automatisk analys av elevtexter har i mindre omfattning förekommit inom elevspråksforskning sedan flera decennier tillbaka och har utvecklats än mer på senare tid (Palmér 2018). Automatisk analys skulle kunna erbjuda fler möjligheter för läraren att få en första övergripande bild av språket i elevtexten, samt tillhandahålla större och mer precisa möjligheter till jämförelser mellan elevtexter, samt mellan elevtexter och texter skrivna av professionella skribenter i olika genrer. De nationella provens kommande digitalisering öppnar än mer upp för sådana möjligheter.

Mer forskning behövs om hur värden för olika mått som framtagits genom automatisk analys kan användas och tolkas, inte minst i fråga om nyare mått som lemmavariationsindex. Mer forskning om vilka kombinationer av mått som leder till de mest givande analyserna på textnivå behövs också. Värdena måste kunna relateras till genre och felkällor får inte inverka på resultaten. För att en texts hela komplexitet ska kunna förstås förefaller det under överskådlig tid dock inte

möjligt att undvika manuell analys av mer kvalitativa textegenskaper såsom innehållets relevans och textens originalitet, samt av texten i dess helhet. Automatiska bedömningsverktyg är dock en intressant tanke och möjligheterna bör undersökas vidare inom forskningen.

En tredje slutsats av studien är att provets design – med populärvetenskapliga källtexter, begränsad skrivtid under provsituationen där textbearbetning i flera led inte är möjlig, samt provets bedömningsmatris med fokus på enhetliga bedömningar utifrån tydligt urskiljbara drag som utmärker akademiska texter – begränsar elevernas möjligheter till självständiga och genomarbetade analyser i förhållande till den frågeställning de skriver om. Som tidigare konstaterats ligger språket i elevtexterna generellt närmare språket i verkliga vetenskapliga artiklar än språket i de källtexter som eleverna använder sig av. Samtidigt finns det i elevtexterna också många exempel på inslag av en mer populärvetenskaplig stil. Det är rimligt att tänka sig att eleverna påverkas av stilen i de källtexter de använder sig av för att skriva sina egna, särskilt om de är osäkra på vad en akademisk stil innebär. Källtexterna är överlag populärvetenskapliga, vilket förefaller rimligt ur innehållsligt hänseende, men vilket kanske ur språkligt och stilistiskt hänseende kan påverka elever i fel riktning, särskilt om de saknar textuella förebilder inom akademiskt/vetenskapligt skrivande.

Det nationella provet i kursen svenska 3 testar elevernas förmåga att skriva en text av vetenskaplig karaktär med avseende på struktur, källhantering, saklighet och språkbruk, men övar kanske inte i lika stor utsträckning deras vetenskapliga tänkande. Samtidigt är det förstås så att de aspekter som bedöms i provet är viktiga beståndsdelar i det akademiska skrivandet. Det kan också anses rimligt att svenskämnet tar ett extra ansvar för sådant som språk, stil, textstruktur och referenshantering. Ansvar för att eleverna tränas i kritiskt och vetenskapligt tänkande vilar på alla ämnen. Provkontextens ramar och strävan efter likvärdiga lärarbedömningar innebär begränsningar av hur fri uppgiften kan vara.

En fjärde, didaktisk slutsats av studien är att elevers förmåga att läsa och skriva akademiska texter är något som måste övas i många ämnen över tid, och att ett framgångsrikt akademiskt skrivande också kräver kunskap om vetenskapligt tänkande. Hur skulle då en fördjupad kunskap om akademiskt skrivande kunna utvecklas hos eleverna i undervisningen i svenska och andra ämnen? Hur kan eleven förstå och använda sig av innehållet i vetenskapliga texter? Enligt Gibbons (2016) modell för genreundervisning i skolan är arbete med modelltexter en viktig del för att lära sig att skriva egna texter inom en viss genre. Det kan handla om att analysera (dekonstruera) texter med fokus på deras uppbyggnad, språkliga drag och det sätt som innehållet framställs på. Därefter påbörjas arbetet med att gradvis skriva egna texter inom den aktuella genren, först kanske genom att skriva en text tillsammans med andra, och sedan så småningom egna texter, med stöttning av sådant som skrivmallar och formativ bedömning. Kanske vore det en god didaktisk idé, om tiden medger det, att i förberedelserna inför det nationella provet i kursen

svenska 3 låta eleverna tillsammans med sin lärare arbeta med några olika vetenskapliga texter från samma ämnesdisciplin som det aktuella provet (språk eller litteratur), med fokus på just det vetenskapliga skrivandet i sig. Ytterligare en idé skulle kunna vara att begränsa provet till en enda uppgift (en frågeställning) som ska diskuteras utifrån ett fåtal autentiska vetenskapliga texter, förslagsvis två eller tre. Eleverna skulle på så sätt inte kunna uppvisa samma självständighet i valet av texter att hänvisa till utifrån den valda uppgiften och deras engagemang i ämnet skulle kanske minska. Samtidigt skulle deras insikter i vad som på ett djupare plan utmärker vetenskaplig analys och vetenskapligt skrivande möjligen öka.

I inledningen (kapitel 1) berördes vikten av att studenter inom högre utbildning bemästrar det akademiska skrivandet för att de ska lyckas med sina studier, liksom det faktum att många lärare på universitet och högskolor upplever brister på området. Vad kan då utifrån resultaten av denna studie sägas om de nya studenternas förmodade kunskaper om akademiskt skrivande när de börjar på högskolan? Frågan är förstås komplex och de individuella variationerna är stora. Utifrån resultaten i denna studie förefaller studenterna kunna mer om vad som utmärker den akademiska stilen, än om vad som mer innehållsligt utmärker vetenskapliga resonemang (jfr Berge m.fl. 2016). Språk och tänkande är intimt sammanlänkade, och komplexa resonemang om komplexa frågeställningar ställer höga krav på språklig korrekthet och precision. Att gymnasieelever genom det nationella provet i kursen svenska 3 får kunskap om vad som utmärker akademiska texter gällande språk, stil, disposition och källhantering är positivt. Det är också rimligt att svenskämnet tar ett särskilt ansvar för sådant som rör texters språk, stil och uppbyggnad i olika genrer, på en mer generell nivå. Dock kan inte svenskämnet, eller i snävare bemärkelse det nationella prov i kursen svenska 3 som här har behandlats, ensamt bära ansvaret för att elever utvecklar ett vetenskapligt tänkande. Här behöver många ämnen bidra, inte minst eftersom vad som förväntas beträffande vetenskapligt förhållningssätt och akademiskt skrivande delvis är ämnesspecifikt. Det gymnasiearbete som eleverna också skriver spelar en viktig roll i sammanhanget. För en mer fullständig förståelse av gymnasieelevers möjligheter att utveckla sitt akademiska skrivande och sin förmåga till kritiskt och självständigt tänkande behövs det fler ämnesöverskridande studier av hur dessa områden behandlas i de olika ämnen som eleverna möter under sin gymnasietid.

