

Forfeltsvariation i elevtekster fra 5. og 9. klassetrin

Morten Tannert, VIA University College

Ulf Dalvad Berthelsen, Aarhus Universitet

Indledning

Diderichsen (1966, s. 24) skriver om udfyldning af forfeltet: ”Behandlingen af dette Indlederled, der paa een Gang er bestemmende for hele Sætningens indre Struktur og det Element der fortrinsvis betegner Sætningens Forhold til det foregaaende, er al Prosas Nerve”. Tydeligere kan det vist ikke siges. Beherskelsen af forfeltets muligheder er et helt afgørende element i beherskelsen af skriftsproget og et vigtigt grundelement i dansk syntaks, idet det fungerer som et vigtigt tekstgrammatisk fundament for en teksts generelle sagsfremstilling, fx som grundlag for tema-rema-strukturer (Jørgensen, 2000; Lindgren, 2021). Selvom der findes få begrænsninger for, hvad man kan placere i forfeltet, er det samtidig klart, at der er visse sætningsled, som forekommer i forfeltet hyppigere end andre, ligesom der er forskelle imellem tekster, hvis man ser på forfeltslængde og graden af variation. Der er derfor også god grund til at interessere sig for, hvad der kendetegner skoleelevers anvendelse af forfeltet på tværs af klassetrin. Forfeltet har af denne grund været et fokuspunkt i den hidtidige internationale skriveforskningslitteratur, som har undersøgt syntaktisk udvikling i tekster skrevet af skoleelever. Her har det bl.a. vist sig, at forfeltslængde er et af de sproglige træk, der bedst kan forudsige kvaliteten af elevernes tekster (typisk forstået som tekstens tildelte karakter) (Crossley, 2020), samt at anvendelsen af bl.a. adverbielle led eller ledsætninger i forfeltet er et tegn på, at eleverne har mere avancerede skrivefærdigheder (Myhill, 2008). Der er imidlertid ikke foretaget lignende systematiske undersøgelser af forfeltsvariation i elevtekster i en dansk kontekst.

I denne artikel præsenterer vi derfor en kvantitativ, eksplorativ analyse af forskelle i grundskoleelevers realisering af forfeltet i et tekstkorpus bestående af elevtekster fra 5. (n = 86) og 9. klassetrin (n = 88). Mere specifikt undersøger vi, om der er statistisk signifikante stigninger mellem de to klassetrin i det gennemsnitlige antal ord i forfeltet samt i antallet af lange forfelter relativt til det totale antal forfelter i teksten. På baggrund af disse kvantitative analyser undersøger vi dernæst kvalitativt seks elevtekster, som har fået forskellige værdier i de kvantitative analyser, med henblik på at beskrive, hvilke syntaktiske træk der kendetegner tekster med

forskellige numeriske værdier i de kvantitative analyser. Analyserne er således designet som mixed methods-analyser og giver både et billede af generelle tendenser i tekstkorpuset samt et kvalitativt og grammatisk indblik i de bagvedliggende syntaktiske strukturer.

Tidligere forskning

At undersøge sproglige strukturer i elevtekster fra et kvantitativt perspektiv er på én gang både en ny og gammel forskningstradition. Det er en gammel tradition i den forstand, at korpuslingvistikken har dybe rødder i 1940'ernes datalingvistik og 1960'ernes og 1970'ernes korpuslingvistiske landvindinger (Hirst 2013; McEnery & Hardie 2013). I en skandinavisk kontekst har dansk og norsk skriveforskning primært fokuseret på skrivedidaktik og læse- og skriveprocesser forstået som sociokulturelle praksisformer, mens man i Sverige har haft en større tradition for korpuslingvistisk og tekstorienteret skriveforskning (Bremholm et al. 2021; Tannert 2022). I et historisk perspektiv var det særligt Hultman & Westmans *Gymnasistsvenska* fra 1977, som markerede et afgørende nybrud, idet projektet med stor gennemslagskraft introducerede kvantitative operationaliseringer af syntaktisk kompleksitet og leksikal variation som nye metodologiske trædesten for den lingvistiske skriveforskning i Sverige. Dette blev op igennem 80'erne og 90'erne fulgt op af en lang række kvantitative studier af skriveudvikling i elevtekster fra grundskolen og forskellige ungdomsuddannelser (fx Chrystal & Ekvall 1999; Garne 1988; Josephson et al. 1998; Larsson 1984; Pettersson 1980, 1989). Kendetegnende for mange af disse studier er imidlertid, at de er baseret på manuelt indtastede og/eller annoterede tekstkorporer. Det vil sige, at både analysearbejdet og den første bearbejdning af teksterne er gjort i hånden.

Den kvantitative skriveforskning, som den ser ud i dag, er imidlertid også etableret i en nyere forskningstradition i den forstand, at nye teknologiske muligheder i form af maskinlæring har gjort værktøjer tilgængelige, der automatisk kan processere og analysere store mængder af tekst. Det er disse værktøjer, som kendes under samlebetegnelsen Natural Language Processing (NLP) (se fx Eisenstein 2019). Sammenholdt med det forhold, at elevtekster i dag som følge af digitaliseringen af uddannelsessystemet ofte eksisterer i digitalt format, er der således kommet gode muligheder for at undersøge skrivning og skriveudvikling i grundskolen fra et digitalt og korpuslingvistisk perspektiv. I den internationale forskningslitteratur inden for området er der to parallelle tendenser. I den ene undersøges det typisk, hvordan forskellige lingvistiske træk i teksterne forandrer sig over tid (et diakront perspektiv), og i den anden undersøges det typisk, hvordan de lingvistiske træk korrelerer med eller er i stand til at forudsige en række ekstralingvistiske forhold, fx en teksts tildelte karakter eller skribentens køn (et synkront perspektiv). Af diakrone studier kan bl.a. nævnes Beard & Burell (2010) og Beard et al. (2016), som undersøger

udviklingen af sætningstyper og -længde i argumenterende og berettende tekster skrevet af 9-11-årige elever, samt Troia et al. (2019), der undersøger klassetrinsrelaterede forskelle i op mod 100 forskellige lingvistiske variable i tekster fra 4.-6. klassetrin. Kendetegnende for den diakrone tilgang er således, at skriveudvikling primært anses som en udviklingsproces, der påvirkes over tid af elevernes alder/klassetrin. Den synkrone tilgang ses derimod hos bl.a. McNamara et al. (2010) og Crossley & Kim (2022), som undersøger, hvilke lingvistiske variable i elevernes tekster der bedst forudsiger teksternes tildelte karakter, samt hos Weber (2022) og Reilly et al. (2019), der undersøger sammenhængen mellem lingvistiske variable i elevernes tekster og henholdsvis elevernes sociale baggrund og køn. Den synkrone tilgang er således kendetegnet ved at interessere sig for årsagerne til den variation, der findes i tekster skrevet af elever med samme alder. I denne artikel er det imidlertid ikke denne synkrone variation, vi interesser os for, men derimod de diakrone forskelle i anvendelsen af forfeltet, der kan identificeres mellem tekster skrevet af elever med fem klassetrins forskel.

