

Tekstkenmerken en tekstkwaliteit van leerlingteksten

Een annotatiestudie

Henk Pander Maat, Kay Raaijmakers, Dennis Vermeulen en Kees de Glopper

TVT 41 (2): 331–361

DOI: 10.5117/TVT2019.2.001.PAND

Abstract

Text features and quality of learner text: an annotation study

Manually annotated corpora of writing products may greatly contribute to writing research: they offer detailed insights in the quality of these texts, in the text features actually attended to by human text raters, in possibilities and difficulties for the use of automatic writing analytics and writing tools, and in the relations between different text quality dimensions. This paper presents the Utrecht System for Annotation of Learner text (USALT), that covers both general features (orthography, punctuation, wording, coherence) and genre-specific elements (such as openings, endings, structuring devices and politeness). The annotations contain up to three items (annotation unit; problem type; part-of-speech tag). USALT reflects various text quality dimensions, notably correctness, comprehensibility and appropriateness (both stylistically and in terms of genre conventions).

We present an USALT analysis of 371 texts produced by Dutch students from grades 7-9 (aged 12-15 years), taken from the so-called *Schrijfmeters*-corpus. The assignment concerned a letter about 'typically Dutch things' to a Swedish girl about to emigrate to The Netherlands. USALT reliabilities were adequate. In terms of problem frequency, we were struck by the pervasiveness of punctuation problems. Furthermore, the orthography and punctuation problems together present considerable difficulties for automatic analysis of original learner texts at this level. A remarkable result regarding relations between various text quality dimensions is that the frequency of orthography problems correlates higher with genre convention problems than with lexico-grammatical problems. We also used the annotations as predictors of

the holistic scores assigned to the texts by human raters. Standardized annotation frequencies by themselves may account for 45% of the score variance, with a prominent role for annotations regarding genre elements; text length by itself explains 52%. The best model includes both text length and annotations (65% explained variance). In ongoing work, USALT is being extended to handle argumentative writing assignments.

Keywords: learner text, writing quality, annotation, text quality ratings

1 Inleiding

In het schrijfonderzoek zijn drie tradities te onderscheiden. In de *cognitieve* traditie staat het schrijfproces centraal: hoe conceptualiseren, formuleren en reviseren schrijvers hun teksten (Rijlaarsdam & Van den Bergh 2006; MacArthur & Graham, 2016), en welke cognitieve vermogens zijn daarbij betrokken (Torrance & Galbraith 2006)? In de *socio-culturele* traditie wordt schrijven gezien als een communicatieve handeling in een sociale situatie die zich kenmerkt door het externaliseren van verschillende cognitieve processen (Prior 2006; Bazerman, 2016). In de *tekstgerichte* traditie gaat het erom schrijfprocessen en de ontwikkeling van schrijfvaardigheid te documenteren via de resulterende teksten (Fayol, 2016; Tolchinsky, 2016). Zo presenteren Sanders en Schilperoord (2006) een methode voor structuuranalyse van leerlingteksten. Van Bochove en Vermeer (2015) laten in een stilistische analyse zien dat kinderen gaande van groep 5 naar groep 8 meer gedetailleerde objectbeschrijvingen en een levendiger woordkeus vertonen. En Jisa, Reilly, Verhoeven, Baruch en Rosado (2002), Verhoeven, Aparici, Cahana-Amitay, Hell, Kriz en Viguié-Simon (2002), Myhill (2008) en Christie (2009) bespreken ontwikkelingen in de zinsbouw in schrijfproducten. Die ontwikkelingen gaan in de richting van toenemende complexiteit en variatie.

In dit artikel kiezen we een andere wijze van tekstgerichte analyse, namelijk annotatie vanuit een normatief perspectief: we richten ons op kenmerken waarvan de aan- of afwezigheid van belang lijkt voor de kwaliteit van de tekst. In onderzoek naar tweedetaalverwerving is normatief annoteren al heel gebruikelijk, met name in de vorm van leerderscorpora met annotaties van 'taalfouten'. Een bekend corpus is het Cambridge Learner Corpus (CLC), dat 45 miljoen woorden telt. Een derde deel van dit corpus is gecodeerd op 88 soorten problemen, variërend van ontbrekende leestekens tot incongruentie tussen onderwerp en persoonsvorm (Nicholls 2003). De

problemannotaties in T2-corpora gaan meestal over grammaticale en lexicale onvolkomenheden die kenmerkend zijn voor taalleerders.

Het is natuurlijk ook mogelijk om een leerderscorpus puur descriptief te analyseren op de frequentie van bepaalde woorden of constructies (Hawkins & Buttery 2010), en zo kwantitatieve verschillen aan het licht te brengen tussen leerderstaal en de taal van ervaren sprekers of schrijvers. Maar het annoteren op tekstkwaliteit, met bijzondere aandacht voor het volgen van schrijftaal- en genrenormen, is om meerdere redenen extra interessant.

Ten eerste documenteren we zo de conflicten tussen de code van de leerling-schrijver en de code die van hem of haar verwacht wordt. Die afwijkingen zijn niet alleen wetenschappelijk maar ook praktisch interessant: de tekstaspecten die in bepaalde leerlingpopulaties voor problemen zorgen, zijn wellicht interessante thema's voor schrijfonderwijs. Ook verschaffen deze gegevens een feitelijke basis aan normatief georiënteerde discussies over het schrijfvaardigheidspeil van jongeren. Daarbij worden zorgen geuit over de negatieve invloed van digitale media, al is de evidentie voor dit verband mager (Verheijen 2013; VandeKerckhove & Sandra 2015).

Ten tweede is annotatie van schrijfproducten interessant voor inzicht in het beoordelen van opstellen. In onderzoek naar de beoordeling van leerlingteksten is veel aandacht besteed aan beoordelings*procedures*, zoals holistisch versus analytisch oordelen en het al of niet gebruiken van benchmarks (Bouwer, Koster & Van den Bergh 2016; Pollman, Prenger & De Gloppe 2012). Vooralsnog is er minder aandacht voor de rol van concrete tekstkenmerken, in ieder geval voor het Nederlandse taalgebied (maar zie De Gloppe 1984 en Boekelder, De Gloppe & Van Wijk, 1991). De vraag is daarbij voor ons niet alleen welke tekstkenmerken van invloed zijn op het opsteloordeel, maar omgekeerd ook in hoeverre schrijfkwaliteit al zichtbaar wordt in eenvoudig te observeren tekstaspecten.

