

De ontwikkeling van syntactische complexiteit in het schrijfproces

Een longitudinaal onderzoek naar de ontwikkeling van de syntactische complexiteit in Nederlandstalige teksten geschreven door middelbare scholieren.

Pim Klaassen

S4853156

30 november 2021

Bachelorscriptie

Begeleider: Prof. dr. J.J.M. Schoonen

Radboud Universiteit



Voorwoord

Voor u ligt de scriptie ‘De ontwikkeling van syntactische complexiteit in het schrijfproces’, waarin de rol en de ontwikkeling van enkele grammaticale componenten is onderzocht met betrekking tot de grammaticale complexiteit van teksten. Deze scriptie is geschreven als onderdeel van mijn afstuderen aan de bacheloropleiding Taalwetenschap aan de Radboud Universiteit te Nijmegen.

Het schrijven van deze scriptie is niet eenvoudig geweest. Gedurende het gehele proces heeft de coronapandemie helaas de voortgang van de scriptie beïnvloed, van de dataverzameling tot aan de laatste loodjes van het schrijven. Na enkele tegenslagen en veel hulp vanuit mijn omgeving en mijn begeleider is het dan toch nog gelukt om de scriptie tot een goed einde te brengen.

Voor alle hulp wil ik ontzettend graag mijn dank uitspreken voor mijn scriptiebegeleider, Rob Schoonen. Ondanks de situatie rondom het coronavirus heeft hij me veel ondersteuning en hulp kunnen bieden bij het schrijven van mijn scriptie. Niet alleen tijdens het bespreken van de inhoud of verwoordingen in de scriptie, maar vooral ook door de aanmoedigingen om de motivatie en het doorzettingsvermogen te vinden om de scriptie af te ronden. Zo is deze scriptie uiteindelijk toch nog tot een goed einde gebracht.

Pim Klaassen

Nijmegen, 30 november 2021

Inhoudsopgave

Inhoudsopgave

Voorwoord	I
Inhoudsopgave	II
Samenvatting	1
Inleiding	1
Schrijfvaardigheid en grammatica.....	1
Meten van complexiteit: de T-unit	2
Grammaticale complexiteitskenmerken	5
Het huidige onderzoek.....	7
Methode.....	9
Onderzoeksontwerp.....	10
Participanten.....	10
Materiaal.....	11
Procedure.....	13
Instrumenten.....	13
Data-analyse	15
Resultaten	16
De lengte van de T-units	16
Het aantal bijzinnen per T-unit.....	17
Het aantal bijwoordelijke bepalingen per deelzin	18
Het aantal NP's per T-unit.....	19
Het aantal woorden per NP	20
Het aantal bijvoeglijke bepalingen per deelzin.....	21
Het aantal attributieve bijvoeglijke bepalingen per deelzin	22
Correlaties	23

Samenvatting resultaten	25
Discussie en conclusie.....	26
Discussie.....	26
Conclusie.....	30
Literatuurlijst.....	31
Bijlage A	33

Samenvatting

Geschreven teksten verschillen in de mate van grammaticale complexiteit. Niet alleen bestaan er verschillen in complexiteit tussen tekstgenres, maar ook tussen schrijvers, bijvoorbeeld tussen verschillende leeftijdscategorieën. Uit eerder onderzoek van Staples et al. (2016) is gebleken dat grammaticale complexiteit in verschillende vormen voor kan komen, namelijk op zinsniveau en op naamwoordelijk niveau. Uit een vergelijking van teksten van studenten bleek dat beginnende studenten grammaticaal anders schrijven dan gevorderde studenten. Gevorderde studenten schreven complexer op naamwoordelijk niveau. In het huidige onderzoek zijn leerlingen van de onder- en bovenbouw van het vwo, de middelbare school, met elkaar vergeleken op basis van het gebruik van grammaticale complexiteitskenmerken op naamwoordelijk dan wel zinsniveau. De naamwoordelijke complexiteitskenmerken bleken nog vrijwel geen rol te spelen in de complexiteit van de door de leerlingen geproduceerde teksten. Twee kenmerken op zinsniveau leken zich wel te ontwikkelen in het schrijfproces van de scholieren.

Inleiding

Schrijfvaardigheid en grammatica

Een goede academische schrijfvaardigheid is belangrijk voor studenten, zowel in hun studieloopbaan als tijdens hun latere carrière. Tijdens de studie krijgen studenten veelvuldig te maken met uitgebreide schrijftaken. Universitaire docenten hechten daarom steeds veel belang aan een goede schriftelijke vaardigheid van studenten. De overtuiging is dat een goed ontwikkelde schrijfvaardigheid zorgt voor kwalitatief betere teksten. De kwaliteit van een geschreven tekst hangt af van verschillende onderdelen, zoals het gebruikte lexicon in een tekst en de syntaxis, oftewel de grammaticale vorm, van de tekst (Lu, 2011). Een goede beheersing en toepassing van de verschillende factoren zou voor kwalitatief sterke teksten moeten zorgen. De vorm en structuur van de zinnen van een tekst, oftewel de grammatica, lijken dus een belangrijke indicatie van de kwaliteit van een tekst te zijn.

Chung en Pullum (z.j.) definiëren het concept grammatica, binnen de taalwetenschap, als een verzameling regels die begrenst op welke manier woorden en zinnen gevormd kunnen worden in een taal. Syntaxis is een subonderdeel van de grammatica dat specifiek gaat over de opbouw van zinnen (Chung & Pullum, z.j.). In dit onderzoek worden ‘grammatica’ en ‘syntaxis’ als hetzelfde beschouwd. Talen, en dus ook teksten, hebben altijd te maken met de grammatica van de taal waarin het is geproduceerd. Dezelfde grammaticale regels gelden voor teksten in dezelfde taal, maar kunnen toch verschillende vormen aannemen. Met name de

complexiteit van de grammatica kan verschillen tussen teksten. Zelfs binnenin dezelfde tekst kunnen grote verschillen bestaan. De mate van complexiteit roept al snel associaties op met de kwaliteit, waarbij een grammaticaal complexere tekst, met bijvoorbeeld een veelvoud aan ingebedde zinnen, waarschijnlijk als een ‘betere’ tekst zal worden beschouwd dan een tekst die slechts uit hoofdzinnen bestaat. Geschreven teksten van een middelbare scholier zullen bijvoorbeeld in de meeste gevallen grote verschillen tonen met teksten van een hoogleraar wanneer de twee teksten eenzelfde onderwerp zouden behandelen. Omdat de kwaliteit van een tekst deels beïnvloed lijkt te worden door de grammaticale complexiteit ervan, is het meten van de grammaticale complexiteit bijzonder nuttig. Als de grammaticale complexiteit van geschreven teksten gemeten kan worden, zou op basis van een assessment een voorspelling of diagnose opgesteld kunnen worden over de vorderingen in de schrijfvaardigheid van een schrijver. Op basis van die gegevens zou dan weer specifieke hulp geboden kunnen worden om te verbeteren in bepaalde grammaticale onderdelen, zodat de schrijfvaardigheid kan verbeteren. Maar het meten van de grammaticale complexiteit van een tekst is geen eenvoudige opgave, want welke kenmerken dienen gemeten te worden om tot een betrouwbaar en representatief resultaat te komen? Dit vraagstuk is reeds in verschillende onderzoeken behandeld. Enkele onderzoeken betreffende dit vraagstuk zullen hieronder meer uiteengezet worden.

Metten van complexiteit: de T-unit

De grammaticale complexiteit van geschreven teksten is in een ruim aantal onderzoeken gemeten aan de hand van de zogeheten *T-unit*. In 1965 introduceert Hunt deze term voor het eerst, als afkorting voor *minimal terminable units*, en hij definieert deze als de kleinst mogelijke, grammaticaal correcte zinnen. De opbouw van een T-unit bestaat altijd uit één hoofdzin (of *main clause*), welke gezien wordt als de kern van die betreffende eenheid, met daarnaast alle tot die hoofdzin behorende bijzinnen. T-unit analyse werd door Hunt (1965) gezien als een effectief instrument voor het meten van grammaticale complexiteit. Het meten van complexiteit aan de hand van de T-unit is mogelijk op verschillende manieren, waaronder kijken naar de verhouding van het aantal zinsdelen (*clauses*) per T-unit of naar de verhouding van het aantal T-units per zin. De sterkste manier van meten is door de gemiddelde lengte van T-units te bekijken. Dit is volgens Hunt (1965) een goede indicatie van het niveau van de schrijver. Omdat de T-unit zich baseert op de zinsdelen van een tekst, kunnen we de T-unit beschouwen als een (complexiteits)kenmerk op zinsniveau. We zullen later zien dat kenmerken

op zinsniveau in onderzoek naar grammaticale complexiteit tegenover (complexiteits)kenmerken op naamwoordelijk niveau kunnen worden geplaatst.

Over de effectiviteit van de T-unit bestaat geen sterke consensus, ondanks de hoge mate van populariteit van de T-unit analyse in onderzoek naar grammaticale complexiteit. Twee manieren in het specifiek zijn erg populair, namelijk de gemiddelde lengte van een T-unit (MLTU) en het gemiddeld aantal zinsdelen, zowel hoofd- en bijzinnen, per T-unit (C/TU) (Biber, Gray, & Poonpon, 2011). In een kritische evaluatie van de T-unit demonstreert Gaies (1980) waarom het toepassen van een analyse aan de hand van T-units niet altijd geschikt is als goede maatstaf. Zo blijkt bijvoorbeeld dat een T-unit analyse niet in alle gevallen het juiste apparaat is om de grammaticale opbouw binnenin een zin accuraat weer te geven. Ter illustratie geeft Gaies (1980) de volgende twee volzinnen:

- 1) Then the rain falls and spring comes. (2 T-units)
- 2) So he went through the woods and pulled the feather out of his hat from the partridges and put a flower in his hat and walked on through the woods. (1 T-unit). (p. 55)

Bij dit voorbeeld is duidelijk waarom de interne grammaticale structuur een vertekend beeld oplevert wanneer alleen op het aantal T-units wordt gelet. De tweede zin is grammaticaal duidelijk een stuk complexer dan de eerste zin en levert waarschijnlijk meer moeite op tijdens het lezen dan de eerste zin. Wanneer we echter gebruikmaken het aantal T-units in een zin zou de eerste zin grammaticaal complexer moeten zijn, omdat deze meer T-units bevat. Zin 1 bevat namelijk twee keer een onderwerp met persoonsvorm, terwijl zin 2 maar één onderwerp bevat. Slechts het hanteren van aantal T-units per volzin blijkt zodoende geen voldoende meetinstrument te zijn voor grammaticale complexiteit, omdat dit in sommige gevallen leidt tot conclusies die de werkelijkheid niet goed weerspiegelen, zoals in het voorbeeld is gedemonstreerd. Wanneer dergelijke zinnen in een tekst voorkomen kan dat namelijk een zeer vertekend beeld opleveren van de C/TU van een tekst.