Der findes et forholdsvis begrænset antal studier med rødder i den lingvistiske og kvantitative skriveforskning, som specifikt har haft forskellige former for kompleksitet og variation i forfeltet som analysegenstand. Dette hænger sandsynligvis sammen med, at der dels er en række tekniske udfordringer forbundet med at isolere forfeltet computationelt, dels at forfeltet ikke er et lige så velfunderet teoretisk fænomen, som det i Danmark er blevet i kraft af Diderichsens (1946) indflydelsesrige introduktion af helsætningsskemaet i dansk strukturel lingvistik. Det amerikansk-udviklede Coh-Metrix, som er et af de største automatiserede, kvantitative værktøjer til tekstanalyse (McNamara et al. 2014), har forfeltet (defineret som 'number of words left of the main verb in the main clause') som et af over 100 lingvistiske variable, og i kraft af dette er analyser af forfeltet inkorporeret i en række amerikanske studier af skrivning og skriveudvikling (se fx Crossley & McNamara 2014; Graesser et al. 2011; Perin & Lauterbach 2018), dog uden at være genstand for særskilt fokus. I flere studier viser forfeltet sig imidlertid at være et af de syntaktiske træk, der har den stærkeste korrelation med tekstkvalitet (Crossley et al. 2011; McNamara et al. 2010, 2013), hvilket altså peger på, at forfeltet er et vigtigt fokuspunkt for at forstå børns skriveudvikling. I et større dansk skriveforskningsprojekt, ATEL, er forfeltet ligeledes inkorporeret som én blandt flere syntaktiske variable i en analyse af et tekstkorpus af elevtekster fra indskolingen (Bremholm et al. 2022; Kabel et al. 2020). Studiet peger bl.a. på, at mere avancerede elevskrivere i indskolingen kan variere forfeltet ved bl.a. at placere direkte objekt, interrogative pronominer eller tidsadverbialer i forfeltet, hvilket står i modsætning til de mest simple tekster, som udelukkende har korte subjektsudfyldninger i forfeltet. Projektet peger således på, at evnen til at variere forfeltet kan ses som et tegn på skriveudvikling på grund-

skolens begyndertrin. I forlængelse heraf undersøger vi i denne artikel, hvordan denne variation ser ud, når man sammenligner tekster skrevet på 5. klassetrin (grundskolens mellemtrin) og 9. klassetrin (grundskolens afsluttende trin) – både fra et kvantitativt, strukturelt perspektiv og fra et kvalitativt, grammatisk perspektiv.

Data og metode

Teksterne, der udgør datamaterialet for denne undersøgelse, stammer fra et ph.d.-projekt om skriveudvikling i grundskolen (Tannert 2022). Tekstkorpuset består af to teksttyper, henholdsvis berettende og beskrivende. Teksterne er skrevet af elever fra 5. og 9. klassetrin, og korpuset indeholder i alt 174 tekster, heraf 87 beskrivende tekster (5. kl. = 43, 9. kl. = 44) og 87 berettende tekster (5. kl. = 43, 9. kl. = 44). Vi vil i det følgende henvise til de fire grupper af tekster på følgende måde: 5D = 5. klassetrins deskriptive tekster, 5N = 5. klassetrins narrative tekster, 9D = 9. klassetrins deskriptive tekster, 9N = 9. klassetrins narrative tekster. Eleverne har skrevet teksterne på baggrund af en forskerinitieret skriveprompt, som består af to opgaver. Skrivepromptens overskrift er ”Sundhedskrise i verden”. I den første opgave skal eleverne vælge mellem to fotografier og lave en beskrivelse af situationen på fotoet. Begge fotos relaterer sig til coronapandemien, som var højaktuel på dataindsamlingsstidspunktet. I den anden opgave skal eleverne fortælle om en typisk dag for dem derhjemme under den første nedlukning af skolerne i foråret 2020.

Som forberedelse til de kvantitative analyser blev teksterne konverteret til .txt-format. De gennemgik desuden en ortografisk standardisering, og syntaktiske perioder (fra punktum til punktum) blev splittet i helsætninger. Dette kan umiddelbart virke som en drastisk transformation af det rå datamateriale, men da formålet med undersøgelsen ikke er at belyse elevernes retskrivningsfærdigheder, vurderede vi, at det var acceptabelt indgreb, fordi det til gengæld forbedrede NLP-modellens performance, der ved de indledende afprøvninger fungerede dårligt på elevtekster præget af ortografiske afvigelser og manglende tegnsætning. Herefter foregik processeringen af teksterne i to skridt. Først blev de rensede og standardiserede tekster annoteret vha. SpaCy v. 3.5.0 (<https://spacy.io>), et NLP-modul til Python, der blev anvendt til at parse teksterne samt tilføje part-of-speech-tags og dependency-tags. Herefter blev forfelterne inklusive POS-tags og DEP-tags udtrukket i listeform for hver tekst vha. af en regelbaseret algoritme baseret på den automatiske annotering af teksterne. At bruge automatiserede metoder til præprocessering, annotering og dataudtræk er forbundet med en vis usikkerhed, idet der er tale om statistisk baserede, probabilistiske sprogmodeller, der kort sagt genererer kvalificerede gæt.

SpaCy-modulet er imidlertid et relativt velfungerende modul med en nøjagtighed for dansk på mere end 95% (for en sammenligning med andre alternativer se Enevoldsen et al. 2021).