Ten derde maken deze coderingen het mogelijk om onderzoek te doen naar de correlatie tussen tekstkenmerken van lagere en hogere orde. Die correlatie is vaak substantieel. Zo zeggen Deane & Quinlan (2010):

[...] [I]t should not be surprising to discover that certain linguistic aspects of an essay relate strongly to rater judgments of overall essay quality. (...) Of themselves, these linguistic aspects may cover only a limited portion of the construct. Generally, educators recognize that “writing skill” goes beyond basic skills (e.g., spelling, grammar, and punctuation), and very much includes abilities to select and organize ideas. In writing, simple things (i.e., words and phrases) are used to build complex things (i.e., documents), with the former to a great extent

controlling the latter. So, the measurement of linguistic aspects can reveal quite a bit about essay quality. They may not capture some essential aspects of quality, such as ideas or voice, but they can measure many of the traits characteristic of stronger writers, including fluency, word choice, adherence to conventions, and use of appropriate discourse structures. (Deane & Quinlan 2010, 155)

Ten vierde geeft normatieve annotatie inzicht in de (on)mogelijkheden van automatische tools voor tekstanalyses van leerlingteksten en eventueel voor ondersteuning van schrijvers. Annotaties op spelling, leestekens en zinsbouw laten zien waar een automatische analyse van leerlingteksten op problemen zal stuiten, omdat door die fouten zinnen, zinsstructuren en woorden niet meer herkend kunnen worden. Verder brengen de annotaties tekstkwaliteitsaspecten aan het licht waarop leerlingsschrijvers in de toekomst mogelijk automatisch ondersteund kunnen worden door software.

In dit artikel presenteren we een annotatiesysteem voor teksten van moedertaalleerders van het Nederlands, en illustreren we de toepassing daarvan op een corpus schrijfopdrachten.

Hieronder bespreken we eerst het tekstmateriaal. Daarna beschrijven we het nieuwe systeem, genaamd USALT (Utrechts Systeem voor de Annotatie van LeerlingTeksten) en gaan we na in hoeverre de USALT-annotaties betrouwbaar zijn. In het tweede deel van het artikel gaan we meer inhoudelijk in op de annotaties. Eerst schetsen we de tekstkwaliteitsproblemen die we aantreffen in schrijfproducten van Nederlandse twaalf- tot veertienjarigen. Vervolgens analyseren we de verbanden tussen verschillende kwaliteitsaspecten die geannoteerd zijn, en tussen deze tekstkwaliteiten en de oordelen over de teksten. Ten slotte verkennen we de invloed van leerjaar en schoolniveau op tekstkwaliteiten en tekstoordelen.

2 Tekstmateriaal

Ons materiaal bestaat uit beschrijvende opdracht in het zogenaamde Schrijfmeterscorpus (De Glopper en Prenger 2013; zie Bijlage 1). Dit corpus is verzameld in het kader een onderzoeksproject dat in het schooljaar 2012-2013 werd uitgevoerd aan het toenmalige Etoc (Expertisecentrum taal, onderwijs en communicatie, Rijksuniversiteit Groningen). Aan dit project werd deelgenomen door een vijftal scholen voor voortgezet onderwijs. De scholen, die zich in reactie op een oproep hadden aangemeld voor het project, vormen een gelegenheidssteekproef uit de scholenpopulatie voor

voortgezet onderwijs, een waarvan de representativiteit onbekend is. Op iedere school werden drie schrijfpoddrachten afgenomen: in november 2012 een beschrijvende tekst, in februari 2013 een instructieve tekst en in april 2013 een argumentatieve tekst. De opdrachten werden afgenomen door de leerkracht Nederlands, binnen één lesuur. Van tijdsdruk was geen sprake en er zijn geen meldingen ontvangen over leerlingen die in tijdnood zijn gekomen. De leerlingteksten bevatten daar ook geen sporen van.

De beschrijvende tekst was een brief aan een Zweeds meisje dat volgende maand met haar ouders naar Nederland gaat verhuizen. De brief beantwoordt de vraag 'wat er allemaal typisch Nederlands is'. De opdracht suggereert enkele themadomeinen (sport, landschap, lekkernijen); er wordt gevraagd om minstens vier Nederlandse kenmerken te noemen en die zo helder mogelijk te beschrijven. Ten slotte wordt de leerling gevraagd de tekst het uiterlijk te geven van een brief.

Er zijn 371 brieven geanalyseerd, afkomstig van vijf scholen: uit Groningen (vwo jaar 1-2-3), Haren (vmbo jaar 1-2-3), Meppel (vmbo jaar 3), Utrecht (vwo jaar 1-2-3) en Winschoten (vmbo jaar 1-2-3, havo jaar 1-2-3). Tabel 1 geeft de verdeling van leerling-schrijvers over leerjaren en schooltypes.

Tabel 1 Leerjaren en schoolniveaus van de leerlingen in de Annika-taak uit het Schrijfmeterscorpus

Leerjaar	1	2	3	Totaal
Schoolniveau				
Vmbo	29	40	60	129
Havo	24	53	43	120
Vwo	28	47	47	122
Totaal	81	140	150	371

De 371 opstellen tellen gemiddeld 209 woorden. De tekstlengtes staan per subgroep in Tabel 2. De lengtes lopen op per leerjaar ($F[2,374] = 11.51$, $p < .001$, $\eta^2 = .058$) en met name per schoolniveau ($F[2,374] = 53.51$, $p < .001$, $\eta^2 = .222$); daarnaast is er een interactie tussen leerjaar en schoolniveau ($F[4,374] = 11.71$, $p < .001$, $\eta^2 = .111$) omdat havoleerlingen in de leerjaren 1 en 2 afwijken van de net genoemde trends.

Tabel 2 De tekstlengtes (afgerond) per leerjaar en schoolniveau

Leerjaar	1	2	3	Totaal
Schoolniveau				
Vmbo	147 (51)	212 (65)	224 (77)	175 (64)
Havo	202 (56)	165 (61)	240 (83)	190 (65)
Vwo	174 (68)	209 (61)	313 (66)	264 (84)
Totaal	190 (72)	201 (74)	227 (88)	209 (81)

De teksten zijn holistisch beoordeeld met behulp van een beoordelingschaal die vijf anker teksten bevatte met de scores 70, 85, 100, 115 en 130 erboven. De kwaliteitsverschillen tussen de vijf teksten betreffen met name de tekstinhoud en -opbouw (inleiding, kern en slot), de alineaopbouw en het gebruik van structuur aanduiders, zoals signaalwoorden. De zwakkere en sterkere teksten verschillen ook van elkaar in de kwaliteit van formuleringen en spelling, maar deze aspecten hebben bij de constructie en de toepassing van de schaal niet op de voorgrond gestaan.