De geschiktheid van de T-unit analyse blijkt ook afhankelijk van genre (Beers & Nagy, 2009; Biber et al., 2011; Staples et al., 2016). Met name het gebruik van de T-unit bij academische teksten lijkt problematisch. De lengte van T-units blijkt volgens Biber, Gray en Poonpon (2011) geschikter te zijn voor alledaagse gesprekken dan voor academische, geschreven teksten, een constatering die Biber ook in 1988 al gemaakt had. De oorzaak van deze problematiek is de manier waarop academische teksten complex zijn. De grammatica in academische teksten is niet complex vanwege uitgebreide zinnen met daarin vele

ondergeschikte bijzinnen, maar de aard van grammaticale complexiteit van academische teksten vinden we in de uitgebreide naamwoordgroepen (Noun Phrase; verder in de tekst: NP) die deze teksten bevatten.

De NP's van academische teksten blijken namelijk met grote regelmaat zeer complex te zijn, wat gerealiseerd kan worden op verschillende manieren. Deze complexiteit betekent in een groot aantal gevallen dat een NP een hoge informatiedichtheid kent. Dit kan een schrijver bijvoorbeeld doen door zijn NP uit te bouwen middels specificieerders en complementen. Dit voorbeeld van Staples et al. (2016) laat zien hoe een zin grammaticaal ontzettend complex kan zijn op basis van naamwoordelijke complexiteit:

Selectivity [of the harvest [on Putauhinu Island]] translates into *large differences* [in *harvest* rates [among *weight classes*]. [. . .] There is evidence [for such **links** [between **characteristics** [of *young individuals*] and *life history traits* [of adults]] [in many taxa]]. (p. 150)

De hoofden van de NP's in bovenstaand voorbeeld zijn **dikgedrukt**. De modificeerders bij de verschillende NP's zijn *schuingedrukt* of staan tussen [haken]. Het voorbeeld toont aan dat (gemodificeerde) NP's de complexiteit van een zin enorm op kunnen schroeven, omdat er slechts twee werkwoorden in voorbeeld staan. Deze zijn onderstreept. Ondanks het feit dat het op basis van de werkwoorden eigenlijk geen complexe zinnen zijn, ervaart een lezer dat door de NP's wel op die manier.

Een mogelijke verklaring voor het feit dat kenmerken op zinsniveau de complexiteit van gesproken taal goed kunnen beschrijven, terwijl naamwoordelijke kenmerken geschikter zijn om academische teksten te beschrijven, is dat men bij gesproken taal minder tijd heeft om over de te produceren taal na te denken. Tijdens het spreken moet de spreker in korte tijd beslissingen nemen over de te produceren tekst, terwijl bij academische geschreven teksten juist veel tijd kan worden besteed aan het nadenken over inhoud en structuur van de tekst. Doordat een spreker minder tijd heeft is het lastiger om compacte, complexe NP's te bouwen, waardoor het bouwen van lange T-units als het ware een oplossing is voor het tijdsgebrek: in plaats van compacte informatie te verschaffen, creëert een spreker met langere T-units tijd voor zichzelf om na te denken over wat en hoe iets gezegd moet worden. Deze nood bestaat niet in eenzelfde mate bij academische teksten, waardoor naamwoordelijke kenmerken volgens Biber et al. (2011) belangrijker zijn voor dat genre tekst. Door dit te stellen zetten zij de meetmanier voor

grammaticale complexiteit middels naamwoordelijke kenmerken tegenover de kenmerken op zinsniveau, waar ook de T-unit onder valt.

Grammaticale complexiteitskenmerken

Hierboven is beschreven hoe naamwoordelijke kenmerken en kenmerken op zinsniveau onder andere academische teksten beïnvloeden. De definities van deze twee categorieën kunnen breed opgevat worden en kennen daardoor ook vele invullingen, maar hier worden de definities gehanteerd zoals Van Rijt, van den Broek, en de Maeyer (2020) beschrijven. Allereerst wordt volgens Van Rijt et al. (2020) veelvuldig in eerder onderzoek de complexiteit van NP's geoperationaliseerd als het aantal modificeerders dat per NP voorkomt. In andere woorden betekent dit dat een tekst waarin NP's in veel gevallen 'lang' zijn, dus naast een naamwoord bestaande uit één of meerdere modificeerders, qua grammaticaliteit complexer zijn. Naamwoordelijke complexiteit heeft dus te maken met NP's en op welke manier deze op- en/of uitgebouwd zijn. Syntactische complexiteit op zinsniveau heeft daarentegen te maken met subordinatie, oftewel ondergeschikte bijzinnen (Van Rijt, van den Broek, & de Maeyer, 2020). Een grammaticaal complexe tekst op basis van dit syntactische kenmerk bevat daarom een groot aantal zinnen waarin één of meerdere (ondergeschikte) bijzinnen voorkomen. In eerder onderzoek zijn pogingen ondernomen om te onderscheiden wanneer welk kenmerk een belangrijke rol speelt, maar een definitief antwoord is tot dusver nog niet gevonden. Er zijn echter wel onderzoeken waaruit is gebleken dat de syntactische kenmerken in verschillende domeinen van belang zijn. In de volgende paragrafen worden dergelijke resultaten bediscussieerd.

Biber et al. (2011) stellen in hun onderzoek een schema voor waarin de ontwikkelingsstadia van de door hen geïdentificeerde complexiteitskenmerken zijn gezet (pp. 30-31). Zij noemen dit schema de *developmental track*, in deze scriptie vertaald als het ontwikkelingstraject. Een gesimplificeerde variant van het ontwikkelingstraject van Biber et al. (2011, pp. 30-31) is weergegeven in Tabel 1. Dit traject reflecteert de bevindingen dat naamwoordelijke kenmerken belangrijker zijn bij academische teksten dan kenmerken op zinsniveau. Het schema bestaat namelijk uit 5 stadia, waarbij stadium 1, de 'laagste', begint met kenmerken op zinsniveau en tekstuele structuren en waar stadium 5, de 'hoogste', met naamwoordelijke kenmerken het schema afsluit. Een complexe naamwoordelijke NP wordt volgens het traject dus pas later gemaakt. Een schrijver zou dus naar verwachting steeds meer

gebruik gaan maken van het uitbouwen van NP's naarmate hij meer ervaren is in de schrijfvaardigheid. Teksten van vergevorderde schrijvers zouden naar verwachting dus veel NP's bevatten die uitgebreid zijn met modificeerders, zoals bijvoeglijke bepalingen of attributieve bijvoeglijke naamwoorden. Kortgezegd worden naamwoordelijke structuren dus later verworven. Dit is een bevinding die de ontwikkeling van grammaticale complexiteit bij het schrijven van academische teksten zou kunnen verklaren.

Tabel 1

Gesimplificeerde variant van het ontwikkelingstraject van grammaticale complexiteitskenmerken voor schrijfvaardigheid naar Biber et al. (2011)

Stadium	Grammaticale structuren
1	Simpele finiete complementszinnen met veelvoorkomende werkwoorden
2	Finiete complementszinnen met een grotere set werkwoorden Infiniete complementszinnen met veelvoorkomende werkwoorden
3	Inbedding van vaste frases en uitdrukkingen Infiniete complementszinnen met een grotere set werkwoorden Simpele frasale inbedding in NP's
4	Infiniete betrekkelijke bijzinnen Meer frasale inbedding in NP's
5	Preposities + infiniete complementszinnen Complementszinnen gecontroleerd door naamwoorden Vergelijkende bepalingen met naamwoorden Uitgebreide inbedding van frasen in NP's

Noot. Bewerkt van "Should we use characteristics of conversation to measure grammatical complexity in L2 writing development?" door D. Biber, B. Gray en K. Poonpon, 2011, *Tesol Quarterly*, 45(1), pp. 30-31. Copyright 2011 door Copyright Clearance Center.

Parkinson en Musgrave (2014) baseren zich in hun onderzoek naar de ontwikkeling van de schrijfvaardigheid onder studenten op de bevindingen van Biber et al. (2011). Zij onderzochten of het voorgestelde ontwikkelingstraject toepasbaar is op zowel L2 als L1 studenten van het Engels, wat inderdaad zo blijkt te zijn. In het onderzoek van Parkinson en Musgrave (2014) zijn twee groepen afstuderende L2-schrijvers vergeleken in hun gebruik van soorten NP-modificeerders. Een groep zat nog in de voorbereiding van hun 'graduate study', terwijl de andere groep al voor de studie ingeschreven was. De nog in te schrijven groep bleek veel meer gebruik te maken van de simpelere modificeerders volgens het ontwikkelingstraject, terwijl het gebruik van de reeds ingeschreven groep gevarieerder en complexer was. Het gehypothetiseerde ontwikkelingstraject van Biber et al. (2011) bleek dus overeen te komen met de resultaten van Parkinson en Musgrave (2014).

Het ontwikkelingstraject wordt ook door de bevindingen van Staples et al. (2016) bevestigd. Zij onderzochten het verband tussen de stijging in academisch niveau en de verandering in het gebruik van kenmerken op naamwoordelijk en zinsniveau in teksten. Dit onderzoek vond plaats bij studenten. Staples et al. (2016) wisten aan te tonen dat naarmate het academische niveau van een student steeg, het gebruik van naamwoordelijke kenmerken ook toenam, terwijl er juist een vermindering van kenmerken op zinsniveau plaatsvond. Daarnaast werd aangetoond dat het soort discipline waarover (bijvoorbeeld natuurwetenschappen, sociale wetenschappen of letterkunde) en het genre waarin geschreven werd invloed had op het gebruik van de verschillende kenmerken. Ook uit dit onderzoek werd daarbij duidelijk dat vooral de naamwoordelijke kenmerken van belang zijn bij de grammaticale complexiteit van geschreven teksten.