På baggrund af dataudtrækket gennemførte vi dels en deskriptiv statistisk analyse med henblik på at sammenligne den gennemsnitlige forfælslængde samt andelen af lange forfæls for de to klassetrin, dels en række permutationstest med henblik på at fastslå om de gennemsnitlige forskelle mellem grupperne var signifikante. En permutationstest indikerer, hvor sandsynligt det er, at et resultatet ikke er et udtryk for tilfældighed. Vi valgte permutationstesten fordi den i modsætning til andre typer af hypotesetest, fx t-test, ikke forudsætter en antagelse om, at datamaterialet er normalfordelt. Signifikansniveauet er fastsat til $p < 0.05$, hvilket vil sige, at der er mindre end 5% sandsynlighed for, at den observerede forskel skyldes tilfældig variation.

Resultater fra de kvantitative analyser

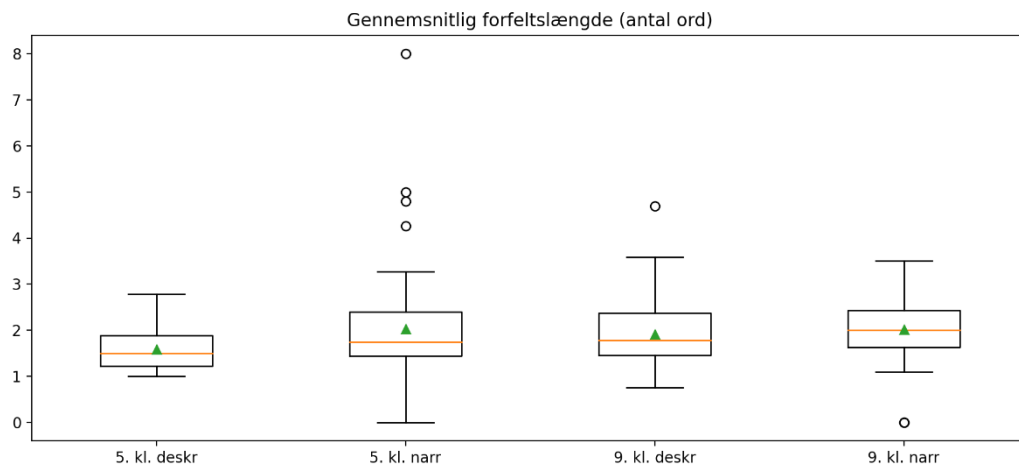
I den deskriptive statistiske analyse har vi fokuseret dels på teksternes gennemsnitlige forfælslængde målt på antal ord, dels på teksternes gennemsnitlige andel af lange forfæls målt som andelen af forfæls, der består af mere end to ord, relativt til det totale antal forfæls i teksten. Grænsen på to ord er valgt, fordi mere end to ord indikerer brugen af andet end pronomener og simple præpositionsforbindelser, der ellers præger mange af elevteksterne. Resultaterne af den deskriptive statistiske analyse er opsummeret i tabel 1.

Tabel 1

Oversigt over resultaterne af den deskriptive statistiske analyse

	Gns. 5. kl. deskriptiv		Gns. 5. kl. narrativ		Gns. 9. kl. deskriptiv		Gns. 9. kl. narrativ		Ratio 5. kl. deskriptiv		Ratio 5. kl. narrativ		Ratio 9. kl. deskriptiv		Ratio 9. kl. narrativ	
Antal	43		43		44		44		43		43		44		44	
Gns.	1,58		2,02		1,91		2,02		0,16		0,22		0,21		0,21	
SD	0,45		1,43		0,77		0,74		0,11		0,23		0,14		0,13	
Min.	1,00		0,00		0,75		0,00		0,00		0,00		0,00		0,00	
25%	1,23		1,44		1,46		1,62		0,09		0,11		0,11		0,12	
50%	1,50		1,75		1,78		2,00		0,14		0,17		0,21		0,20	
75%	1,88		2,40		2,37		2,44		0,25		0,27		0,33		0,29	
Max.	2,79		8,00		4,69		3,50		0,38		1,00		0,50		0,50	
IQR	0,65		0,96		0,91		0,81		0,16		0,16		0,23		0,16	

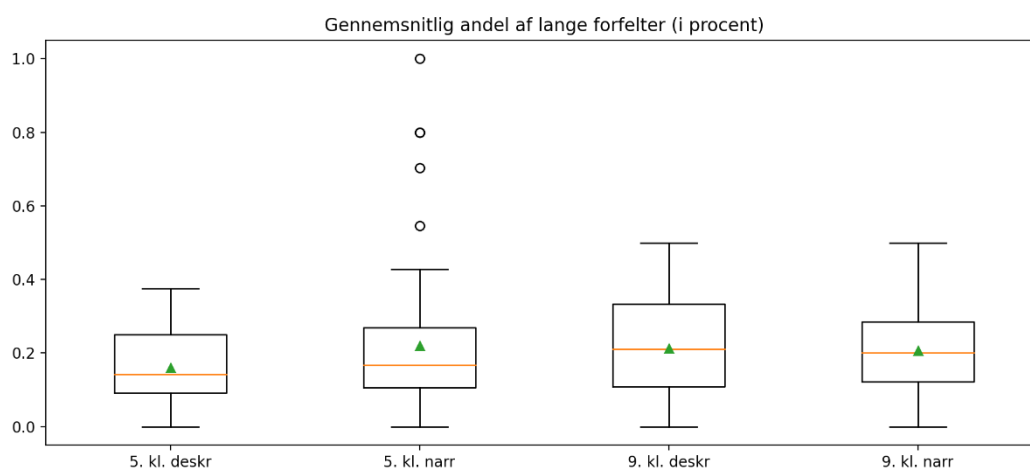
Den gennemsnitlige forfeltslængde (se figur 1) er 1,58 for 5D, 2,02 for 5N, 1,91 for 9D og 2,02 for 9D og svinger således mellem 1,58 som det laveste og 2,02 som det højeste gennemsnitlige antal ord per forfelt. Der er en relativt stor forskel mellem 5D og 5N på 0,44, mens der er en noget mindre forskel mellem 9D og 9N på 0,11 ord. Sammenlignes klassetrinnene, ser vi, at der er en forskel på 5D og 9D på 0,33, mens der ingen forskel er på 5N og 9N. I alle fire grupper er der stort set overensstemmelse mellem middelværdi og median. Dette indikerer sammen med gruppernes lave interkvartilbredde (se IQR i tabel 1), at middelværdierne i rimelig grad karakteriserer grupperne, fordi fordelingen ikke er skævvredet, og spredningen er relativt lille. Den eneste af grupperne, der afviger en lille smule fra dette mønster, er 5N, der, som det fremgår af figur 1, er præget af outlierere, hvilket resulterer i en forskel mellem middelværdi og median på 0,27.