Referenser

Litteratur

- Aitchison, Jean, 2012: *Words in the mind: an introduction to the mental lexicon*. 4 uppl. Oxford: John Wiley & Sons Inc.
- Ask, Sofia, 2005: *Tillgång till framgång: Lärare och studenter om stadiövergången till högre utbildning*. Växjö universitet.
- Ask, Sofia, 2007: *Vägar till ett akademiskt skriftspråk*. Acta Wexionensia, nr 115/2007. Växjö: Växjö University Press.
- Berge, Kjell Lars, 1988: *Skolestilen som genre. Med påtvungen penn*. Skrifter fra Landslaget for norskundervisning 39. Oslo: LNU/Cappelen
- Berge, Kjell Lars, Evensen, Lars Sigfred & Thygesen, Ragnar, 2016: The Wheel of Writing: a model of the writing domain for the teaching and assessing of writing as a key competency. I: *The Curriculum Journal*, vol. 27, nr 2, s. 1–18. DOI:10.1080/09585176.2015.1129980.
- Bergöö, Kerstin, 2005: *Vilket svenskämne? Grundskolans svenskämnen i ett lärarutbildningsperspektiv*. Malmö högskola.
- Bergström, Annika, 2007: *Två olika ämnen? Svenska språket på gymnasiet och på högskolan*. Meddelanden från Institutionen för svenska språket (MISS) 59. Göteborgs universitet.
- Björnsson, Carl Hugo, 1968: *Läsbarhet*. Stockholm: Bokförlaget Liber.
- Blückert, Ann, 2006: Juridiklärarens skrivhandledning och pm-kommentarer som vägledning in i en språklig praktik. I: Ledin, Per (red.), *Svenskans beskrivning 28*, s. 59–66. Örebro universitet.
- Blåsjö, Mona, 2004: *Studenters skrivande i två kunskapsbyggande miljöer*. Acta Universitatis Stockholmiensis. Stockholm Studies in Scandinavian Philology. New Series. 37. Stockholm: Almqvist & Wiksell International.
- Blåsjö, Mona, 2010: *Skrivteori och skrivforskning. En forskningsöversikt*. Meddelanden från Institutionen för nordiska språk vid Stockholms universitet (MINS) 56. 2 uppl.
- Borgström, Eric, 2014: *Skrivbedömning: Uppgifter, texter och bedömningsanvisningar i svenskämnets nationella prov*. Studier från Örebro i svenska språket 11. Örebro universitet.
- Broady, Donald, 1981: Den dolda läroplanen. Nyutgiven 2007 i *Kritisk utbildningstidskrift (KRUT)* 127, s. 15–100.
- Broady, Donald, 2007: Trettio år efteråt – ett återbesök hos den dolda läroplanen. I: *Kritisk utbildningstidskrift (KRUT)* 127, s. 7–12.

- Broman, Andreas & Dalberg, Tobias, 2014: *Resultatrapport kursprov 3 vt 2014 – Hur skriver man?* [pdf]. Gruppen för nationella prov i svenska och svenska som andraspråk, Institutionen för nordiska språk, Uppsala universitet. https://www.natprov.nordiska.uu.se/digitalAssets/526/c_526848-1_3-k_kp3vt14resultatrapport.pdf.
- Broman, Andreas & Dalberg, Tobias, 2014: *Resultatrapport kursprov 3 vt 2015 – Det var en gång* [pdf]. Gruppen för nationella prov i svenska och svenska som andraspråk, Institutionen för nordiska språk, Uppsala universitet. https://www.natprov.nordiska.uu.se/digitalAssets/526/c_526848-1_3-k_kp3vt15resultatrapport.pdf.
- Cassirer, Peter, 1977: *Studier över ordförståelse. Rapport från projektet Svenska och deras ord*. ACTA Nordistica Gothoburgensia 10. Göteborg: Acta Universitatis Gothoburgensis.
- Engblom, Charlotte, 2012: Yngre barns skärmbaserade texthantering. Multimodala texter och nya perspektiv på progression och bedömning. I: Magnusson, Jenny, Malmbjer, Anna & Wojahn, Daniel (red.), *Åttonde nationella konferensen i svenska med didaktisk inriktning: Språket, kroppen och rummet – multimodala och digitala perspektiv på lärande*, Södertörns högskola 25–26 november 2010, s. 35–50. Stockholm.
- Engblom, Charlotte, 2016: Prefabricated Images in Young Children's Text-Making at School. *Designs for Learning*, vol. 8, nr 1, s. 37–47. DOI: 10.16993/dfl.70
- Frick, Nils & Malmström, Sten, 1976: *Språkklyftan: Hur 700 ord förstås och missförstås*. Stockholm: Tidens Förlag.
- Garne, Birgitta & Strömquist, Siv, 1980: Visst kan flickor få femmor! I: Larsson, Kent (red.), *Elevsvenska*, Ord och stil, Språkvårdssamfundets skrifter 12, s. 61–83. Lund: Studentlitteratur.
- af Geijerstam, Åsa, 2006: *Att skriva i naturorienterande ämnen i skolan*. Studia Linguistica Upsaliensia 3. Uppsala: Acta Universitatis Upsaliensis.
- Gibbons, Pauline, 2016: *Stärk språket, stärk lärandet. Språk- och kunskapsutvecklande arbetssätt för och med andraspråkselever i klassrummet*. 4 uppl. Uppsala: Hallgren & Fallgren.
- Gruppen för nationella prov i svenska och svenska som andraspråk, 2014: *Sammanställning av lärarenkät. Hur skriver man? Kursprov 3 vt 2014 i svenska 3 och svenska som andraspråk 3* [pdf]. Institutionen för nordiska språk, Uppsala universitet. https://www.natprov.nordiska.uu.se/digitalAssets/526/c_526722-1_3-k_kp3vt14sammanstallninglararenkat.pdf.
- Gustafsson, Anna & Håkansson, David, 2011: Språkklyftan: 30 år senare. I: Edlund, Ann-Catrine & Mellenius, Ingmarie (red.), *Svenskans beskrivning 31*, s. 123–132. Umeå: Umeå universitet, Institutionen för språkstudier.
- Gustafsson, Anna W., & Håkansson, David, 2017: *Ord på prov: En studie av ordförståelse i högskoleprovet*. Lundastudier i nordisk språkvetenskap, Nr A75. Lund: Språk- och litteraturcentrum, Lunds universitet.