Er is gewerkt met acht beoordelaars, die in 56 verschillende combinaties als driekoppige jury werkzaam zijn geweest. Zij kregen de instructie om elke tekst te scoren op basis van vergelijking met de vijf anker teksten. Er kon gescoord worden tussen 50 (lager dan de zwakste anker tekst) en 150 (hoger dan de beste anker tekst). De leerling teksten waren daarbij geanonimiseerd en van een code voorzien. Zo was het voor de beoordelaars niet duidelijk of de teksten geschreven zijn door (bijvoorbeeld) een vmbo- of gymnasium-leerling of door een jongen of een meisje. De gemiddelde betrouwbaarheid (Cronbach's alfa) van de oordelen van de 56 jury's was 0,86 ($sd=.10$). Het gemiddelde oordeel bedroeg 104, de standaarddeviatie 14.6 (zie verder Tabel 9).

We laten ter illustratie twee teksten volgen. De eerste heeft een lage score van 87, en de tweede een hoge score van 123. In de tekstvoorbeelden worden regelsprongen, voor zover duidelijk in het manuscript, aangegeven met insprongen; witregels zijn behouden.

Tekst 1

Beste Annika,

Holland staat voornamelijk bekend om kaas. In Holland besteedt men veel tijd met het bezig houden van voetbal. Er is veel voetbal op tv, er zijn veel voetbalclubs en bijna iedereen houdt wel van voetbal.

Een typisch Nederlands gerecht is boerenkool: aardappelen met kool en worst en spekjes.

Stroopwafels zijn ook typisch Nederlands, een tussendoor van twee koeken op elkaar met er tussen stroop.

Ik heb voetbal al genoemd, maar ook schaatsen is een bekende sport in Nederland. Nederlanders kijken niet alleen schaatsen op tv, ze doen het ook zelf als er eenmaal ijs op het water is. Na afloop drinken ze warme chocolademelk met slagroom, of kwast.

Kwast is oud-Nederlandse hete drank, met heet water, suiker en citroensap.

Het weer in Nederland is niet altijd even fijn, het is vaak koud.

Maar ondanks dat is Nederland toch wel een fijn land.

Jill

Tekst 2

Utrecht, 26 november 2012

Beste Annika,

Gisteren is aan mij verteld dat ik aan jou een brief moest schrijven over typisch Nederlandse dingen, leuk! In deze brief zal ik dus proberen wat over Nederland te vertellen. Hopelijk zul je alles begrijpen!

Nederland is een plat land, zonder reliëf. De wind waait er het hele jaar door en met regen zijn wij Nederlanders zeer vertrouwd.

Dat het Nederlandse weer niet vaak aangenaam is, is een minpuntje, maar je zult in niet veel landen zulke gezellige en knusse markten vinden zoals wij die hebben! Zeker rond de Kerst is het daar erg aangenaam.

In ons land hebben we, vast ook zoals in jouw land, bepaalde lekkernijen, zoals stroopwafels, pannenkoeken en haring! Bijna elke Nederlander vind dit wel lekker. Bovendien eten wij bij Sinterklaas (een feest in december) pepernoten. Dat zijn een soort kleine koekjes waaraan je je, als je niet oppast, helemaal miselijk kunt eten!

Naast veel eten kunnen Nederlanders ook goed sporten en er met name goed naar kijken. Nu heb ik het over voetbal, dé sport in Nederland. Als je met een Nederlander een gesprekje wilt houden, kun je het daar altijd over hebben!

Ik hoop dat je alles in mijn brief hebt begrepen en ik wens jou en je ouders een hele goede tijd in Nederland toe!

Hartelijke groeten,

Aron

3 Het Utrechtse systeem voor annotatie van leerlingteksten (USALT)

Zoals aangekondigd in de inleiding, benaderen we de kwaliteit van leerlingteksten door de teksten normatief te annoteren, met bijzondere aandacht dus voor het volgen van schrijftaal- en genrenormen. Aan een annotatiesysteem voor tekstproblemen kunnen verschillende eisen worden gesteld (Diaz-Negrillo & Domínguez 2006). Het moet allereerst expliciet zijn in de beschrijving van de annotaties (hier verder ‘tags’ genoemd) zodat betrouwbare codering mogelijk wordt. Daartoe is een helder codeboek bevorderlijk. De tweede eis betreft het detailniveau van de codes: dat moet zodanig zijn dat bekende fouten rechtstreeks herkenbaar zijn in de codering, terwijl anderzijds het aantal codes niet te groot mag worden. Het is ook een voordeel als de tagset hiërarchisch is opgebouwd, zodat zowel naar afzonderlijke soorten annotaties als naar groepen annotaties gezocht kan worden. Verder moet het systeem bruikbaar zijn voor de annotatie van nieuw tekstmateriaal.

Wat betreft de informatie in afzonderlijke probleemtags worden in de literatuur vaak vier soorten gegevens genoemd (Diaz-Negrillo & Domínguez 2006). (De mogelijke waarden van de eerste drie variabelen zijn vooraf bepaald, de vierde vormt een open tekstveld.)

- 1 De linguïstische eenheid c.q. het tekstniveau waarop een tag betrekking heeft (bijvoorbeeld leestekens, zinsbouw, tekstsamenhang). Het kiezen van de annotatie-eenheid is soms complex. USALT gaat daarbij af op het laagste niveau waarop een probleem opgelost kan worden. Als een probleem verholpen kan worden door een woord in te voegen, te schrappen, te verplaatsen of te vervangen, gaat het om een 'woordprobleem'. Daarmee is niet alles gezegd over de oorzaak van het probleem. Een woordprobleem kan immers voortkomen uit een zinsbouwprobleem, zoals later zal blijken.
- 2 Het probleemtype, waaronder we verder verstaan: de aard verandering die nodig is om het probleem op te lossen (bijvoorbeeld vervanging of weglating).
- 3 De woordsoort waarop de tag betrekking heeft, ingeval de annotatie beperkt blijft tot een woord.
- 4 De correcte vorm van de problematische passage, voor zover er één correct alternatief bestaat. Dat is bijvoorbeeld zo bij spellingproblemen, niet bij gepastheidsproblemen.

Dit alles betreft alleen genreonafhankelijke annotaties. We kennen geen annotatiesystemen gericht op genrespecifieke kwesties, maar gezien ons materiaal – dat bestaat uit opdrachten die verschillende genres betreffen – zijn die kwesties zeker van belang. Daarom bevat USALT zowel een algemeen als een genrespecifiek deel.

Bij het algemene deel van USALT hebben we de vier soorten informatie hierboven als uitgangspunt genomen. Aan de eis van explicietheit hebben we proberen te voldoen door gaandeweg een codeerhandleiding te ontwikkelen (Pander Maat, Raaijmakers, Veerbeek & Vermeulen 2017). De combinatie van gedetailleerdheid enerzijds en hiërarchie anderzijds hebben we gerealiseerd door de eerste drie informatietypen voor zover mogelijk met elkaar te kruisen. We onderscheiden 17 eenheden en 9 probleemttypen. Bij de 7 eenheden die het woordniveau niet te boven gaan, worden nog 11 woordsoorten onderscheiden. Een voorbeeld van een geannoteerde tekst is te vinden in Bijlage 2.