Het huidige onderzoek

Wanneer het bovenstaande kort samengevat wordt, zien we dat er in onderzoek naar grammaticale complexiteit van geschreven teksten over het algemeen hoofdzakelijk twee manieren zijn om die complexiteit te meten. Sinds geruime tijd wordt de T-unit gebruikt als meetinstrument, ondanks de afwezigheid van een consensus betreffende de effectiviteit. Bij meerdere onderzoeken is gebleken dat het hanteren van een T-unit analyse niet altijd geschikt is, bijvoorbeeld door het genre van de tekst. Als tegenhanger van de T-unit, een meetinstrument dat gebruikmaakt van de complexiteitskenmerken op zinsniveau van de tekst, is het hanteren

van meetinstrumenten op basis van naamwoordelijke complexiteitskenmerken voorgesteld, waarbij gekeken wordt naar de NP's in teksten en de specificatie en modificatie van deze NP's.

Deze naamwoordelijke kenmerken zijn volgens Biber et al. (2011) en Staples et al. (2016) vooral geschikt als meetinstrumenten bij academische teksten. Echter is veel onderzoek betreffende dit onderwerp met (een van) deze twee meetinstrumenten uitgevoerd bij onderzoek naar moedertaalsprekers van het Engels of participanten met Engels als L2. Onderzoek met deze instrumenten bij participanten met een andere moedertaal dan het Engels is nog schaars. Middels dit onderzoek tracht ik deze ontbrekende kennis ten minste gedeeltelijk in te vullen door de Nederlandse taal in deze onderzoeksrichting te betrekken. Daarnaast wordt onderzoek naar dit onderwerp in de meeste gevallen gedaan bij studenten. De ontwikkeling van het schrijfproces begint echter al op de middelbare school, waar een belangrijke basis voor de hogeschool of universiteit gelegd zou moeten worden. Daarom probeer ik met dit onderzoek ook het gat op te vullen van de schrijfontwikkeling vóóordat een scholier student wordt.

De onderzoeksvraag luidt daarom als volgt:

Is er sprake van een verschil in grammaticale complexiteit tussen geschreven, Nederlandstalige teksten van vwo-onderbouw leerlingen en vwo-bovenbouw leerlingen, waarbij specifiek gekeken wordt naar het verschil tussen kenmerken op naamwoordelijk niveau en kenmerken op zinsniveau?

Op basis van de literatuur die in deze inleiding is besproken is de verwachting dat de waardes voor de kenmerken op naamwoordelijk niveau hoger zullen zijn voor bovenbouw leerlingen dan voor onderbouw leerlingen, oftewel dat teksten complexer zullen zijn met betrekking tot de NP's. Dergelijke kenmerken blijken volgens eerder onderzoek namelijk op latere leeftijd te ontwikkelen. De waardes voor kenmerken op zinsniveau zullen ongeveer gelijk blijven of wat stijgen in de bovenbouw. De teksten worden dus op zinsniveau ook complexer, maar de verwachting is dat teksten op naamwoordelijk niveau meer zullen toenemen in complexiteit.

In dit onderzoek zullen twee manieren om grammaticale complexiteit te meten naast elkaar gelegd worden bij het meten van academische teksten. De kern van dit onderzoek is gebaseerd op Staples et al. (2016) en op basis van dat onderzoek en andere eerdergenoemde

onderzoeken wordt verwacht dat met name de naamwoordelijke kenmerken een duidelijk verschil zullen laten zien tussen de onder- en de bovenbouw. Dit verschil wordt verwacht omdat onder anderen Biber et al. (2011) hebben laten zien dat naamwoordelijke complexiteitskenmerken geschikt zijn om verschillen in academische teksten aan te duiden. In dit onderzoek worden namelijk academische teksten vergeleken. Ondanks geluiden die het meten van complexiteit met T-units als ongeschikt achten voor academische teksten, ligt de verwachting dat hier ook een positieve ontwikkeling te zien zal zijn wat betreft de kenmerken op zinsniveau, al dan niet minder sterk dan de naamwoordelijke kenmerken.

Belangrijk om te onthouden is dat het meten van de grammaticale complexiteit van teksten niet automatisch hetzelfde is als het meten van de kwaliteit van teksten. Er bestaat namelijk geen directe, sterke link tussen die twee tekstonderdelen (Beers & Nagy, 2009). Dit wordt ook benadrukt door Lu (2011). Dat er geen direct verband bestaat betekent niet dat grammaticale complexiteit niet relevant is voor de kwaliteit van een tekst. Een tekst bestaat namelijk uit verschillende factoren, waaronder grammaticale complexiteit, en deze factoren dragen allemaal bij aan de kwaliteit van een tekst (Beers & Nagy, 2009). In dit onderzoek zullen daardoor uitsluitend uitspraken worden gedaan over het niveau van grammaticale complexiteit van teksten en niet over de kwaliteit van desbetreffende teksten.

Methode

Vanwege het coronavirus is de dataverzameling niet volgens het originele plan gegaan. In plaats daarvan is gebruikgemaakt van alternatieve, reeds beschikbare data. Deze data zijn beschikbaar gesteld door Rob Schoonen, en is afkomstig uit het NELSON-project van de Universiteit van Amsterdam. De data zijn onder andere gebruikt in Schoonen et al. (2003) en Schoonen, Van Gelderen, Stoel, Hulstijn, en De Glopper (2011). In deze methodesectie wordt allereerst ingegaan op het originele plan. Daarbij worden in het kort de aspecten beschreven als bij een reguliere methodesectie. Vervolgens gaat de methodebeschrijving verder met het alternatieve plan en daarbij de alternatieve data afkomstig van het NELSON-project, wat allemaal uitgebreid per onderdeel beschreven zal worden. Op enkele plekken zal kort toegelicht worden wat het originele plan en doel was bij de dataverzameling, zodat bijvoorbeeld verduidelijkt kan worden waarom de alternatieve data geschikt bleken om voor dit onderzoek gebruikt te worden.

Onderzoeksontwerp

Om te onderzoeken of er verschillen bestaan in de grammaticale complexiteit van teksten, geschreven door scholieren van het vwo, is een kwantitatief onderzoek uitgevoerd. Het onderzoek is cross-sectioneel, waarbij getracht wordt een longitudinaal onderzoek te simuleren. In dit onderzoek wordt de grammaticale complexiteit van geschreven teksten van leerlingen uit de onderbouw van het vwo vergeleken met leerlingen uit de bovenbouw. Het originele plan was om dit te onderzoeken door aan scholieren van het vwo een digitale vragenlijst voor te leggen waarin zij een korte tekst van minimaal 200 woorden zouden moeten schrijven. Deze vragenlijst zou aan hen voorgelegd worden door een van hun docenten en zou idealiter scholieren uit verschillende delen van Nederland hebben bereikt.

De teksten van de leerlingen worden geanalyseerd op basis van een zevental kenmerken op zinsniveau en kenmerken op naamwoordelijk niveau. Deze zijn gebaseerd op de kenmerken die gebruikt worden in Staples et al. (2016, p. 160). Deze analyse wordt uitgevoerd aan de hand van het programma T-scan (Pander Maat et al., 2014). Omdat er in de literatuur geen eenduidige consensus bestaat over de effectiviteit van deze twee soorten kenmerken, betrek ik in dit onderzoek beide soorten kenmerken.

Participanten

De groep participanten waarbij dit onderzoek idealiter uitgevoerd zou hebben bestaan uit leerlingen op Nederlandse middelbare scholen uit de schoolklassen twee tot en met vijf vwo. Een tweede criterium was dat de participant de Nederlandse taal goed beheerst, waarbij de moedertaal van de participant voor dit onderzoek niet van belang was.

De alternatieve data waarmee uiteindelijk gewerkt zijn benadert de hierboven beschreven participantgroep, maar kent een aantal verschillen. De data zijn verkregen door het NELSON onderzoeksteam en is onder andere gebruikt in Schoonen et al. (2003) en Schoonen, Van Gelderen, Stoel, Hulstijn, en De Gloppe (2011). In totaal hebben 397 scholieren meegedaan aan een longitudinaal onderzoek en deze scholieren zaten op het moment van testen op een van acht middelbare scholen in het westen van Nederland (Schoonen et al., 2011). Aan het begin van dat longitudinale onderzoek zaten de scholieren in de tweede klas van de middelbare school met een leeftijd van ongeveer 13 tot 14 jaar oud (Schoonen et al., 2011).

Ondanks dat deze alternatieve data een geschikte vervanger waren van de data die onder ideale omstandigheden verzameld waren geweest, verschillen deze wel van elkaar op enkele vlakken. De belangrijkste verschillen betreffen het meetmoment en het schooltype. Idealiter zou er vergeleken worden tussen vwo 2 en vwo 5, omdat de scholieren een duidelijke en grote groei door kunnen maken in de drie jaar tussen die twee schooljaren. De data van het NELSON project waar hier gebruik van wordt gemaakt kennen een kortere tijdsspanne tussen de metingen en een verschil qua schooltype. Zo is er een onderscheid gemaakt tussen vmbo en havo/vwo, en is het verschil tussen de testmomenten slechts één jaar. In dit onderzoek vonden de drie meetmomenten, waarbij twee meetmomenten teksten over hetzelfde onderwerp opleverden, plaats in de derde en in de vierde klas (met respectievelijke leeftijden van ongeveer 14 tot 15 jaar en 15 tot 16 jaar).

In dit onderzoek heb ik me beperkt tot een subgroep van 114 leerlingen van de in totaal 397 leerlingen. Hierbij is onder andere gefilterd op basis van schooltype (havo & vwo) en geschiktheid van de data. Daarnaast zijn slechts de data in de analyse meegenomen van participanten die aan beide meetmomenten deel hadden genomen. De uiteindelijke participantengroep bestaat uit 55 jongens (48,2%) en 59 meisjes. Alle participanten maken deel uit van de groep met als schooltype 'havo & vwo'. Voor het grootste deel van de groep was Nederlands de L1 (78,1%). De overige leerlingen (21,9%) spraken Nederlands als L2. De precieze L1 van de rest van de groep is niet bekend, want deze is niet verder gespecificeerd dan 'anderstalig'.

Materiaal

De originele intentie om data te verzamelen was via een digitale enquête. Deze was gemaakt in Qualtrics en participanten konden daarom zowel op de computer als op de mobiele telefoon de vragenlijst beantwoorden. De vragenlijst is verspreid naar leerlingen van het vwo met hulp van docenten van middelbare scholen. Deelnemende participanten vulden de vragenlijst individueel en vrijwillig in. Het betrof een korte vragenlijst, bestaande uit een aantal vragen over persoonlijke gegevens en één *essayvraag*. De instructie die deelnemers kregen is te vinden in bijlage A.

