



PH.D.-AFHANDLING

MORTEN TANNERT

SKRIFTSPROGLIG UDVIKLING I GRUNDSKOLENS DANSKFAG

Et korpusbaseret studie af elevtekster fra 5.-9. klassestrin



AARHUS
UNIVERSITET

DPU - DANMARKS INSTITUT FOR PÆDAGOGIK OG UDDANNELSE

Skriftsproglig udvikling i grundskolens danskfag

*Et korpusbaseret studie af elevtekster
fra 5.-9. klassetrin*

Ph.d.-afhandling ved Morten Tannert

Danmarks institut for Pædagogik og Uddannelse (DPU), Aarhus Universitet

2022

Skriftsproglig udvikling i grundskolens danskfag: Et korpusbase- ret studie af elevtekster fra 5.-9. klassetrin

English title: L1 Writing Development in the Danish Primary and Lower Sec-
ondary School: A corpus-based study of student texts from the 5th-9th grade

Af cand.pæd. i didaktik (dansk) Morten Tannert

Ph.d.-afhandling indleveret til Danmarks institut for Pædagogik og Uddan-
nelse (DPU), Faculty of Arts, Aarhus Universitet, i august 2022.

Hovedvejleder:

Lektor Ulf Dalvad Berthelsen, Institut for Kommunikation og Kultur, Aarhus
Universitet.

Bivejleder:

Lektor Felix Weiss, Danmarks institut for Pædagogik og Uddannelse
(DPU), Aarhus Universitet.

ISBN: 978-87-7507-532-4

DOI: 10.7146/aulsps-e.466



DPU - DANMARKS INSTITUT FOR PÆDAGOGIK OG UDDANNELSE
AARHUS UNIVERSITET

Indholdsfortegnelse

Resumé	11
Summary	15
Forord	19
Kapitel 1: Indledning og forskningsoversigt	20
1.1 Indledning	20
1.2 Baggrund og relevans	23
1.2.1 Hvorfor forske i skriftsproglig udvikling på et kvantitativt grundlag?	25
1.2.2 Hvorfor koble det lingvistiske med et psykologisk og sociologisk perspektiv? ...	27
1.3 Forskningsoversigt.....	28
1.3.1 Rammesætning af forskningsoversigten.....	29
1.3.2 Skandinavisk skriveforskning.....	31
1.3.3 International skriveforskning med et tekstorienteret og kvantitativt fokus	47
Kapitel 2: Afhandlingens teoretiske forudsætninger og centrale begreber .	56
2.1 Skriveudviklingsmodeller i afhandlingens analyser	56
2.2 Hvad er skriftsprog, og hvad er skrivning?.....	58
2.3 Hvad er skriftsproglig udvikling?	59
2.3.1 Skriftsproglig udvikling som modningsproces.....	60
2.3.2 Skriftsproglig udvikling i et sociologisk perspektiv	61
2.3.3 Writers-Within-Community-modellen	62
2.4 Elevtekstens pragmatik: Om teksthandlinger i skolske kontekster	64
2.4.1 Elevteksten i en Griceansk optik	65
2.4.2 Skolens metagenrer og deres rolle i danskfaget	66
2.4.3 Elevers skriftsprog som intersprog	68
2.5 Afhandlingens tekstbegreber	69

2.5.1 Den beskrivende teksttype.....	72
2.5.2 Den berettende teksttype	73
Kapitel 3: Forskningsdesign og metode	75
3.1 Metodisk rammesætning.....	75
3.2 Datamateriale og korpusdesign	77
3.2.1 Skriveprompt	78
3.2.2 Korpusdesign	81
3.2.3 Surveydesign.....	87
3.3 Analysestrategier	88
3.3.1 Natural Language Processing (NLP)	89
3.3.2 Den diakrone analysestrategi	89
3.3.3 Den synkrone analysestrategi	97
3.3.4 Analysestrategierne og kvalitetskriterier for forskning.....	99
3.4 Videnskabsteoretiske perspektiver	105
3.4.1 Strukturalisme som ontologisk og metodologisk forståelsesramme	105
3.4.2 Om operationalisering af sproglige fænomener	107
3.4.3 Computational hermeneutik som fortolkningsramme.....	109
3.5 Forskningsetiske perspektiver	111
3.5.1 Elevtekster som data og principper for anonymisering	111
3.5.2 Analyseernes konsekventielle validitet	112
Kapitel 4: Teoretiske beskrivelser og operationaliseringer af studiets variable.....	115
4.1 Lingvistiske variable	115
4.1.1 Variable knyttet til produktivitet	115
4.1.2 Variable knyttet til leksis	117
4.1.3 Variable knyttet til syntaks	121

4.2 Ekstra-lingvistiske variable.....	126
4.2.1 Askriptive variable	127
4.2.2 Literacy-orienterede variable	130
4.2.3 Øvrige ekstra-lingvistiske variable.....	135
4.3 Overvejelser over konstruktvaliditet i afhandlingens ekstra-lingvistiske variable...	137
Kapitel 5: Den diakrone analysedel.....	140
5.1 Indledende rammesætning.....	140
5.2 Kvantitativ analyse af produktivitetsvariabel	140
5.2.1 Variabel 1: Tekstlængde	141
5.2.2 Opsummering	144
5.3 Tekstudvalg 1: Produktivitet	145
5.3.1 Elevtekst 1 og 2.....	146
5.3.2 Elevtekst 3 og 4.....	148
5.3.3 Elevtekst 5.....	150
5.3.4 Opsummering på analyse af tekstudvalg 1.....	152
5.4 Kvantitative analyser af leksisvariable.....	154
5.4.1 Variabel 2: Ordvariationsindeks (OVIX).....	154
5.4.2 Variabel 3: Gennemsnitlig ordlængde	156
5.4.3 Opsummering	158
5.5 Tekstudvalg #2 (leksis)	159
5.5.1 Elevtekst 6 og 7.....	160
5.5.2 Elevtekst 8 og 9.....	163
5.5.3 Elevtekst 10 og 11	165
5.5.4 Opsummering på analyse af tekstudvalg 2	168
5.6 Kvantitative analyser af syntaksvariable	170

5.6.1 Variabel 4: Gennemsnitlig sætningslængde	170
5.6.2 Variabel 5: Gennemsnitlig dependensafstand	173
5.6.3 Opsummering	175
5.7 Tekstudvalg 3 (syntaks).....	176
5.7.1 Elevtekst 12, 13 og 14.....	177
5.7.2 Elevtekst 15 og 16.....	179
5.7.3 Elevtekst 17, 18, 19 og 20	182
5.7.4 Opsummering på analyse af tekstudvalg 3.....	186
5.8 Delkonklusion på den diakrone analysedel	189
5.8.1 Skriveudviklingspointer.....	189
5.8.2 Metodologiske pointer.....	199
Kapitel 6: Den synkrone analysedel	205
6.1 Indledende rammesætning af analysen	205
6.2 De ekstra-lingvistiske variables betydning i et komparativt perspektiv	206
6.2.1 Tekstlængde	206
6.2.2 Ordlængde	210
6.2.3 OVIX	213
6.2.4 Sætningslængde.....	215
6.2.5 Dependensafstand	218
6.2.6 Opsummering af analysen af de ekstra-lingvistiske variables betydning	220
6.3 Askriptive forskelle i skriftsproglig udvikling.....	222
6.3.1 Betydningen af social baggrund	225
6.3.2 Betydningen af køn.....	232
6.4 Delkonklusion på den synkrone analysedel	239
6.4.1 Skriveudviklingens dobbeltperspektiv	239

6.4.2 Skriveudviklingens ulighedsdimension	240
6.4.3 Skriveudviklingens kønsdimension	243
6.4.4 Skriveudviklingens literacydimension	245
Kapitel 7: Diskussion og konklusion	248
7.1 Afhandlingens implikationer for studier af skriveudvikling.....	248
7.1.1 Potentialer og begrænsninger ved lingvistiske beskrivelser af skriveudv	248
7.1.2 Integreret skriveudviklingsmodel eller metodisk perspektivskifte.....	251
7.2 Afhandlingens danskdidaktiske implikationer	252
7.2.1 Om forholdet mellem lingvistik og didaktik.....	252
7.2.2 Om relevansen af viden om skriftsproglig udv. for skriveevaluering	255
7.2.3 Om relevansen af viden om skriftsproglig udv. for skriveundervisning.....	258
7.3 Konklusion.....	259
7.3.1 Forskningsspørgsmål 1.....	259
7.3.2 Forskningsspørgsmål 2.....	264
7.3.3 Forskningsspørgsmål 3.....	266
7.3.4 Forslag til og overvejelser over fremtidig forskning	268
Referencer	271
Bilag 1: Elevtekster i sætningsskema	294
Bilag 2: Topologisk skema for nominalhelheder	315
Bilag 3: Skriveprompt	320
Bilag 4: Måleegenskaber for 'læsevaner' og 'skrivelyst'	322
Bilag 5: Brev til forældre og samtykkeerklæring	324
Bilag 6: Brev til de deltagende lærere	326

Fortegnelse over figurer og tabeller

Figur 1: Kognitive mekanismer i 'Writers-Within-Community'-modellen	63
Figur 2: Formel for beregning af ordvariationsindeks (OVIX)	118
Figur 3: Eksempel på dependensanalyse af en sætning fra en berettende tekst (6. årg.)	124
Figur 4: Eksempel på dependensanalyse af en sætning fra en berettende tekst (5. årg.)	125
Figur 5: Boksplot for tekstlængde i berettende og beskrivende tekst	141
Figur 6: Boksplot for OVIX i samlet tekst per elev (kombineret besk. og ber. tekst)	154
Figur 7: Boksplot for gennemsnitlig ordlængde i berettende og beskrivende tekst	157
Figur 8: Boksplot for sætningslængde i berettende og beskrivende tekst	171
Figur 9: Boksplot for gennemsnitlig dependensafstand i berettende og beskrivende tekst	173
Figur 10: Model over de ekstra-lingvistiske variables indbyrdes relationer	223
Figur 11: Model af grundlæggende dimensioner af tekstaktiviteter i skolen	250
Figur 12: Grafisk visualisering af Shulmans (1986) 'pedagogical content knowledge'	254
<hr/>	
Tabel 1: Oversigt over afhandlingens tekstbegreber	72
Tabel 2: Overblik over datamateriale	77
Tabel 3: Oversigt over Halliday & Hassans (1976) typer af kohæsive bånd	94
Tabel 4: Hansen & Heltofts (2011a) topologiske skema for nominalhelheder	95
Tabel 5: Adapteret udgave af Diderichsens (1946) helsætningskema	96
Tabel 6: Deskriptiv statistik for det datamateriale, der ligger til grund for OVIX-analyserne	119
Tabel 7: Oversigt over fordelingen af social baggrund i ESeC-grupper (dominansmodel)	128
Tabel 8: Items til skrivelystindekset	132
Tabel 9: Items til variabelen 'skrivevaner'	133
Tabel 10: Fordeling i de to grupper, som udgør variabelen 'skrivevaner'	134
Tabel 11: Items til variabelen 'læsevaner'	134
Tabel 12: Oversigt over koder til variabelen 'billedvalg'	136
Tabel 13: Deskriptiv statistik for fordelingen af variabelen 'tekstlængde' for berettende tekster	142
Tabel 14: Deskriptiv statistik for fordelingen af variabelen 'tekstlængde' for beskrivende tekster	143
Tabel 15: Deskriptiv statistik for fordelingen af variabelen 'OVIX' for kombineret tekst	155
Tabel 16: Deskriptiv statistik for fordelingen af variabelen 'ordlængde' for berettende tekst	157
Tabel 17: Deskriptiv statistik for fordelingen af variabelen 'ordlængde' for beskrivende tekst	158

Tabel 18: Deskriptiv statistik for fordelingen af variabelen 'sætningslængde' for berettende tekst	171
Tabel 19: Deskriptiv statistik for fordelingen af variabelen 'sætningslængde' for beskrivende tekst	172
Tabel 20a: Deskriptiv statistik for fordelingen af variabelen 'gns. dependensafstand' for ber. tekst	173
Tabel 20b: Deskriptiv statistik for fordelingen af variabelen 'gns. dependensafstand' for bes. tekst	174
Tabel 21: Multipel lineær regressionsanalyse for tekstlængde (beskrivende tekst)	207
Tabel 22: Multipel lineær regressionsanalyse for tekstlængde (berettende tekst)	209
Tabel 23: Multipel lineær regressionsanalyse for ordlængde (beskrivende tekst)	211
Tabel 24: Multipel lineær regressionsanalyse for ordlængde (berettende tekst)	212
Tabel 25: Multipel lineær regressionsanalyse for OVIX (kombineret tekst)	214
Tabel 26: Multipel lineær regressionsanalyse for sætningslængde (beskrivende tekst)	216
Tabel 27: Multipel lineær regressionsanalyse for sætningslængde (berettende tekst)	217
Tabel 28: Multipel lineær regressionsanalyse for gns. dependensafstand (beskrivende tekst)	218
Tabel 29: Multipel lineær regressionsanalyse for gns. dependensafstand (berettende tekst)	219
Tabel 30: Lineær regressionsanalyse for social baggrunds betydning for tekstlængde	226
Tabel 31: Lineær regressionsanalyse for social baggrunds betydning for ordlængde	228
Tabel 32: Lineær regressionsanalyse for social baggrunds betydning for sætningslængde	230
Tabel 33: Lineær regressionsanalyse for køns betydning for tekstlængde	233
Tabel 34: Lineær regressionsanalyse for køns betydning for ordlængde	235
Tabel 35: Lineær regressionsanalyse for køns betydning for sætningslængde	237
Tabel 36: Tre koblingspunkter mellem viden om skriveudvikling og evaluering i danskfaget	257
Tabel 37: Elevtekst 12 i sætningsskema	294
Tabel 38: Elevtekst 13 i sætningsskema	295
Tabel 39: Elevtekst 14 i sætningsskema	297
Tabel 40: Elevtekst 15 i sætningsskema	298
Tabel 41: Elevtekst 16 i sætningsskema	303
Tabel 42: Elevtekst 17 i sætningsskema	305
Tabel 43: Elevtekst 18 i sætningsskema	306
Tabel 44: Elevtekst 19 i sætningsskema	307
Tabel 45: Elevtekst 20 i sætningsskema	313
Tabel 46: Nominalfraser i elevtekst 6	315
Tabel 47: Nominalfraser i elevtekst 7	315
Tabel 48: Nominalfraser i elevtekst 8	316

Tabel 49: Nominalfraser i elevtekst 9	316
Tabel 50: Nominalfraser i elevtekst 10	317
Tabel 51: Nominalfraser i elevtekst 11	318
Tabel 52: Faktoranalyse for 'læsevaner'	322
Tabel 53: Faktorloadings for læsevaneitems	322
Tabel 54: Faktoranalyse for 'skrivelyst'	323
Tabel 55: Faktorloadings for skrivelystitems	323

Resumé

I denne afhandling undersøger jeg skriftsproglig udvikling i et korpus af elevtekster fra 5.-9. klassetrin, som er skrevet som en del af danskundervisningen. Korpuset består af to teksttyper: en berettende teksttype og en beskrivende teksttype. Afhandlingens analyse består af to dele. I den første del, som jeg kalder for den diakrone analysedel, undersøges udviklingen af en række sproglige træk i tekstkorpuset henover klassetrin. Denne analyse er dels bygget op om kvantitative analyser af fem lingvistiske variable, dels af analyser af tre tekstudvalg af elevtekster, som belyses kvalitativt gennem leksikogrammatisk analyse. De fem kvantitative, lingvistiske variable er tekstlængde, ordlængde, ordvariationsindeks (OVIX), sætningslængde og dependensafstand. I den anden analysedel, som jeg kalder for den synkrone analysedel, undersøges, hvordan de fem lingvistiske variable hænger sammen med en række ekstra-lingvistiske variable (bl.a. køn, social baggrund, skrivelyst, læsevaner og skrivevaner) knyttet til den enkelte elevskriver. Denne analysedel er således bygget op omkring en række multiple, lineære regressionsmodeller, som estimerer betydningen af de ekstra-lingvistiske variable. Afhandlingen består af syv kapitler.

I kapitel 1 præsenterer jeg afhandlingens formål, forskningsspørgsmål samt baggrund og relevans for studiet. Kapitlet indeholder ligeledes en forskningsoversigt, som falder i to dele. I den første del fokuserer jeg på at beskrive tendenser i den skandinaviske skriveforskning. Forskningsoversigten beskriver således to traditioner i den skandinaviske skriveforskning, som har været dominerende. Der er tale om henholdsvis en lingvistisk samt en etnografisk og didaktisk tradition. Den anden del af forskningsoversigten indeholder en beskrivelse af international skriveforskning med et tekstorienteret og kvantitativt fokus. Her beskriver jeg således to tilgange til skriveforskningen, henholdsvis en diakron og en synkron tilgang. Kapitlet rundes af med en diskussion af dette studies placering og bidrag til den internationale skriveforskning.

I kapitel 2 beskriver jeg afhandlingens skriveudviklingsteoretiske forudsætninger samt begreber, som er centrale for afhandlingens analyser. Kapitlet indledes med en gennemgang af forskellige teoretiske modeller for skriveudvikling. Herefter præsenterer jeg tre teorier, som er vigtige forudsætninger for afhandlingens to analysedele. Det drejer sig dels om en klassisk teori om kognitiv modning, dels om Bernsteins (1971) teori om prog-

lige koder og dels om Grahams (2018) 'Writers-Within-Community'-model. Herefter fokuserer jeg på, hvad der kendetegner elevteksters pragmatik, samt hvilke konsekvenser dette har for, hvordan elevtekster kan forstås som forskningsdata. Til sidst beskriver jeg afhandlingens tekst- og genrebegreber.

Kapitel 3 indeholder beskrivelser af afhandlingens forskningsdesign og metode. Kapitlet indeholder en beskrivelse af afhandlingens datagrundlag samt en beskrivelse af afhandlingens metoder og den måde, de er integreret med hinanden på. Derudover er der en beskrivelse af de to analysestrategier (herunder det sprogteoretiske rammeverk for de kvalitative analyser) samt deres fundering i tre relevante kvalitetskriterier for forskning (trustworthiness, conversion legitimation og intern validitet). Kapitlet afsluttes med en beskrivelse og diskussion af en række videnskabsteoretiske perspektiver på projektets forskningsdesign samt en diskussion af udvalgte forskningsetiske problemstillinger.

I kapitel 4 præsenteres afhandlingens variable. Her fokuseres der på variablenes teoretiske forudsætninger samt deres konkrete operationaliseringer i studiet. Variablerne falder i to overordnede grupper. Den første gruppe er de lingvistiske variable (tekstlængde, ordlængde, OVIX, sætningslængde og dependensafstand), mens den anden gruppe er de ekstra-lingvistiske variable. Sidstnævnte gruppe er inddelt i to undergrupper. Det drejer sig om en gruppe af askriptive variable (social baggrund og køn) samt en gruppe af literacy-orienterede variable (skrivelyst, læsevaner og skrivevaner). Kapitlet rundes af med overvejelser over styrker og svagheder i konstruktvaliditeten af de ekstra-lingvistiske variable.

Kapitel 5 indeholder den diakrone analysedel, hvor jeg undersøger udviklingen af de lingvistiske variable relativt til klasstrin samt de leksikogrammatisk kendetegn ved tre tekstudvalg af elevtekster. Analysen af tekstlængde viste en forholdsvis lineær udvikling i tekstlængde til og med 8. klasstrin, hvorefter der sker et fald til 9. klasstrin. De kvalitative analyser viste efterfølgende, at øget tekstlængde ikke nødvendigvis fører til øget indholdsmæssig variation (flere semantiske felter), men snarere til et mere uddybet indhold (større semantiske felter). Analysen af ordlængde og OVIX viste ikke lineære udviklinger på tværs af klasstrin. OVIX lå forholdsvis stabilt på samme niveau på tværs af klasstrin, mens ordlængde viste en lille forskel ved sammenligning af de to mellemtrinsårgange med de tre udskolingsårgange. De efterfølgende kvalitative analyser indikerede ikke nogen stærk sammenhæng mellem de to leksikale variable og anvendelsen af kohæsive bånd

i teksterne eller kompleksiteten af nominalfraserne. Der var til gengæld markante forskelle relativt til teksttype. Sætningslængde og dependensafstand viste en udvikling i variablerne mellem mellemtrin og udskolingen, mens der ikke var nogen systematisk udvikling i variablerne henover de tre udskolingsårgange. De kvalitative analyser viste, at der særligt ved sætningslængde var stor variation i, hvilke syntaktiske konstruktioner der gemte sig bag de numeriske værdier. Blandt de væsentligste pointer, som kan udledes af den diakrone analysedel, er, at aldersmodning mere er en proxy for udvikling end et strukturerende princip, og at udviklingsplateauer, overforbrug og sproglig variation er vigtige begreber at inddrage for at forstå skriveudvikling i et diakront perspektiv.

I kapitel 6 præsenteres afhandlingens synkrone analysedel, hvor jeg undersøger sammenhængen mellem afhandlingens lingvistiske og ekstra-lingvistiske variable. Analysen indledes med en række multiple lineære regressionsanalyser, hvor de ekstra-lingvistiske variables betydning for de fem lingvistiske variable undersøges og sammenlignes gennem beta-koefficienter. Denne analyse viste, at særligt de to askriptive variable (køn og social baggrund) viste stærk sammenhæng med de fleste af de lingvistiske variable. Blandt de literacy-orienterede variable viste særligt skrivelyst stærk sammenhæng med de lingvistiske variable. Læsevaner viste en lidt mere varierende sammenhæng på tværs af variable, mens skrivevaner kun havde meget begrænset betydning på tværs af variable. Analysen afdækkede desuden relativt stor variation i beta-koefficienterne mellem den berettende og den beskrivende teksttype. I den anden del af den synkrone analysedel undersøges den mere specifikke betydning af køn og social baggrund. Her viste social baggrund stor betydning for tekstlængde, og denne betydning blev kun i lille grad forklaret af de literacy-orienterede variable. Betydningen af social baggrund var lidt mindre ved ordlængde, men også her var betydningen relativt stabil på tværs af de regressionsmodeller, hvor der blev kontrolleret for de literacy-orienterede variable. Der var imidlertid ingen statistisk signifikant betydning ved sætningslængde. Køn havde markant og stabil betydning for tekstlængde, i den forstand at pigerne skriver længere tekster end drengene. Der var ingen markant kønseffekt ved ordlængde, mens der ved sætningslængde sås en omvendt kønseffekt, idet drengene skrev længst sætninger. Analysen afsluttes med en diskussion og fortolkning af disse resultater i henhold til bl.a. Bernsteins kodeteori, 'Writers-Within-Community'-modellen samt relateret forskning.

I kapitel 7, som er afhandlingens sidste kapitel, diskuterer jeg implikationerne af afhandlingens resultater for henholdsvis skriveforskning og danskfagets didaktik. Kapitellet indledes med en diskussion af potentialer og begrænsninger ved at studere skriveudvikling på et lingvistisk grundlag. Diskussionen tager bl.a. udgangspunkt i en model, der viser grundlæggende dimensioner af tekstaktiviteter i skolen (teksten, eleven, skriveprocessen, den didaktiske rammesætning og den bredere sociokulturelle kontekst). Pointen er her, at skriveudvikling er relateret til alle dimensioner af modellen, og at man kan undersøge dimensionerne selvstændigt og på tværs, og i enkelte tilfælde kan dimensionerne også agere som proxy for hinanden (fx kan en elevtekst være en proxy for en elevs kognitive færdigheder). Herefter diskuterer jeg forskellen mellem en integreret model for skriveudvikling over for et metodisk perspektivskifte. Diskussionen af afhandlingens danskdidaktiske implikationer indledes med en generel beskrivelse af forholdet mellem lingvistikken og didaktikken, hvorefter jeg mere specifikt diskuterer relevansen af viden om skriveudvikling for henholdsvis skriveevaluering og skriveundervisning i danskfaget. Til sidst konkluderer jeg på afhandlingens forskningsspørgsmål, og jeg kommer med overvejelser og forslag til fremtidig forskning i skrivning og skriveudvikling.

Summary

In this PhD thesis, I examine writing development in a corpus of student texts from the 5th-9th grade written in the subject of L1 Danish. The corpus consists of two text types, a narrative and a descriptive. The analyses in the thesis are comprised of two main parts. In the first part, I measure the distribution of five linguistic features in the corpus across grade levels: text length (total number of words), word length, lexical variation (OVIX), sentence length and mean dependency distance. These analyses are quantitative and based on digital tools from the field of Natural Language Processing. The quantitative distributional analyses are combined with qualitative, lexicogrammatical analyses of three groups of texts focusing on different linguistic features in each group. The first part of the analyses thus assumes a diachronic approach to writing development, in the sense that I examine linguistic features as they develop across grade levels (cross-sectional data). In the second part of the analyses, I adopt a synchronous approach to writing development. This means that I study how a group of extra linguistic variables (e.g. gender, social background, writing attitude, reading habits and writing habits) affect linguistic features in the texts. I employ multiple regression modelling to estimate this effect, and I integrate beta-coefficients to compare the effect between variables.

The thesis consists of seven chapters. In chapter 1, I present the purpose, research questions and background of the study. The chapter also contains an overview of relevant research. The research overview consists of two main parts. In the first part, I focus on Scandinavian writing research. I describe two dominant traditions, the linguistic and the ethnographic/didactic tradition, and I position the thesis in relation to these dominant traditions. In the second part of the research overview, I focus on international writing research with a text-oriented and quantitative focus. This section is structured around the diachronic/synchronous distinction that I also use to structure the analyses of the present thesis. I round off the chapter with reflections on how the thesis contributes to the existing research field.

In chapter 2, I describe the theoretical foundations and key concepts of the thesis. I start with a description of different theoretical models of writing development, and I discuss how these models relate to the present study. Then, I zoom in on three theories relevant for the different parts of the analyses of the thesis. The first one is a classical theory

on cognitive maturation as described by Hunt (1965, 1970). The second is Bernstein's (1971) theory on restricted and elaborated linguistic codes, and the third one is Graham's (2018) 'Writers-Within-Community' model. Afterwards, I describe how student texts can be conceived as research data with particular emphasis on the pragmatic circumstances that are specific to student writing in a school context. Finally, I describe how the notions of 'text and 'genre' is understood in the context of the present thesis.

Chapter 3 contains descriptions of the methods used in the analyses as well as a description of the overall research design, including how the methods are related in the analysis as well as in the wider interpretation of the results. In the chapter, I describe in detail the two analytical approaches of the thesis (the diachronic and the synchronous approach) as well as the textual data and the survey data that form their empirical basis. I argue for the validity and reliability of the two approaches based on Lincoln & Guba's (1985) notion of 'trustworthiness', Onwuegbuzie & Johnson's (2006) notion of 'conversion legitimation' and the notion of internal validity. I round off the chapter by discussing methodological and epistemological ideas related to structuralist descriptivism, linguistic operationalization and computational hermeneutics as well as ethical concerns related to anonymization and the notion of consequential validity.

In chapter 4, I describe the quantitative variables of the thesis. I focus on both their theoretical foundations as well as their more specific operationalization within the context of the present study. The variables fall in two groups: the linguistic and the extra-linguistic variables. The linguistic variables are based on NLP-analyses of the text corpus, and consist of text length, word length, lexical variation (OVIX), sentence length and mean dependency distance. The extra-linguistic variables have two subgroups. The first is a group of ascriptive variables related to the gender and social background of the participating students. The second group, which I have labeled 'literacy-oriented', consists of writing attitude, reading habits and writing habits. In the last part of the chapter, I discuss the construct validity of the extra-linguistic variables.

Chapter 5 contains the first part of the analysis, the diachronic part. In this chapter, I study the development of the five linguistic variables across grades 5-9, and I also study lexicogrammatical features of three groups of text focusing on semantics, lexis and syntax respectively. The analysis of text length showed a somewhat linear development from grades 5-8 with grade 9 as an interesting outlier representing a average fall in text length.

The qualitative analyses showed that text length is closely connected to semantic elaboration (i.e. increasing size of semantic fields), but not necessarily to semantic variation (i.e. an increased number of semantic fields). The analysis of word length and OVIX showed no linear development across grades. Median values for OVIX were pretty stable across grades, whereas median values for word length showed substantial differences between primary grade levels (5-6) and lower secondary grade levels (7-9). Qualitative analyses of cohesive devices and nominal phrases showed no close connection with variable values, but instead showed great variation between text types. Sentence length and dependency distance also showed substantial differences between grade levels 5-6 and 7-9. However, there were no important developments from grade levels 7-9. Qualitative analyses of syntactic structures in student texts revealed significant variation in the syntactic structures behind the variable values indicating very large-grained syntactic measures. One of the main analytic takeaways from the diachronic analysis is that age maturation is only a proxy for development and not a structuring principle. Another takeaway is that writing development is not necessarily a linear process, but also a process characterized by iterations, development plateaus and hypothesis testing through excessive use of specific linguistic features.

In chapter 6, I present the second part of the analysis, the synchronous part, in which I examine the relation between the linguistic and the extra-linguistic variables. More specifically, I examine the effect that the extra-linguistic variables have on the linguistic variables through multiple linear regression models. In the first part of the synchronous analyses, I compare the effect of all the extra-linguistic variables on the five linguistic variables through beta-coefficients. The results show that the two ascriptive variables (gender and social background) in particular have a significant effect on most of the linguistic variables. In the group of literacy-oriented variables, writing attitude was the one with the biggest effect across all the linguistic variables. Reading habits showed more varying effects, whereas writing habits showed no statistically significant effects. The analysis also revealed a great deal of variation in the beta-coefficients between the narrative and the descriptive text type. In the second part of the synchronous analysis, I examine in more detail the effects of social background and gender. Social background had a substantial effect on text length. The effect was somewhat stable across regression models in which the three literacy-oriented variables were held constant. Social background also

had an effect on word length, although seemingly a bit smaller, that was somewhat stable across models. There were no statistically significant effects on sentence length. Gender had a substantial and stable effect across models on text length in the sense that girls wrote longer texts than boys. There was no gender effect on word length, and in regard to sentence length there was a statistically significant reverse gender effect in the sense that boys wrote longer sentences than girls. I round off the chapter by discussing and interpreting the results in the context of Bernstein's theory of linguistic codes, Graham's 'Writers-Within-Community' model and related research.

In the last chapter of the thesis, chapter 7, I discuss the implications of the present study for writing research as well as for the teaching of writing in the subject of L1 Danish. I start by discussing the potentials and limitations of studying writing from a linguistic perspective. The discussion centers on a model showing fundamental dimensions of text activities in school (the text, the student, the writing process, the instructional framing and the wider sociocultural framing). The main point of the model is that writing development relates to all dimensions of the model, which in turn can be studied independently, across dimensions and in some cases dimensions can act as proxies for other dimensions (e.g. when student texts are taken as proxies for writing competence). I then discuss the difference between an integrated model of writing development and a methodological change of perspective. I then turn to the discussion of how the results of the thesis impact on the teaching and evaluation of writing in the subject of L1 Danish. This discussion starts with a general description of the relation between linguistics and school subjects. I then discuss in more detail the different ways that knowledge of writing development can be used to improve writing evaluation as well as writing instruction. I round off the chapter, by concluding on the research questions, and I provide directions for future research in writing and writing development.

Forord

Tak til den deltagende skole og de dansklærere, som lod mig komme og forstyrre på et kritisk tidspunkt, hvor der i skyggen af den første nedlukning under pandemien blev holdt skole under vanskelige vilkår.

Tak til min hovedvejleder lektor Ulf Dalvad Berthelsen, uden hvem dette projekt aldrig var blevet til noget. Tak for vores utallige snakke og faglige diskussioner og for at overbevise mig om, at statistik og programmering også er en sag for humanister.

Tak til min medvejleder lektor Felix Weiss for yderst kompetent og tålmodig indføring i kvantitativ forskning. Tak for skarpe kommentarer og bistand med de statistiske analyser.

Tak til lektor Anne Palmér og hendes kollegaer ved Institutionen för nordiska språk, Uppsala Universitet, for at tage godt imod mig, da jeg var på besøg. Det var både hyggeligt og fagligt udbytterigt.

Tak til alle kollegaer i Afdeling for Fagdidaktik og i danskgruppen.

Tak til alle mine kollegaer i det daglige på campus Aarhus for at gøre campus til et rart sted at være. Særligt tak til alle de ph.d.-studerende, som jeg har lidt skæbnefællesskab med undervejs. Tak til Anders Helle Malling for godt kollegaskab og for altid at have tid til at hjælpe med at finde løsninger i Python, når jeg selv synes, det var håbløst!

Tak til alle kontormakkere for godt og hyggeligt kontormakkerskab de sidste tre år. En særlig tak til mine mest faste kontormakkere, Merete Munkholm, Maiken Norup og Dorte Vang Eggensen. Det har været en fornøjelse!

Tak til familie og venner for opbakning og for altid at spørge nysgerrigt ind til projektet. Og sidst men ikke mindst tak til Laura for altid at støtte og bakke op. Tak for dig!

Morten Tannert, Aarhus, august 2022.

Kapitel 1: Indledning og forskningsoversigt

I kapitel 1 præsenterer jeg projektets formål og forskningsspørgsmål, og jeg begrundes projektets metodiske forankring. Jeg fokuserer særligt på projektets forankring i nyere tendenser inden for pædagogisk og humanistisk forskning, der fokuserer på anvendelsen af digitale metoder. I anden halvdel af kapitlet præsenterer jeg en forskningsoversigt over relevant og beslægtet forskning. Første del af forskningsoversigten udgøres af en bred præsentation af tendenser og udviklingstræk inden for skandinavisk skriveforskning, mens den sidste del af forskningsoversigten fokuserer lidt mere snævert på international, tekstorienteret skriveforskning, der har en kvantitativ metodisk forankring.

1.1 Indledning

Tilegnelsen af gode skrivefærdigheder er en helt central del af elevernes faglige udvikling i danskfaget såvel som i skolen generelt. Dette skyldes bl.a., at udviklingen af skrivefærdigheder dels har vist sig at være en afgørende forudsætning for tilegnelse af fagligt indhold i alle skolens fag (Klein & Boscolo, 2016; Miller et al., 2018), og dels er skriftsproget fortsat det dominerende modus i elevprodukter i danskfaget (Bremholm et al., 2018). Der er derfor brug for viden om, hvad der karakteriserer elevernes udvikling af skrivefærdigheder inden for en skolekontekst, idet denne type viden kan understøtte såvel skriveevaluering som skriveundervisning i danskfaget (Flowerdew, 2022). Eksempelvis er arbejdet med evaluering og bedømmelse af elevtekster udfordret af, at lærere i meget varierende grad anvender viden om sprog og skriftsproglig udvikling i evalueringen af elevtekster, hvilket har vist sig at være en medvirkende årsag til en lav grad af reliabilitet i lærernes bedømmelser af elevprodukter (Baker, 2014; Skar & Jølle, 2017; Wyatt-Smith et al., 2010). Desuden har forskning vist, at eleverne har særligt gavn af feedback, som bl.a. er tekstorienteret og funderet i analytiske nedslag i de tekster, de har skrevet (Bueie, 2016; Matre & Solheim, 2015; Parr & Timperley, 2010). Der er derfor brug for studier, som netop beskriver, hvad der kendetegner forskellige lingvistiske dimensioner af elevtekster skrevet af elever med forskellige baggrunde og på forskellige klassetrin i grundskolen.

Dette projekt bidrager med viden på området ved at undersøge, hvordan elevens skriftsprog udvikler sig henover grundskolens 5.-9. klassetrin. Projektet fokuserer på

elevtekster produceret i danskfaget, og analyserne gennemføres som kvantitative korpuslingvistiske analyser af et korpus af elevtekster fra danskundervisningen på 5.-9. klassetrin. Korpuslingvistiske metoder har i denne forbindelse et stort og uudnyttet potentiale, idet de kan identificere generelle syntaktiske og leksikale mønstre i digitaliserede elevtekster, som man ellers ikke havde fået øje på gennem manuelle, kvalitative analyser. Tidligere har udviklingen af sådanne tekstkorpuser været kraftigt besværet af, at de papirbaserede elevtekster først skulle digitaliseres og derefter formateres, før det var muligt for en computer at behandle dem (McEnery & Hardie, 2013). Digitaliseringen af skoleområdet har imidlertid medført, at det i dag er reglen snarere end undtagelsen, at elevtekster er tilgængelige i digitalt format, hvilket nu gør det muligt at udvikle tekstkorpuser bestående af elevtekster fra grundskolen.

Der er imidlertid endnu ikke et stort forskningsfelt med fokus på korpuslingvistiske studier af elevtekster fra skolen, hvilket betyder, at der fortsat ikke er udviklet faste standarder og praksisser for, hvordan sådanne analyser gennemføres og fortolkes inden for en didaktisk ramme (Crossley, 2020; Moxley et al., 2017). Formålet med dette projekt er derfor dobbelt. Der er dels et metodisk formål og dels et empirisk formål, hvoraf det empiriske formål er det primære. Det metodiske formål er eksplorativt og knyttet til en undersøgelse af potentialer og begrænsninger ved kvantitative, automatiserede analyser af elevtekster ift. at kunne beskrive skoleelevers skriftsproglige udvikling, og jeg udpeger desuden i denne forbindelse en række fokus- og udviklingspunkter for denne tilgang til skriveforskning. Det empiriske formål er knyttet til en undersøgelse af, hvad der kendetegner elevers skriftsproglige udvikling i danskfaget fra 5.-9. klassetrin. I analyserne fokuseres således på at afdække latente tekstuelle mønstre i elevteksterne med henblik at undersøge, dels i hvilken grad og hvordan de sproglige træk udvikler sig henover klassetrin, dels i hvilken grad sproglige træk i elevernes tekster hænger sammen med en række baggrundsfaktorer knyttet til den enkelte elevskriver, herunder elevernes læse- og skrivevaner i fritiden, køn og sociale baggrund.

Analyserne baseres på to typer af data; tekstdata og surveydata. De tekstuelle data består af elevtekster, som er samlet ind på fem årgange på én skole. Der er som udgangspunkt samlet to tekster ind for hver elev, en beskrivende tekst (n=236) og en berettende tekst (n=228). Teksterne er skrevet på baggrund af et forskerinitieret skriveprompt. De

lingvistiske analyser foretages på de indsamlede tekster, og analyserne tager udgangspunkt i en række lingvistiske variable, som er baseret på automatiserede analyseværktøjer fra det felt, som kaldes Natural Language Processing (Eisenstein, 2019). Surveydataene er indsamlet i forbindelse med indsamlingen af elevteksterne og består af elevbesvarelser på en række items. Disse items har dannet grundlag for en gruppe af variable i studiet, som jeg kalder 'ekstra-lingvistiske variable'. Disse variable dækker bl.a. over elevernes skrivelyst, læse- og skrivevaner, køn og sociale baggrund.

Projektets analyser falder i to dele, en henholdsvis diakron og synkron analysedel. I den diakrone analysedel anlægges en diakron og eksplorativ tilgang til skriftsproglig udvikling. Dette indebærer, at elevteksterne grupperes på klasstrinsniveau med henblik på at undersøge forskelle og udviklinger i de lingvistiske mål relativt til klasstrin. Disse analyser følges op af kvalitative, sproglige analyser af forskellige træk i teksterne med fokus på, hvad der kendetegner elevtekster med forskellige værdier i de kvantitative, lingvistiske variable. I denne analysedel er der både fokus på projektets metodiske og empiriske dimension. I den synkrone analysedel grupperes teksterne ikke efter klasstrin, men derimod anses de som én stor gruppe tekster med annoteret metadata, herunder klasstrin og de føromtalte ekstra-lingvistiske variable. Formålet med den anden del af analyserne er at undersøge, hvilken betydning de ekstra-lingvistiske variable har for de lingvistiske variable. Afhandlingen tager således udgangspunkt i tre overordnede forskningsspørgsmål, som hver udgøres af to eller tre underspørgsmål.

1. Hvilke sproglige karakteristika har et korpus af elevtekster skrevet i danskundervisningen på 5.-9. klasstrin, og hvad siger disse sproglige karakteristika om elevernes skriftsproglige udvikling?
 - a. Hvilke forskelle ses der i fordelingen af de kvantitative, lingvistiske variable relativt til klasstrin og teksttype?
 - b. Hvilke leksikogrammatisk kendetegn har elevtekster, som repræsenterer henholdsvis forskellige variabelværdier og klasstrin?
 - c. På hvilken måde kan de sproglige karakteristika, som de kvantitative og kvalitative analyser af tekstkorpuset identificerer, bidrage til viden om skoleelevers skriftsproglige udvikling i danskfaget?

2. Hvilke metodiske potentialer og begrænsninger er der ved at beskrive skriftsprog-
lig udvikling ved hjælp af automatiserede, kvantitative analyseværktøjer?
 - a. Hvilke metodiske potentialer og udfordringer er der specifikt forbundet
med anvendelsen af de automatiserede analyseværktøjer i denne afhand-
ling?
 - b. Hvilke metodiske fokus- og udviklingspunkter giver anvendelsen af de au-
tomatiserede analyseværktøjer i dette studie anledning til?
3. Hvordan hænger elevernes skriftsproglige udvikling sammen med askriptive (køn
og social baggrund) og literacy-orienterede variable (skrivelyst, læsevaner og
skrivevaner) knyttet til den enkelte elevskriver?
 - a. Hvilke af de ekstra-lingvistiske variable (herunder de askriptive og literacy-
orienterede) viser den stærkeste sammenhæng med de fem lingvistiske va-
riable?
 - b. Hvilken betydning har social baggrund for de lingvistiske variable, når der
kontrolleres for skrivelyst, læse- og skrivevaner og sproglig baggrund?
 - c. Hvilken betydning har køn for de sproglige variable, når der kontrolleres for
skrivelyst samt læse- og skrivevaner?

1.2 Baggrund og relevans

Skriftsproget er på mange måder et allestedsnærværende fænomen. Der findes ikke mange aktiviteter i et moderne samfund, hvor skriftsproget ikke spiller en helt afgørende rolle (Brandt, 2014; Fischer, 2001; Karlsson, 2006), og digitaliseringen af samfundslivet har ikke ført til mindre tekst, men tværtimod til et ekspansivt udviklende og komplekst literacy-landskab, hvor eleverne skal forholde sig til mere tekst og flere forskellige genrer end tidligere (Biber & Egbert, 2018). Gode skriftsproglige færdigheder er derfor afgørende for aktiv deltagelse i et frit og demokratisk samfund (Berge, 2019; Krogh, 2020), og vores evne til at gøre brug af det skrevne sprog er derfor et vigtigt benchmarkingparameter i flere komparative internationale undersøgelser, hvor literacy-niveauet i et givent land ofte er en vigtig variabel ift. at måle trivsel eller faglighed i bred forstand (fx OECD, 2020, 2021). I et lidt større historisk perspektiv er det ligeledes vanskeligt at overvurdere skriftsprogets betydning for menneskets kulturelle og kognitive udvikling, eftersom skriftsproget har været et afgørende eksistensvilkår for udviklingen af en såkaldt teoretisk

kultur, der bl.a. har muliggjort filosofi og videnskab, som vi kender det i dag (Bellah, 2011). Det er derfor heller ikke muligt at forestille sig en moderne skole uden skriftsprog, idet skriftsproget er en vigtig forudsætning for, at mennesket kan overlevere viden mellem generationer, hvilket er det, der gør det muligt for elever at tilegne sig viden i skolen på tværs af den tid (både synkront og diakront) og det sted, denne viden oprindeligt er frembragt under (Donald, 1991). Det er derfor også muligt at se skriftsproget som en vigtig forudsætning for det, Tomasello (2008, s. 341-343) kalder 'shared intentionality', som refererer til menneskets evne til at indgå i kollaborative aktiviteter med andre mennesker, som man deler intentioner og mål med. Skriftsproget muliggør, at disse kommunikative, kollaborative aktiviteter kan ske asynkront, og derfor er gode skriftsproglige færdigheder helt afgørende for elevernes faglige udvikling i skolen.

Sammenlignet med læsning er skrivning imidlertid et relativt understuderet fænomen – særligt hos børn i alderen 11-16 år (Parsons et al., 2020; Parsons & Gallagher, 2016). Dette hænger bl.a. sammen med, at skrivning langt fra er et trivielt fænomen at undersøge. Det er på én gang en kompleks handling forårsaget af et individs kognitive og mentale ressourcer, et sprogligt fænomen med klare kommunikative hensigter og succeskriterier, og samtidig er det en gennemgribende social teknologi, der er konstituerende for vores sociale omgang med hinanden og for den viden, som vi skaber. Dertil kommer, at det i vid udstrækning er i skolen, man lærer at skrive, og derfor er de første mange ord, eleverne skriver, skrevet inden for en skolsk ramme (Hetmar, 2000) – i modsætning til talesproget, som eleverne møder i mange forskellige sammenhænge, hvoraf de fleste er naturlige og funktionelle. Skrivning manifesterer sig derfor på mange forskellige ontologiske niveauer; både som konkrete sproglige tegn, tekster, kommunikative og sociale handlinger, identitetsdannelser, sociale praksisser og diskurser etc. - alt sammen komplekst forbundet til individets kognitive ressourcer og til større sociokulturelle forhold (Baron, 2009; Brandt, 2014; Flower, 1979; Halliday, 2013; Stubbs, 2007). De fleste moderne teoretiske forståelser af skrivning indeholder derfor en mangfoldighed af forskellige kognitive, sociale og kulturelle dimensioner. Nogle skrivemodeller tager udgangspunkt i et mere holistisk literacy-perspektiv og anskuer skrivefærdigheder som en del af en større multimodal økologi, hvor skrivning og læsning er tæt forbundet med hinanden (fx Barton, 2007; Cope & Kalantzis, 2000, 2009; Street, 1984, 1995; The New London Group, 1996), mens andre skrivemodeller er mere specialiserede og har et mere rent fokus

på fx lingvistiske (fx Christie & Derewianka, 2010; Myhill, 2008), kognitive (fx Deane, 2018; Flower, 1979; Hacker, 2018), sociale (fx Krogh & Jakobsen, 2019) eller semiotiske (fx Kress, 1994) dimensioner.

1.2.1 Hvorfor forske i skriftsproglig udvikling på et kvantitativt grundlag?

Et væsentligt spørgsmål i forlængelse heraf er naturligvis, hvorfor det overhovedet er nødvendigt med kvantitative studier af skriveudvikling¹. Den kvalitative skriveforskning har historisk set stået stærkt i nordisk fagdidaktisk forskning (Berge, 2019; Holmberg et al., 2019; Krogh, 2012; Ongstad, 2002) med rødder i bl.a. retorik, socialesemiotik, poststrukturalisme og etnografi. Med undtagelse af en række svenske korpuslingvistiske studier og enkelte norske (se gennemgang af studierne i afsnit 1.3.2) har den kvantitative skriveforskning spillet en meget lille rolle. En mulig forklaring på dette kan være, som jeg også kort var omkring i indledningen, at det tidligere har været besværligt at oprette og udvikle store tekstkorpusser af elevtekster, fordi elevtekster i vid udstrækning kun har været tilgængelige i fysiske papirformater og derfor først skulle digitaliseres gennem manuel indtastning og formatering, før det var muligt at analysere dem computationelt. I et af de største nordiske tekstkorpusser bestående af elevtekster, ”The Uppsala Corpus of Student Writings”, som består af over 2500 elevtekster helt tilbage fra 1996, beretter Megyesi et al. (2016, s. 3193) fx om de store logistiske og tidsmæssige udfordringer, der er forbundet med at digitalisere både håndskrevne og udprintede tekster. Den massive digitalisering af skoleområder har imidlertid gjort, at langt de fleste elevtekster ligger tilgængelige i digitalt format på diverse digitale læringsplatforme, hvilket nu gør det betragteligt lettere at indsamle elevteksterne og oprette dem i et samlet tekstkorpus. Men én ting er naturligvis, at man nu *kan* lave digitale og kvantitative undersøgelser af tekstkorpusser med elevtekster. Noget andet er, hvorfor det er en god idé at gøre det. Svaret på dette spørgsmål er todelt. Dels har svaret en teknologisk dimension, og dels har det en epistemologisk og metodologisk dimension.

Det seneste årti har den kvantitative og automatiserede tekstlæsning vundet indpas i både human- og socialvidenskaberne. Denne tendens har rødder i 1940’ernes tidlige

¹ Jeg anvender i denne afhandling begreberne ’skriftsproglig udvikling’ og ’skriveudvikling’ synonymt.

computationelle lingvistik (Hirst, 2013) og 1960'ernes korpuslingvistik (McEnery & Hardie, 2013). De første computerbaserede studier af skriftsprog fokuserede på at generere simple ordlister med højfrekvente ord (eng: concordances), men var samtidig voldsomt begrænsede af både manglende computerkraft og opbevaringskapacitet (Sula & Hill, 2019). Samtidig krævede det en højt specialiseret teknisk viden om computere og programmering at gennemføre sådanne studier, så det nåede aldrig den store udbredelse uden for de specialiserede forskningsmiljøer (Kelleher & Tierney, 2018). I takt med den drastiske udvikling i computerkraft og med fremkomsten af tilgængelige og mere brugervenlige programmeringssprog såsom Python er computerbaserede studier af skriftsprog for alvor rykket ind i socialvidenskaberne og humaniora under betegnelser som 'digital humanities' (Berry, 2012), 'learning analytics' (Siemens, 2013) og 'writing analytics' (Moxley et al., 2017). Dette skifte kan ifølge Finnemann (2014) beskrives som en bevægelse fra det tidlige *humanities computing* med fokus på at udnytte computerens regnekraft til analyse af digitaliseret materiale til det nuværende *digital humanities*, som er en samlebetegnelse for studiet af digitalt materiale ved hjælp af software-støttede metoder.

Et af de teoretiske grundbegreber i denne bevægelse er Morettis (2000, 2013a) begreb 'distant reading', som henviser til det at anvende digitale metoder til at analysere store mængder af tekst. Begrebet er oprindeligt et litteraturvidenskabeligt begreb og er tænkt som et vigtigt led i etableringen af en såkaldt 'computational criticism' (Moretti, 2013b), som markerer en empirisk og kvantitativ vending inden for litteraturvidenskaben med fokus på at identificere litterære strukturer på tværs af værker, forfatterskaber og tidsperioder. Pointen med 'distant reading'-begrebet er, at kvantitative og automatiserede analyser af store mængder af tekstuelle data kan identificere mønstre (grammatiske, stilistiske, litterære etc.) på tværs af tekster, som man ikke ellers ville have opdaget gennem en klassisk nærlæsning. I den pædagogiske eller uddannelsesrelaterede skriveforskning er potentialet primært knyttet til dels en deskriptiv dimension med fokus på at beskrive og kortlægge sproglige mønstre og strukturer i forskellige typer af elevskrivning, og dels en interesse i at bruge automatiserede analysemetoder til at udvikle fx skrevetests eller automatiserede feedbackværktøjer (Tannert, 2021). Det, som den kvantitative skriveforskning deler med det oprindelige litteraturvidenskabelige begreb, er således interessen i at bruge de sproglige strukturer i teksterne som metodologiske trædesten til at sige noget mere

grundlæggende om de fænomener, der undersøges, samt interessen i at skalere tekstmaterialet, så det bliver muligt at skabe viden om et fænomen, der i store grad kan generaliseres og række ud over den enkelte tekst eller elevskriver.

1.2.2 Hvorfor koble det lingvistiske med et psykologisk og sociologisk perspektiv?

Viden om skriftsproglig udvikling er vigtig, fordi det dels kan belyse en meget vigtig faglig dimension af elevers skolegang, dels kan danne afsæt for udviklingsarbejde knyttet til både skriveundervisning og –evaluering. Den viden om skriftsproglig udvikling, som der produceres i denne afhandling, er både lingvistisk, dvs. den beskriver elevteksterne på et formelt, lingvistisk niveau, og så er den psykologisk og sociologisk, dvs. den kobler de lingvistiske strukturer til psykologiske og sociologiske baggrundsfaktorer ved den enkelte elevskriver. Der er tre årsager til, at det er væsentligt at koble det lingvistiske beskrivelsesniveau med et psykologisk og sociologisk beskrivelsesniveau. Den første årsag er knyttet til en velbeskrevet udfordring i skriveforskningen angående etableringen af et overordnet mål for tekstkvalitet. I mange studier korreleres forskellige sproglige variable med et overordnet endimensionelt mål for tekstkvalitet, fx via rubrics eller karakterbedømmelser. Men som pointeret af Troia et al. (2019) er der flere udfordringer knyttet til denne tilgang. Dels har det ofte vist sig at være krævende at nå et acceptabelt niveau af interterreliabilitet i kvalitative scoringer af elevtekster, og dels har mange rubrics været baseret på scoringer af sproglige træk, som er interdependente, dvs. som ikke er uafhængige af hinanden. Derfor er der i dette projekt ikke fokus på sammenhængen mellem de sproglige variable og mere eller mindre valide scoringer af tekstkvalitet, men derimod på sammenhængen mellem tekst og elevskriver. Dermed udvides perspektivet fra det klassiske fokus på, hvad der kendetegner elevtekster med forskellige kvalitetsbedømmelser, til et fokus på, hvad der karakteriserer tekster skrevet af elever med forskellige forudsætninger og baggrunde. Projektet lægger sig dermed i forlængelse af en klassisk korpuslingvistik tilgang, hvor ”1) *assessing the extent to which a pattern is found, and 2) analyzing the contextual factors that influence variability*” (Biber et al., 1998, s. 3) udgør de to primære forskningsformål.

Den anden årsag til at kombinere det lingvistiske med det sociologiske og psykologiske handler om udviklingsperspektivet. Mange studier, som har anvendt kvantitative

mål af sproglige træk til at studere skriftsproglig udvikling, har taget enten alder eller klasstrin som proxy for kognitiv modning hos eleven. Det betyder, at en øgning af frekvens i et lingvistisk mål henover alder eller klasstrin er blevet betragtet som et tegn på skriftsproglig udvikling. Dette er på mange måder en legitim tilgang, dels fordi alder har vist sig at være en vigtig faktor i udviklingen af skrivefærdigheder (Attali & Powers, 2008; Crossley et al., 2011; Kerz et al., 2020), dels fordi forskelle mellem klasstrin siger noget om den variation og kompleksitet, som er særlig for skrivning i en skolekontekst. Derfor fastholdes denne tilgang også i første del af dette projekts analyser. Men tilgangen har også klare mangler. Aldersmodning kan være én forklaring på, hvorfor skriftsprog udvikler sig, men det vil selvfølgelig være en voldsomt reduceret og singulær kausal forståelse at antage, at det er den eneste årsag. Som det også er set med andre færdigheder i skolen, må der forventes at være sammenhænge mellem skrivekompetence og fx social baggrund, køn eller læse- og skriveerfaring. I anden del af dette projekts analyser betragtes klasstrin derfor blot som en variabel på niveau med de andre sociologiske og psykologiske variable, og det bliver derigennem muligt at sammenligne betydningen af klasstrin med de øvrige ekstra-lingvistiske variable.

Den tredje og sidste årsag angår studiets didaktiske relevans. Klasstrinsrelaterede forskelle i elevernes tekster er som nævnt ovenfor naturligvis interessante didaktisk, fordi de siger noget om den sproglige og tekstuelle variation, som findes både inden for og imellem klasstrin. Men mindst lige så interessant er det, hvordan sproglige træk i elevteksterne hænger sammen med den enkelte elevs baggrund og øvrige forudsætninger. Skoleklasser udgøres langt fra af en homogen elevgruppe, men derimod af elever med meget forskellige sociale baggrunde og med meget forskellige erfaringer med læsning og skrivning hjemmefra. Derfor er det didaktisk set også interessant med en mere fintmasket analysestrategi, der netop undersøger sammenhængen mellem den enkelte elevs baggrund og elevens tekst.

1.3 Forskningsoversigt

Skriveforskningen har historisk set haft en central plads i den nordiske fagdidaktiske forskning - særligt centreret omkring modersmålsdidaktikken (Berge, 2019; Holmberg et al., 2019; Smidt, 2018). Dette hænger naturligvis sammen med, at modersmålsfagene historisk set har haft et særligt ansvar for at varetage elevernes skriveudvikling (Ongstad,

2002; Raahauge, 2020), hvilket betyder, at modersmålsdidaktikken har udviklet en særlig forskningsmæssig tradition for at undersøge og beskrive forskellige dimensioner af elevernes skrivning i bredeste forstand. Denne tradition er imidlertid en heterogen størrelse og er kendetegnet ved i et historisk perspektiv at være præget af forskellige skiftevis dominerede teoretiske og metodiske tilgange. Et samlende særkende for den skandinaviske skriveforskning er dog, at den er udført inden for rammerne af en fagdidaktisk forskningstradition, hvor elevernes skrivning forstås i relation til faglige traditioner, forståelsesmåder og curriculære strukturer. Det fagdidaktiske perspektiv giver derfor et særegent blik på skrivning i fagene, fordi det netop har blik for vekselvirkningen mellem faget og elevernes skrivepraksisser. Indsnævringen til det fagdidaktiske er derfor vigtig, eftersom forskellige fag i skolen er underlagt forskellige curriculære strukturer og didaktiske traditioner, som skaber forskellige skrivesituationer i skolen, og som dermed sætter rammerne for de tekster, eleverne skriver (Carter, 2007). Der er således stor forskel på, om man skriver en fysikrapport eller en reportage i dansk. Skrivekompetence er således kun generisk på et helt grundlæggende niveau, men må derudover forstås som en aktivitet, der er særegen for bestemte kontekster og i en skolesammenhæng i særdeleshed er særegen for skolens fag (Gee, 2007).

Denne forskningsoversigt er således delt op i to dele. Først præsenterer jeg en oversigt over tendenser i nyere skandinavisk skriveforskning med henblik på at placere nærværende studie inden for en modersmålsdidaktisk forskningstradition. Der vil her være fokus på at beskrive forskellige generelle tendenser inden for den skandinaviske skriveforskning, og der vil derfor ikke kun være fokus på kvantitative, lingvistiske studier, men derimod på bredere tendenser på tværs af discipliner og metoder. Derefter snævrer jeg perspektivet ind og præsenterer i et større internationalt perspektiv en oversigt over relevant lingvistisk skriveforskning med et tekstorienteret og kvantitativt fokus.

1.3.1 Rammesætning af forskningsoversigten

De to dele af forskningsoversigten er baseret på systematiske litteratursøgninger i en række databaser. Begge dele af forskningsoversigten er udformet som et induktivt 'mapping review', hvor tendenser i forskningslitteraturen skrives frem på baggrund litteratursøgninger i relevante databaser, gennemgang af referencer fra eksisterende reviews og forskningsoversigter samt løbende litteratur- og referencesøgninger inden for de emner,

som dukker op undervejs (Grant & Booth, 2009, s. 94). Ved den første del af forskningsoversigten med fokus på skandinavisk skriveforskning er denne fremgangsmåde valgt, eftersom den skandinaviske skriveforskning som emneområde er et heterogent og tværdisciplinært område, hvis fællesnævner er geografisk og ikke metodisk, og som derfor ikke vil kunne afdækkes fyldestgørende gennem en systematiseret litteratursøgning alene. Dertil vil den metodiske og emnemæssige variation i forskningsfeltet være for stor. Desuden er formålet med de to dele af forskningsoversigten ikke at lave en udtømmende oversigt over forskningslitteraturen med henblik på hverken at aggregere forskningsresultater eller etablere en 'best practice' på området. Formålet er snarere at placere denne afhandling i relation til dominerende tendenser inden for den skandinaviske og internationale skriveforskningstradition, som afhandlingen skriver sig ind i. Litteratursøgningerne til den skandinaviske del er lavet i følgende databaser: Forskningsdatabasen (DK), CRISTIN (NO) og Swepub (SE). Jeg har med forskellige modifikationer anvendt disse grundlæggende søgestrengene i databaserne.

Dansk: (skriv* OR skriftsprog* OR literac*) AND (dansk* OR modersmål* OR L1)

Svensk: (skriv* OR skriftsprog* OR literac*) AND (svensk* OR modersmål* OR L1)

Norsk: (skriv* OR skriftsprog* OR literac*) AND (norsk* OR morsmål* OR L1)

Søgestrengene er løbende blevet ændret og udbygget gennem litteratursøgningerne i takt med, at bestemte emner eller nationale udtryk begyndte at dukke op. Foruden ovenstående søgninger har jeg konsulteret SkrivBib², som er en bibliografi over nordisk og særligt norsk skriveforskning drevet af Skrivesenteret ved Norges teknisk-naturvitenskapelige universitet samt gennemgået referencer i forskningsoversigter hos Berge (2019), Smidt (2018) og Palmér (2018). Ved screeningen af studierne har inklusionskriteriet været, at studierne skulle undersøge skrivning i bred forstand inden for en modersmålsdidaktisk ramme.

Den anden del af forskningsoversigten, som fokuserer på international, kvantitativ skriveforskning, er baseret på søgninger i databaserne ERIC og Education Research Complete. Jeg har anvendt forskellige variationer af følgende grundlæggende søgestreng, som

² Bibliografien kan findes her: <https://www.zotero.org/groups/4625246/skrivbib/>

igennem iterative processer er blevet udbygget og modereret for at udforske forskellige områder og emner, som er dukket op undervejs i søgningerne.

Engelsk: (writ* OR literac*) AND (quant* OR corp* OR NLP OR Natural Language Processing OR computational linguistic*)

Jeg har foruden ovenstående søgninger foretaget søgninger i følgende internationale tidsskrifter: Journal of Writing Research, Assessing Writing, Linguistics and Education, Reading and Writing og Written Communication. Jeg har desuden gennemgået referencer i reviews af Crossley (2020) og Zheng & Yu (2019). Inklusionskriteriet for de internationale studier har været, at de skulle være tekstorienterede (dvs. indeholde tekstdata), være orienteret mod en uddannelseskontekst og skulle indeholde et kvantitativt element. At jeg har valgt denne induktive og iterative tilgang til forskningsoversigten frem for et systematisk review, skyldes, at det vil det være umådeligt omfangsrigt at lave et systematisk review, som både indfanger den metodiske (kvantitative, korpusbaserede studier af elevtekster) samt den variabelspecifikke dimension (de enkelte lingvistiske og ekstralingvistiske variable) af projektet. Jeg har derfor delt det op således, at den metodiske dimension belyses i forskningsoversigten i dette kapitel, mens oversigten over de studier, som ligger til grund for de lingvistiske variable, fremgår af operationaliseringerne af de enkelte variable i kapitel 4. Ved ingen af de to forskningsoversigter gælder det, at jeg gennemgår eller eksplicit refererer til alle studier, jeg har identificeret i litteratursøgningerne. Begge forskningsoversigter skal ses som narrative fremstillinger af generelle tendenser og emneområder i de identificerede studier.

1.3.2 Skandinavisk skriveforskning

Der findes adskillige forsøg i litteraturen på at kategorisere og beskrive forskellige traditioner og forskningsmæssige tilgange til at undersøge skrivning i skolen. Nogle af disse distinktioner er datadrevne og baseret på systematiske litteratursøgninger af forskningslitteraturen, mens andre på et mere deduktivt grundlag baserer beskrivelsen af forskningsfeltet på teoretiske modeller. Jeg har i denne forskningsoversigt valgt en mere induktiv og iterativ tilgang, eftersom de ontologisk-baserede beskrivelser af forskningsfeltet kan have den indbyggede problemstilling, at de i stedet for at beskrive forskningsfeltets metodiske og teoretiske variation bygger på nogle normative ontologiske antagelser om,

hvordan skrivning (eller literacy i bredere forstand) bør forstås som fænomen, fx ved at adskille mentalistiske og kognitive forståelser fra sociokulturelle forståelser – ofte med den hensigt at promovere én ontologisk forståelse af skrivning frem for en anden. Denne tendens findes fx hos Holm (2006), som meget grundlæggende skelner mellem en kognitiv, en funktionel og en sociokulturel tilgang til literacyforskningen. Holms distinktioner baserer sig på ”de forskellige teoretiske antagelser, der ligger til grund for centrale dele af forskningen” (s. 20) og præsentationen af de tre tilgange bruges som led i en argumentation for en normativ, teoretisk dagsorden – i dette tilfælde at præsentere den sociokulturelle tilgang som det mest passende udgangspunkt for literacyforskningen og –undervisning.

En lidt mindre, men dog stadig delvist ontologisk bundet model findes hos Berge (2019), som skelner mellem fire paradigmer i den nordiske skriftkulturforskning; det mentalistiske, det etnografiske og historiske, det filologisk-lingvistiske samt det skrive- og læsepædagogiske paradigme. Fordelen ved denne model er, at den udpeger nogle tendenser, som er forankrede i skriveforskningen, som den faktisk har udfoldet sig historisk i Norden, men den har stadigvæk snerten af den indbyggede ontologi-problematik, som jeg har beskrevet ovenfor. Berge baserer fx det mentalistiske paradigme på en kognitiv skrivemodell, hvorimod de andre paradigmer primært er knyttet til bestemte metodologiske traditioner eller videnskabelige discipliner. Problemet med de ontologisk-betingede distinktioner er, at de ofte kommer til at forudsætte, at studier, som undersøger bestemte ontologiske dimensioner af skrivning, per definition også abonnerer på nogle meget fastlåste og indbyggede ontologiske forestillinger om skrivning som fænomen, hvilket betyder, at studier med forskellige fokuspunkter i skriveforskningen ikke samarbejder på tværs. Som det bl.a. fremhæves af Deane (2018), er der imidlertid ingen modsætning mellem at interessere sig for kognitive dimensioner af skrivningen, samtidig med at man anerkender, at skrivningen også har sociale, retoriske og strukturelt lingvistiske dimensioner.

For at undgå disse låste ontologiske forestillinger vil jeg i dette afsnit skelne mellem to meget overordnede traditioner i den skandinaviske skriveforskning, som afspejler klassiske og veletablerede videnskabelige discipliner; henholdsvis lingvistikken og antropologien. Disse to discipliner er valgt på baggrund af en indledende gennemlæsning af forskningslitteraturen, hvor det var disse discipliner, som stod ud som de mest dominerende.

De to traditioner skal ses som videnskabelige udgangspunkter, hvorfra forskellige dele af skrivningen kan studeres på baggrund af forskellige metoder og teoretiske udgangspunkter. Min tilgang har således ingen ontologiske implikationer og skal derfor ikke læses som et forsøg på at definere, hvad skrivning er og ikke er, men skal derimod læses som overordnede kategorier, der skal bidrage til at tegne nogle tendenser frem i skriveforskningen, som hver især rummer en stor teoretisk og metodisk variation. Kategorierne er selvfølgelig heller ikke udtømmende, og der vil naturligvis være studier, som har rødder i flere kategorier på én gang, eller som på den ene eller anden måde falder uden for kategori. Jeg afslutter derfor oversigten over den skandinaviske skriveforskning med et kort afsnit om skriveforskning med et eksPLICIT psykologisk eller sociologisk perspektiv.

1.3.2.1 Den lingvistiske tradition

Den lingvistiske tradition er en af de ældste skriveforskningstraditioner i Norden. Med lingvistisk skal her forstås de studier, der undersøger skriftsproglig syntaks, semantik, morfologi, ortografi og/eller tekstlingvistiske forhold i elevtekster. Disse studier kan variere teoretisk og metodisk, men har det tilfælles, at de har lingvistiske strukturer i elevtekster som deres forskningsgenstand. Af ovennævnte områder er staveforskningen det mindste område i den lingvistiske tradition. Staveforskningsstudierne undersøger og klassificerer elevens afvigelser fra forskellige sproglige normer, men disse studier er forholdsvis sjældne inden for modersmålsforskningen³. Som det også fremgår af Wiggens (1982) review af norsk staveforskning fra perioden før 1981, er det imidlertid en smule overraskende, eftersom stavning historisk har haft en central plads både sprogpolitisk og ikke mindst i modersmålsundervisningen i skolen. Staveforskningen har heller ikke stærke rødder i dansk skriveforskning, og der er således kun ganske få studier de sidste 20 år, som har haft skoleelevers stavning som sit primære genstandsfelt (se fx Engmose, 2019; Juul, 2005; Veber, 2016).

Et vigtigt afsæt for den lingvistiske skriveforskning er fremkomsten af korpuslingvistikken og de dertilhørende computationelle analyseværktøjer i 1960'erne og 1970'erne (McEnery & Hardie, 2013), som gav mulighed for at processere og identificere sproglige strukturer og mønstre på tværs af et stort antal elevtekster. Selvom det åbnede for nye

³ Der er imidlertid langt større tradition for staveforskning inden for andet- og fremmedsprogdidaktikken.

muligheder i skriveforskningen, fik det i praksis ikke den store udbredelse inden for den pædagogiske skriveforskning, men snarere inden for bestemte lingvistiske forskningsmiljøer. Dette skyldes sandsynligvis, at det dels var en langsommelig proces at indsamle, digitalisere og manuelt annotere papirbaserede tekster, dels at elevsprog – på et tidspunkt hvor der endnu ikke fandtes maskinlæringsmetoder – var svært at processere og analysere pga. et stort antal stavfejl, slåfejl etc. Der findes således – også den dag i dag - forholdsvis få tekstkorpusser i Norden bestående af elevtekster. Vigtige undtagelser til denne tendens er en række svenske korpusbaserede studier fra 70'erne og frem til i dag. Et af de ældste⁴, men samtidig mest indflydelsesrige af disse studier er Hultman & Westmans *Gymnasistsvenska* fra 1977, som var en del af et stort svensk forskningsprojekt (Skrivsyntax) i gymnasiet. I Hultman & Westmans (1977) lingvistiske delstudie undersøgte de en række leksikale og syntaktiske variable i 151 gymnasiestile samt de lingvistiske variables relation til karakter og køn. Elevteksterne blev desuden sammenlignet med et benchmarkkorpus af journalistiske tekster. Blandt studiets centrale konklusioner var, at mens et simpelt mål som fx tekstlængde var velegnet til at differentiere tekster på de lavere karakterniveauer, var det i højere grad præcision i udtryk (dvs. skrive det samme meningsindhold med færre ord), der adskilte tekster med høje karakterbedømmelser. Samtidig pegede studiet på, at de professionelle journalistiske tekster på nogle punkter faktisk var mindre komplekse end gymnasiestilene (fx ved at anvende færre ledsætninger), hvilket indikerede - hvad der er blevet bekræftet flere gange af senere studier - at øget sproglig kompleksitet ikke altid kan sidestilles med øget tekstkvalitet.

Af lignende svenske studier fra samme periode kan bl.a. peges på Grundins (1975) studie af udviklingen i simple produktivetsmål (antal ord, antal sætninger, antal langord etc.) fra 2. klasse i grundskolen til 3.g i gymnasiet samt Petterson (1980, 1989), Larsson (1984), Garme (1988) og Strömquists (1987) studier fra 1980'erne, som bl.a. undersøgte henholdsvis tekstlængde og syntaks, stavning, skrivestrategier og brugen af afsnitsinddeling i elevtekster. Særligt Larssons (1984) undersøgelse var på daværende tidspunkt et betydeligt bidrag til skriveforskningen, idet han som en af de første kobledes elevernes skrivepræstationer (karakter) til deres sociale baggrund og læsekompetencer. I 90'erne

⁴ Det ældste studie, jeg har kunnet finde, er Björnssons "Uppsatsbedömning och uppsatsskrivning" fra 1960.

er der imidlertid ikke mange lingvistiske studier af elevtekster. Dette sker, som det vil fremgå senere i gennemgangen af den etnografiske tradition, parallelt med en social vending inden for skriveforskningen, hvor fokus i højere grad kommer på de sociale praksisser, som elevernes skrivning indgår i, snarere end på teksterne selv. Vi skal således frem til slut 90'erne/start 00'erne, før der igen begynder at dukke nye lingvistiske studier op (se bl.a. Andersson, 2002; Chrystal & Ekvall, 1999; Josephson et al., 1998; Ledin, 1998, 2001; Nyström, 2000; Olevard, 1999, 2002). Fra denne periode kan bl.a. fremhæves Östlund-Stjärnegårdhs (2002) analyse af sproglige forskelle mellem elevtekster fra gymnasiet, som er vurderet som henholdsvis bestået og ikke-bestået. Hun finder systematiske forskelle mellem de to grupper, hvad angår tekstlængde, brugen af afsnitsinddeling og til dels ift. sætningsstruktur, men finder omvendt ikke systematiske forskelle angående brugen af kohærensmarkører. Et gennemgående træk ved mange af de studier, som er refereret i ovenstående, er, at de undersøger forholdet mellem karakter og sproglige træk i teksterne, og de mestendels er rettet mod gymnasietekster.

Foruden de ovenfor omtalte studier er der de sidste 20 år udkommet studier inspireret af systemisk funktionel lingvistik (fx Folkeryd, 2006; Silén, 2012), en række studier med fokus på forskelle mellem L1- og L2-skrivning (fx Apelgren & Holmberg, 2020; Kokkinakis & Magnusson, 2011; Magnusson & Johansson, 2009), nogle studier med fokus på anvendelsen af multimodale ressourcer i elevtekster (fx Engblom, 2016; Åberg, 2022) samt et par svenske ph.d.-projekter, som har fokuseret på forskellige lingvistiske strukturer i elevtekster. I Johanssons ph.d.-afhandling (2009) undersøges fx udviklingen af både skriftsprog og talt sprog fra grundskoleniveau over gymnasiet til universitetet. Studiet adskiller sig fra tidligere studier ved, at studiet fokuserer på fire aldersgrupper med forholdsvis stor aldersspredning (10 år, 13 år, 17 år og universitetsstuderende), det indarbejder keyboard-logdata fra elevernes skriveproces i analysen, og det sammenligner skriftsproglige tekster med optagede mundtlige tekster. En vigtig konklusion fra studiet er, at de mest synlige forskelle i de skriftsproglige variable (bl.a. antal ord og antal t-units) er mellem de 13- og 17-årige, mens der omvendt ikke er stor forskel mellem 10-årige og 13-årige eller mellem 17-årige og universitetsstuderende. Et andet lingvistisk ph.d.-projekt er Nordenfors (2011), som dels undersøger udviklingen i tekstlængde, ordlængde og ordvariation fra 5.-9. klasse, dels undersøger hvordan konstruktet *voice* udvikler sig henover

de fem klassetrin. En central pointe hos Nordenfors, som bekræfter Hultman & Westmans (1977) konklusion, er, at udviklingen i tekstlængde stagnerer over tid, og at det i højere grad er fx ordvariation, der korrelerer med høj karakter. Nordenfors lægger sig således i forlængelse af en række af de studier, som er angivet ovenfor, idet han fokuserer på forholdet mellem karakter og sproglige strukturer i teksten.

De seneste 5-6 år har den lingvistiske skriveforskning fyldt mindre. En væsentlig undtagelse til denne tendens skal findes på Uppsala Universitet, hvor man både har udviklet et korpus af elevtekster fra grundskolen (Megyesi et al., 2016) og et webbaseret værktøj til automatiseret tekstanalyse på svensk og engelsk, det såkaldte SWEGRAM (Megyesi et al., 2019; Näsman et al., 2017). SWEGRAM består af en række værktøjer til at annotere tekster (dvs. præprocessere teksterne ved fx at normalisere retstavning, fjerne endelser, identificere ordklasser etc.) samt en række værktøjer til at analysere forskellige lingvistiske variable i teksterne (fx ordvariationsindeks, antal ord, ordlængde). Palmér (2018) har brugt SWEGRAM-værktøjet til at undersøge, hvorvidt værktøjet kan bruges til at understøtte bedømmelse og karaktergivning af elevtekster. Palmér viser, at selvom de lingvistiske variable, som anvendes i studiet (tekstlængde, ordlængde, ordvariation og nominalkvoter), generelt korrelerer med karakter, så varierer korrelationskoefficienten forholdsvis meget relativt til den opgave, eleverne besvarer, og hun konkluderer derfor, at de lingvistiske variable skal tilpasses de enkelte opgaveforlæg, samt at antallet af lingvistiske variable bør øges. I et igangværende studie anvender Palmér & Hussenius (komende) ligeledes SWEGRAM til at undersøge skrivning af sagprosa (primært argumenterende tekster) i 6. klasse. Studiet anvender desuden leksikogrammatisk analyse til at undersøge teksterne kvalitativt. Gennem kombinationen af disse to tilgange beskriver Palmér & Hussenius to udviklingsstadier i elevernes argumenterende tekster; ét stadie som kommer til udtryk i en verbal stil med udbyggede verbalfraser og ét stadie som er mere informationstæt med lange nominalfraser og øget sætningskompleksitet.

Som det fremgår af ovenstående, er fortællingen om den lingvistiske skriveforskning i Norden på mange måder fortællingen om den lingvistiske skriveforskning i Sverige. Det betyder selvfølgelig ikke, at der ikke er foretaget relevant lingvistisk skriveforskning i andre nordiske lande, men blot at Sverige historisk set har haft meget stærkere forskningsmiljøer og -traditioner inden for dette felt. I Norge er den lingvistiske tradition forholdsvis ny i den pædagogiske skriveforskning - en pointe som også understreges af Sigmund

Ongstad i 2002, som peger på, at lingvistik og pædagogik på daværende tidspunkt havde været institutionelt og metodologisk adskilt i norsk skriveforskning (Ongstad, 2002). Undtagelser er projekterne *Nord-Skriv* fra 80'erne (Evensen, 1987, 1988) og *Skrive-Puff* fra 90'erne, som bl.a. undersøgte skriveudvikling i elevtekster fra ind- og udskoling (Evensen, 1990, 1991; Lorentzen & Evensen, 1995), samt Berge (2005), Evensen (2005) og Vagles (2005) analyser af tekster indsamlet ifm. det store skrivedidaktiske forskningsprojekt *KAL* (Berge et al., 2005a, 2005b). Særligt Vagles studie er interessant i denne sammenhæng, fordi det baserer sig på kvantitative tekstanalyser, og Vagles konklusioner bekræfter i vid udstrækning de fund angående udviklingen af bl.a. tekstlængde og ordlængde, som den tidlige svenske skriveforskning gjorde i 70'erne og 80'erne. Af nyere norske studier kan nævnes Skjeltens (2013) analyse af sproglige forskelle mellem tekster på de to øverste karakterniveauer ved afgangsprøven i den norske grundskole, Hundals (2017) studie af kohærens- og kohæsionsmarkører i elevtekster fra mellemtrinnet, Kristoffersens (2019) studie af udviklingstræk i ordvalg i elevtekster fra 2.-4. klasse samt Skar & Berge (2017), som undersøgte relationen mellem karakter og en række lingvistiske mål (herunder ordvariation, LIX-tal, antal ord og antal langord mm.) i et mindre korpus af elevtekster fra 8. klassetrin. En central pointe i sidstnævnte studie er, at den skriveopgave, som eleverne skriver på baggrund af, har stor betydning for, hvilke sproglige variable der korrelerer stærkest med karakter. Denne konklusion bekræfter således resultaterne fra bl.a. Vagle (2005) og Palmér (2018).

Et stort norsk igangværende studie, *Functional Writing in the Primary Years* (Skar, Aasen et al., 2020) undersøger en række spørgsmål relateret til bl.a. skriveundervisning og lærerudvikling i indskoling. Det mest interessante i denne sammenhæng er imidlertid, at Skar, Jølle et al. (2020) også har udviklet skalaer til manuelt at bedømme en række forskellige sproglige forhold i elevtekster fra 1.-3. klasse. Disse skalaer måler bl.a. formelle elementer som syntaktisk variation, tegnsætning og stavning samt funktionelle elementer som modtagerbevidsthed og organisering af indhold. Det, der er særligt ved disse skalaer, er, at forskerne gennem en række forandringer ved vurderingsprocessen og mere udfoldede vurderingskriterier opnåede en gennemsnitlig interrater-reliabilitet på 0.86 for skalaerne, hvilket er markant højere, end man normalt ser ved skalaer til skrivebedømmelse. Til sammenligning viser den seneste rapport fra Danmark, at bedømmerreliabiliteten ved karaktergivning til de skriftlige afgangsprøver i dansk ligger på omkring 0.40 (Dolin et al.,

2018). Skar, Jølle et al.'s (2020) studie er således et eksempel på et assessment-perspektiv på skriveudvikling, som befinder sig på grænsen mellem en lingvistisk og en psykologisk tilgang.

Danmark er det land af de tre nordiske, som har den absolut korteste tradition for lingvistisk skriveforskning. Foruden Sandersens (1983) kvalitative studie af sproglige normafvigelser og dialektale træk i en mindre samling elevtekster fra 4. og 9. klassetrin har der ikke været tradition for denne type skriveforskning i Danmark i mange år. Et nyt stort forskningsprojekt bryder dog denne tendens. ATEL (*Automated Tracking of Early Stage Literacy Skills*) er et stort tværdisciplinært projekt med deltagelse af både fagdidaktiske forskere og forskere fra computer science. Projektets formål er at udvikle en sproglig model på baggrund af maskinlæring til beskrivelse af udviklingsniveauer i den tidlige skriveudvikling. Modellen skal inkorporeres i et skrivedidaktisk læremiddel med henblik på automatiseret vurdering af elevteksterne og feedback til lærerne med forslag til skrivedidaktiske fokusområder for den enkelte elev. Modellen baserer sig på en blanding mellem klassiske funktionelle kategorier fra dansk grammatik (fx Diderichsens feltanalyse) og bredere funktionelle kategorier fra socialesemiotikken og den systemisk funktionelle lingvistik med fokus på tekstniveauet (fx attitude og stemme) (Bundsgaard et al., 2022b; Kabel et al., 2022). I projektet identificerer forfatterne fem niveauer i begynder-skriveres skriveudvikling, som beskrives på et tekst-, syntaks og leksisniveau (med fokus på verber og beskrivere) samt et ortografisk niveau. Forfatterne viser bl.a., hvordan elevernes tekster på det laveste niveau består af præfonetiske ord, som ikke eller kun i begrænset omfang danner en sammenhængende sætning, mens eleverne på det øverste niveau anvender mere komplekse syntaktiske former, mere abstrakte verbaler og en større variation af forskellige beskrivere (Bremholm et al., 2022).

Ser man på tværs af de nordiske studier, som jeg har angivet ovenfor, er der en væsentlig tendens, som er værd at fremhæve, som handler om metodologi. Der er mange kvantitative studier i svensk lingvistisk skriveforskning, mens der er en større blanding mellem kvantitative korpusanalyser og kvalitative caseanalyser af enkelttekster i den norske lingvistiske skriveforskning. De tidlige svenske korpusstudier (særligt Hultman & Westman, 1977) fik stor opmærksomhed, dels pga. skalaen af det empiriske materiale, dels fordi de statistiske analyser var baseret på håndannoterede korpusser. Der har imid-

lertid ikke været den store udvikling i, hvilke lingvistiske variable der indgår i de kvantitative studier. Tekstlængde og ordvariation baseret på målet OVIX, som Hultman & Westman introducerede i 1977, anvendes stadig i mange af de nyeste lingvistiske studier, fx Nordenfors (2011), Berge & Skar (2017) og Gustafsson & Lassus (2020). Det danske ATEL-projekt adskiller sig på dette punkt ved at fokusere på en stor mangfoldighed af variable samt ved at fokusere på funktionelle kategorier knyttet til tekstniveauet. Projektet er imidlertid også metodologisk anderledes opbygget end de andre studier, idet teksterne i projektet ikke analyseres med færdige computationelle værktøjer, men snarere på baggrund af et håndannoteret sample af tekster, som så omdannes computationelt til analyseværktøjer vha. maskinlæring. Derudover fokuserer projektet ikke på udviklingen af enkelte lingvistiske variable, som ellers er standarden i meget af den lingvistiske skriveforskning, men snarere på relationen og interaktionen mellem variablerne vha. statistiske Rasch-modeller.

1.3.2.2 Den etnografiske og skrivedidaktiske tradition

Den etnografiske tradition har haft en fremtrædende plads i den nordiske skriveforskning – særligt de sidste 20 år (Berge, 2019; Smidt, 2018). Denne tradition etablerede sig for alvor i Norge i 1980'erne (Ongstad, 2002). Begrebet 'etnografisk' skal her forstås bredt og indbefatter de skrivestudier, som undersøger skrivningens sociale og bredere kulturelle dimensioner. Det er således også inden for denne tradition, at meget af den skrivedidaktiske forskning findes – forstået som empiriske analyser og refleksioner over skriveundervisningens indhold, mål, organisering og evaluering. Det, der adskiller den etnografiske tradition fra den lingvistiske, er således, at man i den lingvistiske tradition primært studerer artefakter/produkter, mens man i den etnografiske tradition i højere grad undersøger de retoriske kontekster og sociale praksisser, hvori artefakterne/produkterne skabes. Selvom etnografien selvfølgelig går langt tilbage (også inden for skriveforskningen), er det værd at starte beskrivelsen af den etnografiske skriveforskning med en kort beskrivelse af den teoretiske udvikling i forståelsen af skrivning som fænomen de seneste 30 år, eftersom den teoretiske udvikling er helt afgørende for den etnografiske traditions fremtrædende position i den nordiske skriveforskning.

Indtil midten af 90'erne (og i praksis også langt senere) var det sædvane i skriveforskningen at skelne forholdsvis skarpt mellem elevers læse- og skrivefærdigheder (Tannert & Berthelsen, 2021). Dette ændrer sig imidlertid gradvist i takt med, at den sociale vending inden for socialvidenskaberne, som langsomt havde fortrængt de individuelt orienterede behavioristiske og kognitivistiske tilgange, begyndte at dominere skriveforskningen til fordel for et sociokulturelt og integreret literacyperspektiv, hvor læse- og skrivepraksisser blev set som interdependente, og hvor literacy i højere grad end tidligere knyttedes til de praksisser, som udgør individets private og offentlige liv (Gee, 1999). Inden for den pædagogiske skriveforskning er et afgørende teoretisk skel i denne forbindelse The New London Groups indflydelsesrige manifest fra 1996, der introducerer begrebet *multiliteracies* som et begreb, der på baggrund af bl.a. funktionel lingvistik og sociokulturel læringsteori udvider literacy-begrebet til også at indbefatte de nye kommunikative muligheder, som bl.a. er opstået som følge af nye digitale teknologier. Dvs. at literacy-begrebet ikke blot er en samlebetegnelse for læsning og skrivning af skriftsproglige tekster, men også en betegnelse for produktion og reception af multimodale tekster i bred forstand (The New London Group, 1996). Disse idéer udfoldes efterfølgende med stor gennemslagskraft dels i et lingvistisk/semiotisk perspektiv hos Kress & van Leeuwen (1996) og Kress (2010), dels i et pædagogisk og didaktisk perspektiv hos bl.a. Cope & Kalantzis (2000, 2009). Selvom denne måde at sammentænke læsning og skrivning med produktion og reception af en bredere mangfoldighed af tekstformer først langsomt begynder at slå igennem i danskfagets policydokumenter i starten af 2010'erne (Tannert & Berthelsen, 2021), spiller det brede literacy-begreb allerede en væsentlig rolle i den nordiske etnografiske skriveforskning fra 00'erne og frem. Det afgørende her er ikke så meget fokuset på det udvidede tekstbegreb, men snarere den sociokulturelle forståelse af læsning og skrivning, som ligger bag literacy-begrebet, og som på mange måder passer godt sammen med etnografiens fokus på sociale praksisser og kontekster. Der er flere årsager til, at literacy-begrebet og den etnografiske skriveforskning har stået stærkt de seneste årtier. Lankshear & Knobel (2011) angiver foruden en række geo- og uddannelsespolitiske forhold, at Paulo Freires kritiske pædagogik med fokus på at understøtte udviklingen af literacy-kompetente borgere med kritisk bevidsthed samt den generelle fremkomst af so-

ciokulturelle teorier inden for human- og socialvidenskaberne (fx Vygotskys kulturhistoriske psykologi) har været væsentlige årsager til, at literacy-begrebet er blevet et dominerende teoretisk begreb i skriveforskningen.

Vender vi blikket mod de nordiske studier, som har undersøgt skrivning inden for denne etnografiske tradition, falder de groft skitseret i to kategorier; deskriptive studier af skrive(r)udvikling og interventionsbaserede studier af skrivedidaktik. Den første kategori fokuserer på udviklingen af skrivefærdigheder og skriveridentitet gennem etnografiske observationer og interviews. Disse studier adskiller sig således fra de lingvistiske studier, som undersøger skriveudvikling i teksterne, ved at fokusere på de sociale og identitetsmæssige forhold, som knytter sig til den enkelte elevskriver. Et af de mest betydningsfulde studier med dette fokus er Hetmars (2000) analyser af det, hun kalder skolske praksisformer i skriveundervisningen. Med 'skolsk' henviser Hetmar til skrivesituationer, som ikke har andre kommunikative formål end at leve op til de formelle krav, som stilles af læreren, og som derfor ikke forekommer andre steder end i skolen. Hetmar viser gennem konkrete caseanalyser af forskellige undervisningssituationer, hvordan elevers udvikling som skrivere ikke understøttes særligt godt af disse klassiske skolske praksisformer som fx stilskrivning (det hun kalder lærerens projekt), men snarere understøttes gennem skrivesituationer, der i højere grad tager udgangspunkt i elevernes egne mål og tekstpraksisser (elevens projekt). Dermed kobler Hetmar udviklingen af skrivefærdigheder og -identitet til et bredere spektrum af tekster og skrivesituationer, end dem eleverne typisk møder i skolens undervisning. En beslægtet pointe findes i Lundgrens (2013) studie af literacypraksisser og identitetsskabelse i den svenske grundskole. På baggrund af en række gruppeinterviews med elever fra 5. og 6. klasse konkluderer Lundgren bl.a., at udviklingen af elevernes identitet som 'skriftkyndige' i særlig grad er knyttet til de tekster og sociale literacy-praksisser, som de selv tilvælger, og som typisk er det, som Lundgren kalder 'hverdaglige' – dvs. tekster og praksisser uden for skolens domæne. Lundgren pointerer derfor – på samme vis som Hetmar – at disse praksisser er vigtige at inddrage, hvis vi skal forstå, hvordan elever udvikler deres identiteter som kyndige skrivere.

Et forholdsvis nyt og stort dansk forskningsprojekt, som også fokuserer på skriveudvikling og -identitet, er projektet *Faglighed og Skriftlighed* (FoS), som er gennemført i regi af forskere fra særligt Syddansk Universitet. Projektet består af en forunderøgelse,

som er gennemført i folkeskolens niende klasse (Christensen et al., 2014), samt to hovedundersøgelser med fokus på, hvordan elever udvikler henholdsvis skrivefærdigheder og skriveridentifikationer i gymnasiet (Krogh et al., 2015; Krogh & Jakobsen, 2016). Projektet består primært af longitudinale etnografiske klasserumsobservationer og interviews med elever samt såkaldt tekstetnografiske studier, hvor elevtekster, lærerrespons og elevinterviews analyseres i sammenhæng. Projektet skiller sig ud dels gennem bredden af det analytiske arbejde, hvor der fokuseres på alt fra tekstkulturforskelle i overgangen fra grundskole til gymnasium, teknologiens rolle i elevers skriverudvikling, kreativ skrivning til genreskrivning i forskellige fag, dels ved at undersøge mange af disse fænomener i et longitudinalt perspektiv, hvilket giver mulighed for at se mønstre og forandringer i disse fænomener over tid. Projektet bidrager - foruden et stort teoretisk rammeværk til at forstå og analysere skriverudvikling - med en række kvalitative empiriske fund om bl.a. betydningen af overgange mellem tekstkulturer i uddannelsessystemet samt betydningen af faglige traditioner og genrer for elevernes mulighed for at udvikle forskellige former for skriveridentiteter i løbet af gymnasiet (Hobel et al., 2019).

I forlængelse af FoS-projektet har Troelsen undersøgt en række forhold relateret til afgangsprøven i skriftlig fremstilling. Blandt andet har Troelsen (2020) undersøgt tekst- og kvalitetsnormer og faglige diskurser i både skriveordrer, elevtekster og danskensorsers bedømmelser. Troelsen (2020) er et eksempel på en metodisk tilgang til skriveudvikling, som - i tråd med FoS-projektet - befinder sig inden for det, som forfatteren selv kalder antropologisk tekstvidenskab, og som dermed er i grænsefladen mellem den lingvistiske og den antropologiske tradition. Dette indebærer fx, at casestudier af enkeltelevers tekster, skriveordrer og lærerfeedback analyseres i sammenhæng med feltobservationer og interviews. På denne måde er der fokus på den kontekst, som teksten konstrueres i, og på de sammenhænge, der er mellem tekst og didaktisk kontekst.

Troelsens studier er således også relateret til den anden kategori af studier inden for den etnografiske tradition, som handler om skrivedidaktik. Den skrivedidaktiske forskningslitteratur er mere omfattende, end det kan lade sig gøre at redegøre for her, og jeg vil derfor i stedet pege på en række store skrivedidaktiske forskningsprojekter, som har haft stor indflydelse for både skriveforskningen og for skrivedidaktisk praksis. I Norge har man gennem en årrække gennemført flere studier, som har arbejdet med at analysere og udvikle evalueringspraksis inden for skriveområdet. KAL-projektet, som også kort er

omtalt i forrige afsnit, havde specifikt fokus på afgangseksamen og på censorernes vurderinger af elevtekster (Berge et al., 2005a, 2005b), mens Normprojektet, som er gennemført i perioden 2012-2016, undersøgte, hvilke normer og standarder for skrivning der var passende i skolen, samt hvilken effekt fælles forventningsnormer til teksterne havde på lærerens vurdering af teksterne og på elevernes skrivekompetencer (Solheim & Matre, 2014). Et væsentligt omdrejningspunkt for Normprojektet har været udviklingen af det såkaldte 'skrivehjul', som er en teoretisk model for skrivning i skolen med særligt fokus på skrivningens sociokulturelle kompleksitet (Berge et al., 2016). Modellen har efterfølgende været anvendt som et centralt element i et stort interventionsstudie, hvor lærerne brugte modellen som et led i planlægning og vurdering af skriveundervisning. Resultaterne af projektet viste, at interventionen havde delvist positive effekter på kvaliteten af elevernes tekster (Berge et al., 2019). Arbejdet med at udvikle og beskrive evalueringspraksis i den norske grundskole er fortsat i det igangværende projekt *Functional Writing in the Primary Years* (Skar, Aasen, et al., 2020), som også er omtalt i forrige afsnit. Foruden at være et stort RCT-studie med fokus på interventionens effekt på elevernes skrivekompetencer indeholder projektet også observationer af skriveaktiviteter i klasserummet samt observationer og interviews med fokus på lærernes undervisning og professionelle udvikling. Projektet bygger således videre på projektet *Skrivning som grundlæggende færdighed og udfordring* (2006-2010), som bl.a. undersøgte skrivning i skolen ud fra et lærerperspektiv og havde særligt fokus på, hvilken form for viden lærerne skal have for at understøtte elevernes skriveudvikling i forskellige fag (Smidt, 2010; Smidt et al., 2013). I forlængelse heraf har Blikstad-Balas et al. (2018) i et stort deskriptivt projekt af skriveundervisningen vist, hvilke funktioner skrivning indtager i norskundervisningen. En central pointe fra studiet er, at længerevarende skriveaktiviteter i skolen stilladseres på mange forskellige måder, og at der arbejdes med et klart proces- og genrefokus.

De ovennævnte norske studier er kendetegnet ved at være meget store tværdisciplinære og multimetodiske projekter. Der er imidlertid også gennemført en række mindre interventionsprojekter i Skandinavien med fokus på skriveundervisning. Matthiesen (2013) undersøgte fx potentialerne i at arbejde med elevstyret imitatio som et retorisk forankret skrivepædagogisk princip, mens Molbæk (2019) undersøgte potentialer og udfordringer i at inddrage autentiske skrivesituationer i undervisningen. Denne type af projekter er interventionsbaserede og er kendetegnet ved kvalitativt at undersøge og afprøve

specifikke og praksisnære skrive-didaktiske fænomener. Kabel (2016), Walldén (2020) og Buch (2015) er imidlertid eksempler på skrive-didaktiske studier, som ikke har et interventionsfokus, men et deskriptivt og analytisk fokus på didaktiske skrivepraksisser og -kulturer, og som derfor i højere grad er forbundet med FoS-projektets deskriptive og sociokulturelle perspektiv.