Tabel 3 bevat de matrix met combinaties van eenheden en probleemttypen, ingevuld met de frequenties waarmee de probleemsituaties zich voordeden in onze 371 opstellen. Onze voorlopige inschatting is dat de lege cellen combinaties betreffen die niet van toepassing zijn; vandaar dat we hier de nullen hebben weggelaten. Zo kan de spelling incorrect zijn of ongepast, maar niet ontbrekend; en onduidelijkheid is het enige probleem dat bij 'referenten' genoteerd is. Toch blijkt op basis van een beperkt aantal variabelen een gedetailleerde codering mogelijk: Tabel 3 bevat 61 verschillende soorten problemen.

Tabel 3 Algemene US-ALT-probleemtypen en -eenheden: frequenties (hoofd. = hoofdletter)

Eenheid	Probleem:	ontbreekt	moet weg	incorrect	volgorde verkeerd	onvolledig	niet parallel	niet duidelijk	stilistisch ongepast	storend herhaald	Totaal
1 Afbreking				11							11
2 Hoofd. naam	556	196									752
3 Symbool	5	1		1					9		16
4 Spatie	67	392		4					11		474
5 Spelling overig				1080					32		1112
6 Leesteken	1890	209		706	10				5		2820
7 Hoofd. zin	458	36		3							497
8 Flexie/affix				516			3				519
9 Woorddeel	14	7		2	2						25
10 Woord	345	105		324	57		8		94	25	958
11 Woordgroep	41	21		76	24		19		8	19	208
12 Zin				61	22	71	11	8	1	8	182
13 Referent								80			80
14 Informatie	15	3		88						6	112
15 Terug-verwijzing								127			127
16 Markering verband	9	23		22						1	55
17 Markering topicovergang	323			26					110	0	459
Totaal	3723	993		2920	115	71	41	215	270	59	8407

De eenheden en probleemttypen zijn grotendeels inductief tot stand gekomen, dus door verkenning van het materiaal. Dat gold alleen niet voor de eerste drie probleemttypen, die in vrijwel alle fout-annotatiesystemen voorkomen (Diaz-Negrillo & Domínguez 2006). In annotaties van teksten van tweede-taalleerders ligt de nadruk veelal op correctheid. Correctheid is, ruim opgevat, ook in ons systeem sterk vertegenwoordigd: de annotaties in de eerste zes probleemttypen betreffen veelal correctheid, namelijk wanneer het gaat om de eerste twaalf eenheden. Voor de eerste drie probleemttypen is dat duidelijk, maar het geldt ook voor incorrecte woordvolgordes, voor onvolledige zinnen (zonder werkwoord) en niet-parallel geformuleerde nevenschikkingen.

Daar waar de eenheden niet talig maar conceptueel van aard zijn (referent, informatie) of coherentie betreffen (terugverwijzing, markering verband, markering topicovergang) komt een andere dimensie van tekstkwaliteit aan de orde, namelijk de begrijpelijkheid van de formulering (van informatie en van de tekststructuur). Ook problemen rond de duidelijkheid van nieuwe referenten en terugverwijzingen vallen onder deze kwaliteitsdimensie.

Daarnaast kunnen zowel inhoudswoorden als verbandmarkeringen problemen vertonen wat betreft stilistische gepastheid, waarbij we doelen op problemen rond het formaliteitsniveau van de tekst (meestal gaat het om een te informele woordkeuze). Ten slotte zijn er soms aantrekkelijkheidsproblemen geannoteerd, waar het gaat om het storend herhalen van formuleringen of van informatie. Daarmee is de dimensie aantrekkelijkheid natuurlijk maar zeer ten dele gedekt; andere aantrekkelijkheidsverschijnselen laten zich minder eenvoudig annoteren in een ja/nee-format. Een tekst moet niet alleen stilistisch gepast zijn, maar ook genrenormen volgen wat betreft opbouw. En hier beperken onze annotaties zich niet meer tot het noteren van afwijkingen van schrijftaalnormen. Ze betreffen de kwaliteiten die specifiek zijn voor het tekstgenre (zie bv. Upton & Cohen 2009).

Een eerste belangrijk kwaliteitskenmerk daarin is het soepel realiseren van topicovergangen, door deze te markeren. Daarom zijn niet alleen problematische topicovergangen geannoteerd maar ook correcte. We komen daarop in paragraaf 5 nog terug. Daarvan is sprake wanneer de schrijver duidelijk maakt dat hij verandert van onderwerp, op terloopse wijze via *ook* (*Nederland hebben we ook veel feestdagen*) of meer expliciet (*het laatste wat je moet weten over Nederland*).

Een ander kwaliteitskenmerk in een brief als deze is dat de schrijver zich verplaatst in zijn lezer. Dat kan door vergelijkingen te maken tussen de Nederlandse en de Zweedse situatie. (Een voorbeeld is *In Nederland is het warmer dan in Zweden, maar in de winter sneeuwt het wel, maar vast niet zo veel als waar je vandaan komt.*)

Verder wordt de kwaliteit van een brief verhoogd door discourse-zetten die we om de briefkern heen aantreffen, die de brief plaatsen in de communicatiecontext, of de inhoud van de brief zelf aankondigen of recapitulieren:

- een aanhef (*Beste Annika,*), gevolgd door een regelsprong of witregel;
- een vermelding van datum en plaats (*Utrecht, 26 november 2012*);
- een zin waarin de schrijver zich voorstelt (*Ik ben Martijn en ik woon in Nederland*);
- een vermelding van de aanleiding voor de brief (*Ik heb gehoord dat je volgende maand naar Nederland verhuist*);
- een zin waarin Annika welkom wordt geheten (*Wat leuk dat je naar Nederland komt!*);
- een vooruitblik op de inhoud van de brief (*Je vroeg wat typisch Nederlands is en dat probeer ik je in deze brief te vertellen*);
- een terugblik op de inhoud van de brief (*Ik hoop dat je nu een beetje weet hoe het er in Nederland aan toe gaat!*);
- een vriendelijke slotzin (*Ik hoop dat je het erg leuk gaat krijgen!*);
- een groet met een naam erbij aan het eind (*Hartelijke groeten en tot snel, Tom Veldwijk*).

Bij de genrespecifieke annotaties gebruiken we niet alle annotatiedimensies (eenheid, probleemtype, woordsoort). Omdat het om uitingen gaat en niet om woorden, is de woordsoortdimensie niet van toepassing. Daarnaast is niet de aanwezigheid van een genre-element een probleem, maar het ontbreken ervan. Wanneer het element slechts als ‘aanwezig’ is gecodeerd, is dat dus goed nieuws. Natuurlijk kan een element minder geslaagd worden gerealiseerd; in dat geval wordt het probleemtype wel geannoteerd, zie Tabel 4. De meest voorkomende problemen betroffen onvolledige dagtekeningen en stilistisch minder gepaste groetformules. Omdat de probleemfrequenties in Tabel 4 laag zijn, durven we hier niet te zeggen of de nullen staan voor ‘onmogelijke’ combinaties. Ze zijn daarom gehandhaafd in de tabel.