Figur 1: Boxplot over fordelingen af gruppernes gennemsnitlige forfeltslængde.

Ser vi herefter på andelen af lange forfelter, tegner der sig et lignende billede. Den gennemsnitlige andel af lange forfelter (se tabel 1) er 16% for 5D, 22% for 5N, 21% for 9D og 21% for 9D og svinger således mellem 16% som den laveste og 22% som den højeste gennemsnitlige andel af lange forfelter. Som det var tilfældet med den gennemsnitlige forfeltslængde, er der en vis forskel på 5D og 5N og mellem 5D og 9D. Til gengæld er der stort set ingen forskel mellem de øvrige grupper. Som det også var tilfældet med den gennemsnitlige forfeltslængde, gælder det for gruppernes gennemsnitlige andel af lange forfelter, at der, bort set fra i 5N der er præget af outlierere, stort set er overensstemmelse mellem median og middelværdi samt en relativt lille

spredning (se figur 2).



Figur 2: Boxplot over fordelingen af gruppernes gennemsnitlige andel af lange forfælder.

På baggrund af den deskriptive statistiske analyse kan det være vanskeligt at vurdere, i hvilket omfang de observerede forskelle er udtryk for reelle forskelle eller blot for tilfældig variation. Vi gennemførte derfor en række permutationstest med henblik på at afgøre, om der er tale om signifikante forskelle. Som det fremgår af tabel 2 (signifikante resultater er markeret med *), er der signifikante forskelle mellem den gennemsnitlige forfælslængde mellem 5D og 5N ($p=0.02$) og mellem 5D og 9D ($p=0.01$). Sammenligner vi gruppernes gennemsnitlige andel af lange forfælder, viser det sig, at der er en signifikant forskel mellem 5D og 9D ($p=0.03$) samt en forskel mellem 5D og 5N (markeret med +), der ikke er signifikant, men dog ligger meget tæt på det fastsatte signifikansniveau ($p=0.07$). De øvrige test viste ingen signifikante forskelle.

Tabel 2

Oversigt over resultaterne af de gennemførte permutationstest

Klasse/teksttype	Forskel i gennemsnitlig forfæls- længde (antal ord)				Forskel i gennemsnitlig andel af lange forfælder (procent)			
	5D/5N	9D/9N	5D/9D	5N/9N	5D/5N	9D/9N	5D/9D	5N/9N
Forskel	0,43	0,11	0,32	0,00	0,06	0,01	0,05	0,01
p-værdi	0,02*	0,24	0,01*	0,50	0,07+	0,58	0,03*	0,6

(* = $p < 0.05$ / + = $p < 0.1$)

De kvantitative analyser giver ikke i sig selv svar på, hvordan de observerede forskelle skal fortolkes. Selvom visse af resultaterne er signifikante, er der ikke tale om store forskelle i absolute tal. Det er heller ikke klart på baggrund af de statistiske beskrivelser af datamaterialet, om

det er klassetrin eller genre, der er årsag til de observerede forskelle. For 5D gælder, at den både er forskellig fra 5N, hvilket kunne indikere en genrebetinget forskel, og forskellig fra 9D, hvilket kunne indikere en klassetrinsrelateret forskel. På den anden side er der ingen forskel på hverken 9D og 9N eller på 5N og 9N, hvilket ikke undersøger disse fortolkniger.

For at komme nærmere på, hvad forskellene dækker over samt på at forstå den interne variation i grupperne, kaster vi i det følgende et kvalitativt blik på seks tekster, der er udvalgt med henblik på at belyse, hvilke tekstlige strukturer der ligger bag resultaterne af den kvantitative beskrivelse af korpusset.

Et kvalitativt blik på seks elevtekster

Tekstudvalget fordeler sig sådan, at der er tre tekster fra hver af de to teksttyper, hvoraf én tekst har korte forfelter, én har en gennemsnitlig forfeltslængde omkring medianværdien for årgangen, og én tekst har en forfeltslængde i tredje/fjerde kvartil for årgangen. Af pladsmæssige hensyn er ikke alle elevtekster gengivet i deres fulde længde. Forfelterne i hver tekst er markeret ved, at der efter hvert forfelt er en kantet parentes, hvori det er angivet, hvilket sætningsled der er i forfeltet. Der er kodet for subjekt [S], direkte objekt [DO], indirekte objekt [IO], adverbial [A], formelt/foreløbigt subjekt [FS], interrogativt pronomen [IP] og for tomme forfelter [T].

Tre berettende elevtekster

Vi starter med at se nærmere på tre berettende tekster. Den første tekst er skrevet af en elev fra 5. årgang. Teksten har en gennemsnitlig forfeltslængde på 0,9. Det vil sige, at teksten er placeret i første kvartil for årgangen. Teksten er gengivet i sin helhed i tabel 3.

Tabel 3		
<i>Skematisk oversigt over berettende elevtekst (5. årgang)</i>		
Interjektion/ konjunktion	Forfelt	Centralfelt/slutfelt
	Nu [A]	vil jeg fortælle om en dag ligesom alle dage
men	[T]	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	[T]	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	[T]	
og	[T]	sov godt.

Teksten består af i alt seksten forfelter. Syv af disse består af forskellige typer af adverbialer, hvoraf de fleste er enkeltstående tidsadverbialer som 'nu', 'efter' og 'derefter'. Eleven anvender således disse adverbialer til at sekvensere tekstens handlingssekvenser kronologisk. Der er fem forfelter med subjekt, hvoraf fire af dem er enkeltstående pronominer, og blot et enkelt forfelt har et nominalsyntaxme med en foranstillet bestemmer ('min mor'). Der er desuden fire tomme forfelter ved imperativformer. At teksten er endt med en gennemsnitlig forfeltslængde på blot 0,9, skyldes således, at der er ni forfelter med ét ord, tre med to ord og fire tomme forfelter. Teksten indeholder således en smule variation i sætningsled i forfeltet, idet der veksles mellem subjekt og adverbielle realiseringer, mens der til gengæld ikke er megen variation i længden og den grammatiske kompleksitet i forfelterne.