1.3.2.3 Øvrig skandinavisk skriveforskning

Jeg vil i dette afsnit kort beskrive to yderligere perspektiver i den nordiske skriveforskning, som er relevante for denne afhandling, men som ikke fylder meget i forskningsfeltet. Den første er den psykologiske tradition, som på den ene side er meget lille, i den forstand at der findes forholdsvis få skriveforskningsprojekter i Skandinavien med et rent psykologisk perspektiv, men på den anden side er den psykologiske tradition latent til stede i mange skriveforskningsprojekter, som undersøger skriveudvikling. Dette gælder især de lingvistiske studier, som enten undersøger udviklingstrin i elevernes skrivning, eller som har udviklet skalaer og værktøjer til skrivebedømmelse. Disse projekter er – selvom deres genstandsfelt er sproglige strukturer - i en vis forstand psykolingvistiske projekter, der undersøger et underliggende kognitivt konstrukt om skrivekompetence, og som dermed også har forbindelser til den psykologiske tradition (Tannert, 2021). Studier, som eksplicit undersøger kognitive dimensioner af skrivningen, er imidlertid sjældne – modsat inden for læseforskningen, hvor der oftere foretages studier af fx hukommelsens betydning for læseforståelsen eller forskelle i afkodning, når der læses på henholdsvis papir og skærm. Ser man bort fra de studier, som enten eksplicit eller implicit undersøger skrivekompetence eller –udvikling som kognitivt konstrukt, er den psykologiske forskningstradition meget sporadisk. Der findes enkelte mindre skandinaviske casestudier af psykologiske fænomener som fx skrivevanskeligheder (se bl.a. Arnbak & Petersen, 2017; Svendsen, 2016), men disse er forholdsvis sjældne. Skrivelyst og –motivation er empirisk underbelyste i nordisk skriveforskning og er primært blevet beskrevet ud fra teoretiske eller praksisanvisende perspektiver (se bl.a. Madsbjerg & Friis, 2013).

Det andet perspektiv, som jeg vil fremhæve, er det sociologiske perspektiv på skriveudvikling, som ligesom det psykologiske er en underbelyst dimension i nordisk skriveforskning. Der er således ganske få studier, som har dette perspektiv. Larsson (1984) er

et eksempel på et studie, hvor sammenhængen mellem sproglige træk elevtekster og elevernes sociale baggrund undersøges, mens Hultman & Westman (1977), Magnusson & Kokkinakis (2009, 2011) og Vagle (2005) er eksempler på studier med fokus på, hvordan kønsforskelle manifesterer sig i elevernes tekster. Ingen af studierne har imidlertid det sociologiske perspektiv som hovedfokus, men har det i højere grad med som et supplerende perspektiv.

1.3.2.4 Om dette studies placering i den skandinaviske skriveforskning

Ovenstående gennemgang af de tre forskningstraditioner i skandinavisk skriveforskning blottelægger tre forhold, som er væsentlige for denne afhandling. For det første viser gennemgangen en national skævvridning i fordelingen af studier mellem de tre forskningstraditioner. Svensk skriveforskning er historisk set stærkt funderet i den lingvistiske tradition, mens norsk og især dansk skriveforskning i langt mindre grad har været optaget af den rent lingvistiske dimension af skrivning og i stedet har fokuseret på de etnografiske og skrivepædagogiske dimensioner. Dette betyder, at vi fra et dansk perspektiv ikke har megen viden om, hvad der sprogligt set kendetegner tekster skrevet i skolen. For det andet viser gennemgangen, at mens en del studier undersøger skriftsproglige udviklingstræk relativt til klassetrin, er der relativt få studier som undersøger sammenhængen mellem de sproglige strukturer og individuelle, askriptive variable knyttet til den enkelte elev såsom social baggrund. De studier, som rent faktisk foretager denne slags analyser, er ofte af ældre dato og har snarere fokus på sammenhængen mellem social baggrund og et endimensionalt mål for tekstkvalitet (fx karakter) end på sammenhængen med mere tekstnære sproglige variable. For det tredje har relativt få skandinaviske studier udnyttet de metodiske muligheder, som følger med nye NLP-værktøjer (se beskrivelse af NLP i afsnit 3.3.1), herunder mulighederne for skalering og for at udvikle modeller og mål, som specifikt er trænet på elevsprog.

Denne afhandling forsøger på forskellige måder at kompensere for de tre ovenstående tendenser – dog med primært fokus på de to første punkter. Afhandlingen er det første danske projekt, som anvender automatiserede digitale værktøjer til at analysere sproglige træk i elevtekster fra mellemtrinnet og udskoling. Det danske ATEL-projekt, som blev beskrevet i afsnit 1.3.2.1, er det eneste andet danske projekt, som undersøger skriveudvikling på et kvantitativt og lingvistisk grundlag. ATEL adskiller sig imidlertid fra

denne afhandling på særligt to punkter. For det første fokuserer ATEL-projektet på begynderbeskrivelse, hvorimod jeg i denne afhandling fokuserer på skriveudvikling i den sidste del af grundskolen. For det andet er projekterne bygget forskelligt op rent metodisk. ATEL-projektet starter med en række teoretiske konstrukter, som herefter operationaliseres og scores manuelt i en stor stikprøve af tekster. Stikprøven danner herefter grundlag for maskinlæring, der kan anvendes og generaliseres til det resterende korpus. Dette kan selvfølgelig kun lade sig gøre, fordi ATEL er et meget stort og tværdisciplinært projekt, som er baseret på et samarbejde mellem et digitalt læremiddel, der bidrager med et stort antal tekster, didaktiske forskere og forskere fra computer science. Denne afhandling minder metodisk mere om de svenske lingvistiske studier af elevtekster, som er beskrevet tidligere, idet jeg anvender de automatiserede analyseværktøjer, der allerede er tilgængelige, som jeg herefter uddyber og konkretiserer gennem kvalitative analyser af teksterne.

Som det fremgår af ovenstående gennemgang, er der forholdsvis få studier, som undersøger relationen mellem tekstnære, sproglige variable og askriptive data knyttet til den enkelte elev. Dette betyder bl.a., at vi ikke ved, om og hvordan den sociale ulighed, som den sociologiske uddannelsesforskning ofte dokumenterer i forhold til bl.a. karakterer og uddannelsesvalg (Skopek et al., 2017), manifester sig helt konkret i elevernes tekster, og vi ved heller ikke, hvad betydningen af social baggrund er sammenlignet med fx alder, køn eller skrivelyst. I denne afhandling undersøger jeg derfor netop disse relationer for at give en mere detaljeret beskrivelse af, hvordan skrivefærdigheder udvikler sig i skolen både diakront og synkront. På denne måde bidrager afhandlingen med viden om elevers skriftsproglige udvikling fra et perspektiv, som indtil videre er meget underbelyst.

Der findes ikke danske studier, som anvender NLP-baserede værktøjer til at beskrive og undersøge elevsprog, og derfor træder jeg i denne afhandling også på forholdsvis ukendt metodisk grund. Som jeg også indledte med i beskrivelsen af projektet, betyder dette, at projektet – foruden det primære empiriske sigte med at beskrive skriftsproglig udvikling i grundskolen - har et metodeudviklende sigte. Dette kommer til udtryk i forskningsspørgsmål 2. Her vil der bl.a. være fokus på at kvalitativt undersøge, hvorvidt og hvordan de kvantitative variable hænger sammen med forskellige grammatiske strukturer i teksterne. Afhandlingen adresserer således ikke direkte behovet for at udvikle nye NLP-baserede mål, der specifikt er trænet på elevsprog. Der er snarere tale om teoretisk metoderefleksion end om konkret teknisk udvikling af NLP-baserede analyseværktøjer.

1.3.3 International skriveforskning med et tekstorienteret og kvantitativt fokus

Som det fremgår af gennemgangen af den nordiske skriveforskning har særligt den lingvistiske tradition en kvantitativ forankring. Dette gælder også den internationale skriveforskning. Dette skyldes blandt andet, at sproglige overfladestrukturer i *tekstartefakter* er lettere at kvantificere end sociale eller kulturelle dimensioner af *tekstpraksisser*. Den korpusbaserede skriveforskning er imidlertid et heterogent forskningsfelt. Der findes således en stor diversitet relativt til bl.a. fag (L1, L2, EFL, STEM etc.), uddannelsesniveau (pre-school, secondary school, higher education etc), sprog, teksttype (narrative, argumentative, expository etc.) og relativt til skrivekonstrukturer og andre variable (early literacy, voice, stance, register etc.). Groft skitseret falder forskningen i to grupper – en henholdsvis synkron og diakron gruppe.

1.3.3.1 Diakrone studier af elevtekster

Inden for den diakrone tradition for skriveforskning er man interesseret i, hvordan sproglige træk i elevernes tekster udvikler sig over tid. Studierne anvender både longitudinale, kohortebaserede designs, hvor bestemte elevgrupper følges over tid, og tværsnitsdata, hvor alder eller klassetrin/uddannelsesniveau bruges til at strukturere korpusset efter tid. Et godt eksempel på en diakron tilgang er Crossley et al. (2011), som på baggrund af tekster fra 9. klasse, gymnasiet og videregående uddannelse analyserede en lang række lingvistiske træk i teksterne med henblik på at identificere de variable, som bedst kunne differentiere mellem uddannelsesniveau, og som derfor kunne tolkes som kendetegn ved skriftsproglig udvikling. Langt de fleste af de lingvistiske variable udviste nonlinear ændringer mellem klassetrin, men forfatterne identificerede dog særligt tekstlængde og antal sætninger som variable, der viste stærke lineære udviklinger, og som derfor kunne ses som generaliserbare tegn på skriveudvikling.

En del af de diakrone studier undersøger skriveudvikling i relation til bestemte genrer og teksttyper. Beard & Burrell (2010) og Beard et al. (2016) undersøger fx udviklings-træk i henholdsvis berettende og argumenterende elevtekster skrevet af 9-11 årige elever. Studiernes kvantitative dimension er baseret på håndscorede træk i elevernes tekster, og de finder bl.a., at der i de argumenterende tekster sker en stigning i variationen af sætningstyper og –længde, mens der i de berettende tekster sker en stigning i anvendelsen af

forskellige narrative træk som fx udfoldede karakter- og miljøbeskrivelser samt konsistent anvendelse af tempus og tredjepersonspronominer. Lignende studier af berettende tekster ses hos Hayes (2011) og Nelson & van Meter (2007). Troia et al. (2019) fokuserer ligeledes på berettende tekster, men skiller sig ud fra de førnævnte studier ved at analysere et meget stort antal lingvistiske variable i teksterne ved hjælp af værktøjet Coh-Metrix⁵, som er et af de største analyseværktøjer til automatiseret tekstanalyse med over 100 lingvistiske variable (se detaljeret beskrivelse hos McNamara et al., 2014). Troia et al. finder klassetrinsrelaterede forskelle fra 4. til 6. klassetrin i næsten alle variable, men de variable, der er knyttet til syntaks, skiller sig dog ud, idet det kun er tegnsætningskorrekthed (punctuation accuracy), der viser udvikling relativt til klassetrin, mens de øvrige syntaksvariable som fx sætningslængde ikke viser systematisk udvikling over tid.

Beers & Nagy (2011) er et eksempel på et diakront skrivestudie, som kombinerer et fokus på skriveudvikling, der går på tværs af både klassetrin og genrer. Forfatterne undersøger, hvordan to syntaktiske variable (antal ledsætninger per helsætning og længden af ledsætninger) forandrer sig på tværs af fire genrer og fem klassetrin (3.-7. klassetrin). Forfatterne konkluderer bl.a., at der er stor variation i de syntaktiske mål mellem genrer, og at denne variation hverken øges eller falder henover klassetrin. Genreeffekten er altså i deres studie konsistent over tid. Et lignende dobbeltfokus på skriveudvikling relativt til både klassetrin og genre ses hos Durrant et al. (2020), som har et meget specialiseret fokus på udviklingen af adverbelle ledsætninger i tekster skrevet af børn mellem 6 og 16 år, hvilket i en dansk kontekst svarer til, at eleverne er fra 0. klasse og op til 9. klasse. Forfatterne viser, at udviklingen i de adverbelle ledsætninger er meget kompleks, og at der er forskelle i udviklingen relativt til både genre og ledsætningstype. I nogle genrer og ledsætningstyper er der forholdsvis lineære udviklinger henover klassetrin, mens der i andre ses såkaldte udviklingsplateauer, hvor udviklingen stopper på et bestemt niveau og herefter flader ud.

Netop det syntaktiske niveau er et af de mest udbredte analytiske fokusområder i den diakrone, kvantitative skriveforskning. Crossley (2020) pointerer i et review af de lingvistiske træk, som oftest undersøges i studier af skriveudvikling, at den underliggende

⁵ Coh-Metrix fungerer imidlertid ikke på danske tekster og kan derfor ikke anvendes på teksterne i dette studie.

grundtanke i mange studier med fokus på syntaks er, at øget syntaktisk kompleksitet indikerer mere avancerede skrivefærdigheder. Debra Myhill's (2008, 2009) forskning i syntaktisk udvikling i elevtekster er i denne forbindelse svær at komme udenom. Hun viser bl.a., at syntaktisk udvikling i højere grad hænger sammen med skrivekompetence end med alder, og hun viser, at syntaksen hos svage skrivere bl.a. er parataktisk (mange sideordnende konjunktioner), indeholder fortrinsvis temporale underordnende konjunktioner og er præget af simple subjekt-verbal-opbygninger. Desuden er de svage skrivere mere orienterede mod indhold end mod skrivningens retoriske dimensioner. Lignende resultater ses hos Gouveia (2013b, 2013a), som viser generelle udviklinger i syntaktiske variable henover klassetrin. Studierne viser desuden, at der er træk, som godt nok ikke viser gennemsnitlig stigning, men hvor der er enkeltteksempler på kompleksitetstræk i tekster fra de ældste klassetrin, som ikke findes hos yngre skrivere.

Den syntaktiske udvikling er forholdsvis veldokumenteret blandt studier med fokus på begynderskrivning. Her viser Berninger et al. (2011) bl.a., at såkaldte 'run-on sentences', dvs. ekstraordinært lange sætninger (typisk parataktiske), samt sætningsfragmenter falder med alderen, mens Wagner et al. (2011) viser, at der i stedet med tiden sker en øgning i længden af begynderskriverses helsætninger (hovedsætning med alle tilhørende ledsætninger) samt en stigning i det, forfatterne kalder 'clause density', som er ratioen af det totale antal af hoved- og ledsætninger divideret med antallet af helsætninger. Dette bekræftes også af en række ældre skriveudviklingsstudier, som ligeledes finder generelle stigninger i både syntaktiske perioder og andre dele af elevernes syntaks (Haswell, 2000; Hunt, 1965, 1966, 1970; Loban, 1976).

Et andet analytisk fokuspunkt i den diakrone skriveforskning er på elevernes produktivitet, dvs. deres evne til at producere lange tekster med mange sætninger (se bl.a. Dockrell et al., 2015; Justice et al., 2006). I disse studier er det et gennemgående fund, at elevtekster bliver længere (dvs. indeholder flere ord) i takt med, at eleverne bliver ældre, og at antallet af sætninger (delvist som et resultat af den øgede tekstlængde) også øges med alderen. Hall-Mills & Apel (2015) er et eksempel på et studie, som godt nok ikke eksklusivt kigger på produktivitet, men alligevel ender med at sige en del om betydningen af produktivitet relativt til andre lingvistiske træk. Forfatterne undersøger en lang række lingvistiske variable i elevtekster fra 2.-4. klasse, hvorefter de gennem eksplorative faktoranalyser identificerer fem funktionelle faktorer i teksterne. Her viste det sig, at den

faktor, som de kalder produktivitet, og som dækker over det totale antal ord, sætninger og antallet af forskellige ord, var den faktor, der havde bedst sammenhæng med klassetrin, mens der til gengæld ikke var den store forskel i produktivitet mellem genrer. Et lignende resultat ses i et studie af Puranik et al. (2008), som også viser, at produktivtetsvariablerne var de eneste, der steg kontinuerligt med alderen (3.-6 klasse), hvorimod forskellige kompleksitets- og korrekthedsvariabler viste ingen eller varierende grad af sammenhæng med alder.

Et tredje analytisk fokuspunkt i diakrone studier er det leksikale niveau. Studier af det leksikale niveau kan fx fokusere på forskellige former for leksikal variation eller leksikal sofistikation (se afsnit 4.1.2 for uddybning). En del studier, der undersøger bestemte lingvistiske konstrukter som 'stance' og 'voice' (dvs. latente fænomener, som kun kan afdekkes indirekte gennem operationaliseringer indeholdende lingvistiske træk i en tekst), har et specifikt fokus på udviklingen af leksikale træk. Dette ses bl.a. i studier af akademisk 'voice' eller stemme, som udgør et stort forskningsfelt i skriveforskning med fokus på videregående uddannelse eller academia (se overblik hos Matsuda & Jeffery, 2012). Mange af disse studier har imidlertid ikke et fokus på, hvordan disse elementer udvikler sig over tid, men har snarere et fokus på forskelle imellem fag og discipliner, hvilket gør, at de hører mere til under gruppen af synkrone studier. De studier, der både har et leksikalt og et diakront udviklingsperspektiv, er derfor mere sparsomme. En undtagelse er Reppen (2009), som har undersøgt forskelle i anvendelsen af leksikale 'bundles' henover klassetrin. Hun finder bl.a., at det er et gennemgående træk på tværs af klassetrin, at eleverne anvender bestemte fraser i indledningen af deres tekster, hvor de kobler an til skrivepromptet. I et lidt større studie af universitetsopgaver har Gardner et al. (2019) vist, at fx lange ord, nominaliseringer og abstrakte substantiver hænger sammen med skriveudvikling over tid, mens Haswell (2000) i et lidt mindre studie af gymnasietekster viser, at antallet af lange ord øges henover klassetrin. Det sidste eksempel, jeg vil nævne, er Struthers et al. (2013), som undersøgte forskelle i anvendelsen af kohæsiionsmarkører fra 4. til 7. klassetrin. Forfatterne viser, at der er meget lille eller ingen udvikling fra 4. til 6. klassetrin, hvorefter der sker en lidt større stigning på 7. klassetrin. Studiet er imidlertid mest interessant, fordi det repræsenterer en vigtig tendens i den kvantitative skriveforskning, hvor der arbejdes med at udvikle rubrics eller lignende instrumenter til manuel scoring af forskellige træk i elevtekster. Modsat studiet i denne afhandling, hvor det er de

automatiserede, computationelle analyser, der udgør det kvantitative datamateriale, er det hos Struthers et al. de manuelle scoringer på den udviklede rubric, der udgør det kvantitative datamateriale. De fleste af denne type af studier har imidlertid ikke fokus på diakron udvikling, men oftere på forholdet mellem rubricscore og tekstkvalitet (karakter).

De tre analytiske fokuspunkter, som jeg har beskrevet (produktivitet, syntaks, leksis), er typisk baseret på velkendte formelle grammatiske kategorier, som er forholdsvis lette at identificere og analysere i store mængder tekst. Der er imidlertid også en gruppe af tekster, som tager udgangspunkt i Hallidays (2013) systemisk funktionelle lingvistik (SFL). Christie & Derewianka (2010) er et godt eksempel på denne tilgang. Forfatterne undersøger, hvordan skoleelevers skrivning udvikler sig inden for forskellige genrer. Forfatterne nøjes imidlertid ikke med et deskriptivt blik på forskellige SFL-baserede sproglige kategorier, men supplerer også disse beskrivelser med funktionelle fortolkninger af de sproglige træks kommunikative funktioner samt deres relation til elevernes kognitive udvikling. Af andre SFL-baserede studier kan nævnes Derewianka (2007) og Christie (2002). Disse SFL-baserede studier adskiller sig fra de øvrige diakrone skriveudviklingsstudier på særligt to måder. For det første er mange af disse studier tæt forbundne med de skrivepædagogiske praksisser, som de analyserede tekster er skrevet inden for. Det vil sige, at de SFL-baserede studier har fokus på, hvordan tekster skabes i en dialektisk relation til undervisningskonteksten. For det andet baserer mange af de SFL-baserede studier sig typisk på et mindre tekstmateriale og mange af studierne er således primært baseret på kvalitative læsninger af elevernes tekster, mens det kvantitative ofte har en mindre eller ingen rolle. Der er således få (med det danske ATEL-projekt som en undtagelse), der har lavet kvantitative, SFL-baserede studier af skriveudvikling.

Ovenstående grupper af diakrone studier er de dominerende, men der findes selvfølgelig også mindre dominerende tendenser. Fx er der studier, som har et komparativt blik på udviklingen af skrivefærdigheder inden for skoleelevers L1- og L2-skriftsprog. Et godt eksempel er Yang & Sun (2015), som er et mindre komparativt studie, hvor fem elevers L1-udvikling sammenlignes med udviklingen inden for deres L2- og L3-sprog. Der laves kvantitative analyser af forskellige former for lingvistiske træk (fx antal korrekte sætninger, sætningslængde, antal hoved- og ledsætninger per helsætning) i teksterne. Forfatterne viser, at udviklingen inden for alle tre sprogområder er kendetegnet ved non-linearitet samt kaotiske og ikke-systematiske iterationer. Et andet eksempel på et lidt

sjældnere diakront studie er Kerz (2020), som anvender forskellige former for maskinlæringsmetoder til at beskrive udviklingen af tekstkompleksitet i elevtekster. Studiet har primært fokus på teknisk metodeudvikling, og har således ikke meget fokus på de lingvistiske resultater. Studiet er imidlertid et eksempel på en tendens inden for skriveforskningen, hvor man fokuserer på at udnytte nyere maskinlæringsalgoritmer og neurale netværk til at beskrive og forudsige forskellige dele af elevers skriveudvikling. Dette er relateret til en nyere tendens inden for skriveforskningen, som bl.a. kendes under betegnelsen *Writing Analytics*, der som en del af sit formål har at udnytte digital computerkraft og nye teknologier til at beskrive skrivning og skriveudvikling på nye måder (Moxley et al., 2017; Palmquist, 2019).

1.3.3.2 Synkrone studier af elevtekster

Det, der kendetegner de synkrone studier, er som bekendt, at de ikke undersøger skriveudvikling relativt til en tidsmæssig struktur, men i stedet undersøger sammenhængen mellem tekst og en række samtidige variable. Et godt eksempel på denne tilgang er Crossley & Kim (2022), som undersøger, hvilke af en lang række variable, herunder både lingvistiske variable i teksterne samt ekstra-lingvistiske variable som fx klasstrin og skriveprompt, der bedst kan forudsige tekstkvalitet i en række tekster skrevet af amerikanske college-studerende. Forfatterne viser, at det er de lingvistiske variable (bl.a. kohærens, bestemte syntaktiske konstruktioner og leksikal sofistikation) og ikke klasstrin og prompt, der forudsiger tekstkvalitet bedst. Crossley & Kim (2022) er således et eksempel på et studie, der ikke fordeler de lingvistiske variable relativt til tid, men som betragter klasstrin som en variabel på samme niveau som andre variable, og de kan derfor sammenligne betydningen af klasstrin med de øvrige variable. Et andet væsentligt skrivestudie med fokus på, hvad der bedst forudsiger tekstkvalitet, er McNamara et al. (2010). Forfatterne har her udelukkende fokus på forholdet mellem tekstkvalitet (score baseret på rubric) og en lang række lingvistiske variable forbundet med bl.a. kohæsion, syntaktisk kompleksitet og leksikal diversitet. Studiet viser, at forfeltslængde, et mål for leksikal diversitet og et logaritmisk mål for ordfrekvens var de tre variable, der viste stærkest sammenhæng med tekstkvalitet. I et tilsvarende studie af Olinghouse & Wilson (2013) undersøges forholdet mellem forskellige leksikale konstrukter (fx diversitet, indholdsspecifikke ord og akademiske ord) og tekstkvalitet i tekster inden for tre forskellige genrer skrevet

af elever i 5. klasse. Studiet viser, at genre er en meget vigtig faktor for de leksikale konstrukter, og at det var forskelligt fra genre til genre, hvilke af de leksikale konstrukter der bedst kunne forudsige tekstkvalitet. De tre nævnte studier er således eksempler på en stærk tendens inden for den synkrone skriveforskning, hvor der fokuseres på forholdet mellem forskellige variable (enten lingvistiske eller andre) og tekstkvalitet (baseret på tekstens tildelte karakter eller en rubricscoring) (se også Crossley et al., 2014). Det er således også inden for denne tendens, man finder en del studier med fokus på skriveevaluering og -test (fx Attali & Powers, 2008; Correnti et al., 2020; Crossley & Kyle, 2018; Wilson et al., 2017). Disse studier er kendetegnet ved at fokusere på validering af forskellige test- og evalueringsinstrumenter, og en stor del af studierne har fokus på skriveevaluering på de videregående uddannelser (Zheng & Yu, 2019).

En anden gruppe af synkrone skrivestudier fokuserer på forholdet mellem fag/disciplin og forskellige træk i elevernes tekster. Her er fokus således ikke på, hvad der kendetegner tekster med en høj score, men snarere på hvad der kendetegner tekster skrevet inden for forskellige skolefag eller fagområder. Hardy & Römer (2013) undersøge genrerelaterede og fagspecifikke forskelle i lingvistiske træk i en række tekster skrevet af amerikanske universitetsstuderende. Forfatterne anvender en multidimensionel faktoranalyse på en lang række lingvistiske træk og identificerer på baggrund heraf fire funktionelle faktorer, som optræder i korpusset på tværs af discipliner og genrer. Forfatterne undersøger herefter, i hvilket omfang tekster fra de forskellige discipliner loader på de forskellige faktorer, og de kan herigennem sige noget om forskellene på at skrive inden for de forskellige discipliner. På samme korpus har Römer & Wulff (2010) lavet et lidt mindre studie, hvor de undersøger forskellene i distributionen mellem discipliner af pronomenet 'this'. Forfatterne viser bl.a., at anvendelsen af 'this' som kohæsiionsmarkør er gennemgående på tværs af discipliner, mens der til gengæld er store forskelle mellem discipliner på, hvordan 'this' er integreret, når det bliver anvendt som demonstrativt pronomen.

En tredje gruppe af synkrone skrivestudier fokuserer på forholdet mellem elevtekster og forskellige former for motivationskonstrukter. Et godt eksempel er Wright et al. (2021), som viser en stærk sammenhæng mellem motivation og elevernes skrivepræstationer (baseret på lærerbedømmelser) i et korpus af elevtekster skrevet af amerikanske elever på et niveau svarende til udskoling og gymnasiet. Forfatterne viser, at særligt 'Writer Self-Beliefs' - dvs. elevens tro på egne evner som skribent - er en stærk prædiktor af

skrivepræstationer, og de konkluderer derfor, at et særligt fokuspunkt for skriveundervisningen bør være at udvikle skriveopgaver, som understøtter elevernes motivation og tro på egne evner. Et lignende eksempel på et synkront studie, som bl.a. fokuserer på motivation, er Troia et al. (2013), som i et stort multivariat studie undersøger forholdet mellem skrivemotivation, skriveperformance, klassetrin, køn og lærerens vurdering af elevens generelle skrivefærdigheder. Forfatterne finder en stærk sammenhæng mellem et multidimensionelt motivationskonstrukt, som indeholder items om bl.a. self-efficacy og interesse i skrivning, og kvaliteten af elevernes tekster. Ligeledes finder de en sammenhæng mellem, hvor meget eleverne skriver i og uden for skolen og elevernes motivation. Studiet er således et eksempel på, hvordan et træk ved elevens tekst (i dette tilfælde dens tildelte karakter) relaterer sig til en række variable knyttet til den enkelte elevskriver. Kendeteggende for de fleste studier med fokus på skrivemotivation er, at de netop fokuserer på tekstkvalitet som den afhængige variabel og ikke sproglige træk i teksten.

Foruden ovenstående tendenser er der stor variation i, hvilke uafhængige variable der sammenholdes med elevens tekster. Der findes fx et studie med fokus på forholdet mellem metakognitiv og genremæssig viden på den ene side og sproglige træk i teksten på den anden (Wen & Coker Jr., 2020), et studie med fokus på genetiske og miljømæssige faktoreres betydning for skriveudvikling (Erbeli et al., 2017) samt et studie af latente faktorer (fx stavning, håndskrivning, mundtlighed) betydning for 'compositional writing' (Abbott & Berninger, 1993). Derudover er der en del studier af elevtekster, som fokuserer på forskelle i tekster skrevet af elever med og uden forskellige former for sprogvanskeligheder (Dockrell et al., 2009; Koutsoftas, 2016; Silverman et al., 2015). Der findes også enkelte studier, der – som i denne afhandling – fokuserer på social baggrund (fx Richardson et al., 1976; Weber, 2022) og køn (fx Reilly et al., 2019; Troia et al., 2013). Disse går jeg dog ikke dybere ind i her, men jeg inddrager dem i stedet i fortolkningen af denne afhandlings analyseresultater i kapitel 6.

1.3.3.3 Om dette studies placering i den internationale skriveforskning med kvantitativt fokus

De to tilgange til tekstorienteret, kvantitativ skriveforskning afspejler sig også i denne afhandling, som netop er bygget op omkring en diakron og en synkron analysedel. I de diakrone studier, som jeg har beskrevet ovenfor, er der særligt to forhold, der træder frem.

For det første så vi, at der var tre dominerende analytiske fokuspunkter: ét på produktivitet, ét på leksikale variable og ét på syntaktiske variable. Inden for disse grupper var der stor variation i, hvilke specifikke variable der blev undersøgt, men kendetegnende for de fleste af dem var, at de undersøgte distributionen af disse variable relativt til klassetrin eller uddannelsesniveau. For det andet viste gennemgangen en dominans af studier, der gjorde brug af tværsnitsdata, mens der til gengæld var få longitudinale, kohortebaserede undersøgelser. Denne afhandling placerer sig således i forlængelse af disse tendenser, i den forstand at den diakrone analysedel ligeledes fokuserer på variable inden for de tre beskrevne grupper, og analyserne baserer sig ligeledes på tværsnitsdata, hvor distributionen af de lingvistiske variable undersøges relativt til klassetrin.

I de synkrone studier så vi, at det, der bandt disse studier sammen, var, at de fokuserede på, hvordan forskellige variable (oftest ekstra-lingvistiske) hang sammen med forskellige træk ved elevernes tekster. Oftest var dette træk knyttet til et begreb om tekstkvalitet (dvs. tekstens karakter eller tildelte scoring på en rubric). Der var desuden meget stor variation i, hvilke variables sammenhæng med tekstkvalitet der blev undersøgt. Denne afhandlings synkrone analysedel adskiller sig fra de beskrevne studier på flere måder. Den afhængige variabel i denne afhandling er således ikke tekstkvalitet, men forskellige lingvistiske variable i elevernes tekster. Dette var tilfældet ved meget få af de beskrevne studier. Samtidig er de uafhængige variable i denne afhandling både knyttet til fænomener, som også findes i de beskrevne studier (fx skrivelyst og køn), men også mere ubeskrevne fænomener som fx social baggrund samt læse- og skrivevaner. I den synkrone analysedel undersøger jeg både den direkte betydning af alle de ekstra-lingvistiske variable gennem multiple regressionsmodeller samt den mere specifikke betydning af køn og social baggrund, hvor skrivelyst samt læse- og skrivevaner er kontrolvariable. Der er således ingen af de beskrevne studier, som undersøger synkron skriveudvikling med denne tilgang. Oftest er betydningen af forskellige former for motivation undersøgt direkte, men i denne afhandling undersøges skrivelyst altså både som direkte uafhængig variabel og som kontrolvariabel for køn og social baggrund. Afhandlingen bidrager dermed med ny viden på dette område.

Kapitel 2: Afhandlingens teoretiske forudsætninger og centrale begreber

I dette kapitel præsenterer jeg afhandlingens skriveudviklingsteoretiske forudsætninger samt en række begreber, som er centrale for afhandlingens analyser. Jeg præsenterer ikke de variable, som danner udgangspunkt for analyserne, ligesom jeg heller ikke beskriver de grammatiske begreber og kategorier, som jeg anvender i de kvalitative analyser. Disse vil fremgå af henholdsvis kapitel 4 og kapitel 3. Jeg indleder i stedet dette kapitel med at præsentere forskellige skriveudviklingsmodeller, som danner teoretisk grundlag for afhandlingens to analysedele. Dernæst fokuserer jeg på, hvad der særligt kendetegner denne afhandlings primære datagrundlag, elevteksten. Jeg vil her fokusere på de retoriske dimensioner, som adskiller elevtekster fra andre typer af tekster, og jeg kommenterer på, hvad det har af betydning for, hvordan vi kan forstå og fortolke elevtekster som forskningsdata. Til sidst vil jeg beskrive afhandlingens grundlæggende tekstbegreber.

2.1 Skriveudviklingsmodeller i afhandlingens analyser

Der findes som tidligere anført en række modeller for skriveudvikling, som tager udgangspunkt i vidt forskellige dimensioner af skriveudviklingen. Ifølge Donahue & Lillis (2014) kan disse modeller opdeles i fire dominerende kategorier; tekstorienterede modeller, didaktiske procesmodeller, sociokognitive modeller og modeller af sociale praksisser. Disse modeller er ikke gensidigt udelukkende ontologiske modeller, der beskriver skriveudvikling i al dens kompleksitet, men skal snarere forstås som forskellige fokale optikker på skriveudvikling. I denne afhandling kombineres en tekstorienteret model med både en sociokognitiv model og en social praksis-model for skriveudvikling. Den tekstorienterede model kommer selvfølgelig til udtryk gennem de kvantitative og kvalitative analyser af elevernes tekster, mens den sociale praksismodel og den sociokognitive model kommer til udtryk i afhandlingens anden analysedel, hvor det lingvistiske fokus kombineres med et fokus på en række individuelle baggrundsvariable knyttet til den enkelte elevskriver.

Det er kendetegnende for den sociale praksis-model, at den orienterer sig mod de måder, hvorpå elever engagerer sig i dagligdagens læse- og skriveaktiviteter (Donahue & Lillis, 2014, s. 68). Afhandlingens empiriske materiale orienterer sig således mod disse

aktiviteter, idet jeg undersøger betydningen af læse- og skrivevaner for sproglige træk i elevernes tekster. Jeg undersøger således ikke specifikt de literacybegivenheder, som har formet de tekster, der indgår i afhandlingens tekstkorpus, men antager et bredere perspektiv og ser på de vaner, som eleverne har i fritiden i relation til læsning og skrivning. Samtidig rækker materialet ud over de literacy-orienterede faktorer og inddrager et bredere sociologisk perspektiv på de baggrundsfaktorer, som latent kan tænkes at påvirke de sproglige ressourcer, som eleverne har tilgængelige. Der er derfor tale om et udvidet praksismodel, som ikke er etnografisk orienteret, hvilket ellers kendetegner de fleste studier med fokus på sociale praksisser, men snarere er sociologisk orienteret mod betydningen af elevernes sociale baggrund. Den sociokognitive model kommer specifikt til udtryk igennem variabelen skrivelyst, idet udviklingen af skriftsproglige færdigheder her kobles til den enkelte elevs motivation for at skrive (Donahue & Lillis, 2014, s. 62).

Den tekstorienterede model, som altså er den dominerende i dette studie, har et deskriptivt sigte. Det vil sige, at målet er at beskrive, hvad der kendetegner elevtekster skrevet af forskellige elever på forskellige alderstrin. Andre tekstorienterede skrivestudier har et mere præskriptivt sigte og er interesseret i at identificere, hvad der kendetegner normativt set gode tekster, og på hvilke måder elevtekster lever op til disse normative træk. Den deskriptive model har imidlertid den svaghed, at den ikke i sig selv knytter an til en teori om skriveudvikling, og derfor må de lingvistiske beskrivelser af, hvordan skriftsprog udvikler sig over tid, forstås og fortolkes inden for rammerne af tre typer af teorier. For det første kræver det en teori om, hvad der overhovedet kendetegner skriftsproget som fænomen, og hvad dets grænseflader fx er til det mundtlige sprog eller andre modaliteter. For det andet skal det knyttes an til en teori om, hvordan skriftsprogstilegnelse og -udvikling påvirkes synkront og udfolder sig diakront. For det tredje kræves en pædagogisk og didaktisk teori om, hvordan denne udviklingsdimension udvikler sig og tilrettelægges inden for rammerne af en pædagogisk og didaktisk kontekst. I det følgende vil jeg derfor gøre rede for afhandlingens forståelse af skriftlighed samt for den teori om skriftsproglig udvikling, som studiet baseres på. Jeg vil desuden beskrive de pragmatiske betingelser, som er særlige for elevtekster, og som er vigtige faktorer i fortolkningen af resultaterne af afhandlingens analyser. Til sidst vil jeg også beskrive afhandlingens tekst- og genrebegreber.

2.2 Hvad er skriftsprog, og hvad er skrivning?

Det har historisk set været omdiskuteret i den lingvistiske forskning, hvilken status der skulle tillægges skriftsproget. Som det bl.a. påpeges af Wyse (2017) og Harris (1995, 2001), blev skriftsproget i de tidlige strukturalistiske og poststrukturalistiske kredse i vid udstrækning set som en skreven optegnelse af det talte sprog, og derfor havde skriftsprog ikke en særskilt forskningsmæssig værdi. Dette hang bl.a. sammen med, at skriftsproglige optegnelser var den eneste kilde til diakrone studier af forandringer i det mundtlige sprog. Dette syn på skriftsproget som subsidiært til det mundtlige har imidlertid ændret sig bl.a. som følge af, at kognitivt orienterede psykologer har beskrevet de komplicerede kognitive processer, som karakteriserer det at producere skrevet sprog, og som adskiller skriveprocessen fra det talte sprog. Flower & Hayes (1981) viste bl.a., at skrivning – i højere grad end tale – trækker på en række bevidste monitorerende kognitive processer, som bl.a. knytter sig til det at organisere, evaluere og revidere tekst under hensyn til tekstens retoriske omstændigheder. Desuden begyndte flere antropologer i 1990'erne at interessere sig for skriftsprogets rolle i samfundets sociale praksisser, og disse studier dokumenterede bl.a. skriftsprogets afgørende betydning for menneskets private og professionelle liv. Barton & Hamilton (1998) viser fx, at skrivning på forskellige måder spiller en afgørende rolle i det de kalder 'local literacies' eller 'vernacular literacies', som dækker over alt fra nære familiekontekster til specialiserede arbejdsdomæner, og svenske Karlsson (2006) har ligeledes i et senere stort etnografisk studie vist, at skrivning er et vigtigt element i stort set alle erhverv – også i de håndværks- og serviceerhverv, man normalt ikke forbinder med skrivning. I tillæg har bl.a. Halliday (1989) påpeget, at skriftsproget har en højere grad af indbygget normativitet og konformitet end det mundtlige, og at det skriftlige tillægges en højere kulturel status. Det skriftlige sprog må derfor ses som en kompleks kognitiv proces med markante sociale og kulturelle implikationer, og som derfor er et fænomen med stor vægtning og en selvstændig forskningsmæssig værdi.

Bloomfield (1933) modstiller skriftsproglig fremstilling med billedskrivning (picture-writing), som angår brugen af billedsymboler for at opnå kommunikative formål, og afgrænser skriftsproglig fremstilling eller skrivning (real writing) til den aktivitet, hvorved der kommunikeres ved hjælp af semiotiske ressourcer, der relaterer sig til bestemte sproglige former. Det, der adskiller skriftsproget fra andre modaliteter, er således, at det

specifikt angår sproglige former, som kan oversættes til mundtligt sprog (fx ved at blive læst højt). Som beskrevet i afsnit 1.3.2.2 har den teoretiske tænkning om skrivning imidlertid bevæget sig fra den indledningsvise samtænkning af skriftsproglig produktion (skrivning) og reception (læsning) i literacybegrebet til en udvidelse mod multimodale kommunikationsformer i multiliteracies-begrebet. Spørgsmålet er derfor, hvorfor det i denne afhandling stadig er interessant at se særskilt på skriftsproglig fremstilling i stedet for på multimodal produktion i bred forstand.

Svaret på dette spørgsmål er todelt. For det første er skriftsproget, som jeg også har beskrevet ovenfor, en vigtig forudsætning for deltagelse i et demokratisk samfund, hvad enten det er gennem deltagelse i den offentlige debat eller gennem varetagelse af forskellige jobfunktioner i samfundet. Skriftsproget fylder desuden meget i vores kommunikation, hvilket også understreges af Bremholm et al.'s (2018) analyse af elevprodukter, som viste, at traditionelle skriftsproglige tekster fylder langt mere i elevernes produktive arbejde end multimodale tekster. Med andre ord er den prototypiske skolske tekst fortsat en skriftsproglig tekst. For det andet er multimodale tekster vanskelige at operationalisere og skalere inden for rammerne af et korpusbaseret studie. Dette hænger sammen med, at multimodale kommunikationsformer ikke er konventionaliserede i samme udstrækning som det skriftsproglige - dvs. normerne for fx læserækkefølge og layout er ved multimodal kommunikation ikke lige så stærke som ved skriftsproget og er dermed sværere at strukturere som datasæt (Allwood, 2008) - og det hænger også sammen med, at multimodale tekster er sværere at analysere computationelt, da de modsat skriftsproget, som næsten altid er lineært og kronologisk, kan udfolde sig synkront i tid og rum (Manovich, 2001).

2.3 Hvad er skriftsproglig udvikling?

Som det bl.a. pointeres af Camp (2012), er det omdiskuteret, hvordan skriveudviklingsstudier relaterer sig til psykologiske teorier om udvikling. Nogle korpusbaserede studier af skriveudvikling rammesættes som en form for eksplorative grounded theory-studier, som er datadrevne og ateoretiske, mens andre studier eksplicit trækker på forskellige kognitive og/eller socialpsykologiske udviklingsteorier. Som beskrevet før er problemet med den ateoretiske tilgang, at disse studier i så fald ikke er egentlige studier af udvikling, men blot deskriptive studier af skrivning på forskellige klassetrin. For at der er tale om studier af udvikling, må der enten eksplicit eller latent foreligge en teori om, hvad udvikling er,

og hvordan den observeres og operationaliseres inden for studiets metodiske rammer (Haswell, 2000). Udfordringen er dog, at en elevs skriveudvikling er et komplekst fænomen. Den består på én gang af socialt initierede, men individuelt udførte tekstbehandlinger, som alle er et led i en kompleks sociokognitiv og sproglig proces med stor variation både diakront og synkront. Det er derfor umuligt at rumme alle elementer af denne udviklingsproces inden for rammerne af et enkelt studie. Der er i ethvert studie metodiske begrænsninger, som nødvendigvis påvirker de teoretiske antagelser, der ligger til grund for studiet. Denne afhandlings analyser er derfor funderet i tre typer af udviklingsteori. Den ene teori knytter sig til begrebet modning, den anden til skrivningens sociologiske ulighedsdimension, mens den tredje teori er knyttet til en bredere sociokognitiv forståelse af skrivning og skriveudvikling.

2.3.1 Skriftsproglig udvikling som modningsproces

I den første del af afhandlingens analyser undersøges de sproglige variable relativt til klassetrin. Disse analyser har foruden et deskriptivt sigte også et udviklingsperspektiv. Dette udviklingsperspektiv er på mange måder funderet i begrebet modning. Modning henviser her til de målte karakteristika ved tekster skrevet af elever på kontinuerligt højere klassetrin eller med højere alder. Hunt (1965, 1970) er blandt de første, der adresserer skriveudvikling som en modningsproces, som kan defineres som observerede forandringer over tid eller på tværs af alder/klassetrin. Modning er ifølge Hunt imidlertid ikke knyttet til et begreb om tekstkvalitet eller skrivekompetence. I stedet hviler modningsbegrebet på en simpel og klassisk idé om, at børn og unge bliver i stand til at agere i mere komplicerede skrivesituationer i takt med, at de bliver ældre, og samtidig fungerer modning som en proxy for akkumuleret skriveerfaring i skolen. Med andre ord ligger det latent i modningsbegrebet, at det siger noget om, hvad der kendetegner elevernes tekster i takt med, at de bliver mere erfarne skrivere og (gennemsnitligt set) kognitivt og socialt mere udviklede, men det kan ikke deraf udledes, at modning i en mere normativ forstand er lig med forbedring. Modning forudsætter således ikke nødvendigvis en lineær skriveudvikling.

Modning er på den ene side et effektivt begreb, der kan forklare, hvordan alder/klassetrin kan agere proxy for udvikling. Dette understreges også af Crossley (2020), som peger på, at klassetrin er en af de variable, der bedst forklarer variansen i mange NLP-base-

rede mål. På den anden side er modning også et begreb, der kraftigt reducerer udviklingsbegrebet ved at tildele alder en central kausal forklaringskraft (Hudson, 2009), og spørgsmålet er således, hvad der sker, hvis man tilføjer flere variable til denne ligning knyttet til sociokulturelle faktorer. Dette gøres i anden del af denne afhandlings analyser, hvor der i stedet for den diakrone analyse af de lingvistiske variable relativt til klassetrin fokuseres på de lingvistiske variables relation til en række synkrone variable relateret til social baggrund, læse- og skrivevaner samt skrivelyst. For at beskrive forholdet mellem skriveudvikling og disse variable inddrager jeg to teorier. Den første teori er knyttet til forholdet mellem social baggrund og skriftsproglig formåen. Den anden teori er knyttet til forholdet mellem læse- og skrivevaner, skrivelyst og skriftsproglig formåen. Særligt forholdet mellem skrivelyst (eller skrivemotivation i bredere forstand) og skrivning er forholdsvis velbeskrevet teoretisk (se afsnit 4.2.2.1 for uddybning). Det er imidlertid forholdsvis sjældent, at skriveforskningen eksplicit adresserer relationen mellem social baggrund og skriftsproglige præstationer, og der er derfor heller ikke mange skriveudviklingsteorier, som direkte modellerer denne sociologiske dimension af skriveudviklingen. Bernsteins (1971) sociolingvistiske teori om sproglige koder er dog en af undtagelserne.

2.3.2 Skriftsproglig udvikling i et sociologisk perspektiv

Ifølge Bernstein (1971) afspejler de sociale omstændigheder i opvæksten sig direkte i individets sproglige formåen. Han skelner her mellem begrænset kode ('restricted code') og elaboreret kode ('elaborated code'). Begrænset kode anvendes i nære situationer, hvor deltagerne deler meget forhåndsviden. Det er således kendetegnet ved at være kortfattet, indforstået og ved at referere til en kontekst, som er velkendt for både afsender og modtager. Elaboreret kode er derimod et udfoldet kommunikativt sprog, som ikke forudsætter en stor mængde delt forhåndsviden, men som tværtimod udfolder og beskriver de dele, som er forudsætninger for, at modtageren kan forstå indholdet. Elaboreret kode har derfor en mere kompleks sætningsstruktur og anvender flere forskellige ord. Det anvendes således i kommunikative situationer, hvor der kommunikeres med modtagere, som ikke har faglige eller sociale/relationelle forudsætninger for at forstå indholdet. Ifølge Bernstein kan disse sproglige koder kobles til elevers sociale baggrunde, i den forstand at børn fra de lavere sociale klasser har større sandsynlighed for kun at tilegne sig begrænset kode, mens børn fra de højere sociale klasser har bedre mulighed for at tilegne sig begge

koder. Alle børn tillærer sig den begrænsede kode som en del af socialiseringsprocessen i opvæksten, hvor familien eller den nærmeste tilknytningsgruppe deler viden og kontekst, som barnet kan navigere i ved at bruge forholdsvis simple og indforståede sproglige ressourcer. Elaboreret kode forudsætter til gengæld, at familien har en større grad af social og kulturel mobilitet, som gør, at barnet eksponeres for en større sproglig variation og bliver placeret i situationer, hvor det skal kommunikere med personer med forskellige baggrunde. Dette gælder for såvel skriftsproget som for det mundtlige sprog. De sproglige koder skal ikke forstås som binære kategorier, men som et kontinuum.

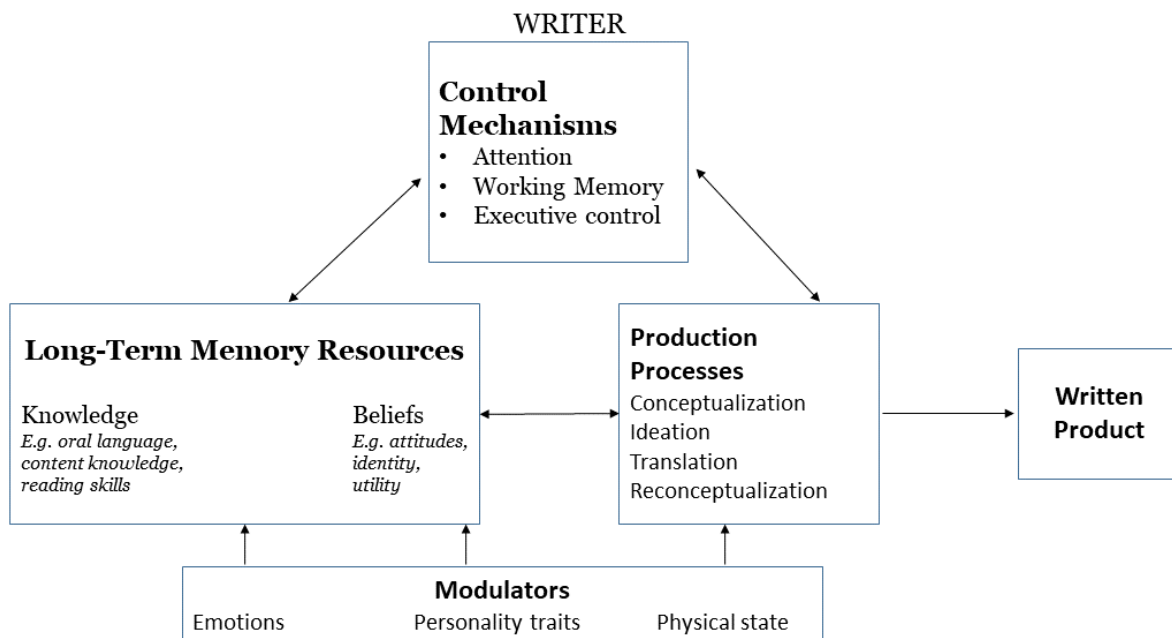
Skriftsproglig udvikling, som det undersøges inden for rammerne af denne afhandling, knyttes altså i denne optik særligt til udviklingen af elaboreret kode. Dette hænger sammen med, at denne kode for en del elever ikke nødvendigvis udvikles i familiekonteksten, men kræver undervisning i skolen. For en del elever er skolen det sted, hvor størstedelen af deres skriveaktiviteter foregår, og for mange elever er skolen det eneste sted, hvor de producerer længere tekster. Det er derfor særligt interessant at undersøge, om denne forskel i sproglige ressourcer, som Bernstein (1971) beskriver, også er synlig, når man undersøger forskellige lingvistiske dimensioner af elevernes tekster. Hvilke forskelle er der med andre ord i de sproglige ressourcer, som er anvendt i elevteksterne, når man undersøger betydningen af social baggrund? Dette undersøger jeg bl.a. i afhandlingens synkron analyse i kapitel 6.

2.3.3 Writers-Within-Community-modellen

Den anden teori, som knytter sig til afhandlingens synkron analyse, modellerer forholdet mellem lingvistiske strukturer i elevernes tekster og deres læse- og skrivevaner samt skrivelyst. Teorien er baseret på Grahams (2018) reviderede 'Writer(s)-Within-Community'-model, som er en kombineret sociokulturel og kognitiv model for skrivning. Modellen består således grundlæggende af to modeller; én som viser, hvilke sociokulturelle komponenter der udgør et skrivefællesskab (writing community), og én der viser de kognitive mekanismer, som er en del af skriveprocessen. Modellerne er ikke adskilte, men indlejrede i hinanden. Modellen er derfor for mangfoldig til, at jeg kan gennemgå alle komponenterne her, så jeg vil i stedet nøjes med at inddrage dem, der er relevante for dette studie. Et skrivefællesskab udgøres ifølge Graham af en række forbundne kontek-

ster, som er indlejrede i hinanden. Pointen er, at en konkret teksthandling sker i en kontekst, hvor skriveren udvikler en tekst ved hjælp af forskellige værktøjer og i samarbejde med andre kollaboratører, herunder læreren. Dette mikrofællesskab er indlejret i et større skrivefællesskab, som bl.a. udgøres af det konkrete fysiske og sociale miljø samt de mere overordnede kommunikative formål, som skrivefællesskabet deler. Denne kontekst er igen indlejret i en række andre fællesskaber, som både er knyttet til de institutionelle rammer og til andre mere uformelle, relaterede skrivefællesskaber. Det er således i disse dialektiske relationer, at elevernes læse- og skrivevaner er interessante i et skriveudviklingsperspektiv. De skrivepraksisser, som eleverne deltager i gennem undervisningen, er således tæt forbundne med de læse- og skrivepraksisser, som eleverne deltager i andre steder. Med Grahams (2018) egne ord: “This means that the functions, practices, and ongoing operation of a writing community are influenced by members’ experiences in other relevant socially derived communities” (s. 264). Derfor kan elevernes skriftsproglige udvikling ikke adskilles fra de læse- og skrivepraksisser, som er knyttet til de skrivefællesskaber, som eleverne er en del af i deres fritid.

Den anden del af Grahams model viser skrivningens kognitive dimension.



Figur 1: Kognitive mekanismer i 'Writers-Within-Community'-modellen. Gengivet ud fra Graham (2018).

Ovenstående model viser, hvordan elevernes kognitive og metakognitive funktioner under skriveprocessen, herunder hukommelse, viden og monitorering, påvirkes og moderes af en række såkaldte modulatorer. Disse modulatorer kalder Graham henholdsvis 'Emotions', 'Personality traits' og 'Physical state'. Her er det særligt kategorien 'Emotions', som er interessant i denne sammenhæng. Emotions henviser nemlig til de affektive aspekter, som er knyttet til den enkelte skrivesituation, og som er med til at påvirke de valg, som en skriver foretager under skriveprocessen. Det er således et aspekt af skriveprocessen som er bundet af situationen og af det enkelte skrivefællesskab. Den affektive dimension har dog også en mere permanent karakter, som Graham kalder 'beliefs', og som er knyttet til en skrifters generelle holdning og værdisættelse af skrivning som aktivitet. Beliefs er vigtige komponenter i skriveudviklingen, fordi det påvirker "whether one engages in writing, how much effort is committed, what actions and tools are applied, and how the writer interacts with collaborators or other members of the writing community" (Graham, 2018, s. 266). Skrivelyst, som måles i denne afhandling, er knyttet til beliefs og ikke emotions, selvom de to dimensioner selvfølgelig påvirker hinanden. Skrivelyst er således en mere permanent sociokognitiv faktor i elevernes skriveudvikling, som går på tværs af de enkelte skrivesituationer.

Jeg har nu beskrevet de tre grundlæggende modeller, som udgør det skriveudviklingsteoretiske grundlag for afhandlingens to analysedele. Modningsteorien er tæt forbundet med den diakrone analysedel i kapitel 5, mens Bernsteins (1971) teori om sociale koder og Grahams (2018) 'Writer(s)-Within-Community'-model i sammenhæng udgør det skriveudviklingsteoretiske grundlag for afhandlingens synkron analyse. Jeg vil derfor vende tilbage til disse teorier i fortolkningen af resultaterne fra de to analysedele. Vi vender nu blikket mod et andet vigtigt teoretisk aspekt af denne afhandling, nemlig hvad der kendetegner elevteksters pragmatik og funktion og dermed de muligheder og begrænsninger, der er forbundet med at fortolke elevtekster som forskningsdata.

2.4 Elevtekstens pragmatik: Om teksthandlinger i skolske kontekster

Skrivning i skolen adskiller sig fra skrivning uden for skolen på flere punkter. Dette skyldes først og fremmest, at skolens tekster er skrevet i en særlig skolsk kontekst, som grundlæggende ændrer på de pragmatiske og retoriske betingelser for kommunikation. Tøgeby

(2015) kalder skolen for et 'øverum', hvor lærer og elev har bestemte asymmetriske statusroller, som bl.a. gør, at læreren både udsteder den skriveordre, som eleverne skriver teksten på baggrund af, og samtidig er modtager af selvsamme tekst med den primære interesse i teksten, at den kan danne grundlag for bedømmelse, hvad enten det er med formativt eller summativt sigte. Dette gør, at skolens praksis kan ses som en række ritualiserede skriftsprogspraksisser, hvor eleven som skribent på den ene side skriver til en virkelig modtager (læreren) og samtidig skriver til en fiktiv eller simuleret læser (fx en fiktiv avislæser hvis de skriver en nyhedsartikel). Ifølge Berge (1988) medfører dette, at skolestilen sjældent er en egentlig kommunikativ handling, men snarere en rituel eller strategisk handling, hvor en vigtig pragmatisk forpligtelse som fx oprigtighed (at eleven mener, at det han/hun skriver, er sandt) træder ud af kraft. En skolestil er således i højere grad knyttet til metafunktionelle succeskriterier, fx læring eller opnåelse af en god karakter, end egentlige kommunikative succeskriterier (Berthelsen, 2021).

2.4.1 Elevteksten i en Griceansk optik

At skolske tekster fungerer efter andre retoriske betingelser end andre tekster, bliver endnu tydeligere, hvis man ser på skrivning i skolen gennem Grices (1989) fire maksimer for kommunikation⁶. Ifølge Grice kan den del af sproglige ytringers mening, som ikke alene kan etableres ud fra den sproglige struktur, men ud fra konteksten, forstås ud fra det, han kalder samarbejdsprincippet. Princippet opdeles i fire maksimer – henholdsvis kvantitet, kvalitet, relation og måde. Kvantitetsmaksimet henviser til, at afsenderen forpligter sig på at give så præcise informationer som muligt og derved hverken give flere eller færre informationer, end det er nødvendigt. Forudsætningen for, at kvantitetsmaksimet fungerer, er, at der er en autentisk kommunikationssituation, hvor afsenderen har et reelt ønske om at viderebringe informationer til en modtager, som enten har brug for dem eller ønsker at modtage dem. I skolestile er dette maksime forvrænget. Den fiktive eller simulerede læser, som er bygget ind i teksten, har ingen funktionel interesse i teksten, og maksimet opretholdes derfor ikke som en funktion af den interpersonelle relation mellem afsender og modtager, men derimod kun i den udstrækning, at læreren pålægger

⁶ Grices maksimer er oprindeligt påtænkt mundtlig kommunikation (deraf *konversationel* implikatur), men jeg følger bl.a. Wyatt & Nunn (2019) og anvender maksimerne til også at beskrive betingelserne for skriftlig kommunikation.

afsenderen et metafunktionelt ansvar for, at teksten potentielt vil kunne fungere, hvis den blev anvendt i en forestillet kommunikativ situation. Kvantitetsmaksimet forvrænges også af et andet klassisk skolsk træk, nemlig at de fleste skriveordrer i skolen indeholder minimumskrav til længden af elevernes tekster, og det er derfor et sandsynligt scenarie, at det i højere grad er dette formelle krav (eller andre didaktiske betingelser som fx tid eller graden af stilladsering), som styrer mængden af informationer i elevteksten, end det er kommunikative hensyn.

Det andet maksime er kvalitet og angår kvaliteten og sandhedsværdien af en ytring. Dette indebærer, at man ikke ytrer, hvad man selv mener er falsk, eller udtaler sig om noget, om hvilket man ikke har tilstrækkeligt belæg. Beder man fx eleverne om at skrive en tekst om, hvad de har lavet i deres sommerferie, pålægger man dem på den ene side kvalitetsmaksimet, fordi de skal beskrive virkelige hændelser, men eleverne er på den anden side ikke mere forpligtet på maksimet, end at andre kriterier kan vægte højere, og de kan derfor også opleve at blive belønnet for at bryde maksimet, hvis det bidrager til en fra lærerens perspektiv normativt set bedre tekst (fx ved at tilføje opdigtede dramatiske situationer til sommerferiebeskrivelsen). En smule anderledes ser det imidlertid ud med relationsmaksimet. Med relation henviser Grice (1989) til det forhold, at afsenderen af en ytring forpligter sig på at være relevant og kun nævne forhold, som giver mening ift. kontekst og budskab. Dette maksime vil i det fleste tilfælde også gøre sig gældende for elevtekster, eftersom denne type af relevans og indholdsmæssig konsistens dels er direkte forbundet med tekstens indholdsmæssige kohærens og mere konkrete sproglige kohæ-sion, og dels ligger det som et implicit kriterie i skriveordren, at elevens besvarelse holder sig inden for det angivne emne. Det samme gælder for det sidste maksime, måde, som henviser til det forhold, at afsenderen forpligter sig på at udtrykke sig klart og forståeligt. Dette maksime er selvfølgelig fortsat gældende ved elevtekster og er meget direkte relateret til elevernes skrivekompetence og dermed et vigtigt element af tekstens kvalitet.

2.4.2 Skolens metagenrer og deres rolle i danskfaget

Ovenstående gennemgang af de fire maksimer viser, at selvom skriveordrer kan simulere situationer, hvor alle fire maksimer træder i kraft, kan andre skolske metafunktionelle kriterier tidsidesætte eller underkende maksimerne (særligt maksimerne kvantitet og kvalitet), fordi der er tale om simulerede kommunikationssituationer (scenarier), og der

kan derfor forekomme situationer, hvor eleverne – modsat kommunikationssituationer uden for skolen – belønnes for at bryde maksimerne. Dette hænger sammen med, at skrivning i skolen er tæt knyttet til en række fagspecifikke traditioner, som ifølge Carter (2007) manifesterer sig i det, han kalder for *metagenrer*. Metagenrerne angiver forskellige måder at udvise faglig gøren og viden på gennem tekster. Carter samler dem i fire metagenrer, som relaterer sig til forskellige faglige situationer i skolen, som kræver en teksthandling fra eleven. De fire metagenrer kalder han for henholdsvis 'problem-solving', 'empirical inquiry', 'research from sources' og 'performance'. Den første handler om at identificere og komme med løsningsforslag på virkelige problemer og er en velkendt genre fra bl.a. projektarbejde. Den anden genre handler om at undersøge hypoteser og spørgsmål på baggrund af empiriske data og findes i skolen i fx biologi- eller fysik/kemi-rapporter. Den tredje metagenre handler om at undersøge og diskutere fænomener baseret på andre tekstkilder og findes i skolen i bl.a. historie og til dels i danskfaget. I den sidste metagenre, performance, er sproget og tekstens komposition et mål i sig selv og kendes særligt fra sprogfagene (herunder danskfaget), hvor det er tekstens sproglige strukturer snarere end tekstens indhold, som danner grundlag for at vurdere tekstens kvalitet. Disse metagenrer manifesterer sig på forskellige måder i skolen, men de deler det indbyggede træk, at de relaterer sig til faglige traditioner i skolen og dermed også til de kriterier for tekstkvalitet, som gør sig gældende inden for den pågældende tradition.

I danskfaget i grundskolen er det særligt de to sidste metagenrer, som gør sig gældende. Dette kommer bl.a. til udtryk i danskfagets policydokumenter, som historisk set har været præget af et teknisk og formelt sprogligt fokus (Tannert & Berthelsen, 2021) samt en forholdsvis lineær forståelse af skriveudvikling, der bevæger henover skoleårene fra en erfaringsbaseret mod en mere analytisk tilgang til skrivning (Jeffery et al., 2018). Afgangsprøven i skriftlig fremstilling er ifølge Krogh (2014) imidlertid også domineret af teksttyper, som kræver, at eleverne tager udgangspunkt i hverdagsviden samt personlige erfaringer og holdninger. Dette indikerer således, at performance-metagenren også præger danskfaget i udskoling, eftersom disse teksttyper ikke har andet funktionelt formål end at udgøre et grundlag, hvor eleven kan demonstrere de lingvistiske ressourcer (herunder stilistiske og genrespecifikke ressourcer), de har tilgængelige, gennem tekstproduktion. Troelsen (2018) pointerer imidlertid i en analyse af den nyeste afgangsprøveform, som er kendt som iPrøven, at eleverne nu kan overskride deres afhængighed af

hverdagserfaringer, fordi det ved visse skriveopgaver nu kræves, at eleverne søger informationer på nettet. Hun pointerer derfor, at iPrøven kan ses som en bevægelse væk fra det tidligere fokus på fiktionsskrivning og hverdagsberetninger mod andre former for fagtagenrer, som bl.a. beror på elevernes informationsøgning. Troelsens analyse indikerer således, at metagenren 'research from sources' også er blevet en vigtig del af danskfagets sigte, og det er således særligt disse to metagenrer, der er med til at definere de kriterier og normer, som gør sig gældende for skrivning i danskfaget.

2.4.3 Elevers skriftsprog som intersprog

At det snarere er de fagspecifikke normer for tekstkvalitet inden for metagenrerne, der har betydning for, om en elevtekst er vellykket eller ej, end det er traditionelle kommunikative succeskriterier, har desuden betydning for måden, vi forstår elevtekster på som forskningsdata. Hvor korpusbaserede studier af fx opslag på de sociale medier eller professionelle nyhedsartikler er med til at sige noget om, hvad der kendetegner naturligt opståede registre og diskurser, som de udfolder og forandrer sig i henhold til bestemte kommunikative kontekster, er et korpus af elevtekster i højere grad et udtryk for et sprog i udvikling, der er produceret inden for en skolsk ramme, som på én gang muliggør og begrænser den sproglige udfoldelse (fx ved at opstille formelle og metafunktionelle krav i skriveordren). Det lingvistiske output i elevtekster repræsenterer således ikke det, Sinclair (1996) kalder "the genuine communications of people going about their normal business" (s. 7), men er snarere et udtryk for en kommunikativ dobbelthed, hvor teksten på én gang skal simulere autentisk kommunikation og samtidig være en øvelse som et led i elevens skriveudvikling og udgøre grundlag for bedømmelse. Dette betyder, at det kan være vanskeligt at koble sproglige træk i elevernes tekster til deres kommunikative funktion. I stedet kan elevteksterne ses som skriftsproglige øvelser, dvs. som bevidste, gentagende tekstaktiviteter med henblik på udvikling af skriftsproglige færdigheder (DeKeyser, 2007, s. 1). At teksterne er udtryk for bevidste handlinger, der udtrykker en elevs forsøg på at mestre bestemte normer, betyder også, at elevernes tekster kan ses som en form for intersprog.

Selinker (1972), som var den første til at anvende intersprogsbegrebet, definerede begrebet som "a separate linguistic system based on the observable output which results

from a learner's attempted production of a TL [target language] norm” (s. 214). Et intersprog er således et systematisk sprog, som afviger fra den intenderede norm, men som samtidig udtrykker en elevs bevidste forsøg på at producere et sprogligt output, der lever op til selvsamme norm. Som det pointeres af Bartholomae (1980), er der klare overlap mellem intersprogstilegnelse på andetsproget og skoleelevers udvikling af skriftsproglige færdigheder på modersmålet. Modsat talesproget, som er forholdsvis veludviklet og kommunikativt set funktionelt i en ret tidlig alder, er skriftsproget langt mere fejlbehæftet og kommunikativt begrænset, men disse normafvigelser er imidlertid systematiske og afspejler mange af de samme karakteristika, som kendetegner et intersprog, fx hypoteseafprøvning og påvirkning fra talesproget. Det er derfor muligt at se elevernes skriftsprog som udtryk for systematiske, interlingvistiske relationer mellem et intersprog og et veludviklet og kommunikativt funktionelt normsprog. Fordelen ved at anskue elevtekster på denne måde er, at det giver mulighed for at betragte elevernes tekster som systematiske forsøg på at levere sproglige former, som anvendes af mere erfarne og dygtigere skrivere. Elevtekster er derfor ikke bare afvigelser fra normsproget, men udtryk for elevens selvstændige systematiske sprog, som derfor kan anskues som en form for performancedata. Det vil sige, at det er bevidste teksthåndlinger, som er et udtryk for kvalitativt forskellige færdighedsniveauer. På denne måde er elevernes tekster udtryk for de sproglige ressourcer, som den pågældende elev har til rådighed. Jeg vender løbende tilbage til, hvordan udviklingstræk i elevernes tekster kan forstås som en form for intersprog i fortolkningen af analyseresultaterne i kapitel 5.

2.5 Afhandlingens tekstbegreber

De to tekster, som eleverne har skrevet til dette korpus, kalder jeg for henholdsvis en 'beskrivende tekst' og en 'berettende tekst'. Disse to tekstbegreber minder således om de basisgenrer, som i skandinavisk kontekst kendes fra genrepædagogikken og den systemisk funktionelle lingvistik (se bl.a. Holmberg, 2019; Mulvad, 2009). Basisgenrerne er et udtryk for teksttyper med tydelige sproglige genrenormer og er derfor tæt knyttet til de fremstillingsformer, hvormed de er realiseret i teksten (Martin & Rose, 2008). Der er imidlertid ikke tale om genrer i gængs retorisk/pragmatisk forstand, hvor det er de funktionelle og ekstra-lingvistiske succeskriterier, der konstituerer genrerne (se fx Swales, 1990). Beskrivende og berettende elevtekster er, hvis man følger denne genretradition, vanskelige

KAPITEL 2

at genrebestemme, eftersom de kommunikative succeskriterier for skrivning i skolen som udgangspunkt er metafunktionelle, jf. ovenstående afsnit om elevtekstens pragmatik. Biber & Conrad (2009) anlægger i forlængelse heraf en mere empirisk og lingvistisk tilgang til genrebegrebet, som derfor er meget udbredt inden for korpuslingvistikken. Biber & Conrad anvender genrebegrebet næsten ækvivalent med registerbegrebet, men angiver samtidig, at den centrale forskel mellem et register og en genre er, at genren bestemmes ved de konventionelle strukturer, som går på tværs af registre, mens registret i højere grad er knyttet til de partikulære lingvistiske træk i den enkelte kommunikationssituation. En genre omgiver således som udgangspunkt en hel tekst, mens registre kan identificeres og knyttes til forskellige enkeltdele i en tekst⁷. Udfordringen med Biber & Conrads tilgang i relation til denne afhandling er imidlertid, at de anser genrer som ” governed by specific conventions, generally recognized by members of a culture, and so the genre itself is named within the culture” (s. 34). I denne afhandling har eleverne skrevet berettende og beskrivende tekster, og selvom både beretninger og beskrivelser er kulturelt genkendelige former for tekst, er de ikke konventionelt etablerede genrer, men snarere et udtryk for fremstillingsformer, som kan anvendes til at realisere forskellige former for tekst, herunder konventionaliserede genrer.

I stedet for at anse teksterne som genrer følger jeg derfor Holmberg (2006) og Nordenfors (2017) og anser beskrivende og berettende tekster som relateret til forskellige tekstaktiviteter knyttet til tekstens forskellige fremstillingsformer snarere end til genrebegrebet. Med begrebet tekstaktivitet understreges, at selvom tekstproduktion i skolen er orienteret mod bestemte udtryksbehov og metafunktionelle situationer, er der fortsat stor variation i, hvordan teksterne struktureres, eller hvilket register de realiseres i. Jeg har derfor ingen forventning om, at elevteksterne i korpusset ser ud på bestemte måder eller følger bestemte normer. Jeg forstår derimod de to slags tekster i korpusset som tekster tilhørende hver sin tekstaktivitet, som lægger op til anvendelsen af bestemte fremstillingsformer, som eleverne i sidste ende kan realisere på forskellige måder. Fremstillingsformer er således ifølge Togeby (2014) ikke nødvendigvis en egenskab ved en hel tekst,

⁷ Biber & Conrad (2009: 33) anerkender dog, at der kan være tilfælde, hvor genrer indlejres i hinanden inden for den samme tekst. Fx kan man betragte en forskningsartikel som en genre, men samtidig kan indledningen i selv samme artikel betragtes som en genre. Genrer er dermed også defineret ud fra ens analytiske fokus.

men et træk ved dele af teksten. Der kan derfor godt være berettende tekster i denne afhandlings korpus, som både indeholder berettende og beskrivende fremstillingsformer, ligesom der kan være beskrivende tekster, som både trækker på en beskrivende, berettende og argumenterende fremstillingsform.

Som betegnelse for den tekst, som er resultatet af en bestemt tekstaktivitet, anvender jeg begrebet teksttype, som er en forholdsvis teorineutral betegnelse for den gruppe af tekster, som er et resultat af bestemte tekstaktiviteter. Når jeg taler om henholdsvis berettende og beskrivende tekster er der således tale om to teksttyper relateret til forskellige tekstaktiviteter i klasserummet, som er initieret af skrivepromptets udformning og formet af de didaktiske, situationelle vilkår, teksten er produceret under. Anvendelsen af disse tekstbegreber knyttet til fremstillingsformerne er desuden i tråd med den internationale skriveforskning, hvor betegnelser som 'narrative writing', 'argumentative writing' og 'expository writing' er standardbetegnelser. Lidt mere skematisk kan afhandlingens tekstbegreber beskrives som i tabel 1.

Tabel 1*Oversigt over afhandlingens tekstbegreber*

Tekstaktivitet	En tekstaktivitet henviser til den proces, hvor en elev producerer en tekst ud fra en prototypisk forestilling om, hvad der normativt set bør udgøre en sådan tekst. Tekstaktiviteten er knyttet til grundlæggende fremstillingsformer, fx forklaringer, beretninger, instruktioner etc. (Nordenfors, 2017). ⁸ Tekstaktiviteten er igangsat af en skriveordre eller et skriveprompt.
Teksttype	En teksttype er en samlebetegnelse for de konkrete tekster, som er resultatet af de enkelte tekstaktiviteter. Berettende skrivning er fx en tekstaktivitet, som resulterer i teksttypen berettende tekster.
Fremstillingsform	Fremstillingsformerne henviser til den prototypiske lingvistiske realisering af teksttyperne. Teksttyperne kan dog godt have en vis variation i fremstillingsformerne (Togeby, 2014).
Elevtekst	Elevteksten er den enkelte realiserede tekst, som en elev har produceret. Elevteksten er altså et konkret eksemplar af en teksttype og et udtryk for en elevs forsøg på at besvare skriveordren eller skrivepromptet.

2.5.1 Den beskrivende teksttype

Beskrivende tekster er i en vis forstand en underbelyst del af skriveforskningen. Dette hænger sammen med, at den beskrivende fremstillingsform er et grundelement i de fleste former for skrivning og derfor sjældent behandles separat. Det er fx en velintegreret del af mange berettende tekster, særligt de fiktive af slagsen, at forskellige former for beskrivelser (fx naturbeskrivelser) indgår, ligesom fagtekster også indeholder forskellige beskrivelser af faktuelle forhold. Ifølge Coker Jr. (2012) er det netop denne udbredelse af den beskrivende fremstillingsform, der bør give den en helt central placering i studier af skriveudvikling i skolen. Foruden det forhold at beskrivende tekster er en af de dominerende

⁸ Det er uklart, hvorvidt Nordenfors (2017) anvender begrebet tekstaktivitet om både skriveproces og elevens færdige tekst. Jeg har imidlertid valgt at skelne mellem tekstaktivitet (selve den handling at skrive teksten) og teksttype (de konkrete tekster som er et resultat af tekstaktiviteten).

teksttyper i indskolingen og derfor allerede en central teksttype i skolen, er beskrivende skrivning ifølge Coker Jr. (2012) også interessant, fordi denne type tekstaktivitet stiller store krav til eleven, som både skal udvælge og beskrive de informationer, som er relevante i henhold til skriveordren, og samtidig beskrive disse informationer på en måde, der både er sprogligt præcis og indholdsmæssigt dækkende. Så selvom beskrivelse typisk forbindes med en aktivitet på et lavt taksonomisk niveau og som en dimension primært knyttet til begynder-skrivningen, er det også noget, der bør undersøges på de senere klassetrin. Schleppegrell (1998) viser fx i en analyse af beskrivende elevtekster fra 7.-8. klassetrin, at teksterne bl.a. er kendetegnet ved at anvende en høj grad af komplekse og udfoldede nominalfraser, og hun argumenterer for, at forskellige grammatiske træk (fx anvendelsen af demonstrative pronomener som 'this' og 'that'), hænger sammen med modtagerbevidsthed i teksterne. Beskrivende skrivning genererer således ikke nødvendigvis simple tekster, men giver mulighed for forskellige former for sproglig kompleksitet, og der må derfor forventes en vis variation i de beskrivende tekster, som indgår i afhandlingens korpus.

2.5.2 Den berettende teksttype

Berettende tekster er et etableret felt i skriveforskningen (Crossley, 2020), og som det bl.a. påpeges af Hall-Mills & Apel (2012), har skoleelever typisk mere erfaring og et større kendskab til berettende tekster end til andre teksttyper. Dette understreges også af Christensen et al. (2014), som viser, at det at skrive tekster med udgangspunkt i egne oplevelser, erfaringer og holdninger ikke bare er et kendetegn ved begynder-skrivningen, men tværtimod strækker sig helt op på udskolingsniveau. Derfor præsterer skoleelever også typisk meget stabilt inden for denne teksttype – særligt når beretningerne tager udgangspunkt i deres egne oplevelser (Bouwer et al., 2015). Berettende tekster er således en oplagt tekstaktivitet at undersøge i et studie af skriveudvikling, fordi beretningerne dels kan tænkes at give eleverne mulighed for at skrive mere frit, idet de er bekendte med teksttypen, dels er en meget udbredt tekstaktivitet i grundskolen og derfor ligner de tekstaktiviteter, som eleverne normalt skriver i relation til i skolen, hvilket øger studiets økologiske validitet. Med andre ord kan man sige, at berettende tekster er en slags kritisk case (Flyvbjerg, 2006), fordi eleverne netop her – måske i højere grad end andre teksttyper - får mulighed for at udfolde forskellige former for sproglig kompleksitet. Beard & Burrell (2010) viser

KAPITEL 2

fx, at berettende tekster giver eleverne mulighed for at anvende flere forskellige stilistiske og sproglige træk som gentagelser, læserhenvendelser, aktive handleverber og dialog, ligesom teksterne stiller store krav til elevernes evne til at skabe kohærens og kohæsion i teksten, fordi den narrative fortællestruktur ofte medfører en mere kompleks informationsstruktur.

Kompleksiteten i elevernes berettende tekster er også belyst kvantitativt i en række studier. Nelson & van Meter (2007) peger fx på, at kvantitative mål som bl.a. tekstlængde, antal unikke ord (types) og ortografisk korrekthed er velegnede til at skelne mellem forskellige kvalitetsniveauer i narrativ elevskrivning, hvilket bekræftes af Troia et al. (2019), som også viser, at bl.a. ordlængde hænger sammen med aldersmæssig udvikling. Dette indikerer således, at længdebaserede mål af forskellige sproglige træk i teksterne kan være med til at opfange forskellige dele af den skriftsproglige udvikling i berettende elevtekster.

Kapitel 3: Forskningsdesign og metode

I dette kapitel beskriver jeg projektets forskningsdesign og de enkelte metoder, som er anvendt i projektet. Jeg indleder kapitlet med en overordnet beskrivelse af, hvordan projektets metoder er integreret med hinanden. Herefter følger en beskrivelse af afhandlingens datagrundlag, herunder udformningen af tekstkorpuset. Efterfølgende beskriver jeg de to analysestrategier, som udgør grundlaget for afhandlingens to analysedele. Beskrivelsen af analysestrategierne afsluttes med en diskussion af strategierne i relation til relevante kvalitetskriterier for forskning. I kapitlets to afsluttende afsnit beskriver og diskuterer jeg først en række videnskabsteoretiske perspektiver og efterfølgende en række forskningsetiske problemstillinger, som rejser sig på baggrund af projektets forskningsdesign og metodiske forankring.

3.1 Metodisk rammesætning

Dette studie er et mixed methods-studie, som falder i tre analysedele: en kvantitativ, deskriptiv analysedel med fokus elevteksternes diakrone udvikling på tværs af klassetrin, en kvalitativ analysedel med fokus på leksikogrammatisk analyse af tre tekstudvalg af elevtekster fra korpuset samt en kvantitativ analysedel bestående af multiple lineære regressionsanalyser med fokus på sammenhængen mellem teksternes sproglige egenskaber og metadata knyttet til den enkelte elevskriver. At jeg betegner studiet som et mixed methods-studie, skyldes ikke alene tilstedeværelsen af både kvantitative og kvalitative metoder, men i lige så høj grad at metoderne er integreret med hinanden og tilbyder forskellige metodiske perspektiver på elevens skriveudvikling (Cresswell & Clark, 2011).

Integrationen er tydeligst mellem analysens første og anden del, idet de kvalitative, leksikogrammatisk analyse bygger videre på de sproglige variable fra den kvantitative, diakrone analyse i første del. De kvalitative analyser skal således ses som en konkretisering og kontekstualisering af de kvantitative, sproglige analyser fra første del. Der er således tale om det, Frederiksen (2013, s. 22-23) kalder henholdsvis data- og metodeintegration. Med dataintegration skal forstås, at det i en vis forstand er samme datamateriale, der ligger til grund for begge analyser, dog med den væsentlige forskel at de kvantitative analyser er gennemført på hele tekstkorpuset, mens de kvalitative analyser er gennem-

ført på tre udvalg af elevtekster fra samme korpus. Der er desuden tale om metodeintegration i den forstand, at det er resultaterne fra første kvantitative analyse, som dels danner udgangspunkt for udvælgelsen af tekster til tekstudvalgene (udddybes i afsnit 3.3.2), dels bidrager til leksikogrammatisk fokuspunkter i de kvalitative analyser (udddybes i afsnit 3.3.2.2). De to metoder er således sekventielt forbundne (Cresswell & Clark, 2011). Selvom der på sin vis er tale om to analysedele med forskellige metodiske tilgange, vil jeg i afhandlingen behandle dem som en samlet analysestrategi, som jeg kalder en diakron analysestrategi (udddybes i afsnit 3.3.2). Dette indebærer således, at begge analysedele fremgår af samme kapitel (kapitel 5), hvor de dog vil være blandet sammen, sådan så de kvalitative analyser fremgår i forlængelse af kvantitative analyser af de sproglige variable, som de er knyttet til.

Den tredje analysedel, hvor jeg fokuserer på de sproglige variables sammenhæng med de ekstra-lingvistiske variable knyttet til den enkelte elev, udgør en selvstændig analysedel i afhandlingen. Dette betyder imidlertid ikke, at denne analysedel ikke hænger sammen med de to andre. Der er således i én forstand tale om data- og analyseintegration imellem den tredje del og de to andre dele forstået på den måde, at analyserne samler sig om den samme overordnede forskningsinteresse; nemlig at beskrive skriftsproglig udvikling i danskfaget fra 5.- 9. klassetrin. I en anden forstand er der imidlertid ikke tale om analyseintegration, men snarere om data- og fortolkningsintegration. Dataintegrationen ses selvfølgelig ved, at det er det samme tekstkorpus, som anvendes i alle tre analysedele, og fortolkningsintegrationen kommer til udtryk ved, at resultaterne fra alle tre analysedele samles og fortolkes i sammenhæng i kapitel 7. Der er således samlet set tale om det, Howe (2012) kalder 'conjunctive triangulation', som viser sig ved, at selvom metoderne er forskellige og har forskellige roller i forskningsdesignet, så knytter de sig samlet set til det samme genstandsfelt. Derfor er formålet ifølge Howe (2012), at man gennem tringuleringen skal "accommodate ostensibly discordant data by bringing it under a more comprehensive explanatory framework" (s. 90). Det er således den skriveudviklingsteoretiske rammesætning af både analyserne og den efterfølgende fortolkning, som er afgørende for, at de enkelte metoder bringes i dialog med hinanden (se også afsnit 3.4.3 om computationel hermeneutik).

3.2 Datamateriale og korpusdesign

Datamaterialet i dette studie baserer sig på to datakilder. Det drejer sig først og fremmest om de indsamlede elevtekster, som udgør studiets tekstkorpus, og så drejer det sig subsidiært om data fra et spørgeskema, som eleverne har besvaret i forbindelse med, at jeg indsamlede teksterne i de respektive klasser. Datamaterialet er samlet ind i 16 klasser fordelt over fem klassetrin på én skole (se overvejelser over dette i afsnit 3.2.2.3). Der er indhentet samtykke fra forældrene til alle deltagende elever (se bilag 5). Studiet baserer sig således på datamateriale fra i alt 239 elever. Datamaterialet er indsamlet i september/oktober 2020. Alle studiets lingvistiske variable er knyttet til elevteksterne, mens de ekstra-lingvistiske variable er dannet på baggrund af data fra spørgeskemaet. Teksterne er skrevet på baggrund af et forskerinitieret skriveprompt, som præsenteres i flere detaljer i næstkommende afsnit (3.2.1).

Tabel 2

Overblik over datamateriale

	Deltagende klasser	Elever på årgangen	Deltagende elever/ procent	Beskrivende tekster	Berettende tekster	Survey – antal svar
5. årg.	2	53	43 / 81%	43	39	41
6. årg.	4	93	41 / 44%	40	37	41
7. årg.	3	69	40 / 58%	40	40	40
8. årg.	4	86	72 / 84%	69	70	68
9. årg.	3	63	44 / 70%	44	42	43
Total /gns.	16	364	239 / 66%	236	228	233

Som det fremgår af ovenstående tabel, varierer deltagelsesprocenten relativt meget mellem klassetrin. Jeg har ikke data, som kan sige noget om årsagerne til dette, så jeg må her bero mig på de oplevelser, jeg havde i forbindelse med dataindsamlingen. For det første er det vigtigt at gøre opmærksom på, at dataene er indsamlet under coronapandemien, som på mange måder har påvirket skolernes praksis. Dataene er indsamlet i en periode, hvor skolen netop var åbnet igen for fysisk fremmøde efter en langvarig nedlukning, og de fleste lærere og elever havde derfor travlt med at finde sig til rette igen i skolens omgivelser. Det er ikke usandsynligt, at dette har påvirket villigheden til at deltage i ekstra

projekter, som forstyrrede den pædagogiske praksis, som lærere og elever i fællesskab var ved at genetablere. Samtidig har indsamlingen af tekster krævet individuelt samtykke fra elevernes forældre, hvilket ofte er et vilkår, der af forskellige praktiske årsager fører til lavere deltagelsesprocent (Pokorny et al., 2001).

3.2.1 Skriveprompt

Elevernes tekster er baseret på et forskerinitieret skriveprompt. Jeg anvender her begrebet skriveprompt frem for skriveordre eller skriveopgave for at understrege, at der ikke er tale om en almindelig skriveopgave udviklet af læreren med henblik på en bestemt undervisningskontekst i relation til en bestemt klasse, men om en eksternt defineret opgave, som de deltagende lærere efter bedste evne har integreret i undervisningen. Skriveordre og særligt skriveopgave er begreber, som tilhører læreren og den daglige undervisningskontekst, mens begrebet skriveprompt understreger det forhold, at opgaven er udarbejdet af en person uden for skolens kontekst med henblik på forskning. Skrivepromptet i dette studie fremgår af bilag 3.

3.2.1.1 Skrivepromptets udformning

Skrivepromptets udformning har afgørende betydning for de tekster, som eleverne skriver. Dette hænger bl.a. sammen med det, der kaldes 'linguistic priming', som handler om, at bestemte sproglige træk i opgaveteksten kan påvirke de sproglige træk i elevernes tekster. Crossley (2020) peger bl.a. på, at syntaktisk komplekse skriveprompts også kan generere gennemsnitligt set mere syntaktisk komplekse elevtekster, og at nogle elever faktisk mimer sproglige strukturer fra opgaveteksten i deres egne tekster. Dette er imidlertid en effekt ved skriveprompts, som ikke er mulig at fjerne, men i større korpusser, hvor teksterne er baseret på mange forskellige prompts, bør der ifølge Crossley (2020) kontrolleres for denne primingeffekt. Eftersom der kun anvendes et enkelt skriveprompt i dette studie, har det ikke på samme måde været relevant her. I stedet har jeg, som jeg uddyber nedenfor, valgt en kort minimalistisk opgavetekst for bl.a. at undgå priming fra skrivepromptet.

Skrivepromptet i dette studie har overskriften "Sundhedskrise i verden" og består af to opgaver. Begge opgaver relaterer sig dermed til det samme emne, hvilket er vigtigt, eftersom emnevariation inden for en opgave kan påvirke de sproglige træk, som eleverne producerer (Huot, 1990). I den første opgave skulle eleverne vælge mellem to billeder og

lave en beskrivelse af det valgte billede. Begge billeder relaterede sig på forskellige måder til promptets overskrift. Eftersom elevernes valg af billede må formodes at have en effekt på de lingvistiske variable, kontrolleres der for dette i regressionsmodellerne i de synkroner analyser. I den anden opgave skulle eleverne skrive en berettende tekst om en typisk dag for dem derhjemme under coronanedlukningen.

Skrivepromptet er således udformet ud fra to grundlæggende kriterier. For det første har det været vigtigt at vælge et emne, som alle elever har nogenlunde lige forudsætninger for at skrive om. Hvis ikke der går læsemateriale eller indholdsspecifik undervisning forud for promptet, skal promptet lægge op til indhold inden for elevernes erfaringsverden (Kroll & Reid, 1994). Derfor valgte jeg at relatere opgaverne til coronapandemien, da dette på tidspunktet for dataindsamlingen var et presserende emne, som alle elever havde personlige erfaringer med. Ved at undgå bias i emnevalget sikres et bedre sammenligningsgrundlag mellem eleverne, i den forstand at der ikke er bestemte elevgrupper, som har en særlig forhåndsviden, der giver dem en fordel ift. at løse opgaverne i skrivepromptet (Tedick, 1990). For det andet har det været et vilkår, at opgaveteksten i begge opgaver skulle være forholdsvis simpel for at sikre, at alle elever fra 5.-9. klassetrin skulle kunne forstå og løse opgaven. Derfor har jeg valgt en minimalistisk tilgang til opgaveteksten og i stedet bedt lærerne om at hjælpe de elever, der havde brug for det, med at kontekstualisere og stilladsere opgaven. Data fra spørgeskemaet, hvor jeg spurgte eleverne om, hvordan de oplevede sværhedsgraden af skrivepromptet, viser, at 53% fandt opgaven ”Hverken let eller svær”, mens 40% fandt den enten ”Let” eller ”Meget let”. Kun 7% oplevede således, at opgaven var enten ”Svær” eller ”Meget svær”. Det tyder således ikke på, at eleverne i hvert fald ikke selv har oplevet den minimalistiske tilgang som en blokering eller udfordring for deres skriveproces.

3.2.1.2 Skrivepromptet som intervention

Promptet skrivning kan ses som en form for intervention. Dette hænger i dette studie sammen med, at skriveprompten dels i sig selv er en forskerinitieret forstyrrelse af den daglige undervisning, dels at eleverne har besvaret skrivepromptet som en del af deres daglige danskundervisning. Der er imidlertid ikke tale om en intervention i snæver forstand, hvor en nøje beskrevet forandring implementeres og gennemføres i praksis, men snarere om det Letiche et al. (2008) kalder ’weak intervention’. Dette indebærer i dette

METODE OG FORSKNINGSDSIGN

tilfælde, at forskeren kommer med et lille input (i dette tilfælde et skriveprompt), som lærerne derefter selv skal integrere i undervisningen på eget grundlag. Jeg har således ikke fastsat de didaktiske rammevilkår for skriveprocessen (fx skrivetid, stilladsering eller andre skrivedidaktiske tiltag), men i stedet bedt lærerne om at integrere skrivepromptet i undervisningen, som var det en hvilken som helt anden skriveopgave, de arbejder med (se brev til lærerne i bilag 6). Dette indebærer også, at lærerne har skullet tage de samme hensyn til enkeltelever, som de normalt tager i undervisningen (fx ved at lade ordblinde elever bruge hjælpeværktøjer eller skriveskabeloner).

Der findes både faglige og praktiske årsager til denne tilgang. Først og fremmest er den manglende forskerstyring med skriveprocessen et udtryk for, hvad der var praktisk muligt at lave en aftale om med den deltagende skole og de involverede lærere på daværende tidspunkt. Tidspunktet for dataindsamlingen var som tidligere nævnt præget af, at skolen netop var åbnet igen efter en lang nedlukning, så både skolens og lærernes fokus var på at genetablere en normal skolehverdag for eleverne, og det var derfor ikke muligt at gennemføre en større forskerstyret intervention i de 16 klasser. Der er imidlertid også forskningsmæssige fordele ved en fri skriveproces, der ikke er styret af forskeren. For det første er det sandsynligt, at en fri skriveproces uden øvre tidsbegrænsning resulterer i mere komplekse tekster. Havde jeg fx bedt lærerne om, at eleverne skulle skrive inden for en fastlagt tidsramme, ville eleverne ikke have haft samme tid til idéudvikling, planlægning og revision af teksten. Dette kan ifølge Crossley (2020, s. 433) resultere i tekster, som er mindre kohærente og mindre komplekse både syntaktisk og leksikalt. For det andet kan den manglende forskerstyring af skriveprocessen ses som en styrkelse af skrivepromptets økologiske validitet. Blikstad-Balas et al. (2018) viser fx i et stort deskriptivt studie af skriveundervisningen i norskfaget, at der er stor variation mellem klasser og lærere i, hvordan der arbejdes didaktisk med skrivning. Ved ikke at fjerne og ensarte denne variation nærmer man sig dermed det, som Bronfenbrenner (1976, s. 5) kalder 'real-life educational settings'. Det vil sige, at teksterne i korpusset er skrevet under forhold, som ligner dem, eleverne normalt producerer tekster under.

3.2.2 Korpusdesign

Tekstkorpusset er designet som et diakront korpus, dvs. det har en indbygget tidsdimension. Dette betyder, at det er muligt at undersøge sproglige forandringer og udviklingstræk på tværs af tid (Davidse & De Smet, 2020). Datamaterialet er dog ikke baseret på et longitudinalt studie (fx en kohorte) af de deltagende elever (paneldata), men derimod på tværsnitsdata fra et sample af elever grupperet efter klassetrin. Tidsdimensionen er derfor baseret på elevernes årgangsvise tilhørsforhold, hvilket er en udbredt tilgang i megen uddannelsesrelateret korpusforskning (Crossley, 2020). Sammenlignet med mange af de nyere såkaldte megakorpusser er tekstkorpusset i denne afhandling relativt lille. Dette hænger selvfølgelig sammen med, at det er vanskeligt at scrape elevtekster automatisk og derfor kræver manuel dataindsamling, hvilket betyder, at det sjældent er muligt at opnå repræsentativitet baseret på et stort randomiseret sample. I stedet er det væsentligste for analysernes validitet, at korpusset ud fra kvalitative kriterier er så repræsentativt som muligt, dvs. at teksterne i korpusset afspejler populationen i højest mulig grad (se diskussion af dette i afsnit 3.2.2.3). Desuden er det vigtigt, at man ved små korpusser kun undersøger forholdsvis højfrekvente lingvistiske træk, dvs. træk som er typiske inden for et bestemt register, og som derfor med al sandsynlighed vil findes mange gange i korpusset. Ellers vil man komme til at over- eller underestimere frekvensen og dermed skabe en analytisk bias i korpusset (Biber, 1993; Koester, 2010). Eftersom korpusset i denne afhandling ikke er stort⁹, undersøger jeg derfor omnipræsente lingvistiske træk (dvs. træk som altid er til stede i en tekst) i elevernes tekster. Det gælder for alle de lingvistiske variable i denne afhandling (med undtagelse af ordvariationsindekset OVIX), at de er til stede og kan måles i alle tekster upåagtet tekstens længde eller kvalitet.

3.2.2.1 Præprocessering af teksterne

Før elevernes tekster er blevet gjort til genstand for computationelle analyser, har de gennemgået en række formateringsprocesser, som gør det muligt at hente information ud fra teksten ved hjælp af automatiserede og computerbaserede analyseværktøjer. For computeren består en tekst nemlig som udgangspunkt af én lang streng af meningsløse tegn (herunder blanktegn/mellemrum). For at denne streng skal give mening, skal der lægges

⁹ Sammenlignet med en del af de studier, jeg har beskrevet i forskningsoversigten i afsnit 1.3, er korpusset dog ret gennemsnitligt af størrelse, hvis man vurderer på antallet af deltagende elever.

en struktur ned over den, der deler strengen op i meningsfulde enheder, og som tagger disse enheder med sproglig information (Jurafsky & Martin, 2009, s. 35-43). Denne proces er i dette studie gjort ved hjælp af sprogpakken DaCy til Python, som er den bedste sprogpakke på dansk (Enevoldsen et al., 2021). Præprocessering med DaCy indebærer, at de sproglige strukturer i elevernes tekster annoteres med forskellige sproglige metadata. Dette indebærer bl.a. en såkaldt tokenization (opsplitning i ord), POS-tagging (tildeling af ordklasser til hvert ord), lemmatization (fjernelse af endelser) og en dependency parsing. Det er således analyser af disse annoterede, sproglige metadata, som danner grundlag for en del af de sproglige variable i dette studie.