Tabel 4 Genrespecifieke USALT-annotaties (genre=brief): frequenties

<i>Probleem:</i> Eenheid	<i>geen probleem (element is aanwezig)</i>	<i>in- correct</i>	<i>on- volledig</i>	<i>volgorde verkeerd</i>	<i>niet duidelijk</i>	<i>stilistisch ongepast</i>	<i>storend herhaald</i>	<i>Totaal</i>
1 Aanhef	340	0	1	0	1	7	0	349
2 Datum-plaats	79	1	40	11	0	0	0	131
3 Voorstellen	58	0	0	1	0	0	0	59
4 Aanleiding	156	0	0	0	0	1	0	157
5 Vooruitblik	197	0	0	0	0	6	1	204
6 Welkom	172	0	0	0	0	0	0	172
7 Empathie	96	1	0	0	0	0	0	97
8 Correcte	2	0	2	0	0	3	0	7
topicovergang								
9 Terugblik	124	0	0	1	0	9	0	134
10 Slotzin	286	0	0	4	0	3	0	293
11 Groet	293	1	0	0	0	36	0	330
Totaal	1803	3	43	17	1	65	1	1933

4 Annotatieprocedure en betrouwbaarheid

Het hele corpus is door twee codeurs doorgewerkt, waarbij twijfelgevallen met een derde codeur zijn besproken. Dit kostte ongeveer 1,25 uren per tekst en beliep een jaar. De codeerovereenstemming van de eerste twee codeurs is na twee maanden en na vijf maanden gemeten.

Het gaat bij het coderen om vier beslissingen. Ten eerste besluit een codeur om een passage al of niet te annoteren. Die keuze kan lastig zijn voor algemene annotaties rond passages waarvan de problematische aard niet evident is. Ten tweede wordt een eenheid toegekend, waarbij gekozen wordt uit 28 opties (17 algemene eenheden en 11 genrespecifieke elementen). Ten derde wordt al of niet een probleemtype toegekend. Ten vierde wordt aan annotaties van wooreenheden een woordsoort toegekend.

De eerste betrouwbaarheidsmeting ging over 430 eenheden (zowel algemene eenheden als genrespecifieke elementen), waarvan er 310 door beide codeurs zijn geannoteerd, 60 eenheden alleen door de eerste codeur, en 60 eenheden alleen door de tweede. Allereerst is de gezamenlijke betrouwbaarheid berekend voor de eerste twee beslissingen: het al of niet annoteren van een eenheid (we spreken verder van selectie) en de benoeming ervan. Dat betekent dat selectieverschillen hier de betrouwbaarheid drukken (daartoe kregen niet-geselecteerde eenheden in de vergelijkingsdata

een eigen code en eenheid). Cohen's Kappa voor de eenheidscode bedroeg inclusief selectieverschillen .62 ($n = 430$). Worden vervolgens de selectieverschillen buiten beschouwing gelaten (er wordt alleen gekeken naar eenheden die twee annotaties krijgen), dan is de Kappa .91 ($n = 310$). Voor de probleemcode is dezelfde tweesporen-benadering gevolgd. We beperken ons hier natuurlijk tot de eenheden waarin minimaal een codeur een probleem signaleert ($n=336$). Laten we zowel eenheden meetellen die slechts door één codeur worden geannoteerd als eenheden die slechts door één codeur problematisch worden geacht, dan bedraagt de Kappa .55. Voor eenheden met twee probleemannotaties is de Kappa voor probleemtypen .91 ($n = 242$). Wanneer we een samengestelde, gedetailleerde code creëren op basis van de codes voor eenheid en probleem, dan is de Kappa inclusief selectieverschillen .61 ($n = 336$); voor eenheden met twee annotaties op zowel eenheid als probleem is deze .88 ($n = 242$). Voor de derde dimensie, woordsoort, is de Kappa .98, waarbij we ons beperken tot eenheden die door beide codeurs geannoteerd zijn én waarbij woordsoorten relevant zijn ($n = 78$).

Dezelfde berekeningen zijn gemaakt bij de tweede meting; die ging over 1086 eenheden waarvan er 851 eenheden door beide codeurs zijn geannoteerd, 134 alleen door de eerste, en 101 alleen door de tweede codeur. Cohen's Kappa voor de eenheidscode bedroeg .73 wanneer selectieverschillen meetellen ($n = 1086$); worden deze buiten beschouwing gelaten, dan is de Kappa .97 ($n = 850$). Voor de probleemcode is de Kappa inclusief selectieverschillen .61 ($n = 736$), waarbij we ons beperken tot eenheden met minstens één probleemannotatie; voor eenheden met twee probleemannotaties is deze .95 ($n = 545$). Voor de code op basis van eenheid plus probleem is de Kappa inclusief selectieverschillen .67 ($n = 736$); voor eenheden met twee annotaties is deze .94 ($n = 545$). Voor woordsoort is de Kappa .93, waarbij we ons beperken tot eenheden met twee annotaties en daarbinnen tot de gevallen waarbij woordsoorten relevant zijn ($n = 193$).

Uit deze analyses blijkt ten eerste dat de betrouwbaarheden een stijgende lijn vertonen, en ten tweede dat de selectie van te annoteren passages de lastigste codeerbeslissing is. Toch zijn de kappa's ook inclusief deze meningsverschillen voldoende tot goed. De overeenstemming over passages met twee annotaties is bijna perfect.

Nadat het corpus gecodeerd is, zijn de annotaties in een database verzameld en nogmaals geïnspecteerd met de geannoteerde passages ernaast. Daarbij zijn duidelijke fouten en inconsistenties verbeterd. Zo is bijvoorbeeld strakker vastgesteld welke groeten 'niet passend' (te informeel) zijn genoemd (bv. *doei*).

5 De tekstkwaliteitsannotaties nader bekeken

We bespreken de gegevens uit Tabel 3 nu inhoudelijk. Een groot aantal annotaties betreft spelling. Twee spellingproblemen zijn bijzonder frequent. Ten eerste zijn er problemen met hoofdletters in namen (zie eenheid nummer 2). Vaak ontbreken die; soms staan ze er ten onrechte, zoals in *Cafetaria*. Ten tweede zijn er problemen met spaties (eenheid 4). Meestal staan die er ten onrechte (*Elf steden tocht*); soms ontbreken ze (*vanalles*). De overige spellingproblemen zijn niet specifiek gecodeerd (eenheid 5). Een analyse naderhand leert dat bij het overgrote deel van die problemen het verschil tussen de goed en fout gespelde versie van het woord niet te horen is (vgl. *s'nachts, bekendste, binnekort, boederij, kadootjes, Anika, 18, me hele leven, heb gehoord*). Bij 16% van de incorrect gespelde woorden gaat het om 'dt-fouten' in werkwoorden. Voor een meer diepgaande analyse van spellingproblemen in leerlingteksten, zie VandeKerckhove & Sandra (2015).