- Hallencreutz, Katharina, 2003: *Särskrivningar och andra skrivningar i elevspråk. Svenska i utveckling nr 20 (FUMS Rapport nr 210)*. Institutionen för nordiska språk, Uppsala universitet.
- Hammarbäck, Solveig, 1989: Skrivutveckling – hinder och möjligheter för eleverna på gymnasieskolans yrkesinriktade linjer. I: Sandqvist, Carin & Teleman, Ulf (red.), *Språkutveckling under skoltiden*, s. 143–158. Lund: Studentlitteratur.
- Heimann Mühlenbock, Katarina, 2013: *I see what you mean. Assessing readability for specific target groups*. Data linguistica 24, Göteborgs universitet.
- Hellspong, Lennart & Ledin, Per, 1997: *Vägar genom texten. Handbok i brukstextanalys*. Lund: Studentlitteratur.
- Holmberg, Per, 2006: Funktionell grammatik för textarbete i skolan. I: Lindberg, Inger & Sandwall, Karin (red.), *Språket och kunskapen – att lära på sitt andraspråk i skola och högskola, Rapport från nordisk konferens den 7–8 oktober 2005 i Göteborg*, Rapporter om svenska som andraspråk (ROSA) 7, s. 129–146. Institutet för svenska som andraspråk, Göteborgs universitet.
- Holmberg, Per, 2009: Text, språk och lärande. Introduktion till genrepedagogik. I: Olofsson, Mikael (red.), *Symposium 2009: Genrer och funktionellt språk i teori och praktik*, s. 13–27. Stockholm: Stockholms universitets förlag.
- Holmlund, Helena, Häggblom, Josefin, Lindahl, Erica, Martinson, Sara, Sjögren, Anna, Vikman, Ulrika & Öckert, Björn, 2014: *Decentralisering, skolval och fristående skolor: resultat och likvärdighet i svensk skola*. IFAU Rapport 2014:25. Uppsala: Institutet för arbetsmarknads- och utbildningspolitisk utvärdering.
- Hultman, Tor G., 1980: Språk och kön i skolan. Några resultat från undersökningar av elevspråk. I: Larsson, Kent (red.), *Elevsvenska, Ord och stil, Språkvårdssamfundets skrifter 12*, s. 29–60. Lund: Studentlitteratur.
- Hultman, Tor G. & Westman, Margareta, 1977: *Gymnasistsvenska*. Lund: Liber Läromedel.
- Johansson, Victoria, 2009: *Developmental aspects of text production in writing and speech*. Travaux de l'institut de linguistique de Lund 48. Lunds universitet.
- Järborg, Jerker, 2007: Om ord och ordkunskap. I: Lindberg, Inger & Johansson Kokkinakis, Sofie (red.) 2007, *OrdiL – en korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år*, Rapporter om svenska som andraspråk (ROSA) 8, s. 61–100. Institutet för svenska som andraspråk, Göteborgs universitet.
- Karlsson, Anna-Malin, 1997: Textnormer i och utanför skolan – att skriva insändare på riktigt och på låtsas. I: Håkansson, Gisela, Lötmarker, Lena, Santesson, Lillemor, Svensson, Jan & Viberg, Åke (red.), *Svenskans beskrivning 22 – Förhandlingar vid Tjugoandra sammankomsten för svenskans beskrivning*, s. 172–186. Lund: Lund University Press.
- Larsson, Kent, 1980: Inledning. I: Larsson, Kent (red.), *Elevsvenska, Ord och stil, Språkvårdssamfundets skrifter 12*, s. 5–8. Lund: Studentlitteratur.

- Larsson, Kent (red.), 1981: *Skola, språk och kön*. Ord och stil. Språkvårdssamfundets skrifter 13. Lund: Studentlitteratur.
- Larsson, Kent, 1984: *Skrivförmåga. Studier i svenskt elevspråk*. Malmö: Liber Förlag.
- Larsson, Kent, 1989: Språket, vetandet och inläringen. I: Sandqvist, Carin, & Teleman, Ulf (red.), *Språkutveckling under skoltiden*, s. 53–78. Lund: Studentlitteratur.
- Lea, Mary R. & Street, Brian V., 1998: Student writing in higher education. An academic literacies approach. I: *Studies in Higher Education*, vol. 23, nr 2, s. 157–172. DOI: 10.1080/03075079812331380364.
- Ledin, Per, 1996: Genrebegreppet – en forskningsöversikt. Nyutgiven 2001 av författaren samt förlaget Studentlitteratur. *Svensk sakprosa, rapport 2*. Lund: Studentlitteratur.
- Liberg, Caroline, af Geijerstam, Åsa & Wiksten Folkeryd, Jenny, 2010: *Utmana, utforska, utveckla! Om läs- och skrivprocessen i skolan*. Lund: Studentlitteratur.
- Liberg, Caroline & Säljö, Roger, 2014: Grundläggande färdigheter – att bli medborgare. I: Lundgren, Ulf P. Liberg, Caroline & Säljö, Roger (red.), *Lärande, skola, bildning: Grundbok för lärare*, 3 uppl., s. 357–378. Stockholm: Natur och Kultur.
- Lindberg, Inger & Johansson Kokkinakis, Sofie (red.), 2007: *OrdiL – en korpusbaserad kartläggning av ordförrådet i läromedel för grundskolans senare år*. Rapporter om svenska som andraspråk (ROSA) 8. Institutet för svenska som andraspråk, Göteborgs universitet.
- Lix räknare, u.å.: *Lix räknare*. www.lix.se. Hämtat 2018-08-31.
- Megyesi, Beáta, Näsman, Jesper & Palmér, Anne, 2016: The Uppsala Corpus of Student Writings: Corpus Creation, Annotation, and Analysis. I: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC 2016)*, s. 3192–3199. Paris: European Language Resources Association (ELRA).
- Melander, Björn, 1997: Ordstudier i svensk gymnasistprosa. I: Håkansson, Gisela, Lötmarker, Lena, Santesson, Lillemor, Svensson, Jan, & Viberg, Åke (red.), *Svenskans beskrivning 22 – Förhandlingar vid Tjugoandra sammankomsten för svenskans beskrivning*, s. 125–139. Lund: Lund University Press.
- Nyström, Catharina, 2000: *Gymnasisters skrivande. En studie av genre, textstruktur och sammanhang*. Skrifter utgivna av Institutionen för nordiska språk vid Uppsala universitet 51. Uppsala.
- Näsman, Jesper, Megyesi, Beáta & Palmér, Anne, 2017: SWEGRAM – A Web-Based Tool for Automatic Annotation and Analysis of Swedish Texts. *Proceedings of the 21st Nordic Conference on Computational Linguistics (NoDaLiDa)*, s. 132–141.
- Ohlsson, Maria, 2001: *Säker stil eller rätt i sak? En studie i gymnasisters analyser av en enkät*. Svenska i utveckling nr 16 (FUMS Rapport nr 204). Institutionen för nordiska språk, Uppsala universitet.

- Palmér, Anne, 2013: *Nationella skrivprov baserade på två olika läroplaner. Genre, kommunikationssituation och skrivdidaktiska diskurser*. Svenska i utveckling nr 31. Institutionen för nordiska språk, Uppsala universitet.
- Palmér, Anne, 2018: Kvantiteter i kvalitativt bedömda elevtexter – framtida verktyg för rättvis bedömning? *Acta Didactica Norge*, vol. 12, nr 4, art. 9, s. 1–23. DOI: 10.5617/adno.6357.
- Palmér, Anne & Östlund-Stjärnegårdh, Eva, 2015: *Bedömning av elevtext – enligt Lgr11 och Lgy11*. Stockholm: Natur & Kultur.
- Pettersson, Åke, Franzén, Lena & Rehman, Ann-Christine, 1981: Känslor och kön. I: Larsson, Kent (red.), *Skola, språk och kön*, Ord och stil, Språkvårdssamfundets skrifter 13, s. 76–107. Lund: Studentlitteratur.
- Pettersson, Åke, 1989: Utvecklingslinjer och utvecklingskrafter i elevernas uppsatser. I: Sandqvist, Carin & Teleman, Ulf (red.), *Språkutveckling under skoltiden*, s. 159–184. Lund: Studentlitteratur.
- Pettersson, Åke, 1997: Blev gymnasisterna bättre skribenter mellan 1984 och 1995? I: Håkansson, Gisela, Lötmarker, Lena, Santesson, Lillemor, Svensson, Jan, & Viberg, Åke (red.), *Svenskans beskrivning 22 – Förhandlingar vid Tjugoundra sammankomsten för svenskans beskrivning*, s. 156–171. Lund: Lund University Press.
- Pilán, Ildikó & Volodina, Elena, 2018: Investigating the importance of linguistic complexity features across different datasets related to language learning, *Proceedings of the Workshop on Linguistic Complexity and Natural Language Processing* [pdf], s. 49–58. Santa Fe, New Mexico, USA, 25 augusti. Association for Computational Linguistics. <https://aclanthology.org/W18-4606>.
- Ribeck, Judy, 2015: *Steg för steg. Naturvetenskapligt ämnesspråk som räknas*. Data linguistica 28, Göteborgs universitet.
- Sofkova Hashemi, Sylvana & Hård af Segerstad, Ylva, 2007: *Elevernas skriftspråkande i skolan och nya medier. Slutrapport från projektet "Att lära sig skriva i IT-samhället"*. Göteborg: Göteborgs universitet.
- Stage, Christina, 1981: Könsskillnader i ordkunskap. I: Larsson, Kent (red.), *Skola, språk och kön*, Ord och stil, Språkvårdssamfundets skrifter 13, s. 108–118. Lund: Studentlitteratur.
- Strömquist, Siv, 1999: Läroboksspråk – mönster eller monster? I: Strömquist, Siv (red.), *Läroboksspråk – Om språk och layout i svenska läroböcker*, Ord och stil, Språkvårdssamfundets skrifter 26, 2 uppl., s. 7–11. Uppsala: Hallgren & Fallgren.
- Teleman, Ulf, 1989: Språkutveckling under skoltiden: ett problemfält. I: Sandqvist, Carin & Teleman, Ulf (red.), *Språkutveckling under skoltiden*, s. 7–13. Lund: Studentlitteratur.
- Teleman, Ulf, 1991: *Lära svenska. Om språkbruk och modersmålsundervisning*. Skrifter utgivna av svenska språknämnden 75. Uppsala: Almqvist & Wiksell.