Aan de deelnemers werd gevraagd een korte, betogende tekst te schrijven over een onderwerp dat nauw aansluit bij hun directe leefwereld, namelijk over het belang van het leren van vreemde talen op de middelbare school. Daarnaast betreft het een onderwerp waarover de

meeste leerlingen zeer waarschijnlijk een mening zullen hebben. Dit werkt het teksttype, een betoog, uiteraard goed in de hand. Het teksttype, een betoog, en het publiek, een docententeam dat beslist over het curriculum, zijn gekozen omdat deze allebei sturen richting een academische tekst. De reden dat gestuurd wordt naar een academische tekst is omdat leerlingen hier eerder op het beste van hen kunnen presteren dan bij andere tekstsoorten. Omdat genre invloed kan hebben op het gebruik van de verschillende complexiteitskenmerken heb ik ervoor gekozen om iedere participant hetzelfde genre en onderwerp te geven.

Bovenstaande manier van dataverzameling verliep dus helaas niet volgens plan. Doordat de vragenlijst precies tijdens het begin van de coronapandemie verstuurd moest worden, en uiteindelijk ook verstuurd is, kwam deze aan bij docenten die met dichte scholen zaten. Veel respons op de vragenlijst kwam er dus niet. Daarom is er gebruikgemaakt van alternatieve data, namelijk data afkomstig van het NELSON project. Deze dataverzameling voldeed (grotendeels) aan de eisen en voorwaarden die ik aan mijn eigen vragenlijst, en dus de gewenste data, had gesteld. De dataset van NELSON kon daarom gebruikt worden voor dit onderzoek.

De data van het NELSON project bestaan uit teksten geschreven, op de computer, door middelbare scholieren uit de derde en vierde klas van het vmbo, havo en vwo. De teksten zijn betogen over onderwerpen waar de middelbare scholieren bekend mee waren ten tijde van schrijven, bijvoorbeeld het verbieden van skaten op de stoep. Daarnaast zijn de teksten ongeveer 100 tot 200 woorden in lengte.

Het materiaal van NELSON is gecontroleerd op aanwezigheid van de participanten. Dat wil zeggen, er is eerst gekeken welke participanten aan beide meetmomenten deelgenomen hebben. Vervolgens is op basis van die lijst gefilterd op basis van schooltype, waarbij alleen de havo & vwo scholieren meegenomen zijn in het onderzoek. Tot slot zijn de teksten van iedere participant doorgenomen en gecontroleerd op geschiktheid. Sommige participanten hebben de taak namelijk niet erg serieus genomen, waardoor niet alle teksten bruikbaar waren.

De NELSON-data zijn met betrekking tot verschillende aspecten aangepast, zodat T-scan accuratere resultaten kan geven. T-scan heeft een specifieke input nodig om accurate en correcte resultaten te kunnen geven. Een voorbeeld van specifieke input is dat T-scan vereist dat zinnen afgesloten worden met een punt. Als een tekst daarom geen punt bevat aan het eind van een zin, zal T-scan dit niet herkennen als zin. Hierdoor kunnen de resultaten vertekenen. Ook is er vanwege deze specifieke input een semiautomatische spellingscontrole uitgevoerd, omdat T-scan de foutief geschreven woorden anders niet herkent als bestaande woorden. Omdat

de geschreven teksten bij het eerste meetmoment de vorm hadden van een brief voor het schoolblad bevatten veel teksten een aanhef, zoals ‘Beste lezers’. Deze aanhef is weggehaald, omdat dit de gemiddelde lengte van zinnen vertekend, vooral vanwege de vaak geringe lengte van de teksten. Daarnaast betreft het in deze gevallen zeer conventionele uitspraken, welke daardoor ook niet veel zouden kunnen zeggen over de grammaticale complexiteit van de tekst. Om dezelfde redenen zijn ook eventueel toegevoegde slotformules, zoals bijvoorbeeld ‘met vriendelijke groet’, weggehaald. Inhoudelijk zijn de geschreven teksten op geen enkele wijze aangepast.

Procedure

In het originele plan zouden scholieren een vragenlijst digitaal verstrekt krijgen. Invullen van de vragenlijst zou volledig vrijwillig gebeuren. De opdrachtbeschrijving van deze vragenlijst is te vinden in bijlage A. Bij het invullen moesten de deelnemers allereerst toestemming geven voor hun deelname. Als zij dit hadden gegeven, gingen ze door naar de rest van de vragenlijst. Voordat deelnemers aan de slag konden met de *essayvraag* werden zij eerst om enkele persoonlijke gegevens gevraagd, waaronder profielpakket, keuzevakken en moedertaal. Na het invullen van de gevraagde persoonlijke gegevens was het aan de deelnemers om de *essayvraag*, te vinden in Bijlage A, te beantwoorden, waarop zij een tekst van minimaal 200 woorden als antwoord moesten geven. Op het globale onderwerp na werden de deelnemers vrij gelaten in hun invulling van de opdracht.

Om tot de teksten te komen zoals deze in de NELSON-data staan werd in het klaslokaal gevraagd aan leerlingen om een kort betoog te schrijven over een vooraf gegeven onderwerp op de computer. Door dit in een klaslokaal te doen in plaats van een willekeurige steekproef van scholieren online, kon twee jaar op rij ongeveer dezelfde groep gevraagd worden voor dit onderzoek, waardoor de teksten van de twee jaren vergeleken kunnen worden.

Instrumenten

Voor het verzamelen van data is gebruikgemaakt van Qualtrics, een programma voor het maken en verspreiden van digitale vragenlijsten. Participanten konden de vragenlijst invullen op een computer of mobiel apparaat en de responsen werden automatisch opgeslagen. Echter leverde deze manier van dataverzameling zeer weinig data op. In plaats daarvan is, zoals eerder

besproken, gebruikgemaakt van data van het NELSON project. De teksten die door de leerlingen geschreven zijn werden door hen getypt op een computer.

Naast apparatuur voor het verzamelen van data is het online programma T-scan gebruikt voor het automatisch analyseren van de geschreven teksten (Pander Maat et al., 2014). T-scan geeft een enorme hoeveelheid informatie over de teksten, bijvoorbeeld het gemiddeld aantal letters per woord, het aantal nevenschikte bijzinnen en het aantal lijdende vormen per deelzin. Van de aanzienlijke hoeveelheid door T-scan onderscheidde tekstkenmerken zijn er zeven geselecteerd om gebruikt te worden in de uiteindelijke analyse. In Tabel 2 staat kort opgesomd welke zeven kenmerken er gebruikt zijn en wat zij meten volgens Pander Maat et al. (2014).

Tabel 2

De zeven T-scan kenmerken van Pander Maat et al. (2014) die gebruikt zijn in de data-analyse. De operationalisatie van deze kenmerken is ook afkomstig van Pander Maat et al. (2014).

Kenmerk	Operationalisatie
<i>Kenmerken op zinsniveau</i>	
Wrd_per_zin	Het gemiddeld aantal woorden per zin
Bijzin_per_zin	Het gemiddeld aantal bijzinnen in een zin (finitie plus infinitie bijzinnen)
Bijw_bep_dz	Het gemiddeld aantal bijwoordelijke bepalingen per deelzin
<i>Naamwoordelijke kenmerken</i>	
Inhwrld_dz	Het gemiddeld aantal inhoudswoorden per deelzin
Wrd_per_nwg	Het gemiddeld aantal woorden per naamwoordgroep
Bijv_bep_dz	Het gemiddeld aantal bijvoeglijke bepalingen per deelzin
Attr_bijv_nw_dz	Het gemiddeld aantal attributieve bijvoeglijke naamwoorden per deelzin

De eerste drie kenmerken zullen gebruik worden om grammaticale complexiteit op zinsniveau te meten, terwijl de laatste vier kenmerken het complexiteitsniveau op naamwoordelijk gebied zullen meten. De in dit onderzoek gekozen kenmerken vormen geen uitputtende lijst, maar het zijn de kenmerken die met het oog op de onderzoeksvraag waarschijnlijk de geschikteste informatie op zullen leveren.

Data-analyse

De getoetste data zijn verkregen via het automatische tekstannotatieprogramma T-scan (Pander Maat et al., 2014). In T-scan zijn alle teksten geanalyseerd op de zeven kenmerken die hier gerapporteerd worden. De uitkomsten worden door T-scan weergegeven als getallen, namelijk als gemiddeld aantal per desbetreffende kenmerk voor de geanalyseerde tekst. De ruwe data van T-scan zijn zowel automatisch als handmatig gecontroleerd op fouten. Hierbij kan gedacht worden aan ‘fouten’ zoals het aangeven van het aantal deelzinnen als slechts één of een ‘n.a.’ als resultaat krijgen bij een van de kenmerken. Fouten die op die manier zijn gevonden zijn ook uit de data gefilterd of gecorrigeerd in het geval dat de geschreven tekst duidelijk aantoonde dat er een fout gemaakt was bij T-scan.

Voordat statistische data-analyse op de resultaten is toegepast zijn de data op basis van een aantal elementen gefilterd. Allereerst zijn de data gefilterd op basis van het schooltype, waarbij alleen scholieren die schooltype havo of vwo volgen geselecteerd zijn om deel uit te maken van de analyse. Daarnaast bevatten de data scholieren die bij slechts één van de twee meetmomenten aanwezig zijn geweest, waardoor er geen vergelijking gemaakt kan worden met de eerste of tweede meting. Ook die participanten zijn uit de data gefilterd. Alle teksten die na deze filtering overbleven zijn handmatig gecontroleerd op geschiktheid van de teksten. Niet alle leerlingen hebben de experimentele taak even serieus genomen. Een voorbeeld hierbij is een tekst waarbij simpelweg willekeurige letters en tekens achter elkaar zijn getypt. Deze teksten zijn uiteraard niet geschikt voor een analyse en dus zijn dergelijke teksten uit de dataset gehaald. Na deze drie processen zijn tot slot enkele leerlingen uit de data weggelaten wegens ontbrekende data bij één of meer van de getoetste syntactische kenmerken, omdat het vergelijken van data onmogelijk was door de ontbrekende waarden.