Een bijzonder spelprobleem is het gebruik van spellingen die gewoon zijn in informele genres. In dit geval is de spelling 'niet passend' genoemd; bij 13% van de spelproblemen gaat het om deze niet-passende vormen. Voorbeelden daarvan zijn verkorte vormen (*Gr. i.p.v. Groeten, dec i.p.v. december, NL i.p.v. Nederland*), het ontbreken van de spatie in *ofzo* en klinker-duplicatie (*heel lang*).

Frequenter dan spelfouten zijn fouten rond leestekens (eenheid 6). Zoals gezegd in par. 3, bevatten onze annotaties waar van toepassing ook een gecorrigeerde versie van de passage. In Tabel 5 vergelijken we voor leestekenfouten de originele tekst ('oud') vergeleken met de gecorrigeerde tekst ('correctie'). Daarbij zijn zeldzame leestekens (ampersand, gedachtestreepje, slash) buiten beschouwing gelaten en geschaard onder 'overig'.

Tabel 5 Problemen met leestekens (oud = originele tekst; nieuw = correctie)

Correctie:	Geen Leestekens	Komma	Punt	Punt- komma	Dubbele punt	En/of	Vraag- teken	Overig	Haakje	Totaal
Origineel										
Geen	-	1182	615	44	70	0	8	1	13	1933
Komma	89	-	341	66	64	39	3	2	1	605
Punt	43	90	-	13	21	3	8	0	0	178
Dubbelpunt	80	3	11	1	-	1	0	0	0	96
En/of	0	22	8	1	0	-	0	0	0	31
Aanhalings- teken	13	0	0	0	0	0	0	0	1	14
Overig	8	9	6	1	7	10	0	6	0	47
Totaal	233	1306	981	126	162	53	19	9	15	2904

De belangrijkste problemen uit Tabel 5 worden nu geïllustreerd, waarbij het ontbreken van een leesteken is aangegeven met een √-symbool en problematische leestekens zijn onderstreept.

- Ontbrekende komma's (*In Nederland zijn ook wat tradities √ zoals: Koninginnedag √ Prinsjesdag en Sinterklaas.*)
- Ontbrekende punten of puntkomma's (*Nederland is verdeeld in provincies √ dat heet weer anders dan in alle andere landen.*) Ontbrekende punten gaan vaak gepaard met ontbrekende hoofdletters aan het begin van de volgende zin.
- Ontbrekende dubbele punten (*Schaatsen √ dit deed je misschien in Zweden ook al.*)
- Ten onrechte geplaatste komma's (*Tot slot wil ik je nog iets vertellen over een hele bekende, Nederlandse bloem: de Tulp.*)
- Komma's die een punt of puntkomma moeten zijn (*Wat Nederlanders ook vaak eten is haring, dat is een soort vis.*)
- Komma's die een dubbele punt moeten zijn (*Het is ieder jaar een hele beslissing, gaat het door of niet.*)
- Voegwoorden die een komma moeten zijn (*Nederland is ook erg groen en we hebben veel weilanden met koeien en schapen en paarden.*)
- Ten onrechte geplaatste dubbele punten (*In Nederland zijn ook wat tradities zoals: Koninginnedag √ Prinsjesdag en Sinterklaas.*)
- Ten onrechte geplaatste punten (*Dus heel veel succes en heel veel plezier. in Nederland.*)
- Punten die een komma moeten zijn (*Niet in een keer natuurlijk, maar in plakjes.*)

Naast het ontbreken van punten is er een ander probleem rond het beginnen van nieuwe zinnen, dat we alleen annoteren wanneer de punt er wél staat: het ontbreken van hoofdletters (Tabel 3, eenheid 7).

Onder 'flexie/affix' (eenheid 8) vatten we verschillende soorten problemen:

- Incongruenties tussen onderwerp en persoonsvorm (*Hockey en voetbal is hier erg populair.*)
- Verkeerde meervouden (*musea's*)
- Verkeerde adjectiefverbuigingen (*typische Nederlandse dingen*)
- Verkeerd geslacht in verwijzwoorden en lidwoorden (*het is een klein snoepje die ...*)

Bij woordproblemen (eenheid 10) valt te onderscheiden tussen niet-passende, ontbrekende, ten onrechte aanwezige en incorrecte woorden. Bij de

niet-passende woorden gaat het om een te informele woordkeuze; voorbeelden zijn *wou / wilde* en *enne / en*.

Bij ontbrekende woorden zijn er twee verschillende soorten problemen aan de orde. Problemen met de constructie vinden we bijvoorbeeld in de vorm van collocatiefouten (*voetbal doen / aan voetbal doen; laag land / laag gelegen land*), ontbrekend existentieel *er* (*schapen zijn √ in verschillende soorten en maten*) en samentrekkingsfouten (*Tijdens Koninginnedag gaat iedereen met oranje kleren, √ haar en andere dingen de straat op.*). In andere gevallen is er geen regelrechte incorrectheid, maar eerder een stijl waarin bochten sterk worden afgesneden. Daarin ontbreken ook schakels als verbindingswoorden (*We hebben totaal geen bergen in ons land. √ geen echte bergen.*) en voorzetsels (*Ik leg hier √ nog een paar dingen uit wat het zijn*). Deze 'kort-door-de-bocht'-stijl vinden we ook bij het ontbreken van lidwoorden (*naar buitenland gaan*) en werkwoorden (*Wat echt een Nederlandse sport is, √ voetbal; ons scholen begint bij de peuterspeelzaal en daarna √ de basisschool, middelbare school en je vervolgopleiding.*) en bij het impliciet laten van contextueel bekende informatie (*Daar hebben we dijken en duinen zodat het √ niet overstroomt als het stormt.*)

Ten onrechte aanwezige woorden kunnen verschillende oorzaken hebben. Soms gaat het om dubbele informatie (*duidelijk typische Nederlandse dingen*), soms om de structuur van uitdrukkingen en naamwoordgroepen (*in de plaats daarvan; volgens een grootmoeders' recept*), soms om de structuur rondom voegwoorden (*aangezien dat*) of werkwoorden (*wat je het vond van mijn brief*).