3.2.2.2 Standardisering af elevsprog

En væsentlig udfordring ved computationelle analyser af elevsprog er, at teksterne indeholder en del ortografiske og syntaktiske fejl. Disse fejl kan selvfølgelig være interessante at studere i et pædagogisk perspektiv, fordi de kan sige noget om, hvordan elever udvikler og eksperimenterer med forskellige sproglige strukturer med henblik på at nå til et niveau, hvor de følger gængse normer for dansk skriftsprog. Fejlene udgør imidlertid også et validitetsproblem for de sproglige variable, eftersom normafvigelser fra dansk retskrivning og syntaks fører til målefejl i de automatiserede analyser. Et eksempel på dette er, at computeren segmenterer sætninger ved punktum, og hvis ikke disse er placeret korrekt i elevens tekst, vil en variabel som fx gennemsnitlig sætningslængde have store målefejl. Det samme gælder de leksikale mål, hvis teksterne har mange ortografiske eller morfologiske fejl. I sådanne tilfælde vil de digitale værktøjer ikke kunne præprocessere de enkelte leksikale enheder og fx tage ordklasser korrekt, hvorfor mål som OVIX eller gennemsnitlig dependensafstand fx vil blive upræcise (Cobb & Horst, 2015). Det er således afgørende, at disse fejl korrigeres, eller som Ledolter & VanderVeldes (2021) formulerer det: ”The quality of one’s analytical results may be directly related to the cleaned quality of the texts” (s. 37).

Standardiseringen af elevernes tekster er således i bund og grund et spørgsmål om at gøre teksterne læsbare for computeren. Målet har i denne forbindelse været at bibeholde elevernes syntaktiske og leksikale valg, så det kun er de forhold, som sænker validiteten af de sproglige mål, som er blevet rettet. Dermed ligger der underforstået en idé om, at syntaktisk og leksikal kompleksitet er latente kvaliteter i elevernes tekster, som er

til stede, uanset om eleverne kan sætte korrekte punktummer eller stave ordene korrekt. Jeg er således ikke interesseret i at måle elevernes normafvigelser, men i at undersøge de syntaktiske og leksikale kvaliteter i elevernes tekster, som de fremtræder latent i teksten.

Der findes ikke fastlagte standarder for, hvordan tekstkorporer standardiseres (især ikke tekstkorporer med elevsprog), og særligt tegnsætning er en underbelyst, men stor udfordring, idet tegnsætning er afgørende for, hvilke syntaktiske enheder der kan måles (Ondelli, 2018). Jeg har derfor udviklet de principper for standardisering, som er anvendt på elevteksterne i dette studies tekstkorpus, ud fra det pejlemærke, at der skal rettes så lidt som muligt så grundigt som muligt (Ledolter & VanderVelde, 2021, s. 41). Principperne har været følgende:

1. Alle punktummer i teksterne er rettet, så de følger reglerne for slutpunktum i Retskrivningsordbogen (Dansk Sprognævn, 2012). Dette indebærer, at slutpunktummer adskiller syntaktiske perioder (se afsnit 4.1.3.1 for definition), og derfor kan der forekomme flere helsætninger inden for én periode, hvis der er anvendt en sideordnende konjunktion. Syntaktiske perioder med overdrevent brug af paratakse er derfor bibeholdt (fx ”jeg stod op og så spiste jeg og så gik jeg i skole og så tog jeg hjem”). Det vil sige, at der er ikke fjernet et ’og’ og indsat et punktum i stedet. Jeg har desuden fjernet alle punktummer, som ikke anvendes til at adskille syntaktiske perioder (fx ved klokkeslætsangivelser og ved forkortelser). Spørgsmålstejn og udråbstegn er bibeholdt, da de på linje med punktum tæller som store skilletejn. Sætningsfragmenter (herunder ufærdige sætninger som fx ”jeg hen skolen”) er adskilt, så de udgør selvstændige perioder, med mindre de meningsfuldt hænger sammen med den foregående eller efterfølgende sætning (fx ”Da jeg havde spist morgenmad gik jeg ovenpå til min computer som stykker sidste uge”).
2. Alle ortografiske (herunder sammensatte ord) og morfologiske fejl er rettet, så de følger normerne for dansk retskrivning.
3. Der er ikke tilføjet ord til teksten (fx ved manglende verbaler som ”det godt”), ligesom der heller ikke er fjernet overskydende ord (fx ved angivelse af to på hinanden efterfølgende verbaler [jeg løb gik nedenunder] eller gentagelser [der er er mange i skole]).

4. Ordrette gengivelser (direkte citater) af opgaveformuleringen fra skrivepromptet er fjernet. Parafraseringer er bibeholdt (fx "I denne opgave skal jeg beskrive, hvad der er på billedet").

3.2.2.3 Repræsentativitet i korpusdesignet

Et centralt spørgsmål i alt korpusforskning er spørgsmålet om repræsentativitet. Som det fremgår af Koplenig (2016, s. 17-39), er der imidlertid mange forskellige tilgange til dette spørgsmål, og disse tilgange varierer relativt til forskningsformål og den forskningsgenstand, som man vil undersøge gennem korpusset. Det er derfor overvejelser over målpopulationen, som er det første skridt i overvejelserne over et korpus' repræsentativitet (Biber, 1993). Der er fx forskel på, om man ønsker, at de sproglige data i et korpus skal generaliseres til en sproglig population (fx det danske sprog) eller til en population af mennesker (fx alle udskolings elever i den danske folkeskole). Korpusbaserede studier af skriveudvikling har imidlertid den indbyggede dobbelthed, at de på én gang er deskriptive, sproglige studier af skriftsprog, som det anvendes inden for en bestemt skolsk situationskontekst, og samtidig er de studier af bestemte elevers evne til at producere skriftsprog (Tannert, 2021). Der er således to spørgsmål, som skal besvares, før man kan sige noget om et korpus' repræsentativitet ift. at studere skriveudvikling. Der er dels det epistemologiske spørgsmål angående, i hvilken udstrækning korpuslingvistiske data kan bruges som empirisk belæg for en persons evne til at producere skriftsprog (skrivekompetence), og dels spørgsmålet om i hvilken grad de korpuslingvistiske data kan generaliseres til en større population.

Det første spørgsmål relaterer sig til en gammel diskussion, som har rødder i 1960'ernes generative grammatik og Chomskys (1965) berømte distinktion mellem 'e-language' og 'i-language', dvs. mellem eksternaliseret og anvendt sprogbrug på den ene side og individets evne til at producere sprog på den anden. Chomsky (1986) kritiserede korpuslingvistikken for ukritisk at drage slutninger fra 'e-language' til 'i-language', som Chomsky ikke nødvendigvis mente, der var hverken teoretisk eller empirisk belæg for at drage. Chomskys kritik har haft en lang virkningshistorie i korpuslingvistikken, men som det fremgår af bl.a. Koplenig (2016), har Chomsky mødt en del modkritik for at opstille en dikotomi mellem den mentale sprogkompetence på den ene side og det observerbare sprog på den anden. Det ligger uden for rammerne af denne afhandling at gå ind i en

længere teoretisk undersøgelse og diskussion af dette forhold, selvom jeg kortvarigt vender tilbage til det i afsnit 3.4.2. I stedet følger jeg bl.a. Leech (1992), som peger på, at der i korpuslingvistiske kredse generelt er konsensus om, at individuel kompetence og sproglig performance er forbundet med hinanden – om end ikke i et 1:1-forhold. Med dette skal forstås, at en elevtekst eller en række elevtekster selvfølgelig ikke kan være et udtryk for en persons samlede kapacitet til at producere skriftsprog i alle tænkelige kontekster, ligesom skriveudvikling også er andet og mere end det, der kan observeres i elevernes tekster. Korpusbaserede studier af elevtekster siger således noget meget specifikt om, hvordan skriveudvikling ser ud inden for en bestemt skolsk kontekst og er påvirket af en række kontekstuelle faktorer knyttet til den enkelte tekstaktivitet, herunder skrivepromptet, teksttypen og de omkringliggende didaktiske vilkår.

Det andet spørgsmål om repræsentativitet angår generaliserbarheden af de indsamlede data. Et traditionelt kriterium for dette vil indebære, at de sproglige data er baseret på en stor, randomiseret stikprøve. Dette er imidlertid et vanskeligt kriterium at efterleve, eftersom det vil kræve nogenlunde fri adgang til landets skoler og alle elevernes tekster. Som alternativ til randomiseringskriteriet foreslår Atkins et al. (1992) derfor, at målet for korpuslingvistikken er at bygge balancerede korpusser. Et korpus er således balanceret, såfremt det udgør "a manageably small scale model of the linguistic material which the corpus builders wish to study" (Atkins et al., 1992, s. 14). Her handler det altså om, at et korpus skal designes sådan, at det ud fra kvalitative kriterier kan siges at opfange al den sproglige variation, som man forventer vil være til stede i målpopulationen, herunder alle genrer og teksttyper (Biber, 1993). Korpusset i dette studie er imidlertid hverken repræsentativt i den gængse forstand med randomisering og sample-størrelse som de vigtigste parametre eller i den alternative forstand med balancering som det vigtigste parameter. Korpusset baserer sig på tekster fra fem årgange på én skole, og det er kun tekster fra de elever, som aktivt har valgt at deltage via samtykke, som er inkluderet i korpusset. Der er således risiko for, at korpusset har en bias ift. elevernes faglige og sociale baggrund. Fordelen ved dette korpus er imidlertid, at denne bias ikke er helt 'black box', da jeg har metadata om eksempelvis elevernes sociale baggrund, der kan være med til at tage højde for nogle af disse forskelle.

METODE OG FORSKNINGSDSIGN

I stedet for et repræsentativt tekstkorpus baseret på enten randomisering eller balancerings er der således tale om et casestudie. Casestudiet forbindes primært med kvalitative studier, men som det pointeres af bl.a. Yin (2018), er casestudiet ikke en metode, men en forskningsstrategi, hvilket betyder, at casestudiet ikke i sig selv udelukker anvendelsen af kvantitative metoder. Det, der i stedet kan ses som et samlende kendetegn ved casestudiet, er, at der anvendes flere forskellige datakilder til at undersøge en specifik case, og at der ikke – modsat ved fx eksperimentel forskning – kontrolleres for kontekst-effekter, som i stedet anses som en vigtig del af forskningsgenstanden (Mills et al., 2010). Blandingen af kvalitative og kvantitative metoder i casestudiet kan ses som et 'embedded case study', hvor der er flere forskellige analysegenstande og metoder på spil (Scholz & Tietje, 2002), og hvor resultaterne fra de forskellige analysedele kan anvendes til gensidigt at belyse hinanden (Creamer, 2018).

I denne afhandling er der tale om et casestudie i mere end én forstand. Den skole, jeg har indsamlet dataene på, kan ses som en case. De klasser, der deltager, kan som cases enten hvert for sig eller samlet, og endelig kan de enkelte elevtekster, jeg analyserer, ses som cases. Denne afgrænsningsproblematik er imidlertid velbeskrevet i metodelitteraturen (se fx Yin, 2018). I relation til spørgsmålet om repræsentativitet er det først og fremmest relevant at se på skolen som en samlet case. Den deltagende skole er en stor folkeskole (+800 elever) fra en mellemstor dansk provinsby. Det særlige ved denne skole er, at den har en forholdsvis skæv fordeling, når det kommer til elevernes sociale og sproglige baggrund. Omkring 73,5% af de deltagende elever har mindst én forælder, som tilhører en af de to øverste grupper i den internationale klassificeringsramme for erhversgrupper, ESeC (se tabel 7 i afsnit 4.2.1.1 for en mere detaljeret beskrivelse af fordelingen af social baggrund). Det betyder, at næsten 3/4 af eleverne har mindst én forælder med en videregående uddannelse. Til sammenligning er det kun omkring 46% af den danske befolkning mellem 25-69 år, der har en videregående uddannelse (Danmarks Statistik, 2022). Vurderet på social baggrund kan casen derfor ses som en kritisk case, idet man kan sige, at i det omfang social baggrund har betydning for elevernes tekster, vil man i høj grad forvente at finde en sådan betydning i denne case (Flyvbjerg, 2006).

Andelen af elever i studiet, som taler et andet sprog derhjemme end dansk, ligger på 16%. Dette er baseret på data fra spørgeskemaet. Til sammenligning viser data fra seneste

PISA-undersøgelse, at omkring 10,7% af de deltagende elever i PISA har indvandrerbaggrund, hvoraf de 8,4% er født i Danmark (Christensen, 2019, s. 13). Selvom tallene i denne afhandling og i PISA ikke dækker over helt det samme, indikerer det dog, at andelen af tosprogede i afhandlingens case er lidt højere. Ser man lidt nærmere på de 16% af eleverne i dette studie, som taler et andet sprog derhjemme, tegner der sig imidlertid et mere nuanceret billede. Ud af de 16%, som udgøres af samlet set 38 elever, angiver 15 elever engelsk som det sprog, de også taler hjemme, og 19 har valgt kategorien "Andet". Af øvrige valgmuligheder var vietnamesisk, arabisk, persisk, polsk, punjabi, serbo-kroatisk/bosnisk, somali og tyrkisk. Her har kun en enkelt valgt vietnamesisk, én har valgt persisk og én har valgt polsk. Enten er der altså en forholdsvis stor overrepræsentativitet i samlet af engelsk som første- eller andetsprog, eller også har en del elever misforstået spørgsmålet og talt deres engelsklektier med, selvom jeg under dataindsamlingen gjorde eleverne opmærksomme på, at engelsklektier ikke var relevant for spørgsmålet. Jeg har ligeledes spurgt de 38 elever, hvor gamle de var, da de begyndte at lære dansk. Her angiver 26 af eleverne, at de begyndte at lære dansk inden for de første tre leveår, hvilket antyder, at de enten er født i Danmark eller er kommet hertil som meget små. De er således sandsynligvis det, man vil kalde 'simultaneous bilinguals', dvs. de har lært deres første- og andetsprog samtidigt fra fødslen. Kun 6 af eleverne (svarende til 2,6% af de elever, der svarede på spørgeskemaet) angiver, at de begyndte at lære dansk, efter de var fyldt 4 år. Det er altså en forholdsvis lille andel, som er det, man kalder 'sequential bilinguals', som dækker over dem, der har lært et andetsprog efter de har udviklet en vis fluency i førstesproget (Ray-Subramanian, 2011). På dette parameter er casen derfor en atypisk case, og derfor indgår elevernes sproglige baggrund primært i regressionsmodellerne som en kontrolvariabel.

3.2.3 Surveydesign

Spørgeskemaet i dette studie er designet som et digitalt online-survey konstrueret og distribueret gennem SurveyXact. Spørgeskemaet består af i alt 35 items. Spørgeskemaet er designet sådan, at generelle spørgsmål om køn og klassetrin kommer først, sværere spørgsmål om læse-/skrivevaner, skrivelyst og forældrenes beskæftigelse (som bl.a. trækker på elevernes hukommelse og andre tungere kognitive processer) i midten og lettere

personlige spørgsmål om bl.a. oplevet sværhedsgrad af skrivepromptet til sidst. Dette design kan ifølge Cohen et al. (2018, s. 492-493) være med til at sikre, at respondenterne bevarer interesse og koncentration gennem den tid, det tager at besvare spørgeskemaet.

Eleverne besvarede spørgeskemaet i forbindelse med, at jeg som forsker besøgte hver enkelt deltagende klasse for at indsamle elevteksterne. At jeg var til stede i lokalet, mens eleverne besvarede spørgeskemaet gav mig mulighed for at adressere og kompensere for misforståelser af items hos respondenterne. Jeg fik fx mulighed for at give en fælles mundtlig rammesætning af spørgeskemaet til hele klassen, inden de gik i gang (hvor jeg bl.a. understregede, at læse- og skrivevanerne angik deres fritid og ikke skoletid), ligesom jeg havde mulighed for at hjælpe de elever undervejs, som ikke forstod spørgsmålene. Omvendt kan min tilstedeværelse under besvarelsen også være med til at skabe en bias hos respondenterne, som indebærer, at eleverne gennem deres svar gerne vil fremstå på en bestemt måde, som anses som mere rigtig inden for en bestemt faglig kontekst. I dette tilfælde kunne det fx være, at eleverne overestimerer deres skrivelyst eller frekvensen af deres læse- og skrivevaner for at fremstå som mere kompetente elever inden for en danskfaglig kontekst, hvor læsning og skrivning forbindes med anerkendelses- og efterstræbelsesværdige faglige normer. Denne effekt kaldes bl.a. for 'social desirability' og angår det forhold, at respondenter kan have en tilbøjelighed til både at fremstå mere konsistente i deres svar, end de egentlig er, og samtidig være mere tilbøjelige til at tilpasse sig de omkringliggende normative strukturer for at fremstå på en bestemt måde (Blasius & Thiessen, 2012). For at kompensere for dette validitetsproblem gjorde jeg eleverne opmærksomme på, at deres svar ville blive behandlet anonymt af mig som forsker (og ikke af deres lærer), og at svarene ikke ville kunne spores tilbage til dem, lige så snart jeg havde oprettet tekstkorpuset med de annoterede metadata fra spørgeskemaet.

3.3 Analysestrategier

I dette afsnit vil jeg beskrive afhandlingens to analysestrategier, henholdsvis den diakrone og den synkrone analysestrategi. Før de to analysestrategier præsenteres, vil jeg kort beskrive de analyseværktøjer, som jeg har anvendt til at lave de sproglige analyser i studiet, eftersom disse er udgangspunktet for begge analysestrategier. Afsnittet rundes af med overvejelser over forskellige former for validitets- og reliabilitetsbegrænsninger i de to analysestrategier.

3.3.1 Natural Language Processing (NLP)

NLP henviser til brugen af digitale, computationelle værktøjer til automatisk analyse af skrevet sprog (Eisenstein, 2019, s. 1). NLP er således baseret på teorier om, hvilke enheder der udgør sprog på forskellige grammatiske niveauer, og det er denne viden, som NLP-værktøjerne bruger til at segmentere og annotere tekst. Dette baserer sig på forskellige lingvistiske, computationelle og matematiske modeller, som bidrager til, at NLP-værktøjerne analyserer sproget så præcist og korrekt som muligt (Jurafsky & Martin, 2009, s. 38-40). Som det pointeres af Kyle (2021), er det imidlertid også disse modeller, som er NLP-værktøjernes svaghed, især når de anvendes til at studere sprog og tekster skrevet af elever, som er på et forholdsvis tidligt stadie af deres skriveudvikling. Mange sprogmodeller til NLP er nemlig baseret på korpusser bestående af professionelle og redigerede tekster, hvilket betyder, at værktøjerne kan have svært ved at genkende afvigende sproglige former, som man ofte finder i elevtekster. NLP-værktøjer er således ikke fejlfri værktøjer. Forskellige grammatiske analyser er altid mulige (også ved manuelt annoterede korpusser), men ved automatiserede analyser gælder det ofte, at de er baseret på probabilistiske modeller og dermed annoterer de enkelte sproglige elementer ud fra, hvad der er mest sandsynligt baseret på de data, modellen er trænet på. Der må derfor forventes en vis grad af målefejl i de kvantitative sproganalyser. Jeg har dog i denne afhandling anvendt den bedst præsterende danske sprogmodel (DaCy-Large), som har vist en høj nøjagtighed, når den er blevet afprøvet på forskellige parametre (Enevoldsen et al., 2021). DaCy-modellen er imidlertid ikke trænet på elevsprog, hvilket sandsynligvis betyder en lidt højere fejlprocent, når den anvendes i korpusset i denne afhandling. De konkrete sproglige analyser er foretaget ved hjælp af Python-pakken TextDescriptives (Hansen, u.å.).

3.3.2 Den diakrone analysestrategi

Den diakrone analysestrategi har til formål at afdække elevernes skriftsproglige udvikling i et diakront perspektiv. Det vil sige, at analyserne undersøger, hvordan de lingvistiske variable varierer henover korpussets tekster relativt til klassetrin. Selvom datasættet som tidligere nævnt består af tværsnitsdata, undersøger jeg imidlertid fortsat datasættet diakront. Den diakrone analyses formål kan forstås ækvivalent med Parris (2010) ord: "Treating the large cross-sectional sample in a longitudinal way allows an enquiry about the

relative rate of progress across year levels” (s. 136). En væsentlig begrænsning ved en del lingvistisk skriveforskning, som netop ser på udviklingstrin henover klassetrin, er imidlertid, at de primært fokuserer på gennemsnitlige forandringer over tid. Dette forudsætter implicit, som det bl.a. pointeres af Bundsgaard et al. (2022), en faglig homogenitet på hvert klassetrin, som selvfølgelig ikke eksisterer i praksis. Jeg anlægger derfor i den dia-krone analysedel en bredere tilgang og ser ikke bare på udviklinger i centraltendenser, men også på den samlede spredning og fordeling af værdier inden for en årgang. Samtidig inddrager jeg tekster i de kvalitative analyser med både meget høje og lave værdier. Dette gør jeg netop for at undersøge den brede udvikling i den samlede variation for hver årgang og ikke kun den typiske og gennemsnitlige udvikling for hver årgang,

Analysestrategien foldes ud i tre dele, som hver består af både kvantitative og kvalitative analyser. De tre dele er struktureret ud fra de tre grupper af lingvistiske variable (produktivitet, leksis og syntaks), som jeg præsenterer senere i kapitel 4. Hver del indledes med en kvantitativ, deskriptiv analyse af de enkelte lingvistiske variable. Der er her fokus på, hvordan de lingvistiske variable udvikler sig kvartilsvist, dvs. i forhold til den samlede fordeling relativt til kvartiler og median. Dernæst følger en række kvalitative analyser af tekstudvalg bestående af elevtekster fra korpusset. Hver analysedel fokuserer på forskellige lingvistiske dimensioner, som jeg vil gøre rede for senere i afsnit 3.3.2.2. Der er tale om tre tekstudvalg, som består af tekster, der er udvalgt på baggrund af to kriterier. For det første er teksterne valgt ud fra repræsentationskriterie, hvilket indebærer, at jeg er interesseret i at undersøge tekster, som repræsenterer så bredt et spektrum af numeriske værdier og klassetrin som muligt. Jeg er altså både interesseret i tekster med en lav værdi i den pågældende variabel, tekster omkring medianværdien og tekster med høj værdi, og jeg er desuden interesseret i at undersøge og sammenligne tekster med samme numeriske værdier skrevet af elever fra forskellige klassetrin. For det andet er teksterne udvalgt ud fra et relevanskriterium, som indebærer, at de udvalgte tekster skal bidrage med relevant viden ift. forskningsspørgsmålene. Udvælgelsen af tekster er således en kvalitativ, iterativ proces. Det varierer derfor også en smule, hvor mange tekster der indgår i hvert tekstudvalg.

Kombinationen af kvantitative og kvalitative analysemetoder i denne analysestrategi kan ses som en kombination af henholdsvis distant reading, hvor de kvantitative analyser viser tendenser og generelle udviklingstræk i hele tekstkorpusset, og close reading,

hvor de generelle udviklingstræk uddybes og sættes i kontekst af mere detaljerede lingvistiske beskrivelser. Der er således tale om en form for zoom in/zoom out-tilgang eller med Shneidermans (1996) kendte mantra: "Overview first, zoom and filter, then details-on-demand" (s. 337). I den diakrone analyse forsøger jeg således at koble teksternes makrostrukturelle niveau, som er repræsenteret af de numeriske data om tekstkorpuset, som NLP-analyserne genererer, og tekstens mikrostrukturelle niveau, som kommer til syne gennem de kvalitative, leksikogrammatisk analyser.

3.3.2.1 Relevante statistiske deskriptorer

De kvantitative analyser i den diakrone analysedel er i udgangspunkt deskriptive. Det betyder, at jeg fokuserer på at præsentere fordelingen af værdier for hver enkelt variabel og sammenligne disse på tværs af klassetrin og relativt til teksttype. Der er således ingen analyser af relationen mellem de forskellige lingvistiske variable. Fordelingen for hver variabel visualiseres gennem et boksplot, som består af fem grundlæggende tal; minimumsværdien, første kvartil (25. percentil), anden kvartil (medianen), tredje kvartil (75. percentil) og maksimumsværdien¹⁰. Derudover er spredningen i dataene for hvert klassetrin angivet som interkvartilsbredde (IQR). Dermed giver boksplottet mulighed for at se udviklinger i fordelingen af data og dermed også, om der sker udviklinger i forskellige grupper af tekster. Med andre ord bliver det bl.a. muligt at se, om der primært sker udviklinger i tekster med høje variabelværdier, eller om der er tilsvarende udvikling i tekster med lave variabelværdier. Boksplottet giver desuden en visualisering af spredningen (IQR) inden for datapunkterne i en variabel. IQR defineres som afstanden fra den 25. til den 75. percentil – dvs. den angiver de midterste 50% af værdierne. IQR siger altså noget om, hvor tæt de midterste værdier ligger på hinanden og kan således bruges som indikator på, hvor meget eller lidt de typiske tekster på årgangen ligner hinanden ift. den pågældende variabel (Dekking et al., 2005).

3.3.2.2 Metoder til kvalitativ, grammatisk analyse

De kvalitative, sproglige analyser følger op på de deskriptive, kvantitative analyser. Formålet med disse analyser er todelt. For det første er formålet med de kvalitative analyser

¹⁰ Outliers vises i boksplottet som prikker over eller under det højeste eller laveste datapunkt. En outlier defineres i et boksplot som enhver værdi, der ligger mere end 1,5 gange længden af IQR fra enten første eller tredje kvartil (Dekking et al., 2005).

METODE OG FORSKNINGSDSIGN

eksplorativt at undersøge de konkrete sproglige realiseringer, der ligger bag de numeriske værdier, som teksterne er tildelt i den kvantitative analyse. Hvad er det fx ved ordvalget i en tekst, der gør, at teksten har en høj gennemsnitlig ordlængde? Denne del er ikke teo-ribundet, men har en eksplorativ, induktiv karakter. For det andet er de kvalitative analyser knyttet til en mere teoretisk og deduktiv del, hvor fokus er på at analysere bestemte sproglige træk i teksterne, og det er denne del, der udgør den substantielt mest fyldige del af de kvalitative analyser.

Analyserne er struktureret sådan, at de knytter sig til to sproglige niveauer i teksterne, henholdsvis det leksikale og det syntaktiske niveau. Opdelingen i disse to niveauer foretages med bevidsthed om, at leksis og syntaks er gensidigt forbundne størrelser. Jeg har derfor også i mine forskningsspørgsmål rammesat de kvalitative analyser som leksikogrammatiske analyser. Med 'leksikogrammatisk' understreges det således, at leksis og syntaks er to forbundne poler langs samme akse (Halliday, 2013, s. 64). Analyserne er imidlertid ikke alene baseret på systemisk funktionel lingvistik, men fokuserer, som det kommer til at fremgå nedenfor, på forskellige grammatiske kategorier med rødder i forskellige sprogvidenskabelige traditioner. Betegnelsen leksikogrammatik anvendes derfor i stedet som paraplybetegnelse for de tre kvalitative analysedele og for at understrege, at analyserne på det leksikale niveau uundgåeligt vil gribe ind i det syntaktiske niveau og omvendt.

Det første tekststudvalg er som bekendt knyttet til de kvantitative analyser af tekstlængde. De kvalitative analyser undersøger i forlængelse heraf, hvordan tekstlængde hænger sammen med indhold i elevteksterne. Indhold forstår jeg her som semantiske felter, og det jeg specifikt undersøger er, hvordan tekstlængde kan bidrage til henholdsvis større eller mindre variation og udbyggethed i indholdet. Indholdsmæssig variation skal her forstås som anvendelsen af flere forskellige semantiske felter i en tekst, mens indholdsmæssig udbyggethed skal forstås som størrelsen på de semantiske felter. Et semantisk felt udgøres ifølge Christensen & Christensen (2019) af ord, "som falder inden for den samme betydningsmæssige og/eller referentielle domæne" (s. 313). Semantiske felter kan bestå af ord fra forskellige ordklasser og kan gå på tværs af syntagmatiske relationer, og deres indbyrdes semantiske relation afgøres ved, om de har fælles betydningskomponenter i relation til den kontekst, de fremgår i. Et ord som 'elev' kan fx i én kontekst tilhøre det semantiske felt 'skole' og i en anden kontekst tilhøre det semantiske felt 'menneske'.

Det afgøres konkret af den tekstuelle sammenhæng (Crystal, 1992, s. 346-347). Når jeg i analysen identificerer semantiske felter i elevernes tekster, vil det altså i sidste ende være et spørgsmål om fortolkning, hvor mange felter der er, og hvilke ord der falder under det enkelte felt (Jackson & Amvela, 2000, s. 15).

Analysen af det andet tekstudvalg af tekster fokuserer på det leksikale niveau i teksterne. De kvantitative analyser undersøger leksikal variation (OVIX) og specifikation (ordlængde), og dette følger jeg op på i de kvalitative analyser ved at fokusere på to kategorier, som jeg henter fra henholdsvis tekstlingvistikken og dansk frasegrammatik. Først vil jeg undersøge, hvordan eleverne anvender kohæsive bånd i teksterne. Kohæsion er interessant i denne sammenhæng som uddybning af særligt OVIX-analyserne, idet leksikal variation kan tænkes at hænge sammen med en stor grad af indholdsmæssig variation, hvilket stiller krav til elevernes anvendelse af kohæsive bånd i teksterne. Det er således interessant at se, om tekster med stor eller lille grad af leksikal variation også anvender flere eller færre kohæsive bånd. Jeg trækker her på Halliday & Hasans (1976) fire typer af kohæsive bånd; referencer, ellipser, konjunktionaler og leksikale bånd¹¹. Referencer, ellipser og leksikale bånd dækker over referentkoblinger/ordkoblinger, mens konjunktionalerne primært refererer til forskellige former for sætningskoblinger. I referentkoblingerne fokuserer jeg på det, som Halliday & Hassan kalder endoforiske referenter, dvs. tekstuelle referenter (i modsætning til eksoforiske referenter, som peger ud af teksten på kommunikationssituationen). De tre første kategorier er grammatiske kategorier, som kan identificeres strukturelt i teksten, mens den sidste kategori, leksikale bånd, er en semantisk kategori, som i højere grad baserer sig på betydningsmæssige relationer mellem ordene i teksten. De fire kohæsive bånd er opsummeret i tabel 3.

¹¹ Jeg har fulgt Ulbæk (2005) og udeladt Halliday & Hassans begreb substitution, idet det i Halliday & Hassans terminologi refererer til forhold, som enten sjældent optræder på dansk, eller som i stedet dækkes af ellipsekonstruktioner. Et eksempel er anvendelsen af et enkeltstående verbum som erstatning for et tidligere verbalsyntagma. Her vil man på dansk ofte sætte et 'det' ind som referentkobling til at akkompagnere den verbale substitution. Fx "Jeg drikker kaffe hver morgen, og **det** tror jeg også, du **gør**". På engelsk ville sætningen lyde "I drink coffee every morning, and I think you **do** as well". Her er verbet 'do' i Halliday & Hassans optik et eksempel på verbal substitution, mens vi altså på dansk både indsætter en pronominel referentkobling og en verbal substitution ('gør'). Jeg følger her Ulbæk og kalder i stedet ovenstående et eksempel på en elliptisk reference, fordi den modsat anaforiske referentkoblinger, som kan strække sig over lange tekststykker, er kortrækkende og dermed antager elliptisk karakter. Begrebet substitution, som det anvendes af bl.a. Bernhardt (2006), hvor det henviser til semantiske relationer (synonyme, hyponyme etc.) mellem ord, falder under Halliday & Hassans kategori 'leksikale bånd'.

Tabel 3

Oversigt over Halliday & Hassans (1976) typer af kohæsive bånd

Grammatisk/ strukturelt	Referencer	Endoforiske referentkoblinger i teksten, dvs. reference tilbage til noget tidligere omtalt i teksten. Der er ofte tale om pronomielle referentkoblinger, som udtrykkes enten anaforisk (fx navn ← pronomenen) eller kataforisk (pronomenen → navn).
	Ellipser	Udeladelse af et eller flere ord, der i stedet forstås ud fra tekstens eller sætningens sammenhæng. Dette ses ofte ved udeladelse af subjekt, hvis der er flere verballed tilknyttet, og subjektet allerede er nævnt (fx ”piggen leger i springvandet og ser glad ud”).
	Konjunktioner	Anvendelse af konjunktioner (både parataktiske og hypotaktiske) samt metatekst (fx adverbelle udtryk som ’for det første’).
Leksikalt/ semantisk	Leksikale bånd	Anvendelse af leksikale enheder til at skabe sammenhæng i teksten. Det kan fx være forskellige former for gentagelser (herunder synonyme og hyponymer) samt gennem forskellige former for kollokation (fx semantiske netværk eller faste leksikale forbindelser).

Efter analysen af kohæsive bånd i teksterne fra det andet tekstudvalg undersøger jeg, hvordan eleverne opbygger nominalfraser. Nominalfraser er særligt interessante, fordi de siger noget om specifikation og informationsrigdom i teksten. På denne måde kan analyserne af nominalfraserne ses som en forlængelse og uddybning af de kvantitative analyser af ordlængde, som ifølge Hultman & Westman (1977) også undersøger netop specifikation. Den kvantitative analyse undersøger det således på ordniveau, mens de kvalitative

analyser undersøger det på fraseniveau. Analysen af nominalfraserne tager udgangspunkt i kategorierne fra Hansen & Heltofts (2011a, s. 518) topologiske skema for nominalhelheder – dvs. med fokus på sammenfattere (kvantorer), bestemmere, tællere, beskrivere, kategori og efterstillede led. Selve skemaet og et par eksempler på, hvordan det udfyldes, fremgår af nedenstående tabel.

Tabel 4					
<i>Hansen & Heltofts (2011a) topologiske skema for nominalhelheder</i>					
Sammenfatter	Bestemmer	Tæller	Beskriver	Kategori	Efterstillede led
	mine	to	bedste	veninder	
hele	min			klasse	fra sidste år
				pigen	til venstre i billedet
		to	lysegrønne	sokker	med kaniner på

Skemaet giver, som det fremgår af ovenstående, information om de enkeltelementer, som udgør nominalfraserne i elevernes tekster. På denne måde bliver det muligt både at sige noget om længden og om variation i opbygningen af nominalfraserne.

I analysen af det tredje tekstudvalg af elevtekster fokuserer jeg på det syntaktiske niveau i elevteksterne. De kvantitative analyser undersøger sætningslængde (syntaktisk periode) og dependensafstand. Særligt analysen af sætningslængde har den ulempe, at det ikke er muligt at sige noget om, hvilke syntaktiske konstruktionsformer der ligger bag sætningslængden. Dette vil jeg kompensere for i de kvalitative analyser ved at fokusere på særligt to forhold. Først vil jeg fokusere på, hvordan eleverne anvender og kombinerer hoved- og ledsætninger. Jeg vil her både fokusere på, hvor udbredt henholdsvis hypotaktiske og parataktiske syntaks konstruktioner er, samt hvilke typer af ledsætninger der er dominerende i teksterne. Analysen af disse dimensioner tager bl.a. udgangspunkt i Diderichsens (1946) helsætnings-skema samt Hansen og Heltofts (2011b, s. 1479-1481) tre hovedgrupper for ledsætninger; nominale, adverbielle og adjektivistiske. Sætnings-skemaet giver på samme måde som skemaet for nominalfraser mulighed for at undersøge sætnin-

METODE OG FORSKNINGSDSIGN

gens enkeltelementer, ligesom den giver et overblik over syntaktisk variation i de analyserede sætninger. Skemaet, som er en adapteret udgave af helsætningskemaet, som det findes hos Diderichsen (1946), fremgår af nedenstående tabel.

Tabel 5								
<i>Adapteret udgave af Diderichsens (1946) helsætningskema</i>								
k <i>konjunktioner</i>	F <i>forfelt</i>	v <i>finite verbaler</i>	n <i>subjekt</i>	a <i>central-adverbialer</i>	V <i>infinite verbaler</i>	N <i>indirekte/direkte objekt +prædikativer</i>	A <i>frie adverbialer</i>	ekstra <i>ledsætninger + infinitiv</i>
	Nu	vil	jeg		fortælle	om en dag ligesom alle dage		men ikke helt fordi der var nedlukning i Danmark
og	det	var				kedeligt		fordi man ikke kunne lege sammen med nogle af sine venner.

Denne overordnede analyse af syntaktiske konstruktioner viser således i mere detaljeret grad, hvad der kendetegner de sætningskonstruktioner, som ligger bag analysen af sætningslængde, og det bliver derfor også muligt i mere detaljeret grad at sige noget om, hvilke klassesrelaterede forskelle der ses i elevernes syntaktiske udvikling. Efter den overordnede syntaktiske analyse vil jeg særligt zoom ind på, hvordan eleverne realiserer forfeltet i teksterne. Forfeltet er interessant af to årsager. For det første har det en vigtig tekstgrammatisk funktion ift. at skabe kohærens og har således betydning for informationsstrukturerne i teksten (Lindgren, 2021), og for det andet er forfeltet et vigtigt område

at fokusere på, hvis man skal sige noget om syntaktisk variation. Dette hænger sammen med, at forfæltet kan realiseres ved hjælp af alle tænkelige sætningsled (inkl. ledsætninger), hvilket gør det til et særligt interessant område set fra et skriveudviklingsperspektiv.

3.3.3 Den synkrone analysestrategi

Vi vender nu blikket mod den synkrone analysestrategi, som udgør anden halvdel af afhandlingens analyser. Her er der ikke kun fokus på skriftsproglig udvikling som en entydig tidsmæssig udvikling, men i højere grad på skriftsproglig udvikling som en kompleks sociokognitiv proces knyttet til både alder, askriptive dimensioner og de literacy-aktiviteter, eleverne deltager i uden for skolen. Denne analysedel udfoldes i to skridt. Først undersøger jeg, hvilke ekstra-lingvistiske variable der har den stærkeste sammenhæng med de lingvistiske variable. Dette gør jeg ved at lave en multipel regressionsanalyse for hver af de lingvistiske variable, hvor de ekstra-lingvistiske variable udgør de uafhængige variable, mens de lingvistiske variable er den afhængige variabel i hver model. For at kunne sammenligne betydningen af de ekstra-lingvistiske variable på tværs af måleenheder, inkluderer jeg standardiserede beta-koefficienter i alle modellerne. Det andet skridt i den synkrone analysestrategi består i at undersøge den specifikke betydning af henholdsvis køn og social baggrund. Den anden del af den synkrone analyse undersøger således køn og social baggrunds betydning for de sproglige variable gennem multiple lineære regressionsmodeller, hvor skrivelyst samt læse- og skrivevaner er indbygget som kontrolvariable.

3.3.3.1 Multiple lineære regressionsmodeller

Den synkrone analysedel er, som det fremgår ovenfor, bygget op omkring en række multiple lineære regressionsmodeller. En regression kan grundlæggende ses som en ligning, der repræsenterer sammenhængen mellem en afhængig variabel (outcome) og en eller flere uafhængige variable. Regressionsanalyser kan have forskellige formål, men de er oftest knyttet til et ønske om enten at kvantificere en uafhængig variabels (mere eller mindre kausale) betydning for en afhængig variabel eller til at forudsige, hvilke faktorer der forudsiger et bestemt resultat (Allison, 1999). Regressionskoefficienten, som angiver den gennemsnitlige forandring i den afhængige variabel, hvis man øger den uafhængige variabel med én enhed, baserer sig på den statistiske metode, der kaldes 'Ordinary Least Squares' (OLS). OLS har til formål at finde den rette linje gennem et datasæt, som bedst

METODE OG FORSKNINGSDSIGN

beskriver datasættets tendens. Dette gør OLS ved at estimere den linje, hvor der er mindst forskel mellem de observerede og de forudsagte værdier (Arkes, 2019, s. 21-23). Regressioner er derfor modeller, fordi de netop modellerer eller simulerer den bedst mulige sammenhæng mellem variablene.

Regressionsmodeller falder i to overordnede kategorier: simple eller multiple. Ved simple modeller foretages der bivariate analyser med kun én uafhængig variabel og ét outcome (afhængig variabel), mens der ved multiple modeller anvendes flere uafhængige variable til ét outcome. Med multiple regressioner kan man således sammenligne betydningen af forskellige uafhængige variable med hinanden. I de tilfælde, hvor man netop ønsker at sammenligne betydningen af flere uafhængige variable, er en vigtig faktor, at koefficienten for hver uafhængige variabel beregnes, mens de øvrige variable holdes konstante. At holde noget konstant betyder her, at man isolerer betydningen af den respektive uafhængige variabel ved at sikre, at de øvrige variable ikke forandrer sig sammen med den uafhængige variabel. Til grund for dette ligger en antagelse om, at ikke bare den uafhængige variabel, men også kontrolvariablerne har effekt på den afhængige variabel og for at sikre, at denne effekt ikke influerer på effekten af den uafhængige variabel, så holdes kontrolvariablenes effekter konstante i modellen (Arkes, 2019, s. 61-62).

I den første del af den synkrone analysedel, sammenligner jeg alle de ekstra-lingvistiske variables betydning for de lingvistiske variable. I disse regressionsmodeller angives regressionskoefficienten, mens alle andre uafhængige variable holdes konstante. Dette gøres netop for at isolere den respektive betydning af hver enkelt variabel. Ulempen ved dette er imidlertid, at man kommer til at holde nogle variable konstante, som ellers teoretisk set må antages at bevæge sig sammen med de andre. Et eksempel kunne være, at en stor del af betydningen af social baggrund kan forklares af elevernes skrivelyst eller læsevaner. Disse påvirkninger skjules imidlertid i de indledende regressionsmodeller, hvor alle variable holdes konstante. Dette kompenseres der til gengæld for i den anden analysedel, hvor jeg undersøger den mere specifikke betydning af nogle af de ekstra-lingvistiske variable. Her inkorporeres de øvrige ekstra-lingvistiske variable som kontrolvariable, hvilket gør det muligt at undersøge, hvordan de uafhængige variable påvirker hinanden.

3.3.3.2 Standardiserede beta-koefficienter

Eftersom gruppen af uafhængige variable i regressionsmodellerne i første halvdel af analysen optræder i forskellige enheder, er det svært at estimere deres relative betydning. Det vil sige, at det er svært at sammenligne betydningen af de uafhængige variable indbyrdes, eftersom de optræder i forskellige måleenheder (fx er social baggrund en binær kategori, mens skrivelyst er et indeks bestående af fem punkter). Derfor inkluderer jeg standardiserede beta-koefficienter i regressionsmodellerne, da de siger noget om sammenhængen mellem de uafhængige variable og den afhængige variabel uafhængigt af de enheder, de optræder i. Standardiserede beta-koefficienter angiver, hvor mange standardafvigelser en afhængig variabel vil forandre sig for hver øget standardafvigelse i den uafhængige variabel. Ved at tage standardafvigelsen som enhed muliggør beta-koefficienterne, at de uafhængige variables relative betydning kan sammenlignes igennem samme enhed (Arkes, 2019, s. 38-39).

3.3.4 Analysestrategierne og kvalitetskriterier for forskning

De to analysestrategier, som jeg har præsenteret i de ovenstående afsnit, undersøger i sammenhæng både forskellige og overlappende dimensioner af skriftsproglig udvikling i danskfaget fra 5.-9. klassetrin. I dette afsnit vil jeg kommentere på tre kvalitetskriterier for forskning, som er relevante for de to analysestrategier. Det drejer sig om 'trustworthiness', 'conversion legitimation' og intern validitet. Dette afsnit indeholder generelle overvejelser over kvalitetskriterierne, som ligger til grund for forskningsdesignet. Det vil selvfølgelig også være relevant at kommentere på begrebet konstruktvaliditet, eftersom dette er grundlaget for kvaliteten af de fænomener, som måles i denne afhandling. Overvejelserne over konstruktvaliditeten af de forskellige variable vil imidlertid ikke fremgå af dette afsnit, men af afsnit 4.3 i næste kapitel, hvor de enkelte variable beskrives. Der er desuden to yderligere kvalitetskriterier, som er relevante i denne sammenhæng. Det ene har jeg allerede kommenteret på i afsnit 3.2.1.2, nemlig økologisk validitet, og det andet (konsekventiel validitet) vil jeg kommentere på senere i dette kapitel i afsnit 3.5.2. I det følgende vil jeg kort kommentere på de tre forskellige kvalitetskriterier, som jeg nævnte ovenfor.

3.3.4.1 'Trustworthiness' i den diakrone analysedel

Det første kvalitetskriterium, vi skal se på, er 'trustworthiness', som er relevant for afhandlingens diakrone analysedel. Den diakrone analysedel er som bekendt bygget op om en deskriptiv kvantitativ del med fokus på de lingvistiske variables kvartilsvise udviklinger relativt til klassetrin samt en række kvalitative leksikogrammatisk analyse af tre tekstudvalg af elevtekster. I og med den kvantitative del er deskriptiv, består den analytiske del af den diakrone analysedel af mine læsninger af tallene og af box-plot-figurerne samt de kvalitative sproglige analyser af de tre tekstudvalg. Formålet med den diakrone analysedel er således at identificere mønstre og strukturer i tekstkorpuset ved hjælp af skiftevis 'distant reading' (se afsnit 1.2.1) og nærlæsninger af det foreliggende tekstkorpus. Der er således ikke tale om en analyse, som har til hensigt hverken at identificere kausale sammenhænge eller generalisere til en større population. Sidstnævnte hænger også sammen med, at tekstkorpuset i denne afhandling ikke er repræsentativt og derfor i stedet skal ses som et casestudie. Det er derfor ikke meningsfuldt at applicere traditionelle kvantitative kvalitetskriterier som fx intern validitet eller generaliserbarhed på denne analysedel. I stedet vil jeg trække på Lincoln & Gubas (1985) klassiske begreb 'trustworthiness', som er et overbegreb, der dækker over epistemologiske og metodologiske kvalitetstræk i kvalitativ forskning. Selvom begrebet typisk anvendes i forbindelse med etnografiske studier, er begrebet imidlertid også relevant i denne sammenhæng, eftersom det bl.a. fokuserer på forskerbias og etableringen af stringens i analyse og fortolkning, som er vilkår for alle typer af kvalitative og hermeneutisk baserede studier.

'Trustworthiness' består af fire underbegreber: credibility, transferability, confirmability og dependability. 'Credibility' refererer til forskerens tillid til sandhedsværdien af de analytiske fund. Denne tillid kan ifølge Lincoln & Guba etableres på forskellige måder, fx gennem triangulering og longitudinale observationer. I kontekst af denne afhandling er det fortrinsvis triangulering, der er relevant. Triangulering kan ifølge Lincoln & Guba udformes på forskellige måder, bl.a. ved anvendelse af forskellige teorier eller analysemetoder på samme genstandsfelt. I den diakrone analysedel anvender jeg en form for metodisk triangulering, idet jeg undersøger de samme overordnede fænomener i elevernes tekster (produktivitet, leksis og syntaks) ud fra to forskellige metodiske perspektiver, dels et

kvantitativt og computationelt perspektiv med fokus fordelingen af de kvantitative variable relativt til klassetrin og dels et leksikogrammatisk perspektiv med fokus på de computationelle analysers bagvedliggende lingvistiske konstruktioner. Credibility i den diakrone analyse forsøges således opretholdt gennem samstillingen af de to metodiske perspektiver, i den forstand at resultaterne fra de to analytiske perspektiver belyser hinanden gensidigt og sammenlagt bidrager til mere troværdige analyseresultater. Jeg sammenstiller og diskuterer desuden løbende mine analyseresultater med øvrig forskningslitteratur inden for samme område for at perspektivere, bekræfte eller udfordre resultaterne. Denne type af kvalitetssikring er således relateret til det, Lincoln og Guba (1985, s. 305-306) refererer til som 'contextual validation' og kan ifølge forfatterne ses som en måde at styrke analysernes credibility på.

Det andet underbegreb 'transferability' angår analyseresultaternes overførbarehed til andre kontekster. Jeg har allerede i afsnit 3.2.2.3 kommenteret på korpussets repræsentativitet, så jeg vil derfor kun kort kommentere på emnet igen i regi af transferability-begrebet. Lincoln & Gubas (1985) transferability-begreb er knyttet til "the thick description necessary to enable someone interested in making a transfer to reach a conclusion about whether transfer can be contemplated as a possibility." (s. 316). Transferability er altså et spørgsmål om at kontekstualisere af de indsamlede data, sådan så andre forskere kan sammenligne med deres respektive tekstkorporer og vurdere, om der er grundlag for at overføre fra én kontekst til en anden. I denne afhandling har jeg foruden de indsamlede tekster samlet metadata ind om de deltagende elever. Det drejer sig bl.a. om køn, klassetrin, social baggrund og sproglig baggrund. Disse data inddrager jeg løbende i afhandlingen til at kontekstualisere og beskrive de deltagende elever, hvilket kan ses som en måde at kvalificere transferability på. Jeg kan imidlertid ikke sige meget om de didaktiske vilkår, som eleverne har produceret teksterne under, hvilket omvendt er en transferability-svaghed i denne afhandling.

Jeg kommenterer på det tredje og fjerde underbegreb, henholdsvis 'dependability' og 'confirmability', i sammenhæng, eftersom de dækker over meget identiske dimensioner. Begreberne angår graden af neutralitet og forskerbias i analyseresultaterne i den forstand, at analyseresultaterne skal være konsistente og transparente. Dette kan ifølge Lincoln & Guba sikres gennem et såkaldt 'audit trail', som er en detaljeret og transparent beskrivelse af dataindsamling, databearbejdning, analyse og fortolkning. Der er to forhold

i den diakrone analysedel, som er med til at styrke confirmability og dependability, og ét forhold, der svækker dem. Det første positive forhold er, at udviklerne af TextDescriptives-pakken til Python, som jeg har anvendt til de kvantitative, sproglige analyser, har gjort pakken open source og Python-koden, som analyserne er baseret på, ligger frit tilgængelig på GitHub (se Hansen, u.å.). Jeg har desuden tidligere i dette kapitel beskrevet vilkårene for dataindsamlingen samt principperne for databearbejdning (herunder standardisering af elevtekster). Det andet forhold, der er med til at styrke confirmability og dependability, er knyttet til de kvalitative, sproglige analyser. Her har jeg bestræbt mig på at gøre analyserne så transparente som muligt ved at gengive alle de analyserede tekster i deres helhed samt ved at dokumentere mine analyser enten gennem farvekoder i teksten (jeg har fx markeret min analyse af de forskellige kohæsive bånd med forskellige farver) eller gennem bilag, hvor teksterne bl.a. er sat ind i sætningskema eller skemaet for nominalhelheder. Alle analytiske betragtninger eller resultater fra den kvalitative analyse kan altså efterprøves på baggrund af materiale, der ligger åbent og frit tilgængeligt i afhandlingen.

Et forhold, der til gengæld kan svække confirmability og dependability i den diakrone analysedel, angår udvælgelsen af tekster til de tre tekstudvalg. Udvalget af tekster til de tre tekstudvalg er en iterativ proces, som, selvom teksterne er valgt med deres variabelværdier som det primære pejlemærke, ikke kan foretages på baggrund af fuldstændig transparente og faste kriterier. Dertil er der for mange andre hensyn, der til tider har overtruffet de primære kriterier, fx har jeg fravalgt tekster med særligt følsomme oplysninger eller tekster, som efter en indledende analyse ikke viste sig at være interessante nok ift. forskningsspørgsmålene (udbydes kort i afsnit 5.1). Der ligger således en mulig forskerbias i udvælgelsen af tekster, som af nævnte årsager har været svær at undgå.

3.3.4.2 'Conversion legitimation' i metodeintegrationen internt og imellem analysestrategierne

Det andet kvalitetskriterium, jeg vil kommentere på i relation til analysestrategierne, er begrebet 'conversion legitimation', som stammer fra Onwuegbuzie & Johnson (2006). Conversion-begrebet er knyttet til kvaliteten af de metainferenser, som drages i projekter, hvor kvalitative data behandles kvantitativt eller omvendt. Forfatterne gør bl.a. opmærk-

som på, at der er fare for at overdrive eller underdrive tilstedeværelsen af bestemte fænomener, hvis man tæller træk i kvalitative data uden at fortolke disse relativt til den kontekst, de optræder i. I denne afhandling tæller jeg lingvistiske træk i elevernes tekster, men de frekvenser, som fremkommer af disse analyser, kan imidlertid ikke siges i sig selv at over- eller underestimere et fænomen, eftersom analyserne er direkte analyser af sproglige træk i teksten (se uddybning i afsnit 3.4.2).

Der er imidlertid fare for over- eller underestimering af betydningen af resultaterne i fortolkningerne. Dette hænger sammen med, at jeg ikke har data om tekstkvalitet eller et benchmarkingkorpus, der kan anvendes til at vurdere, hvornår noget fx er en stor eller lille udvikling i variablene. Jeg forsøger derfor i stedet at fastholde proportionaliteten i fortolkningerne ved at sammenligne resultaterne med øvrig forskning. I de kvalitative analyser forekommer der også eksempler, hvor jeg manuelt tæller bestemte træk i teksterne og sammenligner denne frekvens på tværs af tekster (fx antallet af lange nominalkvoter eller antallet af semantiske felter). Når jeg tæller i de kvalitative analyser er jeg opmærksom på at fortolke tallene i kontekst af den konkrete elevtekst, og når jeg sammenligner manuelt optalte frekvenser på tværs af elevtekster er jeg opmærksom på, at sammenligningen må tage højde for forskelle i fx tekstlængde, teksttype eller tekstens øvrige sproglige og stilistiske sammenhæng.

Problematikken går naturligvis også den anden vej, eftersom de kvantitative data også behandles kvalitativt. Dette hænger bl.a. sammen med, at kvantitative analyser af elevtekster genererer en form for numerisk billede af elevteksterne (Gavin, 2020), som i sig selv kan læses som numeriske tekster, der kræver fortolkning (uddybes i afsnit 3.4.3). Disse billeder er imidlertid også reduktioner af det fænomen, der undersøges, idet de både er baseret på et begrænset antal datapunkter og på gennemsnits- og kvartilsværdier. Der er således en del variation, som ikke kommer med i analysen. Det er derfor vigtigt at undgå det, som Onwuegbuzie & Johnson (2006) kalder ”an over-generalization of the observed numeric data” (s. 59). I denne afhandling indebærer dette for det første, at skriftsproglig udvikling som fænomen ikke kan reduceres til de variable, som undersøges i afhandlingen, og for det andet, at gennemsnitlige udviklinger i lingvistiske variable ikke overfortolkes og generaliseres til generiske udviklingsstadier eller lignende, som gælder for alle elever.

3.3.4.3 Intern validitet i regressionsmodellerne i den synkrone analysedel

Det sidste kvalitetskriterium, jeg vil kommentere på, er intern validitet knyttet til regressionsmodellerne i den synkrone analysedel. Intern validitet er et bredt begreb, der dækker over, hvorvidt en konklusion eller et resultat kan opretholdes på baggrund af de foreliggende data og analyser (Cohen et al., 2018, s. 252). Hanck et al. (2021) peger på en række trusler mod den interne validitet i multiple regressionsmodeller. Jeg vil her kort kommentere på særligt to af disse, nemlig 'omitted variable bias' og 'errors-in-variable bias'.

'Omitted variable bias' dækker ifølge Hanck et al. (2021) over det tilfælde, hvor en regressionsmodel udelader en variabel, som både er korreleret med den afhængige variabel (outcome) og en eller flere af de uafhængige variable, og som derfor forårsager en bias i estimatet af regressionen. Denne bias kan imidlertid være vanskelig at undgå i et studie som dette, hvor den relation, der undersøges (fx betydningen af social baggrund for tekstlængde) er kompleks, og hvor der potentielt kan være mange 'confoundere', der påvirker begge veje i modellen. I den anden analysedel undersøger jeg derfor specifikt betydningen af køn og social baggrund, eftersom disse er askriptive variable, dvs. de i princippet er medfødte. Dette betyder, at det vil være begrænset, hvor mange variable der påvirker disse. Jeg har dog inkluderet sproglig baggrund som kontrolvariabel i de modeller, der har social baggrund som uafhængig variabel, eftersom sproglig baggrund kan tænkes både at påvirke social baggrund og de lingvistiske variable.

'Errors-in-variable bias' dækker over målefejl i variablerne, som kan være med til at skabe en bias i estimatet af regressionskoefficienten. Dette er relateret til, men ikke det samme som konstruktvaliditet, der handler om den konkrete operationalisering af de teoretiske fænomener, der undersøges (diskussion af konstruktvaliditet kommer i afsnit 4.3). Errors-in-variable bias dækker imidlertid over målefejl relateret til selve de instrumenter og data, der anvendes i analyserne. I denne afhandling gælder det dels de NLP-værktøjer, der er anvendt til at analysere tekstkorpuset, og dels det survey, som er anvendt til at indsamle metadata om de deltagende elever. Starter vi med at se på NLP-værktøjerne vil disse uundgåeligt være forbundet med en grad af målefejl, eftersom de er baseret på probabilistiske modeller. Ifølge Kyle (2021) betyder dette, at der stadig er en del usikkerhed forbundet med anvendelsen af disse, eftersom mange NLP-værktøjer er trænet på professionelle eller redigerede tekstkorpuser, som er langt fra de data, som

man typisk arbejder med i elevtekster. Jeg har i denne afhandling forsøgt at afbøde de værste af disse målefejl ved at redigere og standardisere de dele af elevernes tekster, som helt åbenlyst vil føre til målefejl (se afsnit 3.2.2.2). Hvad angår surveyet og de data, der er frembragt på baggrund af det, er der selvfølgelig også forbehold, der må tages. Jeg har allerede kommenteret på en del af disse i afsnit 3.2.3, og jeg vil kommentere mere uddybende på specifikke problemstillinger med de enkelte variable i forbindelse med gennemgangen af disse i næste kapitel. Jeg vil derfor nøjes med at angive her, at disse målefejl selvfølgelig ikke kan undgås helt, og at jeg løbende vil tage de nødvendige forbehold både i fortolkningen af analyseresultaterne samt i den efterfølgende diskussion af disse.

3.4 Videnskabsteoretiske perspektiver

I det følgende vil jeg anlægge et videnskabsteoretisk perspektiv på afhandlingens analyser. Jeg vil først beskrive, hvordan afhandlingen teoretisk set er forbundet med strukturalistisk lingvistisk, samt hvilke konsekvenser dette har for operationaliseringen af de sproglige fænomener i analyserne. Afslutningsvis vil jeg beskrive afhandlingens epistemologiske afsæt i den computationelle hermeneutik.

3.4.1 Strukturalisme som ontologisk og metodologisk forståelsesramme

Et væsentligt spørgsmål, som rejser sig til dette projekts analyser, angår, hvorvidt NLP-baserede korpusanalyser markerer en tilbagevenden til en klassisk deskriptiv, strukturalistisk lingvistik. For at svare på dette spørgsmål er det nødvendigt at skelne mellem to videnskabsteoretiske forståelsesniveauer, et henholdsvis ontologisk og et metodologisk forståelsesniveau. Strukturalisme på det ontologiske niveau er både en teori om sproget og den særlige autonome status, som tillægges den sproglige struktur, og samtidig er det i en videre betydning en teori om, hvad der udgør den sociale virkelighed som sådan (Heydebrand, 2001). En strukturalistisk ontologi indbefatter således et særskilt fokus på strukturelle relationer imellem elementer i et system. I sprogets tilfælde gælder det, at sproget ses som en autonom struktur, hvis elementer får betydning gennem deres indbyrdes relationer til hinanden (Saussure, 1916). Med Chakravartys (2003) ord indebærer en strukturalistisk ontologi således, hvis man sætter det meget på spidsen, at ”we can know structural aspects of reality, and that there is nothing else to know” (s. 867). Strukturalisme på det metodologiske niveau er derimod et udtryk for en mere pragmatisk tilgang,

hvor sproglige strukturer gøres til genstand for analyse, fordi de er observerbare indgange til at sige noget om fænomener og konstrukter, som ikke i sig selv kan observeres direkte (fx skrivekompetence eller -udvikling), eller som ikke i sig selv kan reduceres til selvamme observerbare strukturer. Strukturernes er således epistemologiske indgange til at undersøge bestemte fænomener, men strukturernes udgør ikke i sig selv et ontologisk ud-tømmende beskrivelsesniveau for det pågældende fænomen. Det er således muligt at anvende en strukturalistisk metodologi uden nødvendigvis at abonnere på de ontiske reduktioner, som er forbundet med strukturalisme på det ontologiske niveau.

Dette projekt er strukturalistisk i en metodologisk forstand, idet projektets empiriske genstandsfelt er sproglige overfladestrukturer i elevernes tekster. Projektet er imidlertid ikke strukturalistisk i en mere normativ ontologisk forstand, eftersom skriveudvikling i denne afhandling forstås som andet og mere end de sproglige strukturer, som fremtræder i elevteksterne. De sproglige overfladestrukturer kan derimod ses som en metodologisk trædesten hen mod at undersøge forskningsspørgsmål, som traditionelt anses som værende kvalitative, fx hvad der kendetegner sprog- og tekstproduktion i bestemte kontekster (Berthelsen & Tannert, 2021). Korpusanalyserne i dette projekt er således ikke en tilbagevenden til strukturalismens klassiske dikotomi mellem sprogets mening på den ene side og sprogets autonome struktur på den anden side, som bl.a. ses praktiseret i Harris' (1954) klassiske teori om sproglig distribution, hvor Harris forsøger at vise, at sproget kan ses som et formelt system, der kan gøres til genstand for videnskabelige studier, og hvor regelmæssigheder i distributionen af enkelte sproglige enheder er den primære forskningsgenstand. Harris' centrale pointe er, at sprogets mening er knyttet til dets struktur, som muliggør og begrænser menneskets muligheder for at sammensætte meningsfulde sætninger. Det vil sige, at sprogets struktur ikke er arbitrær, men tværtimod er sprogets enkelte bestanddele knyttet sammen i et strukturelt 'miljø' af enheder, der regelmæssigt optræder sammen (co-occurents), og det er derfor disse strukturelle sproglige miljøer, som lingvistikken bør undersøge og beskrive. Dette passer på sin vis godt sammen med dele af korpuslingvistikens traditionelle formål, hvor der bl.a. er fokus på at identificere gentagende sproglige mønstre (Biber et al., 1998), og dermed også med dette projekts metodiske tilgang. Men som det bl.a. fremhæves af Gastaldi (2021), er problemet med distribueringsstilgangen, at den adskiller den sproglige struktur fra mening

og semantik, og derfor er det ikke tilstrækkeligt med en rent strukturalistisk tilgang til korpusforskning.

Snarere end et udtryk for en strukturalistisk vending er projektet således knyttet til en bredere empirisk og sproglig vending inden for humanvidenskaberne, og her er analyserne af den strukturelle overfladestruktur ikke ensbetydende med en reduceret forståelsehorisont, hvor de sproglige fænomener, som undersøges i projektet, ontologisk set kan reduceres til de observerbare strukturer, der udgør fænomenets overflade. I stedet må disse overfladestrukturer ses som empiriske indgange til at beskrive, forstå og fortolke disse mere komplicerede fænomener inden for dels en funktionel sproglig ramme og dels indenfor den pædagogiske kommunikationskontekst, disse strukturer er produceret under (Tannert, 2021). Dette rejser imidlertid to nye spørgsmål – et spørgsmål om forholdet mellem det sproglige fænomen og dets overfladestruktur (dvs. operationaliseringen) samt et spørgsmål om fortolkningens betydning i analyseprocessen. Dette vil jeg derfor se nærmere på i de to næste afsnit.

3.4.2 Om operationalisering af sproglige fænomener

Operationalisering er et gammelt begreb, som går på tværs af videnskabelige discipliner. Det angår spørgsmålet om, hvordan teoretiske begreber (t-terms) forbindes med observerbare forhold i verden (o-terms) (Nordahl-Hansen & Kvernbekk, 2020). Moretti (2013) rejser derfor også spørgsmålet om operationalisering inden for digital humaniora. Selvom operationalisering angår alle empiriske videnskaber, knytter Moretti særligt operationalisering til anvendelsen af digitale metoder til kvantitative målinger. Ifølge Moretti indbefatter begrebet operationalisering ikke data-drevne tilgange, hvor de empiriske observationer/målinger går forud for teorien. I stedet går operationaliseringsprocessen fra teori over data til den virkelige verden eller med Morettis egne ord: "Measurement does not lead from the world, via quantification, to the constructions of theories; if anything, it leads back from theories, through data, to the empirical world (...) Measurement anchors theories to the world they describe" (Moretti, 2013, s. 106-107). Morettis tilgang er altså funderet i en form for realisme, hvor teoretiske konstrukter kan forbindes til verden ved at indsamle data på baggrund af fastlagte operationelle handlinger, som er begrundede i den teori, de skal repræsentere.

Dette rejser selvfølgelig en række spørgsmål - først og fremmest til forholdet mellem teorien og de forhold i den virkelige verden, de repræsenterer, idet der oplagt er forskel på at skulle operationalisere et fysisk fænomen som fx længden af en meter og et lingvistisk fænomen som fx leksikal specifikation. Dette problem adresseres bl.a. af Stubbs (2007), som specifikt undersøger de ontologiske forskelle, som ligger til grund for korpuslingvistiske undersøgelser. Stubbs skelner – med reference til bl.a. Searles (1995) teori om den sociale virkelighed og Poppers (1979) treverdensmodel – mellem ytringer, individuel viden og social viden. Ytringer eksisterer som det, Searle kalder *brute facts*. Dvs. fakta som eksisterer fysisk i verden, og som dermed har fysiske egenskaber, der ontologisk set eksisterer samstemmigt med, men samtidig uafhængigt af det sociale (adfærd, fysisk bevægelse, planter, sten etc.). Ytringer skal således forstås som sproglige hændelser, som er udfoldet gennem fysisk adfærd (en konkret tekst, en konkret sproglig ytring etc.). Individuel viden er derimod en subjektiv viden, som korresponderer med det, Searle kalder psykologiske fakta. Dvs. fænomener som er private og mentale. Individuel viden er derfor i Stubbs' forstand lig med individets viden om og evne til at producere sprog (det som Chomsky kalder *i-language*, se afsnit 3.2.2.3). Social viden er viden, der hviler på menneskelige (enten eksplicite eller implicite) konventioner. Det korresponderer således med Searles ide om institutionelle fakta, som er dem, der opretholder den sociale virkelighed, og det er ifølge Stubbs her, at de sociale og kulturelle aspekter forbundet med bestemte tilfælde af sprogbrug eksisterer.

Stubbs' (2007) sproglige trevejsmodel viser således det, som udgør korpuslingvistikens ontologiske betingelser. Korpuslingvistikens genstandsfelt er nemlig store mængder af ytringer, eftersom det er de eneste fænomener, som er empirisk observerbare. Individuel viden kan ikke observeres direkte, og ingen former for intro- eller retrospektion (interviews, think-aloud-protokoller etc.) vil kunne afdække det til fulde. Det samme gælder social viden, som hviler på konventionelle aspekter i menneskelige og kulturelle fællesskaber. Disse typer af viden kommer kun indirekte til udtryk gennem ytringer, og forudsætningen for at kunne sige noget om individuel eller social viden er således, at der foretages en valid operationalisering, hvor ytringer eller ytringselementer måles på baggrund af en teoretisk beskrivelse af forholdet mellem det psykologiske eller sociale fænomen på den ene side og de ytringer, som måles, på den anden.

Vender vi blikket mod de lingvistiske og ekstra-lingvistiske fænomener, som måles i denne afhandling, er det tydeligt, at selvom de alle kræver operationaliseringer, så kan operationaliseringen af dem forstås på forskellige måder. I én forstand er der ved flere af de lingvistiske variable tale om det, jeg kalder for 'direct measurement'. Her er der tale om sproglige fænomener, som kan observeres direkte i teksten, og som derfor primært kræver en computationel operationalisering. Det gælder fx tekstlængde eller sætningslængde. Disse fænomener hviler selvfølgelig på social viden (konventionaliserede, sproglige begreber), men har primært fysiske ontologiske egenskaber. Dvs. de observerbare enheder, der måles i teksten, dækker fænomenet helt. Sprog er selvfølgelig social viden (det er fx konventionaliseret, hvornår en sætning tæller som en sætning og et ord som et ord), men forholdet mellem den teoretiske sproglige struktur og de observerbare ytringer i teksten er ved de længdebaserede mål tæt på 1:1. Det er således et spørgsmål om at tælle ordene eller bogstaverne, og så har man dækket variabelen fuldt ud. Dette er imidlertid heller ikke en helt dækkende beskrivelse af de variable, der indgår i analyserne.

I en anden forstand er der nemlig ved alle de lingvistiske variable tale om en form for 'indirect measurement', idet de lingvistiske variable også er proxyer for mere end det, der udgør variabelens operationalisering. For OVIX gælder det fx, at den konkrete operationalisering skal agere som en proxy for et større konstrukt om leksikal variation, og det kræver således en teoretisk begrundelse, hvorfor OVIX kan sige noget om leksikal variation i en tekst. Det samme gælder fx ordlængde, som er et udtryk for leksikal specifikation (uddybes i afsnit 4.1.2.2), og dependensafstand, som er et udtryk for syntaktisk kompleksitet og sværhedsgrad i den kognitive processering af teksten (se afsnit 4.1.3.2). Afhængig af hvilket perspektiv man anlægger i analysen, kræver variablene i denne afhandling således forskellige former for teoretisk begrundede operationaliseringer.

3.4.3 Computational hermeneutik som fortolkningsramme

Som det fremgår af forrige afsnit, er det ikke trivielt, hvordan korpuslingvistiske analyser af bestemte sproglige træk i elevtekster kan bidrage til en beskrivelse af et større og bredere begreb om skriveudvikling, ligesom det heller ikke er trivielt, hvordan man forstår og fortolker de enkelte kvantitative analyser. Derfor er de fortolkningsprocesser, som er forbundet med kvantitative analyser af tekst, blevet et vigtigt emne i litteraturen inden for både korpuslingvistikken og digital humaniora i bredere forstand. Her har der særligt

været fokus på det område, som kaldes computationel hermeneutik, som er en samlebetegnelse for forskellige teoretiske beskrivelser af forholdet mellem computationel og automatiseret analyse på den ene side og humanistisk fortolkning på den anden.

Et vigtigt bidrag til dette felt er Mohr et al.'s (2015) tre typer af tekstlæsning; 'thin reading', 'close reading' og 'thick reading'. Med thin reading refereres der til den type af tekstanalyse, hvor simple manifesterede variable i teksten identificeres (enten kvalitativt eller kvantitativt) og herefter bliver omdannet til numeriske data, som kan gøres til genstand for statistiske analyser. Denne tilgang er bl.a. udbredt i de skriveforskningsstudier, som opererer med forholdsvis små tekstkorpusser, hvor det er muligt at score teksterne manuelt, eller i studier af større tekstkorpusser, hvor der til gengæld kun analyseres ganske få variable. Den anden type tekstlæsning kaldes close reading og refererer til velkendte kvalitative analyser, hvor det er muligt at analysere et stort antal lingvistiske træk, men til gengæld kun i forholdsvis få tekster. Den sidste type tekstlæsning kaldes thick reading og skal forstås som en betegnelse for den type af analyse, hvor så mange meningsfulde variable som muligt undersøges i et stort tekstkorpus for at skabe et rigt datasæt, som kan gøres til genstand for fortolkning. Der er således tale om en form for kvantitativt baseret 'thick description', hvor den epistemiske pluralitet fra close reading kombineres med skalerbarheden i thin reading. Anskuet fra dette perspektiv bør de statistiske resultater af tekstanalyserne ikke bare ses som færdige resultater med iboende manifest mening, men derimod som numeriske tekster, der på samme måde som andre tekster kræver en fortolkning inden for en teoretisk ramme (Agersnap & Johansen, 2021).

Om analyserne i denne afhandling kan kategoriseres som thick reading, kan selvfølgelig diskuteres, men begrebet bidrager alligevel med en meningsfuld beskrivelse af den fortolkningsproces, som også er et vilkår for denne afhandling. De kvantitative, sproglige analyser bliver således først meningsfulde, når de sammenholdes og modstilles med afhandlingens øvrige lingvistiske og ekstra-lingvistiske variable og holdes op imod forskellige teoretiske antagelser og anden relateret forskning. Med Dourish & Cruz' (2018) ord kan man sige, at "Data makes sense only to the extent that we have frames for making sense of it, and the difference between a productive data analysis and a random-number generator is a narrative account of the meaningfulness of their outputs." (s. 8). Det er således denne narrative fremstilling, som er afhandlingens fortolkningsmæssige primat,

og det er ligeledes en sådan tilgang til analysearbejdet, der gør det muligt at koble ”traditional computational approaches involving abstraction and automation with well-known humanistic approaches involving contextualization and interpretation” (Tannert, 2021, s. 19). Disse narrative fremstillinger af analyseresultaterne er således en vigtig del af denne afhandlings analyser.

3.5 Forskningsetiske perspektiver

Den sidste del af dette metodeafsnit handler om de forskningsetiske perspektiver, som rejser sig på baggrund af afhandlingens forskningsdesign og overordnede erkendelsesinteresse. Idet indsamlingen af tekster og metadata ikke direkte har involveret interaktion med de deltagende i felten, har projektet undgået en række af de etiske udfordringer, der normalt møder empirisk forskning i skolekontekster. Det gælder fx forsvarlig fremstilling af elevadfærd og direkte påvirkning af elevernes trivsel (Graham et al., 2015). Der har imidlertid rejst sig en række andre etiske problemstillinger, som angår databehandlingen samt projektets potentielle didaktiske og pædagogiske konsekvenser.

3.5.1 Elevtekster som data og principper for anonymisering

Hvad angår databehandlingen, har særligt to problemstillinger været påtrængende. Først og fremmest har anonymisering, bl.a. i kraft af GDPR (European Parliament, 2016), været et væsentligt fokuspunkt. De mere formelle elementer af dette, fx fjernelse af elevnavne og sidehoved, er foretaget automatisk i forbindelse med formateringen og annoteringen af tekstkorpuset. Der er imidlertid også den særlige problemstilling ved elevtekster, at der i selve brødteksten kan skjule sig personfølsomme og personhenførbare oplysninger (fx om elevens familieforhold, navne eller geografiske oplysninger). Disse oplysninger er blevet fjernet fra de tekster, som er gengivet i deres helhed i denne afhandling. Det vil sige de tekster, som fremgår af de tre tekstudvalg i den diakrone analysedel. Helt konkret er personfølsomme eller –henførbare oplysninger fjernet og erstattet af en firkanten parentes, hvor der i stedet er angivet ordklasserne på de ord, der er fjernet.

Disse tiltag løser imidlertid kun de juridiske forholdsregler, som er forbundet med håndtering af elevdata. Et mindst ligeså væsentligt element angår den personlige karakter af elevernes tekster, som ikke er direkte forbundet med personhenførbare oplysninger, men som i højere grad er relateret til elevernes tekstuelle stemme og skriveridentitet.

Denne tekstuelle stemme udtrykkes i teksten gennem kommissiver og mere konkret gennem elevens anvendelse af eksempelvis adverbelle udtryk, der udtrykker attitude (Biber et al., 1999), og er en ikke bare vigtig, men også meget personlig del af elevernes udvikling som skrivere (Ivanic, 1998). Dette betyder således også, at elevernes tekster ikke bare er neutrale biprodukter af undervisningen, men i stedet må ses som personlige artefakter, som er produceret i henhold til en særlig didaktisk kontrakt mellem lærer og elev.

Problemstillingen er naturligvis velkendt i den pædagogiske skriveforskning, men den nærmere diskussion af, hvordan denne problemstilling mere konkret håndteres, har særligt inden for den kvantitative skriveforskning været knyttet til diskussioner om målings- og begrebsvaliditet. Dette skyldes, at en stor del af den kvantitative skriveforskning er knyttet til udviklingen af pædagogiske tests, og her har devisen således haft en sociologisk karakter og er gået på, at jo mere præcist vi måler forskellige lingvistiske træk i elevernes tekster, desto mindre ulighed er analyserne med til at producere (se fx Elliot, 2016; Slomp, 2016). Men som det pointeres af De Costa (2015), indebærer kvantitative analyser af lingvistiske data også en etisk forpligtelse på at varetage de mere personlige dimensioner af sproget fortroligt samt grundigt at oplyse deltagerne om præmisserne og konsekvenserne ved at afgive data ifm. deltagelsen. Det har derfor været en prioriteret i udformningen af samtykket til dette projekt, at det blev beskrevet grundigt, hvem der i løbet af projektet har adgang til elevernes tekster, samt på hvilken måde resultaterne af analyserne vil blive behandlet og offentliggjort efterfølgende. Eleverne har således vidst, inden de skrev de to tekster, at teksterne ikke alene var til læreren, men først og fremmest var til forskningsbrug.

3.5.2 Analysernes konsekventielle validitet

Det andet etiske fokuspunkt i projektet angår projektets potentielle didaktiske og pædagogiske konsekvenser. Dette hænger sammen med, at projektet gennem den metodiske forankring i digitale og automatiserede analysemetoder er tæt forbundet med voksende forskningsområder som 'Learning Analytics' og 'Automated Essay Scoring', som begge er områder med forholdsvis eksplicite ambitioner om at omsætte forskningsresultater til forskellige datadrevne teknologiske værktøjer til pædagogisk praksis (se fx Nouri et al., 2019; Shermis, 2014). Fx arbejdes der i flere studier, som minder om dette projekt, med at udvikle værktøjer til automatiseret elevfeedback og forskellige adaptive teknologier,

som kan stilladsere elevernes skrivning uden involvering af en lærer (fx Zhu et al., 2020). Dette åbner for en forskningsetisk problemstilling, i den forstand at den individuelle forsker kan tænkes at have et ansvar for de didaktiske og pædagogiske konsekvenser, som den pågældendes forskning potentielt kan medføre i kraft af andre interessenters anvendelse af resultaterne. Messick (1989) pointerer i denne forbindelse, at disse konsekvenser ikke kan adskilles fra gyldigheden af forskningen og foreslår derfor et forenet validitetsbegreb i den kvantitative pædagogiske forskning, der både indeholder klassiske kriterier for forskningskvalitet samt en række kriterier for forskningens konsekvenser for praksis. Sidstnævnte falder under begrebet 'konsekventiel validitet' (eng: consequential validity) og bygger på det argument, at pædagogisk forskning qua sin normative natur har indbyggede konsekvenser for den praksis, den undersøger, og derfor må disse konsekvenser indgå i valideringen af forskningsinstrumenter.

Dette projekt undgår på den ene side en del af dette ansvar, idet jeg i projektet ikke arbejder direkte med udvikling af hverken test eller læringsteknologi, der skal anvendes i en undervisningskontekst. På den anden side er det klart, at arbejdet med automatiserede analyseværktøjer, hvad enten det er i rent forskningsøjemed eller med et mere praksisudviklende sigte, er tæt forbundet med førnævnte ønsker om at udvikle automatiserede teknologier i undervisningen, der på forskellige måder kan understøtte eller endda erstatte noget af lærerens arbejde. Derfor kan det heller ikke ignoreres fuldstændig som en del af denne afhandlings horisont. Det er fx ikke svært at forestille sig en fremtid, hvor automatiserede evaluerings- og scoringsværktøjer, som netop er udviklet på baggrund af lingvistisk forskning i skriveudvikling, spiller en rolle i undervisningen og i lærerens skrivedidaktiske arbejde. Selwyn (2019) peger i denne forbindelse på en række etiske problemstillinger forbundet med denne type af data- og teknologidrevne uddannelsesforskning. Jeg vil ganske kort kommentere på særligt to af disse problemstillinger her, ligesom jeg kort vil vende tilbage til dette i den afsluttende diskussion i kapitel 7.

3.5.2.1 Risikoen for en reduceret forståelse af skrivning og skriveudvikling

Selwyn (2019) peger på, at der næppe findes datapunkter nok til, at teknologier fuldt ud vil kunne afspejle de pædagogiske fænomener, som de forsøger at afdække, og at man derigennem risikerer at komme til at "grossly underestimate the social complexity of classrooms, schools, and the complicated lives that students lead." (s. 12). Hvad angår

skrivning og skriveudvikling er det klart, at i og med de automatiserede analyser i denne afhandling kun undersøger den lingvistiske dimension af skriveudvikling, vil eventuelle digitale værktøjer til lærerne skulle tage højde for, at skriveudvikling ikke kun er en proces, der involverer udvikling af lingvistiske ressourcer, men også er en social og kulturelt approprieret proces (The New London Group, 1996), der bl.a. stiller krav til elevernes forhåndsviden og metakognitive evner (Graham, 2018). At det er lettere at operationalisere forskellige sproglige træk ved *tekstartefakter* end sociokulturelle træk ved *tekstpraksisser* må således ikke føre til, at den ontologiske forståelse af skrivning og skriveudvikling reduceres til de træk og strukturer, der er lettest at måle.

3.5.2.2 Risikoen for ensretning af elevtekster

En anden relevant problemstilling ved at arbejde med automatiseret analyse af elevtekster i en pædagogisk kontekst er knyttet til det, som Selwyn (2019) med reference til det velkendte fænomen 'teaching to the test' kalder 'teaching to the algorithm' (s. 13). Med dette udtryk refererer Selwyn til den mulighed, at et automatiseret evalueringværktøj, som er baseret på et begrænset antal datapunkter, uundgåeligt vil blive et fagligt pejlemærke for både lærere og elever. Det vil med andre ord være en mulighed, at eleverne forsøger at skrive tekster med sproglige træk, som de på forhånd ved belønnes af det automatiserede scoringsværktøj, og som ikke nødvendigvis er normativt eller kommunikativt set hensigtsmæssige. Dette hænger bl.a. sammen med, at overfladestrukturer i tekster kan manipuleres, så de scorer højt på automatiserede analyseværktøjer, men samtidig er semantisk og indholdsmæssigt meningsløse. Et berømt eksempel på dette er den såkaldte BABEL-generator¹² fra 2014, som baseret på et begrænset antal nøgleord er i stand til at generere tekster, som er indholdsmæssigt meningsløse, men som samtidig får tildelt høje scorer af automatiserede bedømmelsværktøjer (Perelman, 2020). Dette viser således, at værktøjer til automatiseret bedømmelse af elevtekster dels kan risikere at føre til ensretning af skolske tekster, dels gøre den sproglige form til det eneste succeskriterie for disse tekster. Dette er således et væsentligt etisk opmærksomhedspunkt i både dette projekt og fremtidig forskning i skriveudvikling, der involverer automatiserede analyser af elevtekster.

¹² BABEL står for Basic Automatic BS Essay Language. BABEL-generatoren kan findes her: <https://babel-generator.herokuapp.com/>

Kapitel 4: Teoretiske beskrivelser og operationaliseringer af studiets variable

I kapitel 4 beskriver jeg de variable, som indgår i studiet. Der er tale om fire grupper af variable. Først præsenterer jeg de lingvistiske variable, som danner udgangspunkt for den første analysedel (den diakrone analyse). Dernæst præsenterer jeg de variable, som er dannet på baggrund af data fra spørgeskemaet, og som indgår i anden analysedel (den synkron analyse). Her præsenterer jeg først en række askriptive variable, som er knyttet til den enkelte elevskriver. Hernæst beskriver jeg en gruppe med tre variable, som jeg kalder literacy-orienterede variable, og som handler om elevernes skrivelyst samt læse- og skrivevaner. Derefter præsenterer jeg kort de øvrige baggrundsvariable. Beskrivelserne af alle variable er både teoretisk baserede og orienteret mod de konkrete operationaliseringer i studiet. Kapitlet rundes af med overvejelser over konstruktvaliditeten i de ekstra-lingvistiske variable.

4.1 Lingvistiske variable

De lingvistiske variable falder i tre overkategorier. Først kommer en kategori knyttet til elevernes produktivitet. Dernæst en kategori med variable knyttet til det leksikale niveau i teksterne og til sidst en kategori med variable knyttet til det syntaktiske niveau.

4.1.1 Variable knyttet til produktivitet

Produktivitet er som sådan ikke en lingvistisk kategori. Det henviser ikke til bestemte grammatiske strukturer i sproget, men jeg har det alligevel med som kategori under de lingvistiske variable, fordi produktivitet dels er en egenskab (med nogen variation) ved den, der producerer sprog, dels er et pragmatisk vilkår for sproget, som det realiseres i bestemte kontekster (se fx gennemgangen af Grices kvantitetsmaksime i afsnit 2.4.1). Hvis ikke man er produktiv, er der ikke meget sprogligt output, og det er netop derfor, jeg anvender overbegrebet 'produktivitet' frem for fx mængde. Produktivitet er knyttet til eleven, hvorimod mængde er en latent og faktisk egenskab ved et stykke sprogligt output. Produktivitet understreger, at der er en kognitiv dimension forbundet med det. Elevtekster i indskoling er typisk kortere end i udskoling, og tekster fra udskoling er typisk

kortere end på universitetet og så fremdeles. Dette hænger sammen med, at vi typisk bliver i stand til at producere længere og mere sammenhængende sprogligt output med alderen, og det er også derfor, produktivetsdimensionen er en vigtig faktor at inddrage i studier af skriveudvikling. Produktivitet har én tilknyttet variabel, tekstlængde.

4.1.1.1 Tekstlængde (antal ord)

Tekstlængde måles som det totale antal ord (tokens) per tekst. 'Ord' forstås her som en sekvens af bogstaver eller cifre med mellemrum/eller interpunktion før og efter. Selvom tekstlængde er den mest simple sproglige variabel, der måles i dette studie, gør det ikke tekstlængde til et uinteressant mål. Ifølge Crossley (2020) kan tekstlængde nemlig ses som "the strongest predictor of writing development and quality" (s. 416), hvilket hænger sammen med, at tekstlængde i mange studier er blandt de variable, som bedst kan forklare tekstkvalitet¹³ (se fx Crossley et al., 2015; Hultman & Westman, 1977; Larsson, 1984) eller stigning i alder (se fx Högg, 2010; Johansson, 2009; Nordenfors, 2011). Én ting er imidlertid, at variabelen har stor empirisk forklaringskraft. En anden ting er selvfølgelig, om tekstlængde dækker over tekstuelle træk eller dimensioner af skriveudvikling.

Ifølge Hultman & Westman (1977) er længden af en tekst et udtryk for "en sprogbehærskning som ytrar sig i att man kan använda språket för att lösa den uppgift man har fått." (s. 55). Det er med andre ord elevens evne til at producere tilstrækkelig med tekst til at løse opgaven, der måles med tekstlængde. Produktivitet kan således ikke forstås som en generisk egenskab hos eleven, men som en egenskab der varierer relativt til den enkelte opgave og skrivesituation. Tekstlængde er desuden et udtryk for et udbygget meningsindhold i teksten, men som bl.a. påvises af Shi & Lei (2022) er relationen mellem tekstlængde og indholdsmæssig udbyggethed (lexical richness) ikke lineær, men derimod når relationen et punkt, hvor den leksikale dimension stagnerer, mens tekstlængde fortsætter med at øges. Tekstlængde dækker derfor ikke direkte over lingvistiske dimensioner, men må snarere ses som et tekststrukturelt træk ved elevernes skrivning, som er en forudsætning for, at eleven kan løse den stillede opgave.

¹³ Det skal dog nævnes, at det er omdiskuteret i forskningslitteraturen, hvorvidt sammenhængen mellem tekstlængde og tekstkvalitet rent faktisk skyldes en kvalitativ sammenhæng mellem de to variable, eller om det snarere er et udtryk for, at bedømmere forfordeler lange tekster (bedømmerbias). Se evt. Fleckenstein et al. (2020) for nærmere overblik.

4.1.2 Variable knyttet til leksis

Den anden gruppe af lingvistiske variable indeholder variable med fokus på det leksikale niveau i elevernes tekster. Skriveforskning med fokus på semantiske dimensioner eller leksikale forandringer i relation til skriveudvikling er forholdsvis udbredt. Dette skyldes ifølge Durrant et al. (2021, s. 117), at ordforråd dels anses for at være et vigtigt element i sprogindlæring og skriveudvikling, dels at ord er forholdsvis lette at identificere og analysere computationelt og dermed gøre til genstand for storskalaundersøgelser. Variable knyttet til dette leksikale niveau falder typisk i tre kategorier, hvoraf de to første er mere udbredt end den tredje. Den første kategori repræsenterer leksikal diversitet. Dette indebærer fx mål for ordvariation eller antallet af unikke ord (types) i en tekst. Den anden kategori repræsenterer mål for leksikal sofistisering og er fx fokuseret på lavfrekvente ord, ordlængde eller forskellige semantiske dimensioner (fx abstraktionsgrad). Den sidste kategori fokuserer på leksikal densitet og er fokuseret på frekvensen af bestemte typer af ord (fx nominelle ord) relativt til andre typer af ord (fx verbale ord). I denne afhandling undersøges to af disse kategorier, henholdsvis diversitet og sofistisering. Der er dels tale om et mål for ordvariation (OVIX) og dels om et mål af ordlængde. Begrebet 'ord' defineres i denne afhandling grafemisk, hvilket indebærer, at et ord defineres som en sekvens af bogstaver adskilt på hver side af mellemrum eller interpunktion (Grzybek, 2015).

4.1.2.1 Ordvariationsindeks (OVIX)

Det er som nævnt en standard i megen korpusforskning at undersøge den leksikale diversitet i teksterne. Det er imidlertid omdiskuteret, hvordan denne leksikale variation bedst måles. Det mest simple mål er den såkaldte type/token-ratio, som undersøger antallet af unikke ord (types) divideret med antallet af ord (tokens) i teksten. Dette mål har imidlertid været kraftigt kritiseret, idet type/token-ratioen er afhængig af tekstlængden. Det vil sige, at jo længere tekster er, desto lavere vil type/token-ratioen typisk også være. Dette hænger sammen med, at chancen for, at et ord i en tekst gentages, selvfølgelig stiger i takt med, at teksten bliver længere (dette gælder fx hyppigt forekommende konjunktioner og præpositioner som 'og', 'i' og 'at'). Ifølge Mitchell (2015) er dette imidlertid et problem, som går igen i langt de fleste mål for ordvariation. I Sverige har man forsøgt at løse problemet ved at udvikle et ordvariationsindeks kaldet OVIX. Dette indeks er udviklet af

Hultman & Westmann (1977) og er efter forfatterens eget udsagn uafhængigt af tekstlængde. Indekset er konstrueret sådan, at jo højere OVIX-værdi, desto flere forskellige ord er der i teksten relativt til tekstlængde. OVIX-målet er primært udbredt i en skandinavisk kontekst, hvor det især anvendes i vid udstrækning i svensk og til dels norsk skriveforskning (se fx Gustafsson & Lassus, 2020; Hultman & Westman, 1977; Jeppsson et al., 2012; Megyesi et al., 2019; Nordenfors, 2011; Nyström, 2000; Pettersson, 1977; Skar & Berge, 2017; Östlund-Stjärnegårdh, 2002). OVIX beregnes ud fra følgende i formel:

$$\text{OVIX} = \frac{\log(\#\text{ord})}{\log\left(2 - \frac{\log(\#\text{unikke ord})}{\log(\#\text{ord})}\right)}$$

Figur 2: Formel for beregning af ordvariationsindeks (OVIX).

Det ligger uden for denne afhandlings rammer at gå ind i en mere teknisk og statistisk diskussion af, i hvilken grad OVIX faktisk løser problemet med tekstlængde. Denne diskussion findes i mere detaljeret grad hos bl.a. Hultman (1994). Argumentet for at inddrage OVIX i dette studie er imidlertid, at det dels er et meget benyttet mål i den store gruppe af svensk forskning i elevtekster, hvilket gør det muligt at sammenligne resultaterne af disse studier med nærværende studie, og dels at OVIX i mange studier faktisk har vist interessante resultater om elevers skriveudvikling. Hultman & Westman (1977, s. 60) pointerer, at OVIX ikke skal ses som et mål for synonymvariation, men snarere som et mål for informationsrigdom. En høj værdi af OVIX indikerer således, at indholdet i en tekst er mere udbygget og indeholder en større leksikal diversitet.

Et studie af Nyström (2000) har imidlertid også udpeget en væsentlig begrænsning ved OVIX, idet hun viser, at en meget høj værdi af OVIX i en elevtekst kan indikere bristede kohærensstrukturer, og derfor kan man ikke nødvendigvis koble en høj OVIX-værdi med en høj grad af tekstkvalitet. Eftersom målet med analyserne i dette studie ikke er at måle tekstkvalitet, men i stedet er at beskrive skriveudvikling ved hjælp af automatiserede, kvantitative mål, er dette dog ikke i samme grad et problem i dette studie. Nyströms (2000) studie viste dog også, at så længe man ikke ser på de tekster, som sammenlignet med resten af datamaterialet har meget høje eller meget lave OVIX-værdier (outliers), så viste OVIX sammenhæng med forskellige former for kohærensstrukturer i teksterne. Tilsvarende har Nordenfors (2011) i en undersøgelse af elevtekster fra 7.-9. klasse vist, at

OVIX faktisk øges henover udskolingsperioden. I tillæg har et sammenlignende studie af forskellige tekstmål for læsbarhed udført af Falkenjack et al. (2013) desuden vist, at OVIX sammenlignet med de andre mål præsterede godt som proxy for læsbarhed, hvilket bekræfter Hultman & Westmans (1977) oprindelige hypotese om, at OVIX måler informationsrigdom i teksten.

Ifølge Hultman (1994) bør OVIX ikke anvendes på korte tekster under 200 ord. Dette hænger sammen med, at korte tekster naturligt har en meget høj andel af types (unikke ord), og derfor vil OVIX blive kunstigt højt, hvis det anvendes på sådanne tekster. Derfor vil analyserne af OVIX – som det eneste af de kvantitative mål i denne afhandling – blive brugt på elevernes samlede tekstproduktion. Det vil sige, at jeg har slået den beskrivende og den berettende tekst sammen til én lang tekst og lavet OVIX-analysen på denne. De tekster, som ikke når op på 200 ord ved at slå begge teksttyper sammen, indgår ikke i analysen. Dette betyder, at OVIX-analyserne baserer sig på følgende datamateriale.

Tabel 6

Deskriptiv statistik for det datamateriale, som ligger til grund for OVIX-analyserne

Klassetrin:	Antal tekster (kombineret deskriptiv og berettende tekst på mere end 200 ord)
5. klassetrin	43
6. klassetrin	41
7. klassetrin	40
8. klassetrin	69
9. klassetrin	44
Total	237

4.1.2.2 Gennemsnitlig ordlængde

Ordlængde defineres som antallet af tegn per ord. Ifølge Hultman & Westman (1977, s. 76) måler gennemsnitlig ordlængde to aspekter af elevens leksikale anvendelse; variation og specifikation. At et simpelt mål som ordlængde kan siges (i hvert fald til en vis udstrækning) at indfange disse to aspekter, har at gøre med, at de mest almindelige ord i

sproget (possessive pronomener, konjunktioner, artikler, relationelle verber etc.) typisk er korte, mens mere specifikke ord (fx proprier, sammensatte substantiver etc.) typisk er længere. Dette bekræftes af Myhill (2009), som viser, at tekster med lav ordlængde anvender simple ord, som er meget lig dem, der anvendes i talesproget, mens tekster med høj ordlængde anvender mere abstrakte ord og udfoldede verbal- og nominalfraser. En høj gennemsnitlig ordlængde i en tekst tyder altså på, at teksten indeholder en højere andel af abstrakte og mere specifikke ord. Den leksikale variation måles i denne afhandling gennem OVIX, som giver et bedre og mere detaljeret billede af variationsaspektet end gennemsnitlig ordlængde, og i denne afhandling anses gennemsnitlig ordlængde derfor først og fremmest som en udtryk for leksikal specifikation. Hos bl.a. Myhill (2009) anvendes gennemsnitlig ordlængde som en proxy for et større konstrukt kaldet 'lexical sophistication', men som det fremgår af Crossley (2020) er der inden for engelsk sprogteknologi i dag udviklet en række mere nuancerede mål for dette, som indfanger en større grad af konstruktet (bl.a. i relation til graden af leksikal abstraktion og akademisk ordvalg). Jeg har derfor valgt at følge Fellbaum (1998) og anse leksikal specifikation som et underbegreb til leksikal sofistisering.