Uiteindelijk zijn hierdoor 120 teksten overgebleven en statistisch geanalyseerd. Over de data bij alle zeven kenmerken is een tweezijdige afhankelijke t-toets uitgevoerd. Vóór die toetsing is gecontroleerd of aan de assumpties voor een dergelijke toets voldaan werd. Aan de assumptie van afhankelijkheid werd voldaan, omdat exact dezelfde groep participanten op twee verschillende tijdstippen werd gemeten. Hierdoor was ook voldaan aan de assumptie van gelijke groepsgrootte. In eerste instantie toonde een Shapiro-Wilk-toets echter aan dat de assumptie van normaliteit van de data geschonden werd. Om deze reden is gecontroleerd op uitbijters door middel van boxplots. In de data bleken veertien uitbijters te zitten, met name bij de kenmerken op zinsniveau en bij het aantal NP's per T-unit. Die uitbijters zijn uit de data

gefilterd, waardoor ook de assumptie van normaliteit voldaan werd. Hier is op gecontroleerd met een tweede Shapiro-Wilk-toets, die geen significante resultaten aantoonde. Naast een t-test zal ook de correlatie tussen de verschillende kenmerken per meetmoment getoetst worden.

Resultaten

Allereerst zal kort toegelicht worden hoe de data eruitzagen waarop de toetsing is toegepast. De analyseresultaten van de grammaticale complexiteitskenmerken worden vervolgens in een zevental paragrafen besproken. Als eerste zullen resultaten voor de grammaticale complexiteitskenmerken op zinsniveau, de *clausal features*, besproken worden. Hierbij komt eerst het kenmerk betreffende de lengte van T-units aan bod, gevolgd door het kenmerk over het aantal bijzinnen per zin. Tot slot wordt het kenmerk over het aantal bijwoordelijke bepalingen per deelzin besproken. Hierop volgend worden de naamwoordelijke complexiteitskenmerken, de *phrasal features*, behandeld. Het eerste kenmerk dat besproken wordt betreft het aantal NP's per deelzin en daaropvolgend het aantal woorden per naamwoordgroep (NP). Vervolgens wordt het aantal bijvoeglijke bepalingen per deelzin besproken en afsluitend wordt een specifieke subcategorie hiervan bekeken, namelijk de attributieve bijvoeglijke bepalingen.

Voor Figuur 1 tot en met 7 geldt dat de foutbalken (*error bars*) een betrouwbaarheidsinterval van 95% aangeven. Een belangrijke noot bij de figuren in deze resultatensectie is dat de Y-assen niet gelijk aan elkaar zijn. Omdat de grammaticale kenmerken een zeer verschillend bereik hebben in hun gemiddelde aantallen zijn figuren met eenzelfde schaal op de Y-as niet duidelijk te interpreteren. Omwille van de leesbaarheid van de figuren is dus gekozen voor verschillende schalen.

De lengte van de T-units

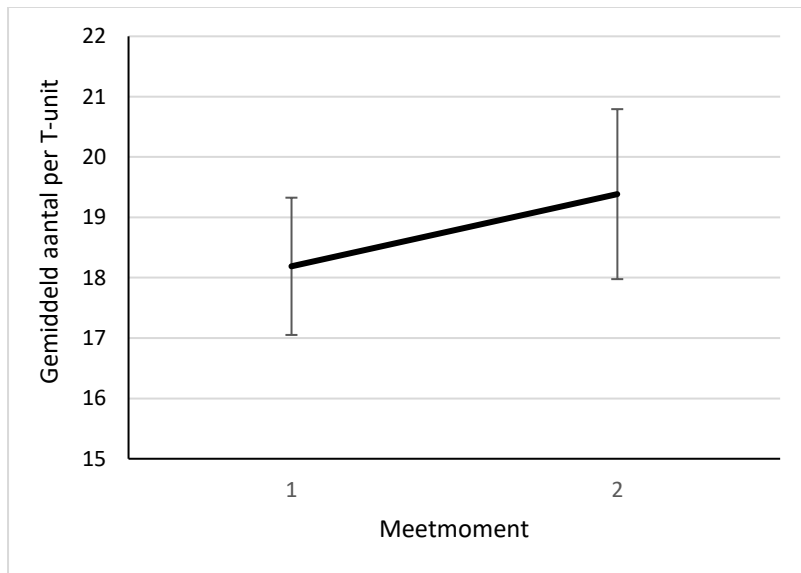
Het eerste geanalyseerde kenmerk is de gemiddelde lengte van T-units in de geschreven teksten. Zoals besproken in de inleiding is deze maat reeds in vele onderzoeken een belangrijk element gebleken. Op basis van die eerdere onderzoeken is hier dan ook de verwachting dat de gemiddelde lengte van een T-unit een positief significant resultaat op zal leveren. In andere woorden wordt dus verwacht dat de scholieren die aan dit onderzoek deelnamen op het tweede

meetmoment hoger zullen scoren dan op het eerste moment. Uit de toetsing is gebleken dat deze verwachting correct was.

Gemiddeld bestaat een T-unit van een tekst geschreven in de vierde klas uit meer woorden ($M = 19.39$, $SD = 7.59$), dan de T-units van geschreven teksten van leerlingen uit de derde klas ($M = 18.19$, $SD = 6.13$). Dit verschil bleek ook significant $t(113) = -1.94$, $p = .027$, $d = 0.17$. Dit verschil wordt weergegeven door Figuur 1.

Figuur 1

Lengte van de T-units



Noot. Het verschil in het gemiddeld aantal woorden per T-unit tussen meetmoment 1 in leerjaar drie en meetmoment 2 in leerjaar vier. De foutbalken geven een betrouwbaarheidsinterval van 95% aan.

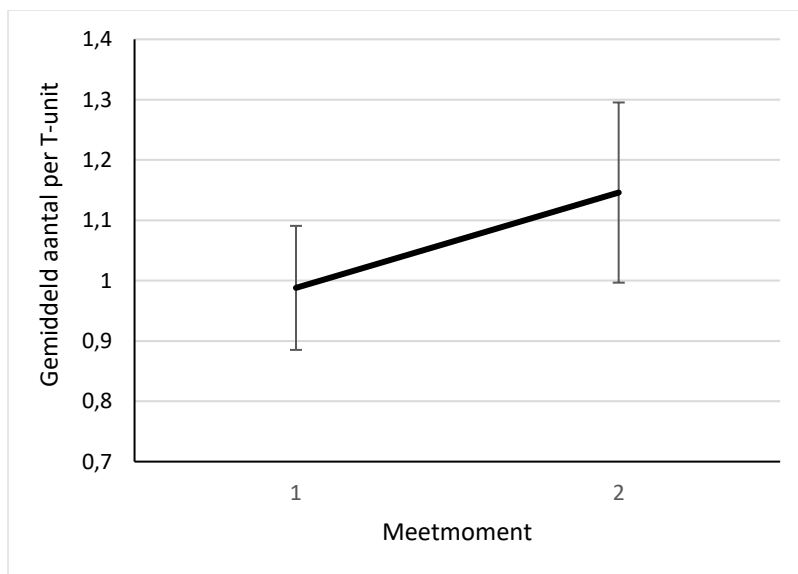
Het aantal bijzinnen per T-unit

Zoals besproken in de inleiding is er in eerder onderzoek gebleken dat het (relatieve) aantal ondergeschikte zinnen vaak een belangrijke maat is om de syntactische complexiteit van een tekst te meten. Om die reden is deze maat meegenomen in dit onderzoek. Naast de lengte van de T-units is dus onderzocht wat het gemiddeld aantal bijzinnen per T-unit is. Hierbij zijn alleen de bijzinnen meegeteld en is de hoofdzin zelf in deze resultaten niet meegeteld. Het aantal

bijzinnen per T-unit ligt hoger voor teksten geschreven in de vierde klas ($M = 1.47$, $SD = 0.81$), dan het geval is bij teksten uit de derde klas ($M = 0.99$, $SD = 0.55$). Opvallend aan deze resultaten zijn de ruime standaarddeviaties. Deze zijn relatief erg groot ten opzichte van de gemiddelden waartoe zij behoren, met name bij de resultaten van meetmoment twee. Toch bleek het verschil in het aantal bijzinnen per T-unit, net als de lengte van de gehele T-units, hier significant te zijn $t(113) = -2.16$, $p = .016$, $d = 0.23$.

Figuur 2

Aantal bijzinnen per T-unit



Noot. Het verschil in het gemiddeld aantal bijzinnen per T-unit tussen meetmoment 1 in leerjaar drie en meetmoment 2 in leerjaar vier. De foutbalken geven een betrouwbaarheidsinterval van 95% aan.

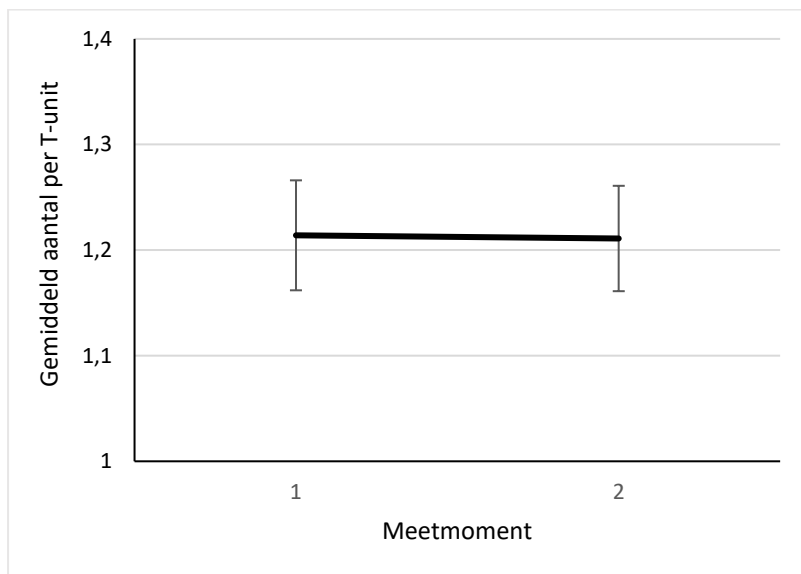
Het aantal bijwoordelijke bepalingen per deelzin

Qua complexiteit op zinsniveau ligt de verwachting dat niet alleen de lengte van de T-units of het aantal bijzinnen van belang zijn. Ook de manier waarop een (bij)zin is opgebouwd heeft invloed op de complexiteit. Bijwoordelijke bepalingen zijn een van deze manieren. Deze manieren hebben weinig betrekking op de naamwoorden in een zin, maar juist op de werkwoordelijke elementen van een zin. Hierdoor kan het gebruik van dergelijke constructies

de complexiteit op zinsniveau beïnvloeden. Het aantal bijwoordelijke bepalingen per deelzin was vrijwel gelijk bij teksten van de vierde klas ($M = 1.211$, $SD = 0.28$), vergeleken met de geschreven teksten van de derde klas ($M = 1.214$, $SD = 0.29$). Dit verschil bleek dan ook niet significant $t(119) = 0.08$, $p = .532$, $d = 0.01$.