Bij incorrecte inhoudswoorden gaat het vaak om kleine semantische misgrepen. Zo wordt Friesland *een streek in Nederland* genoemd, wordt gesproken over *nieuwe* in plaats van *jonge kaas*, en worden *komen* en *gaan* verward (*wat leuk dat je naar Nederland gaat*). Bij incorrecte functiewoorden gaat het vaker om collocatiefouten, zoals bij voorzetsels: *in het gebied van sport, besteedt men veel tijd met het bezighouden van voetbal*).

Deze observaties zijn gebaseerd op een verkenning achteraf. Als het corpus helemaal gecodeerd is, stellen we dit materiaal beschikbaar voor andere onderzoekers. Voorlopig is duidelijk dat bij veel woordproblemen lexicale, collocationale en grammaticale factoren door elkaar lopen.

Wanneer problemen alleen kunnen worden opgelost door meerdere woorden te veranderen, zijn ze op het niveau van de woordgroep geplaatst (eenheid 11). Zo moet er in *Bij ons in Nederland hebben wij veel verkeer en veel file en spits* meer dan één woord veranderd worden. Een probleem wordt op zinsniveau geplaatst wanneer de zin als geheel moet worden aangepakt.

Een voorbeeld is *stop dat plan maar in je achterhoofd*, waarmee iets als *dat kun je wel vergeten* lijkt te worden bedoeld.

Het onderscheid tussen woordgroep- en zinsniveau (eenheid 12) is soms lastig. In *De Nederlandse taal is best moeilijk, want we zijn daar heel slordig in* kun je proberen om een kleine correctie te doen in de tweede deelzin, maar wellicht zou eigenlijk in beide deelzinnen moeten worden ingegrepen, om op een zin als *De Nederlandse taal is best moeilijk te volgen, want Nederlanders spreken die heel slordig uit* te komen.

Bij de eenheden referent, informatie en terugverwijzing (eenheden 13-15) gaat het niet meer om taaleenheden, maar om *conceptuele* eenheden. Bij referentproblemen gaat het om nieuwe entiteiten die eigenlijk introductie behoeven. Daarbij kan het gaan om evenementen (*Sinterklaas*) of locaties (*Afsluitdijk*), maar ook om gewoonten zoals in *de drie zoenen is ook iets typisch*. De eenheid 'informatie' wordt vooral gebruikt bij onjuiste uitspraken (*Nederland wordt geregeerd door Koningin Beatrix*.) Bij onduidelijke terugverwijzingen gaat het meestal om voornaamwoorden met onduidelijke of impliciete antecedenten.

De laatste twee eenheden betreffen de samenhang in de tekst. Bij verbandmarkeringen (eenheid 16) gaat het om relaties tussen uitingen over eenzelfde topic. Een voorbeeld van een ten onrechte geplaatste markering is *maar in het is een typisch Nederlandse sport, maar echt heel populair*. Een incorrecte markering is bijvoorbeeld *doordat in Pannenkoeken zijn typisch Nederlands doordat je overal wel pannenkoeken kan kopen*.

Interessant in de Annika-brief is de manier waarop de topicovergangen worden vormgegeven (eenheid 17). We kijken nog eens naar een fragment uit Tekst 1, waarin topicovergangen verschillende vormen aannemen.

Dat het Nederlandse weer niet vaak aangenaam is, is een minpuntje, maar je zult in niet veel landen zulke gezellige en knusse markten vinden zoals wij die hebben! Zeker rond de kerst is het daar erg aangenaam.

In ons land hebben we, vast ook zoals in jouw land, bepaalde lekkernijen, zoals stroopwafels, pannenkoeken en haring! Bijna elke Nederlander vindt dit wel lekker. Bovendien eten wij met Sinterklaas (een feest in december) pepernoten. Dat zijn een soort kleine koekjes waaraan je je, als je niet oppast, helemaal misselijk kunt eten!

Naast veel eten kunnen Nederlanders ook goed sporten en er met name goed naar kijken.

Als een correcte overgangsmarkering is gecodeerd ‘naast veel eten’ in de laatste zin, waarmee de overgang naar het thema sport wordt aangekondigd. Een wat ruwe overgang is daarentegen die in de eerste zin, die gaat van het weer naar ‘gezellige markten’. Het *maar* wordt op deze plaats als een incorrecte overgangsmarkering geannoteerd. Een voorbeeld van een ontbrekende overgangsmarkering vinden we bij de overgang van kerstmarkten naar Nederlandse lekkernijen. We hebben overigens niet gekeken naar regelsprongen als overgangsmarkeringen, omdat ons materiaal op dit punt niet helemaal betrouwbaar is.

De algemene tekstproblemen overziend, kunnen we de volgende probleemgroepen onderscheiden:

- spellingproblemen in ruime zin (afbreking, hoofdletters binnen zinnen, symbolen, spaties, overige spellingproblemen);
- leestekenproblemen (leestekens, hoofdletters aan het begin van de zin);
- grammaticale en lexicale problemen (flexie/affix, woorddeel, woord, woordgroep, zin);
- coherentieproblemen (onduidelijke referent, terugverwijzing, markering verband, incorrecte topicovergang);
- Inhoudelijke problemen (informatie).

De hoge frequentie van spelling- en leestekenproblemen betekent dat automatische analyses van ruwe leerlingteksten weinig valide zullen zijn, zowel voor lexicale als syntactische kenmerken.

Bezien we nu de genrespecifieke problemen. Uit Tabel 4 kan niet direct worden afgelezen hoeveel brieven een genrespecifiek element vertonen, omdat sommige brieven meerdere voorkomens bevatten van een bepaald element. Uit een analyse per brief blijkt dat twee basale briefconventies massaal gevolgd worden. Van de 371 brieven beginnen er 347 met een aanhef en eindigen er 327 met een groet, al is die groet soms te kortaf (alleen de naam) of te informeel. In een enkel geval ontbreekt een witregel onder de aanhef of boven de groet. De derde formele conventie, de datum en plaats (dagtekening), vinden we daarentegen in slechts 128 brieven, en dan nog regelmatig onvolledig.

De andere briefelementen kunnen we enerzijds beschouwen als tekststructurende elementen en anderzijds als beleefdheidswerk. Wat betreft de structurelementen hebben 250 brieven een of meer slotzinnen en geven 134 brieven een terugblik op de inhoud; 281 brieven hebben minimaal

één van deze slotelementen. 150 brieven noemen de aanleiding voor het schrijven (de verhuizing van Annika) en 193 brieven hebben een vooruitblik; 241 brieven hebben minimaal één van deze openingselementen.

Wat betreft beleefdheid blijkt dat 148 schrijvers Annika welkom heten en 49 schrijvers zich voorstellen. 73 schrijvers verplaatsen zich in Annika door vergelijkingen te maken tussen Nederland en Zweden. Maar ook de 250 slotzinnen kunnen als een beleefdheidselement worden gezien, omdat ze vrijwel altijd een vriendelijke inhoud hebben (vgl. in Tekst 2 *ik wens jou en je ouders een hele goede tijd in Nederland toe*).