Ordlængde har i flere studier af elevtekster vist sig at give interessante resultater. I en analyse af tekster fra de svenske nationale skriveprøver fra gymnasiet viser Palmér (2018), at den gennemsnitlige ordlængde stiger mellem 1.g og 3.g, samt at ordlængde korrelerer med tekster med høj karakterbedømmelse. Dvs. ordlængde kan skelne mellem tekster, der har fået en karakter fra den øverste del af karakterskalaen, mens der ikke er forskel i ordlængde på tekster fra den nederste del af skalaen. Dette hænger godt sammen med et andet studie af et korpus af gymnasietekster udført af Magnusson & Kokkinakkis (2011), som viser, at sammenlignet med tre andre leksikale mål (herunder OVIX) er ordlængde den variabel, som bedst kan forudsige tekstkvalitet. En række studier har dog også vist, at selvom ordlængde generelt korrelerer med tekstkvalitet (forstået som tekstens tildelte karakterbedømmelse), så varierer styrken af korrelationerne relativt til bl.a. genre og skriveordre (Hultman & Westman, 1977; Nordenfors, 2011; Vagle, 2005). Et studie af Berman & Verhoeven (2002) viser desuden en stærk sammenhæng mellem ordlængde og alder/klassetrin. Ordlængde har med andre ord vist sig at være en stærk prædikator af både skriveudvikling (det diakrone aldersperspektiv) og tekstkvalitet.

4.1.3 Variable knyttet til syntaks

Vi vender nu blikket mod den sidste gruppe af lingvistiske variable, de syntaktiske variable. Ifølge Kress (1994) er sætningen en af de væsentligste kendetegn ved skriftsproget, eftersom sætningen - modsat de mange abrupte og indskudte ytringer og gentagelser i det talte sprog - skal danne en selvstændig og sammenhængende helhed. Det er således en af de vanskeligste sproglige træk at tillære sig ved det skrevne sprog, og udvikling i en elevs syntaks er således ifølge Kress en god indikator på en mere generel udvikling i elevens skrivefærdigheder. Det er imidlertid umuligt at adressere sætningsvariable i skriveudvikling uden at inddrage Debra Myhills afgørende arbejde med at udvikle lingvistisk baserede udviklingsmodeller for skrivning med fokus på netop sætningsudvikling. Myhill (2008) undersøgte bl.a. et omfattende materiale bestående af 718 tekster fra 8. og 10. klasse. Her viste hun, at selvom den gennemsnitlige sætningslængde stiger mellem 8. og 10. klasse, falder anvendelsen af både paratakse og hypotakse relativt til både alder og skrivekompetence¹⁴, hvilket står i modsætning til studier af begynderskrivning, som viser, at denne slags syntaktiske kompleksitet stiger med alderen. Gennem kvalitative analyser viser Myhill (2008) imidlertid, at gode skrivere anvender en blanding mellem korte og lange sætninger for at opnå en retorisk effekt, mens de dårligst præsterende elever anvendte lange sætninger med klare meningsforstyrrende syntaktiske fejl. En central pointe hos Myhill er, at dygtige skribenter designer sætninger ud fra en retorisk og kommunikativ bevidsthed om læseren, mens dårlige skribenter skriver sætninger, som minder mere om talesproget. Dette indikerer således på, at kompleksitet ikke altid kan oversættes direkte med kvalitet, og at simple mål for sætningskompleksitet kan sløre for vigtig viden om skriveudvikling, hvis ikke de kombineres med kvalitative analyser.

Norris & Ortega (2009) peger i denne forbindelse på, at syntaktisk kompleksitet på den ene side er et meget anvendt konstrukt, som både anvendes inden for forskning i første- og andetsprogsudvikling, mens det på den anden side også er et omdiskuteret og bredt konstrukt, som dækker over en meget stor variation af forskelligartede mål. Blandt de mest anvendte mål er dem, Norris & Ortega kalder 'length-based measures', som er endimensionelle mål af længden af forskellige syntaktiske enheder. Grundidéen er, at

¹⁴ Myhill (2008) anvender lærerens bedømmelser af teksterne som en proxy for skrivekompetence.

øget længde af en syntaktisk enhed giver mulighed for mere komplekse sætningskonstruktioner, men mål af gennemsnitslængde – om det så er af perioder, t-units (hovedsætning med alle tilhørende bisætninger) eller forfeltet – vil altid være generiske mål for kompleksitet, da de ikke kan sige noget specifikt om, hvad den øgede længde skyldes. At den gennemsnitlige sætningslængde øges over tid, kan altså skyldes alt fra mere udbyggede nominalfraser og verbalfraser til øget brug af underordnede ledsætninger. Dette er imidlertid et vilkår i denne form for analyser, og som det anføres af Norris & Ortega (2009, s. 568) giver denne type af analyser god mening, hvis de anvendes i kombination med mere fintmaskede mål eller kvalitative analyser.

I denne afhandling anvendes to forskellige mål for syntaktisk kompleksitet: gennemsnitlig sætningslængde og gennemsnitlig dependensafstand. Disse mål dækker over to forskellige syntaktiske dimensioner. Gennemsnitlig sætningslængde er det mest grovkornede mål og fungerer som en generel proxy for kompleksitet på det syntaktiske niveau, mens gennemsnitlig dependensafstand angiver en mere specifik form for syntaktisk kompleksitet, som oftest er knyttet til frase- eller syntagmeniveauet.

4.1.3.1 Gennemsnitlig sætningslængde

Gennemsnitlig sætningslængde defineres i denne afhandling som antallet af ord i en syntaktisk periode. Med syntaktisk periode forstås en selvstændig sætning med stort begynderbogsstav i starten og med skilletegn (punktum, udråbstegn eller spørgsmålstegn) i slutningen (Jørgensen, 2000, s. 51). En syntaktisk periode består således af én eller flere hovedsætninger med alle evt. tilhørende ledsætninger. En periode er således ækvivalent med det, der i svensk grammatik meget sigende kaldes en 'grafisk mening', dvs. en meningshelhed som er grafisk adskilt af stort skilletegn (Teleman et al., 1999, s. 117). Modsat helsætningen¹⁵, som kun kan bestå af én hovedsætning plus eventuelle ledsætninger (Christensen & Christensen, 2019, s. 245), kan den syntaktiske periode altså bestå af flere hovedsætninger, hvis disse er koblet sammen med en sideordnende konjunktion (paratakse). Perioden er således den største enhed på det syntaktiske niveau. Fordelen ved at analysere syntaks på det mest generiske niveau er, at det indfanger en stor variation af

¹⁵ Helsætning forstås her ækvivalent med Hunts (1965) begreb t-unit. I den svenske skriveforskning anvendes ofte begrebet makrosyntagme (Larsson, 1984), som altså er tilsvarende det, jeg her kalder en helsætning.

forskellige syntaktiske konstruktioner i elevteksterne. Bag periodeanalysen kan der således ligge alt fra udbyggede fraser og øgede hypotaktiske konstruktioner til lange parataktiske sætninger bestående af korte hovedsætninger forbundet af sideordnende konjunktioner. Det er imidlertid også denne variation, der er periodeanalysens svaghed, idet den ikke giver information om, præcis hvad en evt. øget eller dalende sætningslængde skyldes. Information om dette kan først blive tilvejebragt, når der sammenlignes med andre typer af syntaktiske mål (i denne afhandling gælder det analyserne af dependensafstand), eller når udvalgte tekster med forskellig sætningslængde belyses kvalitativt.

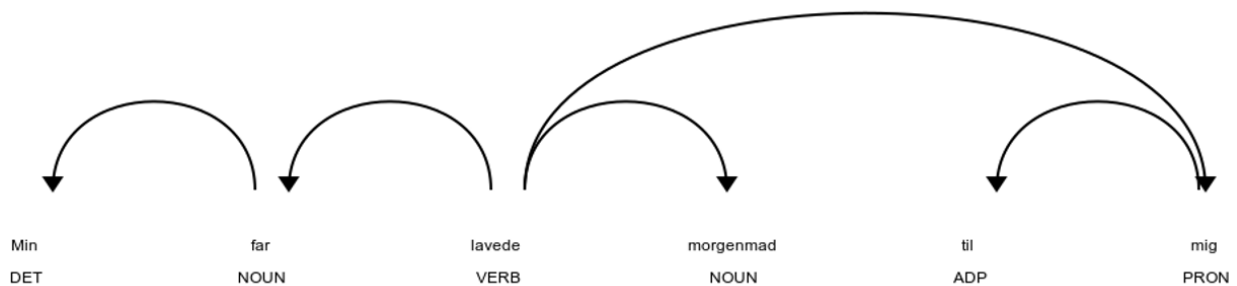
Selvom analyser af længden af syntaktiske perioder både har fordele og ulemper, er det imidlertid et ofte benyttet mål for syntaktisk kompleksitet (Crossley, 2020), ligesom det er et standardmål i det store svenske analyseværktøj SWEGRAM (Megyesi et al., 2019; Näsman et al., 2017). En væsentlig årsag til, at periodeanalysen anvendes i den korpusbaserede skriveforskning, er, at den er kompatibel med sprogteknologier inden for næsten alle sprog. Dette skyldes, at computeren forholdsvis let kan segmentere syntaktiske perioder, idet den blot skal splitte teksten ved stort skilletegn (typisk punktum). Pædagogisk set er det imidlertid også interessant at undersøge perioder, da det siger noget om elevernes evne til at inddele og sammensætte sætninger, idet meget korte eller meget lange sætningslængder siger noget om kohærens og informationsstrukturer i teksten (Ledin, 2001). Elevers skrivning har således en tendens til i starten at være orienteret mod mening frem for syntaks og regler for interpunktion, hvorfor anvendelsen af syntaktiske perioder ifølge Kress (1994) kan være en indikator på, hvordan elever i løbet af skoletiden tilpasser sig til skriftsprogets normer.

Palmér & Hussenius (kommende) viser i en analyse af elevtekster fra 6. klasse, at den gennemsnitlige sætningslængde (periodelængde) faktisk har en svag negativ korrelation med tekstens karakter, og at der generelt er stor spredning i sætningslængden inden for tekster, som har fået samme karakter. Dette kan ifølge forfatterne tænkes at hænge sammen med, at lange syntaktiske perioder er et udtryk for, at eleven ikke evner at bygge sammenhængende sætninger, og at korte sætninger ikke nødvendigvis er et udtryk for lav kompleksitet, men – i tråd med Myhills (2008) pointer - i visse tilfælde snarere kan være et udtryk for en kommunikativ bevidsthed, der kommer til udtryk gennem korte sætninger med retorisk effekt. Dette indikerer således, at sætningslængde er et interessant mål,

ikke fordi det som gennemsnitsmål nødvendigvis er et direkte mål for syntaktisk udvikling, men fordi det indfanger en stor variation af syntaktiske konstruktioner og dermed udpeger interessante fokusområder og tekster, som efterfølgende kan belyses kvalitativt.

4.1.3.2 Gennemsnitlig dependensafstand

Variablen gennemsnitlig dependensafstand har sine rødder i den såkaldte dependensgrammatik. Den grundlæggende antagelse i dependensgrammatikken er, at et sprogs syntaks udgøres af ord, som er koblet til hinanden gennem binære, asymmetriske relationer. Relationerne er asymmetriske, fordi et af de to ord i den binære relation modererer det andet ord. Det første kaldes en 'dependent', mens det andet kaldes et 'head' (Kübler et al., 2009, s. 2)¹⁶. Dette ses fx i nedenstående eksempel, som er taget fra en elevtekst i korpusset.



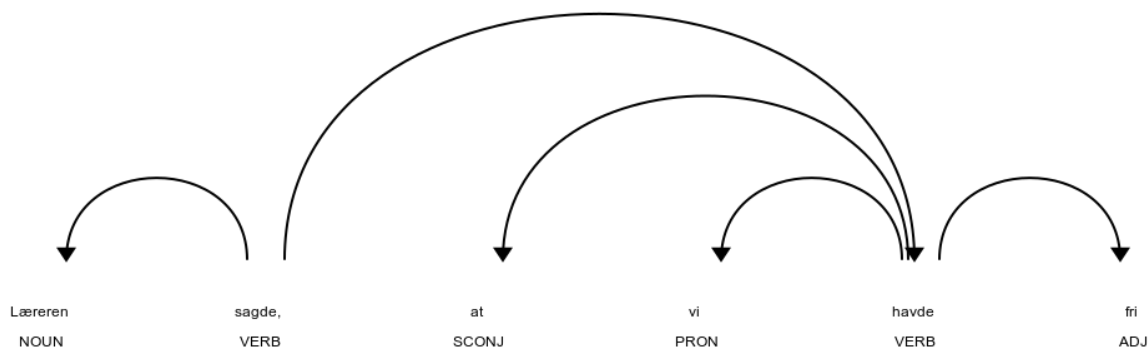
Figur 3: Eksempel på dependensanalyse af en sætning fra en berettende tekst skrevet af en elev fra 6. årg.

Som vi ser i eksemplet ovenfor er der tre heads i sætningen: 'far', 'lavede' og 'mig'. Det finitte verbum 'lavede' gøres til head, mens substantivet 'far' er dets dependent. Det vil sige, at verbet 'lavede' modereres af ordet 'far', eftersom det siger noget, hvem der udfører handlingen. 'Far' er imidlertid også et head, fordi ordet bliver modereret af det possessive pronomen 'min'. Et ord kan altså i en dependensanalyse både være 'dependent' og 'head' på én gang. Vi ser også i figuren, at verbet 'lavede' både modereres af substantivet 'morgenmad' og pronomenet 'mig'. En væsentlig faktor for dependensrelationernes udformning er således, at de er knyttet til valensbårne relationer mellem hovedverbet og bestemte syntaktiske positioner, og det er derfor ofte verberne, der er med til at afgøre,

¹⁶ Der findes mange andre begreber i litteraturen for disse to kategorier, men jeg følger her Kübler et al. (2009) og taler om 'dependencies' og 'heads'.

KAPITEL 4

hvilke dependensrelationer der er mulige (Nikula, 1986, s. 16-17). Det næste eksempel på en dependensanalyse består af en hovedsætning og en ledsætning.



Figur 4: Eksempel på dependensanalyse af en sætning fra en berettende tekst skrevet af en elev fra 5. årg.

I ovenstående eksempel er der kun to heads, nemlig de to finitte verber 'sagde' og 'havde'. Modsat eksemplet i figur 3, hvor der også var et substantiv som 'head', er det altså kun verberne, der er head her. Dette hænger sammen med, at de to nominalfraser i sætningen kun består af ét ord og dermed ikke har nogle dependents. Til gengæld ser vi i ovenstående et eksempel på, at både konjunktioner og adjektiver også kan moderere verberne. Når jeg i denne afhandling måler gennemsnitlig dependensafstand, er det et mål for, hvor langt der i gennemsnit er mellem en dependent og dets head i en hel tekst. I det første eksempel i figur 3 er den gennemsnitlige dependensafstand 1,4, mens det i det andet eksempel i figur 4 er 1,6. Begge værdier er forholdsvis lave, hvis man sammenligner med medianværdien for 5. årgang i tekstkorpusset, som er 2,29.

Dependensanalyser kritiseres ofte for, at de bliver valgt ud fra et bekvemmelighedsargument, idet de er forholdsvis let kompatible med computationelle analyser. Men ifølge Hudson (2018) er der dog en anden mere prægnant årsag til at vælge dependensgrammatikken frem for fx den klassiske Chomsky-inspirerede frasegrammatik, og denne årsag har sine rødder i en kognitiv teori om sproglig processering. Ser man på frasegrammatikken, så opdeles en sætning i et fintmasket net af relationer mellem forskellige indlejrede fraser. En sætning kan fx opdeles i en nominalfrase og en verbalfrase, som hver især så kan opdeles i en ny nominalfrase og en adverbial frase osv. Det betyder, at de samme ord optræder i flere fraser på én gang. Problemet med denne tilgang er ifølge Hudson (2018),

at menneskeret ikke processerer sprog som indlejrede frasestrukturer, men snarere som en lineær sekvens med forskellige assymetriske relationer mellem de enkelte ord. Dependensgrammatikken holder således en flad struktur, idet ordene forbindes med hinanden gennem binære relationer, hvor ordene lægger sig til og moderer hinanden, hvilket altså mimer menneskets kognitive processing af sprog. Det er således også derfor, at dependensafstande kan anvendes som indikator for, hvor svær en tekst er at læse og forstå (Liu, 2008). Dette gør ifølge Hudson (2018), at dependensstrukturer er ”better than phrase structure as a tool for representing the syntax of a corpus because it is cognitively – i.e. psychologically – more realistic” (s. 28).

Modsat længdebaserede syntaksmål, som alene siger noget længden af syntaktisk overfladestruktur, og som derfor ikke kan skelne mellem forskellige grammatiske konstruktioner, er dependensanalysen baseret på en grammatisk teori om relationerne mellem syntaktiske elementer og giver derfor information, som er lettere at forstå og fortolke fra et lingvistisk perspektiv. Dette betyder, at dependensanalysen i højere grad end fx sætningslængde siger noget om teksternes grammatiske kompleksitet (Biber et al., 2020). Flere studier har vist en moderat til stærk korrelation mellem sætningslængde og dependensafstand (se bl.a. Jiang & Liu, 2015; Oya, 2011), men som det understreges af Ouyang et al. (2022) er variansen ofte så stor, at sætninger med samme sætningslængde kan have ganske forskellige dependensafstande, hvilket indikerer, at gennemsnitlig dependensafstand opfanger noget andet ved en teksts syntaktiske kompleksitet end sætningslængde. Begge variable er imidlertid i stand til at skelne mellem forskellige færdighedsniveauer, og derfor fastslår Ouyang et al. (2022), at begge mål er relevante som indikatorer på syntaktisk kompleksitet.

4.2 Ekstra-lingvistiske variable

De ekstra-lingvistiske variable i studiet falder i to grupper; en gruppe med askriptive variable og en gruppe med tillærte literacy-orienterede variable. Distinktionen mellem de to grupper består i, at mens de askriptive variable som udgangspunkt er tildelt individet ved fødslen, så er de literacy-orienterede variable tillærte og kan i højere grad påvirkes og variere over tid.

4.2.1 Askriptive variable

Betegnelsen askriptiv stammer fra Linton (1936) og refererer til de forhold, der som udgangspunkt er arvelige og uden for den enkeltes kontrol. Linton modstiller dette med 'achieved status', som refererer til forhold, som den enkelte har opnået gennem selvstændige handlinger, eller som individet selv har valgt. Askriptive variable er således interessante både i et udviklings- og ulighedsperspektiv, fordi de siger noget om betydningen af de mekanismer, som skolen og didaktikken ikke direkte kan ændre, men kun kan kompensere betydningen af. Ved at se på hvordan disse variable manifesterer sig i elevernes tekster, bliver det muligt at se disse variables betydning på et mere detaljeret grundlag end ved fx karakterer. Modsat karakterer er de lingvistiske variable nemlig ikke afhængige af manuel bedømmelse, og det bliver derfor muligt mere direkte at se, hvordan de askriptive variable påvirker elevernes skrivning uden at skulle tage forbehold for, at forskellige former for bedømmerbias allerede på forhånd skulle have forfordelt bestemte elevgrupper. Der er to askriptive variable i dette studie, henholdsvis social baggrund og køn. Vi starter med at se nærmere på variabelen for social baggrund.

4.2.1.1 Social baggrund

Social baggrund defineres i denne afhandling ud fra elevens forældres beskæftigelse (Wattermann et al., 2016, s. 124), og variabelen er baseret på data fra spørgeskemaet til eleverne. Der er selvfølgelig begrænsninger ved et sådant indirekte mål af forældrenes beskæftigelse, som tager udgangspunkt i elevernes svar og ikke forældrenes egne svar, men som det bl.a. er vist af Vereecken & Vandegheuchte (2003), kan elevsvar være tilstrækkeligt detaljerede til at kunne bruges som grundlag for at måle forældres beskæftigelsesniveau. Variabelen er baseret på to items fra spørgeskemaet, som spurgte eleverne ind til deres forældres (både mor og far) nuværende beskæftigelse. De to items var udformet som open response-besvarelser. Forinden var eleverne blevet spurgt ind til deres forældres højeste opnåede uddannelse ved hjælp af to items, hvor eleverne skulle vælge uddannelsesniveau fra en liste (multiple choice). Det var forventet, at eleverne ville have svært ved at svare på disse to items, og derfor blev de primært inkluderet i spørgeskemaet for at have mulighed for at krydsvalidere dem med elevernes besvarelser i de to open response-items. Forældrenes uddannelsesniveau er således ikke inkluderet direkte i variabelen. Elevernes besvarelser på de to open response-items blev efterfølgende kodet ud fra

ISCO-08, som er en detaljeret, international klassificeringsramme for erhvervsgrupper (International Labour Office, 2012), og efterfølgende automatisk omkodet til "The European Socio-economic Classification" (ESeC) (Harrison & Rose, 2006), som er en mere generel klassificeringsramme. ISCO-kodningen blev dobbeltkodet af to forskere, og forskelle i koderne blev diskuteret, hvorefter den endelige kode blev tildelt. Den endelige kodning er baseret på en dominansmodel, hvor det er den forælder med den højeste rangering, der indgår i analysen. Ikke alle elevbesvarelser kunne imidlertid tildeles en kode, fordi nogle elevbesvarelser var utilstrækkelige, idet der ikke var tilstrækkelig information i hverken open response-besvarelserne eller i multiple choice-besvarelserne til, at en kode kunne blive tildelt på reliabel vis. Dette betyder, at variabelen for social baggrund baserer sig på data fra 208 elever.

ESeC-koderne falder i 10 hovedgrupper. Som det fremgår af nedenstående tabel, er samplet i dette projekt forholdsvis lille og ikke alle hovedgrupper er derfor repræsenterede, hvilket gør, at fordelingen i de 10 grupper er skæv.

Tabel 7

Oversigt over fordelingen af social baggrund i ESeC-grupper (dominansmodel)

ESeC-gruppe	Frekvens	Procent	Kumuleret
1	107	51,44	51,44
2	46	22,12	73,56
3	21	10,10	83,65
7	19	9,13	92,79
8	10	4,81	97,60
9	5	2,40	100
Total	208	100	

Jeg har derfor reduceret de 10 grupper til 2 hovedgrupper. Gruppe 1 består af hovedgruppe 1-2 fra ESeC, som repræsenterer erhverv som pædagoger, ingeniører og direktører. Gruppe 2 består af de resterende hovedgrupper fra ESeC og dækker derfor alt fra ufaglærte erhverv til håndværkere og kontoruddannede. Det, der er den afgørende forskel mellem de to grupper, er altså, at den ene gruppe ofte vil forudsætte en videregående uddannelse, mens den anden gruppe sjældent vil. Den teoretiske antagelse er således, at

elevens familiebaggrund påvirker elevens faglige udvikling i skolen og er med til at stratificere eleverne i grupper. Forældre med højere uddannelsesbaggrund eller højere beskæftigelsesniveau har dels lettere ved at støtte eleverne fagligt, dels udsættes eleverne qua deres opvækst i familien for en større mangfoldighed af sproglige input, hvilket er med til at styrke deres præstationer i skolen (Dämmrich & Triventi, 2018).

En del ulighedsforskning kigger på sammenhængen mellem social baggrund og overordnede outcomemål som karakter eller uddannelsesvalg (fx Bukodi et al., 2014; Machin & Vignoles, 2004; Sirin, 2005). Der er imidlertid ikke megen forskning, som undersøger, hvorvidt og hvordan social baggrund påvirker mere detaljerede outcomemål. PISA-undersøgelserne er en undtagelse til dette, idet undersøgelsen afdækker betydningen af socioøkonomisk baggrund for elevernes læsning, matematik og naturvidenskabelige viden (OECD, 2019), men også her er outcomevariablerne altså knyttet til mere overordnede kompetencer. I denne afhandling er outcomevariablerne i højere grad knyttet til et mikroniveau, idet jeg undersøger, hvordan social baggrund påvirker elevernes tekster helt nede på syntaks- og ordniveau. Dette er der ikke mange studier, der har gjort tidligere. Nogle af de studier, der har faktisk undersøgt denne dimension, vender jeg tilbage til i fortolkningen af analyseresultaterne i afsnit 6.4.2.

4.2.1.2 Køn

Afhandlingens anden askriptive variabel er køn. Information om elevernes køn er indsamlet via spørgeskemaet. Der deltager i alt 141 antal drenge og 156 antal piger i studiet, så der er en forholdsvis jævn fordeling af køn i datamaterialet. Køn er inddraget i dette studie af flere årsager. For det første har kønsforskelle vist sig i flere studier af skrivepræstationer (se fx Reilly et al., 2019; Scheiber et al., 2015), men ligesom med social baggrund er det imidlertid interessant at undersøge, hvorvidt og hvordan de kønsforskelle, der ses i disse overordnede performance outcomes (særligt karakter), manifesterer sig helt nede på de leksikale og syntaktiske niveauer i elevernes tekster. Dette vil således bidrage til at nuancere vores forståelse af, hvordan og i hvilken grad køn påvirker forskellige lingvistiske dimensioner af skriveudvikling i den sidste halvdel af grundskolen. Når jeg i denne afhandling undersøger kønsdimensionen af elevernes skriftsproglige udvikling, er det imidlertid vigtigt at understrege, at jeg ser på overordnede forskelle i kønnenes lingvisti-

ske præstationer. Jeg anser ikke skriftsproglig udvikling som særligt kønnede udviklingsprocesser, men jeg ser snarere på, hvordan køn påvirker de sproglige ressourcer, der er anvendt i elevernes tekster. På denne måde bidrager analyserne af kønsforskelle med vigtig viden i et differentieringsperspektiv.

4.2.2 Literacy-orienterede variable

Den anden gruppe af variable, som er baseret på data fra spørgeskemaet, kalder jeg for literacy-orienterede variable. Med denne betegnelse ønsker jeg at fremhæve to ting ved de tre variable, som falder under denne kategori. For det første dækker disse variable over forhold, som kan påvirkes af den enkelte elev eller dennes omgivelser. Modsat de askriptive variable, som er uden for den enkeltes kontrol, er de literacy-orienterede variable til lærte og kan tænkes at variere diakront henover livet. Disse variable er således tættere på det, Linton (1936) kalder 'achieved status'. For det andet dækker disse variable over forhold, som teoretisk set er meget tæt forbundet med elevernes tekstproduktion. Både skrivelyst samt læse- og skrivevaner er således forbundet med forhold og aktiviteter, som falder under det brede literacybegreb, og som derfor i et eller andet omfang må forventes at have betydning for de tekster, som eleverne producerer. Vi starter med at se nærmere på begrebet skrivelyst, og bagefter ser vi nærmere på variablen for skrivevaner og til sidst læsevaner.

4.2.2.1 Skrivelyst

Begrebet skrivelyst er omdiskuteret i forskningslitteraturen, og der er generelt ikke konsensus om, hvilke teoretiske dimensioner der ligger bag, eller hvordan begrebet operationaliseres (Latif, 2019). En udbredt skriveteoretisk tilgang til skrivelyst tager udgangspunkt i Grahams (2018) anvendelse af begrebet 'attitude towards writing' i skrivemodellen 'Writer(s)-Within-Community Model of Writing', som jeg præsenterede i kapitel 2. Denne model er ifølge Ekholm et al. (2018) særligt interessant som grundlag for empiriske studier af skrivelyst, fordi den som en af få skrivemodeller medtager skriveres emotioner som en del af en større sociokulturel model for skrivning, og fordi den tillader forskellige metodiske tilgange til at undersøge skrivelyst. Graham (2018) knytter skrivelyst til fire affektive værdier hos den enkelte skriver. Skrivelyst er for det første knyttet til en opnåelsesværdi (attainment value), som handler om skriverens ønske om at ville udføre skrivehandlingen på en kvalificeret måde. Dernæst er skrivelyst knyttet til, at skriveren

KAPITEL 4

føler en glæde ved at udføre skrivehandlingen, som har en selvstændig, iboende værdi (intrinsic value). For det tredje er skrivelyst knyttet til følelsen af, at skrivning har en nytteværdi og kan bruges til at opnå kommunikative mål (utility value), og som det sidste er skrivelyst knyttet til det, at den enkelte ønsker at tilvælge skrivning som aktivitet frem for andre aktiviteter (cost).

Skrivelyst er altså ifølge Graham (2018) et multidimensionelt konstrukt, og bag de fire dimensioner ligger en række relaterede underliggende konstrukter, fx 'self efficacy' eller 'writing beliefs'. Som det pointeres af Latif (2019) er den slags konceptuelle overlap en udfordring, fordi det gør det vanskeligt at vurdere konstruktvaliditet, og skalaer, som dækker alle dimensioner af skrivelyst, vil derfor kræve et meget stort antal items. Det har ikke været muligt inden for rammerne af dette studie at anvende et stort multidimensionelt instrument, som indfanger alle fire dimensioner af Grahams (2018) skrivelystbegreb. Derfor har jeg valgt at adaptere en valideret skala udviklet af Graham et al. (2017) med fem items. Skalaen falder dermed under den kategori, som Latif (2019) kalder "pure measures of writing attitude" (s. 4), og har den fordel, at den adresserer en meget specifik og vigtig del af skrivelystkonstruktet. Skrivelyst, som det konstrueres i den valgte skala, er nemlig primært orienteret mod den del af Grahams (2018) skrivelyst-begreb, som angår skrivningens iboende værdi (intrinsic value). Skrivelyst forstås i denne afhandling primært som et affektivt begreb knyttet til elevernes glæde ved at skrive – både i skolen og derhjemme. Skar et al. (2022), som anvender samme skala, angiver, at elevers skrivelyst påvirker "how much effort they apply and what actions they initiate" (s. 513). Dermed er skrivelyst en vigtig mekanisme for at forstå elevernes skriveproces og de sproglige valg, de træffer undervejs.

Graham et al's (2017) skrivelyst-skala er en del af et større instrument, som også indeholder skalaer for skrive-self efficacy og skrivetilgang, men i dette studie er disse skalaer valgt fra. Det er altså skalaen for 'writing attitude', som jeg har oversat til dansk og anvendt i denne afhandling. Nedenstående items har eleverne besvaret ved hjælp af en likert-skala med 5 punkter, som har værdierne 'meget enig', 'enig', 'hverken/eller', 'uenig' og 'meget uenig'.

Tabel 8*Items til skrivelystindekset*

1	Jeg kan godt lide at skrive.
2	Det er sjovt at skrive.
3	Jeg kan godt lide at skrive i skolen.
4	Jeg kan godt lidt at skrive derhjemme.
5	Skrivning er en god måde at bruge tiden på.

En eksplorativ faktoranalyse viste, at alle items loadede på over 0.7 på én samlet faktor. Chronbachs Alpha var på 0.9159, hvilket viser en meget stærk intern konsistens i de indsamlede data (se mere detaljerede faktoranalyser i bilag 4).

4.2.2.2 Skrivevaner

Skrivevaner defineres i denne afhandling som frekvensen af elevernes skriveaktiviteter i fritiden – dvs. alt det der ikke skrives i skoletiden eller som lektier derhjemme. At jeg fokuserer på skrivevaner i fritiden og ikke i skoletiden, skyldes både, at elevernes fritid er en underbelyst dimension af skriveudviklingen, samt at det er en dimension, som kan forventes at have en vis betydning for de sproglige ressourcer, som eleverne har tilgængelige. Downey et al. (2022) viser fx i et studie af kønsforskelle i matematik og læsning, at det i højere grad er ekstra-curriculære aktiviteter end skoleaktiviteter, der er årsagen til forskelle i præstationer mellem drenge og piger. Det er med andre ord i højere grad de aktiviteter, eleverne indgår i uden for skolen, end det er aktiviteterne i skolen, der gør, at piger ofte klarer sig bedre end drenge. Dette er interessant, fordi det netop peger på vigtigheden af at inddrage elevernes læse- og skrivevaner i fritiden i undersøgelser af skoleelevers skriveudvikling. Det er imidlertid en forholdsvis underbelyst dimension i skriveforskningen. En vigtig undtagelse er Camacho et al. (2021), som i et studie af elevtekster fra 5.-8. klasse fandt stærke sammenhænge mellem skrivevaner (writing frequency) og skrivemotivation, mens der omvendt ikke blev fundet stærke sammenhænge mellem skrivevaner og tekstkvalitet (et resultat forfatterne primært tilskriver et utilstrækkeligt mål

for elevernes skrivevaner). Denne sammenhæng findes imidlertid i et metanalyse af skriveundervisning foretaget af Graham et al. (2012), som dog fokuserede på skrivetid i skolen og ikke fritiden. Studiet rapporterer imidlertid om stærke sammenhænge mellem skrivetid og kvaliteten af elevernes tekster, hvilket indikerer, at jo mere eleverne skriver, desto 'bedre' tekster skriver de. Ingen af de to ovenfornævnte studier undersøger imidlertid sammenhængen mellem skrivevaner og mere tekstnære, lingvistiske variable.

Variablen for elevernes skrivevaner er baseret på 4 items fra spørgeskemaet. Disse er adapterede udgaver af fire items, som dækker elevernes læsevaner (se afsnit 4.2.2.3), og er inspireret af de skrivepraksisser, som viste sig at være mest almindelige i en undersøgelse af skrivevaner hos engelske børn og unge (Clark et al., 2020). Oprindeligt dækkede skrivevaner 7 items i spørgeskemaet, men tre af dem er taget ud. Det drejer sig om de tre funktionelle skrivekategorier (opslag på sociale medier, sms og email), eftersom elevernes besvarelser viste, at langt de fleste elever skriver inden for disse tre teksttyper meget hyppigt, hvilket ligeledes bekræftes af resultaterne fra en mindre surveyundersøgelse af Christensen et al. (2014, s. 298-300). Der vil derfor være meget lidt spredning, hvis disse tre items blev inkluderet. De tre skrivekategorier adskiller sig også fra tre af de fire tilbageværende skrivekategorier ved primært at bestå af korte kommunikative tekster og ikke længere, sammenhængende tekster som dem, der indgår i tekstkorpusset i dette studie. De fire tilbageværende items er disse.

Tabel 9

Items til variabelen 'skrivevaner'

- 1 Jeg skriver dagbog.
 - 2 Jeg skriver blogs.
 - 3 Jeg skriver historier (fx små noveller eller fanlitteratur).
 - 4 Jeg skriver digte.
-

Eleverne besvarede de fire items via en Likert-skala med svarmulighederne "Hver dag/næsten hver dag", "Flere gange om ugen", "Flere gange om måneden", "Sjældent" og "Aldrig". Elevernes besvarelser på de fire items viste en skæv fordeling med primær vægt mod svarmulighederne "Sjældent" eller "aldrig". Der er med andre ord ikke mange af de deltagende elever, som skriver inden for de fire kategorier i fritiden. Det var derfor ikke

muligt at foretage en meningsfuld faktoranalyse af dataene. I stedet har jeg manuelt sorteret elevernes svar i to grupper: én gruppe, som aldrig eller næsten aldrig skriver i fritiden, og én bred gruppe, som har svaret sjældent eller derover i mindst to kategorier. Dette er udregnet sådan, at en elev, der svarer "Aldrig" i alle fire items, får maksimumpointet 20, og en elev, der svarer "Hver dag/næsten hver dag", får minimumpointet 4. Den laveste gruppe indeholder således elever med 19-20 point, mens den øverste gruppe indeholder elever med 18 point og derunder. Fordelingen i de to grupper er som følger.

Tabel 10

Fordeling i de to grupper, som udgør variabelen 'skrivevaner'

	Frekvens	Procent
Gruppe 1 (19-20 point)	153	63,75
Gruppe 2 (4-18 point)	87	36,25

4.2.2.3 Læsevaner

Variabelen 'læsevaner' dækker over frekvensen af elevernes læseaktiviteter i fritiden, og variabelen dækker – på samme måde som skrivevaner – derfor ikke over elevernes læsning i skolen eller de læseaktiviteter, som er en del af elevernes lektier. Variabelen er baseret på syv læseitems, som jeg har adapteret fra to danske undersøgelser af læsevaner hos børn (Hansen, 2014; Hansen et al., 2017). De syv items er disse.

Tabel 11

Items til variabelen 'læsevaner'

- | | |
|---|--|
| 1 | Jeg læser tegneserier. |
| 2 | Jeg læser historier eller romaner (enten fysisk eller som e-bog). |
| 3 | Jeg læser bøger, som forklarer ting (fx om sport, idoler, dyr, mode) |
| 4 | Jeg læser blade og magasiner (fx Vi unge, Goal, Illustreret videnskab) |
| 5 | Jeg læser nyheder (på nettet eller på papir). |
| 6 | Jeg læser opslag på de sociale medier. |
| 7 | Jeg læser religiøse tekster (fx Bibelen eller Koranen). |
-

KAPITEL 4

Ligesom ved skrivevaner besvarede eleverne de syv items via en Likert-skala med svarmulighederne ”Hver dag/næsten hver dag”, ”Flere gange om ugen”, ”Flere gange om måneden”, ”Sjældent” og ”Aldrig”. Her var der en jævn fordeling i elevernes svar, hvilket gjorde det muligt at lave et indeks for læsevaner baseret på en faktoranalyse (se faktoranalyse i bilag 4). Jeg har valgt at lave et samlet indeks for læsevaner frem for fx at betragte hvert item som en selvstændig variabel. Dette skyldes, at jeg ikke er interesseret i at skelne mellem betydningen af forskellige typer af læseaktiviteter, men i højere grad er interesseret i, hvad læsning i bred forstand betyder for elevernes skrivning.

Børns læsevaner er et forholdsvis velbelyst område i dansk pædagogisk forskning, og den seneste større læsevaneundersøgelse blandt børn er gennemført så sent som i 2021 (Hansen et al., 2022). Disse undersøgelser har imidlertid et deskriptivt sigte og er orienteret mod at tegne et repræsentativt og synkront billede af børns læsevaner. Der er ikke foretaget mange studier af, hvad læsevaner betyder for elevernes faglige præstationer i skolen (hverken i en dansk eller i en international kontekst), og det er ikke blevet undersøgt, hvad læsevaner i fritiden konkret betyder for de lingvistiske ressourcer, som eleverne anvender i deres tekster. Ved at undersøge betydningen af læsevaner for sproglige træk i elevernes tekster bidrager denne afhandling således med ny viden.

4.2.3 Øvrige ekstra-lingvistiske variable

Den sidste gruppe af ekstra-lingvistiske variable består af to variable: elevernes sproglige baggrund og elevernes valg af billede i den første skriveopgave i skrivepromptet. De to variable er med som kontrolvariable i de specialiserede regressionsmodeller i den synkron analyse i kapitel 6.

4.2.3.1 Sproglig baggrund

Variablen ’sproglig baggrund’ er en binær variabel, som er baseret på et item i spørgeskemaet, hvor eleverne blev spurgt om, hvorvidt de talte et andet sprog end dansk derhjemme. De elever, som svarede ja til dette, blev desuden spurgt ind til, hvilket sprog de talte hjemme, samt hvor gamle de var, da de begyndte at lære dansk. De to sidste items blev inkluderet i surveyet for at få information, som kunne bruges til at kontekstualisere og bidrage til forståelsen af de deltagende elever. Som jeg allerede har beskrevet i forbindelse med diskussionen af repræsentativitet i tekstkorpuset i afsnit 3.2.2.3, har elevernes besvarelser på disse items vist, at der er forholdsvis få af de deltagende elever, som har

dansk som andetsprog. Ud af de 38 elever, som har angivet, at de taler et andet sprog derhjemme end dansk (svarende til 16% af de deltagende elever), har hele 15 elever angivet engelsk som det sprog, det taler derhjemme, og blot 6 af eleverne har angivet, at de først har lært at tale dansk efter de er fyldt fire år. Eftersom der er ganske få af de deltagende elever, der har dansk som andetsprog, har jeg valgt kun at inkludere 'sproglig baggrund' som en kontrolvariabel. Jeg foretager således ingen sammenligninger mellem elever med dansk som henholdsvis første- eller andetsprog, hvilket ellers er udbredt i den korpusbaserede skriveforskning (se forskningsoversigten i afsnit 1.3).

4.2.3.2 Billedvalg

Den sidste variabel er knyttet til elevernes valg af billede i de beskrivende tekster, som er skrevet på baggrund af den første skriveopgave i skrivepromptet (se bilag 3). Denne skriveopgave involverede som bekendt to billeder, som eleverne kunne vælge at skrive om. Variablen 'billedvalg' dækker således over, hvilket billede eleverne har valgt at beskrive, og variabelen er baseret på en manuel kodning, hvor de berettende tekster har fået tildelt en kode fra 1-3 baseret på en manuel screening af teksten. Koderne står for følgende.

Tabel 12

Oversigt koder til variabelen 'billedvalg'

1	Eleven har valgt at beskrive billedet med de to piger, der leger ved springvandet.
2	Eleven har valgt at beskrive billedet med den maskerede kvinde og manden, der ryger.
3	Eleven har valgt at beskrive noget andet end det, der er på billedet, eller også har det ikke været muligt at identificere billedvalget ud fra elevens tekst.

Som nævnt i afsnit 3.2.1.1 er der en del forskning, som viser, at skrivepromptet kan være afgørende for de lingvistiske ressourcer, som en elev anvender til at skrive en tekst. Crossley (2020) peger fx på, at udformningen af selve opgaveteksten kan påvirke teksterne både på det leksikale og det syntaktiske niveau, og Huot (1990) viste, at emnevariation ligeledes påvirkede sproglige træk i teksten. Det er derfor oplagt, at valget af billede i de beskrivende tekster kan påvirke elevernes tekster, fx ved at muliggøre og begrænse de

mulige leksikale valg for eleverne. Variablen 'billedvalg' er derfor oprettet, så denne prompteffekt kan holdes konstant i afhandlingens regressionsmodeller.

4.3 Overvejelser over konstruktvaliditet i afhandlingens ekstra-lingvistiske variable

Jeg runder dette kapitel af med overvejelser over konstruktvaliditeten i de ekstra-lingvistiske variable, som jeg har gennemgået i dette kapitel. Konstruktvaliditet kan defineres som "the legitimacy of the inferences we make about unobservable entities, processes, qualities or states." (Nordahl-Hansen & Kvernbekk, 2020, s. 96). Dette kan ifølge Nordahl-Hansen & Kvernbekk (2020) beskrives som et forhold mellem t-termer (teoretiske termer) og o-termer (observationelle termer). Det er med andre ord et spørgsmål om kvaliteten af sammenhængen mellem de teoretiske begreber og de empiriske observationer, man foretager af den virkelige verden, hvad enten det er kvantitative eller kvalitative observationer. Jeg har allerede i afsnit 3.4.2 diskuteret operationaliseringen af de lingvistiske variable, hvorfor jeg i dette afsnit vil fokusere på gruppen af ekstra-lingvistiske variable. Ved operationaliseringen af de lingvistiske variable skelnede jeg som bekendt mellem *direct measurement*, som angår måling af direkte observerbare fænomener (fx ordlængde eller sætningslængde), og *indirect measurement*, som angår målingen af fænomener, som ikke direkte kan observeres (fx OVIX). Den samme distinktion kan imidlertid ikke appliceres på de ekstra-lingvistiske variable, eftersom jeg har målt alle de ekstra-lingvistiske variable gennem et survey og dermed har beroet mig på elevernes hukommelse og holdninger og de biases, der følger med en sådan tilgang. Jeg vil derfor i det følgende kort kommentere på styrker og svagheder ved konstruktvaliditeten ved fire af de ekstra-lingvistiske variable: social baggrund, skrivelyst, skrivevaner og læsevaner.

Variablen 'social baggrund' er som bekendt baseret på forældrenes beskæftigelse. Som det pointeres af Waterman et al. (2016), er det meget omstridt i den sociologiske forskningslitteratur, hvilke konkrete variable der dækker begrebet, og om social baggrund i det hele taget er et uni- eller multidimensionelt konstrukt. 'Social baggrund' måles i denne afhandling som et unidimensionelt konstrukt baseret på elevernes beskæftigelsesniveau. Dette er en udbredt tilgang i den sociologiske forskning, men det er samtidig klart,

at der er begrænsninger forbundet med et forholdsvis smalt empirisk grundlag. Waterman et al. (2016) peger fx på forældrenes uddannelsesniveau og indkomst som væsentlige variable, der bidrager til at dække en større del af konstruktet. Forældrenes uddannelsesbaggrund er forsøgt målt i surveyet i denne afhandling, men som beskrevet i afsnit 4.2.1.1 er det relaterede item ikke taget med i variabelen, hvilket skyldes manglende kvalitet og detaljegrad i elevernes surveybesvarelser. Der bør derfor drages to væsentlige begrænsninger ved denne variabels konstruktvaliditet. For det første måler jeg på bare én af flere relevante variable, der dækker social baggrund. For det andet måler jeg forældrenes beskæftigelse gennem elevernes besvarelser. Der er således en potentiel bias i målingen, idet den beror på elevernes hukommelse og evne til at udtrykke sig sprogligt præcist i surveybesvarelsen. Det er derfor vigtigt at holde sig for øje, at når jeg i denne afhandling taler om social baggrund, så refererer jeg dels til et unidimensionelt konstrukt, som i andre studier måles som et større multidimensionelt konstrukt, dels til et konstrukt med den mulige bias i målingen, at den beror på elevernes hukommelse.

Jeg har allerede i afsnit 4.2.2.1 beskrevet, at også skrivelyst i denne afhandling refererer til et unidimensionelt konstrukt knyttet til elevernes oplevelse af skrivningens iboende værdi, og den affektive glæde de forbinder med at skrive både i skolen og derhjemme. Når jeg taler om skrivelyst i afhandlingen er det derfor væsentligt at holde sig for øje, at skrivelyst dækker over denne meget specifikke del af elevernes skrivelyst og ikke er knyttet til relaterede aspekter af skrivelyst som fx oplevelsen af skrivningens opnåelses- eller nytteværdi (Graham, 2018). Hvor det ved social baggrund var en klar begrænsning ved variabelen, at den baserede sig på surveydata fra eleverne, er det ved skrivelyst imidlertid en klar fordel. Dette hænger selvfølgelig sammen med, at skrivelyst netop baserer sig på elevernes holdninger, og derfor er surveyet en velegnet måde at indsamle data om dette på. Dette kan således ses som en styrkelse af konstruktvaliditeten.

Jeg vil til sidst kommentere på læse- og skrivevaner i sammenhæng. Vaner forstås i denne afhandling som hyppigheden af bestemte læse- eller skriveaktiviteter i fritiden. Der pågår imidlertid en væsentlig diskussion i forskningslitteraturen om, hvorvidt en vane teoretisk set kan sidestilles med højfrekvent adfærd. Som det pointeres af Rebar et al. (2018, s. 35), findes der en del højfrekvent adfærd i vores dagligdag, som ikke nødvendigvis kan kategoriseres som vanemæssig adfærd, fx indkøb af dagligvarer. Rebar et al. peger i stedet på, at vaner har en dybere kognitiv og motivationel dimension, og at det kræver

KAPITEL 4

en højere grad af teoretisk begrundelse, hvis en bestemt højfrekvent adfærd skal kategoriseres som en vane. Når jeg i denne afhandling refererer til læse- og skrivevaner, refererer jeg således i tråd med de danske læsevaneundersøgelser til et selvrapporteret mål af hyppigheden af forskellige læse- og skriveaktiviteter. Det kan således ses som en begrænsning af konstruktvaliditeten, at jeg ikke baserer variabelen på en mere udfoldet psykologisk teori om de forskellige kognitive og motivationelle aspekter af vaner, men alene på en behavioristisk adfærdsmodel. Mazar & Wood (2018, s. 22) peger imidlertid på, at højfrekvent adfærd er udbredt tilgang til at undersøge vaner inden for mange forskningsområder, hvilket gør det let at sammenligne med den øvrige forskningslitteratur, og desuden har højfrekvent adfærd en høj 'predictive validity', idet selvrapporterede mål af adfærd ofte viser sig at være tæt korelateret med fremtidig adfærd.

Kapitel 5: Den diakrone analysedel

5.1 Indledende rammesætning

I denne analysedel fokuserer jeg på årgangsrelaterede forskelle i de fem sproglige variable: tekstlængde, ordlængde, OVIX, sætningslængde og dependensafstand. Analysen er struktureret sådan, at jeg indleder med den kvantitative, deskriptive analyse af de enkelte sproglige variable. I denne del vil jeg særligt fokusere på, hvordan de enkelte variable udvikler sig i både centraltendens (medianværdi), spredning (IQR) samt i de højeste og laveste værdier. Jeg undersøger alle disse værdier for at kunne identificere ikke bare typiske og gennemsnitlige udviklinger, men også forandringer i den øverste og nederste gruppe af tekster. Jeg vil desuden undersøge eventuelle forskelle i værdierne relativt til teksttype (beskrivende eller berettende tekst). Analyserne er visualiseret gennem boksplotdiagrammer. Foruden boksplottet er de præcise værdier for de enkelte kvartiler samt IQR angivet i en tabel neden for hvert boksplot.

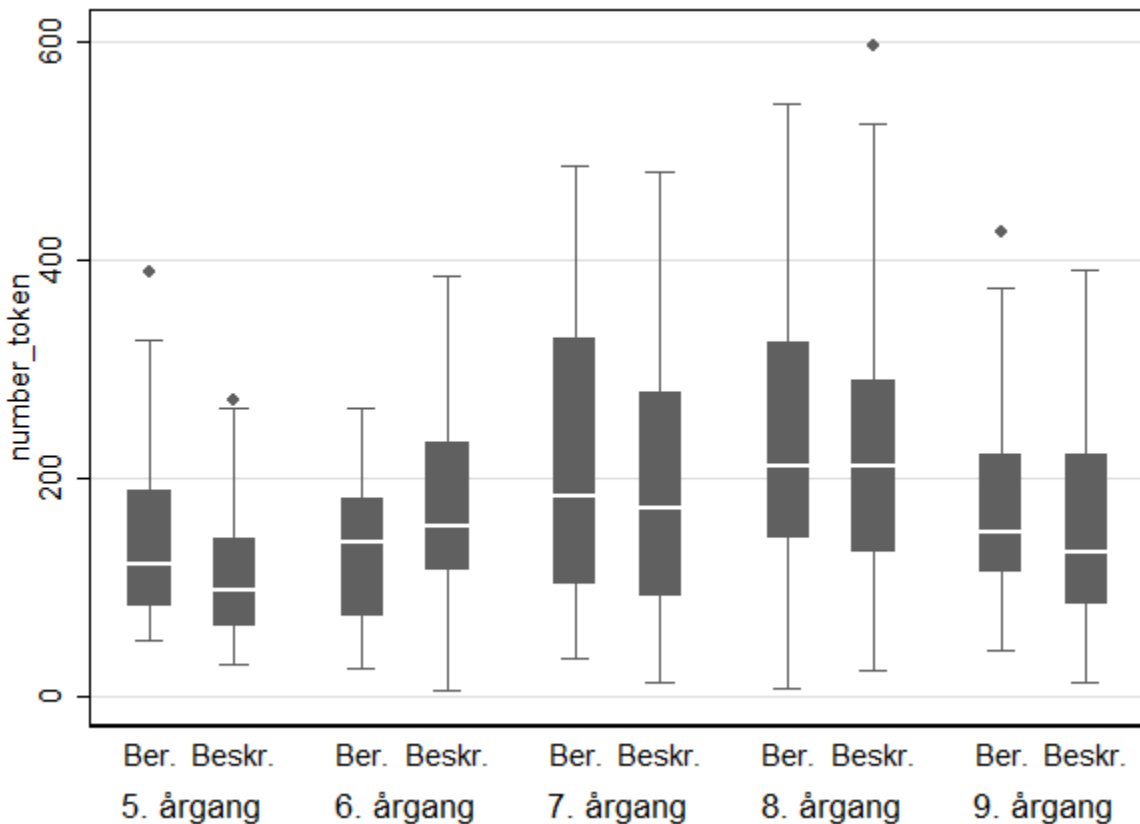
Efter analysen af de kvantitative data analyseres et udvalg af elevteksterne kvalitativt. Der er tale om tre tekstudvalg, som er knyttet til de tre grupper af lingvistiske variable. De kvalitative analyser af hvert tekstudvalg følger således i forlængelse af de kvantitative, deskriptive analyser af hver gruppe af de lingvistiske variable. Elevteksterne i de tre tekstudvalg er udvalgt på baggrund af deres variabelværdier. Der er tre grupper eller tekstpar i hvert tekstudvalg. En gruppe/par med lave variabelværdier for årgangen, en med værdier omkring medianen for årgangen og en gruppe med høje variabelværdier for årgangen. I de kvalitative analyser af disse tekster sammenlignes der på to akser: mellem den beskrivende og berettende teksttype og mellem tekster fra forskellige årgange. Teksterne i de enkelte grupper er valgt, så de bedst muligt repræsenterer begge teksttyper samt har en vis spredning ift. klassetrin. Elevens køn og sociale baggrund har ikke været en faktor i udvælgelsen af tekster.

5.2 Kvantitativ analyse af produktivetsvariabel

Den første variabelgruppe er som bekendt knyttet til produktivitet, og som den eneste gruppe er produktivetsdimensionen i dette studie baseret på én enkelt variabel, tekstlængde.

5.2.1 Variabel 1: Tekstlængde

Som det fremgår af afsnit 4.1.1.1, måles tekstlængde som det samlede antal ord i hver tekst. I nedenstående boksplot fremgår fordelingen af denne variabel for både berettende og beskrivende tekst delt efter klassetrin. Der er indsat en maksimumværdi for boksplottet på 600 ord, hvilket betyder, at fem outliers på henholdsvis 628 (8. årgang, berettende), 602 (8. årg., berettende), 706 (8. årg., berettende), 611 (9. årg., beskrivende) og 613 (7. årg., beskrivende) ikke fremgår af boksplottet. Dette er gjort af hensyn til datavisualiseringen, da disse outliers ellers ville have gjort de enkelte bokse meget små og vanskelige at aflæse.



Figur 5: Boksplot for tekstlængde i berettende og beskrivende tekst. ¹⁷

¹⁷ I alle boksplots i dette kapitel står 'Ber.' for berettende tekst og 'Beskr.' for beskrivende tekst.

Tabel 13*Deskriptiv statistik for fordelingen af variabelen 'tekstlængde' for berettende tekster*

Klassetrin	Minimum	Første kvartil	Median	Tredje kvartil	Maksimum	IQR	Antal tekster
5. årgang	51	82	121	188	389	106	39
6. årgang	26	74	141	182	264	108	35
7. årgang	34	103	184	329	486	226	39
8. årgang	7	152	211	341	706	189	69
9. årgang	42	114	151	222	426	114	42

Vender vi først blikket mod de berettende tekster, ses der en forholdsvis lineær udvikling i medianen fra 5.-8. årgang. Den stiger således mellem 20-40 ord hvert år, hvilket betyder, at den typiske tekst (forstået som datasættets central tendens) bliver omkring 15-30% længere for hvert år mellem 5.-8. årgang. 9. årgang stikker overraskende ud. Her er den typiske tekst kortere end på de to forrige årgange. Kun 5. og 6. årgang ligger lavere i medianværdien. Ser vi på udviklingen i første og tredje kvartil ses ligeledes en udvikling fra 5. til 8. årgang. I første kvartil ser vi – bortset fra et lille fald mellem 5. og 6. årgang – en forholdsvis lineær stigning indtil 8. årgang, mens man i tredje kvartil ser den største forskel mellem 6. og 7. årgang, hvor der sker en stigning på hele 207 ord. Her ses altså en tydelig forskel på tekster fra de to årgange på mellemtrinnet til de to første årgange i udskoling. Stigningerne i de to kvartiler viser, at den halvdel af teksterne på årgangen, som ligger omkring midten (altså mellem første og tredje kvartil), stiger forholdsvis systematisk indtil 8. årgang. 9. årgang stikker også ud her, idet de i begge kvartiler ligger under 7. og 8. årgang. IQR har ikke en lineær stigning, men øges dog betragteligt fra 5. og 6. årgang, hvor den ligger på 106 og 108, til 7. og 8. årgang, hvor den ligger på henholdsvis 226 og 189. Spredningen omkring midten stiger altså mellem mellemtrin og udskoling, hvilket betyder, at afstanden i tekstlængde mellem de midterste 50% af teksterne i udskoling er markant større end på mellemtrinnet. Det samme ses, hvis man ser på hele variationsbredden, dvs. afstanden fra den mindste til den højeste værd. Her er den også markant større i udskoling (7. årg = 452, 8. årg.= 699, 9. årg.= 384) end på mellemtrinnet (5. årg. = 338, 6. årg. = 238). Der er derfor generelt set større spredning i tekstlængde, jo

ældre eleverne bliver. Dette kan forklares, hvis man alene ser på minimums- og maksimumsværdierne. Minimumsværdierne for hver årgang (dvs. de korteste tekster) stiger ikke systematisk for hver årgang, mens de længste tekster derimod stiger markant mellem mellemtrin og udskoling. De længste tekster bliver altså længere, mens de korteste tekster forbliver på nogenlunde samme niveau mellem årgange.

Tabel 14

Deskriptiv statistik for fordelingen af variabelen 'tekstlængde' for beskrivende tekster

Klassetrin	Minimum	Første kvartil	Median	Tredje kvartil	Maksimum	IQR	Antal tekster
5. årgang	29	64	97	145	272	81	43
6. årgang	5	115	156	232	385	117	41
7. årgang	12	93	182	292,5	613	199,5	40
8. årgang	24	132	211	289	595	157	69
9. årgang	13	86	134,5	221,5	611	135,5	44

De beskrivende tekster indeholder mange af de samme overordnede tendenser, som vi så ved de berettende tekster. Også her ses en forholdsvis lineær stigning i medianværdien fra 5. årgang til 8. årgang. Der er tale om en stigning på 62% fra 5. til 6. årgang og herefter en stigning på 16% mellem de efterfølgende to årgange. 9. årgang bryder også tendensen her, idet der ses et fald i medianværdi mellem 8. og 9. årgang på 76,5 ord svarende til ca. 36%. Første kvartil viser ikke lineære stigninger, mens der i tredje kvartil ses en forholdsvis lineær stigning fra 5. til 7. årgang med et lille fald i 8. årgang. 9. årgang ligger konsistent under de tre forrige årgange. IQR stiger indtil 7. årgang, men falder herefter lidt indtil 9. årgang. Både 8. og 9. årgang ligger dog højere end de to årgange på mellemtrinnet. Dermed ser det også ud til i de beskrivende tekster, at der er øget spredning i teksterne, jo ældre eleverne bliver. I hvert fald er der systematisk forskel i spredningen omkring midten mellem årgange på mellemtrinnet og i udskoling. Dette ses også, hvis vi ser på hele variationsbredden. Her ligger mellemtrinnet (5. årg. = 243, 6. årg. = 280) i gennemsnit 56% under årgangene i udskoling (7. årg. = 601, 8. årg. = 571, 9. årg. = 598). Spredningen i tekstlængde er altså markant større i udskoling, både hvis man ser på IQR og på hele variationsbredden. Som vi også så ved de berettende tekster, skyldes forskellene i variationsbredden, at de korteste tekster bliver på nogenlunde samme niveau

på tværs af årgange, mens der sker en stor stigning i de længste tekster fra mellemtrin til udskoling.

5.2.2 Opsummering

Ovenstående gennemgang viser, at tekstlængde generelt ser ud til at være en god indikator for diakron skriveudvikling. Der ses lineære stigninger i medianværdierne fra 5.-8. klassetrin. Derudover er der få lineære stigninger, men der er til gengæld markante forskelle mellem mellemtrin og udskoling i mange af de andre værdier. Særligt interessant er det, at spredningen er større i udskoling end på mellemtrinnet. Dette siger således noget om vilkårene for lærernes skrivepædagogiske arbejde i udskoling, idet lærere i udskoling må arbejde med en større variation af tekstlængder end lærere på mellemtrinnet. Tekster fra 5. årgang er således mere ens i tekstlængde end tekster fra fx 8. klasse.

9. årgang afviger fra tendenserne i det øvrige datasæt på stort set alle parametre. Dette er selvfølgelig et interessant resultat, men det er omvendt også et resultat, der er svært at forklare. Der er imidlertid mindst to mulige forklaringer på denne afvigelse. Den første er en praktisk forklaring. 9. årgang er generelt en travl årgang med stort fokus på afgangsprøven, og derfor er det ikke utænkeligt, at eleverne har nedprioriteret denne skriveopgave, da den ikke er initieret af deres dansklærer, men af mig som forsker. Fra dette perspektiv vil man altså sige, at 9. årgangs afvigelse er et resultat af praktiske omstændigheder og dermed til en vis grad tilfældigt, og derfor vil en anden case kunne give et andet resultat. En anden mulig forklaring kan findes, hvis man sammenligner med lignende undersøgelser i forskningslitteraturen. I Hultman & Westmans (1977) undersøgelse af gymnasistile viste de, at tekstlængde var bedre til at differentiere mellem tekster på de laveste karakterniveauer end på de højeste karakterniveauer. Tekstlængde var altså en større faktor for tekstkvalitet ved de dårligste tekster end ved de bedste. Samme resultat findes også hos Vagle (2005), hvilket indikerer, at de dygtigste elevskrivere altså ikke skriver kontinuerligt længere tekster, jo højere karakter de får. Til gengæld skriver de lidt kortere tekster, som til gengæld har større sproglig præcision. Nordenfors (2011) viser desuden i et studie af elevtekster fra 5.-9. klassetrin, som dermed er meget sammenligneligt med korpuset i denne afhandling, at udviklingen i tekstlængde stagnerer over tid. I Nordenfors' korpus opstår denne stagnering omkring 8. klassetrin. Sammenholdt med resultaterne fra disse forskningsprojekter kan forklaringen på 9. årgangs afvigelse altså

være, at stagnering i tekstlængde i sidste del af grundskolen simpelthen er et tegn på skriftsproglig udvikling. Tekstlængde er altså i denne optik ikke en kontinuerlig funktion af alder, men når i stedet et mætningspunkt, hvorfra det enten stagnerer eller falder en smule.

Vi vender til sidst blikket mod ligheder og forskelle i tekstlængde i de to teksttyper. De to teksttyper ligger meget tæt på hinanden, hvis man ser på den gennemsnitlige medianværdi på tværs af årgange. Her ligger de berettende tekster på 161 ord, mens de beskrivende tekster ligger på 156 ord. Ser man på IQR, ligger de berettende tekster i gennemsnit lidt højere (147) end de beskrivende tekster (138). Variationsbredden er til gengæld gennemsnitligt set højere ved de beskrivende tekster (458) end ved de berettende tekster (422). Man kan dermed sige, at den typiske berettende tekst er lidt længere end den typiske beskrivende tekst, ligesom den berettende tekst har lidt større spredning omkring midten. Til gengæld er der størst spredning ved de beskrivende tekster, hvis man ser over hele datasættet.

5.3 Tekstudvalg 1: Produktivitet

Tekstudvalget med fokus på produktivitet består af fem tekster. Eftersom produktivitet ikke er en egentlig lingvistisk variabel, er der ikke meningsfulde, bagvedliggende sproglige strukturer, som kan belyses kvalitativt. Jeg vil i stedet i fokusere på, hvordan tekstlængde hænger sammen med indholdsmæssig variation (semantiske felter) i teksten samt tekstlængdens relation til opgaveformuleringen i skrivepromptet. Tekstudvalget består af to tekster med meget lav tekstlængde fra henholdsvis 9. og 6. årgang, to tekster med tekstlængde omkring medianen fra henholdsvis 5. og 8. årgang samt en enkelt tekst med meget høj tekstlængde fra 8. årgang. Teksterne gengives i oprindelig, uredigeret format (dog er der af pladmæssige hensyn fjernet linjeskift ved afsnitsinddelinger), hvilket betyder, at der kan være diskrepans mellem den værdi, der angives for variabelen og den konkrete tekst, eftersom de lingvistiske analyser er foretaget på de standardiserede tekster. Analysen af de semantiske felter er visualiseret gennem farvekoder, som er lagt henover ordene i teksten.

5.3.1 Elevtekst 1 og 2

Elevtekst 1 og 2 er eksempler på elevtekster med meget lav tekstlængde. Den første elevtekst er en beskrivende tekst skrevet af en elev fra 9. årgang. Teksten har en længde på bare 17 ord.

På billedet er der to piger. Den ene pige bærer en maske. De leger ved et springvand.

Elevtekst 1: Beskrivende tekst skrevet af en elev fra 9. årgang.

Teksten beskriver nøgternt og kortfattet de væsentligste elementer på billedet¹⁸; to piger, én med maske, de leger i et springvand. Teksten følger en klassisk informationsstruktur med en indledende rammesætning ('På billedet'), hvorefter de to piger introduceres som tekstens diskursreferenter, som herefter kan agere tema, og de to mest påfaldende træk ved billedet, nemlig at den ene pige har maske på, og at de leger ved et springvand, er rema. I løbet af de tre sætninger trækker eleven på fem semantiske felter. Det første felt udgøres af præpositionssyntagmet 'På billedet', som her henviser til en metafunktionelt semantisk felt, der peger ud på opgaveteksten (rød). Hernæst er der et semantisk felt om pigerne, som indeholder nominalfraserne 'to piger' og 'den ene pige' (grøn). Det tredje felt handler om sundhed og udgøres af substantivet 'maske' samt verbet 'bærer' (blå). Eleven udvider ikke dette semantiske felt ved fx at koble det til det skrivepromptets overskrift ("Sundhedskrise i verden"). Det fjerde felt handler om aktiviteterne på billedet og består af verbet 'leger' (lyserød). Det sidste felt handler om fysiske objekter i omgivelserne og indeholder præpositionssyntagmet 'ved et springvand' (orange). Eleven nævner kun springvandet og undlader altså at inddrage den øvrige fysiske kontekst på billedet. Selvom eleven lykkes med at inddrage en vis variation af semantiske felter, er der imidlertid tale om meget små felter, som består af ganske få ord.

På den ene side kan man sige, at eleven løser opgaven i skrivepromptet, idet teksten i hvert fald til en vis udstrækning beskriver situationen på billedet. På den anden side er det klart, at teksten også er indholdsmæssigt mangelfuld og dermed også sprogligt meget simpel. At teksten ikke er længere, kan selvfølgelig skyldes flere ting. For det første kan det skyldes et træk ved det valgte billede, som ikke indeholder en tilstrækkelig variation og detaljeringsgrad til at generere lange tekstbeskrivelser. Dette kan dog ikke være den

¹⁸ Billedet med de to legende piger fremgår af skrivepromptet i bilag 3.

eneste faktor, eftersom der er en forholdsvis stor variation i tekstlængde på tværs af de beskrivende tekster. For det andet kan en forklaring være, at eleven har oplevet, at opgaven er for let og simpel og dermed demonstrativt har skrevet en kort tekst for at udtrykke en form for modstand mod opgaveformuleringen i skrivepromptet. En mindre sandsynlig forklaring er, at eleven kognitivt set ikke er i stand til at producere mere tekst, hvilket dog er usandsynligt elevens alder taget i betragtning. At den lave tekstlængde ikke skyldes en kognitiv begrænsning hos eleven, men snarere førnævnte omkringliggende faktorer, understreges ligeledes af, at samme elev har skrevet en berettende tekst på 281 ord, hvilket placerer teksten i det fjerde kvartil, dvs. blandt de 25% længste på årgangen.

Den anden elevtekst er en berettende tekst skrevet af en elev fra 6. årgang. Teksten har en længde på 26 ord.

Jeg lavede ikke rigtig noget jeg sad enlig bare og spillede hele dagen lang på min computer. Jeg spillede mest Roblox men nogle gange også Minecraft

Eleve tekst 2: Berettende tekst skrevet af en elev fra 6. årgang.

Teksten beskriver på et ganske overordnet niveau nogle enkelte aktiviteter, som eleven har foretaget under nedlukningen. Der etableres tre semantiske felter i teksten. Det første felt handler om agenten (jeg-fortælleren) og dennes generiske handlinger (rød). Feltet udgøres således af pronomerne 'jeg' og 'min' samt verbet 'lavede' og det intransitive verbum 'sad'. Jeg knytter disse verber til samme semantiske felt som pronomerne, fordi verberne ikke knyttes an til bestemte handlinger inden for andre semantiske felter, men i højere grad knytter sig til egenskaber ved tekstens jeg-fortæller. Det andet semantiske felt handler om tid og udgøres af adverbialfrasen 'hele dagen lang' (grøn). Det er dette semantiske felt, som binder tekstens aktiviteter sammen, og som derfor også giver eleven mulighed for at afslutte teksten hurtigt, eftersom de få aktiviteter, der nævnes i teksten, altså ifølge eleven dækker et helt dagsforløb. Det sidste semantiske felt handler om computerspil (blå) og er det største af de tre felter. Her indgår både handleverberne 'spillede', substantivet 'computer' samt de to proprier 'Roblox' og 'Minecraft', som konkretiserer og udbyder de andre ord i feltet.

Som ved elevtekst 1 ser vi også denne gang, at eleven lykkes med at inddrage en vis indholdsmæssig variation (antallet af semantiske felter), men til gengæld udbygges disse felter ikke i særlig høj grad, ligesom det også i denne tekst ses, at eleven på sin vis lykkes

med at besvare opgaveformuleringen i skrivepromptet på trods af den korte tekstlængde. Eleven har til gengæld skrevet en beskrivende tekst på 213 ord, hvilket placerer den i det tredje kvartil, dvs. i den øverste halvdel på årgangen. Det er derfor også sandsynligt, at eleven simpelthen har valgt en skrivestrategi, der indebærer, at der bruges mere tid på den beskrivende tekst og som følge heraf nedprioriteres den berettende tekst.

5.3.2 Elevtekst 3 og 4

Elevtekst 3 og 4 er eksempler på elevtekster med en tekstlængde, der ligger omkring medianen for deres respektive årgange. Jeg har valgt en tekst fra 5. årgang samt en tekst fra 8. årgang for at få en aldersmæssig variation imellem teksterne. Den første elevtekst er en berettende tekst skrevet af en elev fra 5. årgang. Teksten har i sin standardiserede version en tekstlængde på 111 ord.

Under corona nedlukningen da jeg var hjemme skulle jeg lave lektier i ca. 2 timer, og så måtte jeg gå ud og holde pause i ca. 20 min., bagefter skulle jeg lave lektier igen i ca. 1 og en halv time. Da jeg havde det måtte jeg gå ud i et kvarter. Derefter måtte jeg holde "fri" og lave det jeg ville. Jeg ringede tit til mine venner og spillede Roblox med dem. Og efter det gik jeg udenfor og stod på rulleskøjter, skateboard, løbehjul og cyklede. Eller så hoppede jeg på trampolin eller var sammen med mine søskende. Det var lidt kedeligt at vi skulle være hjemme så langt tid.

Elevtekst 3: Berettende tekst skrevet af en elev fra 5. årgang.

Teksten trækker på fem semantiske felter. Det første semantiske felt er knyttet til tid (blå) og består af tidsadverbialer (fx bagefter, da, derefter), som fungerer som kohæsive bånd i teksten, samt mere specifikke tidsangivelser ved brug af numeralier og nominalfraser (fx 'langt tid'). Det andet felt knytter sig, ligesom vi så ved elevtekst 2, til tekstens agent (jeg-fortæller) og dennes generiske handlinger (lyserød). Dette er imidlertid ikke overraskende, eftersom både elevtekst 2 og 3 er berettende tekster. Det tredje felt handler om nedlukningssituationen (grøn) og består af et substantiv, to stedsadverbialer og et adjektiv. Det fjerde semantiske felt er forholdsvis lille og handler om skolen (rød), mens det sidste semantiske felt, som også er det største, er knyttet til elevens fritidsaktiviteter med vennerne (orange). Det sidste felt består mestendels af en række aktive handleverber (spillede, cyklede, hoppede), som knytter sig til konkretiserende substantiver (rulleskøjter, skateboard, løbehjul etc.) og et enkelt proprium (Roblox).

Det interessante ved denne tekst er, at selvom vi har med en tekst at gøre, som er over fire gange så lang som elevtekst 2, ser vi ikke en tilsvarende stigning i antallet af semantiske felter. Til gengæld ser vi en væsentlig udvidelse af felterne. Eleven formår således i denne tekst dels at udvide tekstens indhold med to yderligere semantiske felter knyttet til nedlukningssituationen og skolen, dels at udvide de tre semantiske felter, som de to tekster deler (tid, jeg-fortællers generiske handlinger og fritidsaktiviteter), med en række yderligere ord. Særligt det semantiske felt knyttet til tid udvides, idet ordene i dette felt fungerer som en strukturerende princip i teksten. Vender vi blikket mod den pågældende elevs beskrivende tekst, har den en længde på 64 ord, hvilket placerer den i det første kvartil på årgangen. Der er altså, som vi også har set ved elevtekst 1 og 2, tale om en skæv fordeling mellem de to tekster i den forstand, at den ene tekst er markant længere end den anden, både hvis man sammenligner teksterne imellem og med medianen for årgangen.

Den fjerde elevtekst er en beskrivende tekst skrevet af en elev fra 8. årgang. Teksten har i sin standardiserede version en længde på 218 ord.

På billedet kan jeg se to mindre piger, som skal til at lege med nogle springvand. Man kan se at pigen med de røde shorts helt sikkert har tænkt sig at gå derind og lege. Det kan man fordi at hun har smidt både sine sko og strømper. Til gengæld så kan man se at pigen oven på stenen allerede har været i vandet, fordi at hendes bukser er våde og hun er igang med at tage sine strømper på. Man kan også se at på daværende tidspunkt at covid-19 allerede er udbrudt. Det kan ses ved at pigen oven på stenen sidder med et face shield på, som er et alternativ til de normale face mask. Disse face shields er også til for at beskytte dig mod covid-19. Pigen i de røde bukser virker begejstret, fordi at hendes hår er helt omkuldt og det viser at hun har lavet en hurtig vending. Så jeg tænker at pigen, i de røde bukser, er storesøsteren der har prøvet at få sin lillesøster, som er ked af over at hun fik våde bukser, til at komme med ud at lege med hende. Jeg ville sige at stemningen på billedet er munter fordi at lillesøsteren ser lidt sådan ud, som man nu gør når man er lidt træt af det hele.

Elevtekst 4: Beskrivende tekst skrevet af en elev fra 8. årgang.

Teksten består af fire semantiske felter. Det første felt er knyttet til elevens vurderinger af forskellige forhold på billedet og består bl.a. af et reflekterende 'jeg' eller brugen af 'man' til mere almene vurderinger (blå). Feltet indeholder således en bred variation af både præpositionssyntagmer, personlige og indefinitte pronomener, substantiver, verber og adjektiver. Det andet felt handler om de to piger samt deres handlinger og almindelige beklædning (rød). Der er tale om det proportionelt set største semantiske felt i teksten.

KAPITEL 5

Jeg inkluderer kun pigernes almindelige beklædning i dette felt, da jeg i stedet reserverer ansigtsmasken til det tredje semantiske felt, som knytter sig til selve coronapandemien (**orange**). Her inkluderer jeg både konkrete substantiver knyttet til pandemien (face mask, covid-19 etc.), tidsadverbialet 'på daværende tidspunkt' samt verberne 'udbrudt' og 'beskytte'. Det sidste felt er knyttet til konkrete fysiske objekter i omgivelserne på billedet (**lyserød**). Dette felt er forholdsvis lille og består fortrinsvis af konkrete substantiver som 'stenen', 'vandet' og 'springvand'.

Elevtekst 4 har færre semantiske felter end elevtekst 1 til trods for, at elevtekst 4 er næsten tretten gange så lang som elevtekst 1. Som vi også så ved elevtekst 3, hænger den øgede tekstlængde altså i højere grad sammen med et udbygget meningsindhold (større semantiske felter) i teksterne end med øget indholdsvariation (flere semantiske felter). I denne tekst hænger denne udvidelse af felterne ikke kun sammen med, at der er mere beskrivende indhold, men også at flere af disse beskrivelser begrundes. Der er med andre ord tale om, at der i denne tekst introduceres en form for variation i fremstillingsformerne (Togeby, 2014, s. 97-125) – dvs. at der veksles mellem beskrivende og argumenterende tekst. Sammenligner man længden af elevens to tekster, ses det, at mens den beskrivende tekst altså ligger omkring medianen for årgangen, har eleven til gengæld skrevet en berettende tekst på 164 ord. Dette placerer teksten i andet kvartil for årgangen. Her er tekstlængden på de to teksttyper altså også skævt fordelt, om end forskellen ikke er så stor, som ved de tre forrige elever.

5.3.3 Elevtekst 5

Elevtekst 5 er den længste elevtekst i hele korpuset. Det er en berettende tekst skrevet af en elev fra 8. årgang. Teksten har en længde på 706 ord.

I starten vågnede jeg pga min alarm sat til kl 06.00, men efterhånden som jeg mistede motivationen skiftede jeg alarmen ud med min mor. Hun kom altid ind på mit værelse et minuts tid før mit første meet for at vække mig. De fleste dage kom jeg lidt for sent, men af en eller anden grund var jeg begyndt at være ligeglad. Jeg kan stadig ikke gennemskue hvorfor, for jeg har altid været meget præcis omkring at nå ting til tiden. Jeg plejede at flippe helt ud hvis jeg nu havde glemt at lave en opgave, men det var som om der var noget andet der overtog den slags bekymringer. Under det første meet slog jeg kameraet fra og gjorde mig klar, mens jeg lyttede til de andre sidde og diskutere en opgave. Det var urealistisk at sidde til time hjemme på mit eget værelse. Der var en vis frihed ved det, men det var ikke altid positivt. Jeg mistede koncentration 24/7 og kunne aldrig følge med. Efter første meet kom frikvarteret hvor jeg som regel tog hjem til min bedste veninde fra min klasse. Vi var sammen stort set hele tiden. Jeg boede en uge hos hende og hende en uge hos mig. Vi sad så sammen på meet og lavede opgaverne sammen. Hun var den der motiverede mig til ikke bare at sove hele dagen. Vi havde det rigtig sjovt med det vi lavede af opgaver. I andet frikvarter gik vi en tur. Vi gik som regel forbi skolen eller i netto for at købe snacks. Det var sjovest når vi lige mødte en bekendt. På de mere dovne dage så vi netflix og spiste alt hvad der var i huset. Vi har stadig utroligt mange serier vi begyndte på sammen som vi endnu ikke har set færdige. I de sidste tre lektioner tror jeg nok vi lyttede 5% af tiden hvor vi fik af vide hvordan man skulle løse opgaverne vi så fik for senere. Den tid vi fik til at lave opgaven brugte vi af samme grund til at opklare hvordan opgaven skulle løses. Det gode var at så snart vi faktisk havde fokus på opgaven var vi meget effektive og nåede næsten i mål med det hele inden skoledagen var slut. Ikke et minut senere end skoledagen var slut tog vi op på skolen og sad og snakkede. Skolen blev til et helt andet sted end hvad den plejede at være. Det var hyggeligt bare at sidde på ramperne og snakke uden små børn der løb rundt og larmede over det hele. Vi brugte lang tid udenfor huset efter skole, for det var uudholdeligt at sidde i det sammen hus på det samme værelse så mange timer i streg. Det var som om skolen og hjemmet skiftede roller. Hjemmet blev et sted for koncentration og fokus, mens skolen blev et stille og afslappende sted. Det ændrede også min holdning til skolen nu. Jeg har det generelt mere afslappet på skolen nu. Det var så det vi brugte vores eftermiddag-og aftentimer på. Vi stoppede først når "Vi spiser om 10" beskeden fra forældrene tikkede ind. Vi tog hjem og spiste og snakkede lidt om vores dag med forældrene. De virkede meget uinteresserede i hvad vi gik og lavede eftersom de havde deres egne ting at få styr på, men det kunne vi selvfølgelig godt forstå. Efter vi havde spist gik vi ind på værelset og så nickelodeon bare for mindernes skyld. Vi sad også lidt på vores telefoner. Sådan gjorde vi i et par timer, indtil klokken blev omkring 23.00. Der gik det op for os at vi ikke havde lavet lektier til den næste dag, så vi satte os ud ved spisebordet og lavede lektier til meget sent. Af en eller anden grund koncentrerede vi os allerbedst ved midnat. Det var også lidt hyggeligt at sidde der i mørket med kun lidt lys fra stearinlys og computerskærmen. Vi sad tit og drak kakao eller te bare for hyggens skyld. Da vi endelig blev færdige med lektierne gik vi ind på værelset og snakkede i ca 10 min. Efter det gik vi seng og dagen efter stod på akkurat det samme.

Elevtæst 5: Berettende tekst skrevet af en elev fra 8. årgang.

Teksten består af otte semantiske felter, hvoraf de fleste er forholdsvis store. Det første felt er knyttet til tid (**rød**) og er spredt ud over hele teksten. Det andet felt handler om familien (**lyserød**) og er det mindste af de semantiske felter. Det består fortrinsvis af substantiver (mor, forældre) og personlige pronominer. Det tredje felt er nyt semantisk felt sammenlignet med de to forrige berettende elevtekster (elevtekst 2 og 3). Det knytter sig til følelser, affektive vurderinger (sjovest, utroligt, uudholdeligt, allerbedst etc.) og generelle kognitive processer (koncentration, motivationen etc.) (**mørkeblå**). Dette viser således, at der i denne tekst – modsat de to andre – i højere grad er en jeg-fortæller, som forholder sig vurderende og evaluerende til de hændelser og personer, der indgår i beretningen. Det fjerde semantiske felt er også knyttet til jeg-fortælleren, men består i højere grad af jeg-fortælleren generiske aktiviteter og relationer til forskellige objekter udtrykt gennem possessive pronominer (**mørkegrøn**). Det femte felt er forholdsvis lille og er knyttet til fysiske og digitale objekter i beretningen (**gul**). Her er således konkrete substantiver som 'computerskærmen' og 'stearinlys' samt proprier som 'netflix' og 'nickolodeon'. Det sjette semantiske felt handler om konkrete steder i teksten (**orange**) og er på samme måde som det semantiske felt om tid spredt ud over hele teksten. Feltet består ikke overraskende primært af præpositionssyntagmer (ind på værelset, udenfor huset etc.) og stedsadverbier som 'hjem' og 'hjemme'. De to sidste semantiske felter handler om henholdsvis veninden og deres fælles aktiviteter (**lyseblå**) samt skole og undervisning (**lysegrøn**).

Vi ser altså i denne tekst en fordobling i antallet af semantiske felter i forhold til elevtekst 4. Der er dermed tale om en betydelig større indholdsmæssig variation i denne tekst, ligesom der også sker en udvidelse af meningsindholdet, idet flere af de semantiske felter bliver meget store. Eleven har valgt ikke at indlevere en beskrivende tekst, hvilket betyder, at hun har kunnet allokere al tid og kræfter i denne tekst i stedet. Selvom elevtekst 5 som enkeltstående tekst er den længste i korpusset, er eleven således ikke den, der har produceret mest tekst, hvis man kombinerer den beskrivende og berettende tekst.

5.3.4 Opsummering på analyse af tekstudvalg 1

Analysen af det første tekstudvalg af elevtekster indikerer en sammenhæng mellem et udbygget meningsindhold og tekstlængde. Selvom jeg anerkender, at der er store begrænsninger i generaliserbarheden af et studie på bare fem elevtekster, ligesom analysen af se-

semantiske felter har metodiske begrænsninger ift. at afgrænse de semantiske felter fra hinanden på en transparent og stringent måde, så viser analyserne i hvert fald, at indholdsvariation (antallet af semantiske felter i teksten) ikke er en direkte funktion af tekstlængde, mens snarere at tekstlængde i de fleste tilfælde vil betyde øget indholdsudbygning (størrelsen på de semantiske felter). Dette skyldes bl.a., at de lange tekster også har større variation i fremstillingsformerne (særligt i de beskrivende tekster). Det kan man bl.a. se, hvis man sammenligner elevtekst 4 og elevtekst 1. Her er der det samme antal semantiske felter, selvom den ene tekst er tretten gange længere end den anden. Forskellen på de to tekster ligger i størrelsen på felterne. Elevtekst 5 viser til gengæld, at et udbygget meningsindhold også kan være en katalysator for større indholdsvariation, idet de semantiske felter kan blive tilpas store til, at de enkelte felter kan deles op i flere. Jeg har i disse analyser fokuseret på forholdet mellem indhold og tekstlængde, men det skal selvfølgelig fremhæves, at øget indhold blot er ét aspekt af øget tekstlængde. Et andet vigtigt aspekt, som jeg ikke har undersøgt i dette tekstudvalg, er det syntaktiske aspekt, som selvfølgelig er en mindst lige så vigtig faktor for tekstlængden. Jeg har imidlertid valgt at reservere det syntaktiske aspekt til analysen af tekstudvalg 3, hvor fokus er på forholdet mellem værdierne i de to syntaksvariable og forskellige kvalitative observationer af elevernes syntaktiske valg i teksterne.

Vender vi blikket mod analysen af de to korte elevtekster (elevtekst 1 og 2), viser den, at korte tekster ikke nødvendigvis er et resultat af kognitive begrænsninger hos eleven, men snarere er et udtryk en bevidst skrivestrategi fra elevens side, hvor eleven har prioriteret at bruge tid på at skrive den ene af de to tekster, og som følge heraf har nedprioriteret at skrive den anden. Dette understøttes af det forhold, at der ved de fire første elever, som er inkluderet i dette tekstudvalg, generelt er en skæv fordeling i tekstlængde mellem de to teksttyper.

Den sidste pointe, som kan udledes af analysen af tekstudvalget, angår relationen mellem teksttype og de semantiske felter. Ikke overraskende viser analysen nemlig, at bestemte teksttyper fordrer bestemte semantiske felter. De to beskrivende tekster, som indgår i tekstudvalget, indeholder begge semantiske felter knyttet til de agenter, der fremgår af billedet, de fysiske omgivelser på billedet samt et metafunktionelt semantisk felt, hvor eleven træder frem som beskriver ('På billedet', 'man kan se', 'jeg tænker' etc.). Ved de berettende tekster indgår tid som et semantisk felt i alle tre beretninger, mens det kun

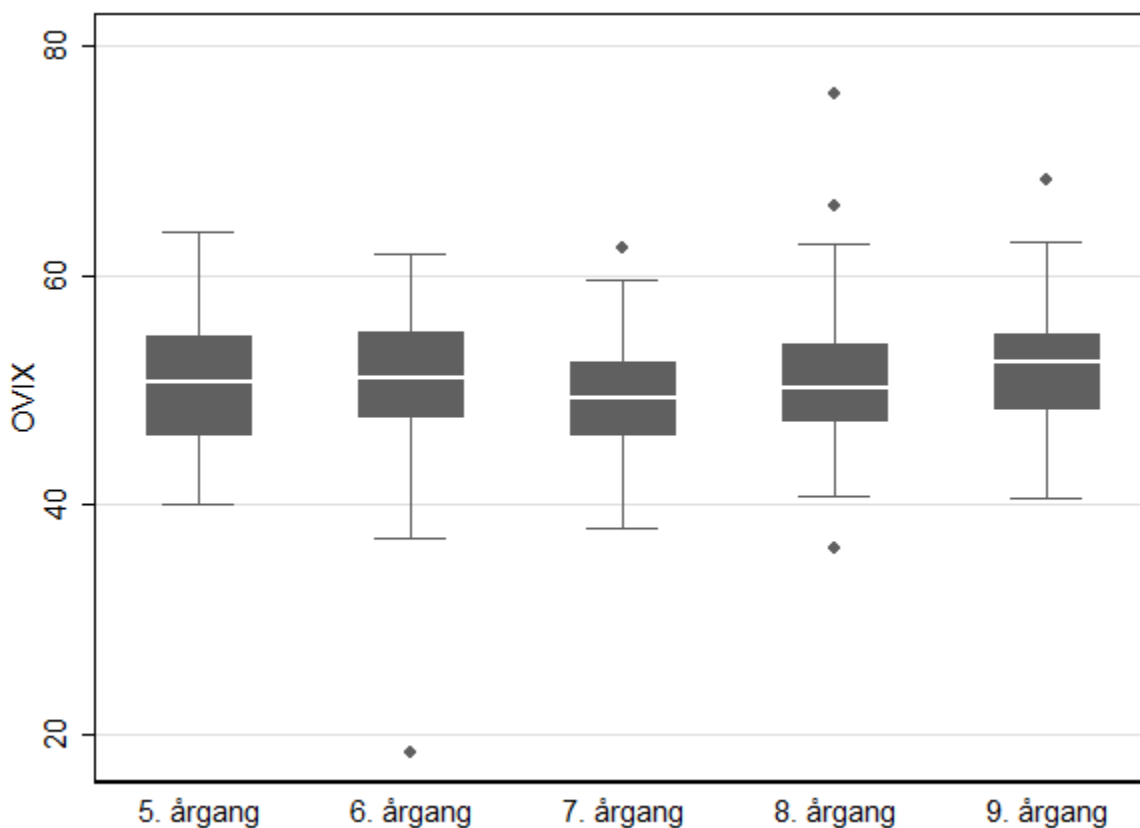
er i elevtekst 5, at sted etablerer sig som et mere udfoldet, selvstændigt semantisk felt. Desuden indgår fritid som semantisk felt i alle tre beretninger, mens skole indgår i to af dem (elevtekst 3 og 5). Disse felter er dog snarere et resultat af skrivepromptet end af teksttypen.

5.4 Kvantitative analyser af lexisvariable

Analyserne på det leksikale niveau er baseret på de to variable OVIX og ordlængde. OVIX dækker som bekendt over ordvariation i teksten og er ifølge Hultman & Westman (1977) et udtryk for informationsrigdom i teksten, mens ordlængde ifølge samme forfattere er et udtryk for leksikal specifikation og variation. Jeg vil først se på fordelingen af OVIX og herefter på fordelingen af ordlængde.

5.4.1 Variabel 2: Ordvariationsindeks (OVIX)

Som den eneste af de lingvistiske variable er OVIX udregnet på elevernes samlede tekstproduktion. Det har ikke været nødvendigt at angive maksimumværdier i boksplottet.



Figur 6: Boksplot for OVIX i samlet tekst per elev (kombineret beskrivende og berettende tekst).

Tabel 15*Deskriptiv statistik for fordelingen af variabelen 'OVIX' for kombineret tekst*

Klassetrin	Minimum	Første kvartil	Median	Tredje kvartil	Maksimum	IQR	Antal tekster
5. årgang	40.10	46.07	50.72	54.63	63.70	8.56	43
6. årgang	18.49	47.60	51.06	55.06	61.82	7.45	41
7. årgang	38.02	45.64	49.28	52.42	62.32	6.78	40
8. årgang	36.21	47.16	50.10	54.06	75.86	6.91	69
9. årgang	40.65	49.05	52.54	54.75	68.32	5.71	44

Ser man på medianværdien, er der ikke systematiske udviklinger i OVIX på tværs af klassetrin. Den stiger en anelse mellem 5. og 6. årgang, hvorefter den falder i 7. årgang. Både 7. og 8. årgang ligger således under medianværdierne for 5. og 6. årgang. 9. årgang ligger dog højest af alle, og der kan derfor ses en stigning på omkring 3,5% fra 5. til 9. årgang. Interessant er det, at IQR falder med alderen med undtagelse af en minimal stigning mellem 7. og 8. årgang. Samlet set falder IQR med over 33% fra 5. til 9. årgang, hvilket er et ganske betragteligt fald. Modsat tekstlængde, hvor spredningen øges med alderen, standardiserer OVIX sig altså over tid. Der er altså mindre spredning omkring midten i elevernes ordvariation. Ser man på den samlede variationsbredde i hele datasættet, er der imidlertid ikke tale om samme (næsten) lineære fald i spredningen. Dog er der forskel på den gennemsnitlige variationsbredde (afstand fra den højeste til den laveste værdi) for mellemtrinårgangene (5. årg. = 23,6, 6. årg. = 43,3), som ligger på omkring 33,5, og udskolingsårgangene (7. årg. = 24,3, 8. årg. = 39,7, 9. årg. = 27,7), som ligger på omkring 30,6. Det svarer til et lille fald på næsten 9% procent i den gennemsnitlige variationsbredde fra mellemtrin til udskoling. Med andre ord betyder dette, at selvom der ikke er nævneværdige udviklinger i ordvariationen i medianværdierne på tværs af årgange, så ser vi en øget standardisering af denne ordvariation over tid. Teksterne bliver altså på dette punkt lidt mere ens med tiden.

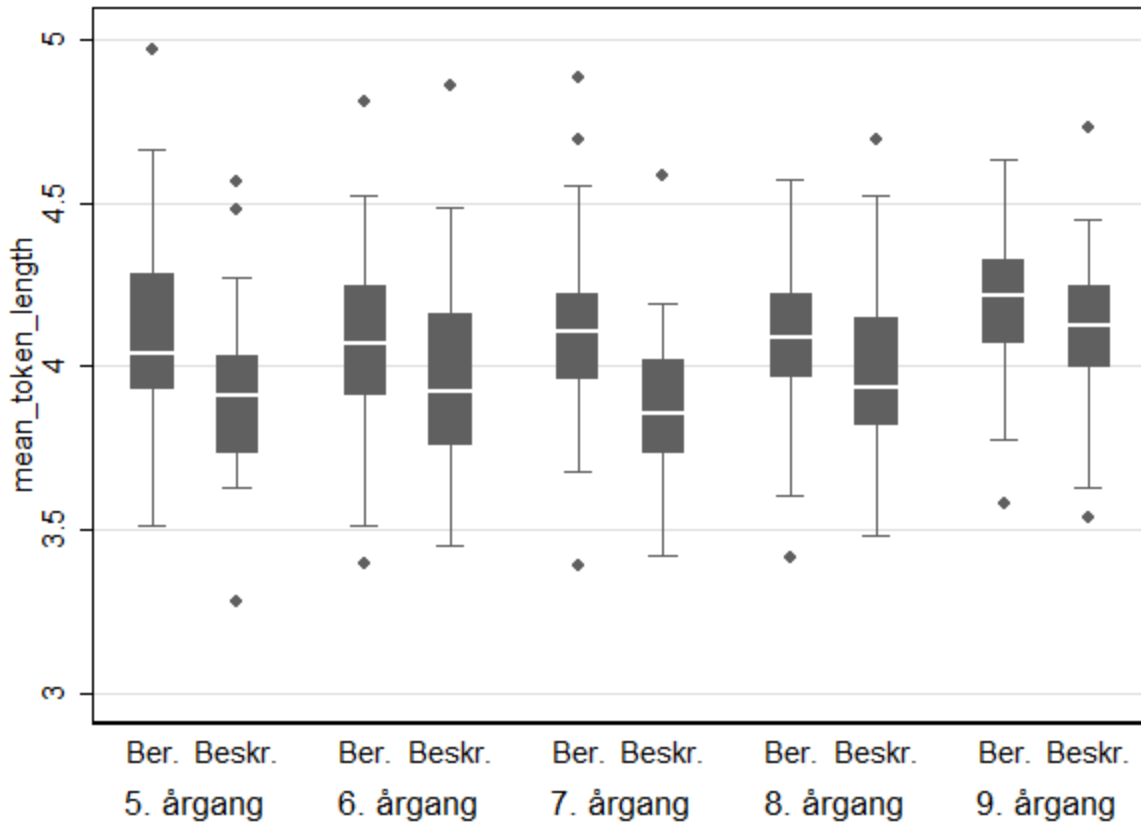
OVIX er, som jeg har beskrevet i afsnit 4.1.2.1, et konstrukt, der dækker over ordvariation i en tekst. Det er dermed ikke et direkte mål, som kan omsættes til meningsfulde lingvistiske strukturer. Det kan derfor være nødvendigt at sætte ovenstående OVIX-værdier i perspektiv ved at sammenligne resultaterne med andre undersøgelser. Hultman &

Westman (1977, s. 56), som oprindeligt udviklede OVIX-målet, fandt i en analyse af gymnasietekster, at medianværdien for OVIX ved de tekster, der havde fået den højeste karakter, lå på 67,6, hvorimod den for de lavest bedømte tekster lå på 58,7. Gennemsnitsværdien for OVIX for alle gymnasieteksterne lå på 64,8, hvilket altså er et godt stykke over medianværdierne for 9. årgang. Nordenfors (2011, s. 111) fandt tilsvarende en stigning i den gennemsnitlige OVIX-værdi på 51 på 5. årgang til en gennemsnitsværdi på 57 på 9. årgang, hvilket er en del over medianværdien for 9. årgang i dette studie¹⁹. Begge studier fandt altså en stigning i OVIX over tid, hvilket ikke stemmer overens med det foreliggende resultat i denne undersøgelse. Resultaterne placerer sig i stedet i forlængelse af Petterson (1977, s. 133), som i en undersøgelse af svenske elevtekster fandt, at OVIX lå stabilt omkring 55 fra 6. til 9. klasse, hvorefter det stiger igen. Petterson taler derfor om et form for udviklingsplateau for OVIX i denne aldersgruppe. I denne undersøgelse ligger medianværdierne for OVIX stabilt på lidt over 50, hvilket dermed bekræfter Pettersons resultater, men altså kontrasterer resultaterne hos Nordenfors og til dels Hultman & Westman.