Figuur 3

Aantal bijwoordelijke bepalingen per deelzin



Noot. Het verschil in het gemiddeld aantal bijwoordelijke bepalingen per deelzin tussen meetmoment 1 in leerjaar drie en meetmoment 2 in leerjaar vier. De foutbalken geven een betrouwbaarheidsinterval van 95% aan.

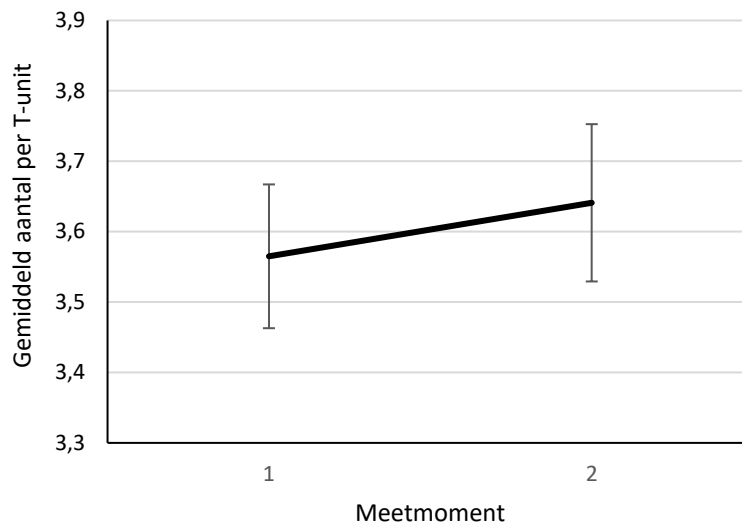
Het aantal NP's per T-unit

In veel gevallen is een toename van een kernelement een aanduiding dat er iets bijzonders aan de hand is. Wanneer het aantal NP's toe zou nemen van de onder- naar de bovenbouw, laat dit zien dat complexiteit sterk op naamwoordelijk niveau ontwikkelt. Uit de resultaten bleek dat het gemiddeld aantal inhoudswoorden in teksten uit de vierde klas hoger lag ($M = 3.64$, $SD = 0.62$), dan het geval is bij teksten uit de derde klas ($M = 3.57$, $SD = 0.57$). Het verschil met betrekking tot het aantal NP's is echter ook hier niet significant $t(119) = -1.16$, $p = .125$, $d =$

0.13. Een ontwikkeling op naamwoordelijk niveau in de complexiteit is dus niet terug te zien wanneer we kijken naar het aantal NP's.

Figuur 4

Aantal NP's per T-unit



Noot. Het verschil in het gemiddeld aantal NP's per T-unit tussen meetmoment 1 in leerjaar drie en meetmoment 2 in leerjaar vier. De foutbalken geven een betrouwbaarheidsinterval van 95% aan.

Het aantal woorden per NP

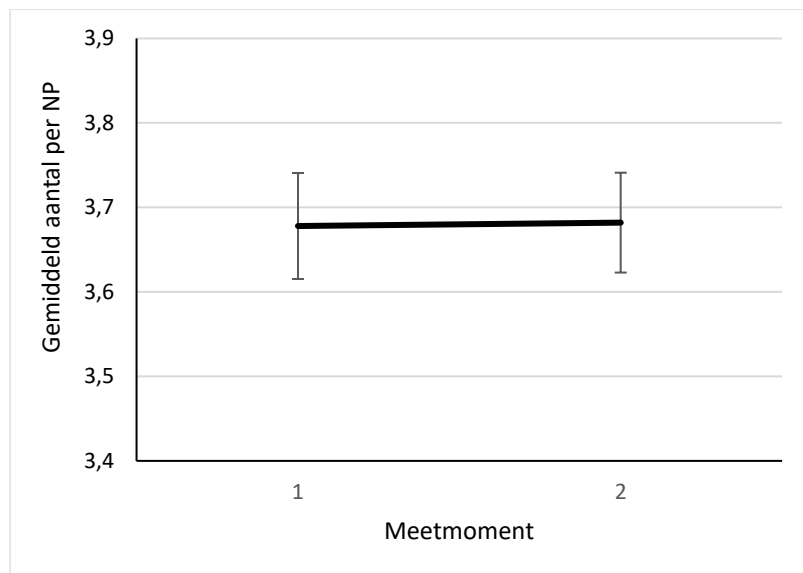
Naast het aantal NP's kijken we naar het aantal woorden per naamwoordgroep. Waar het bij het aantal NP's per T-unit misschien te verwachten was dat er geen sterke toename zou zijn, simpelweg omdat er niet meer NP's in een T-unit 'passen' zonder het overzicht van de betekenis te verliezen, is de hypothese dat we bij dit kenmerk een toename zullen zien bij de vierde klas ten opzichte van de derde klas.

In tegenstelling tot de eerste drie kenmerken is hier geen sprake van een toename tussen de derde en vierde klas. In het geval van aantal woorden per NP zou eerder gesproken kunnen worden van een stagnatie. Het aantal woorden per naamwoordgroep bleek namelijk nagenoeg identiek tussen teksten van de vierde klas ($M = 3.68$, $SD = 0.33$) en de derde klas ($M = 3.68$,

$SD = 0.35$). Het is dan ook niet geheel verrassend te noemen dat dit ‘verschil’ niet significant bleek $t(119) = -0.09, p = .464, d = 0.01$.

Figuur 5

Aantal woorden per NP



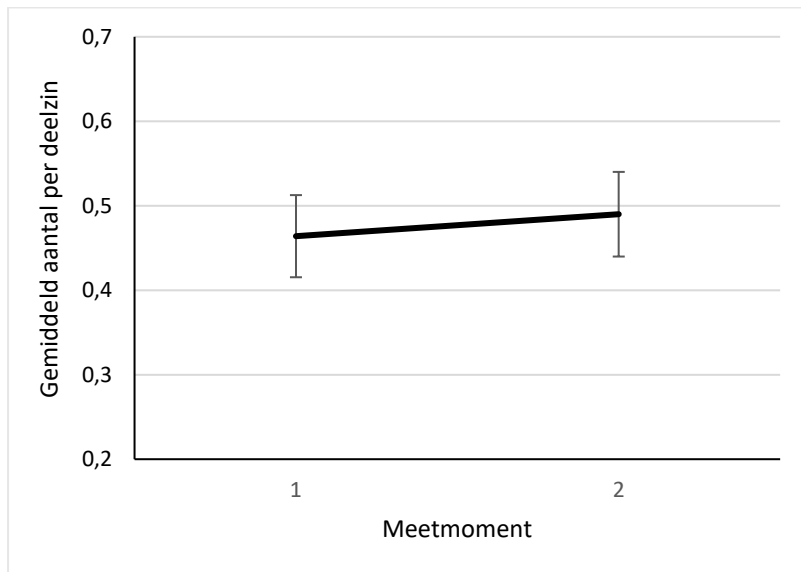
Noot. Het verschil in het gemiddeld aantal woorden per NP tussen meetmoment 1 in leerjaar drie en meetmoment 2 in leerjaar vier. De foutbalken geven een betrouwbaarheidsinterval van 95% aan.

Het aantal bijvoeglijke bepalingen per deelzin

Het aantal woorden per NP is niet significant toegenomen, maar de mogelijkheid bestaat dat het aantal bijvoeglijke bepalingen per deelzin wél toegenomen is. Het aantal bijvoeglijke bepalingen is namelijk een subcategorie van het ‘paraplukenmerk’ over het aantal woorden per NP. Het aantal bijvoeglijke bepalingen per deelzin was wel degelijk groter bij teksten van de vierde klas ($M = 0.49, SD = 0.28$), dan bij de geschreven teksten van de derde klas ($M = 0.46, SD = 0.27$). Maar ook het verschil in aantal bijvoeglijke bepalingen per deelzin blijkt niet significant $t(119) = -0.91, p = .181, d = 0.10$.

Figuur 6

Aantal bijvoeglijke bepalingen per deelzin



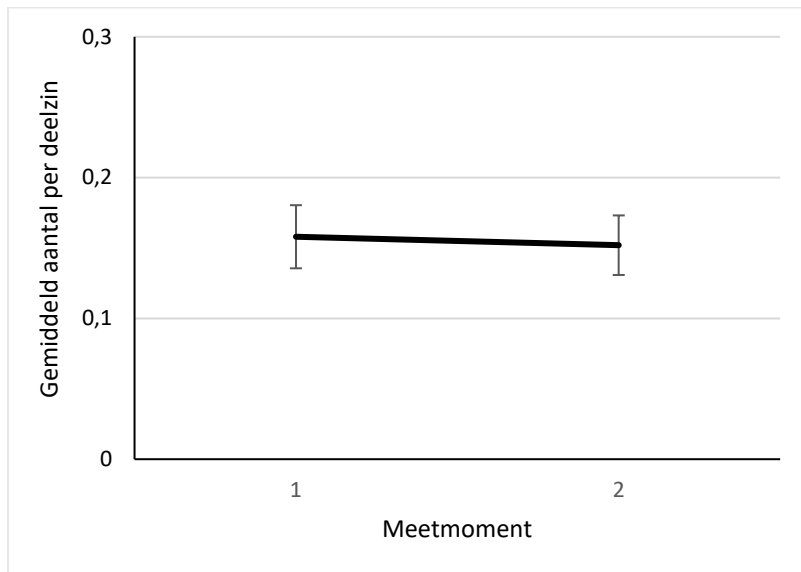
Noot. Het verschil in het gemiddeld aantal bijvoeglijke bepalingen per deelzin tussen meetmoment 1 in leerjaar drie en meetmoment 2 in leerjaar vier. De foutbalken geven een betrouwbaarheidsinterval van 95% aan.

Het aantal attributieve bijvoeglijke bepalingen per deelzin

Het aantal attributieve bijvoeglijke bepalingen per deelzin verschilde minimaal tussen teksten van de vierde klas ($M = 0.15$, $SD = 0.12$) en teksten van de derde klas ($M = 0.16$, $SD = 0.12$), waarbij er zelfs sprake is van een lichte afname van de derde naar de vierde klas. Dit minimale verschil gaat hier gepaard met een grote standaarddeviatie. Er is dus sprake van een sterk variërend kenmerk. Het geringe verschil bleek niet significant $t(119) = 0.51$, $p = .695$, $d = 0.05$.