- Westman, Margareta, 1980: Om textnormer i skolan. I: Larsson, Kent (red.), *Elevsvenska*, Ord och stil, Språkvårdssamfundets skrifter 12, s. 104–117. Lund: Studentlitteratur.
- Wikipedia, 2020: *Läsbarhetsindex*. <https://sv.wikipedia.org/wiki/Läsbarhetsindex>. Hämtat 2020-07-20.
- Östlund-Stjärnegårdh, Eva, 2002: *Godkänd i svenska? Bedömning och analys av gymnasieelevers texter*. Skrifter utgivna av Institutionen för nordiska språk vid Uppsala universitet 57. Uppsala.
- Östlund-Stjärnegårdh, Eva, 2006: *Att förmedla egna och andras tankar. Om gymnasisters källhantering i det nationella provets skrivuppgift*. Svenska i utveckling nr 23 (FUMS Rapport nr 219). Institutionen för nordiska språk, Uppsala universitet.

Material

- Melander, Björn, (red.) 2014: *Språk och stil: Tidskrift för svensk språkforskning*. Ny följd 24. Uppsala: Swedish Science Press.
- Melander, Björn, (red.) 2015: *Språk och stil: Tidskrift för svensk språkforskning*. Ny följd 25. Uppsala: Swedish Science Press.
- Skolverket, 2014. *Nationella provet i svenska och svenska som andraspråk vårterminen 2014*, kursprov 3, omfattande:
- 90 elevtexter ur Uppsala elevtextkorpus. (Uppgift 1, *Skrivregler och sms*: 20 A-texter, 20 C-texter, 20 E-texter. Uppgift 5, *Särskrivning och stavfel*: 10 A-texter, 10 C-texter, 10 E-texter.)
 - *Hur skriver man? Texthäfte till det nationella kursprovet i svenska 3 och svenska som andraspråk 3*.
 - häftet *Delprov A: Hur skriver man?*, uppgiftsinstruktioner
 - häftet *Lärarinformation: Hur skriver man?* ("gröna häftet")
 - häftet *Bedömningsanvisningar: Hur skriver man?* ("röda häftet").

Bilaga 1. Resultattabeller från automatisk analys

Tabell 7. Medelvärden (SVIT) i elevtexter med betyg A, C och E på uppgiften *Särskrivning och stavfel* och *Skrivregler och sms*.

Textmått	<i>Särskrivning och stavfel</i>			<i>Skrivregler och sms</i>		
	Elev- texter betyg A	Elev- texter betyg C	Elev- texter betyg E	Elev- texter betyg A	Elev- texter betyg C	Elev- texter betyg E
Löppord	795	731	672	749	690	649
Antal ord/mening	21	17,4	20,9	20	20,9	22,1
Antal bokst./ord	5,12	5,03	4,94	5,1	4,89	4,77
OVIX	66,7	62	59,5	65,4	63,6	59,3
LIX	48,5	43,3	45,7	46,5	44,5	44
LVIX	55,2	51,9	50,6	55,1	53,4	50,6
Andel bisatser	0,99	0,84	1,04	0,89	0,94	1,02
Träddjup	8,52	7,52	8,51	8,24	8,49	8,63
Subst/Pron-kvot	3,41	2,96	3,07	3,22	2,53	2,39
Nominalkvot	0,97	0,95	0,9	0,98	0,87	0,81

Tabell 8. SVIT-värden i elevtexter, uppgiften *Särskrivning och stavfel*, betyg A.

Elevtext	Löppord	Antal ord/ mening	Antal bokst./ ord	OVIX	LIX	LVIX	Andel bisatser	Träddjup	Subst/Pron- kvot	Nominal- kvot
A1	724	22,62	5,26	72,75	53,0	55,93	0,94	8,3	3,49	1,09
A2	849	22,34	5,24	75,22	50,0	62,36	0,84	8,8	3,53	1,02
A3	926	25,03	5,07	61,32	52,0	53,02	1,59	9,5	2,63	0,89
A4	849	21,77	5,24	75,22	49,4	62,56	0,82	8,8	3,53	1,02
A5	712	22,97	4,81	64,96	45,7	53,71	1,13	9,3	3,08	0,94
A6	781	17,75	4,82	55,31	43,1	47,11	0,84	8,0	1,96	0,79
A7	830	18,86	4,89	60,93	44,0	49,92	0,86	8,1	2,00	0,82
A8	771	15,73	5,08	60,87	42,2	51,58	0,82	7,4	3,91	0,85
A9	694	17,79	5,46	73,98	51,1	59,86	0,95	7,9	3,84	1,15
A10	814	24,67	5,33	66,29	54,4	55,43	1,09	9,1	6,09	1,13
Medelvärden	795	20,95	5,12	66,69	48,49	55,15	0,99	8,52	3,41	0,97
Standard- avvikelse	68,91	3,03	0,21	6,81	4,17	4,91	0,23	0,65	1,12	0,12
Varians	4748,20	9,18	0,05	46,39	17,41	24,11	0,05	0,42	1,25	0,02

Tabell 9. SVIT-värden i elevtexter, uppgiften *Särskrivning och stavfel*, betyg C.