5.4.2 Variabel 3: Gennemsnitlig ordlængde

Vi vender nu blikket mod variabelen 'gennemsnitlig ordlængde'. Maksimumværdi i boksplottet er sat til 5. Det betyder, at en enkelt outlier på hele 8,5 (beskrivende tekst, 8. årg.) ikke fremgår af boksplottet.

¹⁹ For en mere præcis sammenligning med Nordenfors' værdier kan det nævnes, at gennemsnitsværdien for OVIX på 5. årgang i dette studie ligger på 49,8 og for 9. årgang er den 51,2.



Figur 7: Boksplot for gennemsnitlig ordlængde i berettende og beskrivende tekst.

Tabel 16

Deskriptiv statistik for fordelingen af variabelen 'ordlængde' for berettende tekst

Klassetrin	Minimum	Første kvartil	Median	Tredje kvartil	Maksimum	IQR	Antal tekster
5. årgang	3.51	3.93	4.04	4.28	4.97	.35	39
6. årgang	3.39	3.91	4.07	4.24	4.81	.33	35
7. årgang	3.39	3.96	4.10	4.22	4.88	.26	39
8. årgang	3.42	3.97	4.11	4.21	4.57	.24	69
9. årgang	3.58	4.07	4.21	4.33	4.63	.26	42

Ser man på medianværdierne for ordlængde i de berettende tekster, er der en lille stigning for hver øgning i klassetrin på under én procent indtil overgangen fra 8. til 9. årgang, hvor medianen øges med 2,4%. Medianværdien øges samlet set med 4,2% fra 5. til 9. årgang. Nogenlunde samme tendens ses i første og tredje kvartil, hvilket betyder, at ordlængden

i de tekster, som ligger omkring midten, øges for hvert klassetrin. Udviklingen er størst i første kvartil med en øgning på 3,6% fra 5. til 9. årgang, mens den tilsvarende udvikling i tredje kvartil ligger på 1,2%. Dette betyder, at IQR falder nogenlunde systematisk med alderen med undtagelse af en svag stigning mellem 8. og 9. årgang. Vi kan dermed se, at selvom ordlængden stiger med alderen, så standardiseres spredningen omkring midten.

Tabel 17

Deskriptiv statistik for fordelingen af variabelen 'ordlængde' for beskrivende tekst

Klassetrin	Minimum	Første kvartil	Median	Tredje kvartil	Maksimum	IQR	Antal tekster
5. årgang	3.28	3.73	3.91	4.03	4.56	.30	43
6. årgang	3.45	3.76	3.94	4.17	5.60	.41	41
7. årgang	3.42	3.74	3.87	4.02	4.58	.28	40
8. årgang	3.48	3.83	3.94	4.16	8.51	.33	69
9. årgang	3.54	4.00	4.13	4.25	4.73	.25	44

Ved de beskrivende tekster er udviklingen i medianen ikke helt så systematisk. Her ligger den forholdsvis stabilt fra 5. til 8. årgang, hvorefter der ses en stigning fra 8. til 9. årgang på 4,8%. Den samlede stigning fra 5. til 9. årgang ligger på 5,6%, hvilket altså er over den tilsvarende samlede stigning i de berettende tekster, som lå på 4,2%. Medianværdierne for de berettende tekster ligger dog gennemsnitligt set 3,8% højere end ved de beskrivende tekster. Ser på man de udviklinger i første og tredje kvartil i de beskrivende tekster, er der ikke lineære udviklinger, hvilket afspejler sig i en usystematisk udvikling i IQR. Dog er der en gennemsnitlig forskel på IQR for mellemtrinsårgangene, som ligger på 0,36, og udskolingsårgangene, som ligger på 0,29. Der er altså, selvom tendensen ikke er helt så stærk og systematisk som ved de berettende tekster, en let standardisering i spredningen omkring midten. Den samlede variationsbredde er meget påvirket af outliers i maksimumværdierne og dermed vanskelig at tolke på baggrund af.

5.4.3 Opsummering

Hvor den første analyse af tekstlængde viste en generel systematisk og stærk sammenhæng med klassetrin, viser analysen af de to leksisvariable et mere fragmenteret billede.

Dog er tendenserne for de to leksikale variable meget ens. Begge variable har små gennemsnitlige stigninger i medianværdien, hvis man sammenligner værdierne for 5. og 9. årgang. Stigningerne er dog så små, at man særligt ved OVIX snarere kan tale om et udviklingsplateau (Pettersson, 1977) end en egentlig stigning. Særligt interessant er det i denne sammenhæng, at selvom medianværdien gennemsnitligt set stiger lidt, så falder IQR med alderen ved både OVIX og ordlængde. Ved OVIX er der tale om et fald på hele 33% henover de fem årgange. Sammenlignet med tekstlængde, hvor spredningen så ud til at øges over tid, så ser den altså i højere grad ud til at standardiseres over tid i både OVIX og ordlængde. Teksterne bliver på de målte punkter altså mere ens over tid. Dette er interessant i relation til 9. årgang, idet de ligger lavest i tekstlængde, men højest i både OVIX og ordlængde. I den udstrækning, at man accepterer OVIX og ordlængde som udtryk for henholdsvis informationsrigdom og leksikal specifikation, peger resultaterne af denne analyse på, at et fald i tekstlængde ikke påvirker den leksikale kompleksitet i teksterne, men tværtimod fører en lille stigning med sig. På denne måde understøtter ovenstående analyse den forklaring af 9. årgangs fald i tekstlængde, som gik på, at faldet kunne ses som et udviklingstræk, snarere end et resultat af praktiske omstændigheder. 9. årgangs tekster er altså kortere, men til gengæld mere komplekse på det leksikale niveau.

Sammenligner man resultaterne fra de to teksttyper med hinanden, viser det, at de berettende tekster gennemsnitligt set har en højere ordlængde (3,8%), mens IQR til gengæld gennemsnitligt set er 10% lavere ved de berettende tekster (0,29) end ved de beskrivende (0,32). Set i sammenhæng med analyserne af tekstlængde indeholder de berettende tekster altså gennemsnitligt set flere og længere ord end de beskrivende tekster, og de har en smule lavere spredning i ordlængde omkring midten.

5.5 Tekstudvalg #2 (leksis)

Tekstudvalget med fokus på det leksikale niveau består af seks tekster. En særlig omstændighed ved dette tekstudvalg er, at OVIX er udregnet på kombineret tekst, så det er ikke muligt at sammenligne teksttyper med forskellig OVIX-værdi. I dette tekstudvalg fokuserer jeg særligt på kohæsive bånd og opbygningen af nominalfraser (herunder nominalfraser der er indlejret i større præpositionssyntagmer). Jeg har markeret referentkoblinger

med orange, ellipser med **rød**²⁰, konjunktionaler med **blå** og leksikale bånd med **grøn**. Nominalfraser er angivet ved understregning. Analyserne af nominalfraserne tager som bekendt udgangspunkt i Hansen & Heltofts (2011) topologiske skema for nominalhelheder. Jeg har kun indsat de nominalfraser i det topologiske skema, som består af mere end bare én kategori (hovedord), dvs. minimum to ord. Skemaerne fremgår af bilag 2.

5.5.1 Elevtekst 6 og 7

Elevtekst 6 og 7 er skrevet af en elev på henholdsvis 6. årgang og en elev på 9. årgang. Begge tekster er eksempler på tekster med lave værdier i både OVIX og ordlængde. Vi starter med at se på elevtekst 6, som altså er en berettende tekst skrevet af en elev på 6. årgang. Teksten har en gennemsnitlig ordlængde på 3,56, og elevens to tekster har samlet set et OVIX på 37,19. Begge variable er altså i første kvartil for årgangen.

Jeg vågner **og** så tager jeg tøj på, **gik** ind i stuen **og** laver noget mad. Det kunne være noget æg eller brød. Jeg satte mig ned **og** spiser færdig imens jeg ser noget på min telefon, **efter** jeg har spist færdig sætter jeg min ned til vores spisebord, **og** går i gang med at lave lektier. Det gør jeg i cirka 2 timer **og** så går jeg ud **og** spiller basket med min lillebror. **Efter** vi har gjort det i lidt tid går vi i gang med at lave lektier igen, **men** min lillebror brugte alt sin tid på at spille Fortinet **så** han fik ikke lavet så meget. **Efter** vi var færdig med at lave lektier gik min lillebror **og** jeg ud på vores trampolin **og** hoppe der lidt **efter** det gik jeg ind **og** så lidt på min telefon. **Så** lavet min **og** min mor mad **og** **efter** det lagde jeg mig ind ind i min seng **og** kigget på min telefon igen **og** **efter** det gik jeg i seng.

Elevtekst 6: Berettende tekst skrevet af en elev på 6. årgang.

Eleven anvender primært additive konjunktionaler som 'og', der binder handlingsforløbene i teksten sammen, samt temporale konjunktionaler som 'efter', 'så' og 'imens', som sekvenserer handlingsforløbene og sætter dem i rækkefølge. Det er med andre ord velkendte og forventelige kohæsiionsmarkører i en berettende tekst. Der ses enkelte eksempler på adversative konjunktionaler (fx 'men') og kausale konjunktionaler (fx "min lillebror brugte alt sin tid på at spille Fortinet **så** han fik ikke lavet så meget"), men ingen af delene er udbredte i teksten. Der er altså tale om en meget nøgtern beskrivelse af sekventielt forbundne handlingsforløb, og tekstens jeg-fortæller træder (på nær en enkelt gang) ikke frem som en evaluerende og bedømmende fortæller. Den dominerende type af kohæsive bånd i teksten er ikke overraskende simple anaforske referentkoblinger gennem

²⁰ Ved ellipsekonstruktioner markerer jeg det finite verbum, som det manglende subjekt lægger sig til.

anvendelse af personlige og possessive pronominer samt enkelte elliptiske konstruktioner. Der findes også flere eksempler på anvendelsen af leksikale bånd i teksten. Der findes fx et leksikalt bånd mellem 'æg', 'brød', og 'mad', som etableres gennem hyponyme relationer mellem ordene. Et lignende bånd ses mellem 'basket', 'trampolin', 'Fortinet' og 'telefon', som dog ikke er hyponymt forbundne, men i stedet trækker på det samme semantiske felt (underholdningsgenstande/-aktiviteter). Der etableres også leksikale bånd i teksten gennem gentagelser af bestemte ord (fx lektier, lillebror, telefon).

Vender vi blikket mod nominalfraserne i teksten (se bilag 2), er det tydeligt, at der er tale om en tekst med meget simple nominalfraser. Kun 10 ud af i alt 52 nominalfraser (svarende til 19%) består af mere end ét ord, og kun en enkelt nominalfrase indeholder tre ord. I den længste nominalfrase har hovedordet to foranstillede attributter, to tællere ('cirka' og '2'). Der er ingen nominalfraser med efterstillede led. I otte af de ti nominalfraser er det foranstillede attribut placeret i bestemmerkategorien og består af et possessivt pronomen (min telefon, vores spisebord, min lillebror etc.). Nominalfraserne i teksten udbygges således ofte kun i den udstrækning, der er behov for at klargøre hovedordets tilhørsforhold. Dette hænger godt sammen med analysen af kohæsive bånd, som viste et dominerende anvendelse anaforske, personlige referentkoblinger.

Elevtekst 7 er en beskrivende tekst skrevet af en elev fra 9. årgang. Teksten har en gennemsnitlig ordlængde på 3,63, og OVIX er samlet set på 45,45 for elevens to tekster. Dette placerer således begge variabelværdier i nederste kvartil for årgangen. De to variabelværdier ligger dog – særligt i OVIX – et stykke over værdierne fra elevtekst 6.

Jeg har valgt billede 1. Jeg ser to piger. Den ene pige sidder på en sten, og er i gang med at tage en strømpe på. Den anden piger er i gang med at gå væk fra den ene pige. Pigen der sidder på stenen, har grå bukser og en hvid trøje på. Pigen der er i gang med at gå væk, har røde bukser og en lyserød trøje på. Der står også et par lyserøde sko, foran stenen hvor pigen sidder. Der er en restaurant i baggrunden, hvor der ikke er nogle kunder. Der står en gul parasol. Der er også en anden restaurant ved siden af, der er alt mørkebrunt. Der er et springvand, som ser meget flot ud.

Elevtekst 7: Beskrivende tekst skrevet af en elev fra 9. årgang.

Sammenlignet med elevtekst 6 er der i elevtekst 7 en mindre varieret anvendelse af kohæsive bånd. Der anvendes enkelte anaforske referentkoblinger som 'som' og 'der', mens 'der' også anvendes enkelte steder som demonstrativ reference til sted (fx i sætningen "der er alt mørkebrunt", hvor 'der' refererer til nominalfrasen "en anden restaurant" fra forrige

sætning).²¹ Der anvendes ligeledes markant færre konjunktioner i elevtekst 7. Hvor elevtekst 6 er et eksempel på en tekst, hvor der nærmest er tale om et overforbrug af kohæsive bånd, er der i elevtekst 7 snarere tale om, at sammenhængen i teksten skabes gennem leksikale bånd (fx gentagelse af nominalfraser med ordet 'pige' eller 'restaurant', samt semantiske netværk med farver og tøj) samt på det mere overordnede kohærens-niveau, hvor brugen af præpositionsfraser fungerer som deiktiske referencer til billedet, idet teksten har en tydelig jeg-fortæller, som udpeger og beskriver forskellige dele af billedet. En del af forklaringen på denne forskel er selvfølgelig, at de to elevtekster repræsenterer hver deres teksttype, og som det pointeres af bl.a. Coker Jr. (2012), er det typisk for berettende tekster, at de stiller større krav til elevernes anvendelse af forskellige kohæsiionsmarkører, der strukturerer tekstens hændelsesforløb i tid.

Nominalfraserne i elevtekst 7 er generelt længere og mere komplekse end i elevtekst 6. 19 ud af 33 nominalfraser (svarende til 58%) indeholder mindst to ord, og modsat elevtekst 6, hvor beskriverpladsen kun er udfyldt en enkelt gang, er dette tilfældet seks gange i denne tekst. Udfyldningen af beskriverpladsen består dog udelukkende af farveangivelser, så der i den forstand ikke tale om semantisk variation i beskriverpladsen. Vi ser også eksempler på nominalfraser med efterstillede led, der udgøres af ledsætninger (fx "Pigen der er i gang med at gå væk"), hvilket ikke var tilfældet i forrige tekst. Overordnet set kan man sige, at den beskrivende tekst i dette tilfælde indeholder et større grad af kompleksitet i nominalfraserne, hvilket resulterer i flere leksikale bånd i teksten, men til gengæld en mindre varieret anvendelse af de øvrige kohæsive bånd.

²¹ Jeg anser ikke den øvrige brug af pronomenet 'der' som eksempler på demonstrativ (circumstantial) reference (adverbiel reference til sted), men som eksempler på foreløbigt subjekt. Jeg koder dem derfor ikke som kohæsiionsmarkør.

5.5.2 Elevtekst 8 og 9

Elevtekst 8 og 9 er eksempler på tekster med ordlængde- og OVIX-værdier, der ligger omkring medianen for deres respektive årgange. Elevtekst 8 er en berettende tekst skrevet af en elev fra 8. årgang. Teksten har en gennemsnitlig ordlængde på 4,036 og et samlet OVIX for begge elevens tekster på 50,21. Det placerer således begge variabelværdier tæt på medianværdien på 8. årgang.

Da der var Coronanedlukning startede det med at man bare fik lektier for og så kunne man lave dem når man havde lyst. men vi havde ret mange man skulle nogle gange lave flere timers lektier derhjemme end man normalt er i skole, og det blev en gang imellem lidt stressende. der stod jeg op ved 10 om morgen normalt og gik i seng cirka kl et om natten, og var meget sammen med venner. Men da der begynde at komme online klasser stod jeg op 5 til 3 minutter op før skole og startede computeren, og da jeg havde gjort det lavede vi opgaver. Det var ikke noget specielt vi lavede det vi plejede undtagen valgfag men tit hvis vi skulle ud og lave noget fysisk kom min ven på besøg og så var vi i skole sammen resten af dagen efter skole vil vi tit være sammen resten af tiden og nogle gange ville vi sove sammen og være i skole sammen igen næste dag

Elevtekst 8: Berettende tekst skrevet af en elev fra 8. årgang.

Sammenligner man elevtekst 8 med elevtekst 6, som var en berettende tekst skrevet af en elev fra 6. årgang, deler de to tekster det forhold, at der foruden udbredt brug af leksikale bånd primært anvendes referentkoblinger samt temporale konjunktioner, der sekvenserer tekstens handlingsforløb og sætter dem i rækkefølge. Teksterne adskiller sig dog også fra hinanden på dette punkt, idet elevtekst 8 alene i antallet af kohæsive bånd (inkl. referentkoblinger og temporale konjunktioner) ligger væsentligt lavere end elevtekst 6. Særligt antallet af possessive pronomener er lavere. Eleven anvender desuden enkelte ellipsekonstruktioner. Der ses også i denne tekst udpræget brug af leksikale bånd. I denne tekst kommer dette primært til udtryk gennem forskellige semantiske netværk om skole, søvn og tid (fx adverbelle fraser som 'resten af dagen', 'kl et om natten') samt gentagelse af ordet 'ven'. Selvom OVIX-værdien for denne elev ligger højere end for den elev, der skrev elevtekst 6, er der altså i dette tilfælde ikke noget, der tyder på, at den øgede OVIX-værdi fører til anvendelse af flere referentkoblinger eller konjunktioner, men snarere til lidt øget brug af leksikale bånd.

Teksten anvender desuden meget få udbyggede nominalfraser. Bare 5 ud af 44 nominalfraser (svarende til 11%) indeholder to ord eller mere. Der er en enkelt nominalfrase

KAPITEL 5

med et possessivt pronomen i bestemmerfeltet ('min ven'), to med foranstillede attributter på tællerpladsen ('flere timers lektier' og '5 til 3 minutter') samt to med efterstillede led ('resten af dagen' og 'resten af tiden'). Der er således tale om en meget lav grad af variation i nominalfraserne. Ser man på alle nominalfraserne – inkl. dem med kun ét ord – er der i høj grad tale om pronomielle nominaler, mens nominalfraser med substantiver er mindre udbredte, end vi fx så i elevtekst 6.

Elevtekst 9 er en beskrivende tekst skrevet af en elev fra 5. årgang. Teksten har en gennemsnitlig ordlængde på 3,92, og OVIX-værdien for elevens to tekster er på 50,8. Begge variabelværdier ligger således omkring medianen for 5. årgang.

Jeg har valgt at jeg vil skrive om billede 1. På billedet ser jeg to piger der leger i vandet. Stemningen er lidt trist fordi der ikke er mange mennesker på billedet og restauranterne er helt tomme og der er mange tomme stole. Pigerne leger helt alene. Man kan se at den ene af pigerne har et visir foran ansigtet og de leger på en lidt anderledes måde. Jeg tror at det er forår eller sommer fordi at de leger i vand. Jeg tror de to børn er søstre og at de hygger sig men det er alligevel ikke helt normalt som før corona. Måske savner de at lege med de andre børn inden corona og at kunne lege normalt og uden visir. De har nok glædet sig til at komme ud fordi at jeg tror at det er i udlandet og at de har skullet være hjemme meget længe. De må ikke komme i skole endnu så de savner at lege med deres klassekammerater. De venter på deres forældre som handler ind men det er begyndt at blive kedeligt at lege i springvandet så den ene tager sokker på. De tænker meget på at holde afstand til folk der går forbi og til hinanden.

Elevtekst 9: Beskrivende tekst skrevet af en elev fra 5. årgang.

Elevtekst 9 har en mere udbredt anvendelse af possessive og personlige pronomener som referentkoblinger, hvilket hænger sammen med, at teksten har en tydelig jeg-fortæller, som er gennemgående i hele teksten (modsat den forrige beskrivende tekst, elevtekst 7, hvor jeg-fortælleren kun var synlig i de første to sætninger). Elevens egen stemme er således meget tydelig i denne tekst, hvor der anvendes et mere evaluerende og bedømmende sprogbrug. Derfor ses der også en mere udbredt brug af kausale konjunktioner (fx "Stemningen er lidt trist **fordi** der ikke er mange mennesker..." eller "De må ikke komme i skole endnu **så** de savner..."), som eleven anvender i forbindelse med, at hun argumenterer for sine betragtninger og bedømmelser af billedet. Der er ikke eksempler på elliptiske konstruktioner i teksten, og sammenlignet med de øvrige tekster er der forholdsvis få leksikale bånd. Der ses gentagelser af ordene 'billede', 'visir' og 'corona' samt et semantisk

netværk med pigerne som de centrale agenter, men derudover er brugen af leksikale bånd begrænset.

Der er færre udbyggede nominalfraser i elevtekst 9, end vi så i elevtekst 7. I elevtekst 9 er 12 ud af 58 nominalfraser på to ord eller mere (svarende til 21%), mens 19 ud af 33 nominalfraser i elevtekst 7 var på to ord eller mere (svarende til 58%). At elevtekst 9 har en højere ordlængde og OVIX, betyder altså ikke, at teksten har flere udbyggede nominalfraser, hverken hvis man ser på det totale antal eller på den procentvise andel. Der er heller ikke større variation, hvis man ser nærmere på, hvordan nominalfraserne er bygget op. Den typiske udbyggede nominalfrase i begge tekster består af to foranstillede attributter, én på bestemmerpladsen og én på tællerpladsen. Der ses i begge tekster eksempler på brug af efterstillede led i nominalfrasen, men det er ikke udbredt.

5.5.3 Elevtekst 10 og 11

Elevtekst 10 og 11 er eksempler på tekster med meget høj ordlængde- og OVIX-værdier for deres respektive årgange. Elevtekst 10 er en beskrivende tekst skrevet af en elev fra 6. årgang. Teksten har en gennemsnitlig ordlængde på 4,47 og elevens samlede OVIX er på 60,59, hvilket placerer begge variabelværdier i fjerde kvartil for årgangen.

En typisk skoledag i corona-tiden starter med, at jeg bliver vækket omkring kl. 06.20. Jeg går nedenunder og laver min madpakke, skifter til mit hverdagstøj, spiser morgenmad, børster mine tænder og cykler afsted i skole kl. 07.30. Når jeg er mødt ind i klassen, stiller jeg min telefon oppe, hvor de skal stå og spritter hænder. Efter vi har haft undervisning, har vi frikvarter, hvor jeg spiller bold sammen med dem fra min klasse. Sådan forløber de næste timer og andet frikvarter sig for det meste også. Når vi har fri cykler jeg ned på rideskolen, rider og muger ud i boksen. Efter det henter min far mig og min cykel, og så kører vi hjem. Derhjemme tager jeg et bad, og efter noget mad. Efter jeg har spist, går jeg nogle gange en tur med min hund. Indtil vi spiser aftensmad, sidder jeg for det meste oppe på mit værelse og slapper af. Efter vi har spist, ser vi noget TV, og derefter går jeg i seng.

Elevtekst 10: Berettende tekst skrevet af en elev fra 6. årgang.

Som ved de forrige berettende tekster (elevtekst 6 og 8) ses også i elevtekst 10 en udbredt anvendelse af temporale konjunktioner, som binder tekstens handlingsforløb sammen. Enkelte handlingsforløb bindes dog også sammen gennem brug af ellipsekonstruktioner (fx ”Jeg går nedenunder og **laver** min madpakke, **skifter** til (...), **spiser** (...), **børster** (...) og **cykler** afsted...”). Her ser man altså spæde tegn til en begyndende anvendelse af

det, der i tekstlingvistikken kaldes 'ikonicitet', som henviser til sammenhængende handlingsforløb i en tekst, som er relateret til hinanden i tid, uden denne relation er ekspliciteret gennem anvendelse af temporale konjunktioner (Ulbæk, 2005, s. 52-54). Dette ses dog kun i eksemplerne med ellipsekonstruktioner og er ikke udbredt i teksten. Derudover er der - som vi også har set i de andre tekster - udbredt brug af personlige referentkoblinger, hvor de personlige pronomener refererer til eleven selv, mens der ikke udbredt brug af anaforiske referentkoblinger, hvor pronomenet refererer tilbage til en tidligere del af teksten. De leksikale bånd i teksten er primært realiseret gennem semantiske netværk for skole (skoledag, undervisning, klassen, frikvarter etc.), mad (spiser, aftensmad, morgenmad etc.) samt transport (cykel, cykler, går etc.).

Mønsteret for nominalfraserne er det samme som ved de forrige berettende tekster. Der er forholdsvis få udbyggede nominalfraser. Bare 17 ud af 50 nominalfraser (svarende til 34%) er på to ord eller mere. Opbygningen ligner også mønsteret fra de forrige tekster. Der er enten et foranstillet attribut i bestemmer- eller i tællerfeltet, mens beskriverfeltet (på nær enkelte undtagelser) er tomt. Bestemmerfeltet udgøres typisk et possessivt pronomen, hvilket også var en tendens i to de forrige beskrivende tekster. Der ses altså i denne tekst ikke nogen systematisk sammenhæng mellem øget ordlængde eller OVIX og mere (eller mindre) komplekse nominalfraser.

Elevtekst 11 er en beskrivende tekst skrevet af en elev fra 8. årgang. Teksten har en gennemsnitlig ordlængde på 4,25, og elevens to tekster har samlet set et OVIX på 62,92. Det betyder, at teksten i begge variable ligger i fjerde kvartil for årgangen.

Jeg vælger billede 1. Man kan se at billedet er fra corona tiden, fordi pigen til venstre har en maske på. Pigen til venstre har en hvid langærmede blonde trøje på og nogle grå leggings på med lidt mørkere prikker på. Hun er ved at tage nogle grå, blå, gule og sorte sokker af, så hun kan lege i springvandet. Hun ser meget koncentreret ud. Pigen til højre har røde leggings på og en lyserød langærmede på. Hun har brunt hår og hendes hår er i en halv hestehale med et skrig grøn elastik i. Det ligner at hun allerede har været inde ved springvandet, fordi man kan se våde fodspor der før over til hendes fødder. Hun har ikke været helt inde ved springvandet for hele hendes overkrop er tør, men hun har gået på de våde fliser ved siden af. Billedet blev vist taget i Polen, fordi i baggrunden kan man se nogle polske restauranter og ølmærke. Der er ikke andre på billedet end de to piger. Pigen til venstre er det eneste der har maske på, hende til højre har ingen maske på. Pigen med maske på sidder oppe på en sten, og hende uden maske på er på vej hen til springvandet (igen). Der er nogle lyserøde sko ved siden af den sten pigen med masken sidder på, jeg tror det er hendes sko, men jeg er ikke helt sikker. Der er mange farver på billedet, det er alle varme, og hyggelige farver. Det ser meget hyggeligt ud, og det ligner en dejlig dag med solskinsvejr.

Elevtekst 11: Beskrivende tekst skrevet af en elev fra 8. årgang.

Elevtekst 11 har generelt lidt færre referentkoblinger og konjunktioner end de øvrige tekster. Der anvendes primært anaforske referentkoblinger i tredje person samt additive konjunktioner. Der ses ligeledes en mere udbredt anvendelse af leksikale bånd gennem semantiske netværk om fx tøj og farver (fx 'røde leggings', 'nogle lyserøde sko') og kroppsdele (fx 'hendes fødder', 'hendes overkrop', 'brunt hår') samt gentagelser af bestemte ord og nominalfraser som 'pigen til venstre' og 'springvand'. Der anvendes også en del kausale konjunktioner, hvilket hænger sammen med, at eleven blander en beskrivende fremstillingsform med en argumenterede fremstillingsform og kommer med forskellige fortolkninger, som begrundes ud fra elevens iagttagelser på billedet. Dette ses fx i sætningen "Det ligner at hun allerede har været inde ved springvandet, fordi man kan se våde fodspor...", hvor eleven kobler sin analytiske betragtning sammen med sin konkrete iagttagelse af noget på billedet. Der er ingen eksempler på ellipsekonstruktioner.

Elevtekst 11 er den tekst i dette tekststudvalg, som har flest og samtidig de mest komplekse nominalfraser. 32 ud af 67 nominalfraser (svarende til 48%) indeholder to ord eller

mere, hvilket er flest, hvis man ser på det totale antal. Ved de andre tekster var de udbyggede nominaler i højere grad bygget op om bestemmer- og tællerfeltet, men i denne tekst er det i højere grad bygget op om bestemmer- og beskriverfeltet. Særligt beskriverfeltet er i denne tekst udfyldt, og i flere tilfælde er beskriverfeltet meget langt. Der er dog enkelte gange tale om en form for stakato-beskrivelser, hvor eleven opremser en meget lang række adjektiver (i en enkelt sætning er der fx tale om fire farveangivelser), som lægger sig til samme substantiv. Vi ser ligeledes flere eksempler på nominalfraser med lange efterstillede led (fx ”en halv hestehale med et skrig grøn elastik” eller ”en dejlig dag med solskinsvejr”). Denne type af nominalfraser, hvor det efterstillede led udgøres af adverbielle prædikativer, har vi ikke set eksempler på i tekster med lavere værdier i ordlængde og OVIX. Elevtekst 11 bekræfter således den tendens, som også viste sig i analysen af de to forrige beskrivende tekster, at de beskrivende tekster genererer flere udbyggede og mere komplekse nominalfraser.

5.5.4 Opsummering på analyse af tekstudvalg 2

Der er særligt to hovedpointer, som kan udledes af analysen af tekstudvalg 2. For det første viser analysen, at de kohæsive bånd i elevernes tekster er meget påvirkede af teksttype. De berettende tekster er særligt domineret af personlige og possessive referentkoblinger, som placerer eleven selv som tekstens væsentligste agent (eller det man i en tema/remaoptik vil kalde tekstens ’tema’), samt additive og temporale konjunktioner, som forbinde tekstens handlingsforløb og sekvenserede dem i tid. De beskrivende tekster derimod anvender forholds simple anaforiske referentkoblinger samt additive konjunktioner. Der ses dog også brug af personlige pronomener (første person) i alle tre beskrivende tekster, hvilket hænger sammen med, at elevens egen, personlige stemme træder frem i alle tre beskrivelser. I elevtekst 9 og 11 ses dette også ved, at der anvendes kausale konjunktioner i forbindelse med, at de pågældende elever går mere analytisk til værks og argumenterer for forskellige iagttagelser på billedet. Der ses altså en tendens til, at eleverne ikke bare skriver sig selv eksplicit frem som fortæller i de personlige berettende tekster, men også i billedbeskrivelserne.

Der ses ingen systematisk sammenhæng i tekstudvalget mellem antallet af kohæsive bånd og OVIX-værdi, hvilket ellers var en nærliggende antagelse, eftersom en øget informationsrigdom ofte vil stille større krav til intern sammenhæng i teksten. Interessant er

det dog, at elevtekst 6, som var den tekst med den laveste OVIX-værdi, samtidig var en af de tekster med mest udpræget anvendelse af referentkoblinger og konjunktioner. Særligt anvendte eleven mange temporale konjunktioner ('efter', 'så' etc.). Hvad dette skyldes, og om det er et udtryk for en større tendens, kan jeg naturligvis ikke sige noget om baseret på ovenstående analyser. Men der er dog tendenser fra forskningslitteraturen, som indikerer, at et overforbrug af kohæsive bånd kan ses som et tegn på, at eleven befinder sig tidligt i sin skriveudvikling. Crossley & McNamara (2010) viser fx, at tekstkvalitet hænger tæt sammen med holistiske scoringer af kohærens, men til gengæld er det negativt korreleret med korpuslingvistiske frekvensanalyser af forskellige grammatiske kohæsiionsmarkører. Noget kunne således tyde på, at nogle elever overforbruger kohæsiionsmarkører i starten for at skabe fx temporal sammenhæng i teksten, mens de længere fremme i deres udvikling anvender færre kohæsiionsmarkører til fordel for kohærensmarkører (fx ikonicitet som vi så begyndende tegn på i elevtekst 10), eller har en mere varieret anvendelse af forskellige typer af kohæsive bånd (som vi fx så i elevtekst 11). I forlængelse af den intersprogspointe, jeg beskrev i afsnit 2.4.3, kan man se dette træk som en form for hypoteseafprøvning, hvor eleven overforbruger forskellige kohæsiionsmarkører i et forsøg på at tilnærme sig skriftsprogets normer for tekstsammenhæng. Det kræver dog fortsat flere analyser af et større datasæt, end det der er tilgængeligt i denne afhandling, for at afgøre, om dette er en sand hypotese.

Den anden hovedpointe af analysen af tekststudvalg 2 er, at nominalfrasernes længde og kompleksitet er særligt afhængig af teksttype. De kvantitative analyser af ordlængde indikerede, at de berettende tekster generelt havde højere ordlængde, hvilket antyder en højere grad af leksikal specifikation. Men ser vi på de kvalitative analyser af nominalfraserne er tendensen meget tydelig. De beskrivende tekster i tekststudvalget indeholder både længere (forstået som indeholdende flere ord) og mere komplekse (forstået som anvendelse af flere typer af foranstillede attributter) nominalfraser end de berettende tekster. Der er foretaget kvantitative studier af såkaldte nominalkvoter (summen af substantiver, verber og participier delt med summen af pronomener, verber og adverbier) i elevtekster, som har vist, at overgangen fra en verbal stil til en nominal stil kan ses som et tegn på skriveudvikling (Palmér & Hussenius, kommende). Hvis tendensen fra ovenstående analyse er sand i en større og mere generaliserbar forstand, er der imidlertid grund til at overveje, om ikke denne udvikling skal ses relativt til den teksttype, eleven skriver – en

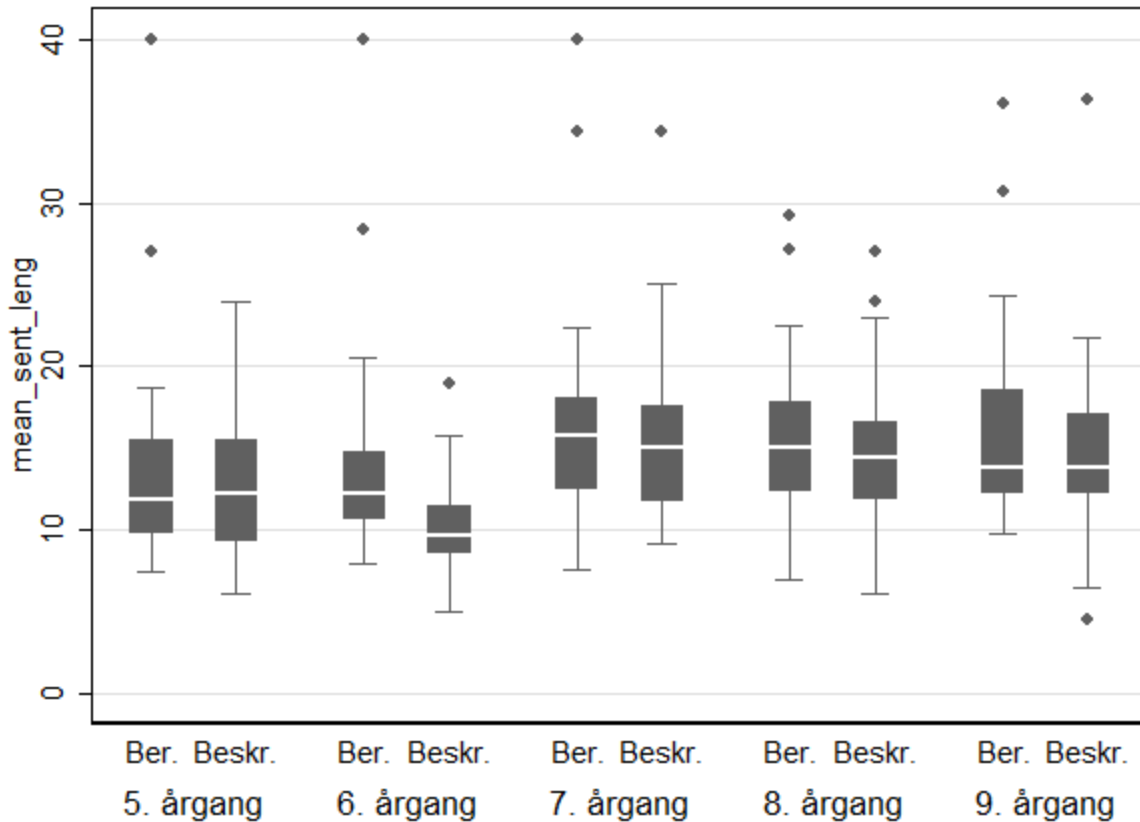
pointe som også understreges af Palmér (2018). Beskrivende tekster er således ikke bare simple tekster forbeholdt begynderkrivningen, men som vi har set, giver denne teksttype også mulighed for at skrive i en mere analytisk stil, hvor beskrivende og argumenterende fremstillingsformer blandes sammen. Det kræver dog også her nye og større kvantitative analyser, hvor man særligt fokuserer på sammenhængen mellem klassetrin, nominaler og teksttype, for at se, om det er en generaliserbar tendens.

5.6 Kvantitative analyser af syntaksvariable

Vi vender nu blikket mod den sidste del af de diakrone analyser, som har fokus på teksternes syntaktiske niveau. De to syntaksvariable udgøres som bekendt af 'gennemsnitlig sætningslængde' og 'gennemsnitlig dependensafstand'. I det følgende vil jeg starte med at se på sætningslængde og derefter på dependensafstand.

5.6.1 Variabel 4: Gennemsnitlig sætningslængde

Sætningslængde defineres som bekendt som det gennemsnitlige antal ord i en syntaktisk periode (se nærmere definition i afsnit 4.1.3.1). Der er indsat en maksimumværdi i boksplottet på 40. Det betyder, at fire outliers ikke fremgår. Det drejer sig om outliers på henholdsvis 53 (5. årg., berettende), 53 (5. årg. berettende), 68 (6. årg., berettende) og 62 (7. årg., berettende).



Figur 8: Bokplot for sætningslængde i berettende og beskrivende tekst.

Tabel 18

Deskriptiv statistik for fordelingen af variabelen 'sætningslængde' for berettende tekst

Klassetrin	Minimum	Første kvartil	Median	Tredje kvartil	Maksimum	IQR	Antal tekster
5. årgang	7.46	9.76	11.80	15.54	53.00	5.78	39
6. årgang	8.00	10.57	12.20	14.80	68.00	4.23	35
7. årgang	7.60	12.50	15.75	18.11	62.00	5.61	39
8. årgang	7.00	12.36	15.07	17.86	29.15	5.49	69
9. årgang	9.81	12.19	13.84	18.57	36.00	6.38	42

Ser man på medianværdierne for sætningslængde i de berettende tekster, ses en stigning fra 5. til 7. årgang, hvor medianværdien er højest, hvorefter der ses et let fald frem mod 9. årgang. Dog falder den i 9. årgang kun til et niveau, der stadig ligger henholdsvis 15,9% og 12,6% over niveauet for 5. og 6. årgang. Der ses altså en tendens til, at sætningslængden

stiger ved overgangen fra mellemtrin til udskoling, hvorefter der ses en stagnering og et let fald. Samme tendens ses i første kvartil, mens der i tredje kvartil ses en fortsat og kontinuerlig stigning (bortset fra små fald i 6. og 8. årgang) frem mod 9. årgang. Dette betyder også, at der kan observeres en stigning i IQR på 10,4% fra 5. til 9. årgang. Modsat leksisvariablerne, hvor spredningen standardiserede sig over tid, så øges spredningen i sætningslængde for de berettende tekster over tid (dog ikke lineært). På denne måde følger sætningslængde altså den samme tendens, som vi så ved tekstlængde. Eftersom de berettende tekster indeholder en del outliers, er det svært at konkludere noget ud fra den samlede variationsbredde.

Tabel 19

Deskriptiv statistik for fordelingen af variabelen 'sætningslængde' for beskrivende tekst

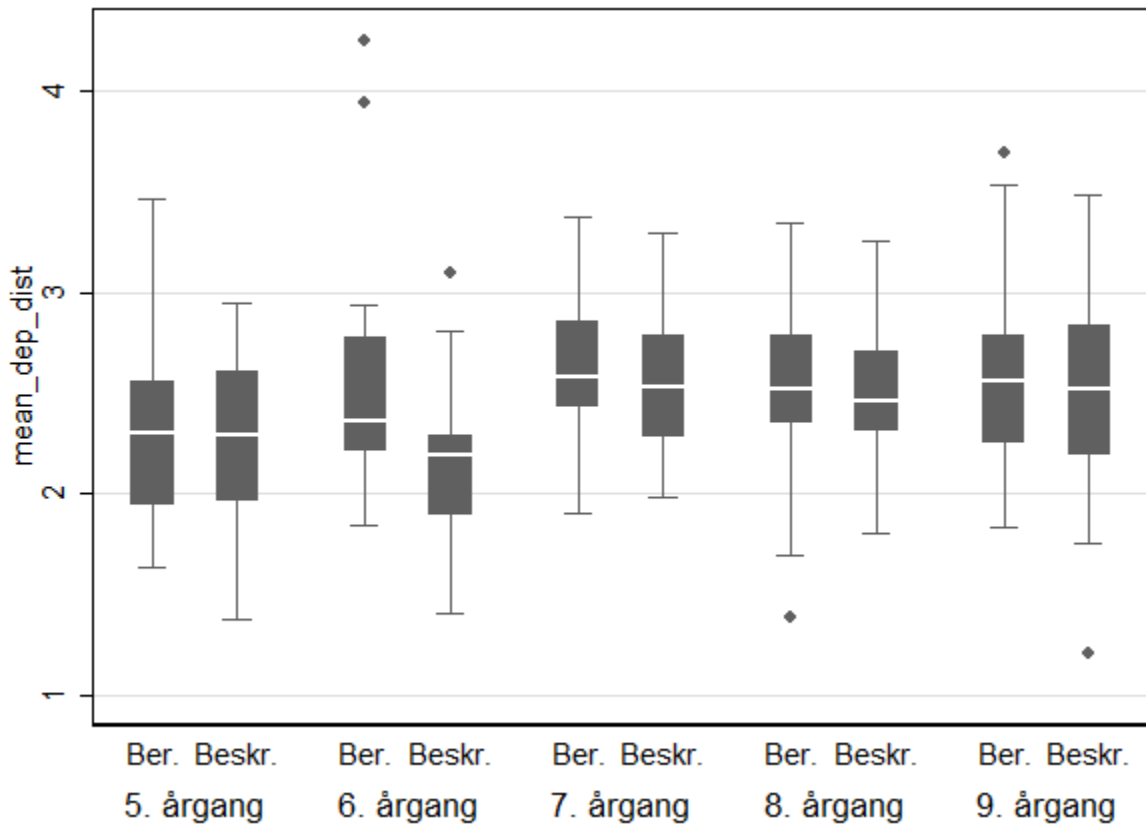
Klassetrin	Minimum	Første kvartil	Median	Tredje kvartil	Maksimum	IQR	Antal tekster
5. årgang	6.13	9.33	12.17	15.53	24.00	6.20	43
6. årgang	5.00	8.50	9.71	11.53	19.00	3.03	41
7. årgang	9.11	11.72	15.08	17.59	34.40	5.87	40
8. årgang	6.13	11.86	14.46	16.67	27.00	4.81	69
9. årgang	4.50	12.18	13.76	17.17	36.33	4.99	44

Medianværdierne for de beskrivende tekster minder på flere måder om de berettende tekster. Endnu engang er medianværdien højest på 7. årgang, hvorefter der ses en stagnering og et let fald frem mod 9. årgang, som dog også her ligger 12,3% over medianværdien for 5. årgang. Der er således markant forskel på mellemtrin og udskoling, idet den gennemsnitlige medianværdi for mellemtrinsårgangene er 10,94 og for udskolingsårgangene 14,43. Det er således en gennemsnitlig forskel på 3,5 ord per sætning. Ved de beskrivende tekster stikker 6. årgang ud med en markant lavere medianværdi end de øvrige årgange (20% lavere end 5. årgang og 35% lavere end 7. årgang). IQR følger ikke samme tendens som ved de berettende tekster. Ved de beskrivende tekster falder IQR derimod fra 5. årgang til 9. årgang med hele 19,5%. Spredningen i sætningslængde for de beskrivende tekster standardiserer sig dermed over tid. Lavest er spredningen ved 6. årgang, som kun har

halvt så stor spredning som 5. årgang. 6. årgang er således en outlier, der er svær at forklare, i stil med det vi så ved 9. årgang i tekstlængde.

5.6.2 Variabel 5: Gennemsnitlig dependensafstand

Vi vender nu blikket mod variabelen gennemsnitlig dependensafstand. Der er ikke angivet maksimumværdier i boksplottene.



Figur 9: Boksplot for gennemsnitlig dependensafstand i berettende og beskrivende tekst.

Tabel 20a

Deskriptiv statistik for fordelingen af variabelen 'gennemsnitlig dependensafstand' for berettende tekst

Klassetrin	Minimum	Første kvartil	Median	Tredje kvartil	Maksimum	IQR	Antal tekster
5. årgang	1.64	1.94	2.31	2.56	3.46	.62	39
6. årgang	1.85	2.21	2.36	2.78	4.25	.57	35
7. årgang	1.90	2.43	2.58	2.86	3.38	.42	39
8. årgang	1.38	2.35	2.52	2.79	3.35	.44	69
9. årgang	1.83	2.25	2.56	2.79	3.69	.55	42

Medianværdierne for dependensafstand i de berettende tekster følger på flere punkter samme udviklingstendens, som vi så for sætningslængde. Der ses en stigning frem til 7. årgang, hvor medianværdien er højest og ligger 8,9% over medianværdien for 6. årgang, hvorefter stigningen flader ud og falder let frem mod 9. årgang. Der er en generel stigning på 10,8% fra 5. til 9. årgang. Der er således også her forskel på mellemtrin og udskoling med en gennemsnitlig medianværdi på henholdsvis 2,33 og 2,55 svarende til en forskel på godt 9%. IQR falder systematisk frem mod 7. årgang, hvor den er lavest, og stiger igen frem mod 9. årgang. IQR for 9. årgang ligger dog fortsat 12% under IQR for 5. årgang. I denne sammenhæng er 7. årgang således interessant, fordi vi her ser den højeste medianværdi og den laveste IQR, hvilket indikerer, at 7. årgang er der, hvor den typiske tekst er syntaktisk mest kompleks og samtidig der, hvor teksterne omkring midten ligner hinanden mest. Maksimumværdierne er dog højere ved både 6. og 9. årgang, hvilket viser, at der er enkelttekster med højere dependensafstand i disse årgange.

Tabel 20b

Deskriptiv statistik for fordelingen af variabelen 'gennemsnitlig dependensafstand' for beskrivende tekst

Klasse-trin	Mini-mum	Første kvartil	Median	Tredje kvartil	Maksi-mum	IQR	Antal tekster
5. årgang	1.38	1.97	2.29	2.61	2.95	.64	43
6. årgang	1.41	1.90	2.19	2.29	3.10	.39	41
7. årgang	1.98	2.29	2.53	2.79	3.30	.51	40
8. årgang	1.80	2.32	2.46	2.71	3.26	.39	69
9. årgang	1.21	2.20	2.52	2.84	3.48	.64	44

Ser man nærmere på fordelingen i de beskrivende tekster, ses samme tendens som ved de berettende tekster. Medianværdien stiger indtil 7. årgang, hvorefter den stagnerer og falder let. Der er dog stadig en generel stigning fra 5. til 9. årgang på 10%, hvilket er lidt højere end ved de berettende tekster. IQR har ingen systematisk udvikling. Den skiftevis stiger og falder for hvert klassetrin. 9. årgang er således på samme niveau som 5. årgang. Maksimumværdierne stiger dog for hvert klassetrin med 8. årgang som undtagelse. Der dukker således enkelttekster op med kontinuerligt højere dependensafstand for hver øgning i klassetrin. Dette afspejler sig også, hvis man ser på den samlede variationsbredde. Her ligger 9. årgang øverst med 2,27, hvilket er 36,5% højere end 5. årgang, som har en variationsbredde på 1,57. Der er dermed størst spredning i dependensafstand i 9. årgang set over hele datasættet.

5.6.3 Opsummering

Ovenstående analyse af de syntaktiske variable viser, at selvom der ikke kan observeres lineære udviklinger i nogle af variablerne, så er der dog substantielle forskelle mellem tekster skrevet af elever på mellemtrinnet og i udskolingen. Sætningslængden (medianværdien) ligger for berettende og beskrivende tekster gennemsnitligt set henholdsvis 24% og 32% højere i udskolingen end på mellemtrinnet, og i dependensafstand er den tilsvarende forskel henholdsvis 9% og 12% for berettende og beskrivende tekster. At der ikke er en lineær udvikling skyldes, at der for begge variable sker en stagnering og et lille fald efter 7. årgang. Så selvom der for begge variable er en generel stigning fra 5. til 9. årgang, så viser den årgangsvise fordeling, at der er tale om en form for udviklingsplateau for de

to variable i udskolingen. Den største udvikling i de syntaktiske variable sker således i overgangen fra mellemtrin til udskoling, dvs. mellem 6. og 7. årgang. IQR viste ikke nogen særlig generel systematik, eftersom tendenserne så forskellige ud relativt til både variabel og teksttype.

Sammenligner man de to teksttyper med hinanden, kan der kun observeres ganske små forskelle. Udviklingstendenserne i begge variable er meget ens, og det samme er medianværdierne. I gennemsnitlig sætningslængde har de berettende tekster generelt højere medianværdi, men forskellen er ved de fleste klassetrin på under ét ord. Den mest bemærkelsesværdige forskel mellem de to teksttyper er, at man ved de berettende tekster ser en generel stigning i IQR med alderen, mens man i de beskrivende tekster ser et fald. Vender vi blikket mod gennemsnitlig dependensafstand, er billedet det samme som ved sætningslængde. Medianværdierne er en smule højere ved de berettende tekster, men forskellen er ved de fleste klassetrin dog kun på 2%. Det er vanskeligt at sammenligne spredningen mellem de to tekster, idet der i de berettende tekster ses et fald i IQR frem mod 7. årgang, hvorefter det stiger en smule igen, mens der ikke kan observeres systematiske forandringer i IQR for de beskrivende tekster.

5.7 Tekstudvalg 3 (syntaks)

Selvom der ikke er lineære stigninger i nogle af de to syntaksvariable, er der som beskrevet ovenfor systematiske forandringer i variablene mellem mellemtrin og udskoling, ligesom der kan observeres en generel stigning i begge variable fra 5. til 9. årgang. Dette rejser imidlertid en række væsentlige spørgsmål angående de syntaktiske strukturer og konstruktioner, der ligger bag disse mønstre. Hvilke syntaktiske forskelle er det, der forårsager en stigning i de to variable fra mellemtrin og udskoling? Dækker det udviklingsplateau, som der kan observeres på tværs af de tre udskolingsårgange, over andre syntaktiske udviklingstræk i elevernes tekster, som først bliver synlige, når man ser på teksterne kvantitativt? Som jeg beskrev i afsnit 4.1.3.1, er der særligt en begrænsning knyttet til analyserne af sætningslængde, eftersom disse tager udgangspunkt i syntaktiske perioder. Syntaktiske perioder er imidlertid det mest overordnede syntaktiske niveau, hvilket betyder, at de kvantitative analyser af sætningslængde er et stykke fra at kunne sige noget om, hvad der kendetegner teksterne på et mere deskriptivt grammatisk niveau. Dette spørgsmål vil jeg imidlertid belyse i det kommende afsnit.

Tekstudvalget af tekster med fokus på syntaks består af 7 tekster. Elevtekst 12, 13 og 14 er eksempler på tekster med lave værdier i sætningslængde og dependensafstand for deres respektive årgange. Elevtekst 15 og 16 repræsenterer tekster omkring medianen, mens elevteksterne 17-20 repræsenterer tekster med meget høje værdier. Analysen fokuserer som beskrevet i afsnit 3.3.2.2 på syntaktiske konstruktioner med særligt henblik på hypotakse/paratakse-variation, ledsætningstyper og forfelsesrealiseringer. Ledsætningstyperne er farvemarkeret. Nominale ledsætninger er markeret med grøn, adverbielle med rød og adjektiviske med blå. Forfelterne er understreget, og sætningsledet i forfettet er markeret i en firkantet parentes før det finitte verbum. Her står "S" for subjekt, "O" for objekt, "IO" for indirekte objekt, "A" for adverbial, "IP" for interrogativt pronomen, "FS" for formelt eller foreløbigt subjekt²², og "SF" står for sætningsfragment. Derudover er alle sætninger i de 9 elevtekster indsat i sætningskemaer, som kan findes i bilag 1.

5.7.1 Elevtekst 12, 13 og 14

Elevtekst 12 er som bekendt et eksempel på en tekst med lave værdier i både sætningslængde og dependensafstand. Der er tale om en berettende tekst skrevet af en elev fra 5. årgang. Teksten har en gennemsnitlig sætningslængde på 7,40 ord og en gennemsnitlig dependensafstand på 1,75, hvilket placerer teksten i første kvartil på årgangen for begge variable.

Nu [A] vil jeg fortælle om en dag ligesom alle dage men ikke helt **fordi der var nedlukning i Danmark** og det [S] var kedeligt **fordi man kunne ikke lege sammen med nogle af sine venner**. Nå men lad os starte jeg [S] stod op klokken 7. Min mor [S] sagde **der var morgenmad**. Jeg [S] spiste cornflaeks ligesom alle dage. Efter [A] havde vi "Skole" hjemme. Vi [S] skulle starte med at lave matematik. Efter der [A] lavet vi dansk. Til Sidst [A] var det madkunstsab. Der [A] skulle vi lave banan kage. Efter [A] skulle vi hag aftensmad. Derefter [A] skulle i seng. Godnat og sov godt.

Elevtekst 12: Berettende tekst skrevet af elev fra 5. årgang.

Selvom elevtekst 12 indledes med en syntaktisk periode på hele 34 ord, er den gennemsnitlige sætningslængde lav på grund af en række syntaktiske perioder på bare 4-5 ord. Den første syntaktiske periode er ikke bare den længste, men også den syntaktisk set mest

²² I tilfælde med foreløbigt subjekt følger jeg Cramer et al. (1996: 23) og placerer det foreløbige subjekt i forfettet og det egentlige subjekt på N-pladsen.

komplicerede. Den består af to hovedsætninger, et sætningsfragment²³ og to adverbielle ledsætninger, der indledes med en kausal konjunktional ('fordi'). Der anvendes således både hypotakse og paratakse ('og det var kedeligt') inden for samme periode. I den resterende del af teksten anvendes en mere simpel syntaks. Foruden et enkelt eksempel på en nominal ledsætning, der fungerer som objekt, består de resterende perioder af enkelte hovedsætninger adskilt af punktum. Der er udpræget brug af tidsadverbialer i forfeltet (6 ud af 12 forfelter består således af tidsadverbialer), mens de resterende forfelter er realiseret af subjekt. Ligesom vi så ved analysen af kohæsive bånd, er det således karakteristisk for berettende tekster, at der anvendes en stor andel af tidsadverbielle led.

Elevtekst 13 er skrevet af en elev fra 8. årgang. Teksten har en gennemsnitlig sætningslængde på 6,3 ord og en gennemsnitlig dependensafstand på 1,8. Begge variable er i første kvartil for årgangen.

2 Piger [S] ved springvand Det [S] ser fortabte ud de [S] kan mærke Corona tæt på de [S] gå med masker og vi siger. Den ene pige [S] sidder på en stor kugle den anden [S] står ved springvandet. Der [FS] er ingen mennesker i baggrunden pigernes [S] ser ud til at være helt alene i byen.

Elevtekst 13: Beskrivende tekst skrevet af en elev fra 8. årgang.

Elevtekst 13 er en kort tekst bestående af bare 8 syntaktiske perioder. Alle perioder består af én hovedsætning. Der er ingen anvendelse af ledsætninger. Der anvendes en enkelt prædikativ sætning ("Det ser fortabte ud"), samt et par objekter på N-pladsen, men ellers består hovedsætningerne af subjekt i forfeltet (mest pronomener) og en præpositionsforbindelse på A-pladsen. De syntaktiske perioder varierer fra 4 og op til 10 ord per periode. Modsat elevtekst 12, hvor vi så et enkelt eksempel på en kompleks syntaktisk periode på 34 ord med både hypotaktiske og parataktiske konstruktioner, er der meget lidt syntaktisk variation i denne tekst og dette til trods for, at teksterne ligger på nogenlunde samme niveau i gennemsnitlig dependensafstand, mens elevtekst 12 har en anelse højere sætningslængde.

²³ Jeg anser ikke 'men ikke helt' som en ledsætning, selvom der anvendes en underordnende konjunktion, men som et sætningsfragment, eftersom der mangler et finit verbal i sætningen.

Vender vi blikket mod elevtekst 14, ser vi et eksempel på en beskrivende tekst, som ligger en anelse højere i både sætningslængde og dependensafstand. Elevtekst 14 er en beskrivende tekst skrevet af en elev fra 5. årgang. Teksten har en gennemsnitlig sætningslængde på 9,14 og en gennemsnitlig dependensafstand på 2,13, hvilket placerer sætningslængde i første kvartil og dependensafstand i andet kvartil for årgangen.

Der [FS] er helt lukket i byen. Der [FS] er kun få **der går rundt dernede**. Det [S] er pga coronavirussen. Det [FS] er en meget trist stemning i Tyskland i år. Restauranterne [S] er helt tomme men der [FS] er to piger **der er fornuftige**. De [S] leger i et springvand men de [S] har mundbind på og de [S] holder afstand mellem hinanden. Det [FS] er sommer så de [S] bader i springvandet.

Elevtekst 14: Beskrivende tekst skrevet af en elev fra 5. årgang.

Teksten består af 6 syntaktiske perioder, som indeholder i alt 11 hovedsætninger og to ledsætninger. Ledsætningerne i teksten er adjektiviske og relative, og de lægger sig således som uddybende efterstillede led til to nominalfraser. Den længste syntaktiske periode indeholder 16 ord og består af tre hovedsætninger forbundet af de parataktiske konjunktioner 'men' og 'og'. Der findes i alt fire eksempler på parataktiske konstruktioner og to eksempler på hypotaktiske konstruktioner. Alle forfelterne er, som vi også så i elevtekst 13, realiseret gennem subjekt eller foreløbigt/formelt subjekt. Der er således en del ligheder mellem de to beskrivende tekster, men det, der særligt adskiller de to, er, at elevtekst 14 binder en del af hovedsætningerne sammen med parataktiske konjunktioner. Det er således særligt den øgede brug af paratakse, som bidrager til den øgede sætningslængde. Øgningen i dependensafstand kan derimod tænkes at hænge sammen med brugen af hypotakse, som bl.a. bidrager til en smule mere udbyggede nominalfraser og derimod større afstand mellem 'head' og 'dependencies'.

5.7.2 Elevtekst 15 og 16

Elevtekst 15 og 16 er eksempler på tekster med værdier i de syntaktiske variable, der ligger omkring medianen for deres respektive årgange. Vi starter med elevtekst 15, som er en berettende tekst skrevet af en elev fra 7. årgang. Teksten har en gennemsnitlig sætningslængde på 15,36 ord og en gennemsnitlig dependensafstand på 2,6.

Jeg [S] vågnede ca 7:30. Så jeg kunne nå at få morgenmad og alt det der inden vi havde det første meet møde. Det første jeg gjorde da jeg vågnede [S+A] var at jeg tjekkede min telefon. Da jeg havde ligget lidt med den [A] gik jeg nedenunder hvor min far sad og arbejde, fordi han arbejde hjemmefra. Bagefter [A] gik jeg hen og lavede noget morgenmad. Jeg [S] spiser normalt ikke morgenmad fordi jeg føler ikke at jeg har tid til at lave morgenmad inden jeg skulle i skolen. Med i lockdown [A] havde jeg masser tid til at både lave og spise morgenmad. For det meste [A] fik jeg et rundstykke eller en bolle og måske en smoothie. Efter jeg havde spist min morgenmad [A] gik jeg ovenpå igen og børstede mine tænder. Der [A] var klokken ca blevet 7:45. Så der [A] havde jeg ca et kvarter til at enten se noget youtube eller netflix inden det første meet møde. Det første meet møde [S] startede for det meste kl 8:00. Efter første time [A] havde vi pause hvor jeg gik ned og tog noget frokost. Det [S] var bare en bolle eller noget rugbrød. Så [A] havde jeg ca en halv time tilbage til jeg skulle have min næste time. I den halv time [A] sad jeg enten med min telefon eller min computer. Efter 2 time [A] havde vi pause igen. Der [A] sad jeg bare med min telefon eller computer igen. Da skoledagen var ovre [A] tog jeg tit hen til nogle af mine venner eller de [S] kom over til mig, eller vi [S] kunne mødes et sted. I corona tiden [A] stod jeg meget på rulleskøjter med [*proprium, proprium, proprium, proprium, proprium, og proprium*]. Så [A] vi mødes tit og stod på rulleskøjter sammen på skolen. Vi [S] kørte også nogle gang i netto eller fakta. Nogle gang [A] bestilte vi en pizza hos [*proprium*] og spiste den på skolen, men for det meste [A] spiste vi hjemme. Da jeg så kom hjem igen [A] var aftensmaden klar. Efter jeg havde spist aftensmad [A] gik jeg ind på mit værelse og sad på min telefon eller min computer indtil jeg skulle sove.

Elevtekst 15: Berettende tekst skrevet af en elev fra 7. årgang.²⁴

Elevtekst 15 består af 18 syntaktiske perioder bestående af 32 hovedsætninger (herunder ellipser) og 17 ledsætninger. Langt hovedparten af ledsætningerne er adverbielle ledsætninger, men der ses også et eksempel på både en adjektivisk og en nominal ledsætning. Der ses bare 7 eksempler på parataktiske konjunktioner, mens der er langt mere udbredt brug af hypotakse. Den korteste periode i teksten er på 8 ord, mens den længste er på 25, så der er stor spredning i længden af sætningerne internt i teksten. Ser man på sætnings-skemaet i bilag 1 (tabel 40), er det desuden tydeligt, at der er stor variation i, hvor tyngden i sætningerne placeres. En del ledsætninger er placeret til sidst i ektrafeltet, men der er også gentagne eksempler på, at ledsætninger placeres i forfeltet, hvilket giver en balance mellem venstre- og højretyngde i sætningsopbygningerne. Hvor der i den forrige berettende tekst (elevtekst 12) kun blev brugt tidsadverbielle enkeltord eller korte tidsadverbialfraser i forfeltet, ses der altså i denne tekst gentagne eksempler på hele tidsadverbielle

²⁴ De dele af teksten, som er anonymiseret, fremgår med en kursiveret angivelse af ordklasser i en firkanter parentes.

ledsætninger i forfeltet. Så selvom elevtekst 15 og 12 deler det træk, at de i høj grad er præget af tidsadverbialer, så adskiller elevtekst 15 sig ved længden og udbyggetheden af tidsadverbialerne, hvilket i høj grad er med til at forklare forskellene mellem de to tekster i sætningslængde og dependensafstand.

Vender vi blikket mod den næste elevtekst, har vi med en tekst at gøre, som dels er markant kortere (tekstlængde) end elevtekst 15, dels ligger lidt lavere i både sætningslængde og dependensafstand. Begge værdier ligger dog meget tæt på medianen for 9. årgang. Elevtekst 16 er nemlig en beskrivende tekst skrevet af en elev fra 9. årgang. Teksten har en gennemsnitlig sætningslængde på 13,71 ord, og den gennemsnitlige dependensafstand ligger på 2,51.

Man [S] kan se en dreng, som er i midten af billedet i mellemgrunden. Der [FS] er ikke andre ansigter på billedet end pigens, mandens og drengens. Dog [A] kan man se andre folk bagfra, hvis man kigger godt efter i mellemgrunden på billedet. Jeg [S] lægger også hurtigt mærke til, at det er en travl gade, da jeg ser mange folk, som går på gaden. Der [FS] er et blåt skilt, som viser en mand, som går ned af nogle trapper. Dette [S] kunne måske henvise til en metro og være medvirkende til, at dette sted er en travlt gade.

Elevtekst 16: Beskrivende tekst skrevet af en elev fra 9. årgang.

Elevtekst 16 består af 6 syntaktiske perioder bestående af 7 hovedsætninger og 8 ledsætninger. Selvom teksten er kort, er der således eksempler på alle tre typer ledsætninger, hvoraf der er flest indlejrede, nominale ledsætninger. Kun en enkelt gang anvendes der en parataktisk konjunktion i forbindelse med en ellipse, men derudover består teksten i udpræget grad hypotaktiske konstruktioner. Som vi også så ved elevtekst 13 og 14, der ligeledes er beskrivende tekster, er forfelterne i teksten primært realiseret af subjekt eller formelt/foreløbigt subjekt. Der ses ingen eksempler på ledsætninger i forfeltet, hvorfor der er udpræget højretyngde i teksten. Det er således især disse hypotaktiske konstruktioner, som kan forklare forskellen i sætningslængde mellem de to tekster. Den øgede dependensafstand kan især tænkes at hænge sammen med det øgede antal nominale ledsætninger, som indlejres i og dermed udbygger længden og afstanden mellem 'head' og 'dependents' i nominalfraserne.

5.7.3 Elevtekst 17, 18, 19 og 20

De sidste fire elevtekster i dette tekstudvalg er eksempler på tekster med meget høje værdier i sætningslængde og dependensafstand. Den to første tekster, vi skal se nærmere på, er elevtekst 17 og 18, som har en gennemsnitlig sætningslængde på henholdsvis 66 og 62 ord. De er således blandt de tekster i korpusset med den højeste sætningslængde. En gennemsnitlig sætningslængde på 66 og 62 er selvfølgelig outliers og dermed langt over, hvad der er rimeligt for tekster på dette niveau. Selv professionelle akademiske tekster er ikke i nærheden af at have værdier på dette niveau, men ligger i et studie af Chen et al. (2020, s. 1779) i stedet mellem 22 og 27 ord per sætning. Teksterne er således primært taget med som ekstreme cases og eksempler, der skal illustrere, hvilke former for syntaktiske konstruktioner der kan skabe kunstigt høje værdier i sætningslængde. Elevtekst 17 er en berettende tekst skrevet af en elev fra 6. årgang. Foruden den høje sætningslængde har den en gennemsnitlig dependensafstand på hele 3,94, hvilket placerer værdien i fjerde og øverste kvartil for årgangen.

en typisk dag i corona [S] ville være **at jeg vågnede kl 8.30 eller 9.00**, og så [A] spiste jeg morgenmad og så [S] går jeg ud **hvis der ikke er et goggle meet** og **når jeg er færdig med det** [A] og så laver jeg lektier og så [A] spiller jeg måske unu og så [A] skal vi spise aftensmad og så [A] ser jeg måske youtube og så [A] skal jeg sove.

Elevtekst 17: Berettende tekst skrevet af en elev fra 6. årgang.

Elevtekst 17 består af bare én lang syntaktisk periode bestående af 8 hovedsætninger og 3 ledsætninger. Ledsætningerne er placeret forskellige steder i sætningskemaet (se bilag 1, tabel 42); de to adverbielle ledsætninger er placeret i henholdsvis forfeltet og ekstrafeltet, mens den nominale ledsætning er placeret i ekstrafeltet, selvom den faktisk er direkte objekt i sætningen.²⁵ Der er altså en vis variation i anvendelsen af ledsætninger i teksten. Der er til gengæld ikke megen variation i den resterende del af teksten. Den parataktiske konjunktion 'og' anvendes hele syv gange til at binde hovedsætningerne sammen, og i seks af de syv tilfælde akkompagneres konjunktionen med et tidsadverbielt 'så' i forfeltet.

²⁵ Det er omdiskuteret, om objektledsætninger skal placeres på N-pladsen, i ekstrafeltet eller alternativt i et nyt felt kaldet 'tungtledsposition', se fx Christensen & Christensen (2019: 239-240). Jeg følger imidlertid Hansen (1970: 119-120) og placerer for overskuelighedens skyld objektledsætninger i ekstrafeltet.

Der er altså tale om en form for overforbrug af paratakse, hvor eleven binder hovedsætningerne sammen ved at gentage den samme konjunktivfrase ”og så” i stedet for at adskille hovedsætningerne med punktum. Det er således parataksen, der primært er med til at skabe den høje sætningslængde. Den høje dependensværdi skyldes derimod højst sandsynligt de hypotaktiske konstruktioner, idet dependensrelationer typisk går på tværs hoved- og ledsætninger.

Ser man nærmere på elevtekst 18, ses samme tendens. Elevtekst 18 er en berettende tekst skrevet af en elev fra 7. årgang. Teksten har en sætningslængde på 62 ord og en dependensafstand på 3,29, hvilket placerer værdien i fjerde kvartil for årgangen.

Jeg [S] står op og tager noget tøj på og spiser morgenmad og så [A] ser jeg en video og så [A] lavede jeg nogle lektier og så [A] spiste jeg noget middagsmad og så [A] tog jeg på arbejde og **da jeg fik fri fra arbejde** [A] lavede jeg lidt lektier og så [A] spiste jeg aftensmad og så [A] spillede jeg lidt og så [A] gik jeg i seng

Elevtekst 18: Berettende tekst skrevet af en elev fra 7. årgang.

Elevtekst 18 minder på mange måder om elevtekst 17. Den består af én lang syntaktisk periode med 11 hovedsætninger og bare én ledsætning. Som ved elevtekst 17 ses der 10 anvendelser af den parataktiske konjunktion ’og’, som i 7 tilfælde akkompagneres af tidsadverbialet ’så’ i forfeltet. Det er altså samme mønster som i elevtekst 17. Elevtekst 18 adskiller sig dog ved at anvende tre ellipsekonstruktioner i starten af teksten, hvilket vi ikke så eksempler på i forrige tekst. Forfeltet i første sætning er realiseret ved subjekt, men herefter er forfeltet enten tomt (ellipse) eller realiseret gennem samme konjunktivfrase (”og så”). Der ses kun et enkelt eksempel på en ledsætning i forfeltet. Også i denne tekst er det overforbrug af parataktiske konstruktioner, der skaber den høje sætningslængde. Det er imidlertid værd at bemærke, at det kun er de berettende tekster, der har de meget høje outliers i sætningslængde (se tabel 18). Der er således grund til at antage, at det er den berettende teksttype og dennes krav til tidsmæssig kohærens og kohæsion, der gør, at nogle elever forfalder til overforbrug af paratakse gennem anvendelse af konjunktivfrasen ”og så”, fordi denne frase binder handlingsforløb sammen i tid. I hvert fald har de to elevers beskrivende tekster en kortere sætningslængde på henholdsvis 12,4 og 34,4, hvilket altså – især i tilfældet med elevskriveren til elevtekst 17 – indikerer, at den berettende teksttype er en vigtig faktor i at muliggøre de meget høje outlierværdier.

Vi vender nu blikket mod to tekster, som også repræsenterer høje sætnings- og dependensværdier, men som – i hvert fald i den ene variabelværdi - kun ligger lidt højere end tredje kvartils-værdierne for de respektive årgange, og som dermed ikke på samme måde som de to forrige tekster er eksempler på outliers. Elevtekst 19 er en beskrivende tekst skrevet af en elev fra 8. årgang. Teksten har en gennemsnitlig sætningslængde på 19,06 ord og en gennemsnitlig dependensafstand på 2,79, hvilket placerer begge værdier i fjerde kvartil for årgangen.

Illustrationen [S] viser en ung voksen, med et ubehageligt ansigtstræk. Eller det [S] viser det af ansigtet **man kan se over beskyttelsen**. Hun [S] er omringet af mennesker uden beskyttelse, og føler sig nok overdramatisk. Manden bag hende [S] ignorere alle former for varsler **der har været mod spredningen af den nye virus**. Enten [A] vælger han at undgå virkeligheden eller er bare ikke klar til at tage imod den. Nyheden om **at virussen rammer ekstra hårdt mod rygere med dårligere lunger** [S] er nok ikke nået hans vej. De mange andre mennesker **som bevæger sig rundt om den anderledes pige, som faktisk tænker på sig selv og dem omkring hende** [SF]. Hun [S] prøver at beskytte de tætte mennesker mod den ukendte virus. **Hvad der er sket før billedet blev taget** [S] er svært at vurdere, Men med den kvindeliges ansigtsudtryk [A] kan man blive ledt frem til **at hun ikke har det super behageligt i den situation hun er i, i øjeblikket**. Således [A] kan hendes ansigtsudtryk også have noget med **hvad det er på hendes elektroniske device, som hun holde om beskyttende med begge hænder**. De slørede ansigter i baggrunden [S] viser et stort smil, på den mandlige person **som nærmer sig**. Han [S] er iført en dunjakke med en meget blændende farve på indersiden. Efter billedet [A] ville jeg regne med **at den kvindelige figur ville gå sin egen vej uden at bevæge sig for tæt på de andre rundt om hende**. Den rygende man bag kvinden [S] ville nok blive stående og færdiggøre hans forgiftning af sine lunger eller fortsætte gående **indtil det foldede papir med nikotin er lille og ligegyldig nok til at smide på jorden og kvase under sin sko**. Stemning [S] afhænger af de forskellige personer. De [S] har nok alle deres eget synspunkt for situationen.

Elevtekst 19: Beskrivende tekst skrevet af en elev fra 8. årgang.

Teksten består af 15 syntaktiske perioder bestående af 20 hovedsætninger, ét sætningsfragment og 13 ledsætninger. Der ses blot 5 eksempler på parataktiske konjunktioner, som binder hovedsætninger sammen, mens der til gengæld er tre eksempler på ellipsekonstruktioner. Der er meget begrænset brug af adverbelle ledsætninger i teksten, mens der til gengæld er meget udbredt brug af både adjektiviske ledsætninger, som er integreret i større nominalfraser, samt nominale ledsætninger, som primært agerer som objekt, men i et enkelt tilfælde ses også, at to nominelle ledsætninger står i forfæltet som subjekt. Det, der særligt kendetegner syntaksen i elevtekst 19, er, at der er en stor variation i, hvordan de

enkelte syntaktiske perioder er skruet sammen. Vi ser således både venstre- og højre-tyngde i sætningerne, og ser man alene på længden af de enkelte syntaktiske perioder, er der ligeledes en vis variation. De fleste perioder er på 15 ord eller mere, men der ses også eksempler på perioder med en længde på 6 og 9. Det, der adskiller denne tekst fra de øvrige tekster i tekstudvalget, er således, at den indeholder en mere varieret syntaks, og at den indeholder nogle former for kompleksitet, som vi ikke har set i de øvrige tekster i dette tekstudvalg, fx anvendelse af nominelle ledsætninger som subjekt eller flere indlejrede relativsætninger i en nominalfrase.

Den sidste elevtekst i dette tekstudvalg er elevtekst 20, som er en berettende tekst skrevet af en elev fra 7. årgang. Teksten har en gennemsnitlig sætningslængde på 18,57 ord og en dependensafstand på 2,98, hvilket placerer begge værdier i fjerde kvartil for årgangen. Både sætningslængde og dependensafstand ligger tæt på de tilsvarende værdier for elevtekst 19.

Jeg [S] har valgt en hverdag i corona krisen, **hvor vi var begyndt at have hjemmeundervisning**. Jeg [S] vågnede altid kl 7:50 ca og stod op. Så [A] gik jeg ned og fik noget morgenmad og derefter [A] fandt jeg mine skoleting frem og var klar til at lave skolearbejde 8.05. Så [A] gik jeg i skole til 12-12:30 tiden **hvor vi fik pause** og derefter [A] var jeg i skole til 14-15 ca og fik derefter fri. så [A] plejede jeg at spise frokost og derefter løbe en tur og gå over at spille fodbold. [*pos. pronomen + substantiv*] [S] er [*præposition + substantiv*] så jeg [S] måtte slet ikke være sammen med nogle under lockdown så derfor [A] dyrker jeg meget motion. Om aftenen [A] spiste jeg aftensmad og var f.eks sammen med familie og så [A] gik jeg i seng ved 23-24 tiden.

Elevtekst 20: Berettende tekst skrevet af en elev fra 7. årgang.

Der er betragtelige forskelle på elevtekst 19 og 20, selvom de ligger tæt på hinanden i de to variabelværdier. Hvor elevtekst 19 indeholdt en stor variation af hoved- og ledsætninger, er der i elevtekst 20, som indeholder syv perioder, blot to ledsætninger (begge adverbialle) og 19 hovedsætninger, hvilket er en betragteligt lavere frekvens af ledsætninger, også hvis man tager højde for forskelle i tekstlængde imellem de to tekster. Teksten indeholder hele 12 parataktiske konjunktioner, og der er 6 ellipser, hvilket forklarer den høje sætningslængde. Sætningerne har en forholdsvis ensartet opbygning med et tidsadverbial eller subjekt i forfeltet samt et objekt på N-pladsen eller en præposition på N-pladsen. Elevtekst 20 bekræfter desuden den tendens ved de berettende tekster, at der er udpræget

brug af tidsadverbialer i forfeltet, og at beretningerne giver mulighed for anvendelse af mange parataktiske konjunktioner.

5.7.4 Opsummering på analyse af tekstudvalg 3

Jeg indledte analysen af tekstudvalg 3 ved at stille to spørgsmål, som havde rejst sig på baggrund af de kvantitative analyser af de syntaktiske variable. Hvilke syntaktiske forskelle kan være med til at forklare stigningen i sætningslængde og dependensafstand mellem mellemtrin og udskoling? Og gemmer der sig skjulte syntaktiske udviklinger bag den stagnering, der kunne observeres i udskolingsårgangene i de to kvantitative variable? Disse spørgsmål er imidlertid store og vanskelige spørgsmål, som det naturligvis ikke kan lade sig gøre at besvare hverken entydigt eller fyldestgørende på baggrund af kvalitative analyser af 9 tekster. Analyserne giver imidlertid et konkret indblik i, hvad der ligger bag tallene og variationerne og kan derfor give en indikation på, hvad der kunne være i hvert fald en del af svaret på de to spørgsmål.

Ud fra analysen er det klart, at variabelen sætningslængde ikke øges som en direkte funktion af bestemte syntaktiske træk. I stedet er der flere forskellige syntaktiske konstruktioner, som kan føre til høje værdier i sætningslængde. Til trods for denne variation, er det imidlertid muligt at udpege to overordnede veje til øget sætningslængde. Den første vej er en ”kunstig” vej, hvor sætningslængden øges gennem overforbrug af paratakse. Dette så vi i særlig grad et eksempel på i elevtekst 17 og 18, som begge havde ekstraordinært høje værdier i sætningslængde, men også delvist i elevtekst 20, som havde en sætningslængde omkring tredje kvartil for årgangen. I de to førstnævnte tekster var der kun ganske få ledsætninger, og i stedet bestod teksten primært af hovedsætninger, som blev koblet sammen gennem repetition af den samme konjunktivfrase ”og så”. Teksterne er således ikke syntaktisk set mere komplicerede end elevtekst 13 og 14, som begge havde meget lave sætningslængder, men forskellen ligger her i, at hovedsætningerne i elevtekst 17 og 18 blev bundet sammen med brug af parataktiske konjunktioner frem for adskilt af punktum. I elevtekst 20 sås samme tendens med brug af paratakse, dog lidt mindre udbredt, men til gengæld var der flere eksempler på ellipsekonstruktioner, hvor handlingsforløbene i hovedsætningerne blev koblet sammen gennem brug af konjunktionen ’og’ og med samme subjekt som reference for hovedsætningernes verballed. Det er kendetegnende for alle tre tekster, at det er berettende tekster, og sammenligner man tabel 18 og

19, som viser fordelingen af sætningslængde inden for de to teksttyper, er det klart, at det særligt er den berettende teksttype, som muliggør outliers med høj sætningslængde og dermed overforbrug af paratakse. Dette hænger sandsynligvis sammen med, at berettende tekster udfolder sig i tid, og de stiller dermed store krav til elevernes anvendelse af temporale kohæsive bånd, til hvilket formål eleverne altså kan anvende konjunktivfrasen ”og så”. Dette bekræftes også, hvis man ser på forskelle i realiseringerne af forfeltet i de to teksttyper, hvor de berettende tekster er domineret af temporale adverbialer, mens de beskrivende tekster primært har subjekt eller foreløbigt/formelt subjekt i forfeltet.

Den anden vej til øget sætningslængde er i højere grad knyttet til forskellige former for syntaktisk kompleksitet i teksterne. Hvor den første vej til øget sætningslængde, som jeg netop har beskrevet, siger mere om begrænsningerne ved den måde, som sætningslængde måles på i denne afhandling, siger den anden vej til øget sætningslængde mere om elevernes skriftsproglige udvikling. Starter man med at sammenligne elevtekst 14 (5. årg.) og 16 (9. årg.), som begge er beskrivende tekster med værdier i sætningslængde og dependens i henholdsvis første kvartil og omkring medianen, er det særligt to træk, der adskiller de to tekster. For det første ser vi eksempler i elevtekst 16 på, at alle tre typer af ledsætninger er integreret i teksten. Mens elevtekst 14 kun har to adjektiviske relativsætninger, der lægger sig til to nominalfraser, ser vi altså en langt mere udpræget brug af hypotaktiske konstruktioner i elevtekst 16. Alle ledsætningerne i elevtekst 16 er imidlertid placeret sidst i sætningen, så der er udpræget højretyngde i teksten. Vi skal i stedet frem til elevtekst 19 (8. årg.), som har høje variabelværdier i fjerde kvartil, før vi ser eksempler på ledsætninger, der er placeret i forfeltet. Her ser vi både eksempler på nominelle ledsætninger, som agerer som selvstændigt subjekt, og vi ser eksempler på adjektiviske relativsætninger, der er integreret i en større nominalhelhed som subjekt i forfeltet. Det, der adskiller elevtekst 16 og 19 er således ikke bare, at syntaksen er mere kompliceret, men derimod at der er større variation (forstået som flere forskellige kombinationer af forskellige syntaktiske træk) i den syntaktiske kompleksitet. Der er således noget, der tyder på, at syntaktisk variation kan være med til at forklare stigninger i både sætningslængde og dependensafstand. En væsentlig faktor for variation er selvfølgelig tekstlængde, hvilket også ses ved, at den gradvise øgning i syntaktisk kompleksitet og variation, som vi har set mellem de tre tekster, følges af en gradvis øgning i tekstlængde på 35% mellem elevtekst 14 og 16 og 200% mellem elevtekst 16 og 19.

Ser man på tilsvarende udviklingstendenser i de berettende tekster, der er inkluderet i tekstudvalget, ses flere overlap. Der ses således også en gradvis øgning af både syntaktisk kompleksitet og variation – i hvert fald mellem elevtekst 12 og 15. Det, der særligt adskiller den syntaktiske udvikling i de to teksttyper, er imidlertid, at den primære udvikling i de berettende tekster sker i de adverbielle led. Elevtekst 12 (5. årg.), som har lav sætningslængde og dependensafstand, består af to adverbielle ledsætninger, der har en kausal og argumenterende funktion som belæg til en forudgående påstand. De er således placeret efter hovedsætningen. De temporale adverbialer i teksten består fortrinsvis af enkeltord og er typisk placeret i forfeltet (fx ”Efter havde vi skole” og ”Derefter skulle i seng”). Ser man herefter på elevtekst 15 (7. årg.), som har variabelværdier omkring medianen, er det tydeligt, at den indeholder en betragtelig udvidelse af antallet af ord i de adverbielle led. Der ses stadig eksempler på temporale enkeltadverbialer i forfeltet, men teksten har langt mere udpræget anvendelse af adverbielle ledsætninger, herunder de temporale, og der ses ligeledes flere eksempler på temporale ledsætninger i forfeltet. Elevtekst 20, som er et eksempel på en berettende tekst med høje variabelværdier, bryder dog tendensen. Her ses bare to adverbielle ledsætninger i hele teksten. Til gengæld ses hele 12 parataktiske konjunktioner og 6 ellipser, hvorfor elevtekst 20 altså snarere er et eksempel på en tekst, der følger den ”kunstige” udviklingsvej med overforbrug af paratakse, som jeg beskrev før. Desuden er elevtekst 20 hele 61% kortere end elevtekst 15, hvilket kan forklare en del af den manglende variation. Opsummerende om forskellen på de to teksttyper kan man dog sige, at de beskrivende tekster anvender en større grad af især adjektiviske og nominelle ledsætninger end de berettende tekster, hvilket bl.a. fører til længere nominalfraser og dermed længere sætninger og dependensafstand, mens det i udpræget grad er anvendelsen af de adverbielle ledsætninger, der fører til øget sætningslængde og dependensafstand i de berettende tekster.

Dependensværdierne i de 9 tekster i tekstudvalget har vist en stærk sammenhæng med sætningslængde, hvilket indikerer, at de syntaktiske udviklinger, som har været med til at skabe en høj sætningslængde, også har bidraget til at øge dependensafstanden. Dette giver god mening i den forstand, at øget hypotakse og nominalhelheder med efterstillede adjektiviske ledsætninger selvfølgelig bidrager til en mere kompleks og integrerede syntagmehelheder og dermed længere afstande mellem hovedordet og dets dependenter. Der

er til gengæld også noget, der tyder på, at den automatiske dependensanalyse har begrænsninger ift. meget lange parataktiske sætninger. Dette ses fx i elevtekst 17, som har en gennemsnitlig dependensafstand på hele 3,94, hvilket er blandt de højeste i hele korpusset. Selvom teksten har en enkelt lang nominalfrase og tre ledsætninger, hvoraf en er nominal og agerer som objekt, er der imidlertid ingen yderligere grund til, at elevtekst 17 har en dependensværdi, der er henholdsvis 52% og 41% højere end elevtekst 15 og 19, som er de to tekster, der i de kvalitative analyser viste sig at have mest kompleks og varieret syntaks. Dependensværdien for elevtekst 17 siger således mere om begrænsningerne ved analyseværktøjet end om dependensstrukturerne i elevens tekst.

5.8 Delkonklusion på den diakrone analysedel

I den diakrone analysedel har jeg undersøgt, hvordan de kvantitative, sproglige variable har udviklet sig relativt til klassetrin, og jeg har foretaget kvalitative, grammatiske analyser af tre tekststudvalg af tekster, som tilsammen har indeholdt 20 elevtekster. Denne delkonklusion på analysen falder i to dele. Den første del er knyttet til forskningsspørgsmål 1 og handler specifikt om de resultater af analysen, som siger noget om elevernes skriftsproglige udvikling. Den anden del af resultaterne fra den diakrone analysedel er knyttet til forskningsspørgsmål 2 og handler om de metodologiske potentialer og udfordringer ved at undersøge skoleelevers skriftsproglige udvikling ved hjælp af automatiserede analyseværktøjer.

5.8.1 Skriveudviklingspointer

Den første del af delkonklusionen handler specifikt om de analyseresultater, der beskriver elevernes skriveudvikling, og som er afstedkommet af både de kvantitative analyser af hele korpusset samt de leksikogrammatiske analyser af de tre tekststudvalg. De tre spørgsmål, som er knyttet til forskningsspørgsmål 1, hænger imidlertid tæt sammen, og svarene på dem vil derfor gribe ind i hinanden. Det giver derfor ikke mening at besvare dem et ad gangen, så jeg vil i stedet bygge denne delkonklusion op omkring fem overordnede pointer om skriftsproglig udvikling, som samlet set er afstedkommet af den diakrone analysedel. For mere udspecificerede svar på forskningsspørgsmålenes enkelte spørgsmål henvises til afhandlingens samlede konklusion i kapitel 7.

5.8.1.1 Aldersmodning som proxy snarere end som strukturerende princip

Den første pointe angår den modningsteori, som jeg knyttede den diakrone analysedel op på (se afsnit 2.3.1). De kvantitative analyser af de fem lingvistiske variable viser kun delvist og i varierende grad tegn på lineære udviklinger. Tydeligst er det ved tekstlængde, som stiger forholdsvis lineært frem til 8. årgang, hvorefter der ses et fald. Leksis- og syntaksvariable viser ikke samme tendens, men viser til gengæld markante forskelle mellem mellemtrin og udskoling. At netop tekstlængde hænger godt sammen med aldersudvikling, er ikke så overraskende. Som jeg også har pointereret tidligere, er det velbeskrevet i forskningslitteraturen, at tekstlængde har meget stærk forklaringskraft både ift. aldersmæssig udvikling og ift. at forudsige tekstkvalitet på tværs af uddannelsesniveauer (Crossley, 2020). Hudson (2009) beskriver i denne forbindelse, at en del aldersmæssig udvikling kan (enten direkte eller indirekte) tænkes at hænge sammen med den naturlige kognitive udvikling i arbejdshukommelse, som elever undergår i skoleårene. Jo større arbejdshukommelse, desto længere tekster kan eleverne skrive. Dette er selvfølgelig ikke helt så enkelt, og det gælder således heller ikke for de øvrige lingvistiske variable. Ser man nærmere på fordelingen af variabelen tekstlængde inden for de to teksttyper, er det imidlertid også klart, at der er stor variation i tekstlængde inden for hver årgang, og denne variation ser ud til at øges over tid. Den største spredning i sætningslængde ses således i udskolingen, så det, der adskiller udskolingsårgange fra mellemtrinsårgange, er, at selvom de korteste tekster forbliver på samme niveau på tværs af årgange, så bliver de længste tekster længere i udskolingen. Der er således eksempler på tekster med stor aldersmæssig spredning, som i de kvantitative variable ligner hinanden meget.

Det giver på den ene side god mening fra et didaktisk perspektiv at betragte skriftsproglig udvikling som en diakron proces struktureret efter klassetrin. Vi tænker intuitivt faglig udvikling i skolen som en gradvist udviklende proces i tid, hvor eleverne bliver kontinuerligt dygtigere og i stand til at klare og løse mere komplicerede opgaver for hvert nyt skoleår. En analyse af Jeffery et al. (2018) viser, at dette også gør sig gældende i måden, skriveudvikling konciperes på i Fælles Mål for danskfaget. Her forstås skriveudvikling som en gradvis bevægelse mod øget kontekstualisering og en forholdsvis lineær bevægelse fra narrative til mere analytiske genrer. Det er altså på denne måde naturligt at tænke

skriftsproglig udvikling som en funktion af aldersmodning, eftersom det er denne struktur, som er udgangspunkt for skrivning inden for en skolekontekst og dermed også et vilkår for dansklærernes skrivedidaktiske arbejde. Men som det har fremgået af denne diakrone analyse, er skriftsproglig udvikling, som det forstås inden for rammerne af de lingvistiske variable, ikke en fuldstændig lineær proces – hverken hvis man ser på medianværdier eller på den øvrige fordeling. Dertil er den interne variation på årgangene for stor. I stedet for at betragte klassetrin som et strukturerende princip for skriftsproglig udvikling, giver det i stedet mening at betragte klassetrin enten som en indirekte proxy for akkumuleret skriveerfaring i skolen eller som en proxy for en mere snæver kognitiv dimension hos eleven som beskrevet af Hudson (2009). I den synkrone analysedel, som følger i forlængelse af denne delkonklusion, betragter jeg således klassetrin som en variabel på linje med de øvrige ekstra-lingvistiske variable.

5.8.1.2 Udviklingsplateauer og overforbrug som træk ved skriftsproglig udvikling

I forlængelse af ovenstående pointe om den manglende lineære udvikling i variablene, er den anden pointe knyttet til nogle af de strukturer og tendenser, som træder frem i analysen i stedet for den lineære udvikling. Der er her særligt to tendenser, som er interessante i et udviklingsperspektiv. Den første angår muligheden for udviklingsplateauer i bestemte sproglige træk i teksterne, dvs. muligheden for, at bestemte træk stagnerer og enten bibeholder samme værdi over en periode eller måske endda falder lidt. Den anden tendens angår elevernes overforbrug og gentagelse af bestemte sproglige træk.

Udviklingsplateauet som et kendetegn ved skriftsproglig udvikling er bl.a. tydeligt ved OVIX, idet analyserne viser, at der her er tale om et udviklingsplateau på tværs af alle fem årgange. Variablens medianværdier ligger og svinger stabilt omkring 50 (laveste er 49,28 ved 7. årgang og højeste er 52,54 ved 9. årgang), og selvom der kan observeres en svag stigning på 1,8 (svarende til 3,6%) mellem 5. og 9. årgang, er der ingen systematik mellem årgangene. Hvor udviklingsplateauet for OVIX ser ud til at være gennemgående på alle fem årgange i indeværende studie, er udviklingsplateauet for sætningslængde og dependensafstand særligt for udskolingen. Her ses tydelig forskel på mellemtrins- og udskolingsårgangene, men til gengæld ses der ikke systematiske stigninger fra 7.-9. årgang. Tværtimod ses der for sætningslængde en lille fald fra 7. til 9. årgang i begge teksttyper.

Disse udviklingsplateauer er interessante, fordi de viser, at selvom skriftsproglig udvikling selvfølgelig i et vist omfang er et lineært fænomen, som udfolder sig diakront i tid, så kan en del af denne udvikling bestå af plateauer, hvor der ikke sker udvikling inden for bestemte sproglige træk²⁶.

Dette kan selvfølgelig skyldes en form for komplementaritet mellem variablene, hvor en stagnering i én variabel fører til udvikling i en anden variabel (der ses fx delvise stigninger i udskolingen i variablene tekst- og ordlængde), men det kan, som jeg senere vil uddybe i afsnit 5.8.1.5, også skyldes, at den ubestemte entalsform af fænomenet 'udvikling' ikke er tilstrækkelig. Skriftsproglig udvikling kan i stedet tænkes som et multidimensionelt og pluralistisk fænomen, hvor udviklingstrækkene bliver sløret og kommer til at fremstå som stagnerende, når det som i denne analyse undersøges som et samlet konstrukt struktureret efter klassetrin.

Den anden tendens i elevernes tekster, som jeg vil fremhæve her, er knyttet til elevernes overforbrug af bestemte sproglige træk. Hvor udviklingsplateauet som et træk ved elevernes skriveudvikling blev tydeligt gennem de kvantitative analyser, blev overforbruget som træk tydeligt gennem de kvalitative analyser. Dette kommer særligt til udtryk i elevtekst 6 og 18. I elevtekst 6 er overforbruget knyttet til de kohæsive bånd, som eleven anvender til at binde handlingsforløbene i beretningen sammen. Her er det særligt anvendelsen af temporale konjunktioner ('efter', 'så' etc.) og personlige referentkoblinger, som eleven anvender i meget høj grad, til trods for at teksten har en meget lav grad af ordvariation (OVIX). I de efterfølgende tekster i tekstudvalget, hvor variabelværdierne stiger, falder anvendelsen af de specifikke kohæsive bånd til gengæld. I elevtekst 18 er overforbruget knyttet til anvendelsen af parataktiske konjunktioner, hvor især konjunktivfrasen "og så" bliver gentaget og anvendt til at binde hovedsætningerne sammen. Begge elevtekster overforbruger og gentager således bestemte sproglige træk, som i kraft af den redundante brug bidrager til en mindre varieret og simple tekst.

At dette er et træk ved elevers sprogudvikling, er velkendt i forskning i elevers intersprog inden for andetsprogstilegnelse. Her kendes fænomenet som 'overgeneralization'

²⁶ Udviklingsplateauet skal ikke her forveksles med begrebet 'fossilization' fra intersprogsforskningen (Se-linker, 1972), hvor udviklingsplateauet i intersproget refererer til det afsluttende niveau, hvorfra sproget – til trods for fejl og mangler – ikke udvikler sig yderligere. Det udviklingsplateau, jeg refererer til i ovenstående, er mere dynamisk og refererer til en periode uden udvikling i det lingvistiske træk, hvorefter det øges igen på et senere tidspunkt i elevens udvikling (jf. Pettersson, 1977).

og refererer til det fænomen, hvor en elev kommer til at overforbruge et nyligt tillært sprogligt træk eller et bestemt sprogligt mønster. Idéen i intersprogsteorien er dog også, at overgeneraliseringen sker på baggrund af indflydelse fra L1-sproget, og at eleven derigennem kommer til at anvende trækket i situationer, hvor det ikke hører hjemme (Selinker, 1972). Et sådant tilfælde er der selvfølgelig ikke tale om her. De sproglige træk, som overforbruges, er således ikke placeret grammatisk forkert, men er blot anvendt redundant og med høj frekvens. Dog er der især i tilfældet med overforbrug af paratakse en klar parallel til det oprindelige intersprogsbegreb, idet gentagelsen af bestemte konjunktivfraser er et velkendt træk ved talte narrativer (Juzwik, 2012), og især ved yngre børn er det velkendt, at der kan være meget lidt variation i anvendelsen af konjunktionsfraser i talte fortællinger (Geva, 2007). Det er således sandsynligt, at gentagelsen af konjunktivfrasen ”og så” i elevtekst 18 er en reminiscens af elevens talesprog, hvilket på mange måder kan ses som en parallel til påvirkning fra L1-sproget, som kendes fra intersprogsforskningen i andetsprogstilegnelse. Denne pointe findes også hos Myhill (2008), som peger på, at der er visse syntaktiske træk i elevtekster, som befinder sig i det, hun kalder ”the speech-writing-interface”, hvilket indikerer, at eleven endnu ikke mestrer skriftsprogets særlige retoriske dimension og derfor må forlade sig på sine erfaringer fra talesproget. Overforbrug af bestemte træk kan derfor også forstås som en form for compensation, idet eleven mangler de skriftsproglige ressourcer til at udtrykke sig mere præcist og/eller varieret.