Figuur 7

Aantal attributieve bijvoeglijke bepalingen per deelzin



Noot. Het verschil in het gemiddeld aantal attributieve bijvoeglijke bepalingen per deelzin tussen meetmoment 1 in leerjaar drie en meetmoment 2 in leerjaar vier. De foutbalken geven een betrouwbaarheidsinterval van 95% aan.

Correlaties

Naast het uitvoeren van een t-test is ook getoetst of er correlaties bestaan tussen de verschillende kenmerken op de twee verschillende meetmomenten. De resultaten voor het eerste meetmoment zijn te zien in Tabel 3 en de resultaten voor meetmoment twee zijn weergegeven in Tabel 4. In de tabellen zijn zowel de waarden van Pearson's r weergegeven als de bijbehorende p -waarden. Pearson's r staat rechtsboven de diagonaal en p staat linksonder de diagonaal.

Tabel 3*Correlaties tussen de zeven kenmerken op meetmoment 1*

Kenmerk	Lengte T-units	Bijzinnen per T-unit	Bijw. bep. per deelzin	NP's per T-unit	Woorden per NP	Bijw. bep. per deelzin	Attr. bep. per deelzin
Lengte T-units	1	0.726**	0.273*	0.387**	0.209*	0.365**	0.296*
Bijzinnen per T-unit	< .001	1	0.149	0.030	0.284*	0.189*	0.120
Bijw. bep. per deelzin	0.003	0.107	1	0.472**	0.445**	0.181*	0.074
NP's per T-unit	< .001	0.747	< .001	1	0.199*	0.674**	0.567**
Woorden per NP	0.022	0.002	< .001	0.029	1	-0.021	0.066
Bijw. bep. per deelzin	< .001	0.040	0.048	< .001	0.821	1	0.795**
Attr. bep. per deelzin	0.001	0.195	0.424	< .001	0.476	< .001	1

Noot. Het deel rechtsboven de diagonaal weergeeft Pearson's r . Hiervoor geldt * $p < .05$, ** $p < .001$. Het deel linksonder de diagonaal weergeeft de bijbehorende p -waarde.

Tabel 4*Correlaties tussen de zeven kenmerken op meetmoment 2*

Kenmerk	Lengte T-units	Bijzinnen per T-unit	Bijw. bep. per deelzin	NP's per T-unit	Woorden per NP	Bijw. bep. per deelzin	Attr. bep. per deelzin
Lengte T-units	1	0.805**	0.308**	0.168	0.048	0.324**	0.226*
Bijzinnen per T-unit	< .001	1	0.162	-0.097	0.154	0.133	0.063
Bijw. bep. per deelzin	< .001	0.083	1	0.481**	0.165	0.194*	0.193*
NP's per T-unit	0.073	0.302	< .001	1	0.012	0.656**	0.535**
Woorden per NP	0.612	0.099	0.072	0.897	1	-0.105	0.039
Bijw. bep. per deelzin	< .001	0.156	0.034	< .001	0.253	1	0.756**
Attr. bep. per deelzin	0.015	0.502	0.035	< .001	0.673	< .001	1

Noot. Het deel rechtsboven de diagonaal weergeeft Pearson's r . Hiervoor geldt * $p < .05$, ** $p < .001$. Het deel linksonder de diagonaal weergeeft de bijbehorende p -waarde.

Samenvatting resultaten

Twee van de drie grammaticale kenmerken op zinsniveau toonden een verschil in de verwachte, namelijk positieve, richting. Het betreft hier de kenmerken 'de lengte van de T-units' en 'het aantal bijzinnen per T-unit'. Scholieren in de vierde klas scoorde significant hoger bij die twee categorieën. Bij het derde zinsniveau-kenmerk 'het aantal bijwoordelijke bepalingen per deelzin' bleek het verwachte verschil niet aanwezig te zijn. De naamwoordelijke grammaticale kenmerken, waarvan de resultaten in figuur 4 tot en met 7 worden getoond, bleken geen van alle het verwachte verschil aan te geven.

Voor ieder complexiteitskenmerk is ook de effectgrootte berekend. De resultaten hiervan zijn weergegeven in Tabel 5. Bij de eerste twee kenmerken, die ook significante verschillen aantoonde, zien we een effectgrootte van respectievelijk $d = 0.17$ en $d = 0.23$. Naar Field (2013) kunnen beide effectgroottes geïnterpreteerd worden als kleine tot medium effecten.

Veel van de andere kenmerken kennen naast een niet significant verschil ook een zeer kleine effectgrootte, maar twee van de naamwoordelijke kenmerken kennen toch een kleine effectgrootte. We zien deze kleine effectgroottes bij ‘het aantal NP’s per T-unit’, $d = 0.13$, en bij ‘het aantal bijvoeglijke bepalingen per deelzin’, $d = 0.09$.

Tabel 5

Effectgroottes van de zeven complexiteitskenmerken

Kenmerk	Effectgrootte (d)
<i>Kenmerken op zinsniveau</i>	
De lengte van de T-units	0.17
Het aantal bijzinnen per T-unit	0.23
Het aantal bijwoordelijke bepalingen per deelzin	0.01
<i>Naamwoordelijke kenmerken</i>	
Het aantal NP’s per T-unit	0.13
Het aantal woorden per NP	0.01
Het aantal bijvoeglijke bepalingen per deelzin	0.09
Het aantal attributieve bijvoeglijke bepalingen per deelzin	0.05

Discussie en conclusie

Discussie

De grammaticale complexiteit van geschreven teksten kan op vele manieren gemeten worden. In dit onderzoek is die complexiteit geoperationaliseerd aan de hand van een zevental elementen, onderverdeeld in een tweetal categorieën, welke zijn weergegeven in Tabel 2. Uit de resultaten is gebleken dat slechts twee van deze zeven elementen een significant verschil aantonen in de ontwikkeling van de grammaticale complexiteit van geschreven teksten. Het betreft hier de lengte van de zogeheten T-units die schrijvers produceren en het aantal bijzinnen dat per T-unit voorkomt in een tekst.

De T-units van schrijvers uit het vierde jaar van het vwo bleken in dit onderzoek langer te zijn dan de T-units van leerlingen uit het derde jaar. In andere woorden betekent dit betekent

dat zinnen van de vierdejaars leerlingen gemiddeld genomen langer waren dan hun zinnen in het derde jaar. Daarnaast toonden de resultaten aan dat teksten van leerlingen uit de bovenbouw significant meer ingebedde zinnen bevatten. De hoofdzinnen binnen T-units waren dus niet alleen langer geworden, maar de T-units bleken uitgebreid te zijn middels bijzinnen. De andere vijf onderzochte kenmerken lieten geen significant verschil zien in de schrijfvaardigheidsontwikkeling.

In de hypothese is de verwachting gesteld dat het verschil in grammaticale complexiteit met name goed zichtbaar zou worden aan de hand van de naamwoordelijke kenmerken. Op basis van de zeven gebruikte grammaticale kenmerken kunnen we stellen dat die hypothese niet ondersteund kan worden. Geen van de vier naamwoordelijke kenmerken bleek namelijk significant en slechts twee van de kenmerken op zinsniveau bleken significant te verschillen. De verwachting was dat naamwoordelijke kenmerken een belangrijkere rol zouden gaan spelen naarmate een auteur meer ervaring heeft in het schrijven, omdat die kenmerken zich sterk beginnen te ontwikkelen naarmate een auteur vaker bezig is met schrijven en er dus (waarschijnlijk) steeds vaardiger in wordt (Parkinson & Musgrave, 2014). De naamwoordelijke kenmerken ontwikkelen zich pas later omdat deze ook lastiger toe te passen zijn. Dit wordt onderbouwd door het ontwikkelingstraject van Biber et al. (2011). Een toenemende complexiteit via zinsniveau ontwikkelt vroeg en is gemakkelijker te gebruiken, onder andere omdat een toevoeging op die manier in een groot aantal voorkomens op ongeveer dezelfde manier gebeurt. Een bijzin neemt namelijk vrijwel altijd eenzelfde syntactische vorm aan en is daardoor een soort invuloefening. Het modifieren van een NP vergt meer van een auteur én van een lezer. Een gemodificeerde NP bevat al snel een hoge informatiedichtheid. Waar normaliter een geheel nieuwe bijzin nodig is, kan nu in één NP dezelfde hoeveelheid informatie overgedragen worden.

Om deze reden is het opmerkelijk dat geen van de naamwoordelijke kenmerken van omvangrijk genoeg belang lijkt te zijn bij de geschreven teksten van de bovenbouwleerlingen dat die kenmerken zich ontwikkelen in de schrijfvaardigheid. In de resultaten zien we wel een zeer lichte stijging in de waardes omtrent die kenmerken, maar de stijgingen zijn te klein om een daadwerkelijk verschil aan te tonen. Een mogelijke reden voor dit opmerkelijke resultaat zou kunnen zijn dat het verschil qua leeftijd en ontwikkeling tussen de hier onderzochte groepen te klein is. Er zou te weinig tijd kunnen zijn geweest om een omslag aan te kunnen tonen in de manier van schrijven. Daarnaast is het ook een mogelijkheid dat niet alleen het verschil tussen

de twee groepen te klein is, maar dat de ‘oudste’ groep nog te jong is waardoor er nog geen ontwikkeling wat betreft de naamwoordelijke kenmerken plaats heeft kunnen vinden.

Naast significantiewaardes zijn ook de effectgroottes geanalyseerd. De twee zinsniveaus kenmerken die significant bleken, hadden ook een klein tot medium effect op de complexiteit van de geschreven teksten. Opvallend is bij de effectgroottes dat twee van de naamwoordelijke kenmerken een klein effect hadden op de complexiteit van de tekst. Het betreft hier de kenmerken ‘aantal NP’s per T-unit’ en ‘aantal bijvoeglijke bepalingen per deelzin’. Het is duidelijk zo dat die twee kenmerken toch invloed hebben op grammaticale complexiteit. Het zou kunnen zijn dat deze kenmerken zich pas net beginnen te ontwikkelen op de leeftijd van de onderzochte participanten. Hierdoor verschillen de resultaten van de vierde klas niet veel met die van de derde klas, omdat de kenmerken pas aan het begin van hun ontwikkeling staan en nog niet in een dusdanig stadium zitten dat het effect op de tekst als geheel erg groot is. Vooral de kleine effectgrootte van het aantal NP’s per T-unit valt in dezelfde lijn zoals we zien in Staples et al. (2016), waarin voorbeelden worden genoemd van lange, complexe zinnen met slechts één werkwoord maar met veel en uitgebreide NP’s. De hypothese dat naamwoordelijke complexiteitskenmerken van groter belang worden naarmate een schrijver gevorderder is, zou op basis van de gevonden effectgroottes dus niet direct verworpen hoeven te worden.