Elevtext	Löppord	Antal ord/ mening	Antal bokst./ ord	OVIX	LIX	LVIX	Andel bisatser	Träddjup	Subst/Pron- kvot	Nominal- kvot
C1	749	25.83	4.48	53.26	41.8	44.93	1.10	9.0	1.49	0.62
C2	629	17.97	5.29	64.33	46.4	52.26	0.77	7.5	3.66	0.94
C3	709	15.41	4.90	60.23	39.4	50.85	0.78	7.4	2.84	1.00
C4	763	13.87	5.04	73.37	40.5	60.83	0.67	6.5	2.89	0.95
C5	644	21.47	4.84	50.27	45.2	44.30	1.40	8.4	2.19	0.87
C6	740	14.80	5.07	68.23	41.4	55.69	0.62	6.9	2.79	0.96
C7	777	15.24	5.42	79.04	45.7	65.43	0.57	7.3	4.02	1.32
C8	760	18.10	4.85	53.39	42.0	44.40	1.05	8.2	2.62	0.86
C9	797	16.96	5.02	57.33	44.3	48.95	0.89	7.6	2.95	0.90
C10	741	14.25	5.41	60.80	46.4	50.82	0.52	6.4	4.10	1.10
Medelvärden	730,90	17,39	5,03	62,03	43,31	51,85	0,84	7,52	2,96	0,95
Standard- avvikelse	52,25	3,56	0,27	8,82	2,45	6,71	0,26	0,79	0,76	0,17
Varians	2729,89	12,64	0,08	77,84	6,02	45,08	0,07	0,62	0,58	0,03

Tabell 10. SVIT-värden i elevtexter, uppgiften *Särskrivning och stavfel*, betyg E.

Elevtext	Löpdord	Antal ord/ mening	Antal bokst./ ord	OVIX	LIX	LVIX	Andel bisatser	Träddjup	Subst/Pron- kvot	Nominal- kvot
E1	719	22.47	4.85	66.03	46.7	55.65	1.25	9.4	3.43	1.13
E2	412	24.24	5.12	56.75	52.1	48.23	1.24	9.9	3.48	0.92
E3	551	19.68	4.73	49.61	43.8	44.79	0.89	8.2	1.82	0.65
E4	602	24.08	5.02	54.94	49.2	46.09	1.32	7.9	2.41	0.72
E5	707	20.20	5.05	56.25	45.0	48.62	1.11	8.6	3.49	1.03
E6	885	16.09	4.73	55.31	36.1	46.14	0.80	7.6	2.43	0.73
E7	797	15.63	4.96	66.87	41.3	56.40	0.43	6.7	2.29	0.82
E8	575	30.26	4.78	48.45	51.5	42.06	1.84	11.0	3.17	0.87
E9	802	20.05	4.92	66.43	44.9	56.65	0.68	8.4	2.79	0.84
E10	672	16.39	5.21	74.23	45.9	61.05	0.80	7.4	5.39	1.27
Medelvärden	672,20	20,91	4,94	59,49	45,65	50,57	1,04	8,51	3,07	0,90
Standard- avvikelse	133,37	4,30	0,16	7,99	4,52	6,01	0,38	1,21	0,95	0,18
Varians	17787,76	18,53	0,02	63,83	20,41	36,14	0,14	1,47	0,90	0,03

Tabell 11. SVIT-värden i elevtexter, uppgiften *Skrivregler och sms*, betyg A.

Elevtext	Löp-ord	Antal ord/ mening	Antal bokst./ord	OVIX	LIX	LVIK	Andel bisatser	Träddjup	Subst./Pron- kvot	Nominal- kvot
A1	773	21,47	5,16	68,9	51,1	60,3	0,69	8,1	4,87	1,12
A2	798	21	4,93	61,17	44,3	51,8	1,18	8,8	2,26	0,89
A3	771	15,42	5,22	70,36	43	58,61	0,72	7,3	3,18	1,02
A4	751	20,86	5,3	61,99	51,4	52,43	1,11	8,9	4,62	1,27
A5	762	21,17	4,73	56,84	42,4	48,02	1,28	8,6	2,29	0,77
A6	654	17,68	4,93	59,97	41,4	50,48	0,51	7,9	3,14	0,96
A7	713	14,85	5,66	72,1	47,5	61,74	0,44	7,2	6,23	1,32
A8	778	20,47	5,07	65,34	44,1	54,75	0,89	8,4	2,29	0,92
A9	856	27,61	5,62	74,97	59,3	63,64	1,48	10	3,74	1,14
A10	750	19,23	5,33	79,16	49,6	66	0,77	8,6	2,56	0,94
A11	761	21,14	5,01	61,41	45,4	51,39	0,86	8,3	2,36	0,94
A12	748	15,58	5,3	71,3	45	58,58	0,58	7,3	4,34	1,21
A13	681	20,64	4,6	53,25	42,2	45,31	0,79	8,1	1,58	0,61

A14	632	19,75	4,97	70,26	46	59,74	0,88	7,9	3,18	0,9
A15	694	27,76	4,69	65,49	50	53,64	1,48	9,9	2,16	0,79
A16	802	19,56	5,23	64,87	47,7	54,67	0,8	8,1	3,38	1,01
A17	774	17,59	5,11	70,04	44,3	58,41	1,02	7,9	2,28	1
A18	759	15,49	5,32	68,06	44,9	55,67	0,37	6,5	5,46	1,13
A19	727	24,23	5,07	58,84	50,6	50,46	1,3	8,8	2,07	0,77
A20	889	21,17	4,95	59,11	44,9	50,13	0,98	8,6	2,47	0,85
Medelvärden	748,90	20,02	5,10	65,40	46,50	55,06	0,89	8,24	3,22	0,98
Standard- avvikelse	61,86	3,45	0,27	6,39	4,22	5,30	0,32	0,81	1,21	0,17
Varians	3827,23	11,88	0,07	40,85	17,81	28,06	0,10	0,65	1,46	0,03

Tabell 12. SVIT-värden i elevtexter, uppgiften *Skrivregler och sms*, betyg C.

Elevtext	Löp-ord	Antal ord/ mening	Antal bokst./ord	OVIX	LIX	LVIX	Andel bisatser	Träd- djup	Subst/Pron- kvot	Nominal- kvot
C1	771	21,42	4,96	68,11	46,3	58,81	1,22	8,9	2,41	0,84
C2	760	20,54	4,93	67,72	46,5	56,97	0,81	9,2	2,26	0,9
C3	622	17,77	4,71	57,64	39,3	48,95	0,74	7,4	1,62	0,73
C4	668	25,69	4,8	62,93	46,5	51,05	0,88	8,3	2,87	0,95
C5	604	17,26	5,02	66,45	40,8	56,48	0,66	7,3	2,36	0,79
C6	773	17,57	5,25	74,1	44,6	62,64	0,52	7,5	2,74	1,05
C7	655	18,19	5,27	62,93	47,8	53,92	0,97	8,1	5,54	1,37
C8	869	25,56	4,58	57,12	44	49,73	1,03	8,6	2,31	0,82
C9	757	25,23	4,53	57,76	45	48,55	1,23	9,7	1,15	0,64
C10	649	15,45	4,99	74,34	40,9	61,7	0,5	6,6	1,99	0,74
C11	628	23,26	5,09	55,7	48,4	48,02	1,19	9,6	3,45	1,08
C12	548	21,92	4,51	56,99	41,8	47,32	1,2	9,5	1,66	0,79
C13	768	15,06	4,78	59,84	36,7	47,93	0,75	7,1	1,74	0,68

C14	757	22,94	4,86	66,06	46,5	55,8	1,45	9,3	2,72	0,87
C15	649	22,38	4,93	67,17	46	55,23	0,55	8,8	2,81	0,9
C16	586	27,9	4,77	53,34	50,1	46,55	1,62	10,2	1,78	0,61
C17	760	19	5,11	67,96	45,1	56,57	0,82	8,2	4,14	1,17
C18	660	20,62	5,1	72,1	48,4	57,79	0,81	8,5	3,9	1,05
C19	657	23,46	4,8	61,95	45,8	52,32	1,18	9,5	1,52	0,73
C20	651	16,27	4,73	62,02	38,7	51,26	0,75	7,5	1,62	0,7
Medel- värdet	689,60	20,87	4,89	63,61	44,46	53,38	0,94	8,49	2,53	0,87
Standard- avvikelse	79,30	3,62	0,21	6,02	3,52	4,77	0,30	0,99	1,04	0,19
Varians	6288,74	13,14	0,05	36,22	12,41	22,76	0,09	0,97	1,08	0,04

Tabell 13. SVIT-värden i elevtexter, uppgiften *Skrivregler och sms*, betyg E.