5.8.1.3 Fra længde som proxy for kompleksitet til sproglig variation

Den tredje pointe fra analysen er knyttet til elevteksternes grammatiske kompleksitet. Især i tekststudvalg 3, hvor der var fokus på syntaks, var det tydeligt, at de elevtekster, som havde den mest komplicerede syntaks, ikke nødvendigvis var dem, der havde de længste syntaktiske perioder eller de højeste dependensafstande, men derimod dem der havde størst variation i sætningsopbygningen. Elevtekst 15, som havde variabelværdier omkring medianen for 7. årgang, er et godt eksempel på dette. Her så vi en meget stor variation i måden, sætningerne var opbygget på. Der var både eksempler på lange adverbelle led og ledsætninger i ekstra-feltet, adverbelle prædikativer på N-pladsen, tidsadverbelle ledsætninger i forfeltet samt udpræget vekslen mellem paratakse og hypotakse. Der var ligeledes stor variation, hvis man ser på længden af sætningerne samt på længden af forfel-

terne. Samme syntaktiske kompleksitet kunne kun ses i elevtekst 19, som havde høje variabelværdier, mens elevtekst 20, som også havde høje variabelværdier i fjerde kvartil, var syntaktisk meget simple end elevtekst 15 og primært domineret af enkeltadverbialer i forfæltet eller hovedsætninger med klassisk SVO-opbygning (subjekt-verbal-objekt).

At variation i syntaksen er et tegn på skriftsprog udvikling, er også en central pointe hos Myhill (2008), som peger på, at gode elevtekster bl.a. er karakteriseret ved brugen af sætninger med varieret længde, en mere varieret anvendelse af forskellige hypotaktiske konjunktioner, hyppigere subjekt/verbal-inversion og flere passivkonstruktioner. Selvom jeg ikke har et mål for tekstkvalitet i dette studie, korresponderer Myhills resultater på mange måder godt med fundene i de kvalitative analyser. Med variation som kvalitativt benchmark betyder det samtidig også, at der er naturlige grænser for, hvor høj en sætningslængde og dependensafstand gode elevtekster kan have, uden at dette går ud over, hvad der er kommunikativt hensigtsmæssigt for teksten. Som jeg senere vil uddybe i afsnit 5.8.2.1, er der således intervaller i de lingvistiske variable, inden for hvilke de fleste tekster befinder sig, og hvis variation skal indfanges af automatiserede analyseværktøjer kræver det naturligvis, at der udvikles lingvistiske mål, som ikke alene baserer sig på gennemsnitlig længde af en syntaktisk struktur, men som tværtimod er designet til specifikt at indfange syntaktisk variation.

5.8.1.4 Elevernes stemme som gennemgående træk

Den fjerde pointe, som viste sig på baggrund af de kvalitative analyser, er, at, selvom det ikke var et eksplicit fokuspunkt i analyserne, så viste det sig, at elevernes egne stemmer var eksplicit til stede i de fleste tekster. Det var naturligvis forventet ved de berettende tekster, eftersom skrivepromptet lagde op til, at eleverne skulle skrive med udgangspunkt i egne erfaringer og oplevelser. Mere overraskende var det til gengæld, at elevernes stemmer også var præsente i flere af de beskrivende tekster. Elevernes stemmer kommer til udtryk på to måder: enten som eksplicit førstepersonsfortæller eller gennem implicit evaluerende sprogbrug. Den første måde er der eksempler på i 6 ud af 9 beskrivende elevtekster fra de tre tekstudvalg. Dette ses typisk allerede i den første sætning, hvor der indledes med en metakommentar fra eleverne om deres valg af billede (fx ”Jeg har valgt billede 1.”). Den anden måde, elevernes stemme kommer til udtryk på, er gennem anvendelsen af evaluerende sprogbrug, hvor elevernes egne, personlige holdninger skinner igennem.

Som vi så i analysen af tekstudvalg 2 kom dette bl.a. til udtryk gennem anvendelsen af den kausale konjunktion 'fordi', som eleverne anvendte som led i et argument, og i mere epistemiske udsagn som fx "Man kan se at pigen med de røde shorts helt sikkert har tænkt sig..." eller vurderende udsagn som "Stemningen er lidt trist". I første eksempel er "helt sikkert" et eksempel på en såkaldt epistemisk booster, som eleverne anvender til at markere, at vedkommende er sikker på sandhedsværdien i sit udsagn (Hyland, 2018, s. 62). Disse eksempler er således forbundet med det, Martin (1992) kalder 'appraisal', som henviser til de lingvistiske markører i en tekst, som er udtryk for skribentens holdning eller vurdering. Ifølge Martin er realiseringerne af appraisal tæt knyttet til tekstens retoriske og kommunikative effekt, eftersom det hænger sammen med skribentens troværdighed og den kontrakt, som indgås mellem afsender og modtager. Dermed er elevens udvikling af stemme en vigtig del af skriveudviklingen (se Matsuda & Jeffery (2012) for overblik).

At tilstedeværelsen af elevens stemme er et gennemgående træk ved flere af teksterne fra de tre tekstudvalg, er ikke en overraskelse. Det ligger således i forlængelse af en af de centrale pointer fra Faglighed og Skriftlighed-projektet (Christensen et al., 2014; Krogh & Jakobsen, 2016), hvor forfatterne sammenlignede skrivekulturer i udskolingen og på gymnasiet. Her viste en af de væsentligste forskelle sig at være, at skrivning i de ældste klasser i grundskolen var karakteriseret ved, at elevernes stemmer havde fokus på egne holdninger og erfaringer, hvorimod der på gymnasiet var en mere analytisk og argumenterende stemme på spil. Det er således en nærliggende tanke, at netop forandringer i måden, hvorpå elevernes stemmer kommer til udtryk i teksten, er en vigtigt fokuspunkt for skriveudvikling i uddannelseskontekster. I analysen af tekstudvalg 2 sås det fx, at anvendelsen af kausale konjunktioner var et træk, som først var tilstede ved tekster med OVIX- og ordlængdeværdier fra omkring medianen og derover. Nordenfors (2011) har i denne forbindelse undersøgt, hvordan forskellige former for stemme forandrer sig i elevtekster fra 5.-9. klasses trin. Han konkluderer bl.a., at selvom anvendelsen af stemmemarkører overordnet stiger fra 5.-8. klasses trin og falder igen på 9. klasses trin, så sker der en markant forandring i elevernes anvendelse af stemme imod slutningen af udskolingen, hvor eleverne begynder at blande forskellige typer af stemme samt anvende en mere kommenterende stemme.

Thomson (2020) har i forlængelse heraf i en analyse af gymnasietekster vist, hvordan anvendelsen af stemmemarkører (Thomson kalder det dog ikke stemme, men 'meta-diskurs') varierer relativt til genren, og at udviklingen af stemme ikke er en lineær proces fra lav frekvens af markører mod høj frekvens af markører, men snarere er præget af nogle af de nonlineære tegn på fx overforbrug af bestemte træk, som jeg også selv har beskrevet i dette kapitels analyser. Skar et al. (2022) har desuden vist, at stemme i form af modtagerbevidsthed i teksterne ikke bare er et anliggende for skrivning på mellemtrinnet og i udskoling, men også findes i tekster på begynderstadiet/indskoling.

Jeg har ikke haft eksplicit fokus på stemme i analysen og resultaterne kan derfor ikke bidrage til en yderligere nuancering af, hvordan denne dimension udvikles. Der er således brug for flere undersøgelser af, hvordan elevernes stemme udvikler sig og kommer til udtryk i teksten, især fordi fagenes skrivepraksisser står tydeligere frem i de ældre klasser, og dermed kommer de fagspecifikke metagenrer (se afsnit 2.4.2 for uddybelse af Carters begreb om metagenrer) til at spille en endnu større rolle for, hvordan elevernes stemmer udvikler sig i et diakront perspektiv.

5.8.1.5 Fra skriftsproglig udvikling til skriftsproglige udviklinger

Den sidste skriveudviklingspointe fra den diakrone analyse angår entalsformen af begrebet skriveudvikling. Som vi så ved både de kvantitative og de kvalitative analyser var der en del sproglige forskelle relativt til teksttype. Ved de kvantitative variable var forskellene måske ikke store, men til gengæld forholdsvis entydige. De berettende tekster var baseret på medianværdierne både længere (indeholdt flere ord), indeholdt gennemsnitligt set længere ord, længere syntaktiske perioder og havde en højere dependensafstand. Udviklingen i IQR viste ingen systematik mellem teksttyper. Baseret på dette kan man sige, at de berettende tekster indeholdt en større grad af leksikal specifikation og syntaktisk kompleksitet end de beskrivende tekster. Dette blev imidlertid nuanceret gennem de kvalitative analyser af de tre tekstudvalg. Her viste analysen af tekstudvalg 2, at de beskrivende tekster i tekstudvalget havde længere og mere komplekse nominalfraser, hvilket bl.a. hang sammen med, at eleverne anvendte nominalfraserne til at udpege og beskrive forskellige elementer af det valgte billede (fx "pigen med de røde bukser", "restauranten der er i baggrunden"). Analysen af teksterne i tekstudvalg 3 viste ligeledes, at teksttype er en vigtig faktor for, hvilke syntaktiske elementer i teksterne der udvikler sig. I de berettende

tekster er det i høj grad forandringer i de adverbielle led, der er med til at øge værdierne i sætningslængde og dependensafstand. Dette hænger især sammen med, at det netop er de adverbielle led, der er med til at sikre dels temporal kohæsion, dels placere beretningerne i sted og rum. Så hvor det i de beskrivende tekster i høj grad er nominalfraserne, der forandrer sig og udbygges, er det i de berettende tekster altså de adverbielle led (herunder adverbielle ledsætninger), der forandrer sig. Dette ses også i forfeltetsrealiseringerne, hvor der meget tydeligt var tale om subjekt-realiseringer af forfeltet i de beskrivende tekster, mens der i meget udpræget grad blev anvendt temporale adverbialer i forfeltet i de berettende tekster.

Disse forskelle mellem teksterne indikerer således, at skriftsproglig udvikling på de undersøgte klassetrin ikke er et endimensionelt og lineært fænomen, som kan beskrives gennem generiske stadiebeskrivelser. Tværtimod er skrivning en funktionel aktivitet, hvor eleverne anvender forskellige sproglige ressourcer til at realisere deres tekster alt efter teksttype og kontekst. Det er således tvivlsomt, om det giver mening at beskrive generiske udviklingsniveauer for elevernes tekster uden at tage hensyn til de forskellige måder, som skrivning udvikler sig på alt efter teksttype. Det er med andre ord værd at overveje, om man i stedet for at tale om elevens skriftsproglige udvikling eller skriveudvikling i ental bør tale om elevernes skriftsproglige udviklinger i flertal. Dette skal imidlertid ikke forstås sådan, at skrivekompetence er et fuldstændig partikulært og fragmenteret fænomen. En dygtig elevskriver fra 9. årgang skriver jo næppe berettende tekster, som ligger over middel for årgangen, mens samme elevs beskrivende tekster svarer til en middeltekst på 5. årgang. Pointen er snarere, at den lingvistiske beskrivelse af elevernes skriftsproglige udvikling ikke kan forudsætte homogenitet på tværs af tekster og tekstaktiviteter, men må snarere tage udgangspunkt i en antagelse om, at eleverne på mellemtrinnet og i udskoling har nået et punkt i deres skriveudvikling, hvor deres skriftsproglige færdigheder er på et funktionelt stadié, hvor eleverne bevidst eller ubevidst anvender sproglige ressourcer målrettet teksttype og kontekst.

Selvom der implicit i mange skriveudviklingsstudier ligger en forudsætning om homogenitet på tværs af tekster, er det imidlertid en udbredt pointe i forskningslitteraturen, at gode elevtekster kan se ud på mange forskellige måder. I Palmérs (2018) undersøgelse af gymnasietekster, hvor hun undersøger kvantitative, lingvistiske træk ved teksterne, er en af de væsentligste konklusioner, at selvom de kvantitative træk generelt korrelerer med

tekst kvalitet, så varierer korrelationskoefficienten så meget relativt til skriveordren, at kvantitative analyser af tekster fremover bør målrettes de enkelte skriveordrer og genrer, som teksterne er skrevet i relation til. Samme pointe ses hos Allison et al. (2002), som viser, at elevteksternes genremæssige tilhørsforhold påvirker helt nede på ordniveau, hvordan og i hvilket omfang eleverne anvender forskellige hypotaktiske konjunktioner i teksten. Dette ligger således godt i tråd med analyserne i denne analysedel, som bl.a. viste, at adverbelle ledsætninger i udpræget grad var et træk ved de berettende tekster, mens relative, adjektiviske ledsætninger primært fandtes i de beskrivende tekster. Crossley et al. (2014) dokumenterede ligeledes i en clusteranalyse af højt bedømte tekster fra klassetrin svarende til 9. klassetrin og op til første år på videregående uddannelse, at gode elevtekster kan realiseres på mange forskellige måder, og at traditionelle lineære tilgange til skriveudvikling ”may lock writers and raters into a single method of developing and scoring a successful writing sample that may not represent the breadth and depth of successful writing samples available” (s. 206). Med andre ord er en generisk og lineær skriveudviklingsteori ikke i overensstemmelse med, hvordan skriveudvikling udfolder sig empirisk i forskellige situationer. Dette korresponderer godt med Fogal (2020), som kritiserer en del andetsprogsforskning for at være nomotetisk i den forstand, at den er orienteret mod at beskrive generaliserbare udviklingstrin for sprogudvikling, hvorved den overser ”the dynamic and highly contextualized nature of language development” (s. 576). I stedet foreslår Fogal ’variability’ – dvs. evnen til at variere sprog – som et vigtigt fokus i studier af sproglig udvikling. Fogal viser bl.a., hvordan en andetsproglørners ’learning trajectory’ udfolder sig som perioder med fremgang og perioder med tilbagegang. Derfor er skriveudvikling ikke lineær. Tværtimod kan ’variability’ indikere adaptiv adfærd, som er funktionelt og kommunikativt betinget, dvs. knyttet til skrivningens formål.

Den nonlineære opfattelse af skriveudvikling, som Crossley et al. (2014) og Fogal (2020) lægger op til, er imidlertid en central del af det såkaldte ”Skrivehjulet” (Berge et al., 2016), som er udviklet af norske forskere i forbindelse med Normprosjektet (se afsnit 1.3.2.2). Omdrejningspunktet for skrivehjulet er, at skrivning er en funktionel handling, som er drevet af teksternes formål og situationelle og kulturelle indlejring. Det er således disse dimensioner, der bl.a. er med til at afgøre elevernes valg af tekststruktur, modaliteter, ordforråd og grammatiske konstruktioner. Set fra dette perspektiv er det det, der bør kendetegne forskning i skrivning og elevtekster, således et blik for teksternes diversitet

og fleksibilitet, snarere end det er at undersøge bestemte normative former for kompleksitet i teksterne, som er afkoblet teksternes formål og kontekstuelle indlejring.

Dette er selvfølgelig lettere sagt end gjort. Som jeg beskrev i afsnit 2.4, er elevtekster retorisk set tvetydige, og den lingvistiske realisering af elevernes tekster drives således ikke nødvendigvis af et autentisk kommunikativt formål, men tværtimod kan særlige metafunktionelle og skolske vilkår i den situationelle og kulturelle indlejring være med til at påvirke, hvordan eleverne udformer sin tekst. I intersprogsforskningen har dette forhold fået Selinker (1972, s. 210) til at afvise brugen af tekster produceret i skolske kontekster som relevante forskningsdata, fordi de ikke er et udtryk for spontane, kommunikative ytringer og dermed ikke en afspejling af elevens reelle intersprogskompetence. Dette kritiske forbehold er på sin vis relevant, men samtidig er det klart, at i relation til skriftsprogsudvikling på modersmålet er skolen ikke en uvæsentlig kontekst, som kan ignoreres. Tværtimod er skolen et sted, hvor eleverne skriver meget og dermed en kontekst, hvor en meget stor del af elevernes skriveudvikling faciliteres. Det er derfor vigtigt at undersøge, hvordan skriveudvikling ser ud specifikt i en skolekontekst, men det er samtidig også vigtigt at have Selinkers forbehold in mente og huske på, at skriftsproglig udvikling studeret gennem elevtekster fra skolen er et udtryk for skriftsproglig udvikling i netop skolen, og havde man studeret elevernes funktionelle tekster skrevet uden for skolen, kunne resultaterne derfor have set anderledes ud. Det er imidlertid vigtigt, at vi har viden om, hvad der karakteriserer de tekster, som eleverne producerer i skolen, før vi fra et skrivepædagogisk perspektiv kan understøtte udviklingen af skriftsproglige kompetencer, der gør eleverne i stand at overgå fra at skrive skolske tekster på baggrund af stilladserede tekstaktiviteter til at skrive autentiske, kommunikative tekster uden for skolen.

5.8.2 Metodologiske pointer

Den sidste del af denne delkonklusion er knyttet an til forskningsspørgsmål 2 og er fokuseret på de metodologiske pointer, som er afstedkommet af den diakrone analyse. Som vi også så ved forskningsspørgsmål 1, griber undersøgelsesproblemerne til forskningsspørgsmål 2 ind i hinanden, og jeg besvarer derfor ikke spørgsmålene sekventielt, men samler i stedet svarene i fire overordnede metodologiske pointer.

5.8.2.1 Begrænsninger ved lingvistiske deskriptorer som kontinuerlige skalaer

Den første metodologiske pointe er knyttet til de lingvistiske variable. Ligesom det viste sig ikke at være meningsfuldt at betragte udvikling i de lingvistiske variable som en entydigt lineær proces, viste analysen også, at der ligeledes er begrænsninger ved at betragte de lingvistiske variable som kontinuerlige skalaer i sig selv. Dette var særligt tydeligt ved analysen af sætningslængde, hvor den ene af de to syntaktisk mest komplekse tekster (elevtekst 15) havde værdier i sætningslængde og dependensafstand omkring medianen, mens to af de syntaktisk set mest simple tekster (elevtekst 13 og 18) havde henholdsvis meget lave og ekstraordinært høje variabelværdier. Dette tyder således på, at syntaktisk kompleksitet kun øges med sætningslængde op til et vist niveau, hvorefter sætningslængde udtrykker noget andet end kompleksitet. Helt samme tendens kunne ikke ses i de kvalitative analyser af de øvrige tekstudvalg, men Nyström (2000) har blandt andet fundet lignende resultater for OVIX, som hun viste kun hang sammen med tekstkvalitet op til et vist niveau, hvorover OVIX i højere grad var udtryk for mere negative tekstegenskaber som fx bristede kohærensstrukturer.

5.8.2.2 Fra teknocentrisk til fagspecifik og forskningsdrevet udvikling af lingvistiske variable

Den anden metodologiske pointe er knyttet til selve forskningsdesignet i den diakrone analyse. Jeg har i analysen anvendt en eksplorativ tilgang med fokus på dominerende mønstre på tværs af teksterne (Tahmasebi et al., 2019), som jeg har fulgt op med kvalitative analyser af forskellige lingvistiske strukturer i teksterne. Disse analyser har haft det dobbelte formål dels at gøre os klogere på elevernes skriveudvikling, dels at gøre os klogere på hvilke sproglige konstruktioner der gemmer sig bag de numeriske billeder af teksterne, som er tegnet af de kvantitative analyser. At det har været interessant at undersøge kvalitativt, hvilke sproglige konstruktioner der egentlig ligger bag de kvantitative analyser, hænger sammen med to forhold. For det første har den kvalitative dimension været nødvendig, fordi de fleste af de lingvistiske variable, der er anvendt i denne afhandling, ikke i sig selv siger noget om den bagvedliggende grammatiske struktur i de tekster, de er anvendt på. De er med andre ord ikke på et tilstrækkeligt detaljeret sprogligt niveau til, at man meningsfuldt kan anvende dem til at beskrive elevens skriveudvikling uden supplerende kvalitativ viden. For det andet var det på forhånd klart, at der var begrænsninger i,

hvilke sproglige træk det var muligt at undersøge med de nuværende danske sprogteknologier, hvilket betød, at de lingvistiske variable ikke kun er fagligt motiverede, men også begrundet ud fra, hvad der var praktisk muligt med den foreliggende sprogteknologi.

Denne problematik relaterer sig direkte til en vigtig diskussion i forsknings- og metodelitteraturen inden for digital humaniora, som drejer sig om, hvorvidt digitalt medierede forskningsanalyser eller projekter er drevet af en form for teknocentrisme. Med teknocentrisme menes, at det i højere grad er potentialerne og mulighederne ved nye former for digitale data og softwareværktøjer, der driver forskningen, end det er faglige spørgsmål og agendaer. Van Es et al. (2018) understreger bl.a. behovet for et større fokus på det, de kalder 'digital methodology', som er en kritisk og metodologisk reflekterende tilgang til digitale analyseværktøjer, hvor der bl.a. fokuseres værktøjernes epistemiske påvirkning. Dette modstiller forfatterne med et isoleret fokus på 'digital methods', hvor fokus alene er på de digitale softwareværktøjers analytiske potentialer. I denne afhandling svarer det til et behov for at vende den teknocentriske tilgang og lade udviklingen af de lingvistiske variable drives af kvalitative analyser, sådan så det kvalitative går forud for det kvantitative og ikke er et efterfølgende nødvendigt empirisk supplement. Der er med andre ord brug for, at kvalitative, lingvistiske studier af skriveudvikling identificerer lingvistiske træk i teksterne, som ud fra et skriveudviklingsperspektiv er interessante at skalere og operationalisere kvantitativt. På denne måde er det ikke softwareværktøjerne, der definerer, hvilke lingvistiske træk der undersøges, men i højere grad hypoteser om skriftsproglig udvikling baseret på kvalitative studier af elevtekster, der er udgangspunktet for kvantitative undersøgelser.

5.8.2.3 Fra adskilte længdebaserede mål til grammatisk variation og kompleksitet
Den tredje metodologiske pointe ligger i forlængelse af ovenstående og er knyttet til variablenes meningsfuldhed. Mere specifikt handler det om forholdet mellem computationel analyse og humanistisk fortolkning. Computationelle analyser af sprog er hyppigt blevet kritiseret for at være inkompatible med klassiske hermeneutiske processer. Særligt litteraturen Stanley Fish (2018b, 2018a) har lanceret en kritik af forskning inden for digital humaniora under den meget sigende overskrift "The Interpretive Poverty of Data". Pointen hos Fish er, at selvom korpuslingvistiske analyser kan blotlægge mønstre, som ikke vil fremkomme gennem manuelle, kvalitative analyser, så er disse mønstre frigjorte fra den

semantiske, kommunikative og institutionelle kontekst, som de er frembragt under, og som giver et meningsgrundlag for fortolkningen, og dermed forbliver fortolkningen af de korpuslingvistiske mønstre på en form for spekulativt niveau, fordi "the data, just sitting there in all its empty bulk, can be made to support anything." (Fish, 2018b).

Det er selvfølgelig ikke korrekt, at korpuslingvistiske analyser kan understøtte enhver mulig fortolkning, men Fishs kritik er alligevel interessant i relation til denne afhandling, fordi den understreger, at en stor del af det fortolkningsmæssige grundlag, som en analyse af en elevtekst hviler på, er knyttet til konteksten; dvs. til afsenderens intention, de didaktiske rammevilkår, elevens forhåndsviden og sociale baggrund, de omkringliggende fysiske, psykiske og sociale rammevilkår i skriveprocessen osv. Det er et vilkår for ethvert studie (også et kvalitativt), at disse kontekstuelle faktorer kun kan inddrages i varierende grad og i den udstrækning, man har data, der kan understøtte det. Der er imidlertid to yderligere forhold i dette studie, som taler imod Fishs kritik. Den første er knyttet til teoriens rolle i fortolkningen. Videnskabelige teorier er abstrakte og generaliserede fremstillinger af den empiriske virkelighed, og i den forstand fortolkes korpuslingvistiske analyser ikke frit og retningsløst, men derimod inden for de rammer som en teori stiller op. I denne afhandling skabes disse narrative fortolkninger af de korpuslingvistiske resultater både inden for rammerne af de teorier om sprog, skrivning og skriftsprogstilgenese, som jeg har præsenteret i kapitel 2, samt inden for rammerne af relevante forskningsresultater, som jeg har inddraget til at bekræfte, kontrastere eller uddybe analyseresultaterne (se desuden afsnit 3.4.3 for beskrivelse af afhandlingens relation til computationel hermeneutik). En del af de kontekstuelle faktorer, som giver grundlag for fortolkningen, er altså beskrevet i afhandlingens teoretiske rammeværk (fx i form af teorier om elevtekstens pragmatik eller skrivemodeller), ligesom de i varierende omfang inddrages i den forskningslitteratur, jeg har inddraget undervejs til at kontekstualisere analyseresultaterne.

Den anden indvinding mod Fishs kritik er knyttet til de lingvistiske variable, som Fish anser som meningsløse, når de frigøres fra deres oprindelige tekstuelle og kommunikative kontekst. Som det pointeres af Kilgarriff (2005) i artiklen "Language is never, ever, ever, random", er sprog imidlertid aldrig tilfældigt (forstået som arbitrært), men altid et resultat af bestemte formål hos afsenderen, og det er netop det, der ifølge Gries (2009) udgør selve grundlaget for korpuslingvistikken; "...formal differences reflect, or

correspond to, functional differences.” (Gries, 2009, s. 1228). Dette betyder, at korpuslingvistiske studier netop knytter an til funktionelle (herunder semantiske og diskursive) forhold gennem studiet af distributioner af formelle sproglige træk, og dermed er form og indhold ikke adskilt, som det ellers er i Fishs kritik. Hvor kvalificeret koblingen mellem form og funktion er, afhænger imidlertid af kvaliteten af de lingvistiske variable. I dette studie har jeg anvendt en række lingvistiske mål, som siger noget om længden af forskellige sproglige enheder i teksterne, men karakteristisk for flere af dem er, at de ikke i sig selv giver grammatisk information om de strukturer, de dækker over. Sætningslængde siger således noget om antallet af ord mellem to punktummer, men ikke noget om hvilke grammatiske træk der er anvendt til at konstruere sætningen. På samme måde siger ordlængde noget om antallet af bogstaver i ordet, men ikke direkte noget om ordets semantik. Derfor har det krævet mere uddybende viden gennem de kvalitative analyser at forstå og fortolke de kvantitative mønstre, som analyserne af de lingvistiske variable blotlagde. Denne viden kan således danne udgangspunkt for udvikling og validering af fremtidige kvantitative mål.

Netop dette er også en central pointe hos Biber et al. (2020), som peger på, at ønsket om at forstå og beskrive skriveudvikling er fundamentalt forskelligt fra ønsket om at forudsige tekstkvalitet og skrivekompetence. Til sidstnævnte er det tilstrækkeligt med det, forfatterne kalder omnibusmål, som er gode til at forudsige fx karakterniveauer, men som er svære at forstå og fortolke kvalitativt. For at forstå og beskrive skriveudvikling kræver det imidlertid, at der kommer et større fokus på det, de kalder 'grammatical complexity', som er lingvistiske mål, der meningsfuldt og eksplicit referer til forskellige grammatiske træk i teksten. I forlængelse af den tidligere pointe om behovet for at vende den tekno-centriske tilgang til automatiseret tekstanalyse er det således klart, at en måde at tilbagevise Fish og andres kritik af korpuslingvistikken er at udvikle nye og bedre lingvistiske mål. Ser man på resultaterne fra denne analysedel er det fx tydeligt, at selvom de berettende tekster ligger højest i de kvantitative analyser, så er det, når man ser på den kvalitativt, også dem, der er mest ensformige. Der er større variation i de beskrivende tekster både leksikalt og syntaktisk, hvorfor der er brug for lingvistiske mål, som undersøger variation i stedet for gennemsnitlig længde.

5.8.2.4 Om den didaktiske relevans af at undersøge spredning og fordeling

Den sidste metodologiske pointe er knyttet til det analytiske fokus i lingvistiske studier af skriveudvikling. Langt de fleste skriveudviklingsstudier, som undersøger lingvistiske variables diakrone udvikling, fokuserer på gennemsnitsforskelle i variablene. Dette er selvfølgelig interessant ud fra et skriveudviklingsperspektiv, fordi det siger noget om de typiske kendetegn ved tekster på forskellige niveauer eller aldersgrupper. Men fokus på gennemsnitsforskelle kan også være med til at sløre for udviklinger eller manglende udviklinger i bestemte grupper af tekster, hvilket betyder, at viden, der fra et didaktisk perspektiv er vigtig, overses. I denne analyse så vi fx ved tekstlængde, at selvom der var stigninger i medianen, så var der til gengæld ikke tilsvarende stigninger i teksterne i første kvartil, men til gengæld markante stigninger ved fjerde kvartil. Dette viser altså, at spredningen i teksterne på en årgang bliver større med tiden, hvilket fra et didaktisk perspektiv er mindst lige så interessant, som at de typiske tekster omkring midten udvikler sig. Der er således brug for, at der stadig fokuseres på IQR, som siger noget om spredningen i de typiske tekster, men også tekster i de yderste kvartiler og outliers er interessante i et didaktisk perspektiv, fordi de siger noget om elevskrivere eller grupper af tekster, som er stagneret i skriveudviklingen, hvilket er vigtig didaktisk viden i et differentieringsperspektiv.

Kapitel 6: Den synkrone analysedel

6.1 Indledende rammesætning af analysen

I det forrige kapitel undersøgte jeg, hvordan de sproglige variable forandrede sig relativt til klasstrin, og jeg undersøgte kvalitativt, hvilke lingvistiske konstruktioner der lå bag de sproglige variable i tre tekstudvalg af elevtekster. Denne tilgang tillagde således tidsmæssig forandring (stigning i klasstrin) den centrale forklaringskraft i elevernes skriftsproglige udvikling. I denne analysedel vender jeg blikket mod de lingvistiske variables sammenhæng med afhandlingens ekstra-lingvistiske variable. Jeg undersøger, hvilken betydning disse ekstra-lingvistiske variable har for de forskellige sproglige træk i elevernes tekster, og hvor stor denne betydning i så fald er. Med andre ord undersøger jeg i denne analysedel elevernes skriftsproglige udvikling i et synkront perspektiv, dvs. uden klasstrin som det strukturerende princip.

Analysen udfoldes i to dele. Først undersøger jeg i afsnit 6.2, hvilke ekstra-lingvistiske variable der betyder mest for de lingvistiske variable. Dette gør jeg ved at lave en multipel regressionsanalyse for hver lingvistisk variabel med alle ekstra-lingvistiske²⁷ variable indbygget som uafhængige variable i modellen. Her fokuserer jeg både på hver variables individuelle betydning samt variablernes relative betydning i forhold til hinanden. Til sidstnævnte formål er der derfor bygget standardiserede beta-koefficienter ind i modellen. I anden del af analysen (afsnit 6.3) undersøger jeg mere specifikt, hvilken betydning henholdsvis køn og social baggrund har for elevernes tekster, når der kontrolleres for de øvrige ekstra-lingvistiske variable. Jeg runder kapitlet af med at diskutere og fortolke resultaterne af den synkrone analysedel i relation til writers-within-community-modellen, Bernsteins kodeteori samt relateret skriveforskning.

²⁷ Sproglig baggrund medtages ikke i de indledende, komparative analyser, eftersom der er tale om en meget skæv fordeling i de to grupper. Sproglig baggrund er til gengæld inddraget som kontrolvariabel i analyserne i afsnit 6.3. Billedvalg er inddraget i regressionsmodellerne i både afsnit 6.2 og 6.3 for at holde eventuelle prompteffekter konstante.

6.2 De ekstra-lingvistiske variables betydning for elevernes tekster i et komparativt perspektiv

I regressionsmodellerne i denne analyse betragtes klassetrin som en variabel på niveau med de øvrige ekstra-lingvistiske variable. Eftersom der generelt ikke er linearitet i de lingvistiske variable relativt til klassetrin, er klassetrin bygget ind i regressionsmodellerne som en differentieret effekt med 5. årgang som referencekategori. Det bliver derigennem også muligt at sammenligne betydningen af de enkelte klassetrin med de øvrige ekstra-lingvistiske variable. Som i forrige kapitel gennemføres to analyser for hver variabel; én for beskrivende tekst og én for berettende tekst.

6.2.1 Tekstlængde

Den første sproglige variabel, vi skal se på, er tekstlængde. Den diakrone analyse viste som bekendt, at tekstlængde var en god indikator for skriftsproglig udvikling relativt til klassetrin med 9. årgang som en væsentlig undtagelse²⁸. Vi starter med at se på tekstlængde for de beskrivende tekster.

²⁸ Regressionsanalyserne baserer sig på gennemsnitsberegninger, og tallene kan derfor afvige lidt fra de kvartilsvise fordelinger, som var udgangspunktet for forrige analysedel.

Tabel 21

*Multiple lineær regressionsanalyse for tekstlængde
(beskrivende tekst)*

	Tekstlængde	P-værdi	Beta-koef.
Klassetrin			
(5. årgang er ref.):			
6. årgang	52.72	0.041	.175
7. årgang	94.12	0.000	.299
8. årgang	114.94	0.000	.440
9. årgang	51.14	0.039	.175
Lit.-orient.:			
Skrivelyst	16.42	0.088	.128
Skrivevaner	-10.84	0.549	-.044
Læsevaner	.33	0.901	.009
Billedvalg			
(1 er ref.):			
2	-11.58	0.476	-.047
3	-60.59	0.146	-.102
Askriptive:			
Køn (piger)	34.15	0.042	.146
Social baggrund (lav)	-30.28	0.090	-.115
Konstant	91.70	0.035	.

Kilde: Egen dataindsamling og egne beregninger.

Som det fremgår af ovenstående, viser klassetrin ikke overraskende en stærk sammenhæng med tekstlængde. Regressionsanalysen bekræfter i den forstand tendenserne fra den diakrone analysedel. Det gælder også faldet ved 9. årgang, hvor regressionen her viser, at effekten af at hoppe fra 5. til 9. årgang er den samme som at hoppe fra 5. til 6. årgang. Alle årgangseffekter er statistisk signifikante. Vender vi blikket mod gruppen af literacy-orienterede variable, er det særligt væsentligt at bemærke betydningen af elevernes læse- og skrivevaner. Læsevaner har ingen betydning for længden af de beskrivende tekster, mens skrivevaner faktisk har en lille negativ betydning. Her skal det dog bemærkes, at den høje p-værdi indikerer, at betydningen næsten lige så godt kan være nul end de -10,84, der angives i koefficienten. Skrivelyst har derimod en relativ stor betydning for tekstlængden. Her øges tekstlængden med 16 ord for hver øgning i skrivelystindekset, når

man holder alle andre variable i modellen konstante. Ser man på beta-koefficienten, ligger sammenhængsstyrken af skrivelyst med 0,128 lige lidt under betydningen af en øgning i klassetrin fra 5. til 6. årgang. Selvom resultatet med en p-værdi på 0,09 ikke er helt signifikant, indikerer det alligevel, at skrivelyst potentielt kan være en betydningsfuld faktor for tekstlængde i beskrivende tekster. Elevernes billedvalg har ikke den store betydning for tekstlængde. Dog har de elever, som ikke har beskrevet billedet (3)²⁹, skrevet markant kortere tekster (60 ord kortere), men eftersom der er tale om meget få elever i denne kategori, er resultatet ikke signifikant.

De askriptive variable viser generelt en stærk sammenhæng med tekstlængde i de beskrivende tekster. Køn har her den største betydning. Pigernes beskrivende tekster er i gennemsnit 34 ord længere end drengenes tekster med alle andre variable holdt konstante (herunder læse- og skrivevaner). Der er altså tale om en betragtelig kønseffekt, som endda må formodes at være underestimeret i denne model, da jeg holder andre variable konstante, som også må formodes at have kønsforskelle. Med en beta-koefficient på 0,146 ligger betydningen af køn således tæt på effekten af at gå fra 5. til 6. klassetrin. Social baggrund, der som bekendt her er målt som relativt store kategorier af social klasse, har ligeledes en stærk sammenhæng, selvom resultatet her ikke er helt signifikant. Koefficienten er negativ, fordi den højeste sociale gruppe er anvendt som reference. Elever fra den nederste sociale gruppe skriver således gennemsnitligt set omkring 30 ord kortere tekster end elever fra den højeste gruppe. Beta-koefficienten ligger på -0,115 og er således lige lidt under betydningen af køn.

Vender vi blikket mode de berettende tekster, tegner der sig nogenlunde samme billede.

²⁹ Se hvad koderne for billedvalg dækker over i tabel 12 i afsnit 4.2.3.2.

Tabel 22

*Multipel lineær regressionsanalyse for tekstlængde
(berettende tekst)*

	Tekstlængde	P-værdi	Beta-koef.
Klassetrin			
(5. årgang er ref.):			
6. årgang	-10.47	0.693	-.032
7. årgang	72.18	0.006	.221
8. årgang	141.47	0.000	.519
9. årgang	51.62	0.040	.167
Lit.-orient.:			
Skrivelyst	19.18	0.065	.139
Skrivevaner	36.17	0.054	.137
Læsevaner	-.82	0.759	-.023
Askriptive:			
Køn (piger)	40.77	0.020	.164
Social baggrund (lav)	-25.12	0.170	-.090
Konstant	76.90	0.087	.

Kilde: Egen dataindsamling og egne beregninger.

Ligesom ved de beskrivende tekster ses generelt en stærk sammenhæng mellem klassetrin og tekstlængde. At 6. årgang har en negativ koefficient, skyldes i denne kontekst både, at regressionerne tager udgangspunkt i gennemsnitsværdier, hvorimod analyserne i forrige analysedel tog udgangspunkt i kvartilsværdier, og at de øvrige variable holdes konstante. Resultatet er desuden ikke statistisk signifikant. Blandt de literacy-orienterede variable er det endnu engang skrivelyst, der viser den stærkeste sammenhæng med tekstlængde. Med de øvrige variable holdt konstante er der tale om en øgning på 19 ord per øgning i skrivelystindekset, og ser man på beta-koefficienten, er der tale om, hvad der næsten svarer til effekten af 9. årgang. Hvor skrivevaner ved den beskrivende tekst havde en decideret negativ betydning for tekstlængde, forholder det sig helt modsat i de beskrivende tekster. Her har skrivevaner en positiv betydning på omkring 36 ord med en beta-koefficient på samme niveau som skrivelyst. Det er sandsynligt, at denne forskel hænger sammen med den individuelle variation i tekstlængde mellem teksttyper, som jeg udpegede i den diakrone analyse, som handler om, at eleverne skriver den ene tekst længere

end den anden. De elever, der skriver mest i fritiden, ser altså ud til at skrive lange berettende tekster, men korte beskrivende tekster. Læsevaner har imidlertid samme lave betydning for de berettende tekster som for de beskrivende. Ved de berettende tekster er der endda tale om en lille negativ betydning, men eftersom p-værdien er meget høj, kan der ikke konkluderes noget på baggrund af resultatet.

Vender vi blikket mod de askriptive variable, ser vi samme billede som ved de beskrivende tekster. De askriptive variable viser endnu engang en stærk sammenhæng med tekstlængde. Køns effekten er endda lidt større ved de berettende tekster end ved de beskrivende. Pigernes tekster er gennemsnitligt set 41 ord længere end drengenes, og beta-koefficienten ligger også lidt højere her på 0,164 mod 0,146 i de beskrivende tekster. Social baggrund har en lidt lavere betydning ved de berettende tekster med en koefficient på -25,12 og en beta-koefficient på -0,090. Betydningen af social baggrund er dog ikke signifikant, så resultatet skal læses med et vis forbehold. Ser man på tværs af de to teksttyper, er det imidlertid tydeligt, at de askriptive variable er vigtige faktorer for længden af elevernes tekster.

6.2.2 Ordlængde

Den næste variabel, vi skal se på, er ordlængde. Det er vigtigt at holde for øje, at ordlængde har antal bogstaver som enhed, og derfor vil der være tale om små decimalforskelle i koefficienterne. Vi lægger ud med at se på ordlængde for de beskrivende tekster.

Tabel 23

*Multipl lineær regressionsanalyse for ordlængde
(beskrivende tekst)*

	Ordlængde	P-værdi	Beta-koef.
Klassetrin			
(5. årgang er ref.):			
6. årgang	.051	0.602	.047
7. årgang	.032	0.743	.029
8. årgang	.224	0.012	.239
9. årgang	.214	0.024	.204
Lit.-orient.:			
Skrivelyst	.001	0.977	.002
Skrivevaner	.094	0.174	.106
Læsevaner	.014	0.176	.109
Billedvalg			
(1 er ref.):			
2	-.021	0.730	-.024
3	.214	0.179	.010
Askriptive:			
Køn (piger)	.110	0.084	.132
Social baggrund (lav)	-.000	0.995	-.000
Konstant	3.705	0.000	.

Kilde: Egen dataindsamling og egne beregninger.

Som det fremgår af ovenstående tabel har de ekstra-lingvistiske variable moderate effekter på ordlængde i de beskrivende tekster, og mange af resultaterne er derfor også statistisk insignifikante. Hvis vi starter med at se på klassetrin, er der her tale om en differentieret effekt. En øgning fra 5. årgang til 6. og 7. årgang giver en meget lille stigning i ordlængde på under 0,05 bogstav, mens 8. og 9. årgang derimod har en statistisk signifikant effekt på henholdsvis 0,22 og 0,21 bogstav. Ser vi på de literacy-orienterede variable, viser skrivevaner stærk sammenhæng. Her skriver eleverne fra den øverste gruppe næsten 0,1 bogstav længere ord end dem fra den nederste gruppe. Skrivelyst har ingen betydning for ordlængde i de beskrivende tekster med en regressionskoefficient på 0,001 og en beta-koefficient på langt under en tyvendedel af de øvrige literacy-orienterede variable. Læsevaner er baseret på beta-koefficienten den variabel med størst betydning for ordlængde

og har en øgning på 0,014 ord per øgning i læsevaneindekset, når alle øvrige variable i regressionsmodellen holdes konstante. Ingen af de literacy-orienterede variable er dog signifikante. Elevernes billedvalg har ikke den store betydning for ordlængde. Dog har eleverne med billedvalg 3, dvs. dem som ikke har skrevet om et af de to billeder fra skrivepromptet, skrevet 0,2 bogstav længere ord, hvilket svarer til betydningen af 8./9. årgang. Ved billedvalg 3 er der dog ikke tale om en egentlig prompteffekt, men snarere om en effekt af ikke at følge promptet. De askriptive variable har varierende effekt på ordlængde. Størst betydning har køn, idet piger i gennemsnit skriver 0,11 bogstav længere ord end drengene, mens social baggrund ikke har nogen betydning for ordlængde. Baseret på beta-koefficienterne er det således - foruden effekten af en øgning i klassetrin på 3. og 4. årgange - køn samt læse-og skrivevaner, der har den stærkeste sammenhæng med ordlængde. Effekterne er dog forholdsvis små.

Tabel 24

*Multipl lineær regressionsanalyse for ordlængde
(berettende tekst)*

	Ordlængde	P-værdi	Beta-koef.
Klassetrin			
(5. årgang er ref.):			
6. årgang	-.060	0.309	-.090
7. årgang	.034	0.555	.052
8. årgang	.039	0.456	.071
9. årgang	.133	0.018	.211
Lit.-orient.:			
Skrivelyst	.015	0.521	.053
Skrivevaner	-.042	0.318	-.078
Læsevaner	.014	0.019	.193
Askriptive:			
Køn (piger)	.024	0.531	.048
Social baggrund (lav)	-.076	0.065	-.134
Konstant	3.959	0.000	.

Kilde: Egen dataindsamling og egne beregninger.

Betydningen af de ekstra-lingvistiske variable er generelt mindre ved de berettende tekster end ved de beskrivende. Særligt betydningen af klassestrin er en del mindre. Her er eneste signifikante resultat ikke overraskende ved 9. årgang, som også har den højeste effekt på 0,13 bogstav. Selvom der altså er en lille årgangseffekt, varierer den en del mellem de to teksttyper. Dette korresponderer godt med resultaterne fra Nordenfors' (2011, s. 105-111) studie, som, selvom han påviste en svag øgning af ordlængde over tid (fra 4,0 i 5. årg. til mellem 4,2 og 4,9 i 9. årg.), samtidig påviste stor variation i denne øgning relativt til teksttype. Ved de literacy-orienterede variable har skrivelyst og læsevaner størst effekt på henholdsvis 0,015 og 0,014 bogstav. Skrivelyst har dog en høj p-værdi, hvilket betyder, at det er vanskeligt at konkludere noget om skrivelyst ud fra modellen. Det samme gælder betydningen af skrivevaner. Kønseffekten er ligeledes mindsket fra 0,11 ved de beskrivende tekster til 0,02 i de berettende tekster, hvilket betyder, at resultatet er insignifikant. Social baggrund har til gengæld en signifikant effekt på 0,076 i de berettende tekster modsat ved de beskrivende tekster, hvor der ikke var nogen effekt. 9. årgangseffekten og læsevaner er vurderet på beta-koefficienterne de variable med størst sammenhæng med ordlængde på henholdsvis 0,211 og 0,193.

6.2.3 OVIX

Den tredje variabel er OVIX. Ved analyser, der involverer OVIX, er det som bekendt et opmærksomhedspunkt, at analyserne er lavet på elevernes kombinerede tekstproduktion (>200 ord), hvilket i denne sammenhæng betyder, at regressionsanalysen skaber en delvist kunstig sammenhæng mellem elevernes billedvalg i de beskrivende tekster og OVIX, eftersom modellen antager, at billedvalget gælder hele elevens tekstproduktion. Der skal derfor tages forbehold for de koefficienter, der er forbundet med billedvalg. Disse er taget med for at kontrollere for eventuelle prompteffekter.

Tabel 25

*Multipl lineær regressionsanalyse for OVIX
(kombineret tekst)*

	OVIX	P-værdi	Beta-koef.
Klassetrin			
(5. årgang er ref.):			
6. årgang	-0.902	0.535	-0.055
7. årgang	-1.602	0.274	-0.093
8. årgang	1.384	0.295	.097
9. årgang	1.827	0.192	.115
Lit.-orient.:			
Skrivelyst	.879	0.108	.126
Skrivevaner	.383	0.709	.028
Læsevaner	.233	0.123	.122
Billedvalg			
(1 er ref.):			
2	-1.865	0.044	-.140
3	-4.044	0.087	-.124
Askriptive:			
Køn (piger)	1.827	0.055	.143
Social baggrund (lav)	.322	0.750	.022
Konstant	45.935	0.000	.

Kilde: Egen dataindsamling og egne beregninger.

Hvis vi starter med at se på betydningen af klassetrin, ser vi samme overordnede billede, som vi så ved den diakrone analyse. Klassetrin har ikke noget særlig betydning for OVIX-værdierne, hvilket altså bekræfter det OVIX-plateau, som jeg beskrev i kapitel 5. At der er negativ koefficient ved 6. årgang, selvom der i den diakrone analyse i forrige kapitel var en lille stigning fra 5. til 6. årgang, skyldes, at regressionsanalysen tager udgangspunkt i gennemsnitsværdier (50,84 for 5. årgang og 50,12 for 6. årgang), mens stigningen i den diakrone analyse tog udgangspunkt i medianværdier. De literacy-orienterede variable viser også en varierende grad af sammenhæng med OVIX. Alle tre variable har ganske små koefficienter på under én OVIX-værdi, og ingen af resultaterne er signifikante. Den største effekt ses ved skrivelyst (0,88), som også er den variabel, der har den andenhøjeste

beta-koefficient af alle variable i modellen. Skrivevaner og læsevaner har med koefficienter på henholdsvis 0,38 og 0,23 forholdsvis små effekter. Skrivevaner har en meget høj p-værdi, hvilket bl.a. hænger sammen med, at der er forholdsvis få elever i den øverste gruppe. Ved de askriptive variable er det værd at bemærke, at der er en betragtelig køns-effekt på hele 1,83, når alle variable holdes konstante, og en beta-koefficient på 0,14, hvilket er den største af alle variable. Køn har således en betydning for OVIX svarende til lidt over effekten af 9. årgang. Social baggrund viser ikke nogen særlig sammenhæng med OVIX. Både regressions- og betakoefficienten ligger forholdsvis lavt sammenlignet med køn, og med en p-værdi på hele 0,75 er det vanskeligt at udlede nogen sammenhæng af resultatet. Resultaterne af OVIX ligger således godt i forlængelse af resultaterne fra den anden lexisvariabel, ordlængde, som ligeledes viste lav sammenhæng med social baggrund, men til gengæld stærkest sammenhæng med køn.

6.2.4 Sætningslængde

Vi vender nu blikket mod de syntaktiske variable. Den første af de syntaktiske variable, vi skal se på, er sætningslængde. Som ved de foregående variable starter vi med at se på de beskrivende tekster.

Tabel 26

Multipel lineær regressionsanalyse for sætningslængde (beskrivende tekst)

	Sætnings- længde	P> t	Beta
Klassetrin			
(5. årgang er ref.):			
6. årgang	-2.93	0.005	-.248
7. årgang	1.49	0.152	.121
8. årgang	1.11	0.238	.108
9. årgang	1.83	0.065	.160
Lit.-orient.:			
Skrivelyst	.35	0.368	.069
Skrivevaner	.19	0.796	.019
Læsevaner	-.12	0.248	-.090
Billedvalg			
(1 er ref.):			
2	.15	0.813	.016
3	.60	0.721	.025
Askriptive:			
Køn (piger)	-.89	0.183	-.097
Social baggrund (lav)	-.57	0.430	-.055
Konstant	13.53	0.000	.

Kilde: Egen dataindsamling og egne beregninger.

Klassetrin har kun delvist sammenhæng med sætningslængde. Som vi har set ved flere af de foregående variable, er der ikke statistisk signifikante positive effekter ved de første tre årgange, men først ved 9. årgang. Dette er også tilfældet med sætningslængde. Her er der ved 9. årgang en positiv effekt på 1,83 ord per sætning. P-værdien ligger på 0,065 og er derfor lige på kanten af at være signifikant. 9. årgangseffekten er klart den mest betydningsfulde af alle de ekstra-lingvistiske variable med en beta-koefficient på 0,16. Ser vi på de literacy-orienterede variable, er der kun små insignifikante effekter på 0,35 ord per sætning og derunder. Elevernes billedvalg ser desuden ikke ud til at have nogen betydning for teksternes sætningslængde, ligesom de to askriptive variable heller ikke viser tydelige sammenhænge med sætningslængde. Køn viser stærkest sammenhæng af de to askriptive

variable. Her skriver drengene i gennemsnit 0,89 ord længere sætninger end pigerne, men resultatet er med en p-værdi på 0,18 et stykke fra at være signifikant.

Tabel 27

*Multipl regressionsanalyse for sætningslængde
(berettende tekst)*

	Sætnings- længde	P-værdi	Beta-koef.
Klassetrin			
(5. årgang er ref.):			
6. årgang	.87	0.660	.041
7. årgang	1.85	0.341	.088
8. årgang	-.07	0.970	-.004
9. årgang	1.02	0.582	.051
Lit.-orient.:			
Skrivelyst	.92	0.232	.104
Skrivevaner	-1.48	0.287	-.087
Læsevaner	-.07	0.707	-.032
Askriptive:			
Køn (piger)	-2.70	0.039	-.169
Social baggrund (lav)	.20	0.883	.011
Konstant	14.65	0.000	.

Kilde: Egen dataindsamling og egne beregninger.

Ser man på resultaterne for de berettende tekster, er de lige så uklare som ved de beskrivende tekster. Klassetrin viser ikke systematisk effekt på sætningslængde. Størst er effekten ved 7. årgang med en koefficient på 1,85 ord, men resultatet er med en p-værdi på 0,34 et godt stykke fra at være signifikant. Effekten af de literacy-orienterede variable er også forholdsvis små og usikre. Skrivelyst er mest interessant, idet den har den højeste beta-koefficient. Effekten ligger på en gennemsnitlig stigning på 0,92 ord for hver øgning i skrivelystindekset, når alle andre variable holdes konstante, hvilket er en ganske betragtelig effekt. Resultatet er dog med en p-værdi på 0,23 forbundet med en del usikkerhed. Det eneste statistisk signifikante resultat er kønseffekten. Her skriver piger i gennemsnit 2,7 ord kortere sætninger end drenge. Det forholder sig altså omvendt end ved variabelen tekstlængde, hvor pigerne generelt skrev markant længere (41 ord) berettende tekster end

drengene. Pigerne skriver altså generelt lange berettende tekster, men med lidt kortere sætninger end drengene. Social baggrund har ikke stor betydning her. Den ligger med en beta-koefficient på 0,011 svarende til en lidt over en fjerdedel af beta-koefficienten for 6. årgang. Resultatet er desuden meget langt fra at være statistisk signifikant.

6.2.5 Dependensafstand

Den sidste af de fem variable er dependensafstand. Dependensafstand måles som bekendt i antal ord, hvilket betyder, at det vil være decimalforskelle, der kan aflæses i koefficienterne. Vi starter med at se på de beskrivende tekster.

Tabel 28

Multipel lineær regressionsanalyse for gennemsnitlig dependensafstand (beskrivende tekst)

	Dependens-afstand	P-værdi	Beta-koef.
Klassetrin			
(5. årgang er ref.):			
6. årgang	-.189	0.039	-.178
7. årgang	.214	0.020	.192
8. årgang	.225	0.007	.243
9. årgang	.281	0.002	.272
Lit.-orient.:			
Skrivelyst	.045	0.191	.099
Skrivevaner	.004	0.951	.004
Læsevaner	-.011	0.256	-.086
Billedvalg			
(1 er ref.):			
2	-.057	0.323	-.066
3	.012	0.938	.005
Askriptive:			
Køn (piger)	-.019	0.752	-.023
Social baggrund (lav)	-.084	0.188	-.090
Konstant	2.268	0.000	.

Kilde: Egen dataindsamling og egne beregninger.

Med undtagelse af 6. årgang er der statistiske signifikante klassetrinseffekter ved alle klassetrin. Effekterne er på mellem 0,214 og 0,281 ord, hvilket er forholdsvis moderate effekter. De tre udskolingsårgange er desuden de tre variable med den højeste beta-koefficient. Det er altså klassetrin, der i denne model viser den stærkeste sammenhæng med dependensafstand. Vender vi blikket mod de literacy-orienterede variable, er der ingen signifikante sammenhænge. Skrivelyst viser den stærkeste sammenhæng, men med en beta-koefficient på 0,099 er den blot lige over en tredjedel af beta-koefficienten for 9. årgang. Både billedvalg og de askriptive variable viser også lav sammenhæng med dependensafstand. Køn har højest sammenhæng af de askriptive variable, idet drengenes beskrivende tekster gennemsnitligt set har 0,019 kortere dependensafstand end pigernes, men p-værdien er meget høj, hvilket gør det vanskeligt at udlede noget om kønseffekten. Social baggrund har også en lille effekt. Eleverne i den laveste gruppe skriver beskrivende tekster med 0,08 kortere dependensafstand, men også her er beta-koefficienten lav og p-værdien høj.

Tabel 29

Multipel lineær regressionsanalyse for gennemsnitlig dependensafstand (berettende tekst)

	Dependens- afstand	P-værdi	Beta-koef.
Klassetrin			
(5. årgang er ref.):			
6. årgang	.258	0.014	.221
7. årgang	.312	0.003	.271
8. årgang	.188	0.043	.196
9. årgang	.308	0.002	.282
Lit.-orient.:			
Skrivelyst	.075	0.065	.155
Skrivevaner	-.102	0.166	-.110
Læsevaner	.004	0.739	.0275
Askriptive:			
Køn (piger)	-.111	0.106	-.127
Social baggrund (lav)	-.060	0.407	-.061
Konstant	2.154	0.000	.

Kilde: Egen dataindsamling og egne beregninger.

Ser man på dependensafstanden i de berettende tekster, ses der lignende tendenser. Klassesettrin viser stærkest sammenhæng. Den er endda lidt stærkere her end ved de beskrivende tekster. Regressionskoefficienten ligger her på mellem 0,188 og 0,312, og beta-koefficienterne er ligeledes højere ved de berettende tekster end ved de beskrivende. 6. årgang, som stak ud med negative koefficienter i de beskrivende tekster, har positiv og stærkere sammenhæng ved de berettende tekster. De literacy-orienterede variable viser som ved de beskrivende tekster mere varieret sammenhæng. Skrivelyst ligger også her højest med en beta-koefficient på 0,155, hvilket svarer til lidt over halvdelen af effekten af 9. årgang. Regressionskoefficienten ligger på 0,075, hvilket vil sige, at hver øgning i skrivelystindekset medfører en øgning på 0,075 i dependensafstanden med de øvrige variable holdt konstante. Resultatet er tæt på at være statistisk signifikant. Ligesom ved de beskrivende tekster er det køn, der viser stærkest sammenhæng med dependensafstand. Pigerne berettende tekster har 0,11 kortere dependensafstand end drengenes, hvilket falder godt i tråd med resultaterne af sætningslængde, hvor pigerne også lå lavere end drengene. Beta-koefficienten placerer køns effekten tæt på samme niveau som skrivelyst. Social baggrund viser heller ikke ved de berettende tekster en særlig sammenhæng med dependensafstand i teksterne.

6.2.6 Opsummering af analysen af de ekstra-lingvistiske variables betydning

I ovenstående analyse har jeg undersøgt sammenhængen mellem de lingvistiske og de ekstra-lingvistiske variable. Analysen har fokuseret på den isolerede betydning af de ekstra-lingvistiske variable, idet regressionsmodellerne har holdt de øvrige variable konstante. Det har derfor været muligt gennem beta-koefficienterne at se, hvilke variable der isoleret set havde stærkest sammenhæng med de lingvistiske variable.

De klassetrinsrelaterede variable havde varierende effekt på de lingvistiske variable. Ikke overraskende var sammenhængen stærkest ved tekstlængde, hvilket falder godt i tråd med resultaterne af den diakrone analyse. Derudover havde klassetrin betragtelig effekt på ordlængde i de beskrivende tekster, dependensafstand i både berettende og beskrivende tekster og delvist på sætningslængden i de beskrivende tekster. Alle disse resultater afspejler resultaterne af den diakrone analysedel i forrige kapitel. Selvom ikke

alle resultaterne var signifikante, var der dog ved alle de lingvistiske variable en betragtelig effekt ved 9. årgang, hvilket indikerer en generel stigning (med varierende hældning) i alle de lingvistiske variable fra 5. til 9. årgang.

De literacy-orienterede variable viste ligeledes varierende effekt på de lingvistiske variable. Skrivelyst viste stærk sammenhæng med flere af de lingvistiske variable, hvilket navnlig var tydeligt ved tekstlængde for både beskrivende og berettende tekst samt OVIX. Læsevaner havde også en betragtelig effekt på ordlængde i både de beskrivende og berettende tekster samt OVIX, men til gengæld ikke store effekter på de øvrige lingvistiske variable. Skrivevaner havde moderat effekt på ordlængde i de beskrivende tekster og OVIX, men særligt i OVIX-modellen havde skrivevaner en meget høj p-værdi, hvilket gør det vanskeligt at udlede noget på baggrund af. Samlet set var skrivelyst den af de tre literacy-orienterede variable med stærkest sammenhæng med de lingvistiske variable efterfulgt af læsevaner og med skrivevaner som den med lavest sammenhæng.

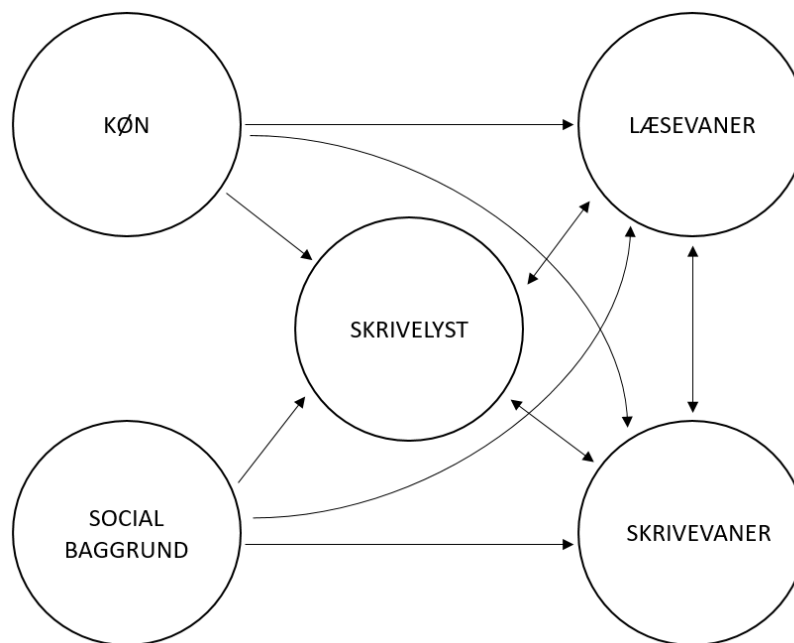
Ved de askriptive variable var køn den variabel med størst effekt. Køn var således, hvis man fraregner nogle af de højeste klassetrinseffekter, en af de variable med højest beta-koefficient ved alle de lingvistiske variable undtagen ordlængde i de berettende tekster og dependensafstand i de beskrivende tekster. I tekstlængde, ordlængde og OVIX var det pigerne, der havde de højeste koefficienter, mens drengene generelt havde højere koefficienter i de syntaktiske variable. Social baggrund viste kun delvist sammenhæng med de lingvistiske variable. Stærkest var effekten ved tekstlængde, hvor elever fra den øverste sociale gruppe skrev mellem 25 og 30 ord længere tekster, men der var også betragtelige effekter på ordlængde i de berettende tekster. Ved de øvrige lingvistiske variable var effekten meget lille.

Der er to yderlige forhold fra ovenstående komparative analyse, som er værd at bemærke. For det første viser analysen, at der fortsat er betragtelige forskelle relativt til teksttype. I den diakrone analysedel så vi, at de lingvistiske variable varierede forholdsvis meget relativt til teksttype, og i ovenstående analyse har vi således set, at denne variation relativt til teksttype også er synlig i effekten af de ekstra-lingvistiske variable. Dette er interessant, selvom en stor del af forklaringen på det selvfølgelig hænger sammen med variationen i de lingvistiske variable. For det andet viser analysen, at de ekstra-lingvistiske variable i særlig grad har betydning for tre af variablene. Den diakrone analyse har allerede afdækket betydningen af klassetrin, om end på baggrund af et mere deskriptivt

design, men i denne analysedel har det også været muligt at undersøge betydningen af klassetrin relativt til de øvrige variable. Her har det været tydeligt, at særligt tekstlængde og delvist ord- og sætningslængde viste tæt sammenhæng med andre variable end klassetrin. Analysen har således også vist, at skriftsproglig udvikling ikke blot udvikler sig diakront relativt til klassetrin med en del variation til følge, men også at elevernes skriveudvikling påvirkes synkront af en række ekstra-lingvistiske variable, herunder særligt køn, skrivelyst og læsevaner.

6.3 Askriptive forskelle i skriftsproglig udvikling

Vi har nu kastet et overordnet og forholdsvis grovkornet blik på sammenhængen mellem de lingvistiske variable i elevernes tekster og de ekstra-lingvistiske variable, som er knyttet til den enkelte elevskriver. Den forudgående analyse har således forudsat, at alle de ekstra-lingvistiske variable var på samme niveau. Dette ændrer sig i den følgende analyse. Her vil jeg i stedet fokusere på betydningen af de to askriptive variable, og i hvilken udstrækning de literacy-orienterede variable forklarer betydningen af de askriptive variable. Jeg går således fra i den forrige analysedel at betragte alle variable som værende på samme niveau til i denne analysedel at indlægge teoretiske forudsætninger i regressionsmodellerne om de fem ekstra-lingvistiske variables indbyrdes relationer. Denne analysedel besvarer således forskningsspørgsmål 3. De følgende analyser tager udgangspunkt i følgende model over variablernes indbyrdes relationer.



Figur 10: Model over de ekstra-lingvistiske variables indbyrdes relationer.

Modellen viser forholdet mellem de to askriptive variable og de tre literacy-orienterede variable. Køn og social baggrund går forud for de tre literacy-orienterede variable, fordi køn og social baggrund ikke er tillærte, men derimod uden for den enkelte elevs kontrol. De tre literacy-orienterede variable er derimod tillærte og må derfor formodes at være påvirket af de askriptive. Desuden viser særligt køn stærk empirisk sammenhæng med de lingvistiske variable i de forudgående modeller og er derfor en interessant variabel at undersøge nærmere. I de forudgående regressionsmodeller kan betydningen af disse variable imidlertid være understimeret, eftersom modellerne har holdt alle de øvrige variable konstante, som også kan tænkes at have askriptive forskelle. Derfor er det væsentligt med en mere integreret regressionsmodel, der mere specifikt undersøger betydningen af de askriptive variable samt, hvorvidt og i hvilken udstrækning de øvrige variable kan forklare en del af betydningen af de askriptive variable. I denne analysedel undersøger jeg med andre ord, hvorvidt og i hvilken grad effekten af køn og social baggrund hænger sammen med, at pigerne samt eleverne fra den højeste sociale gruppe har bedre læse- og skrivevaner samt højere skrivelyst.

I figur 10 ser vi, at køn og social baggrund påvirker alle de tre øvrige variable, og at påvirkningen er ensidig, dvs. at køn og social baggrund ikke påvirkes tilbage af de literacy-orienterede variable. Skrivelyst, læsevaner og skrivevaner er alle gensidigt forbundne i

den forstand, at de potentielt påvirker hinanden, og det er således ikke muligt at sige, at nogle af de literacy-orienterede variable kausalt forårsager hinanden. Om skrivelyst er et resultat af gode læse- og skrivevaner eller omvendt, eller om skrivelyst udvikles diakront i dialektisk vekselvirkning med, at børns læse- og skrivevaner udvikles og manifesteres i løbet af skoleårene, er således svært at sige. Implikationen af figur 10 er således ikke en entydig udpegning af kausale relationer mellem variablerne, men snarere en udpegning af de variable, der kan forklare dele af betydningen af de askriptive variable, og som der derfor må kontrolleres for i regressionsmodellerne. En udpegning af en entydig kausal relation vil dels kræve flere variable, og dels en meget mere elaboreret og systematisk teoretisk model, der kan forklare sammenhængen mellem de enkelte variable. Dette kendes fx fra de såkaldte 'directed acyclic graphs' (DAGs), som er grafiske modeller, der anvendes i socialvidenskaberne til at illustrere de teoretiske antagelser om relationerne mellem variablerne i et datasæt, der er grundlaget for, at der kan drages kausale konklusioner på baggrund af analyserne (Elwert, 2013). At etablere kausalitet er imidlertid ikke en del af denne afhandlings formål. Elevernes skriftsprog udvikler sig i en kompleks økologi, hvor mange forskellige faktorer påvirker og bidrager til udviklingen. Writers-within-community-modellen, som jeg kort præsenterede i kapitel 2, indeholder alene over 60 variable, som alle spiller ind på elevernes skrivefærdigheder og dermed også på de tekster, de producerer. Selvom der selvfølgelig er en vis kausal logik på spil (det er fx den teoretiske årsag til, at man i det hele taget holder andre variable konstante i regressionsmodellerne), er formålet med analyserne specifikt at zoome ind på de askriptive variables betydning og undersøge, hvorvidt og i hvilket omfang disse variables betydning forklares af elevernes skrivelyst samt læse- og skrivevaner.

Regressionsmodellerne i denne analysedel er bygget op omkring tre outcomevariable: tekstlængde, ordlængde og sætningslængde. Der er således medtaget én variabel for både produktivitet, leksis og syntaks. Det betyder også, at jeg ikke ser på OVIX og dependensafstand. Der er to årsager til dette. For det første er det lettere at fortolke regressionskoefficienterne ved de tre valgte variable, end det er ved OVIX og dependensafstand. Dette hænger sammen med, at de tre variable er længdebaserede og derfor kobler direkte til observerbare strukturer i teksten. For det andet er de to variable fravalgt på baggrund af resultaterne fra den diakrone analysedel, hvor ingen af dem viste særlige udviklings-træk eller sammenhænge med de kvalitative observationer i teksten. OVIX er desuden

ikke lavet på det samlede datasæt, men kun en mindre del. Det er derfor ved OVIX heller ikke muligt at sammenligne imellem teksttyper.

6.3.1 Betydningen af social baggrund

Vi starter med at se på social baggrund, der som bekendt i denne afhandling er bestemt ud fra forældrenes nuværende beskæftigelse. Den første regressionsmodel har en af de tre lingvistiske variable som afhængig variabel, social baggrund som uafhængig variabel samt de tre literacy-orienterede variable og sproglig baggrund som kontrolvariable. Ved de beskrivende tekster kontrolleres der ligeledes for elevernes billedvalg. De tre øverste horisontale kategorier i modellerne angår de berettende tekster, mens de sidste tre kategorier angår de beskrivende tekster.

6.3.1.1 Social baggrund og tekstlængde

Vi starter med at se på social baggrunds betydning for tekstlængde. I de fulde regressionsmodeller, hvor alle de øvrige variable blev holdt konstante (inkl. klassetrin), så vi, at social baggrund havde relativt stor betydning for tekstlængde. I den følgende model (tabel 30) vil jeg estimere, hvor stor betydningen reelt set er, og i hvilken grad denne betydning forklares af de øvrige variable.

Tabel 30*Lineær regressionsanalyse for social baggrunds betydning for tekstlængde*

	(1)	(2)	(3)	(4)	(5)	(6)
	Tekstlængde, berettende	Tekstlængde, berettende	Tekstlængde, berettende	Tekstlængde, beskrivende	Tekstlængde, beskrivende	Tekstlængde, beskrivende
Social baggrund (høj)						
Social baggrund (lav)	-48.16* (19.79)	-41.07* (19.83)	-43.90* (20.03)	-49.07* (18.17)	-44.49* (18.38)	-45.00* (18.54)
Dansk som andetsprog (ja)						
Dansk som andetsprog (nej)	10.12 (24.95)	15.08 (24.79)	14.73 (24.82)	-5.96 (22.59)	-3.66 (22.58)	-2.98 (22.64)
Skrivelyst		21.94* (9.73)	20.16+ (11.14)		13.00 (8.87)	16.81+ (9.91)
Skrivevaner			26.88 (20.49)			-15.52 (18.67)
Læsevaner			-1.97 (2.96)			-0.95 (2.80)
Billedvalg 1						
Billedvalg 2				-4.19 (17.16)	-4.81 (17.11)	-3.98 (17.28)
Billedvalg 3				-81.38+ (41.85)	-79.87+ (41.74)	-78.04+ (42.35)
Konstant	198.24* (24.07)	125.96* (39.94)	136.96* (40.85)	204.00* (23.16)	162.04* (36.79)	161.11* (37.89)
Observationer	195	195	195	204	204	204

Standardfejl i parentes

+ $p < .10$, * $p < .05$

Som det fremgår af regressionsmodellen i tabel 30, har social baggrund stor betydning for tekstlængde. Elever fra den laveste gruppe skriver i gennemsnit 48 ord kortere beretende tekster end elever fra den højeste gruppe, når man blot kontrollerer for sproglig baggrund³⁰. Kontrolleres der også for skrivelyst falder betydningen af social baggrund til 41 ord, mens den ligger på ca. 44 ord, når der også kontrolleres for læse- og skrivevaner. Social baggrund er med andre ord en meget konsistent faktor for tekstlængde i de beretende tekster, og kun en meget lille del af betydningen af social baggrund kan forklares ud fra elevernes skrivelyst samt læse- og skrivevaner. Sammenlignet med tabel 22, hvor vi så på social baggrunds relative betydning sammenlignet med de øvrige variable, estimeres betydningen af social baggrund i denne regressionsmodel altså til at være en anelse højere. Alle resultaterne for social baggrund i tabel 30 er statistisk signifikante.

Vender vi blikket mod de beskrivende tekster ses samme tendens. Her skriver eleverne fra den laveste gruppe i gennemsnit 49 ord kortere tekster end elever fra den højeste gruppe. Skrivelyst og læse- og skrivevaner har også her kun et lille bidrag til forklaring af effekten, og når der kontrolleres for alle de tre literacy-orienterede variable, er der stadig en forskel imellem de to grupper for social baggrund på 45 ord i de beskrivende tekster. Det er således tydeligt, at social baggrund er en meget betydningsfuld faktor for længden af de tekster, som eleverne har produceret, og lidt overraskende hænger kun en forholdsvis lille del af denne betydning sammen med, at elever fra den højeste social gruppe også har større skrivelyst og bedre læse- og skrivevaner.

6.3.1.2 Social baggrund og ordlængde

Vi vender nu blikket mod social baggrunds betydning for ordlængde i de to teksttyper. I tabel 23 og 24 så vi, at social baggrund i de fulde regressionsmodeller, hvor alle de øvrige variable blev holdt konstante, havde størst betydning for de berettende tekster, men at betydningen generelt var forholdsvis lille.

³⁰ Det er nødvendigt at kontrollere for sproglig baggrund som en 'confounding variable' for at estimere betydningen af social baggrund korrekt, eftersom sproglig baggrund kan influere på både den uafhængige variabel (social baggrund) og den afhængige variabel (tekstlængde). Dette er kun tilfældet ved social baggrund og ikke ved køn, eftersom køn og sproglig baggrund ikke påvirker hinanden.

Tablel 31
Lineær regressionsanalyse for social baggrunds betydning for ordlængde

	(1) Ordlængde, beretende	(2) Ordlængde, beretende	(3) Ordlængde, beretende	(4) Ordlængde, beskrivende	(5) Ordlængde, beskrivende	(6) Ordlængde, beskrivende
Social baggrund (høj)						
Social baggrund (lav)	-0.10* (0.04)	-0.09* (0.04)	-0.08+ (0.04)	-0.05 (0.07)	-0.04 (0.07)	-0.03 (0.07)
Dansk som andetsprog (ja)						
Dansk som andetsprog (nej)	-0.00 (0.05)	0.00 (0.05)	0.00 (0.05)	0.03 (0.08)	0.04 (0.08)	0.04 (0.08)
Skrivelyst		0.03 (0.02)	0.02 (0.02)		0.04 (0.03)	0.01 (0.04)
Skrivevaner			-0.05 (0.04)			0.07 (0.07)
Læsevaner			0.01* (0.01)			0.01 (0.01)
Billedvalg 1						
Billedvalg 2				0.00 (0.06)	0.00 (0.06)	0.00 (0.06)
Billedvalg 3				0.19 (0.15)	0.19 (0.15)	0.16 (0.15)
Konstant	4.14* (0.05)	4.03* (0.08)	4.00* (0.08)	3.99* (0.08)	3.85* (0.13)	3.83* (0.14)
Observationer	195	195	195	204	204	204

Standardfejl i parentes

+ $p < .10$, * $p < .05$

Tabel 31 viser, at social baggrund også skaber forskelle i ordlængden i de berettende tekster. Eleverne fra den laveste gruppe skriver 0,1 bogstav kortere ord end eleverne fra den øverste gruppe. Sproglig baggrund har i denne forbindelse ikke nogen forklarende effekt. Som ved tekstlængde har de literacy-orienterede variable også forholdsvis begrænsede forklarende effekter. Kontrolleres der alene for skrivelyst, falder regressionskoefficienten til 0,09, og den falder til 0,08, hvis der også kontrolleres for læse- og skrivevaner. Ved sidstnævnte er resultatet imidlertid ikke signifikant. Resultatet viser således, at selvom effekten af social baggrund er lille for ordlængde i de berettende tekster, er den imidlertid meget stabil. At der er en del resultater, der ikke er signifikante, hænger både her og generelt i afhandlingens analyser selvfølgelig sammen med, at samplet er forholdsvis lille. Med et større sample havde flere resultater sandsynligvis været signifikante.

Ved de beskrivende tekster er betydningen af social baggrund endnu mindre. Regressionskoefficienten er på 0,05, hvis der alene kontrolleres for sproglig baggrund. Det vil sige, at den halvt så stor som ved de berettende tekster. Som ved de berettende tekster falder koefficienten med 0,01, når der kontrolleres for skrivelyst, og med yderligere 0,01, når de kontrolleres for alle tre literacy-orienterede variable. Ingen af koefficienterne er imidlertid signifikante. Sammenholdt med tekstlængde har social baggrund således en mindre betydning, men til gengæld er betydningen forholdsvis stabil, eftersom de literacy-orienterede variable kun har en forholdsvis begrænset forklarende effekt. At elever fra den højeste sociale gruppe skriver længere ord end dem fra den laveste, skyldes altså kun delvist, at eleverne fra den højeste gruppe har større skrivelyst og bedre læse- og skrivevaner.

6.3.1.3 Social baggrund og sætningslængde

Den sidste variabel, vi skal se på, er sætningslængde. Ved den forrige komparative analyse viste social baggrund forholdsvis lille og varierende sammenhæng med sætningslængde.

Tabel 32
Lineær regressionsanalyse for social baggrunds betydning for sætningslængde

	(1) Sætningsl., beretende	(2) Sætningsl., beretende	(3) Sætningsl., beretende	(4) Sætningsl., beskrivende	(5) Sætningsl., beskrivende	(6) Sætningsl., beskrivende
Social baggrund (lav)	0.56 (1.29)	0.61 (1.31)	0.66 (1.32)	-0.38 (0.73)	-0.37 (0.74)	-0.43 (0.74)
Dansk som andetsprog (ja)						
Dansk som andetsprog (nej)	-0.07 (1.63)	-0.03 (1.64)	0.05 (1.64)	0.29 (0.90)	0.30 (0.91)	0.32 (0.91)
Skrivelyst		0.15 (0.64)	0.55 (0.74)		0.01 (0.36)	0.20 (0.40)
Skrivevaner			-1.75 (1.36)			-0.41 (0.75)
Læsevaner			-0.04 (0.20)			-0.08 (0.11)
Billedvalg 1						
Billedvalg 2				0.49 (0.69)	0.49 (0.69)	0.49 (0.69)
Billedvalg 3				-2.04 (1.67)	-2.04 (1.68)	-1.86 (1.70)
Konstant	15.51* (1.57)	15.00* (2.64)	14.56* (2.70)	13.38* (0.92)	13.35* (1.48)	13.46* (1.52)
Observationer	195	195	195	204	204	204

Standardfejl i parentes
⁺ $p < .10$, * $p < .05$

Tabel 32 bekræfter det billede, som viste sig i de komparative analyser af de ekstra-lingvistiske variables betydning. (se tabel 26 og 27) Social baggrund har ikke nogen særlig stor betydning for sætningslængden i de berettende tekster. Når der kontrolleres for sproglig baggrund, viser modellen, at eleverne i den lave sociale gruppe skriver 0,56 ord længere sætninger end eleverne i den højere gruppe. Der er altså tale om en omvendt effekt end den, vi har set ved de to forrige variable. Som vi så i den diakrone analyse, kan dette tænkes at hænge sammen med, at sætningslængde ikke hænger kontinuerligt sammen med syntaktisk kompleksitet. Meget lange sætninger er således ikke nødvendigvis et udtryk for mere komplekse sætningskonstruktioner, men kan være et tegn på overforbrug af bl.a. paratakse eller ellipsekonstruktioner. Når der kontrolleres for skrivelyst, øges betydningen af lav social baggrund. Dette understøtter således, at høj sætningslængde i denne model ikke skal fortolkes som et positivt udviklingstegn. Det samme ses, når der kontrolleres for alle de tre literacy-orienterede variable. Her øges regressionskoefficienten til 0,66.

Ved de beskrivende tekster er betydningen af lav social baggrund negativ, hvilket vil sige, at elever fra den højeste gruppe her skriver 0,38 ord længere sætninger end eleverne fra den laveste gruppe. Dette passer således godt med resultaterne af den diakrone analyse, som viste, at problemerne med meget høje sætningslængder med lav grad af kompleksitet særligt var gældende for de berettende tekster. Ved de beskrivende tekster ser vi således, at regressionskoefficienten falder med 0,01, når der kontrolleres for skrivelyst, mens den til gengæld stiger med 0,06 til 0,43, når der kontrolleres for alle tre literacy-orienterede variable. Der er altså ikke nogen særlig forklarende effekt af literacy-orienterede variable.

6.3.1.4 Opsummering af betydningen af social baggrund

Betydningen af social baggrund for de tre lingvistiske variable kan kort opsummeres således. Social baggrund har stor betydning for længden af de tekster, eleverne producerer. Både ved de berettende og de beskrivende tekster ses en stabil effekt af social baggrund, som er relativt upåvirket på tværs af de tre modeller, hvor de literacy-orienterede variable gradvist indbygges. Forskellen fra den første model, som er bivariat, til den tredje model, som holder alle relevante kontrolvariable konstante, er således 8-9%. Ved ordlængde så

vi en meget lille, men dog statistisk signifikant betydning ved de berettende tekster. Også her var betydningen upåvirket på tværs af modeller, hvilket betyder, at de literacy-orienterede variable ikke havde nogen stor forklarende effekt for social baggrund. Ved analysen af sætningslængde så vi ingen statistisk signifikante effekter af social baggrund. Dog så vi en omvendt effekt ved de berettende tekster, hvor dem fra den nederste gruppe skrev længere sætninger end dem fra den øverste gruppe. Resultatet var dog ikke signifikant. Social baggrund påvirker således primært elevernes produktivitet og i mindre grad de leksikale og syntaktiske niveauer i teksterne.

6.3.2 Betydningen af køn

Vi gentager nu analysen, men denne gang med den anden askriptive variabel 'køn' som den uafhængige variabel. Eftersom køn ikke influerer på sproglig baggrund og omvendt, er sproglig baggrund ikke inkluderet i regressionsmodellerne. Som ved forrige analyse er de tre literacy-orienterede variable integreret som kontrolvariable, sådan så det er muligt at undersøge, hvorvidt og i hvilket omfang de tre literacy-orienterede variable forklarer betydningen af køn for de tre lingvistiske variable.

6.3.2.1 Køn og tekstlængde

Vi starter med at se på tekstlængde som den afhængige variabel. De komparative analyser viste, at køn var den af de ekstra-lingvistiske variable, der havde den højeste beta-koefficient og dermed udviste den stærkeste sammenhæng med tekstlængde i både de beskri vende og berettende tekster. Der må således forventes en betragtelig kønseffekt i denne regressionsmodel også, men spørgsmålet er, i hvilket omfang denne kønseffekt kan forklares af de tre literacy-orienterede variable.

Tablet 33
Lineær regressionsanalyse for køns betydning for tekstlængde

	(1) Tekstlængde, berettende	(2) Tekstlængde, berettende	(3) Tekstlængde, berettende	(4) Tekstlængde, beskrivende	(5) Tekstlængde, beskrivende	(6) Tekstlængde, beskrivende
Drenge (ref.)						
Piger	33.17 ⁺ (17.01)	18.99 (17.57)	16.27 (17.81)	35.48 [*] (15.34)	31.33 [*] (15.88)	34.64 [*] (16.21)
Skrivelyst		26.21 [*] (9.70)	22.55 [*] (11.08)		11.07 (8.68)	12.35 (9.73)
Skrivevaner			22.01 (19.99)			-19.31 (18.09)
Læsevaner			-0.19 (2.75)			1.02 (2.60)
Billedvalg 1						
Billedvalg 2				0.25 (16.13)	1.71 (16.15)	3.64 (16.30)
Billedvalg 3				-51.20 (39.32)	-49.47 (39.24)	-51.05 (39.87)
Konstant	173.54 [*] (12.95)	102.55 [*] (29.22)	109.25 [*] (30.34)	162.27 [*] (13.11)	131.36 [*] (27.11)	124.61 [*] (28.24)
Observationer	214	214	214	226	225	225

Standardfejl i parentes

⁺ $p < .10$, ^{*} $p < .05$

Køn har en forholdsvis stor betydning for tekstlængde i de berettende tekster. Ser man på den første simple, bivariate analyse af køns-effekten uden kontrolvariable ligger koefficienten på 33,17 i pigernes favør. Det vil sige, at pigerne i gennemsnit skriver 33,17 ord længere berettende tekster end drengene. Resultatet er imidlertid et stykke fra at være statistisk signifikant. Kontrollerer man for skrivelyst falder koefficienten med hele 43% til 18,99. Det vil sige, at en meget stor del af køns-effekten i de berettende tekster kan forklares med, at pigerne har større skrivelyst end drengene. Kontrolleres der yderligere for læse- og skrivevaner, falder den til 16,27. Det betyder altså, at de tre literacy-orienterede variable til sammen forklarer over halvdelen af betydningen af køn. Her skal dog tages forbehold for, at hverken model 2 eller 3, som indeholder kontrolvariablene, er statistisk signifikante.

Vender vi blikket mod de beskrivende tekster, tegner der sig et lidt andet billede. Den bivariate køns-effekt ligger på nogenlunde samme niveau som ved de berettende tekster. Koefficienten er her på 35,48 i pigernes favør, hvilket betyder, at pigerne i gennemsnit skriver 35,48 ord længere beskrivende tekster. Den rene køns-effekt er her på nogenlunde samme niveau som ved de berettende tekster. Forskellen er blot, at resultatet ved de beskrivende tekster er statistisk signifikant. Kontrolleres der for skrivelyst ser vi et markant mindre fald i koefficienten, end vi så ved de berettende tekster. Her falder koefficienten blot med 12% til 31,33 ord. Kontrolleres der yderligere for læse- og skrivevaner og billedvalg, stiger koefficienten en smule til 34,64. Den samlede forklarende effekt af de tre literacy-orienterede variable er således forholdsvis lille ved de beskrivende tekster. Forskellen mellem model 4 og model 6 er således blot på 0,84 ord svarende til et samlet fald på blot 2,4%.

6.3.2.2 Køn og ordlængde

Den næste variabel, vi skal se på, er ordlængde. I de komparative regressionsmodeller viste køn en forholdsvis stærk sammenhæng med ordlængde i de beskrivende tekster, men ikke med ordlængde i de berettende tekster.

Tabel 34
Lineær regressionsanalyse for køns betydning for ordlængde

	(1) Ordlængde, berettende	(2) Ordlængde, berettende	(3) Ordlængde, berettende	(4) Ordlængde, beskrivende	(5) Ordlængde, beskrivende	(6) Ordlængde, beskrivende
Drenge (ref.)						
Piger	0.04 (0.04)	0.02 (0.04)	0.03 (0.04)	0.12* (0.05)	0.10+ (0.06)	0.10+ (0.06)
Skrivelyst		0.04* (0.02)	0.03 (0.02)		0.03 (0.03)	0.01 (0.03)
Skrivevaner			-0.07+ (0.04)			0.04 (0.06)
Læsevaner			0.02* (0.01)			0.01 (0.01)
Billedvalg 1						
Billedvalg 2				0.01 (0.06)	0.01 (0.06)	0.01 (0.06)
Billedvalg 3				0.18 (0.14)	0.19 (0.14)	0.16 (0.14)
Konstant	4.08* (0.03)	3.97* (0.06)	3.92* (0.06)	3.92* (0.05)	3.85* (0.10)	3.84* (0.10)
Observationer	214	214	214	226	225	225

Standardfejl i parentes

+ $p < .10$, * $p < .05$

Tabel 34 bekræfter resultatet fra den komparative analyse, idet den viser, at køn ikke har nogen effekt på de berettende tekster. Model 1 med den bivariante analyse viser en effekt på kun 0,04 bogstav i pigernes favør, men med en standardfejl på samme niveau er det ikke muligt at udlede noget af resultatet. Skrivelyst har en signifikant koefficient på 0,04, og den forklarer også kønseffekten en smule, men resultatet for køn er stadig ikke signifikant. Kontrolleres der for alle de tre literacy-orienterede variable ses samme billede. Læsevaner har en signifikant betydning på 0,02, men kønseffekten er stadig kun på 0,03 og derfor meget lille, og resultatet er heller ikke signifikant.

Lidt større er betydningen af køn, hvis vi ser på de beskrivende tekster. Model 4, som indeholder den simple bivariante analyse, viser en signifikant effekt af køn på 0,12 bogstav. Kontrolleres der for skrivelyst falder betydningen af køn til 0,10, og resultatet er heller ikke længere helt signifikant, men p-værdien er dog stadig under 0,1. Kontrolleres der også for de to øvrige literacy-orienterede variable ses ingen forklarende effekt på betydningen af køn. Hvor køn altså ikke havde nogen effekt på de berettende tekster, ser vi altså her, at køn har en lille effekt på ordlængden i de beskrivende tekster, og vi ser samtidig, at de literacy-orienterede variable ikke har nogen væsentlig forklarende effekt af denne betydning. Kønseffekten på ordlængde i de beskrivende tekster er således lille, men stabil. At piger gennemsnitligt set skriver lidt længere ord end drengene i de beskrivende tekster, skyldes altså i nogen grad, at de har større skrivelyst end drengene, men ikke at de har bedre læse- og skrivevaner.

6.3.2.3 Køn og sætningslængde

Den sidste variabel, vi skal se på, er sætningslængde. Den komparative analyse viste, at køn havde en varierende effekt på sætningslængde. Som de eneste af de lingvistiske variable viste de to syntaktiske variable en omvendt kønseffekt, hvor det var drengene der lå højest. Mest markant var det ved de berettende tekster, hvor køn var den variabel af alle de ekstra-lingvistiske variable med den højeste beta-koefficient.

Tabel 35*Lineær regressionsanalyse for køns betydning for sætningslængde*

	(1) Sætningsl., berettende	(2) Sætningsl., berettende	(3) Sætningsl., berettende	(4) Sætningsl., beskrivende	(5) Sætningsl., beskrivende	(6) Sætningsl., beskrivende
Drenge (ref.)						
Piger	-2.46* (1.05)	-2.71* (1.10)	-2.61* (1.11)	-1.39* (0.60)	-1.50* (0.62)	-1.54* (0.64)
Skrivelyst		0.46 (0.61)	0.86 (0.69)		0.32 (0.34)	0.46 (0.38)
Skrivevaner			-1.08 (1.25)			0.13 (0.71)
Læsevaner			-0.11 (0.17)			-0.10 (0.10)
Billedvalg 1						
Billedvalg 2				0.39 (0.63)	0.45 (0.64)	0.40 (0.64)
Billedvalg 3				-2.48 (1.54)	-2.43 (1.54)	-2.18 (1.57)
Konstant	16.93* (0.80)	15.69* (1.83)	15.54* (1.90)	14.43* (0.51)	13.54* (1.07)	13.78* (1.11)
Observationer	214	214	214	226	225	225

Standardfejl i parentes

+ $p < .10$, * $p < .05$

Som det fremgår af tabel 35, viser køn en forholdsvis stærk sammenhæng med sætningslængde, og som vi så ved de komparative analyser, er det denne gang drengene, der ligger højest. Den bivariate analyse i model 1 viser således, at drengene i gennemsnit skriver 2,46 ord længere sætninger i de berettende tekster end pigerne. Holdes skrivelyst konstant, øges betydningen til 2,71, mens den falder til 2,61, hvis man kontrollerer for alle de tre literacy-orienterede variable. Kønsforskellene er altså meget stabile, og de kan ikke i særlig høj grad forklares ud fra hverken skrivelyst eller læse- og skrivevaner. Alle tre regressionskoefficienter for køn er statistisk signifikante.

Ved de beskrivende tekster ses samme tendens. Dog er kønseffekten en del mindre. Her viser model 4, som indeholder den simple, bivariate analyse, at drengene i gennemsnit skriver 1,39 ord længere sætninger i de beskrivende tekster end pigerne. Der er således tale om tæt på en halvering af den effekt, vi så ved de berettende tekster. Resultatet er imidlertid også her statistisk signifikant. Kontrolleres der for skrivelyst, øges betydningen af køn til 1,5 ord per sætning, og i model 6, hvor der kontrolleres for alle literacy-variable samt billedvalg, øges betydningen til 1,54. At koefficienten øges i takt med, at der kontrolleres for de literacy-orienterede variable, hænger sammen med, at det i højere grad er pigerne, der har høj skrivelyst og gode læse- og skrivevaner. Den omvendte tendens, hvor drengene ligger højest i sætningslængde, men lavest i ord- og tekstlængde, ligger godt i forlængelse af resultatet af betydningen af social baggrund, hvor det var eleverne fra den laveste sociale gruppe, der skrev længst sætninger, mens eleverne fra den øverste gruppe havde længst tekster og højest ordlængde.

6.3.2.4 Opsummering på betydningen af køn

Betydningen af køn var generelt en anelse større end den, vi så ved social baggrund. Ved tekstlængde var betydningen størst ved de beskrivende tekster. Her så vi en forholdsvis konstant effekt på tværs af de tre modeller, hvilket betyder, at kønseffekten i meget lille grad forklares af de literacy-orienterede variable. Forskellen mellem model 4, som indeholder den bivariate analyse, og model 6, som indeholder alle kontrolvariable, er således 0,84 ord. De literacy-orienterede variable havde meget større forklarende effekt ved de berettende tekster, men resultatet var imidlertid ikke statistisk signifikant. Ved ordlængde havde køn ingen betydning i de berettende tekster, mens der var en lille, men

stabil effekt på tværs af modeller i de beskrivende tekster. Som vi også så ved social baggrund var der omvendt effekt af køn ved sætningslængde, idet det her var drengene, der gennemsnitligt set havde de højeste værdier. Her var det i særlig grad i de berettende tekster, at drengenes sætninger var længere. De literacy-orienterede variable øgede effekten af køn henover modellerne, hvilket indikerer, at lav skrivelyst og dårlige læse- og skrivevaner faktisk kan forklare en del af de lange sætningslængder hos drengene. Alle køns-effekterne ved sætningslængde var statistisk signifikante.

6.4 Delkonklusion på den synkrone analysedel

I den synkrone analysedel har vi som bekendt set, hvordan sproglige træk i elevernes tekster påvirkes af en række baggrundsvariable knyttet til den enkelte elevskriver. I det følgende vil jeg konkludere på baggrund af analyseresultaterne samt fortolke disse i relation til nogle af de teorier, som jeg præsenterede i kapitel 2, samt i relation til øvrig relevant forskning, som kan understøtte, uddybe, kontekstualisere eller kontrastere resultaterne fra denne undersøgelse. Delkonklusionen er bygget op om fire pointer om skriveudvikling. De fire skriveudviklingspointer er knyttet til både det analytiske dobbeltperspektiv i afhandlingen (koblingen af det synkrone og diakrone perspektiv), til et kønsperspektiv, til et ulighedsperspektiv og til et bredere literacyperspektiv.

6.4.1 Skriveudviklingens dobbeltperspektiv

Den første pointe er forholdsvis simpel og kort, men alligevel væsentlig at fremhæve. Analyserne i dette kapitel har fremhævet, at skriveudvikling i elevtekster ikke entydigt kan beskrives hverken synkront eller diakront, men må belyses fra begge perspektiver. I den diakrone analyse så vi fx, at selvom der var få lineære udviklingstræk i de lingvistiske variable relativt til tid, så var der alligevel generelle stigninger i de fleste variable, hvis man sammenlignede medianværdien for 5. og 9. årgang. Analysen viste samtidig også en stigende spredning i flere variable, som i et didaktisk perspektiv er interessant, fordi det siger noget om den variation af elevtekster, som læreren skal rumme i undervisningen. Den diakrone tilgang til at studere skriveudvikling er desuden interessant, fordi skriveudvikling udfolder sig i tid, og især i skolen er klassetrin og alder afgørende for, hvilke typer af opgaver og forventninger der bliver stillet til eleverne. Som det har fremgået af denne analysedel må skriveudvikling imidlertid også forstås synkront. En hel del varians

i de lingvistiske variable kunne forklares af de ekstra-lingvistiske variable, og særligt de to askriptive variable samt elevernes skrivelyst viste sammenhæng med flere af de lingvistiske variable. Samtidig viste analysen, at en del af den variation relativt til teksttype, som blev afdækket i den diakrone analyse, også resulterer i, at de ekstra-lingvistiske variable har forskellig betydning relativt til teksttype. Der er derfor vigtigt at fastholde, at skriveudvikling er et multidimensionelt fænomen, som må beskrives fra både diakrone og synkrone perspektiver, eftersom elevernes skriveudvikling påvirkes i begge retninger.