Een andere analyse die uitgevoerd is heeft betrekking op de eventuele correlaties tussen de zeven kenmerken per meetmoment. Deze resultaten zijn te zien in Tabel 3 en Tabel 4. Sommige correlaties vielen sterk in de lijn der verwachting. Een voorbeeld hiervan is de sterke, positieve correlatie tussen de lengte van T-units en het aantal bijzinnen per T-unit. Wanneer het aantal bijzinnen stijgt, zal ook de lengte van de gehele T-unit toenemen en andersom zal een verlenging van een T-unit toch vaak gebeuren via het toevoegen van bijzinnen. Deze correlatie was dan ook aanwezig tijdens beide meetmomenten. Er zijn echter ook correlaties die bij slechts een van de twee meetmomenten aanwezig zijn. Eén van deze correlaties is tussen lengte van T-units en het aantal NP’s per T-unit. In tegenstrijdigheid met de verwachting bestaat er een positieve correlatie tussen de twee bij het eerste meetmoment, maar niet bij het tweede meetmoment. Dit betekent voor het eerste moment dat wanneer de lengte van een T-unit toeneemt, het aantal NP’s per T-unit ook een toename kent. Bij meetmoment twee is dit dus niet het geval op basis van correlaties, terwijl we daar juist verwachtten dat er een effect zou bestaan. Een correlatie op meetmoment één die afwezig is bij meetmoment twee zien we vaker in de data. Dit komt bijvoorbeeld ook voor tussen de kenmerken ‘lengte van de T-units’ en ‘aantal woorden per NP’, ‘aantal woorden per NP’ en ‘aantal bijwoordelijke bepalingen per

deelzin', en 'aantal bijvoeglijke bepalingen per deelzin' en 'aantal bijzinnen per t-unit'. Waarom zijn deze correlaties tegenstrijdig met de verwachtingen? Een mogelijke verklaring voor deze resultaten zouden kunnen liggen in het onderwerp van de tekst. Misschien leende het onderwerp van de tekst tijdens het eerste meetmoment zich beter voor het gebruik van een veelvoud aan (gemodificeerde) NP's, terwijl het tweede onderwerp dit meer in de weg stond. Ondanks een kleine mogelijkheid dat deze verklaring klopt, is het natuurlijk erg opvallend dat deze correlaties verdwijnen bij het tweede meetmoment, terwijl we ze daar eigenlijk juist zouden verwachten. Het zou ook zo kunnen zijn dat het verdwijnen van de correlaties juist betekent dat de verschillende kenmerken meer op zichzelf kunnen staan en geen andere kenmerken nodig hebben. In andere woorden zou het aan kunnen duiden dat de NP kenmerken zich wat 'onafhankelijker' van andere kenmerken beginnen te ontwikkelen.

In dit onderzoek is geen concluderend antwoord naar voren gekomen wat betreft de verschillen tussen naamwoordelijke en zinsniveau kenmerken. Van de kenmerken op zinsniveau bleken twee van de drie een belangrijke factor van de complexiteit te zijn, terwijl op naamwoordelijk niveau geen kenmerk belangrijk bleek (op basis van *p*-waardes). Een mogelijke verklaring is dat op deze leeftijd de grammaticale complexiteit nog te veel berust op zinsniveau kenmerken. Wanneer we kijken naar effectgroottes en correlaties lijkt er wel degelijk sprake te zijn van een bepaalde ontwikkeling, maar op deze leeftijd lijken de verschillende kenmerken zich nog onvoldoende ontwikkeld te hebben om een grote rol te spelen in grammaticale complexiteit. Daarnaast zou het ook nog kunnen zijn dat de tijdsspanne waarin de teksten door de scholieren zijn geschreven een te korte tijd is geweest om een duidelijk verschil tussen meetmoment 1 en meetmoment 2 aan te kunnen geven. Een andere factor van invloed zou natuurlijk kunnen liggen bij de gekozen grammaticale kenmerken. Misschien waren de gekozen kenmerken niet helemaal de juiste om te onderzoeken en te vergelijken op dit punt in het schrijfproces en waren er andere aanwezige kenmerken die een duidelijkere schets van de situatie op hadden kunnen leveren.

Om meer zicht te krijgen op wanneer de verschillen tussen de verschillende soorten complexiteit een rol spelen, zal meer onderzoek uitgevoerd moeten worden tussen verschillende leeftijdsgroepen. Bij dergelijk onderzoek moet ervoor gezorgd worden dat er voldoende tijd zit tussen de verschillende meetmomenten om een eventueel resultaat duidelijk waar te kunnen nemen. Daarnaast zou het ook bevorderlijk zijn voor het onderzoek om meer (of andere) kenmerken mee te nemen in de analyses van de teksten, omdat het zou kunnen zijn dat de in dit onderzoek gekozen kenmerken het verschil in complexiteit niet duidelijk genoeg weer kunnen

geven. Veel onderzoek omtrent dit onderwerp gaat over studenten, maar om de ontwikkeling van beginnende academische schrijvers goed te kunnen ondersteunen is meer onderzoek nodig bij jongere auteurs. Meer onderzoek waarbij de grijze gebieden die in dit onderzoek overblijven in de gaten gehouden worden zou dus een goede stap zijn in de richting van het kunnen voorspellen van de ontwikkeling van grammaticale complexiteit bij schrijvers.

Conclusie

In dit onderzoek zijn twee soorten complexiteitskenmerken van geschreven teksten tegenover elkaar gezet. Op deze wijze is getracht significante verschillen te vinden tussen de verschillende soorten kenmerken en te laten zien dat leeftijd en ontwikkeling van schrijfvaardigheid een effect hebben op de manier waarop teksten complex zijn. De gekozen complexiteitsmaten blijken echter geen goede voorspellers te zijn van de complexiteit van een geschreven tekst, ten minste niet voor dit leeftijdsverschil. In deze leeftijdscategorie lijken alleen kenmerken op zinsniveau een rol te spelen, wat suggereert dat de schrijfontwikkeling van het gebruik van naamwoordelijke kenmerken nog niet van start is gegaan óf nog van onvoldoende kwaliteit is om effect te hebben op de grammaticale complexiteit van de teksten.

Literatuurlijst

- Beers, S. F., & Nagy, W. E. (2009). Syntactic complexity as a predictor of adolescent writing quality: Which measures? Which genre? *Reading and Writing*, 22(2), 185-200.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge University Press.
- Biber, D., Gray, B., & Poonpon, K. (2011). Should we use characteristics of conversation to measure grammatical complexity in L2 writing development? *Tesol Quarterly*, 45(1), 5-35.
- Chung, S., & Pullum, G. (z.j.). Grammar. Geraadpleegd op 28 juli 2021, op <https://www.linguisticsociety.org/resource/grammar/>
- Field, A. (2013). *Discovering statistics using IBM SPSS statistics*. sage.
- Gaies, S. J. (1980). T-unit analysis in second language research: Applications, problems and limitations. *TESOL quarterly*, 14(1), 53-60.
- Hunt, K. W. (1965). A synopsis of clause-to-sentence length factors. *The English Journal*, 54(4), 300-309.
- Lu, X. (2011). A corpus-based evaluation of syntactic complexity measures as indices of college-level ESL writers' language development. *TESOL quarterly*, 45(1), 36-62.
- Pander Maat, H., Kraf, R., Van den Bosch, A., Dekker, N., Van Gompel, M., Kleijn, S., Sanders, T. J. M., & Van der Sloot, K. (2014). T-Scan: a new tool for analyzing Dutch text. *Computational Linguistics in the Netherlands Journal* 4, 53-74.
- Parkinson, J., & Musgrave, J. (2014). Development of noun phrase complexity in the writing of English for academic purposes students. *Journal of English for Academic Purposes*, 14, 48-59.
- Schoonen, R., Van Gelderen, A., De Glopper, K., Hulstijn, J., Simis, A., Snellings, P., & Stevenson, M. (2003). First language and second language writing: the role of linguistic knowledge, speed of processing, and metacognitive knowledge. *Language Learning*, 53(1), 165-202.

- Schoonen, R., Van Gelderen, A., Stoel, R. D., Hulstijn, J., & De Glopper, K. (2011). Modeling the development of L1 and EFL writing proficiency of secondary school students. *Language Learning, 61*(1), 31-79.
- Staples, S., Egbert, J., Biber, D., & Gray, B. (2016). Academic writing development at the university level: phrasal and clausal complexity across level of study, discipline, and genre. *Written Communication, 33*(2), 149-183.
- Van Rijt, J. H., Van den Broek, B., & De Maeyer, S. (2021). Syntactic predictors for text quality in Dutch upper-secondary school students' L1 argumentative writing. *Reading and Writing, 34*(2), 449-465.

Bijlage A

De opdrachtbeschrijving van de originele vragenlijst:

PAGINA 1:

De Opdracht

Je gaat zo meteen beginnen aan een schrijfopdracht. Je krijgt bij de opdracht een bepaald onderwerp en een bepaald publiek aangewezen, waarbij ik van je zou willen vragen dat je een tekst van minimaal 200 woorden schrijft over dat onderwerp. Schrijf de tekst alsof je deze voor een schoolopdracht maakt..

Deze vragenlijst wordt tussendoor automatisch opgeslagen. Mocht deze vragenlijst dus tussentijds uitvallen, dan kun je verder gaan waar je gebleven was door op dezelfde computer en browser naar dezelfde link te gaan.

Het schrijven van de tekst in één sessie heeft echter wel de voorkeur.

Veel succes!

PAGINA 2:

Op de middelbare school maak je kennis met verschillende talen, zoals Engels, Frans en Duits. Het Engels wordt door velen als een zeer belangrijke taal in de moderne maatschappij gezien, maar hoe zit het met Frans en Duits? Is het belangrijk dat deze talen onderwezen worden op het voortgezet onderwijs in Nederland?

De schrijfopdracht is: **schrijf een korte betogende tekst van minimaal 200 woorden, waarbij je beargumenteert waarom jij het wel of niet belangrijk vindt dat je Frans en Duits leert op de middelbare school.**

Het publiek van je tekst is een team van docenten dat beslist over welke vakken op school gegeven moeten worden. Probeer helder en in goed Nederlands te betogen, en onthoudt: er bestaan geen foute antwoorden!

Je schrijft je tekst in het onderstaande vak. Als je klaar bent met het invullen klik je op het pijltje rechtsonder om je antwoorden definitief op te slaan en te versturen.

Succes!