Elevtext	Löp-ord	Antal ord/ mening	Antal bokst./ord	OVIX	LIX	LVIX	Andel bisatser	Träd- djup	Subst/Pron- kvot	Nominal- kvot
E1	826	16,52	4,64	56,53	37,6	49,47	0,92	7,4	2,05	0,71
E2	581	17,09	4,86	57,81	42,6	50,93	1,12	7,8	2,73	0,83
E3	567	21	4,66	56,22	39,2	47,01	0,74	8,9	2,29	0,75
E4	617	19,9	4,6	49,86	40,3	43,34	1	8,1	1,48	0,6
E5	754	19,33	5,38	75,9	49,7	64,92	0,85	8,2	4,81	1,1
E6	578	34	4,64	47,84	55,3	41,93	0,88	9,2	3,69	1,03
E7	764	23,15	5,03	69,82	48	58,89	1,15	9,2	2,41	0,86
E8	641	18,31	4,82	59,91	40,6	52,52	0,8	7,3	2,28	0,79
E9	589	21,04	4,73	61,97	41,2	52,41	0,93	8,5	2,05	0,74
E10	602	18,81	4,36	52,83	34,9	46,67	1	8,1	1,59	0,66
E11	693	23,1	4,49	56,39	41,6	46,84	1,27	8,8	1,51	0,61
E12	717	22,41	4,55	55,99	42,5	49,54	1,22	9	1,45	0,67
E13	635	20,48	5,05	67,4	46,5	58,22	0,84	8,8	2,42	0,94

E14	803	26,77	4,67	59,24	48,2	50,75	1,1	8,9	1,36	0,63
E15	595	20,52	4,84	63,33	41,4	51,52	0,93	7,8	2,02	0,64
E16	666	26,64	4,65	52,57	48	45,61	1,2	11,4	1,89	0,96
E17	636	25,44	4,65	63,02	44,8	54,18	0,8	8,5	2,15	0,82
E18	809	32,36	4,51	48,92	49,8	43,43	2	10,9	2,04	0,65
E19	509	18,18	4,91	63,97	42,7	52,22	1,11	8,4	2,04	0,84
E20	399	16,62	5,34	65,56	44,7	51,97	0,54	7,4	5,5	1,39
Medel- värdet	649,05	22,08	4,77	59,25	43,98	50,62	1,02	8,63	2,39	0,81
Standard- avvikelse	104,98	4,74	0,26	7,09	4,74	5,51	0,29	1,02	1,06	0,19
Varians	11020,55	22,47	0,07	50,23	22,46	30,31	0,08	1,04	1,12	0,04

Tabell 14. SVIT-värden i texthäftet.

Källtext	Löppord	Antal ord/ mening	Antal bokst./ ord	OVIX	LIX	LVIX	Andel bisatser	Träddjup	Subst/Pron- kvot	Nominal- kvot
Källtext 1	1241	18,8	4,8	65,72	40,6	56,91	0,82	7,8	1,97	0,77
Källtext 2	1009	12,77	5,42	90,09	41,2	76,85	0,32	5,8	3,76	1,01
Källtext 3	1246	14	5,09	72,24	39,9	61	0,4	6,6	2,93	0,96
Källtext 4	1442	17,37	4,79	58,66	39,1	50,03	0,75	7,6	1,95	0,79
Källtext 5	719	17,54	5,18	85,54	44,9	71,09	0,41	7,2	3,4	1,06
Källtext 6	935	21,25	5,15	72,46	46,6	62,71	0,93	8,4	2,03	0,76
Källtext 7	1574	12,9	4,81	74,91	35,3	61,37	0,47	6,2	2,16	0,89
Källtext 8	1981	11,65	5,11	84,44	37,5	72,03	0,28	5,8	3,49	1,17
Källtext 9	275	11,96	5,32	86,9	40	67,57	0,48	5,5	3,05	1,07
Källtext 10	402	14,36	5,28	91,65	41,2	75,57	0,21	6,2	6,62	1,4
Källtext 11	632	15,05	4,92	66,88	38,5	57,02	0,4	6,7	1,47	0,78
Källtext 12	658	16,45	4,73	64,48	37,7	55,01	0,88	7,5	1,39	0,78
Källtext 13	381	14,65	4,7	62,21	34,1	52,17	0,46	7,2	1,22	0,6
Källtext 14	356	9,37	5,17	76,69	33	61,72	0,26	4,8	1,3	0,66
Medel- värdet	917,93	14,87	5,03	75,21	39,26	62,93	0,51	6,66	2,62	0,91
Standard- avvikelse	502,02	3,05	0,23	10,54	3,62	8,23	0,23	0,97	1,39	0,21
Varians	252019,92	9,32	0,05	111,17	13,11	67,73	0,05	0,94	1,92	0,04

Tabell 15. SVIT-värden i *Språk och stil*-artiklarna.

Texter	Löp- ord	Antal ord/ mening	Antal bokst./ ord	OVIX	LIX	LVIX	Andel bisatser	Träddjup	Subst/Pron- kvot	Nominal- kvot
SoS 1	7573	18,16	5,52	67,99	49,8	59,31	0,54	7,5	3,77	1,36
SoS 2	8380	20,59	5,78	59,94	57,7	53,43	0,5	7,7	8,67	1,99
SoS 3	8929	17,17	5,93	67,7	52	57,09	0,56	7,2	10,68	1,69
SoS 4	4283	19,74	5,86	69,07	56,1	58,65	0,45	8,2	5,78	1,49
SoS 5	3190	19,45	5,46	72,49	48,6	60,88	0,76	7,8	5,06	1,33
SoS 6	8135	20,54	5,6	52,95	55,3	47,33	0,7	8	6,63	1,58
SoS 7	7799	18,66	5,44	68,9	49,8	59,81	0,51	7,4	8,07	1,71
SoS 8	9098	18,38	5,61	72,24	50,7	60,99	0,59	7,8	4,87	1,44
SoS 9	7101	28,29	5,89	77,88	63,4	67,02	0,73	9,3	5,35	1,65
SoS 10	4110	22,58	5,97	62,67	59,6	53,46	0,57	8,2	10,25	2,02
SoS 11	7520	17,9	5,31	69,3	47,3	59,68	0,45	7	3,88	1,29
SoS 12	8966	16,39	5,71	67,18	50,7	58,31	0,52	7	5,14	1,37
SoS 13	3377	16,89	5,47	69,4	47,9	58,75	0,35	7,3	6,1	1,73
Medel- värdet	6805	19,60	5,66	67,52	52,99	58,05	0,56	7,72	6,48	1,59
Standard- avvikelse	2135,93	3,00	0,21	5,94	4,79	4,50	0,11	0,60	2,18	0,23
Varians	4562176,1	9,0	0,0	35,3	23,0	20,3	0,0	0,4	4,7	0,1

Bilaga 2. Tablåer över analyserade texter

Tablå 5. Förteckning över elevtexter som analyseras i studien.