6.4.2 Skriveudviklingens ulighedsdimension

Social baggrund viste sig som bekendt i dette kapitels analyser at være en forholdsvis vigtig faktor for en del af de lingvistiske variable. I de komparative regressionsmodeller var sammenhængen stærkest med tekstlængde i begge teksttyper og ordlængde i de beretende tekster, hvilket blev bekræftet i de specialiserede regressionsmodeller, hvor det imidlertid viste sig, at de tre literacy-orienterede variable kun havde en lille forklarende effekt på social baggrund. Social baggrund viste sig således at være en meget stabil faktor på tværs af alle regressionsmodeller. Dette betyder, at social baggrund i dette kapitels analyser ikke hænger særligt stærkt sammen med øget skrivelyst hos eleven eller bedre læse- og skrivevaner. Dette er på sin vis et overraskende resultat, bl.a. fordi variablerne læsevaner og skrivevaner netop dækker over elevernes vaner i fritiden, hvor effekten af social baggrund må formodes at være størst.

Som jeg tidligere har beskrevet (se kapitel 2), er der ifølge Bernstein (1971) tætte sammenhænge mellem social baggrund og de sproglige ressourcer, som en elev har tilgængelig. Han skelner derfor mellem en elaboreret kode, som er et udfoldet og fuldt funktionelt sprog, der ikke kræver delt viden mellem afsender og modtager, og en begrænset kode, som er et begrænset og indforstået sprog, der kun kan anvendes i mindre sociale grupper. Analyserne i dette kapitel understøtter på sin vis Bernsteins distinktion, idet den viser, at der er forskelle i de sproglige ressourcer, som elever med henholdsvis høj og lav social baggrund kommer med, og at denne forskel er synlig i elevernes tekster. Mest synligt er det ift. elevernes produktivitet, hvor eleverne fra den øverste sociale gruppe skriver næsten 50 ord længere tekster end eleverne fra den nederste gruppe svarende til næsten en tredjedel af den gennemsnitlige medianværdi for alle årgange i de to teksttyper. Læn-

gere tekster giver, som vi så i den diakrone analysedel, mulighed for udbygget meningsindhold i teksterne, hvor eleverne kan uddybe de semantiske felter i teksten. Et kendetegn ved den elaborerede kode er netop, at den indebærer modtagerbevidsthed og indeholder flere eksplicite metakommunikative markeringer, hvor den forhåndsviden, som læseren skal bruge for at forstå indholdet i teksten, præsenteres og udfoldes. Øget tekstlængde giver selvfølgelig større mulighed for at gøre brug af netop disse træk. At social baggrund også viser delvis sammenhæng med ordlængde, er heller ikke overraskende. Ordlængde indikerer ifølge Hultman & Westmann (1977) en større grad af leksikal specifikation (fx gennem anvendelsen af flere proprier og sammensatte ord), og Myhill (2009) viser, at ordlængde i elevtekster også hænger sammen med en større andel af abstrakte ord. Ifølge Bernstein er abstraktion og specifikation netop noget af det, som de elever, der kun mestrer den begrænsede kode, mangler i deres sprog. Narrative former i den begrænsede kode, udgøres fx typisk af simple sekvenseringer af erfaringer eller hændelsesforløb, mens narrative former i den elaborerede kode også muliggør et abstrakt reflekterende lag oven på erfaringerne (Bernstein, 1971, s. 112-113).

Interessant ved analyserne af ordlængde og sætningslængde var også, at betydningen af social baggrund var lavest ved de beskrivende tekster. Dette kan tænkes at hænge sammen med, at eleverne i kraft af den stærke binding til billederne i skrivepromptet er mere begrænsede i de ord, de kan vælge, hvilket er en fordel for eleverne fra den nederste gruppe. En mulig fortolkning er, at konkrete og mere stilladserede skriveopgaver støtter elever fra den nederste gruppe i deres skriveproces ved på forhånd at begrænse og udvælge nogle af de lingvistiske ressourcer, som eleverne skal bruge til at løse opgaven i skrivepromptet, hvilket i sidste ende er med til at mindske forskellene mellem de to sociale grupper. Eleverne fra den øverste gruppe kan således tænkes i lidt mindre grad at have mulighed for at anvende lingvistiske ressourcer fra den elaborerede kode i de beskrivende tekster, eftersom skrivepromptet er med til at begrænse disse på forhånd.

Flere studier har fundet lignende resultater for betydningen af social baggrund, som dem vi har set i dette studie. Richardson et al. (1976) finder fx forskelle i tekstlængde relativt til social baggrund, men ikke i sætningslængde (t-unit), mens Larsson (1984) finder, at både socioøkonomisk baggrund samt moderens uddannelse har betydning for elevernes skrivepræstationer. Weber (2022) finder i en analyse af over 8000 britiske elevtekster skrevet af 11-årige, at der er forskelle i antallet af stavfejl og leksikal diversitet

relativt til social baggrund, og Weber finder ligeledes, at de lingvistiske variable medierer en del af relationen mellem social baggrund og uddannelsesniveau. Et lidt andet perspektiv ses hos Korat et al. (2007), som undersøger betydningen af social baggrund for relationen mellem begyndende literacyfærdigheder hos børn i alderen 5-6 år (fx fonetisk bevidsthed, ordgenkendelse og bogstavgenkendelse) og to uafhængige variable, henholdsvis literacymiljøet i hjemmet og moderens evne til at mediere indholdet i børnebøger under højtlesning (dvs. moderens evne til at understøtte afkodning, læseforståelse og refleksion i dialog med barnet under højtlesningen). Forfatterne finder bl.a., at der er tæt sammenhæng mellem mødrenes evne til at mediere indhold under højtlesning og børns literacyfærdigheder, når man kun ser på børn fra den øverste sociale gruppe. I den nederste gruppe var der ikke samme relation. Literacymiljøet i hjemmet havde betydning for begge grupper, men størst betydning ved børnene i den laveste sociale gruppe. Dette er interessant i relation til denne afhandling, fordi det i modsætning til resultaterne fra dette kapitels analyser viser en stærk sammenhæng mellem literacyaktiviteter i fritiden og børnenes sproglige færdigheder.

Det er imidlertid svært at give en forklaring på, hvorfor skrivelyst og læse- og skrivevaner ikke forklarer betydningen af social baggrund bedre, end den gør i dette studie. En mulig forklaring findes hos Bernstein (1975), som peger på, at skolen er en vigtig ud-lignende institution, som kan være med til at kompensere og modificere skellet mellem de sociale kontekster, som eleverne er opvokset i. Dette skyldes, at skolens "rituals, its ceremonies, its authority relations, its stratification, its procedures for learning, ... can modify or change the pupil's role as this has been initially shaped by the family." (Bernstein, 1975, s. 48-49). Om det er denne kompenserende effekt af skolen og undervisningen, der er med til at mindske den forklarende effekt af elevernes fritidslæsning og -skrivning, er selvfølgelig svært at sige, eftersom dette vil kræve data, der tilsvarende dokumenterer elevernes læse- og skriveaktiviteter i skolen. Et andet bud er i forlængelse af Korat et al.'s studie, som jeg beskrev ovenfor, at grundstenen til ulighedsdimension ligger tidligere i elevernes aldersudvikling end den, jeg har undersøgt i dette studie. Med andre ord kan det være, at det i højere grad er forskelle i læse- og skrivevaner i den tidlige barndom, der stadig manifesterer sig i den sidste del af grundskolen. Endelig kan et tredje bud være, at der er målefejl og andre utilstrækkeligheder i de literacy-orienterede variable. Dette beskrev jeg også kort i afsnit 4.3.

6.4.3 Skriveudviklingens kønsdimension

Køn viste sig generelt som en af de ekstra-lingvistiske variable med den stærkeste sammenhæng med de lingvistiske variable. I de fulde, komparative regressionsmodeller var køn, hvis man fraregner de højeste klassetrinseffekter, den variabel med de højeste beta-koefficienter på tværs af de fleste modeller (dog ikke ordlængde i de berettende tekster og dependensafstand i de beskrivende tekster). Der tegnede sig ligeledes en klar tendens imellem kønnene, idet pigerne havde de højeste koefficientværdier i tekstlængde, ordlængde og OVIX, mens drengene lå højest i de to syntaksvariable - sætningslængde og dependensafstand. At det fordeler sig sådan imellem kønnene, indikerer således, at de syntaktiske variable i dette studie ikke nødvendigvis dækker over komplicerede syntaktiske konstruktioner, men i lige så høj grad over syntaktiske konstruktioner, som overskrider det kommunikativt hensigtsmæssige. Dette var som bekendt også en central konklusion i den diakrone analysedel (se fx afsnit 5.7.4 og 5.8.1.3). I de specialiserede regressionsmodeller i afsnit 19, så vi at kønseffekten var meget stabil på tværs af modeller, hvilket viser, at de literacy-orienterede variable heller ikke her havde nogen særlig forklarende effekt.

Analysens resultater afspejler således flere tendenser i den eksisterende forskningslitteratur. En del forskning har eksempelvis dokumenteret, at piger generelt klarer sig bedre i skriverelaterede aktiviteter end drenge (Reilly et al., 2019; Troia et al., 2013; Voyer & Voyer, 2014). I den skandinaviske skriveforskning er kønseffekten også dokumenteret, men her er det mere divergerende, om det er drengene eller pigerne, der klarer sig bedst. I Hultman & Westmans (1977) studie af tekster fra gymnasiet er det drengene, der klarer sig bedst i den forstand, at deres tekster gennemsnitligt set vurderes højere end pigernes. Forfatterne sammenligner også sproglige træk i elevteksterne med en benchmarkingkorpus af journalistiske tekster og konkluderer, at drengenes tekster minder mere om de journalistiske tekster end pigernes. Magnusson & Kokkinakkis (2009, 2011) finder ligeledes kønsrelaterede forskelle i et korpus af elevtekster, og også her er det drengene, der ligger øverst i ordlængde, OVIX og nominalkvote, mens det hos Vagle (2005) er pigerne, der ligger højest i tekstlængde. Vagle viser imidlertid også, at drengenes ordlængdeværdier skiller sig ud, hvis man ser på spredningen, hvor drengene havde flere tekster med meget lave eller meget høje værdier.

Det er vanskeligt at give en teoretisk forklaring på, hvorfor køn har så forholdsvis stor betydning for udformningen af elevernes tekster. Køn er ikke en vigtig faktor i teorier om sprogudvikling eller i nogle af de skriveudviklingsmodeller, jeg refererede til i kapitel 2. Det ser i højere grad ud til at være en latent faktor, som man kan måle den empiriske betydning af, men som ikke er en ekspliciteret askriptiv faktor som fx sproglig baggrund eller social baggrund ofte er. En ofte anvendt teori til at forklare kønsforskelle er den såkaldte 'expectancy-value theory', som i uddannelsesforskningen oftest forbindes med Jacquelynne Eccles. Pointen med denne teori er, at der er sammenhæng mellem elevens vurderinger af egne evner inden for bestemte fag og de faglige valg, som eleven træffer, og dermed også elevens faglige præstationer (Eccles, 1983). Heyder et al. (2017) anvender fx denne teori i et studie, hvor de viser, at det er drengens selvforståelse (ability self-concept) snarere end deres verbale sproglige evner, der bedst forklarer deres karakterer i sprogfag. I denne optik reflekterer kønsforskelle med andre ord snarere forskelle i faglig selvforståelse og motivation end kønsforskelle i elevernes egentlige evner. Der er imidlertid to forhold ved analyserne i dette kapitel, der ikke understøtter, at en lignende konklusion kan drages i dette studie. For det første har jeg konkrete sproglige træk i elevernes tekster som outcomevariable og ikke elevernes karakterer, hvilket betyder, at mine outcomevariable er upåvirkede af de biases, som opstår i manuel vurdering. Dette betyder, at det er elevernes færdigheder målt gennem træk i teksterne, der er outcome, hvorimod færdighed er en medierende effekt hos Heyder et al. Kønseffekten i dette studie kan således ses helt ned på tekst-, lexis- og syntaksniveau og siger noget om elevernes evne til at konstruere en tekst. For det andet kontrollerer jeg i studiet for skrivelyst, som, selvom det ikke er identisk med 'ability self-concept', må formodes at være et relateret psykologisk motivationskonstrukt. Skrivelyst har en forholdsvis lille forklarende effekt på køn, hvilket indikerer, at kønseffekten på elevernes tekster kun i mindre grad reflekterer kønsforskelle i elevernes motivation.

Læse- og skrivevaner viste sig ligeledes kun at have en lille forklarende effekt på kønseffekten. Dette står i modsætning til et nyere stort amerikansk studie af Downey et al. (2022), som viser, at det i højere grad er de ekstra-curriculære aktiviteter uden for skolen, end det er aktiviteter i skolen, der er årsag til forskelle i faglige præstationer mellem kønnene. Piger udvikler sig således fagligt mere end drengene, når de ikke er i skole, men denne forskel udlignes i skolen, hvilket får forfatterne til at pege på, at det i højere

grad er aktiviteterne uden for skolen, der giver pigerne en faglig fordel frem for drengene. Studiet siger ikke noget om, hvilke ekstra-curriculære aktiviteter det er, pigerne gavner af, men det virker oplagt at læse- og skriveaktiviteter i fritiden er en væsentlig del af det. Der er dog ikke noget i denne afhandlings analyse, der peger på, at læse- og skriveaktiviteter i fritiden er en væsentlig forklarende faktor.

6.4.4 Skriveudviklingens literacydimension

Selvom de tre literacy-orienterede variable havde en forholdsvis begrænset forklarende effekt på de askriptive variable, viste de komparative regressionsmodeller imidlertid, at de tre variable i varierende grad i sig selv havde mere direkte betydning for de lingvistiske variable. Særligt skrivelyst viste sig at hænge stærkt sammen med tekstlængde i begge teksttyper og OVIX, mens læsevaner viste en sammenhæng med ordlængde i både de beskrivende og de berettende tekster samt OVIX. Skrivevaner viste en lille moderat effekt på ordlængde i de beskrivende tekster, men ellers viste skrivevaner ingen sammenhæng med de øvrige variable.

Graham (2018) pointerer i Writers-within-community-modellen, at skrivning er "fueled at the individual level by beliefs about writing, including the value and utility of the writing assignment and perceptions of competence, as well as one's beliefs, identity, and commitment to the community in which writing is situated..." (s. 270). Skrivelyst er i denne forstand både et individuelt fænomen, som påvirker elevernes indsats og de valg, som eleverne træffer under skrivningen, og samtidig et kollektivt betinget fænomen, som påvirkes af det skrivefællesskab, skriveprocessen foregår i. Det er således ikke overraskende, at analyserne i dette kapitel viser en forholdsvis stærk sammenhæng mellem skrivelyst og flere af de lingvistiske variable. Særligt sammenhængen med tekstlængde var forventet. I Writers-within-community-modellen placerer Graham således skrivelyst (det Graham kalder 'beliefs') tæt på de kognitive kontrolmekanismer, herunder arbejdshukommelse, som er en vigtig del af det at skrive lange tekster. Analyserne i dette kapitel bekræfter således, at der er en stærk sammenhæng mellem individuel skrivelyst og antallet af ord, som eleverne er i stand til at producere. Det er imidlertid væsentligt at pointere, at skrivelyst i denne afhandling forstås som et generisk fænomen, og det er derfor ikke bundet af den specifikke skrivesituation eller skrivefællesskab, sådan som det eller kon-

ceptualiseres hos Graham. Selvom den generiske tilgang er standard i forskningslitteraturen (Ekholm et al., 2018), viser Jeffery & Wilcox (2014) i et kvalitativt studie af skrivelyst hos elever på udskolings-/gymnasieniveau, at der er relativt stor variation i skrivelyst mellem fag og genrer og endda også mellem de enkelte skriveordrer. Det er således en mulighed, at sammenhængen ville have været endnu stærkere, hvis jeg havde målt skrivelyst som et mere partikulært konstrukt. Wright et al. (2021) viser desuden i et mindre studie af forholdet mellem forskellige motivationskonstrukter og læreres bedømmelser af elevernes generelle skrivefærdigheder, at det i højere grad er elevernes selvbillede/selvforståelse (writer self-belief eller self-efficacy) end elevernes skrivelyst (value of writing), der bedst kan forudsige kvaliteten af elevernes skrivefærdigheder. Havde jeg undersøgt denne selvkonceptuelle del snarere end et generisk konstrukt som skrivelyst, kunne det ligeledes tænkes, at der var fundet en endnu stærkere sammenhæng mellem motivation og elevernes tekster.

Læsevaner var som bekendt den af de tre literacy-orienterede variable, der viste næststærkest sammenhæng med de lingvistiske variable, og her var sammenhængen stærkest ved de to leksikale variable. Sammenhængen mellem læsevaner i en fritidskontekst og skrivepræstationer i skolen er ikke modelleret direkte i Grahams (2018) model, eftersom modellen har et eksklusivt fokus på skriveaktiviteter. At læseaktiviteter et tæt forbundet med skrivekompetence er imidlertid udbredt og konventionelt accepteret i literacyforskningen (Barton, 2007; Gee, 2015). Det, som Grahams model alligevel viser om forholdet mellem fritidslæsning og skoleskrivning, er, at skrivefællesskaber, som vi her lige så godt kan kalde literacyfællesskaber (eller skriftsprogsfællesskaber), er forbundet med hinanden gennem sociale, institutionelle eller historiske relationer, og at individer krydser mellem disse fællesskaber på tværs af tid og sted. Så selvom fællesskaberne også afgrænser sig fra hinanden, fordi de har forskellige formål og/eller er situeret i forskellige sociale eller fysiske kontekster, så påvirker de også hinanden – både på det sociale niveau og på det individuelle, kognitive niveau. Der er imidlertid forsket ganske lidt i relationen mellem læsevaner i fritiden og skrivekompetence i skolen, og det har primært været med fokus på enten fremmedsproglæsning og –skrivning (se fx Lee & Krashen, 2002) eller begynderlæsning og –skrivning (se fx Aram et al., 2013). Graham et al. (2018) viser imidlertid i en metaanalyse, at læsestyrkende indsatser i undervisningen også har effekt på

elevernes skrivefærdigheder – både på et overordnet tekstkvalitetsniveau og på mere specifikke tekstuelle træk som tekstlængde og stavning. Selvom der her ikke er tale om transfer mellem literacyfællesskaber (som i denne afhandling), viser studiet imidlertid, at læseaktivitet er en vigtig faktor for udviklingen af skrivekompetencer. Dette bekræftes også delvist i dette studie, selvom det er svært at give en teoretisk forklaring på, hvorfor læsevaner kun hænger sammen med de leksikale variable og ikke tekstlængde eller de syntaktiske variable.

Den sidste literacy-orienterede variabel er skrivevaner, som samtidig er den variabel med de mindst interessante resultater. Ud over en moderat effekt på ordlængde viste skrivevaner ikke nogen særlig sammenhæng med de lingvistiske variable. Der er flere mulige forklaringer på dette, men den mest oplagte er, at eleverne primært skriver længere tekster, når de er i skolen. Skrivevanevariablen dækker som bekendt over længere tekster (dagbøger, blogs og historier) samt æstetiske tekster som digte og dermed ikke over funktionelle tekster som sms'er eller opslag på sociale medier. Skriveudvikling begynder i de helt tidlige år i de nære relationer i hjemmet gennem læseaktiviteter, men når børnene når skolealderen bliver skolen den primære kontekst for elevernes skriveudvikling (Bazerman et al., 2018). Camacho et al. (2021) rapporterer ligeledes om vanskeligheder ved at finde et validt og dækkende mål for elevers skrivevaner. Forfatterne peger på, at udfordringerne både er forbundet med invalide selvrapporteringer fra skoleelever samt med at finde et antal items, der dækker alle relevante skriveaktiviteter. Dertil kan tilføjes, at mål for skrivevaner, som er baseret på cluster- eller faktoranalyser, ofte forudsætter, at de målte skriveaktiviteter er på samme niveau og har samme betydning for udviklingen af elevernes skrivefærdigheder, hvilket må formodes at være en tvivlsom hypotese.

Kapitel 7: Diskussion og konklusion

Kapitel 7 består af tre dele. Den første del er en diskussion af analyseresultaternes implikationer for studier af skriveudvikling. Jeg fokuserer her dels på potentialer og begrænsninger ved at studere skriveudvikling fra et lingvistisk perspektiv, dels på forskellen mellem en *integreret ontologisk model* for skriveudvikling over for et *metodisk perspektivskifte* på skriveudvikling. Kapitlets anden del fokuserer på de danskdiraktiske implikationer af analyserne. Diskussionen starter med en beskrivelse af forholdet mellem lingvistik og didaktik, hvorefter jeg diskuterer analysernes implikationer for først skriveevaluering og dernæst skriveundervisning i danskfaget. Den sidste del af dette kapitel udgøres af en samlet konklusion på afhandlingens forskningsspørgsmål samt forslag til videre forskning i skrivning og skriveudvikling i skolen.

7.1 Afhandlingens implikationer for studier af skriveudvikling

Afhandlingens analyser har som bekendt vist, hvordan skriveudvikling kan beskrives lingvistisk både diakront gennem kombinationen af kvantitative og kvalitative analyser og synkront gennem multiple regressionsmodeller, der undersøger relationen mellem sproglige træk i teksten og en række ekstra-lingvistiske variable. Studiet har derigennem belyst en række potentialer og begrænsninger ved at belyse skriveudvikling fra et lingvistisk perspektiv, hvilket selvfølgelig også giver anledning til en række overvejelser over, hvad disse potentialer og begrænsninger evt. måtte betyde for fremtidige studier af skriveudvikling. I dette afsnit vil jeg derfor kort diskutere to forhold. For det første vil jeg diskutere nogle af de epistemiske begrænsninger, som er forbundet med at studere skriveudvikling fra et lingvistisk perspektiv, og for det andet vil jeg beskrive og diskutere, hvordan analyserne giver anledning til at tænke skriveudvikling ud fra en samlet og integreret ontologisk model frem for gennem metodiske perspektivskifter.

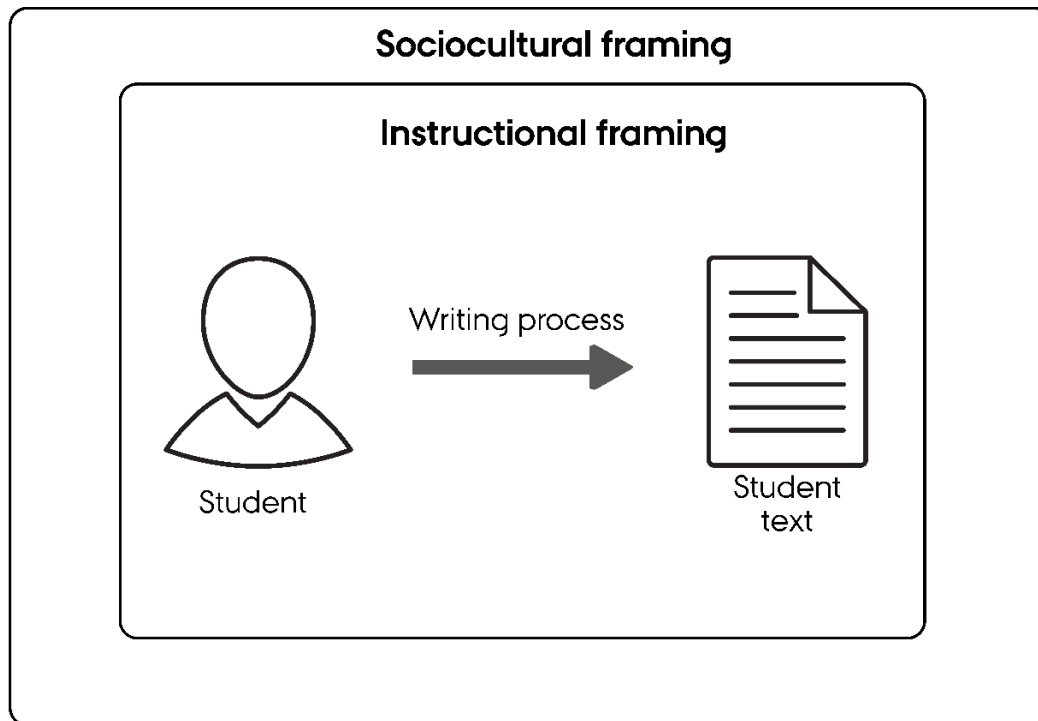
7.1.1 Potentialer og begrænsninger ved lingvistiske beskrivelser af skriveudvikling

Digitale, automatiserede analyser af sproglige træk i elevtekster har flere potentialer, som jeg også har beskrevet i kapitel 1. For det første er de skalérbare. Det vil sige, at det er muligt at analysere meget store mængder af tekst, så modsat manuelle analyser er der

altså ingen begrænsninger i størrelsen af datasættet. For det andet er det muligt at gøre de computationelle værktøjer transparente, sådan så det er åbent og gennemskueligt, hvordan analyserne er udført (fx ved at afrapportere Python-kode). På dette punkt er automatiske analyseværktøjer baseret på forskellige former for neurale netværk og maskinlæring til gengæld mere udfordrede. Transparens er imidlertid med til at sikre ekstern reliabilitet i analyserne, ligesom det bidrager til åbenhed i et større open science-perspektiv. For det tredje giver de automatiserede analyser mulighed for at skabe viden om skriveudvikling, som ellers vil være vanskeligt at skabe gennem manuelle analyser (jf. distant reading-pointen fra afsnit 1.2.1).

Der er imidlertid også en række begrænsninger forbundet med denne type af analyse. Man kan spørge, hvad det er, som de lingvistiske analyser ikke siger noget om? Hvad er det for en viden om skriveudvikling, som man ikke kan generere ved blot at se på sproglige træk i elevernes tekster? Noget af det, der forsvinder, når en tekst konverteres til et numerisk billede, er den visuelle fremtræden af teksten (layout etc.) på siden, som ellers er en væsentlig del af det, som elever fokuserer på som et led i deres skriveudvikling (Kress, 1994). Der er ligeledes klare begrænsninger ved lingvistiske analyser ift. til at sige noget om indholdet af teksten (jf. Perelman-pointen i afsnit 3.5.2.2), ligesom lingvistiske analyser heller ikke kan sige noget om den multimodale kompleksitet, som fx er en del af nyere typer af digitale elevprodukter i danskfaget som fx Scratch-tekster eller weblogs.

Der er imidlertid også mere grundlæggende dimensioner af skriveudviklingen, som ikke indfanges af lingvistiske beskrivelser af elevtekster. Analyser af sproglige træk dækker på den ene side over et meget reduceret skriveudviklingskonstrukt, idet de alene undersøger de lingvistiske ressourcer, som eleven har anvendt til at skabe sin tekst i den konkrete skrivesituation. På den anden side er de elever, som er i stand til at producere komplicerede tekster, ifølge Deane (2013) typisk også dem, der har mange af de øvrige sociokognitive evner, som man forbinder med dygtige skrivere. På denne måde kan de lingvistiske strukturer i teksterne ses som en form for proxy for en række sociokognitive kompetencer. Dette kan fx illustreres gennem figur 11, som er gengivet ud fra Tannert (2021).



Figur 11: Model af grundlæggende dimensioner af tekstaktiviteter i skolen. Gengivet ud fra Tannert (2021).

Modellen viser en grundlæggende tekstaktivitet i skolen. I midten ser vi eleven og elevens tekst, som er et resultat af en skriveproces indlejret i en bestemt skriveidaktisk kontekst, som igen er indlejret i en større sociokulturel ramme (herunder de institutionelle rammer). Pointen er, at skriveudvikling er relateret til alle dimensioner af modellen, og alle dimensioner af modellen kan i princippet undersøges uafhængigt af hinanden. Man kan fx undersøge den skriveidaktiske rammesætning gennem observationsstudier, skriveprocessen gennem logdata (fx Leijten et al., 2015) eller den sociokulturelle rammesætning af skriveudvikling gennem policyanalyser (fx Elf & Troelsen, 2021) etc. En anden tilgang er at undersøge, hvordan de forskellige dimensioner i modellen påvirker hinanden. Et eksempel er Rousse-Malpat (2019), som undersøger, hvordan forskellige didaktiske tilgange til skriveundervisning påvirker morfosyntaktiske træk i en række elevtekster.

Man kan imidlertid også i enkelte tilfælde undersøge nogle af dimensionerne indirekte ved at anvende en dimension som en proxy for en anden. Den mest åbenlyse proxyrelation er mellem teksten og eleven, hvor sproglige træk i teksten tages som et udtryk for elevens skriftsproglige færdigheder. Dette har jeg fx gjort i denne afhandling. Som pointeret i afsnit 3.2.2.3 anser jeg elevernes eksternaliserede sprog (det sproglige output) som værende forbundet til deres internaliserede sprog (dvs. deres skrivefærdigheder), hvilket

betyder, at jeg kan sige noget om elevernes skrivefærdigheder ved at se på teksten. Her er dog det klare forbehold, at forholdet mellem sprogligt output og færdighed selvfølgelig ikke er 1:1, eftersom elevernes præstationer selvfølgelig varierer relativt til en lang række faktorer (herunder teksttype, motivation etc.), og dermed udtrykker en enkelt tekst selvfølgelig ikke elevens fulde færdighedsniveau. Deane (2013) foreslår derfor også, at automatiserede analyser af lingvistiske træk kombineres med andre typer af undersøgelsesmetoder, sådan så man etablerer en form for undersøgelsesøkologi, hvor forskellige metoder og måleinstrumenter supplerer hinanden og undersøger forskellige dimensioner af skriveudviklingskonstrukten (fx dem som er angivet i figur 11).

7.1.2 Integreret skriveudviklingsmodel eller metodisk perspektivskifte

Dette bringer os til det andet forhold, jeg vil fremhæve på baggrund af dette studie. Donahue & Lillis (2014) pointerer, som jeg også har beskrevet tidligere i kapitel 2, at der findes forskellige skriveudviklingsmodeller, som udgør forskellige fokale optikker, hvorfra skriveudvikling kan beskrives og undersøges. Modellerne er ikke gensidigt udelukkende, men udgøres netop af forskellige supplerende perspektiver på skrivning som en multidimensionel og kompleks aktivitet. Der er imidlertid også, som det også pointeres af Bazerman (2013, s. 427-430), en tendens til at disse optikker på skriveudvikling behandles som dikotomisk adskilte metodiske perspektiver på skriveudvikling. Dette indebærer, at forskellige ontologiske dimensioner af skrivning (sociokulturelle, kognitive, lingvistiske etc.) manifesterer sig i forskellige gensidigt uafhængige traditioner (ofte med eksplicit anvendelse af paradigmebegrebet) for skriveforskning, og disse dimensioner har, som der fx ses tendenser til i Lankshear & Knobel's (2011) klassiske tekst om literacybegrebets historie, ofte et implicit hierarki, hvor sociokulturelle eller økologiske modeller rangeres som værende bedre end fx kognitive eller lingvistiske modeller. Det er imidlertid vigtigt at fastholde, at de integrerede, holistiske og økologiske modeller ikke nødvendigvis fører til en udvidet forståelse af skriveudvikling, men lige så godt kan føre til et metodisk perspektivskifte, der erstatter (og ikke supplerer eller udvider) de eksisterende perspektiver, herunder det lingvistiske.

En bevægelse fra det lingvistiske til det sociokulturelle implicerer således ikke nødvendigvis en bevægelse fra en snæver literacyforståelse til en bred literacyforståelse, men kan i lige så høj grad tænkes at markere en overgang fra én reduceret literacyforståelse til

en anden. I denne afhandling har jeg forsøgt at kombinere en lingvistisk model for skriveudvikling med en psykologisk og sociologisk model, og jeg har forsøgt at fortolke disse perspektiver i sammenhæng i afhandlingens synkrone analysedel. Hvis man skal fastholde et integreret perspektiv på skriveudviklingen må forskellige datatyper således kombineres i fortolkningsprocessen, hvilket imidlertid stiller store krav til den enkelte forsker såvel som til den samlede stringens og transparens i fortolkningen.

7.2 Afhandlingens danskdidaktiske implikationer

I denne afhandling har jeg undersøgt skriftsproglig udvikling i danskfaget, i den forstand at jeg har beskrevet, hvordan skolske tekster kan se ud i kontekst af danskfaget. Denne viden bidrager selvfølgelig til danskfagets didaktik, men den gør det på forskellige måder og i forskellig grad. I det følgende vil jeg derfor diskutere, på hvilke måder afhandlingens resultater kan informere forskellige områder af danskfagets didaktik. Jeg indleder med en kort beskrivelse af forholdet mellem lingvistikken og didaktikken, hvorefter jeg fokuserer på henholdsvis skriveevaluering og skriveundervisning i danskfaget.

7.2.1 Om forholdet mellem lingvistik og didaktik

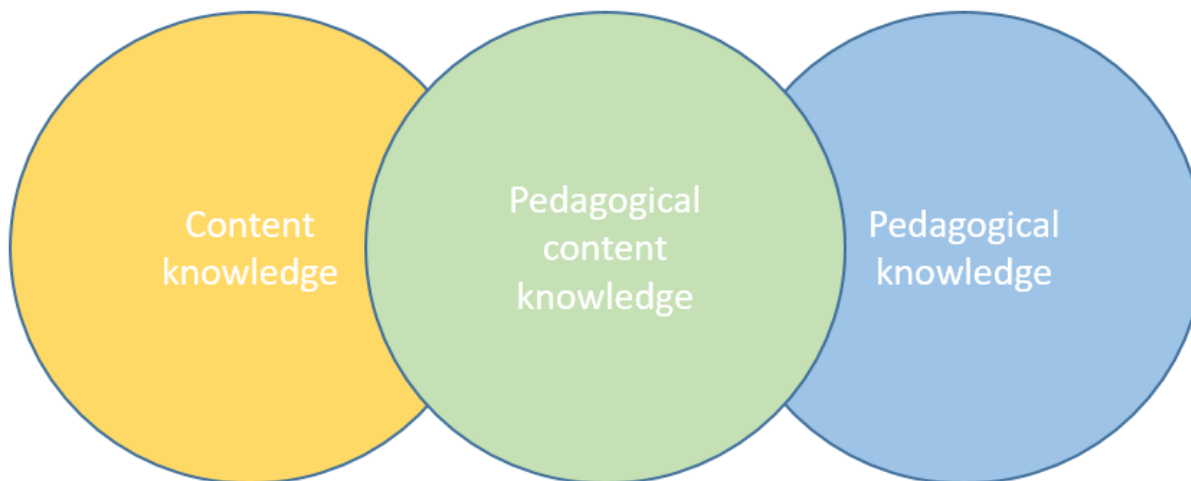
Teleman (1978) peger på, at en central begrænsning ved lingvistiske skriveudviklingsstudier er, at ”De frågor vi kunde ställa och ställde bestämde de svar vi fick. Och svaren bestämde i sin tur vilka pedagogiska konklusioner som kunde dras” (s. 39). Det, Teleman her adresserer, er den tendens, hvor den lingvistiske forskning i skrivning forveksles med en præskriptiv didaktik, hvor der sættes lighedstegn mellem forskningsresultater og undervisningsindhold. Det kunne fx være, hvis et studie har fundet ud af, at kompleksitet i nominalfraserne indikerer skriveudvikling, og den didaktiske konsekvens derfor bliver, at vi må undervise formalistisk i frasekompleksitet i danskundervisningen. En sådan slutning implicerer imidlertid en voldsom reduktion af både lingvistikken og didaktikken. Halliday et al. (1964) peger i stedet på, at lingvistikkens relation til didaktikken går gennem læreren. ”... the place for both phonetics and linguistics is *behind* the classroom teacher, in the training that he received for his job ... in the preparation of the syllabus ... and in the preparation of the teaching materials ...” (s. 187). Lingvistikkens indflydelse – i hvert fald i sin mere deskriptive form – er altså i denne optik i højere grad knyttet til

lærerens viden om fag og fagdidaktiske forudsætninger for undervisningsplanlægning end til selve undervisningssituationen og -indholdet.

Et forsøg på at bygge bro mellem lingvistikken og de pædagogiske videnskaber findes hos Hudson (2020), som med termen 'pedagogical linguistics' peger på, at lingvistikken og didaktikken står i en dialektisk relation til hinanden. Didaktikken påvirker lingvistikken, idet lingvistikkens genstandsfelt, sprogbrug, påvirkes af den undervisning, folk modtager i løbet af deres skolegang. Lingvistikken påvirker imidlertid også didaktikken i den forstand, at lingvistikken tilbyder viden om sprog, som er en vigtig del af sprogundervisningen i både modersmål og fremmedsprog. Hudsons ærinde er imidlertid primært at etablere en forbindelse fra pædagogikken til lingvistikken, og han gør således ikke meget ud af at beskrive, hvordan lingvistikkens bidrag til didaktikken mere konkret ser ud. Dette adresseres imidlertid eksplicit af Widdowson (2020), som også specifikt kommenterer på korpuslingvistikens bidrag til didaktikken. Widdowson forholder sig imidlertid kritisk til korpuslingvistikens didaktiske potentiale. Selvom korpuslingvistikken kan tilbyde objektive analyser af autentiske sprogdata, som de udfolder sig i klasserummet, så kan den ikke " ... reconcile the disparity between the objective and the process of learning..." (Widdowson, 2020, s. 56). Der er således ikke noget ved korpuslingvistiske analyser, der siger noget om, hvordan en elevs skriveudviklingsproces faciliteres didaktisk henover skoleårene. På denne måde er Widdowson på linje med Halliday et al. (1964) og Teleman (1978). Der er således ikke nogen direkte linje fra lingvistikken til didaktisk praksis. Relationen kan kun være indirekte gennem lærerne og læremiddeludviklerne, som kan kombinere lingvistisk viden med andre typer af viden med henblik på at planlægge, gennemføre og evaluere undervisning.

Relationen kan således beskrives ud fra et lærerprofessionalitetsperspektiv, fx gennem Shulmans (1986) 'pedagogical content knowledge' (PCK). PCK betegner en lærers professionelle viden. Den består grundlæggende af to typer af viden, 'content knowledge' og 'pedagogical knowledge', som tilsammen former PCK. Content knowledge refererer til lærerens viden om fagets indhold. For en dansklærer vil det fx dække over viden om sprog, litteratur, kunst, digitale medier etc. Pedagogical knowledge refererer derimod til lærerens viden om pædagogik og didaktik, dvs. undervisningens planlægning, gennemførelse og evaluering. Når disse videnstyper kombineres opstår der det, som Shulman (1986) kal-

der PCK, som refererer til lærerens evne til at undervise i sit fags indhold på en pædagogisk set kvalificeret måde. PCK dækker med andre ord over lærerens fagdidaktiske viden og dømmekraft (Kansanen, 2009).



Figur 12: Grafisk visualisering af Shulmans (1986) 'pedagogical content knowledge'.

Afhandlingens analyseresultater er på sin vis både relateret til 'pedagogical knowledge' og 'content knowledge'. Det er pædagogisk viden i den forstand, at det siger noget om elevens skriftsprog, som det udfolder sig inden for den særlige pædagogiske kontekst, som er skolen og danskfaget. På denne måde bidrager afhandlingens resultater med viden om, hvad der er særligt ved tekster skrevet inden for en pædagogisk ramme. Resultaterne siger imidlertid ikke noget om, hvordan udviklingen af skriftsproglige færdigheder understøttes didaktisk. Man kan derfor også anse viden om skriftsproglig udvikling som en del af lærerens danskfaglige viden forstået på den måde, at man kan se det som en mere almen viden om sprog og sprogbrug. Upåagtet om man fortolker viden om skriftsproglig udvikling som pædagogisk eller faglig viden, så viser PCK-modellen imidlertid en mulig relation mellem lingvistikken og didaktikken, som altså her går gennem den professionelle lærer og fagdidaktikken. Spørgsmålet er imidlertid, hvordan afhandlingens analyser mere specifikt kan bidrage til forskellige dimensioner af danskfaget. Jeg vil i det følgende kort kommentere på dette spørgsmål i forhold til først skriveevaluering og -bedømmelse og til sidst i forhold til mere konkrete skrivedidaktiske tiltag i danskundervisningen.

7.2.2 Om relevansen af viden om skriftsproglig udvikling for skriveevaluering i dansk

Den seneste rapport fra følgegruppen for én bedømmer ved folkeskolens afgangsprøver viser, at der ved afgangsprøverne i skriftlig fremstilling i dansk kun er mellem 40-50% sandsynlighed for, at en tekst får samme karakter, hvis den bedømmes af to forskellige bedømmere (Dolin et al., 2018). Det er derfor særligt relevant at diskutere, hvorvidt og på hvilken måde viden om skriftsproglig udvikling kan være med til at understøtte og informere evalueringspraksisserne, der knytter sig til elevtekster skrevet i danskfaget.

I et studie af de rationaler, som lærere anvender, når de bedømmer elevtekster, identificerer Jølle & Skar (2020) to forskellige tilgange eller rationaler til skrivebedømmelse, som befinder sig i hver sin ende af et kontinuum. Den ene tilgang, som forfatterne kalder for 'professional raters', indebærer, at bedømmeren entydigt baserer sin bedømmelse på de formelle evalueringskriterier, mens den anden tilgang, som kaldes 'independent teacher experts', indebærer, at bedømmeren selektivt anvender de officielle evalueringskriterier og i højere grad beror sig på implicite, personlige og subjektive kriterier (ofte baseret på fortolkninger om den enkelte elevskrivners motivation eller intentioner). Et lignende fund ses hos Troelsen (2020), som beskriver, hvordan "elevteksten forsvinder" (s. 80), når censorer enten forholder sig instrumentelt til evalueringskriterierne eller mere normativt til implicite tekstnormer.

Viden om lingvistiske udviklingstræk i elevernes tekster kan informere begge de beskrevne bedømmerpositioner. Den kan informere og kvalificere bedømmere, der anvender en normativ og fortolkende tilgang til bedømmelsen, ved at skærpe deres "reflekterade och välgrundade läsningar av elevernas texter" (Borgström & Ledin, 2014, s. 162) med forskningsmæssig viden om, hvad der typisk kendetegner tekster på forskellige niveauer og klassetrin. Dette er på sin vis relateret til PCK-pointen i forrige afsnit, idet det beror på den forudsætning, at man kan kvalificere evalueringen af elevernes tekster ved at styrke lærernes dømmekraft med forskningsmæssig viden. Viden om skriftsproglig udvikling bliver på denne måde en del af lærerens fagdidaktiske vidensgrundlag og dermed en del af grundlaget for lærerens dømmekraft i den enkelte evalueringssituation. Proble-

met med denne tilgang er, som det også pointeres af Jølle & Skar (2020), at det ikke bidrager til at løse problemet med lav bedømmerreliabilitet, men snarere forstærker problemet, eftersom bedømmelsen vil blive endnu mere afhængig af den enkelte lærer.

En anden tilgang vil derfor være, at viden om skriftsproglig udvikling anvendes som en del af grundlaget for udviklingen af skalaer og evalueringskriterier. Denne tilgang understøttes bl.a. af Sandiford & Macken-Horarik (2020), som på baggrund af kvalitative analyser af en række berettende elevtekster viser, at sprognære og lingvistisk-informerede analyser baseret på etablerede skalaer eller rubrics kan opfange forskellige former for skriveudvikling i elevernes tekster, herunder genrespecifikke udviklinger, ligesom analyserne fx funktionelt kan kobles til tekstens retoriske dimensioner. Dette kræver imidlertid en meget detaljeret grad af analyse, som kan være vanskelig at realisere inden for rammerne af mere formelle evalueringssituationer. Desuden er der selvfølgelig det problem, at tekstkvalitet ikke kan reduceres til lingvistiske strukturer alene, og som denne afhandlings analyser også har vist, er det vanskeligt at finde generiske udviklingsstadier, idet udviklingen i de lingvistiske variable er karakteriseret ved bl.a. udviklingsplateauer og nonlinearitet. Der vil således være tale om at finde en balance mellem at udvikle kriterier, der er tilpas generelle til, at den kan rumme variation på tværs af teksterne, og samtidig specifikke nok til, at de meningsfuldt kan appliceres og danne grundlag for bedømmelse af den enkelte tekst.

En tredje vej er, at viden om skriftsproglig udvikling anvendes ifm. udviklingen af forskellige automatiserede bedømmelsesværktøjer, som enten helt kan erstatte manuel bedømmelse eller supplere og støtte lærerens manuelle bedømmelse. Dette kendes, som jeg også har beskrevet i afsnit 3.5.2, bl.a. fra områder som 'automated essay scoring', der især er udbredt i den amerikanske forskningslitteratur (se fx Correnti et al., 2020; McNamara et al., 2014; Shermis, 2014). Foruden de validitetsbegrænsninger, der kan være forbundet med automatiseret bedømmelse af tekster (jf. pointen med BABEL-generatoren i afsnit 3.5.2.2), er det ligeledes en udfordring, at der er stor variation i de sproglige realiseringer af elevernes tekster relativt til teksttype og skriveordre. I denne afhandling har vi fx set en relativt stor variation mellem teksttyper ved analyserne af bl.a. ordlængde og sætningslængde. Samme pointe ses hos Palmér (2018), som derfor pointerer, at automatiserede tekstanalyser bør tilpasses og målrettes de enkelte skriveprøver, og at en egentlig

bedømmelse af en elevtekst bør bero på en blanding af kvantitative og kvalitative læsninger af teksten. Med Palmérs (2018) egne ord: ”Med stöd av en automatisk analys borde dock bedömningens kvalitet kunna höjas.” (s. 21).

I ovenstående har jeg peget på tre mulige koblingspunkter mellem viden om skriveudvikling og evaluering af elevtekster i danskfaget. De tre koblingspunkter kan opsummeres skematisk således.

Tabel 36

Tre koblingspunkter mellem viden om skriveudvikling og evaluering i danskfaget

Kvalificering af lærerens dømmekraft	Viden om skriveudvikling som en integreret del af lærerens fagdidaktiske viden (PCK). Det bliver således gennem lærerens professionelle dømmekraft i bedømmelsessituationen, at viden om skriveudvikling bringes i spil.
En del af grundlaget for udvikling af skalaer og evalueringskriterier	Viden om skriveudvikling som grundlag for at udvikle forskningsbaserede skalaer og evalueringskriterier til evaluering af elevtekster. Lingvistiske beskrivelser af skriveudvikling skal således inkluderes som en del af en række tekstnære evalueringskriterier.
Grundlag for udvikling af automatiserede analyseværktøjer til støtte af manuel bedømmelse	Viden om skriveudvikling danner grundlag for udviklingen af automatiserede analyseværktøjer, som kan give information om kvantitative strukturer i teksten, der kan supplere lærernes kvalitative læsninger af elevteksterne. Analyseværktøjerne må imidlertid i et vist omfang målrettes de enkelte skriveprøver/skriveordrer.

Der er ikke empirisk belæg i denne afhandling for at anbefale hverken den ene eller den anden af de tre beskrevne tilgange. Det er imidlertid klart, at de alle tre er forbundet med både potentialer og begrænsninger, og viden om skriveudvikling kan (og muligvis også bør) derfor spille en rolle ved alle tre punkter.

7.2.3 Om relevansen af viden om skriftsproglig udvikling for skriveundervisning i dansk

Det sidste del af diskussionen af de danskidaktiske implikationer angår relevansen af viden om skriveudvikling for skriveundervisning i danskfaget. Denne afhandlings analyse har som bekendt ikke berørt de didaktiske vilkår, som teksterne er skrevet under, eftersom det ikke har været muligt at indsamle informationer om dette som en del af afhandlingens datagrundlag. Afhandlingens analyseresultater har derfor som sådan ingen direkte konsekvenser for skriveundervisningen i danskfaget. Jeg vil derfor i stedet kort kommentere på to punkter, hvor korpusbaserede studier baseret på automatiseret tekstanalyse i mere generel forstand kan informere skrivedidaktisk praksis i danskfaget.

Modsat den modersmålsdidaktiske forskning er der i andet- og fremmedsprogdidaktikken mere udbredte overvejelser i forskningslitteraturen, der beskriver forskellige former for anvendelse af tekstkorpusser i undervisningen. Der er her typisk tale om tekstkorpusser bestående af tekster skrevet af enten elever eller professionelle skribenter. Et eksempel kunne være at inddrage autentiske eksempler i undervisningen på sætninger eller større tekstdele skrevet af elever i steder for de konstruerede teksteksempler, som eleverne typisk møder i læremidlerne. På denne måde vil de sproglige eksempler, som eleverne arbejder med i undervisningen, ligne elevens eget sprog mere. Her vil lærerne selv kunne konsultere forskellige tekstkorpusser og identificere eksempler, som er målrettet egne elever (Cobb & Boulton, 2015, s. 480). Alene i de forholdsvis få tekster, jeg analyserede i de tre tekstudvalg i kapitel 5, er der flere eksempler på sætninger (fx dem med overforbrug af paratakse) eller større tekstdele, som kunne være interessante at anvende som eksempler i skriveundervisningen, ligesom man kunne forestille sig, at eleverne sammenlignede og diskuterede bestemte retoriske virkemidler i elevtekster i forskellige genrer, arbejdede med ordlister eller på anden måde arbejdede med tekster, som dansklæreren forinden har identificeret i et tilgængeligt tekstkorpus, hvor teksterne er tagget med forskellig lingvistiske informationer baseret på automatiserede analyser (se eksempler på dette hos Reppen, 2010).

Et andet punkt, hvor automatiseret tekstanalyse kan spille en rolle i undervisningen, er i relation til forskellige former for automatiseret feedback og teknologisk støtte og vejledning til eleverne, mens de skriver. Graham et al. (2015) viser fx i en metaanalyse, at

computer-genereret formativ feedback, der genereres undervejs, mens eleverne skriver, kan bidrage til en væsentlig forbedring af kvaliteten af teksterne (gennemsnitlig vægtet effektstørrelse på 0,38). Denne form for feedback er baseret på viden om skriveudvikling samt store mængder af tekst, som danner grundlag for udviklingen af forskellige maskinlæringsbaserede værktøjer til skriveundervisningen. Det danske ATEL-projekt, som jeg har beskrevet i afsnit 1.3.2.1, har fx også haft som en del af studiets sigte, at det skulle bidrage til udviklingen af automatiseret feedback-værktøjer til læreren (Bremholm et al., 2022). På denne måde kan viden om skriftsproglig udvikling bidrage meget direkte til udviklingen af forskellige pædagogiske teknologier i skriveundervisningen, der kan informere både lærerens arbejde med at understøtte den enkelte elevs skriveudvikling samt den enkelte elevs arbejde med at konstruere en tekst.

7.3 Konklusion

Jeg har i denne afhandling undersøgt skriftsproglig udvikling fra et diakront og et synkront perspektiv i et korpus af elevtekster fra 5.-9. klassetrin. Analyserne tog udgangspunkt i tre overordnede forskningsspørgsmål, som alle havde to eller flere underspørgsmål. I det følgende vil jeg konkludere på de tre forskningsspørgsmål et ad gangen. For mere uddybende beskrivelser og fortolkninger af afhandlingens analyseresultater henvises der til delkonklusionerne for de to analysedele (henholdsvis afsnit 5.8 og 6.4). Jeg runder konklusionen af med forslag til og overvejelser over fremtidig forskning i skrivning og skriveudvikling.

7.3.1 Forskningsspørgsmål 1

Afhandlingens første forskningsspørgsmål lød: Hvilke sproglige karakteristika har et korpus af elevtekster skrevet i danskundervisningen på 5.-9. klassetrin, og hvad siger disse sproglige karakteristika om elevernes skriftsproglige udvikling? Dette spørgsmål er således besvaret i afhandlingens diakrone analysedel (kapitel 5). Forskningsspørgsmålet besvares gennem tre underspørgsmål, som jeg i det følgende vil besvare et ad gangen.

7.3.1.1 Hvilke forskelle ses der i fordelingen af de kvantitative, lingvistiske variable relativt til klassetrin og teksttype?

Den første af dem fem lingvistiske variable, tekstlængde, havde forholdsvis lineære stigninger i medianværdien fra 5. til 8. klassetrin med en øgning i medianværdien på minimum 15-30% for hver øgning i klassetrin. 9. klassetrin viste et overraskende fald i medianværdien, og ved begge teksttyper var det kun 5. årgang, der lå lavere. Analyserne viste desuden, at spredningen (IQR) var højere i udskoling end på mellemtrinnet, hvilket hang sammen med, at de korteste tekster i korpusset forblev på nogenlunde samme niveau, mens de længste tekster til gengæld blev markant længere henover klassetrin. Forskellene i tekstlængde mellem de to teksttyper var meget små. De berettende tekster havde lidt højere medianværdier og lidt højere spredning, men til gengæld var den samlede variationsbredde størst ved de beskrivende tekster.

De to leksikale variable, ordlængde og OVIX, viste små gennemsnitlige stigninger ved sammenligning af medianværdien for 5. og 9. årgang. Der var imidlertid ingen systematiske stigninger imellem de øvrige klassetrin. Der ses imidlertid ved begge variable et markant fald i spredningen (IQR) henover klassetrin. Modsat tekstlængde, hvor spredningen øges over tid, så falder den altså ved begge de leksikale variable. De typiske tekster (forstået som tekster med variabelværdier blandt de 50% midterste for årgangen) i udskoling er altså mere ensartede på det leksikale niveau end på mellemtrinnet, men til gengæld er der større spredning i længden af teksterne. De berettende tekster havde en lidt højere medianværdi i ordlængde end de beskrivende tekster. OVIX er udregnet på baggrund af elevens samlede tekstproduktion i begge teksttyper, og der kan derfor ikke sammenlignes mellem teksttyper.

De to syntaksvariable, sætningslængde og dependensafstand, viste heller ingen lineære udviklinger mellem klassetrin, men til gengæld kunne der observeres substantielle forskelle mellem tekster fra mellemtrin (5.-6. klassetrin) og udskoling (7.-9. klassetrin). Sætningslængden lå mellem 24-32% højere ved de tre udskolingsårgange, og ved dependensafstand var forskellen 9-12%. Ved begge variable skete den største udvikling mellem 6. og 7. årgang (dvs. i overgangen fra mellemtrin til udskoling), mens der til gengæld ikke var egentlige stigninger i medianværdien på tværs af de tre udskolingsårgange (7.-9. klassetrin). Forskellene imellem teksttyper var forholdsvis små. De berettende tekster havde

en anelse højere sætningslængde og dependensafstand end de beskrivende tekster, men forskellene var imidlertid ved begge variable meget små (ved sætningslængde er forskellen fx på under ét ord).

7.3.1.2 Hvilke leksikogrammatisk kendetegn har elevtekster, som repræsenterer henholdsvis forskellige variabelværdier og klassetrin?

De leksikogrammatisk analyser baserede sig på analyser af tre tekstudvalg af elevtekster. Analysen af de første tekstudvalg fokuserede på forholdet mellem tekstlængde og semantiske felter i teksten. Analysen viste, at tekstlængde i særlig grad hang sammen med uddybning af indholdet (forstået som øget størrelse af de semantiske felter i teksten) og i mindre grad med variation i indholdet (forstået som øget antal af semantiske felter i teksten). Samtidig viste analysen, at der var forholdsvis stor variation på individniveau i tekstlængde mellem de to teksttyper, hvilket indikerede, at tekstlængde ikke alene kan tages som en proxy for kognitive egenskaber hos eleven, men i lige så høj grad hænger sammen med en bevidst og pragmatisk skrivestrategi fra elevens side. Analysen viste desuden, at visse semantiske felter hang nøje sammen med teksttypen (fx det semantiske felt 'tid' og teksttypen beretning).

Analysen af det andet tekstudvalg havde fokus på forholdet mellem de leksikale variable, ordlængde og OVIX, og elevernes anvendelse af kohæsive bånd i teksten samt længde og kompleksitet i tekstens nominalfraser. Analysen viste, at anvendelsen af kohæsive bånd og længde og kompleksitet i nominalfraserne hang tættere sammen med teksttype end med variabelværdierne i ordlængde og OVIX. Der var særligt stor forskel på længde og kompleksitet i nominalfraserne mellem de to teksttyper, idet de beskrivende tekster generelt indeholdt flere lange og mere varierede nominalfraser end de berettende tekster. Dette hang bl.a. sammen med, at nominalfraserne i de beskrivende tekster havde en deiktisk funktion, hvor de blev brugt til at udpege forskellige dele af det billede, eleverne beskrev (fx 'pigen der er i gang med at gå væk'). Analysen indikerede ligeledes, at overforbrug af kohæsive bånd kan være et tegn på, at eleven befinder sig på et tidligt niveau i sin skriveudvikling.

Analysen af det tredje og sidste tekstudvalg fokuserede på forholdet mellem sætningslængde og dependensafstand på den ene side og hypotakse/paratakse-variation, ledsætningstyper og forfælsrealiseringer i teksterne på den anden side. Analysen viste, at

sætningslængde ikke øges som en direkte funktion af bestemte syntaktiske konstruktioner eller karakteristika, men at en teksts sætningslængde kan afhænge af mange forskellige syntaktiske forhold i teksten. Analysen identificerede to veje til øget sætningslængde: en ”kunstig” vej og en vej karakteriseret ved øget kompleksitet. Den kunstige vej er kendetegnet ved, at sætningslængden øges gennem overforbrug af parataktiske konjunktioner/konjunktivfraser (fx ’og så’), som anvendes til at forbinde hovedsætninger med hinanden. Dette var i særligt grad tilfældet i de berettende tekster, hvor tekstens enkelte handlingsforløb blev koblet samlet i lange parataktiske syntaktiske perioder (fx ”... og så ser jeg en video og så lavede jeg nogle lektier og så spiste jeg noget middagsmad og så tog jeg på arbejde...”). I disse tilfælde er der tale om sætninger, som syntaktisk set er meget simple og ensartede. Den anden vej til øget sætningslængde er i højere grad knyttet til øget variation og kompleksitet i sætningsopbygningerne. Her skyldes en høj sætningslængde og dependensafstand, at der fx er mere udpræget brug af hypotakse, større variation i forfæltet (fx gennem anvendelse af ledsætninger eller adverbelle led) eller større variation i anvendelsen af forskellige ledsætningstyper. Analysen pegede desuden på, at den primære syntaktiske udvikling i de berettende tekster sker gennem de adverbelle led, idet tekster med højere variabelværdier typisk også har længere og mere forskelligartede adverbelle led. I de beskrivende tekster var den syntaktiske udvikling i højere grad knyttet til øget længde af nominalfraserne, fx gennem integrationen af forskellige former for adjektiviske og nominelle ledsætninger.

7.3.1.3 På hvilken måde kan de sproglige karakteristika, som de kvantitative og kvalitative analyser af tekstkorpusset identificerer, bidrage til viden om skoleelevers skriftsproglige udvikling i danskfaget?

Det sidste underspørgsmål til forskningsspørgsmål 1 angår de lingvistiske analysers implikationer for forståelsen af, hvad der kendetegner skoleelevers skriftsproglige udvikling i danskfaget. I afsnit 5.8 1 identificerede jeg fem grundlæggende pointer om skoleelevers skriveudvikling, som kunne identificeres på baggrund af den diakrone analysedel. Den første pointe handler om, at aldersmodning (forstået som øget klassetrin) snarere er en proxy for akkumuleret skriveerfaring i skolen eller for kognitive dimensioner af elevernes skriveproces (fx arbejdshukommelse), end det er et strukturerende princip for elevers

skriveudvikling. På denne måde kan aldersmodning ikke tildeles en central kausal udsigelseskraft, men må snarere forstås som én blandt flere variable, der påvirker og former elevernes skriveudvikling. Den anden pointe er knyttet til overforbrug og udviklingsplateauer som vigtige begreber til at beskrive skriveudvikling fra et lingvistisk perspektiv. Udviklingsplateauerne er empirisk bestemte, idet der i afhandlingens datasæt kan identificeres perioder med fremgang i variabelværdierne og perioder med stagnering. Overforbrug er selvfølgelig også baseret på empiriske betragtninger i analyserne (fx overforbruget af parataktiske konjunktivfraser), men er i lige så høj grad også et teoretisk begreb lånt fra intersprogsforskningen, hvor det refererer til de situationer, hvor eleven mangler lingvistiske ressourcer til at udtrykke sig mere præcist og varieret og derfor kompenserer for det ved at gentage og derigennem overforbruge de sproglige ressourcer, som eleven har tilgængelig.

Den tredje pointe handler om sproglig variation som et tegn på skriftsproglig udvikling. De kvalitative analyser viste bl.a., at de tekster, der sprogligt set var mest komplekse, også var de tekster, der havde størst sproglig variation – især på det syntaktiske niveau. Dette bekræftes også af tidligere forskning. Der var imidlertid ikke nogen klar sammenhæng mellem variation og de kvantitative, lingvistiske variable i denne afhandling, eftersom disse måler gennemsnitlig længde af de sproglige strukturer. Den fjerde pointe er knyttet til begrebet 'stemme' som en vigtig del af skriveudviklingen. De kvalitative analyser viste, at elevernes stemme var til stede i mange af teksterne på tværs af teksttyper. Dette kom dels til udtryk gennem brugen af en eksplicit førstepersonsfortæller, dels gennem anvendelsen af kausale konjunktionaler (fx 'fordi'), epistemiske adverbialer ('helt sikkert') eller vurderende udsagn (fx 'stemningen er lidt trist'). Særligt sidstnævnte er knyttet til velkendte områder af den lingvistiske skriveforskning som fx appraisal (Martin, 1992) eller såkaldte 'hedges' og 'boosters' (Hyland, 2018), og tilstedeværelsen af disse på tværs af teksttyper gør det til et interessant fokuspunkt for fremtidig skriveforskning.

Den sidste og femte pointe handler om nonlinearitet og pluralitet som gennemgående træk ved skriftsproglig udvikling i grundskolen. Analyserne viste en relativt stor sproglig variation imellem teksttyper, hvilket giver anledning til at overveje, om man i stedet for skriftsproglig udvikling i entalsform bør tale om skriftsproglige udviklinger i flertal. Dette hænger sammen med, at entalsformen lægger op til at forstå skriveudvikling

som et generisk og lineært fænomen, mens flertalsformen åbner op for at forstå skriveudvikling som et partikulært og multidimensionelt fænomen præget af nonlinearitet og sproglig variation alt efter teksttype, kontekst, skriveformål etc. Dermed bliver diversitet og fleksibilitet vigtige begreber for at forstå, hvad skriveudvikling er, og hvordan det understøttes i en skolekontekst.

7.3.2 Forskningsspørgsmål 2

Forskingsspørgsmål 2 lød som bekendt: Hvilke metodiske potentialer og begrænsninger er der ved at beskrive skriftsproglig udvikling ved hjælp af automatiserede, kvantitative analyseværktøjer? Dette forskningsspørgsmål er således knyttet til den metodiske del af afhandlingen og fokuserer specifikt på potentialer og begrænsninger ved analysernes metodiske forankring. Denne del er således knyttet til pointer fra den diakrone analysedel. Forskingsspørgsmål 2 består af to underspørgsmål, som jeg nu vil besvare én ad gangen.

7.3.2.1 Hvilke metodiske potentialer og udfordringer er der specifikt forbundet med anvendelsen af de automatiserede analyseværktøjer i denne afhandling?

Analyserne i den diakrone analysedel viste, at en vigtig begrænsning forbundet med at studere skriveudvikling i dette studie var, at jeg ikke havde data om tekstkvalitet, og at de lingvistiske variable ikke kan tolkes som kontinuerlige skalaer for hverken kvalitet eller kompleksitet. Dette betyder, at man ikke nødvendigvis kan fortolke høje variabelværdier som et udtryk for hverken normativt set bedre tekster eller lingvistisk set mere komplicerede tekster. Tværtimod viste de kvalitative analyser, at en tekst med variabelværdier omkring medianen for årgangen også kan være en af de syntaktisk set mest komplicerede tekster. Dette tyder altså på, at der ved visse variable er et punkt, hvor tekster med en værdi derover udtrykker noget andet end kompleksitet. I denne afhandlings analyser var dette tydeligst ved variabelen sætningslængde, men tidligere forskning har vist samme tendens ved OVIX-analyser. Dermed viste analysen, at de automatiserede analyseværktøjer i sig selv primært er deskriptive analyseværktøjer. I den udstrækning analyserne skal sige noget om kvalitet, skal der suppleres med yderligere data, og skal analyserne sige noget om sproglig kompleksitet, er det langt fra alle sproglige mål, der kan bruges uden supplerende kvalitative analyser.

Et andet vigtigt punkt angår det didaktiske perspektiv i, at afrapporteringen og fortolkningen af automatiserede analyser fastholder et blik for både, hvad der er typisk for

en årgang (dvs. gennemsnits- og medianværdier), og hvor stor diversiteten er på en årgang (IQR, variationsbredde etc.). Skriftsproglig udvikling kan altså ikke alene reduceres til gennemsnitlige forandringer på en årgang, men må også knyttes an til udviklinger inden for tekstgrupper fra de øverste eller nederste kvartiler. Dette hænger sammen med, at årgange og skoleklasser langt fra udgøres af en homogen gruppe af elever, og derfor er det interessant, hvorvidt variationen i elevernes tekster (spredningen) bliver større, mindre eller forbliver uforandret over tid. I den diakrone analysedel så vi fx, at spredningen i tekstlængde øgedes forholdsvis meget henover klassetrin, mens OVIX modsat standardiserede sig over tid. Dette er interessant fra et didaktisk perspektiv, fordi det siger noget om de elevforudsætninger, som dansklæreren arbejder med på et pågældende klassetrin.

Et tredje vigtigt punkt angår udviklingen af lingvistiske mål og er relateret til en central metodisk diskussion inden for digital humaniora, som drejer sig om, hvorvidt udviklingen af lingvistiske mål og automatiserede analyseværktøjer er drevet af en teknocentriske tilgang, hvor det er teknologiens muligheder og begrænsninger frem for faglige agender, der styrer udviklingen. I denne afhandling har det fx været nødvendigt at supplere de kvantitative analyser med kvalitative læsninger af elevernes tekster for at få en dybere forståelse for, hvad de lingvistiske variable dækker over. En vigtig pointe er derfor, at der er behov for at vende tilgangen om, sådan så det er kvalitative og detaljerede analyser af elevernes tekster, der går forud for udviklingen af de kvantitative mål, og som derigennem er med til at definere og udpege lingvistiske områder, som set fra et skriveudviklingsperspektiv er interessante at måle. Der er altså brug for en tilgang, hvor vi måler det, der fagligt set er mest interessant, og ikke bare det, der teknologisk set er muligt. Dette kræver et større tværdisciplinært samarbejde mellem bl.a. lingvister, sprogdidaktikere og folk fra computer science.

7.3.2.2 Hvilke metodiske fokus- og udviklingspunkter giver anvendelsen af de automatiserede analyseværktøjer i dette studie anledning til?

Ovenstående pointe om den forskningsdrevne udvikling af lingvistiske mål, rejser naturligvis spørgsmålet om, hvad det er for en type af mål, der skal udvikles. Med udgangspunkt i en diskussion af Stanley Fishs (2018a, 2018b) kritik af korpuslingvistikens fortolkningsmæssige begrænsninger peger jeg i den diakrone analysedel på nødvendigheden

af, at fremtidige lingvistiske mål ikke blot er længdebaserede gennemsnitsmål af grovkornede lingvistiske strukturer, men i højere grad siger noget om det, Biber et al. (2020) kalder for 'grammatical complexity'. Dette indebærer, at de lingvistiske mål forholdsvis utvetydigt refererer til bestemte grammatiske strukturer i teksten, som meningsfuldt kan fortolkes uden supplerende kvalitative læsninger af teksten. Som eksempel peger jeg i kapitel 5 på behovet for at udvikle mål, der fx kan opfange syntaktisk variation, eftersom analyserne viste, at netop dette var en vigtig dimension af elevernes skriveudvikling.

7.3.3 Forskningsspørgsmål 3

Det tredje forskningsspørgsmål lød: Hvordan hænger elevernes skriftsproglige udvikling sammen med askriptive (køn og social baggrund) og literacy-orienterede variable (skrivelyst, læsevaner og skrivevaner) knyttet til den enkelte elevskriver? Dette forskningsspørgsmål besvares således i afhandlingens synkrone analysedel. Forskningsspørgsmålet har tre tilknyttede underspørgsmål, som jeg i det følgende besvarer ét ad gangen.

7.3.3.1 Hvilke af de ekstra-lingvistiske (herunder de askriptive og de literacy-orienterede) variable viser den stærkeste sammenhæng med de lingvistiske variable?

Det første skridt i den synkrone analysedel var at sammenligne den relative betydning af alle de ekstra-lingvistiske variable (herunder klasstrin) for de fem lingvistiske variable gennem regressionsmodeller med indbyggede beta-koefficienter. Analyserne viste, at klasstrin havde varierende effekt på de lingvistiske variable. Dog var der særlig stærk sammenhæng mellem klasstrin og tekstlængde, hvilket altså bekræfter resultaterne af den diakrone analysedel. Desuden var der statistisk signifikante effekter ved 9. årgang for alle de fem variable, hvilket indikerer en generel stigning med varierende hældning i alle de lingvistiske variable, hvis man kun ser på 9. årgang med 5. årgang som referencegruppe. Ved de tre literacy-orienterede variable var det skrivelyst, der viste den stærkeste sammenhæng med de lingvistiske variable (særligt ved tekstlængde og OVIX), mens læsevaner generelt havde mere begrænset effekt, men dog relativt stor effekt ved ordlængde i begge teksttyper og OVIX. Skrivevaner var den af de tre literacy-orienterede variable, der viste den laveste sammenhæng med de lingvistiske variable, idet den kun viste en lille moderat betydning for ordlængde og til dels OVIX. De to askriptive variable, køn og social

baggrund, viste generelt stærk sammenhæng med de lingvistiske variable. Særligt køn havde forholdsvis stor effekt på de lingvistiske variable. Analysen viste, at pigerne lå højest i tekstlængde, ordlængde og OVIX, mens drenge lå højest i de syntaktiske variable. Social baggrund havde også relativt stor betydning for de lingvistiske variable, men dog lidt mindre end køn. Ved social baggrund var effekten således størst ved tekstlængde og ordlængde, men meget lille ved de øvrige variable.

7.3.3.2 Hvilken betydning har social baggrund for de lingvistiske variable, når der kontrolleres for skrivelyst, læse- og skrivevaner og sproglig baggrund?

De specialiserede regressionsmodeller viste, at social baggrund i særlig grad havde betydning for tekstlængde og ordlængde, og at denne betydning var relativt stabil på tværs af de modeller, hvor de literacy-orienterede variable blev holdt konstante. Dette viser således, at kun en lille del af betydningen af social baggrund kan forklares ved, at eleverne fra den øverste gruppe havde højere skrivelyst eller bedre læse- og skrivevaner. Der var ingen statistisk signifikante resultater ved analysen af social baggrunds betydning for sætningslængde. Analysen bekræfter således til dels Bernsteins (1971) distinktion mellem elaboreret og begrænset kode, idet den viser, at der er forskelle i de lingvistiske ressourcer, som eleverne fra de to sociale grupper har anvendt i teksterne, og at disse forskelle bl.a. handler om at kunne producere mere tekst og at kunne anvende en større grad af specifikke og abstrakte ord (målt med ordlængde som proxy). At de literacy-orienterede variable ikke havde større forklarende effekt på betydningen af social baggrund, var imidlertid mere overraskende. Dette kan skyldes flere ting. Det kan fx skyldes målefejl eller andre begrænsninger i variablerne. Det kan også skyldes, at skolen kompenserer for betydningen af elevernes fritidslæsning og –skrivning, og det kan også tænkes at hænge sammen med, at elevernes læse- og skrivevaner ikke har så stor betydning for skriveudviklingen på de undersøgte klassetrin, men i stedet kan tænkes at have betydning på et tidligere stadie af skriveudviklingen.

7.3.3.3 Hvilken betydning har køn for de sproglige variable, når der kontrolleres for skrivelyst samt læse- og skrivevaner?

Køn viste sig generelt at have lidt større betydning for de lingvistiske variable end social baggrund. Endnu engang var betydningen størst ved tekstlængde, hvor kønseffekten var

forholdsvis konstant på tværs af regressionsmodeller, hvilket viste, at også her kunne de tre literacy-orienterede variable kun forklare en lille del af effekten. Ved ordlængde havde køn kun en forholdsvis begrænset effekt ved de beskrivende tekster, men også her var effekten forholdsvis stabil i de modeller, hvor de literacy-orienterede variable blev holdt konstante. Kønsforskellene i elevernes tekster reflekterer således kun i mindre grad kønsforskelle i kønnenes motivation eller literacy-aktiviteter i fritiden. Ved sætningslængde var der tale om en omvendt kønseffekt, idet drengene her skrev længere sætninger end pigerne, og interessant nok øgedes kønseffekten, når de literacy-orienterede variable blev holdt konstante, hvilket indikerer, at lav skrivelyst og dårlige læse- og skrivevaner faktisk har en forklarende effekt hos kønseffekten ved drengene. Dette uddyber således de kvalitative analyser fra den diakrone analysedel, der viste, at sætningslængde ikke nødvendigvis er en direkte funktion af syntaktisk kompleksitet, men også kan indikere overforbrug af paratakse eller ellipsekonstruktioner. De dokumenterede kønseffekter i den synkron analyse er interessante, fordi studiet undersøger konkrete træk i elevernes tekster, og dermed er studiets outcomevariable upåvirkede af eventuelle kønsbiases i manuel bedømmelse af elevtekster.

7.3.4 Forslag til og overvejelser over fremtidig forskning i skrivning og skriveudvikling

Jeg runder denne afhandling af med forslag og overvejelser over fremtidig skriveforskning. Jeg har allerede tidligere i denne konklusion samt i afsnit 5.8.2.3 peget på behovet for at udvikle nye lingvistiske mål, som ikke er drevet af teknologiens muligheder alene, men som i lige så høj grad er drevet af eksisterende viden om skrivning og skriveudvikling, og som udtrykker det, Biber et al. (2020) kalder for 'grammatical complexity'. Jeg vil derfor i dette afsnit fokusere på tre andre områder, som vil være relevante at undersøge set i kontekst af denne afhandlings resultater.

Det første område, jeg vil fremhæve, angår vigtigheden af flere genre- eller teksttypespecifikke studier af skriveudvikling. Når eleverne når et vist niveau i deres skriveudvikling begynder deres tekster at antage en mere funktionel karakter, og eleverne bliver i højere grad i stand til at variere deres sproglige ressourcer alt efter teksttype og kontekst. Det giver derfor ikke nødvendigvis mening at betragte skriveudvikling på mellemtrin og udskoling som bestående af generiske udviklingstrin. Dette giver i højere grad mening

ved begynderskrivningen, hvor eleverne stadig er ved at lære at mestre mere grundlæggende syntaks og ortografi. Derfor bør fremtidige lingvistiske studier af skriveudvikling tage højde for, at elevtekster, når de når forbi begynderstadiet, i forskellig udstrækning er funktionelle tekster, som kun afspejler en del af elevens sproglige repertoire. Der er derfor brug for både store korpusbaserede studier af udvikling inden for bestemte teksttyper samt mere longitudinale studier af enkelte elevskrivere, hvor man kan følge udviklingen af enkelte elevers skrivefærdigheder på tværs af teksttyper. Derigennem får man blik for, hvordan eleverne tilpasser deres sproglige ressourcer til de enkelte teksttyper, skriveordrer og skrivekontekster.

Det andet område, jeg vil fremhæve, er knyttet til begrebet 'stemme' som et vigtigt analytisk fokuspunkt i studier af skriveudvikling. Jeg har tidligere i afsnit 5.8.1.4 beskrevet, hvordan de kvalitative analyser viste, at elevernes stemme var et gennemgående træk ved flere af elevernes tekster. Tidligere forskning i stemme har enten fokuseret på akademisk skrivning på videregående uddannelser eller fokuseret mere fragmenteret på forskellige lingvistiske dimensioner knyttet til elevernes stemme eller stillingtagen i forskellige teksttyper, fx brugen af imperativer (Neiderhiser et al., 2016) eller bestemte adverbier, enkeltord og fraser (Aull, 2019). Der er imidlertid brug for mere samlede lingvistiske beskrivelser af, hvad der kendetegner elevernes stemme på forskellige klasse- og uddannelsesstrin. Dette vil på mange måder kunne bidrage ikke alene til en dybere forståelse af en vigtig dimension af skriveudviklingen, men det vil også bidrage til en dybere forståelse af de forskelle, der er på at skrive på forskellige niveauer af uddannelsessystemet. I analyserne af sætningslængde og dependensafstand i den diakrone analysedel skete de tydeligste klassetrinsrelaterede forskelle fx i overgangen fra mellemtrin til udskoling, og forskning i skrivekulturer har fx vist, at overgangen fra udskoling til gymnasiet også markerer en overgang fra en personlig og erfaringsbaseret stemme til en mere analytisk og reflekterende stemme (Christensen et al., 2014; Krogh & Jakobsen, 2016). Det vil således være interessant at undersøge fra et lingvistisk perspektiv, hvorvidt og hvordan overgangene mellem mellemtrin, udskoling og gymnasiet manifesterer sig i forskellige typer af stemme i elevernes tekster.

Det tredje og sidste område, jeg vil fremhæve, angår behovet for mere viden om synkron variable betydning for elevernes skriveudvikling. Jeg har i denne afhandling vist, at særligt de askriptive variable har betydning for elevernes tekster, men også at elevernes

generelle skrivelyst og til dels deres læse- og skrivevaner har betydning for forskellige sproglige elementer i teksterne. Dette studie har imidlertid været begrænset af et forholdsvis lille korpus fra bare én skole. Der er derfor behov for flere og større studier, som kan undersøge, om afhandlingens analyseresultater også kan overføres til et større korpus. Forholdet mellem skrivepræstationer og skrivemotivation er vigtigt at belyse, fordi skrivemotivation er en vigtig forudsætning for, at eleverne engagerer sig i de skriveprocesser, der sættes i gang i undervisningen, og der er derfor brug for mere viden, der kan belyse, hvordan forskellige former for skrivemotivation påvirker elevernes evne til at komponere og sammensætte tekster (Wright et al., 2021). Viden om læse- og skrivevaners betydning er også vigtig at belyse yderligere, eftersom området er forskningsmæssigt underbelyst, og der er - trods de begrænsede resultater i dette studie – teoretisk belæg for at antage, at disse dimensioner bør have en stor betydning for elevernes skrivefærdigheder. Den sociologiske skriveforskning er også meget lille, og der er især meget få studier, der kobler et sociologisk perspektiv med et lingvistisk og tekstorienteret perspektiv. Det sociologiske perspektiv bidrager imidlertid – som det forhåbentlig har fremgået af denne afhandlings analyser – med vigtige perspektiver på elevernes skriveudvikling, idet de dels kan være med til at identificere og tydeliggøre forskelle, som i et større ligheds- og demokratisk perspektiv er problematiske, dels kan identificere områder som i et differentieringsperspektiv kræver yderligere didaktisk opmærksomhed. Der er derfor behov for flere studier, som undersøger, hvordan disse askriptive faktorer påvirker og interagerer med elevernes skrivefærdigheder.

Referencer

- Abbott, R. D., & Berninger, V. W. (1993). Structural Equation Modeling of Relationships Among Developmental Skills and Writing Skills in Primary- and Intermediate-Grade Writers. *Journal of Educational Psychology, 85*(3), 478–508. <https://doi.org/10.1037/0022-0663.85.3.478>
- Agersnap, A., & Johansen, K. H. (2021). From Linguistic Features to Cultural Patterns: Perspectives from Large-Scale Text Studies of Danish Sermons. *Scandinavian Studies in Language, 12*(1), 29–49. <https://doi.org/10.7146/sss.v12i1.130067>
- Allison, P., Beard, R., & Willcocks, J. (2002). Subordination in Children's Writing. *Language and Education, 16*(2), 97–111. <https://doi.org/10.1080/09500780208666822>
- Allison, P. D. (1999). *Multiple Regression: A Primer*. Pine Forge Press.
- Allwood, J. (2008). Multimodal corpora. I A. Lüdeling & M. Kytö (Red.), *Corpus Linguistics. An International Handbook* (s. 207–225). Mouton de Gruyter.
- Andersson, H. (2002). *Svenska som första- och andraspråk: En jämförande studie av texter från skolår 9*. Uppsala Universitet.
- Apelgren, B.-M., & Holmberg, P. (2020). Upper secondary students' discursive writing in two languages. *Writing and Pedagogy, 12*(1), 47–71. <https://doi.org/10.1558/wap.36367>
- Aram, D., Korat, O., & Hassunah-Arafat, S. (2013). The contribution of early home literacy activities to first grade reading and writing achievements in Arabic. *Reading and Writing, 26*(9), 1517–1536. <https://doi.org/10.1007/s11145-013-9430-y>
- Arkes, J. (2019). *Regression Analysis: A Practical Introduction*. Routledge. <https://doi.org/10.4324/9781351011099>
- Arnbak, E., & Petersen, D. K. (2017). *Projekt it og ordblindhed: En undersøgelse af it-støtte til ordblinde elever på mellemtrinnet*. DPU, Aarhus Universitet.
- Atkins, S., Clear, J., & Ostler, N. (1992). Corpus design criteria. *Literary and Linguistic Computing, 7*(1), 1–16. <https://doi.org/10.1093/lc/7.1.1>
- Attali, Y., & Powers, D. (2008). A Developmental Writing Scale. *ETS Research Report Series, 1*, 1–59. <https://doi.org/10.1002/j.2333-8504.2008.tb02105.x>
- Aull, L. (2019). Linguistic Markers of Stance and Genre in Upper-Level Student Writing. *Written Communication, 36*(2), 267–295. <https://doi.org/10.1177/0741088318819472>
- Baker, N. L. (2014). “Get it off my stack”: Teachers' tools for grading papers. *Assessing Writing, 19*, 36–50. <https://doi.org/10.1016/j.asw.2013.11.005>
- Baron, D. (2009). *A Better Pencil: Readers, Writers, and the Digital Revolution*. Oxford University Press.
- Bartholomae, D. (1980). The Study of Error. *College Composition and Communication, 31*(3), 253–269. <https://doi.org/10.2307/356486>
- Barton, D. (2007). *Literacy: An introduction to the ecology of written language* (2. udg.). Blackwell Pub.
- Barton, D., & Hamilton, M. (1998). *Local literacies: Reading and writing in one community*. Routledge.
- Bazerman, C. (2013). Understanding the lifelong journey of writing development. *Journal for the Study of Education and Development, 36*(4), 421–441. <https://doi.org/10.1174/021037013808200320>

- Bazerman, C., Applebee, A. N., Berninger, V. W., Brandt, D., Graham, S., Jeffery, J. V., Matsuda, P., Murphy, S., Rowe, D. W., Schleppegrell, M., & Wilcox, K. C. (2018). Towards an Understanding of Writing Development across the Lifespan. I *The Lifespan Development of Writing* (s. 20–51). National Council of Teachers of English.
- Beard, R., & Burrell, A. (2010). Investigating narrative writing by 9–11-year-olds. *Journal of Research in Reading, 33*(1), 77–93. <https://doi.org/10.1111/j.1467-9817.2009.01433.x>
- Beard, R., Burrell, A., & Homer, M. (2016). Investigating persuasive writing by 9–11 year olds. *Language and Education, 30*(5), 417–437. <https://doi.org/10.1080/09500782.2015.1133638>
- Beers, S. F., & Nagy, W. E. (2011). Writing development in four genres from grades three to seven: Syntactic complexity and genre differentiation. *Reading and Writing, 24*(2), 183–202. <https://doi.org/10.1007/s11145-010-9264-9>
- Bellah, R. (2011). The Axial Age I: Introduction and Ancient Israel. I *Religion in Human Evolution* (s. 265–323). The Belknap Press of Harvard University Press.
- Berge, K. L. (1988). *Skolestilen som genre: Med påtvungen penn*. Landslaget for norskundervisning LNU. Cappelen.
- Berge, K. L. (2005). Tekstkulturer og tekstkvalitet. I K. L. Berge, L. S. Evensen, F. Hertzberg, & W. Vagle (Red.), *Ungdommers skrivekompetanse. Bind II: Norskeksamen som tekst* (s. 11–190). Universitetsforlaget.
- Berge, K. L. (2019). Skriftekulturforskning i Norden. Nokre overordna perspektiv på forskningsfeltet og forskningsresultat på 2000-tallet. I S. J. Helset & E. Brunstad (Red.), *Skriftekulturstadiar i ei brytningstid* (s. 21–51). Cappelen Damm Akademisk. <https://doi.org/10.23865/noasp.67>
- Berge, K. L., Evensen, L. S., Hertzberg, F., & Vagle, W. (2005a). *Ungdommers skrivekompetanse. Bind I: Norskensuren som kvalitetsvurdering*. Universitetsforlaget.
- Berge, K. L., Evensen, L. S., Hertzberg, F., & Vagle, W. (2005b). *Ungdommers skrivekompetanse. Bind II: Norskeksamen som tekst*. Universitetsforlaget.
- Berge, K. L., Evensen, L. S., & Thygesen, R. (2016). The Wheel of Writing: A model of the writing domain for the teaching and assessing of writing as a key competency. *The Curriculum Journal, 27*(2), 172–189. <https://doi.org/10.1080/09585176.2015.1129980>
- Berge, K. L., Skar, G. B., Matre, S., Solheim, R., Evensen, L. S., Otnes, H., & Thygesen, R. (2019). Introducing teachers to new semiotic tools for writing instruction and writing assessment: Consequences for students' writing proficiency. *Assessment in Education: Principles, Policy & Practice, 26*(1), 6–25. <https://doi.org/10.1080/0969594X.2017.1330251>
- Berman, R., & Verhoeven, L. (2002). Cross-linguistic perspectives on the development of text-production abilities: Speech and writing. *Written Language & Literacy, 5*(1), 1–43. <https://doi.org/10.1075/wll.5.1.02ber>
- Bernhardt, L. (2006). Kohærens og kohæsjon i hypertekster. Om sammenheng i internettekster. *Hermes - Journal of Language and Communication, 36*, 157–180. <https://doi.org/10.7146/hjlcb.v19i36.25843>
- Berninger, V. W., Nagy, W., & Beers, S. (2011). Child writers' construction and reconstruction of single sentences and construction of multi-sentence texts: Contributions of syntax and transcription to translation. *Reading and Writing, 24*(2), 151–182. <https://doi.org/10.1007/s11145-010-9262-y>
- Bernstein, B. (1971). *Class, Codes and Control. Volume 1: Theoretical Studies towards a Sociology of Language*. Routledge & Kegan Paul.
- Bernstein, B. (1975). *Class, Codes and Control. Volume 3: Towards a Theory of Educational Transmissions*. Routledge & Kegan Paul.

REFERENCER

- Berry, D. (2012). *Understanding Digital Humanities*. Palgrave Macmillan.
- Berthelsen, U. D. (2021). Digitale tekster og skriftlig fremstilling i gymnasiet—Et curriculumperspektiv. *Tidsskriftet Læring og Medier (LOM)*, 13(23), 1–17. <https://doi.org/10.7146/lom.v13i23.120963>
- Berthelsen, U. D., & Tannert, M. (2021). Introduction: The Role of Computational Methods within the Humanities and Social Sciences. *Scandinavian Studies in Language*, 12(1), i–iv. <https://doi.org/10.7146/sss.v12i1.130065>
- Biber, D. (1993). Representativeness in Corpus Design. *Literary and Linguistic Computing*, 8(4), 243–257. <https://doi.org/10.1093/lc/8.4.243>
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge University Press.
- Biber, D., & Egbert, J. (2018). *Register Variation Online*. Cambridge University Press. <https://doi.org/10.1017/9781316388228>
- Biber, D., Gray, B., Staples, S., & Egbert, J. (2020). Investigating grammatical complexity in L2 English writing research: Linguistic description versus predictive measurement. *Journal of English for Academic Purposes*, 46, 1–15. <https://doi.org/10.1016/j.jeap.2020.100869>
- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Pearson Education.
- Björnsson, C. H. (1960). *Uppsatsbedömning och uppsatsskrivning*. Almqvist & Wiksell.
- Blasius, J., & Thiessen, V. (2012). *Assessing the Quality of Survey Data*. SAGE Publications.
- Blikstad-Balas, M., Roe, A., & Klette, K. (2018). Opportunities to Write: An Exploration of Student Writing During Language Arts Lessons in Norwegian Lower Secondary Classrooms. *Written Communication*, 35(2), 119–154. <https://doi.org/10.1177/0741088317751123>
- Bloomfield, L. (1933). *Language*. George Allen & Unwin Ltd.
- Borgström, E., & Ledin, P. (2014). Bedömarvariation. Balansen mellan teknisk och hermeneutisk rationalitet vid bedömning av skrivprov. *Språk & stil*, 24, 133–165.
- Bouwer, R., Béguin, A., Sanders, T., & van den Bergh, H. (2015). Effect of genre on the generalizability of writing scores. *Language Testing*, 32(1), 83–100. <https://doi.org/10.1177/0265532214542994>
- Brandt, D. (2014). *The Rise of Writing: Redefining Mass Literacy*. Cambridge University Press. <https://doi.org/10.1017/CBO9781316106372>
- Bremholm, J., Bundsgaard, J., & Kabel, K. (2022). Proficiency scales for early writing development. *Writing and Pedagogy*, 13(1–3), 121–154. <https://doi.org/10.1558/wap.21490>
- Bremholm, J., Hansen, R., & Slot, M. F. (2018). Student work and the multimodal challenge: A mixed methods study of students' productive work in L1, science and mathematics. *L1 - Educational Studies in Language and Literature*, 18, 1–27. <https://doi.org/10.17239/L1ESLL-2018.18.01.12>
- Bronfenbrenner, U. (1976). The Experimental Ecology of Education. *Educational Researcher*, 5(9), 5–15. <https://doi.org/10.3102/0013189X005009005>
- Buch, B. (2015). *Tekster og genrer på erhversuddannelserne—Med særligt henblik på overgangen fra grundskole til erhversuddannelse*. Ph.d.-afhandling. DPU, Aarhus Universitet.
- Bueie, A. A. (2016). Nyttige og mindre nyttige lærercommentarer – slik elevene ser det. *Nordic Journal of Literacy Research*, 2(1), 1–28. <https://doi.org/10.17585/njlr.v2.188>

- Bukodi, E., Erikson, R., & Goldthorpe, J. H. (2014). The effects of social origins and cognitive ability on educational attainment: Evidence from Britain and Sweden. *Acta Sociologica*, 57(4), 293–310. <https://doi.org/10.1177/0001699314543803>
- Bundsgaard, J., Kabel, K., & Bremholm, J. (2022a). Skriveudvikling i de tidlige skoleår. *Læsepædagog*, 3, 23–30.
- Bundsgaard, J., Kabel, K., & Bremholm, J. (2022b). Validating scales for the early development of writing proficiency. *Writing and Pedagogy*, 13(1–3), 89–120. <https://doi.org/10.1558/wap.21491>
- Camacho, A., Alves, R. A., De Smedt, F., Van Keer, H., & Boscolo, P. (2021). Relations among motivation, behaviour, and performance in writing: A multiple-group structural equation modeling study. *British Journal of Educational Psychology*, 91(4), e12430. <https://doi.org/10.1111/bjep.12430>
- Camp, H. (2012). The psychology of writing development—And its implications for assessment. *Assessing Writing*, 17(2), 92–105. <https://doi.org/10.1016/j.asw.2012.01.002>
- Carter, M. (2007). Ways of Knowing, Doing, and Writing in the Disciplines. *College Composition and Communication*, 58(3), 385–418.
- Chakravartty, A. (2003). The Structuralist Conception of Objects. *Philosophy of Science*, 70(5), 867–878. <https://doi.org/10.1086/377373>
- Chen, B., Deng, D., Zhong, Z., & Zhang, C. (2020). Exploring linguistic characteristics of highly browsed and downloaded academic articles. *Scientometrics*, 122(3), 1769–1790. <https://doi.org/10.1007/s11192-020-03361-4>
- Chomsky, N. (1965). *Aspects of the theory of syntax*. MIT Press.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. Praeger.
- Christensen, L. H., & Christensen, R. Z. (2019). *Dansk grammatik* (4. udg.). Syddansk Universitetsforlag.
- Christensen, T. S., Elf, N. F., & Krogh, E. (2014). *Skrivekulturer i folkeskolens niende klasse*. Syddansk Universitetsforlag.
- Christensen, V. T. (2019). *PISA 2018—Danske unge i en international sammenligning*. VIVE - Det Nationale Forsknings- og Analysecenter for Velfærd.
- Christie, F. (2002). The Development of Abstraction in Adolescence in Subject English. I M. J. Schleppegrell, M. C. Colombi, & J. Baugh (Red.), *Developing Advanced Literacy in First and Second Languages: Meaning with Power* (s. 45–66). Taylor & Francis Group.
- Christie, F., & Derewianka, B. (2010). *School discourse: Learning to write across the years of schooling*. Continuum.
- Chrystal, J.-A., & Ekvall, U. (1999). Planering och revidering i skolskrivande. I L.-G. Andersson (Red.), *Svenskans beskrivning 23* (s. 57–76). Lund University Press.
- Clark, C., Picton, I., & Lant, F. (2020). “More time on my hands”: Children and young people’s writing during the COVID-19 lockdown in 2020. National Literacy Trust.
- Cobb, T., & Boulton, A. (2015). Classroom applications of corpus analysis. I D. Biber & R. Reppen (Red.), *The Cambridge Handbook of English Corpus Linguistics* (s. 478–497). Cambridge University Press. <https://doi.org/10.1017/CBO9781139764377.027>
- Cobb, T., & Horst, M. (2015). Learner corpora and lexis. I F. Meunier, G. Gilquin, & S. Granger (Red.), *The Cambridge Handbook of Learner Corpus Research* (s. 185–206). Cambridge University Press. <https://doi.org/10.1017/CBO9781139649414.009>
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research Methods in Education* (8. udgave). Routledge.

REFERENCER

- Coker Jr., D. L. (2012). Descriptive Writing. I E. Grigorenko, E. Mambrino, & D. D. Preiss (Red.), *Writing: A Mosaic of New Perspectives* (s. 159–172). Psychology Press.
- Cope, B., & Kalantzis, M. (2000). *Multiliteracies: Literacy learning and the design of social futures*. Routledge.
- Cope, B., & Kalantzis, M. (2009). “Multiliteracies”: New Literacies, New Learning. *Pedagogies: An International Journal*, 4, 164–195. <https://doi.org/10.1080/15544800903076044>
- Correnti, R., Matsumura, L. C., Wang, E., Litman, D., Rahimi, Z., & Kisa, Z. (2020). Automated Scoring of Students’ Use of Text Evidence in Writing. *Reading Research Quarterly*, 55(3), 493–520. <https://doi.org/10.1002/rrq.281>
- Cramer, J., Kunøe, M., Larsen, E. V., & Tøgeby, O. (1996). 699 varmer termer. *Leksikon til sprogvitenskap*. Institut for Nordisk Sprog og Litteratur, Aarhus Universitet.
- Creamer, E. G. (2018). Paradigms in Play: Using Case Studies to explore the Value-Added of Divergent Findings in Mixed Methods Research. *International Journal of Multiple Research Approaches*, 10(1), 30–40. <https://doi.org/10.29034/ijmra.v10n1a2>
- Cresswell, J., & Clark, V. L. P. (2011). Choosing a Mixed Methods Design. I *Designing and conducting mixed methods research* (2. udg., s. 53–102). SAGE.
- Crossley, S. A. (2020). Linguistic features in writing quality and development: An overview. *Journal of Writing Research*, 11(3), 415–443. <https://doi.org/10.17239/jowr-2020.11.03.01>
- Crossley, S. A., & Kim, M. (2022). Linguistic Features of Writing Quality and Development: A Longitudinal Approach. *Journal of Writing Analytics*, 6, 59–93. <https://doi.org/10.37514/JWA-J.2022.6.1.04>
- Crossley, S. A., & Kyle, K. (2018). Assessing writing with the tool for the automatic analysis of lexical sophistication (TAALES). *Assessing Writing*, 38, 46–50. <https://doi.org/10.1016/j.asw.2018.06.004>
- Crossley, S. A., Kyle, K., & McNamara, D. S. (2015). To aggregate or not? Linguistic features in automatic essay scoring and feedback systems. *Journal of Writing Assessment*, 8(1).
- Crossley, S. A., & McNamara, D. S. (2010). Cohesion, coherence, and expert evaluations of writing proficiency. I S. Ohlsson & R. Catrambone (Red.), *Proceedings of the 32nd Annual Conference of the Cognitive Science Society* (s. 984–989). Cognitive Science Society.
- Crossley, S. A., Roscoe, R., & McNamara, D. S. (2014). What Is Successful Writing? An Investigation Into the Multiple Ways Writers Can Write Successful Essays. *Written Communication*, 31(2), 184–214. <https://doi.org/10.1177/0741088314526354>
- Crossley, S. A., Weston, J. L., Sullivan, S. T. M., & McNamara, D. S. (2011). The Development of Writing Proficiency as a Function of Grade Level: A Linguistic Analysis. *Written Communication*, 28(3), 282–311. <https://doi.org/10.1177/0741088311410188>
- Crystal, D. (1992). *An encyclopedic dictionary of language and languages*. Basil Blackwell.
- Danmarks Statistik. (2022, maj). *Befolkningens højest fuldførte uddannelse*. <https://www.dst.dk/da/Statistik/emner/uddannelse-og-forskning/befolkningens-uddannelsesstatus/befolkningens-hoejst-fuldfoerte-uddannelse>
- Dansk Sprognævn. (2012). *Retskrivningsordbogen* (4. udgave). Alinea.
- Davidse, K., & De Smet, H. (2020). Diachronic Corpora. I M. Paquot & S. Th. Gries (Red.), *A Practical Handbook of Corpus Linguistics* (s. 211–233). Springer International Publishing. https://doi.org/10.1007/978-3-030-46216-1_10
- De Costa, P., Paltridge, B., & Phakiti, A. (2015). Ethics and Applied Linguistics Research. I *Research methods in applied linguistics: A practical resource* (s. 245–257). Bloomsbury.