Den næste elevtekst, vi skal se på, er skrevet af en elev på 9. årgang. Teksten har en gennemsnitlig forfeltslængde på 2,1, hvilket placerer teksten lige lidt over medianen for årgangen. Teksten er forholdsvis lang (391 ord), så i tabel 4 nedenfor er der blot gengivet et uddrag af den første halvdel af teksten.

Tabel 4		
Skematisk oversigt over berettende elevtekst 2 (9. årgang)		
Interjektion/ konjunktion	Forfelt	Centralfelt/slutfelt
	Da vi skulle starte med at have skole hjemmefra [A],	stod jeg op meget sent. Omkring kl 10 eller 11.

Men	jeg [S]	fandt hurtigt ud af, at jeg skulle stå tidligere op.
	Ellers [A]	ville jeg ikke have en chance for at nå at lave mine lektier, før det blev sent.
	Derfor [A]	begyndte jeg at stå op omkring 8-10 tiden.
	Jeg [S]	ville spise morgenmad, glemme at børste tænder og tjekke hvad dagens lektier byder på.
	I nogle tilfælde [A]	tog jeg ud og løb som det første nogle morgener før morgenmad,
og	[T]	gik derefter i bad efterfølgende.
	Lektierne [S]	tog det meste af dagen,
men	jeg [S]	snakkede nogle gange med mine venner, som hurtigt kunne hjælpe mig.
	Vi [S]	havde også møder nogle gange, som gjorde det lidt mindre kedeligt, og sjovere fordi jeg fik lov at se mine klassekammerater igen.
	Som sagt [A]	tog jeg ud at løbe nogle gange,
men	jeg [S]	dansede også Just Dance
eller	[T]	gik ture med min mor [...]

Den samlede tekst består af niogtyve forfelter. Ni af disse er adverbialer enten i form af enkeltstående adverbialer eller adverbielle ledsætninger. Der er tre tomme forfelter ved ellipsekonstruktioner, mens de resterende sytten forfelter består af subjekt. Den væsentligste årsag til, at vi i denne tekst ser mere end en fordobling af den gennemsnitlige forfeltslængde sammenlignet med den forrige tekst, er, at der er flere eksempler på lange adverbialer i forfeltet, fx ledsætninger som "Da vi skulle starte med at have skole hjemmefra" eller tidsadverbielle udtryk som "i nogle tilfælde". Det er altså i høj grad i længden og variationen af de adverbielle udtryk, at man ser de største forskelle på de to berettende elevtekster. Vi ser ligeledes enkelte eksempler på nogle andre typer af adverbialer i forfeltet i denne tekst, som i højere grad er knyttet til betingelser og årsagsforhold i teksten (fx 'ellers' og 'derfor'), hvorimod de i forrige tekst forholdsvis entydigt var knyttet til kohæsion gennem temporale sætningskoblinger. Der er til gengæld ikke forskel på de sætningsled, som fremgår af forfeltet i de to tekster. Begge steder varieres der mellem forskellige adverbielle udtryk og subjektsudfyldninger med enkeltstående personlige pronominer.

Den sidste berettende elevtekst, vi skal se på, er ligeledes skrevet af en elev på 9. årgang. Teksten har en gennemsnitlig forfeltslængde på ca. 3,1, hvilket placerer teksten i fjerde og øverste kvartil på årgangen. Som ved den forrige tekst er der af pladsmæssige hensyn kun gengivet et uddrag af teksten (se tabel 5).

Tabel 5		
<i>Skematisk oversigt over berettende elevtekst 3 (9. årgang)</i>		
Interjektion/ konjunktion	Forfelt	Centralfelt/slutfelt
	En typisk hverdag hjemme hos mig under coronanedlukningen [S]	ville cirka foregå følgende.
	Jeg [S]	står op klokken 6:30 (nogle gange også klokken 7:00), som jeg normalt også ville have gjort, hvis vi havde været i skole som normalt.
	Så [A]	går jeg i bad,
	[T]	tager tøj på
	[T]	spiser morgenmad.
	Så [A]	er den normale morgenrutine egentligt overstået.
	Nu [A]	kommer vi til alt det, som har ændret sig betydeligt meget ved min hverdag i forhold til før.
	I stedet for at tage i skole på det her tidspunkt [A],	går jeg ovenpå, op på mit værelse med en kop te og
	[T]	starter computeren.
	Først [A]	kigger jeg lektierne igennem
og	[T]	begynder måske på en enkelt individuel opgave vi har fået for.
	Når klokken bliver cirka 8:30 [A]	kontakter jeg min studiegruppe [...]

Teksten består af i alt sytten forfelter, hvoraf hele ti af dem består af forskellige adverbielle led. Tre forfelter består af subjekt og fire forfelter er tomme, idet de forekommer ved ellipsekonstruktioner. Der er altså en større andel af adverbielle led i forfeltet i denne elevtekst end i den forrige, og ser man nærmere på disse adverbielle led, er det tydeligt, at det især er disse led, der er med til at øge forfeltslængden. Det kommer bl.a. til udtryk ved tidsadverbielle ledsætninger som ”Når klokken bliver cirka 8:30” og længere fraser som ”I stedet for at tage i skole på det her tidspunkt” og ”Så hver morgen”. I tekstens første sætning er der ligeledes et meget langt nominalsyntaxme på otte ord (”En typisk dag hjemme...”) i forfeltet, som er med til at øge den gennemsnitlige forfeltslængde i teksten. Det, der adskiller denne elevtekst, fra de to forrige er således, at der er flere eksempler på meget lange forfelter i teksten blandet med korte forfelter, som typisk består af enkeltstående pronominer eller tidsadverbialer.

Tre beskrivende elevtekster

Vi vender nu blikket mod tre beskrivende elevtekster. Den første elevtekst har en gennemsnitlig forfeltslængde på 1,0 og er skrevet af en elev på 5. klassetrin. Teksten er gengivet i sin helhed i tabel 6.

Tabel 6 <i>Skematisk oversigt over beskrivende elevtekst 1 (5. årgang)</i>		
Interjektion/ konjunktion	Forfelt	Centralfelt/slutfelt
	Jeg [S]	ser nogle piger der godt ville lege også selvom der er corona.
	De [S]	er i byen næsten helt alene.
	Der [FS]	er ingen mennesker på restauranterne i baggrunden.
	De [S]	er på vej ud at lege i springvandet.
	Jeg [S]	tror det er ved at være forår fordi pigerne ikke har jakker på.
	Jeg [S]	tror ikke det er så varmt
	det [FS]	er heller ikke isnende koldt.
	Det [FS]	er nok sådan lidt lundt.
	De [S]	er nok glade men kede af at de skal holde afstand og bære maske.