<i>Uppgift Särskrivning och stavfel</i>		<i>Uppgift Skrivregler och sms</i>	
Beteckning i studien (betyg och siffra)	Kod i Uppsala elev- textkorpus	Beteckning i studien (betyg och siffra)	Kod i Uppsala elevtextkorpus
A1	C167	A1	C165
A2	C168	A2	C170
A3	C175	A3	C171
A4	C180	A4	C174
A5	C181	A5	C176
A6	C182	A6	C177
A7	C183	A7	C178
A8	C184	A8	C179
A9	C185	A9	C192
A10	C186	A10	C193
C1	C115	A11	C194
C2	C116	A12	C195
C3	C118	A13	C196
C4	C120	A14	C197
C5	C124	A15	C198
C6	C127	A16	C202
C7	C128	A17	C204
C8	C129	A18	C210
C9	C134	A19	C212
C10	C138	A20	C213
E1	C74	C1	C117
E2	C84	C2	C123
E3	C86	C3	C125
E4	C88	C4	C126
E5	C92	C5	C130
E6	C95	C6	C131
E7	C99	C7	C132
E8	C100	C8	C133

E9	C101	C9	C136
E10	C103	C10	C137
		C11	C139
		C12	C142
		C13	C144
		C14	C147
		C15	C149
		C16	C153
		C17	C154
		C18	C155
		C19	C156
		C20	C158
		E1	C65
		E2	C66
		E3	C67
		E4	C68
		E5	C69
		E6	C70
		E7	C71
		E8	C72
		E9	C75
		E10	C76
		E11	C78
		E12	C79
		E13	C80
		E14	C81
		E15	C82
		E16	C83
		E17	C85
		E18	C87
		E19	C89
		E20	C90

Tablå 6. Förteckning över källtexter som analyseras i studien (provets texthäfte vårterminen 2014).

Beteckning i studien	Text
Källtext 1	Lindström, Fredrik, 2000: ett utdrag ur <i>Världens dåligaste språk</i> . Stockholm: Albert Bonniers förlag.
Källtext 2	Borg, Martin, 2004: Hur en djup fryst kyckling lever. <i>www.ne.se</i> (publicerad 20.2.2004).
Källtext 3	Melin, Lars, 2007: Vad är det för fel på ett fel? <i>Forskning & Framsteg</i> 2007:1.
Källtext 4	Andersson, Lars-Gunnar, 2013: Därför blir vi så arga på språkfel. <i>Språktidningen</i> 2013:4.
Källtext 5	Melin, Lars, 2000: ett utdrag ur <i>Människan och skriften</i> . Stockholm: Norstedts.
Källtext 6	Lotsson, Anders, 2007: Sms är kort, inte snålt. <i>Computer Sweden</i> 2007:52.
Källtext 7	Bellander, Theres, 2009: Intensiv ordväxling kräver känsla. <i>Språktidningen</i> 2009:6.
Källtext 8	Engstrand, Olle, 2012: ett utdrag ur <i>Hur låter svenskan, ejenglien?</i> Stockholm: Norstedts.
Källtext 9	Wittensten, Lars, 2012: Schyst/juste/schysst/sjyst/just. <i>Upsala Nya Tidning</i> 9.12.2012.
Källtext 10	Norlander, Göran, Kristiansson, Kerstin & Stenberg, Hans, 1999: Motion 1999/2000:Kr221. Stavningsreform. <i>www.riksdagen.se</i> (hämtad 16.10.2013).
Källtext 11	Svenska språknämnden, 2005: utdrag ur <i>Språkkriktighetsboken</i> . Stockholm: Norstedts akademiska förlag.
Källtext 12	Strömquist, Siv, 2011: ett utdrag ur <i>Vart är vart på väg?</i> Stockholm: Norstedts.
Källtext 13	Bergman-Claeson, Görel, 2001: De, dem eller dom? <i>Ergo</i> 2001:7.
Källtext 14	Josephson, Olle, 2010: Det finns en utväg ur dem-problemet. <i>Svenska Dagbladet</i> 29.11.2010.

Tablå 7. Förteckning över *Språk och stil*-artiklar som analyseras i studien.

Beteckning i studien	Tidskrifts-nummer	Artikel
SoS 1	Språk och stil, NF 24, 2014	Näslund, Shirley. ”Jag orkar inte längre” – ”Du orKAr”: Språkliga bemötanden av föderskors modlöshet och förtvivlan.
SoS 2	Språk och stil, NF 24, 2014	Bohnacker, Ute & Lindgren, Josefin. Fundament, formellt subjekt och frekvens: Ordföljdsmonster i svenska, nederländska och hos vuxna inlärare av svenska.
SoS 3	Språk och stil, NF 24, 2014	Vogel, Anna. Laddade ord: Hur nya uttryck, som kategoriserar människor, tas in i svenskan.
SoS 4	Språk och stil, NF 24, 2014	Hagren Idevall, Karin. ”Politiskt korrekt” och normalisering av rasism: En diskursanalys av positioneringar och underliggande perspektiv i ett kommentarsfält.
SoS 5	Språk och stil, NF 24, 2014	Borgström, Erik & Ledin, Per. Bedömarvariation: Balansen mellan teknisk och hermeneutisk rationalitet vid bedömning av skrivprov.
SoS 6	Språk och stil, NF 24, 2014	Tolvanen, Eveliina. Auktoritet i finlandssvenska och sverigesvenska myndighetstexter: En studie av lexikogrammatiska val och ergativa roller.
SoS 7	Språk och stil, NF 24, 2014	Stenberg-Sirén, Jenny. Skriftspråksnära talat standardspråk: En studie av uttalsdrag i finlandssvenska tv- och radionyheter 1970–2009.
SoS 8	Språk och stil, NF 25, 2015	Ledin, Per & Machin, David. Universitet som en multimodal marknadsplats: Designen av en nyliberal managementdiskurs.
SoS 9	Språk och stil, NF 25, 2015	Eriksson, Olof. Kontrastiv språkforskning på översättningsanalytisk grund – exemplet presens particip.
SoS 10	Språk och stil, NF 25, 2015	Lindqvist, Yvonne. Det skandinaviska översättningsfältet – finns det?
SoS 11	Språk och stil, NF 25, 2015	Fremer, Maria. <i>Men kör såhär, så slipper du göra bort dig</i> : Hur svenskans du-reform återspeglas i reklamfilmer.
SoS 12	Språk och stil, NF 25, 2015	Hillbom, Annika. Känseladjektiv i svenskan: Kategorier och figurativa användningar.
SoS 13	Språk och stil, NF 25, 2015	Scheglov, Andrej. Magnus Erikssons landslag: Handskrifterna i Sankt Petersburg och Moskva.