- Deane, P. (2013). On the relation between automated essay scoring and modern views of the writing construct. *Assessing Writing*, 18(1), 7–24. <https://doi.org/10.1016/j.asw.2012.10.002>
- Deane, P. (2018). The Challenges of Writing in School: Conceptualizing Writing Development Within a Sociocognitive Framework. *Educational Psychologist*, 53(4), 280–300. <https://doi.org/10.1080/00461520.2018.1513844>
- DeKeyser, R. M. (2007). Introduction: Situating the concept of practice. I R. M. DeKeyser (Red.), *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology* (s. 1–18). Cambridge University Press.
- Dekking, F. M., Kraaikamp, C., Lopuhaä, H. P., & Meester, L. E. (2005). Exploratory data analysis: Numerical summaries. I *A Modern Introduction to Probability and Statistics* (s. 231–243). Springer.
- Derewianka, B. (2007). Using appraisal theory to track interpersonal development in adolescent academic writing. I A. McCabe, M. O’Donnell, & R. Whittaker (Red.), *Advances in Language and Education* (s. 142–165). Continuum.
- Diderichsen, P. (1946). *Elementær dansk grammatik*. Gyldendal.
- Dockrell, J. E., Connelly, V., Walter, K., & Critten, S. (2015). Assessing children’s writing products: The role of curriculum based measures. *British Educational Research Journal*, 41(4), 575–595. <https://doi.org/10.1002/berj.3162>
- Dockrell, J. E., Lindsay, G., & Connelly, V. (2009). The Impact of Specific Language Impairment on Adolescents’ Written Text. *Exceptional Children*, 75(4), 427–446. <https://doi.org/10.1177/001440290907500403>
- Dolin, J., Nielsen, K., & Rangvid, B. S. (2018). *Rapport fra Følgegruppen for én bedømmer ved folkeskolens prøver*. Undervisningsministeriet. <https://www.uvm.dk/publikationer/folkeskolen/2018-foelgegruppen-for-e-n-bedoemmer-ved-folkeskolens-proever>
- Donahue, C., & Lillis, T. (2014). Models of writing and text production. I E.-M. Jakobs & D. Perin (Red.), *Handbook of Writing and Text Production* (s. 55–78). De Gruyter.
- Donald, M. (1991). *Origins of the Modern Mind: Three Stages in the Evolution of Culture and Cognition*. Harvard University Press.
- Dourish, P., & Cruz, E. G. (2018). Datafication and data fiction: Narrating data and narrating with data. *Big Data & Society*, 5(2), 1–10. <https://doi.org/10.1177/2053951718784083>
- Downey, D. B., Kuhfeld, M., & van Hek, M. (2022). Schools as a Relatively Standardizing Institution: The Case of Gender Gaps in Cognitive Skills. *Sociology of Education*, 95(2), 89–109. <https://doi.org/10.1177/00380407211070319>
- Durrant, P., Brenchley, M., & Clarkson, R. (2020). Syntactic development across genres in children’s writing: The case of adverbial clauses. *Journal of Writing Research*, 12(2), 419–452. <https://doi.org/10.17239/jowr-2020.12.02.04>
- Durrant, P., Brenchley, M., & McCallum, L. (2021). *Understanding Development and Proficiency in Writing: Quantitative Corpus Linguistic Approaches*. Cambridge University Press. <https://doi.org/10.1017/9781108770101>
- Dämmrich, J., & Triventi, M. (2018). The dynamics of social inequalities in cognitive-related competencies along the early life course – A comparative study. *International Journal of Educational Research*, 88, 73–84. <https://doi.org/10.1016/j.ijer.2018.01.006>
- Eccles, J. (1983). Expectancies, Values and Academic Behaviors. I J. T. Spence (Red.), *Achievement and Achievement Motives: Psychological and Sociological Approaches* (s. 78–146). W. H. Freeman and Company.
- Eisenstein, J. (2019). *Introduction to natural language processing*. The MIT Press.

REFERENCER

- Ekholm, E., Zumbrunn, S., & DeBusk-Lane, M. (2018). Clarifying an Elusive Construct: A Systematic Review of Writing Attitudes. *Educational Psychology Review*, 30, 827–856. <https://doi.org/10.1007/s10648-017-9423-5>
- Elf, N., & Troelsen, S. (2021). Between joyride and high-stakes examination: Writing development in Denmark. I J. V. Jeffery & J. M. Parr (Red.), *International Perspectives on Writing Curricula and Development* (s. 169–191). Routledge. <https://doi.org/10.4324/9781003051404-9>
- Elliot, N. (2016). A Theory of Ethics for Writing Assessment. *Journal of Writing Assessment*, 9(1).
- Elwert, F. (2013). Graphical Causal Models. I S. L. Morgan (Red.), *Handbook of Causal Analysis for Social Research* (s. 245–273). Springer. https://doi.org/10.1007/978-94-007-6094-3_13
- Enevoldsen, K., Hansen, L., & Nielbo, K. L. (2021). DaCy: A unified framework for Danish NLP. I M. Ehrman, F. Karsdorp, M. Wevers, T. L. Andrews, M. Burghardt, M. Kestemont, E. Manjavacas, Piotrowski, & J. van Zundert (Red.), *Proceedings of the Conference on Computational Humanities Research 2021* (s. 206–216). CEUR Workshop Proceedings.
- Engblom, C. (2016). Prefabricated Images in Young Children's Text-Making at School. *Designs for Learning*, 8(1), 37–47. <https://doi.org/10.16993/dfl.70>
- Engmose, S. (2019). *(IT)-støttet børnestavning: Studier af børnestavnings rolle i den tidlige skriftsproglige udvikling*. Ph.d.-afhandling. Københavns Universitet.
- Erbeli, F., Hart, S. A., Kim, Y.-S. G., & Taylor, J. (2017). The effects of genetic and environmental factors on writing development. *Learning and Individual Differences*, 59, 11–21. <https://doi.org/10.1016/j.lindif.2017.08.005>
- European Parliament. (2016). Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation). *Official Journal of the European Union*, 1–119.
- Evensen, L. S. (1987). Elevtekster og elevers tekststrategier. *Trondheimsskrifter i anvendt språkvitenskap* 3, 55–83.
- Evensen, L. S. (1988). Lokal og global koherens i elevers skrivning. *Trondheimsskrifter i anvendt språkvitenskap* 4, 31–50.
- Evensen, L. S. (1990). *What do the structures mean?: Developing argumentative student writing*. Senter for samfunnsforskning, Sesam, Universitetet i Trondheim.
- Evensen, L. S. (1991). *Utvikling av skriftspråklig kompetanse: Forskningsbakgrunn og kunnskapsutfordringer* (2. rev. utg.). ALLFORSK.
- Evensen, L. S. (2005). Perspektiv på innhold? Relieff i ungdomsskoleelevers eksamensskrivning. I K. L. Berge, L. S. Evensen, F. Hertzberg, & W. Vagle (Red.), *Ungdommers skrivekompetanse. Bind II: Norskeksamen som tekst* (s. 191–236). Universitetsforlaget.
- Falkenjack, J., Mühlenbock, K. H., & Jönsson, A. (2013). Features indicating readability in Swedish text. I S. Oepen, K. Hagen, & J. B. Johannessen (Red.), *Proceedings of the 19th Nordic Conference of Computational Linguistics (NODALIDA 2013)* (s. 27–40). Linköping University Electronic Press.
- Fellbaum, C. (1998). *WordNet: An electronic lexical database*. MIT Press.
- Finnemann, N. O. (2014). Digital humanities and networked digital media. *MedieKultur: Journal of Media and Communication Research*, 30(57), 94–114. <https://doi.org/10.7146/mediekultur.v30i57.15592>

- Fischer, S. R. (2001). *A history of writing*. Reaktion Books.
- Fish, S. (2018a). If You Count It, They Will Come. *New York University Journal of Law and Liberty*, 12(2), 333–351.
- Fish, S. (2018b). The Interpretive Poverty of Data. *Balkinization*. <https://balkin.blogspot.com/2018/03/the-interpretive-poverty-of-data.html>
- Fleckenstein, J. J., Meyer, J. J., Jansen, T. T., Keller, S. S., & Köller, O. O. (2020). Is a Long Essay Always a Good Essay? The Effect of Text Length on Writing Assessment. *Frontiers in Psychology*, 11, 1–10. <https://doi.org/10.3389/fpsyg.2020.562462>
- Flower, L. (1979). Writer-Based Prose: A Cognitive Basis for Problems in Writing. *College English*, 41(1), 19–37. JSTOR. <https://doi.org/10.2307/376357>
- Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4), 365–387. <https://doi.org/10.2307/356600>
- Flowerdew, L. (2022). Using corpora for writing instruction. I A. O’Keeffe & M. McCarthy (Red.), *The Routledge Handbook of Corpus Linguistics* (2. udg., s. 443–455). Routledge. <https://doi.org/10.4324/9780367076399-31>
- Flyvbjerg, B. (2006). Five Misunderstandings About Case-Study Research. *Qualitative Inquiry*, 12(2), 219–245. <https://doi.org/10.1177/1077800405284363>
- Fogal, G. G. (2020). Investigating Variability in L2 Development: Extending a Complexity Theory Perspective on L2 Writing Studies and Authorial Voice. *Applied Linguistics*, 41(4), 575–600. <https://doi.org/10.1093/applin/amz005>
- Folkeryd, J. W. (2006). *Writing with an Attitude: Appraisal and student texts in the school subject of Swedish*. Ph.d.-afhandling. Uppsala Universitet. <https://www.diva-portal.org/smash/get/diva2:169419/FULLTEXT01.pdf>
- Frederiksen, M. (2013). Integration i “mixed methods” forskning: Metode eller design? *Metode & Forskningsdesign*, 1(1), 17–40.
- Gardner, S., Nesi, H., & Biber, D. (2019). Discipline, Level, Genre: Integrating Situational Perspectives in a New MD Analysis of University Student Writing. *Applied Linguistics*, 40(4), 646–674. <https://doi.org/10.1093/applin/amy005>
- Garme, B. (1988). *Text och tanke: Om skrivstrategier i elevuppsatser*. Liber.
- Gastaldi, J. L. (2021). Why Can Computers Understand Natural Language? *Philosophy & Technology*, 34(1), 149–214. <https://doi.org/10.1007/s13347-020-00393-9>
- Gavin, M. (2020). Is there a text in my data? (Part 1): On counting words. *Journal of Cultural Analytics*. <https://doi.org/10.22148/001c.11830>
- Gee, J. P. (1999). The Future of the Social Turn: Social Minds and the New Capitalism. *Research on Language and Social Interaction*, 32(1–2), 61–68. <https://doi.org/10.1080/08351813.1999.9683608>
- Gee, J. P. (2007). Semiotic domains: Is playing video games a “waste of time”? I *What video games have to teach us about learning and literacy* (revised and updated version, s. 17–43). Palgrave Macmillan.
- Gee, J. P. (2015). *Social Linguistics and Literacies: Ideology in Discourses* (5. udg.). Routledge.
- Geva, E. (2007). Conjunction Use in School Children’s Oral Language and Reading. I R. Horowitz (Red.), *Talking Texts: How Speech and Writing Interact in School Learning* (s. 271–294). Lawrence Erlbaum Associates.
- Gouveia, C. (2013a). Evolving in confidence: Writing across basic schooling. I VIAN JR., Orlando e Cida Caltabiano (eds.), *Língua(gem) e suas múltiplas faces – Estudos em Homenagem a Leila Barbara*. São Paulo: Mercado de Letras: 93-108.

REFERENCER

- Gouveia, C. (2013b). L1 writing development and the criteria to assess it: Insights from a portuguese pilot project. I «A Scholar for all Seasons». *Homenagem a João Almeida Flor*. (s. 209–221). Departamento de Estudos Anglísticos/Centro de Estudos Anglísticos da Universidade de Lisboa.
- Graham, A., Powell, M. A., & Taylor, N. (2015). Ethical Research Involving Children: Encouraging Reflexive Engagement in Research with Children and Young People. *Children & Society*, 29(5), 331–343. <https://doi.org/10.1111/chso.12089>
- Graham, S. (2018a). A Revised Writer(s)-Within-Community Model of Writing. *Educational Psychologist*, 53(4), 258–279. <https://doi.org/10.1080/00461520.2018.1481406>
- Graham, S. (2018b). A Writer(s)-within-Community Model of Writing. I C. Bazerman, A. N. Applebee, V. W. Berninger, D. Brandt, S. Graham, J. V. Jeffery, P. Matsuda, S. Murphy, D. W. Rowe, M. Schleppegrell, & K. C. Wilcox (Red.), *The Lifespan Development of Writing* (s. 272–325). National Council of Teachers of English.
- Graham, S., Hebert, M., & Harris, K. R. (2015). Formative Assessment and Writing: A Meta-Analysis. *The Elementary School Journal*, 115(4), 523–547. <https://doi.org/10.1086/681947>
- Graham, S., Kihara, S. A., Harris, K. R., & Fishman, E. J. (2017). The Relationship among Strategic Writing Behavior, Writing Motivation, and Writing Performance with Young, Developing Writers. *The Elementary School Journal*, 118(1), 82–104. <https://doi.org/10.1086/693009>
- Graham, S., Liu, X., Bartlett, B., Ng, C., Harris, K. R., Aitken, A., Barkel, A., Kavanaugh, C., & Talukdar, J. (2018). Reading for Writing: A Meta-Analysis of the Impact of Reading Interventions on Writing. *Review of Educational Research*, 88(2), 243–284. <https://doi.org/10.3102/0034654317746927>
- Graham, S., McKeown, D., Kihara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology*, 104(4), 879–896. <https://doi.org/10.1037/a0029185>
- Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal*, 26(2), 91–108. <https://doi.org/10.1111/j.1471-1842.2009.00848.x>
- Grice, P. (1989). *Studies in the Way of Words*. Harvard University Press.
- Gries, S. Th. (2009). What is Corpus Linguistics? *Language and Linguistics Compass*, 3(5), 1225–1241. <https://doi.org/10.1111/j.1749-818X.2009.00149.x>
- Grundin, H. U. (1975). *Läs- och skrivförmågans utveckling genom skolåren*. Liber Läromedel.
- Grzybek, P. (2015). Word length. I J. R. Taylor (Red.), *The Oxford Handbook of the Word* (s. 89–119). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199641604.013.37>
- Gustafsson, A. M., & Lassus, J. (2020). Ordförrådet i starka och svaga essäprovssvar. I M. Forskåhl, A. M. Gustafsson, J. Hellgren, J. Lassus, B. Silén, S. Stolt, & A. Westerlund (Red.), *Modersmålsprovet i brännpunkten. Studentexamensprovet i modersmålet svenska på 2000-talet* (Bd. 848, s. 163–196). Svenska Litteratursällskapet.
- Hacker, D. J. (2018). A Metacognitive Model of Writing: An Update From a Developmental Perspective. *Educational Psychologist*, 53(4), 220–237. <https://doi.org/10.1080/00461520.2018.1480373>
- Halliday, M. A. K. (1989). *Spoken and Written language* (2. udg.). Oxford University Press.
- Halliday, M. A. K. (2013). *Halliday's Introduction to Functional Grammar* (4. udg.). Routledge.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. Routledge.

- Halliday, M. A. K., McIntosh, A., & Stevens, P. (1964). *The Linguistic Sciences and Language Teaching*. Longman.
- Hall-Mills, S., & Apel, K. (2012). Narrative and expository writing of adolescents with language-learning disabilities: A pilot study. *Communication Disorders Quarterly*, 34, 135–143. <https://doi.org/10.1177/1525740112465001>
- Hall-Mills, S., & Apel, K. (2015). Linguistic Feature Development Across Grades and Genre in Elementary Writing. *Language, Speech & Hearing Services in Schools*, 46(3), 242–255.
- Hanck, C., Arnold, M., Gerber, A., & Schmelzer, M. (2021). Threats to Internal Validity of Multiple Regression Analysis. I *Introduction to Econometrics with R*. Universität Duisburg-Essen. <https://www.econometrics-with-r.org/>
- Hansen, E. (1970). Sætningsskema og verbalskemaer. *NyS, Nydanske Sprogstudier*, 2, 116–141. <https://doi.org/10.7146/nys.v2i2.10320>
- Hansen, E., & Heltoft, L. (2011a). *Grammatik over det Danske Sprog. Bind II: Syntaktiske og semantiske helheder*. Det Danske Sprog- og Litteraturselskab & Syddansk Universitetsforlag.
- Hansen, E., & Heltoft, L. (2011b). *Grammatik over det Danske Sprog. Bind III: Sætningen og dens konstruktion*. Det Danske Sprog- og Litteraturselskab & Syddansk Universitetsforlag.
- Hansen, L. (u.å.). *TextDescriptives*. Hentet 19. april 2022, fra <https://github.com/HLasse/Text-Descriptives>
- Hansen, S. R. (2014). *Når børn vælger litteratur: Læsevaneundersøgelse perspektiveret med kognitive analyser*. Ph.d.-afhandling. DPU, Aarhus Universitet.
- Hansen, S. R., Gissel, S. T., & Puck, M. R. (2017). *Børns læsning 2017—En kvantitativ undersøgelse af børns læse- og medievaner i fritiden*. Læremiddel.dk & Tænketanken Fremtidens Biblioteker.
- Hansen, S. R., Hansen, T. I., & Petterson, M. (2022). *Børn og unges læsning 2021*. Tænketanken Fremtidens Biblioteker, UCL - Center for Anvendt Skoleforskning & Aarhus Universitetsforlag.
- Hardy, J. A., & Römer, U. (2013). Revealing disciplinary variation in student writing: A multi-dimensional analysis of the Michigan Corpus of Upper-level Student Papers (MICUSP). *Corpora*, 8(2), 183–207. <https://doi.org/10.3366/cor.2013.0040>
- Harris, R. (1995). *Signs of writing*. Routledge.
- Harris, R. (2001). *Rethinking writing*. Continuum.
- Harris, Z. S. (1954). Distributional Structure. *WORD*, 10(2–3), 146–162. <https://doi.org/10.1080/00437956.1954.11659520>
- Harrison, E., & Rose, D. (2006). *The European Socio-economic Classification: (ESeC) User Guide*. University of Essex.
- Haswell, R. H. (2000). Documenting Improvement in College Writing: A Longitudinal Approach. *Written Communication*, 17(3), 307–352. <https://doi.org/10.1177/0741088300017003001>
- Hayes, J. R. (2011). Kinds of Knowledge-Telling: Modeling Early Writing Development. *Journal of Writing Research*, 3(2), 73–92. <https://doi.org/10.17239/jowr-2011.03.02.1>
- Hetmar, V. (2000). *Elevens projekt, lærerens udfordringer: Om skriveundervisning og skriveudvikling i folkeskolen*. Dansk lærerforening.
- Heydebrand, W. V. (2001). Theories of Structuralism. I N. J. Smelser & P. B. Baltes (Red.), *International Encyclopedia of the Social & Behavioral Sciences* (s. 15230–15233). Pergamon. <https://doi.org/10.1016/BO-08-043076-7/01976-8>

REFERENCER

- Heyder, A., Kessels, U., & Steinmayr, R. (2017). Explaining academic-track boys' underachievement in language grades: Not a lack of aptitude but students' motivational beliefs and parents' perceptions? *British Journal of Educational Psychology*, 87(2), 205–223. <https://doi.org/10.1111/bjep.12145>
- Hirst, G. (2013). Computational Linguistics. I K. Allan (Red.), *The Oxford Handbook of The History of Linguistics* (s. 707–726). Oxford University Press.
- Hobel, P., Krogh, E., & Piekut, A. (2019). Writing and writer development: Pedagogical challenges and perspectives. I E. Krogh & K. S. Jakobsen (Red.), *Understanding Young People's Writing Development* (s. 181–201). Routledge.
- Holm, L. (2006). Forskningsmässige tilgange til literacy. I M. Ellwin (Red.), *Ord och bild ger mening. Om literacy i förskola och skol* (s. 20–34). Svenskläraryöningen.
- Holmberg, P. (2006). Funktionell grammatik för skolans textarbete. I I. Lindberg & K. Sandwall (Red.), *Språket och kunskapen—Att lära på sitt andraspråk i skola och högskola. Rapport från nordisk konferens den 7-8 oktober 2005 i Göteborg* (s. 129–146). Göteborgs Universitet.
- Holmberg, P. (2019). Grammatik i funktionellt perspektiv. I C. Liberg & J. Smidt (Red.), *Att bli lärare i svenska* (s. 103–116). Liber Förlag.
- Holmberg, P., Krogh, E., Nordenstam, A., Penne, S., Skarstein, D., Karlskov Skyggebjerg, A., Tainio, L., & Heilä-Ylikallio, R. (2019). On the emergence of the L1 research field. A comparative study of PhD abstracts in the Nordic countries 2000–2017. *L1-Educational Studies In Language And Literature*, 19, 1–27. <https://doi.org/10.17239/L1ESLL-2019.19.01.05>
- Howe, K. R. (2012). Mixed Methods, Triangulation, and Causal Explanation. *Journal of Mixed Methods Research*, 6(2), 89–96. <https://doi.org/10.1177/1558689812437187>
- Hudson, R. (2009). Measuring Maturity. I R. Beard, D. Myhill, J. Riley, & M. Nystrand (Red.), *The SAGE Handbook of Writing Development* (s. 349–362). SAGE Publications Ltd. <https://doi.org/10.4135/9780857021069>
- Hudson, R. (2018). Dependency, Corpora and Cognition. I J. Jiang & H. Liu (Red.), *Quantitative Analysis of Dependency Structures* (s. 1–30). De Gruyter Mouton.
- Hudson, R. A. (2020). Towards a pedagogical linguistics. *Pedagogical Linguistics*, 1(1), 8–33. <https://doi.org/10.1075/pl.19011.hud>
- Hultman, T. G. (1994). Hur gick det med OVIX? I N. Jörgensson, C. Platzack, & J. Svensson (Red.), *Språkbruk, grammatik och språkförändring: En festskrift till Ulf Teleman* (s. 55–64). Lunds Universitet.
- Hultman, T. G., & Westman, M. (1977). *Gymnasistsvenska*. Liber Läromedel.
- Hundal, A. K. (2017). *Skriveutvikling på mellomtrinnet: En undersøkelse av sammenhengsmekanismer i elevtekster og hvilke didaktiske forhold som bidrar til utvikling av sammenheng i tekst*. Ph.d.-afhandling. NTNU, Norges teknisk-naturvitenskapelige universitet.
- Hunt, K. W. (1965). *Grammatical Structures Written at Three Grade Levels*. NCTE Research Report No. 3.
- Hunt, K. W. (1966). Recent measures in syntactic development. *Elementary English*, 43, 732–739.
- Hunt, K. W. (1970). Syntactic Maturity in Schoolchildren and Adults. *Monographs of the Society for Research in Child Development*, 35(1), 1–67.
- Huot, B. (1990). The Literature of Direct Writing Assessment: Major Concerns and Prevailing Trends. *Review of Educational Research*, 60(2), 237–263. <https://doi.org/10.3102/00346543060002237>

- Hyland, K. (2018). *Metadiscourse: Exploring Interaction in Writing*. Bloomsbury Publishing Plc.
- Högg, C. N. (2010). *Mot ökad diskursivitet? Skrivutveckling speglad i provtexter från årskurs 5 och årskurs 9* (Svenska i utveckling nr 27, s. 1–66). Uppsala Universitet.
- International Labour Office. (2012). *International Standard Classification of Occupation (ISCO-08). Volume 1: Structure, group definitions and correspondence tables*. International Labour Organization.
- Ivanic, R. (1998). *Writing and identity: The discursal construction of identity in academic writing*. John Benjamins Publishing Company.
- Jackson, H., & Amvela, E. Z. (2000). *Words, meaning and vocabulary: An introduction to modern English lexicology*. Cassell.
- Jeffery, J. V., Elf, N., Skar, G. B. U., & Wilcox, K. C. (2018). Writing development and education standards in cross-national perspective. *Writing & Pedagogy*, 10(3), 333–370. <https://doi.org/10.1558/wap.34587>
- Jeffery, J. V., & Wilcox, K. (2014). ‘How do I do it if I don’t like writing?’: Adolescents’ stances toward writing across disciplines. *Reading and Writing*, 27(6), 1095–1117. <https://doi.org/10.1007/s11145-013-9493-9>
- Jeppsson, A., Johnson, S., Karlsson, E., Malmberg, C., Sjölin, V., & Svensson, Å. (2012). *SimSum. En studie om automatisk sammanfattning och omskrivning av texter*. Linköpings Universitet.
- Jiang, J., & Liu, H. (2015). The effects of sentence length on dependency distance, dependency direction and the implications—Based on a parallel English–Chinese dependency tree-bank. *Language Sciences*, 50, 93–104. <https://doi.org/10.1016/j.langsci.2015.04.002>
- Johansson, V. (2009). *Developmental Aspects of Text Production in Writing and Speech*. Lund University.
- Josephson, O., Melin, L., & Oliv, T. (1998). *Elevtext: Analyser av skoluppsatser från åk 1 till åk 9*. Studentlitteratur.
- Jurafsky, D., & Martin, J. H. (2009). *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition* (2. udg.). Prentice Hall.
- Justice, L. M., Bowles, R. P., Kaderavek, J. N., Ukrainetz, T. A., Eisenberg, S. L., & Gillam, R. B. (2006). The Index of Narrative Microstructure: A Clinical Tool for Analyzing School-Age Children’s Narrative Performances. *American Journal of Speech-Language Pathology*, 15(2), 177–191. [https://doi.org/10.1044/1058-0360\(2006/017\)](https://doi.org/10.1044/1058-0360(2006/017))
- Juul, H. (2005). Knowledge of context sensitive spellings as a component of spelling competence: Evidence from Danish. *Applied Psycholinguistics*, 26(2), 249–265. <https://doi.org/10.1017/S0142716405050162>
- Juzwik, M. M. (2012). Spoken narrative. I J. P. Gee & M. Handford (Red.), *The Routledge Handbook of Discourse Analysis* (s. 326–341). Routledge. <https://doi.org/10.4324/9780203809068.ch23>
- Jølle, L., & Skar, G. B. (2020). “Digging for Gold” or “Sticking to the Criteria”: Teachers’ Rationales When Serving as Professional Raters. *Scandinavian Journal of Educational Research*, 64(2), 283–299. <https://doi.org/10.1080/00313831.2018.1541821>
- Jørgensen, H. (2000). *Indføring i Dansk Syntaks* (3. udg.). Århus Universitet.
- Kabel, K. (2016). *Danskfagets litteraturundervisning: Et casestudie af elevers skriftsproglige måder at skabe stillingtagen på i udskolingen*. Ph.d.-afhandling. DPU, Aarhus Universitet.

REFERENCER

- Kabel, K., Bremholm, J., & Bundsgaard, J. (2022). A framework for identifying early writing development. *Writing and Pedagogy*, 13(1–3), 51–87. <https://doi.org/10.1558/wap.21467>
- Kansanen, P. (2009). Subject-matter didactics as a central knowledge base for teachers, or should it be called pedagogical content knowledge? *Pedagogy, Culture & Society*, 17(1), 29–39. <https://doi.org/10.1080/14681360902742845>
- Karlsson, A.-M. (2006). *En arbetsdag i skriftsamhället. Ett etnografiskt perspektiv på skriftanvändning i vanliga yrken*. Norstedts Akademiska Forlag.
- Kelleher, J. D., & Tierney, B. (2018). *Data Science*. The MIT Press.
- Kerz, E., Qiao, Y., Wiechmann, D., & Ströbel, M. (2020). Becoming Linguistically Mature: Modeling English and German Children's Writing Development Across School Grades. I J. Burstein, E. Kochmar, Leacock, N. Madnani, I. Pilán, H. Yannakoudakis, & T. Zesch (Red.), *Proceedings of the Fifteenth Workshop on Innovative Use of NLP for Building Educational Applications* (s. 65–74). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2020.bea-1.6>
- Kilgarriff, A. (2005). Language is never, ever, ever, random. *Corpus Linguistics and Linguistic Theory*, 1(2), 263–276. <https://doi.org/10.1515/cllt.2005.1.2.263>
- Klein, P., & Boscolo, P. (2016). Trends in Research on Writing as a Learning Activity. *Journal of Writing Research*, 7(3), 311–350. <https://doi.org/10.17239/jowr-2016.07.03.01>
- Koester, A. (2010). Building small specialized corpora. I A. O'Keeffe & M. McCarthy (Red.), *The Routledge Handbook of Corpus Linguistics* (s. 66–79). Routledge.
- Koplenig, A. (2016). *Analyzing lexical change in diachronic corpora*. Ph.d.-afhandling. Universität Mannheim.
- Korat, O., Klein, P., & Segal-Drori, O. (2007). Maternal mediation in book reading, home literacy environment, and children's emergent literacy: A comparison between two social groups. *Reading and Writing*, 20(4), 361–398. <https://doi.org/10.1007/s11145-006-9034-x>
- Koutsoftas, A. D. (2016). Writing Process Products in Intermediate-Grade Children With and Without Language-Based Learning Disabilities. *Journal of Speech, Language, and Hearing Research*, 59(6), 1471–1483. https://doi.org/10.1044/2016_JSLHR-L-15-0133
- Kress, G. (1994). *Learning to write* (2. udg.). Routledge.
- Kress, G. (2010). *Multimodality: A social semiotic approach to contemporary communication*. Routledge.
- Kress, G., & van Leeuwen, T. (1996). *Reading Images: The Grammar of Visual Design*. Routledge.
- Kristoffersen, C. (2019). Where do my words come from? Towards methods for analyzing word choice in primary level writing. *Apples - Journal of Applied Language Studies*, 13(3), 59–75. <https://doi.org/10.17011/apples/urn.201907163639>
- Krogh, E. (2012). Skrivning i modersmålsfagene i skriftkyndighedens æra. I N. F. Elf & P. Kaspersen (Red.), *Den nordiske skolen—Fins den?: Didaktiske diskurser og dilemmaer i skandinaviske modersmålsfag* (s. 59–91). Novus forlag.
- Krogh, E. (2014). Overgangen fra grundskole til gymnasium. En elevskrivners håndtering af skiftende skrivekulturer og positioneringer. I P. Andersson, P. Holmberg, A. Lyngfelt, A. Nordenstam, & O. Widhe (Red.), *Mångfaldens möjligheter: Litteratur- och språkdidaktik i Norden* (s. 155–180). Institutionen för svenska språket, Göteborgs Universitet.
- Krogh, E. (2020). Bildung and Literacy in Subject Danish: Changing L1 Education. I B. Green & P.-O. Erixon (Red.), *Rethinking L1 Education in a Global Era: Understanding the (Post-*

-)*National L1 Subjects in New and Difficult Times* (s. 157–176). Springer International Publishing. https://doi.org/10.1007/978-3-030-55997-7_8
- Krogh, E., Christensen, T. S., & Jakobsen, K. S. (2015). *Elevskrivere i gymnasiefag*. Syddansk Universitetsforlag.
- Krogh, E., & Jakobsen, K. S. (Red.). (2016). *Skriverudviklinger i gymnasiet*. Syddansk Universitetsforlag.
- Krogh, E., & Jakobsen, K. S. (Red.). (2019). *Understanding Young People’s Writing Development: Identity, Disciplinarity, and Education*. Routledge.
<https://doi.org/10.4324/9781351010894>
- Kroll, B., & Reid, J. (1994). Guidelines for designing writing prompts: Clarifications, caveats, and cautions. *Journal of Second Language Writing*, 3(3), 231–255.
[https://doi.org/10.1016/1060-3743\(94\)90018-3](https://doi.org/10.1016/1060-3743(94)90018-3)
- Kübler, S., McDonald, R., & Nivre, J. (2009). *Dependency Parsing*. Morgan & Claypool.
<https://doi.org/10.2200/S00169ED1Vo1Y200901HLT002>
- Kyle, K. (2021). Natural language processing for learner corpus research. *International Journal of Learner Corpus Research*, 7(1), 1–16. <https://doi.org/10.1075/ijlcr.00019.int>
- Lankshear, C., & Knobel, M. (2011). From “reading” to “new” literacies. I *New Literacies: Everyday Practices and Social Learning* (3. udg., s. 3–31). McGraw-Hill Education.
- Larsson, K. (1984). *Skrivförmåga: Studier i svenskt elevspråk*. Liber Förlag.
- Latif, M. M. M. A. (2019). Unresolved issues in defining and assessing writing motivational constructs: A review of conceptualization and measurement perspectives. *Assessing Writing*, 42, 1–11. <https://doi.org/10.1016/j.asw.2019.100417>
- Ledin, P. (1998). Att sätta punkt: Hur elever på låg- och mellanstadiet använder meningen i sina uppsatser. *Språk & Stil*, 8, 5–48.
- Ledin, P. (2001). To put a full stop: The use of the sentence in children’s composition. I W. Vagle & K. Wikberg (Red.), *New Directions in Nordic Text Linguistics and Discourse Analysis* (s. 187–198). Novus förlag.
- Ledolter, J., & VanderVelde, L. S. (2021). *Analyzing Textual Information: From Words to Meanings through Numbers* (Bd. 188). SAGE Publications, Inc.
- Lee, S.-Y., & Krashen, S. (2002). Predictors of success in writing in English as a foreign language: Reading, revision behavior, apprehension, and writing. *College Student Journal*, 36(4), 532–543.
- Leech, G. (1992). Corpora and theories of linguistic performance. I J. Svartvik (Red.), *Directions in Corpus Linguistics. Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991* (s. 105–122). De Gruyter Mouton.
- Leijten, M., van Waes, L., & Horenbeeck, E. van. (2015). Analyzing writing process data: A linguistic perspective. I G. Cislaru (Red.), *Writing(s) at the Crossroads: The process-product interface* (s. 277–302). John Benjamins Publishing Company.
<https://doi.org/10.1075/z.194.14lei>
- Letiche, H., Boeschoten, R. V., & Dugal, S. (2008). Special issue on interventionist research. *Journal of Organizational Change Management*, 21(5).
<https://doi.org/10.1108/jocm.2008.02321eaa.002>
- Lincoln, Y. S., & Guba, E. S. (1985). Establishing Trustworthiness. I *Naturalistic Inquiry* (s. 289–331). SAGE Publications.
- Lindgren, J. (2021). Fundamentet: Syntaktisk struktur möter textgrammatisk funktion. I J. Brandtler & M. Kalm (Red.), *Nyanser av grammatik: Gränser, mångfald, fördjupning* (s. 179–192). Studentlitteratur AB.

REFERENCER

- Linton, R. (1936). *The Study of Man*. D. Appleton-Century.
- Liu, H. (2008). Dependency Distance as a Metric of Language Comprehension Difficulty. *Journal of Cognitive Science*, 9(2), 159–191. <https://doi.org/10.17791/jcs.2008.9.2.159>
- Loban, W. (1976). *Language Development: Kindergarten through Grade Twelve*. NCTE Committee on Research Report No. 18. National Council of Teachers of English, Urbana, Ill.
- Lorentzen, R. T., & Evensen, L. S. (1995). *Studies in early writing*. Senter for samfunnsforskning, Sesam, Universitetet i Trondheim.
- Lundgren, B. (2013). Föränderliga text-praktikers betydelse för ungas identitetsskapande. I A.-C. Edlund & S. Haugen (Red.), *Människor som skriver: Perspektiv på vardagligt skriftbruk och identitet* (s. 159–179). Umeå Universitet och Kungl. Skytteanska Samfundet.
- Machin, S., & Vignoles, A. (2004). Educational inequality: The widening socio-economic gap. *Fiscal Studies*, 25(2), 107–128. <https://doi.org/10.1111/j.1475-5890.2004.tb00099.x>
- Madsbjerg, S., & Friis, K. (Red.). (2013). *Skrivelyst i fagene*. Dansk Psykologisk Forlag.
- Magnusson, U., & Kokkinakis, S. J. (2009). Quantitative measures on student texts. I P. Juvenon (Red.), *Språk och lärande. Rapport från ASLA: nr. 22* (s. 43–56). Stockholm University.
- Magnusson, U., & Kokkinakis, S. J. (2011). Computer based quantitative methods applied to first and second language student writing. I R. Källström & I. Lindberg (Red.), *Young Urban Swedish: Variation and change in multilingual settings* (s. 105–124). University of Gothenburg.
- Manovich, L. (2001). *The Language of New Media*. MIT Press.
- Martin, J. R. (1992). *English Text: System and Structure*. John Benjamins Publishing Company.
- Martin, J. R., & Rose, D. (2008). *Genre relations: Mapping culture*. Equinox Publishing Ltd.
- Matre, S., & Solheim, R. (2015). Writing education and assessment in Norway: Towards shared understanding, shared language and shared responsibility. *L1 Educational Studies in Language and Literature*, 15, 1–33. <http://dx.doi.org/10.17239/L1ESLL-2015.15.01.05>
- Matsuda, P. K., & Jeffery, J. V. (2012). Voice in Student Essays. I K. Hyland & C. S. Guinda (Red.), *Stance and Voice in Written Academic Genres* (s. 151–165). Palgrave Macmillan UK. https://doi.org/10.1057/9781137030825_10
- Matthiesen, C. (2013). *Elestyret imitatio: En retorisk skrivepædagogik i teori og praksis*. Ph.d.-afhandling. Københavns Universitet.
- Mazar, A., & Wood, W. (2018). Defining Habit in Psychology. I B. Verplanken (Red.), *The Psychology of Habit: Theory, Mechanisms, Change, and Contexts* (s. 13–29). Springer International Publishing. https://doi.org/10.1007/978-3-319-97529-0_3
- McEnery, T., & Hardie, A. (2013). The History of Corpus Linguistics. I K. Allan (Red.), *The Oxford Handbook of The History of Linguistics* (s. 727–745). Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780199585847.013.0034>
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic Features of Writing Quality. *Written Communication*, 27(1), 57–86. <https://doi.org/10.1177/0741088309351547>
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014). *Automated Evaluation of Text and Discourse with Coh-Matrix*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511894664>
- Megyesi, B., Näsman, J., & Palmér, A. (2016). The Uppsala Corpus of Student Writings: Corpus Creation, Annotation, and Analysis. I N. Calzolari, K. Choukri, T. Declerk, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odijk, & S. Piperidis (Red.),

- Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)* (s. 3192–3199). European Language Resources Association.
- Megyesi, B., Palmér, A., & Näsman, J. (2019). *SWEGRAM - Annotering och analys av svenska texter*. Institutionen för lingvistik och filologi och Institutionen för nordiska språk, Uppsala Universitet.
- Messick, S. (1989). Validity. I R. L. Linn (Red.), *Educational Measurement* (s. 13–103). American Council on Education.
- Miller, D. M., Scott, C. E., & McTigue, E. M. (2018). Writing in the Secondary-Level Disciplines: A Systematic Review of Context, Cognition, and Content. *Educational Psychology Review*, 30(1), 83–120. <https://doi.org/10.1007/s10648-016-9393-z>
- Mills, A., Durepos, G., & Wiebe, E. (2010). Quantitative Analysis in Case Study. I *Encyclopedia of Case Study Research* (s. 761–764). SAGE Publications, Inc. <https://doi.org/10.4135/9781412957397>
- Mitchell, D. (2015). Type-token models: A comparative study. *Journal of Quantitative Linguistics*, 22(1), 1–21. <https://doi.org/10.1080/09296174.2014.974456>
- Mohr, J. W., Wagner-Pacifici, & Breiger, R. L. (2015). Toward a computational hermeneutics. *Big Data & Society*, 2(2), 1–8. <https://doi.org/10.1177/2053951715613809>
- Molbæk, M.-L. (2019). *Skrivning i skolen som autentisk praksis – potentialer og udfordringer: Udvikling og afprøvning af et didaktisk design med udgangspunkt i retoriske situationer*. Ph.d.-afhandling. DPU, Aarhus Universitet.
- Moretti, F. (2000). Conjectures on World Literature. *New Left Review*, 1.
- Moretti, F. (2013a). *Distant reading*. Verso.
- Moretti, F. (2013b). “Operationalizing”: Or, the function of measurement in modern literary theory. *Literary Lab, Pamphlet 6*, 1–13.
- Moxley, J., Elliot, N., Eubanks, D., Vezzu, M., Elliot, S., & Allen, W. (2017). Writing Analytics: Conceptualization of a Multidisciplinary Field. *Journal of Writing Analytics*, 1, v–xvii. <https://doi.org/10.37514/jwa-j.2017.1.1.02>
- Mulvad, R. (2009). *Sprog i skole: Læseudviklende undervisning i alle fag*. Alinea.
- Myhill, D. (2008). Towards a Linguistic Model of Sentence Development in Writing. *Language and Education*, 22(5), 271–288. <https://doi.org/10.1080/09500780802152655>
- Myhill, D. (2009a). Becoming a Designer. Trajectories of Linguistic Development. I D. Myhill, M. Nystrand, J. Riley, & R. Beard (Red.), *The SAGE Handbook of Writing Development* (s. 402–414). SAGE.
- Myhill, D. (2009b). From talking to writing: Linguistic Development in Writing. *British Journal of Educational Psychology Monograph Series II*, 6, 27–44.
- Neiderhiser, J. A., Kelley, P., Kennedy, K. M., Swales, J. M., & Vergaro, C. (2016). ‘Notice the Similarities between the Two Sets ...’: Imperative Usage in a Corpus of Upper-Level Student Papers. *Applied Linguistics*, 37(2), 198–218. <https://doi.org/10.1093/applin/amu017>
- Nelson, N. W., & van Meter, A. M. (2007). Measuring Written Language Ability in Narrative Samples. *Reading & Writing Quarterly*, 23(3), 287–309. <https://doi.org/10.1080/10573560701277807>
- Nikula, H. (1986). *Dependensgrammatik*. Liber Förlag.
- Nordahl-Hansen, A., & Kvernbekk, T. (2020). Construct Validity in Scientific Representation: A Philosophical Tour. *Nordisk Tidsskrift for Pedagogikk Og Kritikk*, 6. <https://doi.org/10.23865/ntp.k.v6.1704>

REFERENCER

- Nordenfors, M. (2011). *Skriftspråksutveckling under högstadiet*. Ph.d.-afhandling. Göteborgs Universitet.
- Nordenfors, M. (2017). *Elevernas texter. Redskap för textanalys, textsamtal och bedömning*. Studentlitteratur.
- Norris, J. M., & Ortega, L. (2009). Towards an Organic Approach to Investigating CAF in Instructed SLA: The Case of Complexity. *Applied Linguistics*, 30(4), 555–578. <https://doi.org/10.1093/applin/amp044>
- Nouri, J., Ebner, M., Ifenthaler, D., Saqr, M., Malmberg, J., Khalil, M., Bruun, J., Viberg, O., González, M. Á. C., Papamitsiou, Z., & Berthelsen, U. D. (2019). Efforts in Europe for Data-Driven Improvement of Education – A Review of Learning Analytics Research in Seven Countries. *International Journal of Learning Analytics and Artificial Intelligence for Education (IJAI)*, 1(1), 8–27. <https://doi.org/10.3991/ijai.v1i1.11053>
- Nyström, C. (2000). *Gymnasisters skrivande: En studie av genre, textstruktur och sammanhang*. Ph.d.-afhandling. Institutionen för nordiska språk, Uppsala Universitet.
- Näsman, J., Megyesi, B., & Palmér, A. (2017). SWEGRAM - A Web-Based Tool for Automatic Annotation and Analysis of Swedish Texts. I J. Tiedeman (Red.), *Proceedings of the 21st Nordic Conference on Computational Linguistics* (s. 132–141). Linköping University Electronic Press.
- OECD. (2019). Students' socio-economic status and performance. I *PISA 2018 Results (Volume II): Where All Students Can Succeed*. PISA, OECD Publishing. <https://doi.org/10.1787/f7986824-en>
- OECD. (2020). *Early Learning and Child Well-being: A Study of Five-year-Olds in England, Estonia, and the United States*. OECD.
- OECD. (2021). *21st-Century Readers: Developing literacy skills in a digital world*. PISA, OECD Publishing.
- Olevar, H. (1999). "Tonårsliv": En pilotstudie av 60 elevtexter från standardproven för skolor 9 åren 1987 och 1996. Enheten för sociolingvistik, FUMS, Institutionen för nordiska språk vid Uppsala universitet.
- Olevar, H. (2002). Skrivuppgift i repris—Presentation av ett forskningsprojekt. I B. Garme (Red.), *Språk på väg: Om elevers språk och skolans möjligheter* (s. 86–100). Hallgren & Fallgren.
- Olinghouse, N. G., & Wilson, J. (2013). The relationship between vocabulary and writing quality in three genres. *Reading and Writing*, 26(1), 45–65. <https://doi.org/10.1007/s11145-012-9392-5>
- Ondelli, S. (2018). Treat Texts as Data but Remember They Are Made of Words: Compiling and Pre-processing Corpora. I A. Tuzzi (Red.), *Tracing the Life Cycle of Ideas in the Humanities and Social Sciences* (s. 133–150). Springer International Publishing.
- Ongstad, S. (2002). Positioning Early Research on Writing in Norway. *Written Communication*, 19(3), 345–381. <https://doi.org/10.1177/074108802237749>
- Onwuegbuzie, A. J., & Johnson, R. B. (2006). The Validity Issue in Mixed Research. *Research in the Schools*, 13(1), 48–63.
- Ouyang, J., Jiang, J., & Liu, H. (2022). Dependency distance measures in assessing L2 writing proficiency. *Assessing Writing*, 51, 100603. <https://doi.org/10.1016/j.asw.2021.100603>
- Oya, M. (2011). Syntactic Dependency Distance as Sentence Complexity Measure. *Proceedings of the 16th Conference on Pan-Pacific Association of Applied Linguistics*, 313–316.
- Palmér, A. (2018). Kvantiteter i kvalitativt bedömda elevtexter—Framtida verktyg för rättvis bedömning? *Acta Didactica Norden*, 12(4), 1–23. <https://doi.org/10.5617/adno.6357>

- Palmér, A., & Hussenius, S. (kommende). *Att skriva diskursivt i 12-årsalderen: Skrivutveckling vid övergangen fra mellanstadiet til högstadiet*. Institutionen för nordiska språk, Uppsala Universitet.
- Palmquist, M. (2019). Directions in Writing Analytics: Some Suggestions. *Journal of Writing Analytics*, 3, 1–12. <https://doi.org/10.37514/JWA-J.2019.3.1.02>
- Parr, J. M. (2010). A dual purpose data base for research and diagnostic assessment of student writing. *Journal of Writing Research*, 2(2), 129–150. <https://doi.org/10.17239/jowr-2010.02.02.3>
- Parr, J. M., & Timperley, H. S. (2010). Feedback to writing, assessment for teaching and learning and student progress. *Assessing Writing*, 15(2), 68–85. <https://doi.org/10.1016/j.asw.2010.05.004>
- Parsons, S. A., & Gallagher, M. A. (2016). A Content Analysis of Nine Literacy Journals, 2009–2014. *Journal of Literacy Research*, 48(4), 476–502. <https://doi.org/10.1177/1086296X16680053>
- Parsons, S. A., Gallagher, M. A., Leggett, A. B., Ives, S. T., & Lague, M. (2020). An Analysis of 15 Journals' Literacy Content, 2007–2016. *Journal of Literacy Research*, 52(3), 341–367. <https://doi.org/10.1177/1086296X20939551>
- Perelman, L. (2020). The BABEL generator and E-rater: 21st Century Writing Constructs and Automated Essays Scoring (AES). *Journal of Writing Assessment*, 13(1), 1–9.
- Pettersson, Å. (1977). Barnspråk. I Å. Pettersson & L. Badersten (Red.), *Språk i utveckling* (s. 88–137). Liber.
- Pettersson, Å. (1980). *Nittonårssvenska: Språket i 1977 års centrala prov i svenska för gymnasiets årskurs 3*. Lärarhögskolan.
- Pettersson, Å. (1989). Utvecklingslinjer och utvecklingskrafter i elevernas uppsatser. I C. Sandqvist & U. Teleman (Red.), *Språkutveckling under skoltiden* (s. 159–184). Studentlitteratur.
- Pokorny, S. B., Jason, L. A., Schoeny, M. E., Townsend, S. M., & Curie, C. J. (2001). Do Participation Rates Change when Active Consent Procedures Replace Passive Consent. *Evaluation Review*, 25(5), 567–580. <https://doi.org/10.1177/0193841X0102500504>
- Popper, K. (1979). *Objective Knowledge: An Evolutionary Approach* (2. udg.). Oxford University Press.
- Puranik, C. S., Lombardino, L. J., & Altmann, L. J. P. (2008). Assessing the Microstructure of Written Language Using a Retelling Paradigm. *American Journal of Speech-Language Pathology*, 17(2), 107–120. [https://doi.org/10.1044/1058-0360\(2008/012\)](https://doi.org/10.1044/1058-0360(2008/012))
- Ray-Subramanian, C. E. (2011). Bilingualism. I S. Goldstein & J. A. Naglieri (Red.), *Encyclopedia of Child Behavior and Development* (s. 240–241). Springer US. https://doi.org/10.1007/978-0-387-79061-9_335
- Rebar, A. L., Gardner, B., Rhodes, R. E., & Verplanken, B. (2018). The Measurement of Habit. I B. Verplanken (Red.), *The Psychology of Habit: Theory, Mechanisms, Change, and Contexts* (s. 31–49). Springer International Publishing. https://doi.org/10.1007/978-3-319-97529-0_3
- Reilly, D., Neumann, D. L., & Andrews, G. (2019). Gender differences in reading and writing achievement: Evidence from the National Assessment of Educational Progress (NAEP). *American Psychologist*, 74(4), 445–458. <http://dx.doi.org/10.1037/amp0000356>
- Reppen, R. (2009). Exploring L1 and L2 Writing Development through Collocations: A Corpus-based Look. I A. Barfield & H. Gyllstad (Red.), *Researching Collocations in Another Language: Multiple Interpretations* (s. 49–59). Palgrave Macmillan UK. https://doi.org/10.1057/9780230245327_4

REFERENCER

- Reppen, R. (2010). *Using Corpora in the Language Classroom*. Cambridge University Press.
- Richardson, K., Calnan, M., Essen, J., & Lambert, L. (1976). The linguistic maturity of 11-year-olds: Some analysis of the written compositions of children in the National Child Development Study. *Journal of Child Language*, 3(1), 99–115. <https://doi.org/10.1017/S0305000900001331>
- Rousse-Malpat, A., Steinkrauss, R., & Verspoor, M. (2019). Structure-based or dynamic usage-based instruction: Long-term effects on (morpho)syntactic and lexical complexity in writing samples. *Instructed Second Language Acquisition*, 3(2), 181–205. <https://doi.org/10.1558/isla.38054>
- Römer, U., & Wulff, S. (2010). Applying corpus methods to writing research: Explorations of MICUSP. *Journal of Writing Research*, 2(2), 99–127. <https://doi.org/10.17239/jowr-2010.02.02.2>
- Raahauge, J. (2020). Danskfaget i skolen. I E. Hjørth, H. G. Jacobsen, B. Jacobsen, M. K. Jørgensen, & L. K. Fahl (Red.), *Dansk sproghistorie. Bind 4: Dansk i brug* (s. 321–336). Aarhus Universitetsforlag.
- Sandersen, V. (1983). *Skolestile og skriftsprogsnormen: En undersøgelse af udvalgte stile fra 9. klasse og 4. klasse*. Danmarks Lærerhøjskole.
- Sandiford, C., & Macken-Horarik, M. (2020). Changing stories: Linguistically-informed assessment of development in narrative writing. *Assessing Writing*, 45, 100471. <https://doi.org/10.1016/j.asw.2020.100471>
- Saussure, F. de. (1916). *Cours de linguistique générale* (C. Bally & A. Sechehaye, Red.). Payot.
- Scheiber, C., Reynolds, M. R., Hajovsky, D. B., & Kaufman, A. S. (2015). Gender Differences in Achievement in a Large, Nationally Representative Sample of Children and Adolescents. *Psychology in the Schools*, 52(4), 335–348. <https://doi.org/10.1002/pits.21827>
- Schlepppegrell, M. J. (1998). Grammar as Resource: Writing a Description. *Research in the Teaching of English*, 32(2), 182–211.
- Scholz, R. W., & Tietje, O. (2002). *Embedded case study methods: Integrating quantitative and qualitative knowledge*. Sage Publications.
- Searle, J. R. (1995). *The Construction of Social Reality*. Free Press.
- Selinker, L. (1972). Interlanguage. *International Review of Applied Linguistics*, 10(1–4), 209–232. <https://doi.org/10.1515/iral.1972.10.1-4.209>
- Selwyn, N. (2019). What's the Problem with Learning Analytics? *Journal of Learning Analytics*, 6(3), 11–19. <https://doi.org/10.18608/jla.2019.63.3>
- Shermis, M. D. (2014). State-of-the-art automated essay scoring: Competition, results, and future directions from a United States demonstration. *Assessing Writing*, 20, 53–76. <https://doi.org/10.1016/j.asw.2013.04.001>
- Shi, Y., & Lei, L. (2022). Lexical Richness and Text Length: An Entropy-based Perspective. *Journal of Quantitative Linguistics*, 29(1), 62–79. <https://doi.org/10.1080/09296174.2020.1766346>
- Shneiderman, B. (1996). The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. *Proceedings of the IEEE Symposium on Visual Languages (1996)*, 336–343.
- Shulman, L. S. (1986). Those Who Understand: Knowledge Growth in Teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, 57(10), 1380–1400. <https://doi.org/10.1177/0002764213498851>

- Silén, B. (2012). Uttryck för attityd i niondeklassares texter. I H. Lehti-Eklund, A. Slotte-Lüttge, B. Silén, & R. Heilä-Ylikallio (Red.), *Skriptpraktiker hos barn och unga* (s. 154–174). Vasa.
- Silverman, R. D., Coker, D., Proctor, C. P., Harring, J., Piantedosi, K. W., & Hartranft, A. M. (2015). The Relationship between Language Skills and Writing Outcomes for Linguistically Diverse Students in Upper Elementary School. *The Elementary School Journal*, *116*(1), 103–125. <https://doi.org/10.1086/683135>
- Sinclair, J. (1996). *Preliminary recommendations on Corpus Typology*. Expert Advisory Group on Language Engineering Standards.
- Sirin, S. R. (2005). Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. *Review of Educational Research*, *75*(3), 417–453. <https://doi.org/10.3102/00346543075003417>
- Skar, G. B., & Jølle, L. J. (2017). Teachers as raters: Investigation of a long term writing assessment program. *L1-Educational Studies in Language and Literature*, *17*, 1–30. <https://doi.org/10.17239/L1ESLL-2017.17.01.06>
- Skar, G. B. U., & Berge, K. L. (2017). *Elevers skriveförmåga och texters kvantitativa egenskaper*. Nasjonalt senter for skriveopplæring og skriveforskning.
- Skar, G. B. U., Jølle, L., & Aasen, A. J. (2020). Establishing Rating Scales to Assess Writing Proficiency Development in Young Learners. *Acta Didactica Norden*, *14*(1), 1–30. <https://doi.org/10.5617/adno.7909>
- Skar, G. B. U., Lei, P.-W., Graham, S., Aasen, A. J., Johansen, M. B., & Kvistad, A. H. (2022). Handwriting fluency and the quality of primary grade students' writing. *Reading and Writing*, *35*, 509–538. <https://doi.org/10.1007/s11145-021-10185-y>
- Skar, G. B. U., Aasen, A. J., & Jølle, L. (2020). Functional Writing in the Primary Years: Protocol for a Mixed-Methods Writing Intervention Study. *Nordic Journal of Literacy Research*, *6*(1), Article 1. <https://doi.org/10.23865/njlr.v6.2040>
- Skar, G. B., Aasen, A. J., Kvistad, A. H., & Johansen, M. B. (2022). Audience awareness in elementary school students' texts: Variations within and between grades 1–3. *Writing and Pedagogy*, *13*(1–3), 155–179. <https://doi.org/10.1558/wap.21541>
- Skjelten, S. M. (2013). *Jakta på kvalitetsforskjellar i elevane sine tekstar. Kva skil gode tekstar frå middels gode?* Ph.d.-afhandling. Det utdanningsvitenskapelige fakultet. Universitetet i Oslo.
- Skopek, J., van de Werfhorst, H., Rözer, J., Zachrisson, H. D., & van Huizen, T. (2017). *Inequality in Various Stages of the Educational Career: Patterns and Mechanisms—Literature Review*. Inclusive Education and Social Support to Tackle Inequalities in Society (ISO-TIS).
- Slomp, D. (2016). Ethical Considerations and Writing Assessment. *Journal of Writing Assessment*, *9*(1).
- Smidt, J. (2010). *Skrijving i alle fag: Innsyn og utspill*. Tapir akademisk forlag.
- Smidt, J. (2018). Tekster, skrivere og skrivekultur i samspill—Et økologisk blik på utviklingstræk i nordisk skriveforskning. I T. S. Christensen, N. Elf, P. Hobel, A. Qvortrup, & S. Troelsen (Red.), *Didaktik i udvikling* (s. 59–74). Klim.
- Smidt, J., Solheim, R., & Aasen, A. J. (Red.). (2013). *På sporet af god skriveundervisning: En bog for lærere i alle fag*. Klim.
- Solheim, R., & Matre, S. (2014). Forventninger om skrivekompetanse: Perspektiver på skrijving, skriveopplæring og vurdering i “Normprosjektet”. *Viden om Læsning*, *15*, 76–89.
- Street, B. V. (1984). *Literacy in theory and practice*. University Press.

REFERENCER

- Street, B. V. (1995). *Social literacies: Critical approaches to literacy in development, ethnography and education*. Longman.
- Struthers, L., Lapadat, J. C., & MacMillan, P. D. (2013). Assessing cohesion in children's writing: Development of a checklist. *Assessing Writing*, 18(3), 187–201. <https://doi.org/10.1016/j.asw.2013.05.001>
- Strömqvist, S. (1987). *Styckevis och helt: Om styckeindelningens roll i skrivprocessen och bruket av nytt stycke i svenska elevuppsatser*. Liber.
- Stubbs, M. (2007). On texts, corpora and models of language. I M. Hoey, M. Mahlberg, M. Stubbs, & W. Teubert (Red.), *Text, Discourse and Corpora: Theory and Analysis* (s. 127–161). Continuum.
- Sula, C. A., & Hill, H. V. (2019). The early history of digital humanities: An analysis of Computers and the Humanities (1966–2004) and Literary and Linguistic Computing (1986–2004). *Digital Scholarship in the Humanities*, 34(Supplement_1), i190–i206. <https://doi.org/10.1093/llc/fqz072>
- Svendsen, H. B. (2016). *Teknologibaseret læsning og skrivning i folkeskolen*. Ph.d.-afhandling. DPU, Aarhus Universitet.
- Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. University Press.
- Tahmasebi, N., Hagen, N., Brodén, D., & Malm, M. (2019). A Convergence of Methodologies: Notes on a Data-Intensive Research Methodology. *DHN2019*, 437–449.
- Tannert, M. (2021). Computing Numerical Images of Student Writing: Perspectives from Educational Research. *Scandinavian Studies in Language*, 12(1), 1–29. <https://doi.org/10.7146/sss.v12i1.130066>
- Tannert, M., & Berthelsen, U. D. (2021). Literacy og digitalisering i skolen: Et historisk perspektiv. *Viden om Literacy*, 29, 8–16.
- Tedick, D. J. (1990). ESL writing assessment: Subject-matter knowledge and its impact on performance. *English for Specific Purposes*, 9(2), 123–143. [https://doi.org/10.1016/0889-4906\(90\)90003-U](https://doi.org/10.1016/0889-4906(90)90003-U)
- Teleman, U. (1978). Om skrivtråning i grundskola och på gymnasium. *Forskning om utbildning*, 2, 38–49.
- Teleman, U., Hellberg, S., & Andersson, E. (1999). *Svenska Akademiens grammatik*. Svenska Akademien.
- The New London Group. (1996). A Pedagogy of Multiliteracies: Designing Social Futures. *Harvard Educational Review*, 66(1), 60–92. <https://doi.org/10.4324/9780203979402-6>
- Thomson, J. J. (2020). Metadiscourse in upper secondary pupil essays: Adapting a taxonomy. *Nordic Journal of Literacy Research*, 6(1), Article 1. <https://doi.org/10.23865/njlr.v6.1720>
- Togebly, O. (2014). *Bland blot genrerne—Ikke teksterne! Om sprog, tekster og samfund*. Samfundslitteratur.
- Togebly, O. (2015). Den dobbelte genreforventning: Genreforhold og sproglige valg i elevtekster. I E. Krogh, T. S. Christensen, & K. S. Jakobsen (Red.), *Elevskrivere i gymnasiefag* (s. 267–292). Syddansk Universitetsforlag.
- Tomasello, M. (2008). *Origins of Human Communication*. MIT Press.
- Troelsen, S. (2018). En invitation man ikke kan afslå – analyse af afgangsprøven i skriftlig fremstilling med særligt fokus på skriveordren. *Nordic Journal of Literacy Research*, 4(1), 142–166. <https://doi.org/10.23865/njlr.v4.1267>

- Troelsen, S. (2020a). *Forkastet eller anerkendt? Et eksplorativt studie i folkeskoleelevers afgangseksamen i "Dansk, skriftlig fremstilling"*. Ph.d.-afhandling. Institut for Kulturvidenskaber, Syddansk Universitet.
- Troelsen, S. (2020b). Sig det med sprogløse blomster! - Censurers kvalitetsnormer og positioneringer ved bedømmelsen af afgangseksamen i 'Dansk, skriftlig fremstilling'. *Nordisk Tidsskrift för Allmän Didaktik*, 6(1), 62–84.
- Troia, G. A., Harbaugh, A. G., Shankland, R. K., Wolbers, K. A., & Lawrence, A. M. (2013). Relationships between writing motivation, writing activity, and writing performance: Effects of grade, sex, and ability. *Reading and Writing*, 26(1), 17–44. <https://doi.org/10.1007/s11145-012-9379-2>
- Troia, G. A., Shen, M., & Brandon, D. L. (2019). Multidimensional Levels of Language Writing Measures in Grades Four to Six. *Written Communication*, 36(2), 231–266. <https://doi.org/10.1177/0741088318819473>
- Ulbæk, I. (2005). *Sproglig tekstanalyse—Introduktion til pragmatisk tekstlingvistik*. Academia.
- Vagle, W. (2005). Tekstlængde + ordlængdesnitt = kvalitet? I K. L. Berge, L. S. Evensen, F. Hertzberg, & W. Vagle (Red.), *Bind II: Ungdommers skrivekompetanse* (s. 303–387). Universitetsforlaget.
- van Es, K., Wieringa, M., & Schäfer, M. T. (2018). Tool Criticism: From Digital Methods to Digital Methodology. I E. Reyes, M. Bernstein, G. Ruffo, & I. Saleh (Red.), *WS.2 2018: Proceedings of the 2nd International Conference on Web Studies* (s. 24–27). Association for Computing Machinery. <https://doi.org/10.1145/3240431.3240436>
- Veber, A.-M. (2016). *God, bædre, best: Om ortografisk viden og ortografisk indlæring hos danske børn på skolens begynder- og mellemtrin*. Københavns Universitet.
- Vereecken, C., & Vandegehuchte, A. (2003). Measurement of parental occupation: Agreement between parents and their children. *Arch Public Health*, 61, 141–149.
- Voyer, D., & Voyer, S. D. (2014). Gender Differences in Scholastic Achievement: A Meta-Analysis. *Psychological Bulletin*, 140(4), 1174–1204. <https://doi.org/10.1037/amp0000356>
- Wagner, R. K., Puranik, C. S., Foorman, B., Foster, E., Wilson, L. G., Tschinkel, E., & Kantor, P. T. (2011). Modeling the development of written language. *Reading and Writing*, 24(2), 203–220. <https://doi.org/10.1007/s11145-010-9266-7>
- Walldén, R. (2020). "It was that Trolle thing" Negotiating history in Grade 6: A matter of teachers' text choice. *Linguistics and Education*, 60, 100884. <https://doi.org/10.1016/j.linged.2020.100884>
- Watermann, R., Maaz, K., Bayer, S., & Roczen, N. (2016). Social Background. I S. Kuger, E. Klieme, N. Jude, & D. Kaplan (Red.), *Assessing Contexts of Learning: An International Perspective* (s. 117–145). Springer International Publishing. https://doi.org/10.1007/978-3-319-45357-6_5
- Weber, M. (2022). Language skills in student essays: Social disparities and later educational attainment. *Longitudinal and Life Course Studies*, 13(2), 239–261. <https://doi.org/10.1332/175795921X16244506861047>
- Wen, H., & Coker Jr., D. L. (2020). The role of discourse knowledge in writing among first-graders. *Journal of Writing Research*, 12(12). <https://doi.org/10.17239/jowr-2020.12.02.05>
- Widdowson, H. (2020). *On the Subject of English: The Linguistics of Language Use and Learning*. De Gruyter Mouton. <https://doi.org/10.1515/9783110619669>
- Wiggen, G. (1982). *Rettskrivningsstudier I. Talemålsskorrelative normavvik i grunnskoleelevers skriftlige arbeid: Norsk forskningsstatus 1981*. Novus forlag.

REFERENCER

- Wilson, J., Roscoe, R., & Ahmed, Y. (2017). Automated formative writing assessment using a levels of language framework. *Assessing Writing*, 34, 16–36. <https://doi.org/10.1016/j.asw.2017.08.002>
- Wright, K. L., Hodges, T. S., Enright, E., & Abbott, J. (2021). The relationship between middle and high school students' motivation to write, value of writing, writer self-beliefs, and writing outcomes. *Journal of Writing Research*, 11(3), 601–623. <https://doi.org/10.17239/jowr-2021.12.03.03>
- Wyatt, M., & Nunn, R. (2019). Exploring academic writing with Grice. *Journal of Teaching English for Specific and Academic Purposes*, 6(3), 365–375. <https://doi.org/10.22190/JTESAP1803365W>
- Wyatt-Smith, C., Klenowski, V., & Gunn, S. (2010). The centrality of teachers' judgement practice in assessment: A study of standards in moderation. *Assessment in Education: Principles, Policy & Practice*, 17(1), 59–75. <https://doi.org/10.1080/09695940903565610>
- Wyse, D. (2017). Thinking about Writing and Language. I *How Writing Works: From the Invention of the Alphabet to the Rise of Social Media* (s. 25–54). Cambridge University Press.
- Yang, & Sun, Y. (2015). Dynamic Development of Complexity, Accuracy and Fluency in Multilingual Learners' L1, L2 and L3 Writing. *Theory and Practice in Language Studies*, 5(2), 298–398. <http://dx.doi.org/10.17507/tpls.0502.09>
- Yin, R. K. (2018). *Case Study Research and Applications: Design and Methods* (6. udg.). SAGE Publications, Inc.
- Zheng, Y., & Yu, S. (2019). What has been assessed in writing and how? Empirical evidence from Assessing Writing (2000–2018). *Assessing Writing*, 42, 1–11. <https://doi.org/10.1016/j.asw.2019.100421>
- Zhu, M., Liu, O. L., & Lee, H.-S. (2020). The effect of automated feedback on revision behavior and learning gains in formative assessment of scientific argument writing. *Computers & Education*, 143, 1–15. <https://doi.org/10.1016/j.compedu.2019.103668>
- Östlund-Stjärnegårdh, E. (2002). *Godkänd i svenska?: Bedömning och analys av gymnasieelevers texter*. Ph.d.-afhandling. Institutionen för nordiska språk, Uppsala Universitet.
- Åberg, E. S. (2022). Upper Secondary School Students' Design of Arguments in Essays. *Designs for Learning*, 14(1), 85–98. <https://doi.org/10.16993/dfl.174>

Bilag 1: Elevtekster i sætningskema

Tabel 37									
<i>Elevtekst 12 i sætningskema</i>									
Konj.	F	V	n	a	V	N	A	ekstra	
<i>Konjunktioner</i>	<i>Forfelt</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverber</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverber vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>	
	Nu	vil	jeg		fortælle	om en dag ligesom alle dage		men ikke helt fordi der var nedlukning i Danmark	
og	det	var				kedeligt		fordi man ikke kunne lege sammen med nogle af sine venner.	
Nå men		lad				os		starte	
	jeg	stod op					kløkken 7		
	Min mor	sagde						der var morgenmad.	
	jeg	spiste				cornflaeks	ligsom alle dage		
	Efter	havde	vi			"Skole"	hjemme.		
	Vi	skulle			starte		med at lave matematik,		

Tabel 37 (fortsat)

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Forfælt</i>	<i>Finitte verber</i>	<i>subjekt</i>	<i>Centraladverbier</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverbier vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
	Efter der	lavet	vi			dansk.		
	Til Sidst	var	det			madkunsts-kab.		
	Der	skulle	vi		lave	banan kage.		
	Efter	skulle	vi		hag	aftensmad.		
	Derefter	skulle					i seng.	
	Godnat							
	og	sov					godt.	

Tabel 38

Elevtækst 13 i sætningsskema

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Forfælt</i>	<i>Finitte verber</i>	<i>subjekt</i>	<i>Centraladverbier</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverbier vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
	2 Piger						ved	
	Det	ser				fortabte ud	springvand	
	de	kan			mærke	Corona	tæt på	

Tabel 38 (fortsat)

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Forfeit</i>	<i>Finitte verber</i>	<i>subjekt</i>	<i>Centraladverber</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverber veedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
	de	gå					med masker og vi siger [læs: visiter]	
	Den ene pige	sidder					på en stor kugle	
	den anden	står					ved springvandet	
	Der	er				ingen mennesker	i bagerunden	
	pigernes	ser					ud til at være helt alene i byen.	

Tablet 39

Eleutektst 14 i sætningsskema

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Forfelt</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverber</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverber vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
		Der	er			helt lukket	i byen	
	der	er		kun		ganske få, der går rundt derne		
	det	er					pga corona virussen	
	det	er				en meget trist stemning	i Tyskland i år	
	Restauranterne	er		helt		tomme		
men	der	er				to piger der er fornuftige		
	de	leger					i et springvand	
men	de	har				mundbind	på	
og	de	holder				afstand	mellem hinanden	
	det	er				sommer		
så	de	bader					i springvandet	

Tablel 40
Eleuttekst 15 i sætningsskema

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Forfælt</i>	<i>Finitte verber</i>	<i>subjekt</i>	<i>Centraladverbier</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverbier vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
	Jeg	vågnede					ca 7:30.	Så jeg kunne nå at få morgenmad og alt der der inden vi havde det første meet møde.
	Det første jeg gjorde da jeg vågnede	var				at jeg tjekkede min telefon.		
	Da jeg havde ligget lidt med den	gik	jeg				nedenunder	hvor min far sad og arbejde, fordi han arbejde hjemme fra.
	Bagetter	gik	jeg				hen	
og		lavede				noget morgenmad.		

Tabel 40 (fortsat)

Konj.	F	V	n	a	V	N	A	ekstra
Konjunktioner	Forfelt	Finite verber	subjekt	Centraladverber	Infinitte verber	Indirekte objekt + direkte objekt + prædikativer	Adverber vedr. tid, sted og omstændigheder	Ledsætninger + Infinitiver
Så	der	havde	jeg	ca	blevet	7:45. et kvarter til at enten se noget youtube eller netflix inden det første meet møde.		
	Der	var	klokken	ca				
	Det første meet møde	startede		for det meste			kl. 8:00.	
	Efter første time	havde	vi			pause		Hvor jeg gik ned og tog noget frokost.
	Det	var		bare		en bolle eller noget rugbrød.		
	Så	havde	jeg			ca en halv time	tilbage	til jeg skulle have min næste time.

Tabel 40 (fortsat)

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Førfelt</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverber</i>	<i>Infinitive verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverber vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + Infinitiver</i>
	I den halv time	sad	jeg	enten			med min telefon eller min computer.	
	Efter 2 time	havde	vi			pause	igen.	
	Der	sad	jeg	bare			enten med min telefon eller min computer igen.	
	Da skoledagen var ovre	tog	jeg	tit			hen til nogle af mine veninder	
eller	de	kom					over til mig, et sted.	
eller	vi	kunne			mødes			

Table 40 (continued)

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Førfelt</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverbier</i>	<i>Infinite verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverbier vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitever</i>
	I corona tiden	stod	jeg	meget			på rulleskøjter med [<i>proprium</i> , <i>proprium</i> , <i>proprium</i> , <i>proprium</i> , <i>proprium</i> og <i>proprium</i>].	
Så og	vi	mødes stod		tit			på rulleskøjter sammen på skolen.	
	Vi	kørte		også nogle gange			i netto eller fakta.	
	Nogle gange	bestilte	vi			en pizza	hos [<i>proprium</i>]	
og		spiste	den				på skolen	
men	for det meste	spiste	vi				hjemme.	

Tabel 40 (fortsat)

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Forfeit</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverber</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverber vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + Infinitiver</i>
	Da jeg så kom hjem igen	var	aftensmaden				klar.	
	Efter jeg havde spist aftensmad	gik	jeg				ind på mit værelse	
og		sad					på min telefon eller min computer	indtil jeg skulle sove.

Tablel 41
Eleutekst 16 i sætningsskema

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Forfælt</i>	<i>Finitte verber</i>	<i>subjekt</i>	<i>Centraladverbier</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverbier vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
	Man	kan			se	en dreng, som er i midten af billedet i mellemgrunden.		
	Der	er		ikke		andre ansigter på billedet end pigens, mandens og drengens		
	Dog	kan	man	se		andre folk	bagfra,	hvis man kigger godt efter i mellemgrunden på billedet.
	Jeg	lægger		også hurtigt	mærke til,			at det er en travl gade, da jeg ser mange folk, som går på gåden.

Tabel 41 (fortsat)

Konj.	F	V	n	a	V	N	A	ekstra
Konjunktioner	Føfelt	Finite verber	subjekt	Centraladverber	Infinitte verber	Indirekte objekt + direkte objekt + prædikativer	Adverbier vedr. tid, sted og omstændigheder	Ledsætninger + infinitiver
	Der	er				et blåt skilt, som viser en mand, som går ned af nogle trapper.		
	Dette	kunne		måske	henvise til	en metro		
og		være			medvirkende til,			at dette sted er en travlt gade.

Tablel 42
Eleutektst 17 i sætningsskema

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Forfælt</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverbier</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverbier vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
	en typisk dag i corona	ville			være			at jeg vågnede kl 8.30 eller 9.00,
og	så	spiste	jeg			morgennemad		
og	så	går	jeg				ud	hvis der ikke er et goggle meet
og	når jeg er færdig med det og så	laver	jeg			læktier		
og	så	spiller	jeg	måske		unu		
og	så	skal	vi		spise	aftensmad		
og	så	ser	jeg			youtube		
og	så	skal	jeg		sove.			

Tablet 43
Eleutkest 18 i sætningsskema

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Forfælt</i>	<i>Finitte verber</i>	<i>subjekt</i>	<i>Centraladverber</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverber vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
	Jeg	står					op	
og		tager				noget tøj	på	
og		spiser				morgennemad		
og	så	ser	jeg			en video		
og	så	lavede	jeg			nogle lektier		
og	så	spiste	jeg			noget middagsmad		
og	så	tog	jeg				på arbejde	
og	da jeg fik fri fra arbejde	lavede	jeg			lidt lektier		
og	så	spiste	jeg			aftensmad		
og	så	spillede	jeg				lidt	
og	så	gik	jeg				i seng	

Tabel 44
Elevtækst 19 i sætningsskema

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Forfælt</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverbier</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverbier vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
	Illustrationen	viser				en ung mad med et ubehageligt ansigtstræk.		
Eller	det	viser				det af ansigtet, man kan se over beskyttelsen.		
	Hun	er				omringet	af mennesker uden beskyttelse	
og		føler				sig nok overdramatisk		
	Manden bag hende	ignorerer				alle former for varsler der har været mod spredningen af den nye virus.		

Tabel 44 (fortsat)

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Foøfeit</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverbier</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverbier vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
	Enten	vælger	han			at undgå virkeligheden		
eller		er		bare ikke		klar	til at tage imod den.	
	Nyheden om at virussen rammer ekstra hårdt mod rygere med dårligere lunger	er		nok ikke	nået	hans vej.		

Tabel 44 (fortsat)

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Foøfelt</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverbier</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverbier vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
	De mange andre mennesker som bevæger sig rundt om den anderledes pige, som faktisk tæmker på sig selv og dem omkring hende.							
	Hun	prøver						at beskytte de tætte mennesker mod den ukendte virus.

Tabel 44 (fortsat)

Konj.	F	v	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Førfelt</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverbier</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverbier vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
	Hvad der er sket før billedet blev taget	er				svært		at vurdere
Men	med den kvindeliges ansigtsudtryk	kan	man		blive ledt		frem til	at hun ikke har det super behageligt i den situation hun er i, i øjeblikket.
	Således	kan	hendes ansigtsudtryk	også	have		noget med	hvad det er på hendes elektroniske device, som hun holde om beskyttende med begge hænder.

Tabel 44 (fortsat)

Konj.	F	v	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Foøfelt</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverbier</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverbier veør: tid, øted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
	De slørede ansigter i baggrunden	viser				et stort smil	på den mandlige person som nærmer sig.	
	Han	er			iført	en dunjakke	med en meget blændende farve på indersiden.	
	Efter billedet	ville	jeg		regne		med	at den kvindelige figur ville gå sin vej uden at bevæge sig for tæt på de andre rundt om hende.
	Den rygende man bag kvinden	ville		nok	blive	stående		

Tabel 44 (fortsat)

Konj.	F	v	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Førfelt</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverbier</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverbier vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
og					færdiggøre	hans forgiftning af sine lunger		
eller					fortsætte	gående		indtil det foldede papir med nikotin er lille og ligeysldig nok til at smide på jorden og kvase under sin sko.
	Stemming	afhænger					af de forskellige personer.	
	De	har		nok alle		deres eget synspunkt for situationen.		

Tablel 45
Eleutekst 20 i sætningsskema

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Forfælt</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverbier</i>	<i>Infinitte verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverbier vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + Infinitiver</i>
	Jeg	har			valgt	en hverdag	i corona krisen	hvor vi var begyndt at have hjemmeundervisning.
	Jeg	vågnede		altid			kl 7:50 ca	
og		stod					op.	
	Så	gik	jeg				ned	
og		fik				noget morgenmad		
og	derefter	fandt	jeg			mine skoleting	frem	
og		var				klar	til at lave skolearbejde	
	Så	gik	jeg				8.05.	
								hvor vi fik pause
og	derefter	var	jeg				i skole til 12- 12:30 tiden	
og		fik		derefter		fri.	i skole til 14- 15 ca	
	så	plejede	jeg					at spise frokost
og	derefter	løbe				en tur		

Table 45 (continued)

Konj.	F	V	n	a	V	N	A	ekstra
<i>Konjunktioner</i>	<i>Forfelt</i>	<i>Finite verber</i>	<i>subjekt</i>	<i>Centraladverber</i>	<i>Infinite verber</i>	<i>Indirekte objekt + direkte objekt + prædikativer</i>	<i>Adverber vedr. tid, sted og omstændigheder</i>	<i>Ledsætninger + infinitiver</i>
og	[pos. <i>pronomen + substantiv</i>]	er				[<i>preposition + substantiv</i>]	over at spille fodbold.	
så	jeg	måtte		slet ikke	være		sammen med nogle under lockdown	
så	derfor	dyrker	jeg			meget motion.		
og	Om aftenen	spiste	jeg			aftensmad	f.eks. sammen med familie	
og	så	gik	jeg				i seng ved 23-24 tiden.	

Bilag 2: Elevtekster i skema for nominalhelheder

Tabel 46					
<i>Nominalfraser i elevtekst 6</i>					
Sammenfatter	Bestemmer	Tæller	Beskriver	Kategori	Efterstillede led
	min			telefon	
	vores			spisebord	
		cirka 2		timer	
	min			lillebror	
		lidt		tid	
	min			lillebror	
	min			lillebror	
	vores			trampolin	
	min			telefon	
	min			mor	

Tabel 47					
<i>Nominalfraser i elevtekst 7</i>					
Sammenfatter	Bestemmer	Tæller	Beskriver	Kategori	Efterstillede led
				billede	1
		to		piger	
	den	ene		pige	
	en			sten	
	en			strømpe	
	den	anden		piger	
	den	ene		pige	
				Pigen	der sidder på stenen
			grå	bukser	
	en		hvid	trøje	

Tabel 47 (fortsat)

				Pigen	der er i gang med at gå væk
			røde	bukser	
	en		lyserød	trøje	
	et	par	lyserøde	sko	
	en			restaurant	
	nogle			kunder	
	en		gul	parasol	
	en	anden		restaurant	
	et			springvand	

Tabel 48

Nominalfraser i elevtekst 8

Sammenfatter	Bestemmer	Tæller	Beskriver	Kategori	Efterstillede led
		flere timers		lektier	
		5 til 3		minutter	
	min			ven	
				resten	af dagen
				resten	af tiden

Tabel 49

Nominalfraser i elevtekst 9

Sammenfatter	Bestemmer	Tæller	Beskriver	Kategori	Efterstillede led
				Billede	1
		to		piger	der leger i vandet
		mange		mennesker	
		mange	tomme	stole	

Tabel 49 (fortsat)

	den	ene af		pigerne	
	en	lidt	anderledes	måde	
	de	to		børn	
	de	andre		børn	
	deres			klassekammer- ater	
	deres			forældre	
	den	ene			
				folk	der går forbi

Tabel 50
Nominalfraser i elevtekst 10

Sammenfatter	Bestemmer	Tæller	Beskriver	Kategori	Efterstillede led
	en		typisk	skoledag	i corona-tiden
	min			madpakke	
	mit			hverdagstøj	
	mine			tænder	
	min			telefon	
	min			klasse	
	de	næste		timer	
		andet		frikvarter	
	min			far	
	min			cykel	
	et			bad	
		noget		mad	
		nogle		gange	
	en			tur	
	min			hund	
	mit			værelse	

Tabel 50 (fortsat)

		noget		tv	
--	--	-------	--	----	--

Tabel 51*Nominalfraser i elevtekst 11*

Sammenfatter	Bestemmer	Tæller	Beskriver	Kategori	Efterstillede led
				billed	1
				pigen	til venstre
	en			maske	
				pigen	til venstre
	en		hvid langærmede blonde	trøje	
	nogle		grå	leggings	
		lidt	mørkere	prikker	
	nogle		grå, blå, gule og sorte	sokker	
				pigen	til højre
			røde	leggings	
	en		lyserød langærmede		
			brunt	hår	
	hendes			hår	
	en	halv		hestehale	med et skrig grøn elastik
			våde	fodspor	
	hendes			fødder	
hele	hendes			overkrop	
	de		våde	fliser	

Tabel 51 (fortsat)

	nogle		polske	restauran- ter og øl- mærke	
	de	to		piger	
				pigen	til venstre
	det		eneste		
				hende	til højre
				pigen	med maske
	en			sten	
				hende	uden maske
	nogle		lyserøde	sko	
	den			sten	pigen med masken sid- der på
	hendes			sko	
		mange		farver	
alle			varme, og hyggelige	farver	
	en		dejlig	dag	med solskins- vejr

Bilag 3: Skriveprompt

SKRIVEOPGAVE

Du skal både skrive opg. 1 og opg. 2.

Sundhedskrise i verden...

1) Vælg ét af billederne. Beskriv den situation, du ser på billedet.



Billede: Pixabay.com

(se næste side)



Billede: Unsplash.com

2) Fortæl om en typisk dag for dig derhjemme under coronanedlukningen.

Bilag 4: Måleegenskaber for 'læsevaner' og 'skrivelyst'

Læsevaner

Tabel 52*Faktoranalyse for 'læsevaner'*

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	1.92027	0.73833	0.3200	0.3200
Factor 2	1.18194	0.24311	0.1970	0.5170
Factor 3	0.93883	0.17934	0.1565	0.6735
Factor 4	0.75949	0.11299	0.1266	0.8001
Factor 5	0.64650	0.09354	0.1078	0.9078
Factor 6	0.55296	.	0.0922	1.0000

Note: N = 230, principal-component factors, unrotated, 2 retained factors, 11 params.

Tabel 53*Faktorloadings for læsevaneitems*

Factor	Factor 1	Factor 2	Uniqueness
Jeg læser religiøse tekster (fx Bibelen eller koranen).	1.92027	0.73833	0.3200
Jeg læser nyheder (på nettet eller på papir).	1.18194	0.24311	0.1970
Jeg læser blade og magasiner (fx Vi unge, Goal, Illustreret videnskab).	0.93883	0.17934	0.1565
Jeg læser bøger, som forklarer ting (fx sport, idoler, dyr, mode).	0.75949	0.11299	0.1266
Jeg læser historier eller romaner (enten fysisk eller som e-bog).	0.64650	0.09354	0.1078
Jeg læser tegneserier.	0.55296	.	0.0922

Note: Scale reliability coefficient (Chronbachs Alpha) = 0.5406 ³¹

³¹ Chronbachs Alpha er forventet lav. Dette skyldes, at der teoretisk set ikke er nogen nødvendig sammenhæng i aktivitetsniveauet imellem de læseaktiviteter, som de seks items dækker over. At en elev læser mange tegneserier, hænger fx ikke nødvendigvis sammen med, at eleven også læser mange romaner.

Skrivelyst**Tabel 54***Faktoranalyse for 'skrivelyst'*

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	3.43488	3.34771	1.0411	1.0411
Factor 2	0.08717	0.11572	0.0264	1.0676
Factor 3	-0.02855	0.04171	-0.0087	1.0589
Factor 4	-0.07026	0.05384	-0.0213	1.0376
Factor 5	-0.12410	.	0.0376	1.0000

Note: N=230, principal factors, unrotated, 2 retained factors, 9 params.

Tabel 55*Faktorloadings for skrivelystitem*

Factor	Factor 1	Factor 2	Uniqueness
Jeg kan godt lide at skrive.	0.9223	-0.0318	0.1483
Det er sjovt at skrive.	0.9115	-0.0310	0.1683
Jeg kan godt lide at skrive i skolen.	0.7652	-0.2003	0.3743
Jeg kan godt lide at skrive derhjemme.	0.7794	0.1988	0.3530
Skrivning er en god måde at bruge tiden på.	0.7486	0.0747	0.4340

Note: Scale reliability coefficient (Chronbachs Alpha) = 0.9159

Bilag 5: Brev til forældre og samtykkeerklæring

Forskningsprojekt

SkrivL1: Skriftsproglig udvikling i grundskolens danskfag

■■■■■■■■■■ deltager i forskningsprojektet SkrivL1, og i denne forbindelse har vi brug for dit samtykke til at behandle dit barns personoplysninger. Dit barns personoplysninger behandles i overensstemmelse med Databeskyttelsesforordningen (GDPR).

Projektets formål

Formålet med forskningsprojektet SkrivL1 er at undersøge, hvordan skoleelevers skriftsprog udvikler sig i løbet af de sidste fem år i grundskolen. Til det formål vil vi gerne indsamle tekster fra danskfaget (danskstile) skrevet af elever på 5.-9. klassetrin. Dette gør vi, så vi kan opbygge en digital samling af tekster, der gør det muligt at bruge digitale analyseværktøjer til at beskrive, hvordan teksterne og skriftsproget udvikler sig, fx i forhold til sætnings- og tekststrukturer, ordforråd mm. Der vil derfor ikke være fokus på at nærlæse indholdet eller måle den enkelte elevs niveau, men derimod på at finde generelle udviklingsmønstre på tværs af tekster og klassetrin.

Hvilke data indsamles?

Vi indsamler én tekst fra hver elev skrevet i danskundervisningen i indeværende skoleår (20/21). Teksten skrives på baggrund af en opgave, som er udarbejdet særligt til dette projekt. I forbindelse med indsamlingen af teksterne svarer eleverne på et spørgeskema, hvor der indsamles metadata om bl.a. klassetrin, alder, fag samt skrive- og læsevaner. Vi indsamler **ikke** data om navn, fødselsdato, adresse, cpr-nr. eller andre personhenførbare personoplysninger. I det omfang denne type oplysninger indsamles tilfældigt, fx hvis et elevnavn optræder i et sidehoved, slettes de i forbindelse med den indledende formatering af teksterne. Det er aftalt med skoleleder, ■■■■■■■■■■, at alle elever løser skriveopgaven som en del af undervisningen, men at det kun er ved de elever, der er givet tilsagn til, at tekster og metadata indsamles.

Hvem er dataansvarlig?

Aarhus Universitet (CVR nr. 31119103) er dataansvarlig for behandlingen af dit barns personoplysninger. Projektleder Morten Tannert er ansvarlig for projektet og kan kontaktes på mota@edu.au.dk.

Hvordan behandles og opbevares teksterne?

Alle forskningsprojekter på Aarhus Universitet skal efterleve retningslinjerne i den nye europæiske forordning om databeskyttelse (GDPR), hvilket sikrer en høj grad af datasikkerhed. Dette betyder, at de indsamlede tekster bliver fuldstændigt anonymiserede. De anonymiserede tekster samles til et digitalt tekstkorpus, hvorefter de originale tekster slettes. Når formateringsprocessen er gennemført, vil det ikke længere være muligt at identificere forfatteren til en given tekst, ligesom det også vil være umuligt at genskabe denne relation mellem forfatter og tekst.

Hvad bruges de indsamlede data til?

Det anonymiserede tekstkorpus vil blive brugt i forbindelse med forskning og undervisning. Det vil desuden på et senere tidspunkt blive inkorporeret i et større korpus af danske elevtekster udviklet med henblik på forskning og undervisning. Dette betyder, at også forskere og studerende fra samarbejdende institutioner på sigt vil få adgang til det anonymiserede korpus.

På forhånd tak.

Samtykkeerklæring

Frivillighed

Det er frivilligt at deltage i projektet, men vi håber, du har lyst til at deltage, idet kvaliteten af forskningsresultaterne i høj grad afhænger af volumen. Kort sagt: Jo flere tekster, jo bedre.

Du kan trække dit samtykke tilbage ved at kontakte projektansvarlig Morten Tannert på mota@edu.au.dk. Hvis du tilbagetrækker dit samtykke, får det først virkning fra dette tidspunkt og påvirker ikke lovligheden af vores behandling af de indsamlede data op til dette tidspunkt.

Du skal dog være opmærksom på, at samtykket ikke kan trækkes tilbage efter formateringsprocessen og oprettelsen af tekstkorpuset. Herefter vil det som følge af anonymiseringsprocessen være umuligt at identificere dit barns tekst i korpuset.

Samtykke

Jeg bekræfter at have modtaget, læst og forstået ovenstående som baggrund for mit samtykke til følgende:

- 1) at mit barns tekst og de indsamlede metadata anvendes til oprettelse af digitalt tekstkorpus, der skal bruges til undersøgelse af skriftsproglig udvikling ved hjælp af automatiserede analysemetoder.
- 2) at mit barns tekster og de indsamlede metadata, som en del af SkrivL1-korpuset, på sigt vil blive en del af et større korpus af danske elevtekster.

Elevens navn: _____

Klasse: _____

Dato: _____

Forældreunderskrift: _____

AU's databeskyttelsesrådgiver kan kontaktes på dpo@au.dk.

Bilag 6: Brev til de deltagende lærere

Forskningsprojekt

SkrivL1: Skriftsproglic udvikling i grundskolens danskfag

Kære dansklærere fra 5.-9. årgang på [REDACTED]

Tak for jeres deltagelse i forskningsprojektet SkrivL1.

I projektet undersøger jeg, hvordan skoleelevers skriftsprog udvikler sig i de sidste fem år af grundskolen. Jeg skal derfor indsamle danskstile skrevet af elever fra 5.-9. klassetrin. Formålet med dette er at opbygge et såkaldt tekstkorpus (en digital samling af tekster), som gør det muligt at anvende digitale analyseværktøjer til at finde sproglige udviklingsmønstre på tværs af tekster og klassetrin (fx i forhold til syntaks- og tekststrukturer, ordforråd, semantiske netværk mm.).

Projektet indebærer følgende:

- 1) Uddeling og indsamling af samtykkeerklæringer i jeres klasse.
- 2) Jeres elever skal skrive én danskstil på baggrund af en opgave, som jeg formulerer. I må meget gerne afsætte tid i undervisningen til, at eleverne kan løse opgaven.
- 3) I forbindelse med afleveringen af danskstilen udfylder eleverne et kort spørgeskema, der bl.a. spørger ind til alder, køn og læse- og skrivevaner.

Skriveopgaven er ens for alle klassetrin. Det er den bl.a. for at sikre, at elevernes tekster kan sammenlignes på tværs af klassetrin. Derfor er skriveopgaven også formuleret forholdsvis åbent, hvilket selvfølgelig kan være en udfordring for nogle elever. Ved de elever, som har svært ved at løse opgaven, skal I selvfølgelig bare støtte og vejlede dem undervejs, på samme måde som I plejer at gøre i forbindelse med skriftlige afleveringer. Opgaven skal ikke ses som en test, men som en almindelig skriftlig aflevering. Det betyder, at fx elever med læse- og skrivevanskeligheder selvfølgelig kan bruge de hjælpemidler, som de normalt bruger i undervisningen.

Det er ligeledes vigtigt for kvaliteten af projektet, at så mange elever som muligt deltager. Det er derfor vigtigt at understrege, at elevernes tekster bliver fuldstændig anonymiserede, så snart teksterne er formateret og samlet i et digital tekstkorpus. Eleverne behøver derfor ikke at være nervøse for, om man

BILAG

på et senere tidspunkt vil kunne finde frem til, hvem der har skrevet den enkelte tekst. På samme måde er det vigtigt at understrege, at projektet ikke handler om at vurdere den enkelte elevs niveau, men om at finde generelle mønstre på tværs af elever, klassetrin, køn etc.

Tidsplanen for projektet ser således ud:

Tidsperiode	Aktivitet
Uge 34	Uddeling og indsamling af samtykkeerklæringer
Uge 35-36:	7.-9. kl.: Eleverne besvarer skriveopgaven
Uge 35-38	5.-6. kl.: Eleverne besvarer skriveopgaven.
Fredag d. 4/9	7.-9. kl.: Jeg kommer ud i de enkelte klasser og hjælper med indsamling af tekster og spørgeskemaer.
Fredag d. 18/9	5.-6. kl.: Jeg kommer ud i de enkelte klasser og hjælper med indsamling af tekster og spørgeskemaer.

Teksterne afleveres i docx-format. I må meget gerne hjælpe eleverne med at downloade deres tekst i dette format, hvis de fx har skrevet den i Google Docs. Ellers er jeg selvfølgelig også behjælpelig med dette, når jeg kommer ud i klasserne.

Hvis I har spørgsmål eller kommentarer, er I altid meget velkomne til at skrive til mig på mota@edu.au.dk eller ringe på tlf. [REDACTED].

På forhånd tusind tak for hjælpen!

Bedste hilsner

Morten Tannert
Ph.d.-studerende
DPU, Aarhus Universitet
Afdeling for Fagdidaktik
Jens Chr. Skous Vej 4, 8000 Aarhus C
E-mail: mota@edu.au.dk

August/september 2020