Af de tekster, vi har set på indtil videre, er denne tekst den med mindst variation både i forhold til længde og sætningsled. Alle ni forfelter består af præcis ét ord, og alle forfelter har enten subjekt eller formelt/foreløbigt subjekt i forfeltet. Flere af sætningerne i teksten indeholder imidlertid andre former for syntaktisk kompleksitet såsom efterstillede ledsætninger eller indskudte sætninger, så et kort forfelt er i denne tekst ikke det samme som, at teksten har en simpel syntaks eller korte syntaktiske perioder. Eleven anvender forfelterne til dels at positionere sig selv som tekstens fortæller, der kommenterer og vurderer forhold på billedet, dels til deiktisk at referere til pigerne på billedet. Den tilsvarende berettende elevtekst, som havde en gennemsnitlig forfeltslængde på 1,0, havde en større variation i sætningsled, idet den narrative form i den berettende teksttype både kræver en sammenhænge temporal struktur og en aktivt agerende og vurderende fortæller. I denne beskrivende tekst kan eleven til gengæld uproblematisk udføre skriveordren uden variation i, hvilket sætningsled vedkommende placerer i forfeltet.

Den næste beskrivende elevtekst, vi skal se på, har en gennemsnitlig forfeltslængde på 1,3, som altså er en anelse højere end den forrige tekst og samtidig er placeret i andet kvartil

lige under medianværdien. Teksten er ligeledes skrevet af en elev på 5. klassetrin. Teksten er gengivet i sin helhed i tabel 7.

Tabel 7		
<i>Skematisk oversigt over beskrivende elevtekst 2 (5. årgang)</i>		
Interjektion/ konjunktion	Forfelt	Centralfelt/slutfelt
	Jeg [S]	har valgt billede nummer 1.
	Her [A]	ser man to piger på cirka 5 og 6 år stå midt ude på en hel tom gågade.
	Der [FS]	er ikke nogle mennesker fordi at corona eller covid 19 hersker verden over.
	Det [FS]	er en ny helt ukendt virus
og	forsker hele verden over [S]	kæmper på livet løs for at få en vaccine.
	Der [FS]	er normalt liv i gaderne
men	nu [A]	er virkeligheden sat på pause.
	De [S]	har visir på for at beskytte sig selv og andre.
	De [S]	står inde på torvet uden for en restaurant.
	Det [FS]	ligner at der står at restauranten hedder Janne.
	Den ene af pigerne [S]	sidder på en sten.
	De [S]	har nok vejret hjemsendt i et lille stykke tid.
	Nu [A]	er de blevet træet af at være indenfor
og	[T]	våger sig ud på gaden.
	Jeg [S]	håber at det snart er over.
	Det [S]	bliver nok først når de har fundet en vaccine.

Teksten indeholder i alt seksten forfelter, hvoraf tolv består af subjekt eller foreløbigt/formelt subjekt, tre består af et adverbialt led og et enkelt forfelt er tomt ved en ellipse. Anvendelsen af adverbialle led giver således en anelse større variation i udfyldningen af forfeltet i denne tekst sammenlignet med den forrige. Det er dog stadig subjekt og formelt/foreløbigt subjekt, der i overvejende grad er de dominerende sætningsled i forfelterne. Ser man på længden af forfelterne, består tretten af dem kun af ét ord. Det, der øger tekstens gennemsnitlige forfeltslængde

sammenlignet med forrige tekst, er således to nominalsyntagmer på fire ord ('forsker hele verden over' og 'Den ene af pigerne'). Teksten har desuden generelt en større syntaktisk kompleksitet forstået på den måde, at vi ser større grad af variation mellem parataktiske og hypotaktiske sætningskonstruktioner, hvilket altså også viser sig ved, at eleven formår at anvende mere komplekse nominalsyntagmer i forfeltet.

Den tredje og sidste beskrivende elevtekst har en gennemsnitlig forfetslængde på hele 3,0 og er skrevet af en elev på 9. klassetrin. Teksten er forholdsvis lang, så i tabel 8 er blot gengivet et uddrag af første halvdel af teksten.

Tabel 8		
<i>Skematisk oversigt over beskrivende elevtekst 3 (9. årgang)</i>		
Interjektion/ konjunktion	Forfelt	Centralfelt/slutfelt
	Jeg [S]	har valgt billede nr. 2.
	Grunden til at jeg har valgt billede nr. 2 [S]	er fordi, at jeg føler at der er mere betydning i det billede end det billede nr. 1.
	Selvfølgelig [A]	kan jeg ikke lade vær med at referere til corona krisen, som er det der hovedsageligt er fokus på i medierne for tiden.
	Først og fremmest [A]	vil jeg beskrive miljøet og hvordan stedet ser ud.
	Der hvor billedet er taget, [S]	kunne godt ligne en opgang fra en togstation eller et sted inde i en by, hvor der er trafikeret.
	Der [FS]	er mange mennesker - især mange UDEN mundbind i det her offentlige område.
	Helt bagerst i baggrunden [A]	er der nogle træer samt en flok af mennesker.
	På bebyggelserne [A]	er der skilte, som kan tyde på at der er et offentligt sted inde i byen.
	På billedet [A]	ser man en mand i højre side, der ryger en cigaret.
	Det [FS]	ligner at manden er fra en middelklasse, da hans tøj ligner noget, man kunne se fra en normal hverdag på et kontorarbejde.
	Men selvom at det tyder på at hans fornuft er i top og at	kan man stadig ikke lade vær med at se cigaretten han ryger[...]

	han har et ordentligt arbejde [A],	
--	------------------------------------	--

Teksten består af i alt enogtyve forfelter, hvoraf tretten indeholder subjekt eller formelt/foreløbigt subjekt, syv indeholder et adverbialt led og kun et enkelt forfelt er tomt ved en ellipse. Elleve af forfelterne består blot af et enkelt ord (typisk et pronomen), så den øgede gennemsnitlige forfeltslængde i denne tekst sammenlignet med de forrige skyldes især, at der er enkeltteksempler på meget lange forfelter. Disse forfelter indeholder typisk enten nominalsyntagmer med indbyggede relativsætninger (fx ”Der hvor billedet er taget”) eller længere adverbialle fraser med deiktisk funktion (”Helt bagerst i baggrunden” eller ”Ud af alle dem omkring hende”), der peger ud på selve billedet på skriveordren. Tekstens længste forfelt består af atten ord og består af hele tre ledsætninger (”... selvom at det tyder på at hans fornuft...”), hvilket viser, at teksten således har en meget stor spredning og variation i længden af forfelterne. Når den gennemsnitlige forfeltslængde øges, tyder det altså på, at det skyldes enkelte meget lange og komplekse forfelter snarere end en generel øgning i længden af forfelterne generelt i teksten. De korte forfelter med enkeltstående pronomener eller adverbier er således dominerende på tværs af alle de elevtekster, vi har set på.