Rapportserien Svenska i utveckling

1. Margareta Andersson, 1995: Prov i modersmålet. En översikt över centralt utarbetade prov i Sverige och nio andra länder. (= FUMS Rapport nr 176.) 40 s.
2. Catharina Nyström, 1996: Skrivandet, kursplanerna och läromedlen. En studie av gymnasieskolans läromedel i svenska 1970–1995. (= FUMS Rapport nr 178.) 52 s.
3. Birgitta Garne, 1996: Att pröva skrivförmågan med nationella prov. En presentation av proven i svenska för skolår 2 och 5. (= FUMS Rapport nr 180.) 55 s.
4. Birgitta Garne & Anne Palmér, 1996: Samspel och Kommunikation – om utvecklandet av nationella kursprov i svenska för gymnasieskolan. (= FUMS Rapport nr 182.) 73 s.
5. Orla Vigsø, 1996: Valgplakaten som kommunikation og marketing. (= FUMS Rapport nr 183.) 56 s.
6. Eva Östlund-Stjärnegårdh, 1997: Skriva debattartiklar i skolan – går det? Om texttypen debattartikel i nationella prov. (= FUMS Rapport nr 186.) 40 s.
7. Hanna Sofia Öberg, 1997: Referensbindning i elevuppsatser. En preliminär modell och en analys i två delar. (= FUMS Rapport nr 187.) 109 s.
8. Björn Melander, 1998: Inskickat och registrerat – sammanställning av uppgifter rörande de nationella proven i svenska vår- och höstterminen 1996 samt vårterminen 1997. (= FUMS Rapport nr 188.) 53 s.
9. Annika Persson, 1998: Texten och inspirationskällan. En studie av förlagans betydelse för elevers fria textproduktion i skolår 5. (= FUMS Rapport nr 191.) 42 s.
10. Anne Palmér, 1999: Tankar om tal – lärares och elevers syn på muntlig framställning i undervisning och bedömning. (= FUMS Rapport nr 193.) 57 s.
11. Helena Olevard, 1999: ”Tonårsliv”. En pilotstudie av 60 elevtexter från standardproven för skolår 9 åren 1987 och 1996. (= FUMS Rapport nr 194.) 28 s.
12. Eva Östlund-Stjärnegårdh, 1999: Principen och praktiken. En enkätundersökning av lärares syn på bedömning av gymnasieelevers texter. (= FUMS Rapport nr 195.) 40 s.
13. Catharina Nyström & Maria Ohlsson (red.), 1999: Svenska på prov. Arton artiklar om språk, litteratur, didaktik och prov. (= FUMS Rapport nr 196.) 150 s.
14. Catharina Nyström, 2000: Ledfamiljer och referentrelationer. En modell för analys av referensbindning tillämpad på gymnasisttexter. (= FUMS Rapport nr 197.) 59 s.
15. Kerstin Lagrell, 2000: Växa i skrivandet. Om förhållningssättet till de yngre skolbarnens skrivutveckling. (= FUMS Rapport nr 199.) 70 s.
16. Maria Ohlsson, 2001: Säker stil eller rätt i sak? En studie i gymnasisters analyser av en enkät. (= FUMS Rapport nr 204.) 45 s.
17. Anne Palmér, 2002: Röster i samspel. Analys av gymnasieelevers radioprogram. (= FUMS Rapport nr 205.) 59 s.
18. Helena Andersson, 2002: Svenska som första- och andraspråk – en jämförande studie av texter från skolår 9. (= FUMS Rapport nr 208.) 97 s.
19. Catharina Nyström, 2003: Argumentera! En presentation av gymnasisters texter i databasen ARGUS. (= FUMS Rapport nr 209.) 61 s.
20. Katharina Hallencreutz, 2003: Särskrivningar och andra skrivningar i elevspråk. (= FUMS Rapport nr 210.) 101 s.
21. Maria Eklund Heinonen, 2005: Godkänd eller underkänd? Hur processbarhetsteorin kan tillämpas vid muntliga språktester av andraspråksinlärare. (= FUMS Rapport nr 215.) 59 s.

22. Inger Gröning, 2006: Interaktion och lärande i flerspråkiga klasser. (= FUMS Rapport nr 218.) 89 s.
23. Eva Östlund-Stjärnegårdh, 2006: Att förmedla egna och andras tankar. Om gymnasisters källhantering i det nationella provets skrivuppgift. (= FUMS Rapport nr 219.) 52 s.
24. Karin Wesslén, 2008: Processkrivande – en etablerad metod på gymnasiet? (= FUMS Rapport nr 225.) 55 s.
25. Ciolek Laerum, Beatrice, 2009: Elever skriver och lärare bedömer – en studie av elevtexter i åk 9. (= FUMS Rapport nr 226.) 60 s.
26. Palmér, Anne, 2010: Att bedöma det muntliga. Utvärdering av ett delprov i gymnasieskolans nationella kursprov, Svenska B. (= FUMS Rapport nr 227.) 61 s.
27. Nyström Höög, Catharina, 2010: Mot ökad diskursivitet? Skrivutveckling speglad i provtexter från årskurs 5 och årskurs 9. (= FUMS Rapport nr 228.) 62 s.
28. Hagberg-Persson, Barbro, Berg, Elisabeth & Lagrell, Kerstin, 2010: Ämnesprov i svenska och svenska som andraspråk för årskurs 3 – en utvärderingsomgång. (= FUMS Rapport nr 229.) 89 s.
29. Nordberg, Olle, 2013: Att finnas till som läsare – skönlitterär läsning i ett elevperspektiv. Didaktiska tillämpningar av en empirisk studie baserad på elevers egna texter om sin läsning. 91 s.
30. Hagberg-Persson, Barbro & Wiberg, Cecilia, 2013: Elever visar vad de kan. Två studier kring ämnesprovet i svenska och svenska som andraspråk för årskurs 3. 61 s.
31. Palmér, Anne, 2013: Nationella skrivprov baserade på två olika läroplaner. Genre, kommunikationssituation och skrivdidaktiska diskurser. 72 s.
32. Schüssler, Eija L., 2014: Visuell kommunikation – en modell för bedömning av gymnasieelevers bruk av bildspel vid muntliga anföranden. 46 s.
33. Hagberg-Persson, Barbro, 2015: Nationellt prov i årskurs 3 – en redovisning av ämnesprovet i svenska och svenska som andraspråk. 53 s.
34. Mark, Malin & Palmér, Anne, 2017: En utvecklande tolkningsgemenskap? Matrisanvändning, interaktion och kontext i bedömningsamtal om ett nationellt prov i muntlig framställning. 91 s.
35. Nordberg, Olle, 2019: Berättelser som förändrar. Utvärdering och didaktisk diskussion kring ett nationellt läsprojekt för ungdomar. 161 s.
36. Dalberg, Tobias, 2019: Samstämmighet i skrivbedömning. Statistisk analys vid bedömning av två nationella skrivprov. 47 s.
37. Dalberg, Tobias, Zachiu, Martina, Shahsavari, Negin, Eriksson, Kristina & Hussenius, Siri, 2020: Samstämmighet i läsbedömning. Statistisk analys vid bedömning av ett nationellt läsförståelseprov. 47 s.
38. Hillbom, Annika, 2022: En fråga om ord? Automatisk och manuell analys av gymnasieelevers språkanvändning i nationella provet i kursen svenska 3. 121 s.

Beställning

För beställning av rapporter i serien Svenska i utveckling, kontakta någon av provgruppens administratörer. Kontaktuppgifter finns på följande webbplats: natprov.nordiska.uu.se. Där finns även vissa av rapporterna nätpublicerade.