Forfeltet og elevers skriveudvikling

At det netop er forfeltet, som har været genstand for denne undersøgelse, er som anført i indledningen ikke et tilfælde. Forfeltet kan indeholde stort set alle sætningsled og har som følge heraf en vigtig funktionel rolle i de syntaktiske valg, som eleverne foretager i skriveprocessen. De kvantitative og kvalitative analyser peger samlet set på, at der findes to typer af forfeltsvariation, og at disse to typer af variation kan fungere uafhængigt af hinanden og på forskellig vis gøre os klogere på elevers skriveudvikling. Den første type af variation handler om variation i det sætningsled, som placeres i forfeltet. Her indikerede de kvalitative analyser, at det i høj grad er teksttypen og skriveordrens udformning, der afgør denne type af variation. I de beretende tekster så vi fx, at den narrative fortællestruktur i teksterne gjorde, at der ofte var tidsadverbialer i forfeltet, mens der omvendt ofte var subjektsudfyldninger i de beskrivende tekster, idet eleverne enten anvendte forfeltet til at positionere sig selv som fortællere (fx ”Jeg ser nogle piger...”) eller til deiktisk at pege ud på forskellige elementer af det billede, de skulle beskrive (”De står inde på torvet...”). Der var dog også eksempler på adverbialle led i forfeltet i de beskrivende tekster, men også her havde de i høj grad en deiktisk funktion, idet de pegede ud og identificerede de områder af billedet på skriveordren, som den pågældende sætning beskrev

(fx ”Helt bagerst i baggrunden er der nogle træer...”). Det er altså tydeligt ved de beskrivende tekster, at skriveordren sætter eleverne i en bestemt kommunikationssituation, hvor de eksplicit skal forholde sig til et billede, og her er det i høj grad forfeltet, som eleverne anvender til at løse den specifikke kommunikative udfordring, det er at henvise til et billede på en eksternt givet skriveordre.

Den anden type af forfeltsvariation er knyttet til kompleksiteten og længden af forfelterne. Her viste de kvantitative analyser, at der var statistisk signifikante forskelle mellem 5. og 9. klassetrin på både forfeltslængde og ratioen af lange forfelter i de beskrivende tekster, mens der omvendt ikke var signifikant forskel mellem de to klassetrin, når man sammenlignede de berettende tekster. Fra et rent kvantitativt og strukturelt perspektiv ligner de to grupper af berettende tekster altså hinanden. At der til gengæld er signifikante forskelle ved de beskrivende tekster, kan skyldes flere ting. Der er eksempelvis tegn i de kvalitative analyser, der indikerer, at nogle elever omfortolker den beskrivende teksttype og gør den til en mere analyserende og vurderende tekst, hvor eleven fx kommer med begrundelser for valg af billede (fx ”Grunden til at jeg har valgt billede nr. 2 er fordi...”) eller kommer med mere vurderende udsagn om forhold på billedet (fx ”Men selvom at det tyder på at hans fornuft er i top og at han har et ordentligt arbejde, kan man...”). Denne fortolkning af skriveordren gør altså, at nogle elever kan bruge forfeltet til mere komplekse syntaktiske forhold som fx at opstille betingelsesledsætninger eller anvende mere vurderende adverbialle udtryk. Hvor variation i sætningsleddet i høj grad er noget, der afgøres af teksttypen og skriveordren, er der i de kvalitative analyser noget, der tyder på, at variation i kompleksitet og længde af forfeltet er noget, der kendetegner mere erfarne skrivere. I hvert fald er det primært i tekster fra 9. årgang, at vi ser eksempler på meget lange forfelter. De kvalitative analyser viser således, at det i høj grad er enkeltteksempler på meget lange forfelter, der får den gennemsnitlige forfeltslængde til at stige, snarere end det er en generel kvantitetsforøgelse af alle forfelterne i teksten. Det prototypiske forfelt på tværs af alle de kvalitativt analyserede tekster består således af ét enkelt ord (oftest et pronomen eller et tidsadverbium), og det, der adskiller tekster med forskellig gennemsnitlig forfeltslængde, er således i hvilken grad disse prototypiske forfelter varieres i sammenhæng med enkeltteksempler på lange forfelter med længere nominalsyntagmer, tidsadverbialer eller ledsætninger.

Fra et forfelperspektiv kommer skriveudvikling således ikke entydigt til udtryk som gennemsnitligt set mere kvantitet over tid, men som en funktionel evne til at variere og tilpasse forfeltet til teksttype og skriveordre. Med andre ord er variabilitet, dvs. evnen til at tilpasse skriftsproget til situationen, et tegn på skriveudvikling, idet denne type af adaptive adfærd er funktionelt og kommunikativt betinget (Fogal 2020). Der er således ikke noget i dette studie,

der indikerer, at bestemte længder eller typer af forfelter er udtryk for bedre skrivefærdigheder eller mere komplekse tekster end andre, ligesom det heller ikke kan udledes af analyserne, at lange forfelter i sig selv er bedre end korte. Analyserne bekræfter således fundene fra andre studier af syntaktisk udvikling i elevtekster, der viser, at skriveudvikling ikke kommer til udtryk som en lineær og kontinuerligt øget længde af en bestemt syntaktisk struktur, men som evnen til at variere mellem forskellige former for grammatisk kompleksitet i teksterne som svar på funktionelt betingede forhold i skrivesituationen eller skriveordren (se fx Crossley et al. 2014; Hardy & Römer 2013; Tannert 2022), og her er forfelter et af flere steder, hvor denne adaptive evne kommer til udtryk i teksterne.

Afsluttende betragtninger

Skriveudvikling, som det kommer til udtryk i grundskolen, er et komplekst fænomen, og det kan som følge heraf studeres på forskellige niveauer og fra mange forskellige metodiske perspektiver. I denne artikel har vi undersøgt en lille afgrænset del af dette felt ved at sætte særskilt fokus på, hvordan elever anvender og varierer forfelter på tværs af to klassetrin og to teksttyper. Ved blandt andet at fokusere på betydningen af forfelterlængde har vi forsøgt at bidrage til den voksende gruppe af forskningslitteratur, der undersøger tekster ved hjælp af automatiserede og digitale analyseværktøjer. En væsentlig metodologisk pointe i forskningslitteraturen er, at selv om længdebaserede mål af forskellige sproglige træk i elevtekster godt kan identificere udviklingstræk i elevernes tekster og ofte kan være med til at forudsige, hvilken karakter en tekst vil få (dvs. har en høj prædiktiv validitet), så kan de være vanskelige at fortolke lingvistisk (se fx diskussion heraf hos Norris & Ortega 2009). I denne artikel har vi derfor suppleret de kvantitative analyser med kvalitative analytiske nedslag i seks elevtekster for at få en dybere og grammatisk detaljeret analyse og forståelse for, hvordan eleverne anvender forfelter. Et væsentligt fokuspunkt fremadrettet er imidlertid at forskningsfeltet for automatiseret tekstanalyse udvikler nye og mere detaljerede kvantitative mål for forskellige dele af elevernes tekster. Det svarer til det, som Biber et al. (2020) omtaler som en bevægelse fra prædiktive mål til mål af grammatisk kompleksitet. For eksempel er det i denne artikels analyser blevet klart, at teksttype og skriveordre har stor betydning for, hvilket sætningsled eleverne placerer i forfelter, idet skriveordren tvinger eleverne til at anvende forfelter deiktisk, mens den berettende teksttype stiller krav til, at forfelter særligt anvendes til at skabe kohæsion i teksten gennem temporale sætningskoblinger. Tilsvarende kan man i forlængelse af Togeby (2014:146-150) forestille sig, at forskellige former for grammatisk kompleksitet i forfelter varierer systematisk relativt til kompositions-mønstrene i en tekst – altså at bestemte sætningsled og kompleksitetsformer er knyttet til be-

stemte dele af en teksts generelle sagsfremstilling. Disse forhold kan således være fokuspunkter i fremtidige studier af forfeltsvariation i elevtekster fra grundskolen.

Litteratur

- 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>.
- 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>.
- 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., Kabel, K., Liberg, C., & Skar, G. B. (2021) A review of Scandinavian writing research between 2010 and 2020. *Writing and Pedagogy*, 13(1–3), 7–49. <https://doi.org/10.1558/wap.21637>.
- 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.
- 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., & McNamara, D. S. (2014) Does writing development equal writing quality? A computational investigation of syntactic complexity in L2 learners. *Journal of Second Language Writing*, 26, 66–79. <https://doi.org/10.1016/j.jslw.2014.09.006>.
- 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>.
- Diderichsen, P. (1946) *Elementær dansk grammatik*. Gyldendal.
- Diderichsen, P. (1966) Prolegomena til en metodisk dansk Syntax. I *Helhed og struktur: Udvalgte sprogvidenskabelige afhandlinger*, s. 21–24. Gads Forlag.
- Eisenstein, J. (2019) *Introduction to natural language processing*. The MIT Press.
- Enevoldsen, K., Hansen, L., & Nielbo, K. (2021) DaCy: A unified framework for Danish NLP. *Ceur Workshop Proceedings*, 2989, 206–216.
- 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>.

- Garne, B. (1988) *Text och tanke: Om skrivstrategier i elevuppsatser*. Liber.
- Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011) Coh-Metrix: Providing Multilevel Analyses of Text Characteristics. *Educational Researcher*, 40(5), 223–234.
<https://doi.org/10.3102/0013189X11413260>.
- 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>.
- Hirst, G. (2013) Computational Linguistics. I K. Allan (Red.), *The Oxford Handbook of The History of Linguistics* (s. 707–726). Oxford University Press.
- Hultman, T. G., & Westman, M. (1977) *Gymnasistsvenska*. Liber Läromedel.
- Josephson, O., Melin, L., & Oliv, T. (1998) *Elevtext: Analyser av skoluppsatser från åk 1 till åk 9*. Studentlitteratur.
- Jørgensen, H. (2000) *Indføring i Dansk Syntaks* (3. udg.). Århus Universitet.
- Kabel, K., Bremholm, J., Bundsgaard, J., & Berthelsen, U. D. (2020) *Elevers tidlige skriveudvikling: Identificerede trin* (Working paper i ATEL). DPU, Aarhus Universitet.
- Larsson, K. (1984) *Skrivförmåga: Studier i svenskt elevspråk*. Liber Förlag.
- 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.
- 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/oxford-hb/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., Crossley, S. A., & Roscoe, R. (2013) Natural language processing in an intelligent writing strategy tutoring system. *Behavior Research Methods*, 45(2), 499–515. <https://doi.org/10.3758/s13428-012-0258-1>.
- McNamara, D. S., Graesser, A. C., McCarthy, P. M., & Cai, Z. (2014) *Automated Evaluation of Text and Discourse with Coh-Metrix*. Cambridge University Press. <https://doi.org/10.1017/CBO9780511894664>
- 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>.
- 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>.
- Perin, D., & Lauterbach, M. (2018) Assessing Text-Based Writing of Low-Skilled College Students. *International Journal of Artificial Intelligence in Education*, 28(1), 56–78. <https://doi.org/10.1007/s40593-016-0122-z>.
- 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.
- 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>.

- Tannert, M. (2022) *Skriftsproglig udvikling i grundskolens danskfag: Et korpusbaseret studie af elevtekster fra 5.-9. klassetrin*. Ph.d.-afhandling. DPU, Aarhus Universitet. <https://doi.org/10.7146/aulsps-e.466>.
- Togeby, O. (2014) *Bland blot genrene—Ikke teksterne! Om sprog, tekster og samfund*. Samfundslitteratur.
- 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>.
- 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>