6 De samenhang tussen tekstkwaliteiten en tekstoordelen

Eerst hebben we voor elke tekst het aantal woorden geteld, uitgaande van versies met gecorrigeerde spelling. Vervolgens zijn per brief de volgende grootheden bepaald:

- het aantal spellingproblemen per 100 woorden (incl. spatieproblemen);
- het aantal leestekenproblemen per 100 woorden (incl. hoofdletterproblemen aan het begin van de zin);
- het aantal lexicale problemen (woorddeel- en woordproblemen) per 100 woorden;
- het aantal grammaticale problemen (woordgroep- en zinsproblemen) per 100 woorden;
- het aantal coherentieproblemen per 100 woorden (problemen rond referenten, terugverwijzingen, topicovergangen en relatiemarkeringen);
- het aantal correct gemarkeerde topicovergangen per 100 woorden;
- het aantal structurerende brieuelementen (een schets van de aanleiding voor de tekst, een aankondiging van de globale inhoud, een terugblik op de inhoud of andere afsluitzin);
- het aantal formele brieuelementen (aanhef, plaats en datum, groet);
- het aantal beleefdheidselementen in de brief (schrijver stelt zich voor, heet Annika welkom, of maakt vergelijkingen tussen Zweden en Nederland).

Bij de brieuelementen hebben we de (schaarse)problemen in de realisatie buiten beschouwing gelaten; de aanwezigheid ervan vatten we op als aanwijzing dat de schrijver beseft heeft van de betreffende conventie. Inhoudsproblemen laten we voorlopig buiten beschouwing, omdat die geen tekstkwaliteit in strikte zin betreffen.

Zoals par. 2 beschreven, zijn de teksten voorzien van holistische kwaliteitsscores. Het is nu interessant om na te gaan welke tekstkenmerken helpen deze oordelen te voorspellen. Andersom gesteld, kunnen onze annotaties meer concrete informatie geven over de kenmerken van teksten die als goed zijn beoordeeld. We doen regressieanalyses waarin de voornoemde grootheden als voorspeller voor het oordeel fungeren. Daarbij kijken we behalve naar annotatiekenmerken ook naar tekstlengte. Dat kenmerk is namelijk veelal de krachtigste voorspeller voor tekstoordelen, zoals al blijkt uit De Gloppe (1984). Ook in ons geval is goed voorstelbaar dat een brief van 100 woorden als minder geslaagd, want minder informatief, wordt gezien dan een van 200 woorden.

We verkennen eerst de correlaties tussen de verschillende tekstkwaliteiten en tekstoordelen, en die tussen tekstkwaliteiten. Tabel 6 laat zien dat het tekstoordeel positief correleert met tekstlengte en de toegevoegde genre-elementen, en negatief met alle probleemttypen, uitgezonderd het aantal correct gemarkeerde topicovergangen. De hoogste correlatie vertoont het oordeel met de tekstlengte ($r=.72$), die dus op zich al 52% van de oordeelsvariantie verklaart.

De correlaties tussen tekstkenmerken vertonen een samenhangend beeld. Tekstlengte correleert negatief met alle probleemttypen, uitgezonderd het correct markeren van topicovergangen; tekstlengte correleert daarentegen positief met de toegevoegde genre-elementen. De probleemttypen correleren positief met elkaar, wederom uitgezonderd het correct markeren van topicovergangen. De correlaties tussen toegevoegde elementen enerzijds en probleemttypen zijn meestal lager of niet significant. Een uitzondering vormen de correlaties tussen briefconventies (formele elementen) en structurerende elementen enerzijds en spelling anderzijds ($-.28$).

We concluderen dat het aantal correct gemarkeerde topicovergangen zich eigenaardig gedraagt. Bij nader inzien zijn we waarschijnlijk te scheutig geweest met deze code, door bijvoorbeeld een markering als *ook* al voldoende te achten.

We onderzoeken de onderlinge relaties verder met een factoranalyse met Promaxrotatie, waarbij correlatie tussen factoren is toegestaan (overigens geeft een Varimaxrotatie soortgelijke resultaten); daarbij laten we tekstlengte en de topicovergangen buiten beschouwing. Tabel 7 presenteert een oplossing met drie factoren, die 56.7% van de variantie verklaart.

Tabel 6 Correlaties tussen tekstoordeel, tekstlengte en annotatiekenmerken

	Tekst-oordeel	Tekst-lengte	Spelling	Inter-punctie	Lexicaal	Grammaticaal	Coherentie	Topic-overgang	Formeel element	Structurend element	Beleefdheid
Score	-										
Tekstlengte	0,72	-									
Spellingprobleem	-0,37	-0,34	-								
Interpunctieprobleem	-0,40	-0,39	0,42	-							
Lexicaal probleem	-0,28	-0,27	0,26	0,31	-						
Grammaticaprobleem	-0,17	-0,14	0,19	0,34	0,29	-					
Coherentieprobleem	-0,38	-0,34	0,23	0,35	0,23	0,23	-				
Correcte markering topicovergang	-0,08	-0,14	0,13	0,25	0,10	0,12	0,15	-			
Formeel element	0,36	0,24	-0,28	-0,21	-0,18	-0,10	-0,21	-0,08	-		
Structurend element	0,51	0,34	-0,28	-0,14	-0,11	-0,05	-0,17	0,15	0,21	-	
Beleefdheidselement	0,14	0,15	-0,14	-0,12	-0,11	-0,02	-0,06	-0,05	0,14	-0,02	-

Voor correlaties boven de .10 geldt $p < .05$, boven de .13 $p < .01$, boven de .17 $p < .001$

Tabel 7 Principal Components-analyse (Promaxrotatie) op annotatieaspecten

Kwaliteitsaspect	Factor 1 'lokale problemen'	Factor 2 'structuur'	Factor 3 'beleefdheid'
Spellingprobleem	0,29	-0,52	-0,01
Leestekenprobleem	0,67	-0,31	-0,06
Lexicaal probleem	0,65	0,01	-0,09
Grammaticaprobleem	0,84	0,22	0,18
Coherentieprobleem	0,45	-0,27	0,03
Formeel element	0,03	0,58	0,31
Structurerend element	0,12	0,87	-0,29
Beleefdheidselement	0,04	-0,07	0,94

Er valt voorzichtig te onderscheiden tussen een factor met 'lokale tekstproblemen' enerzijds (leestekens, grammaticaal, lexicaal, coherentie) en twee factoren die primair genregebonden zijn anderzijds, namelijk 'structuur' (formeel element, structurerend element) en 'beleefdheid'. Het valt op dat spellingproblemen meer laden op de factor 'structuur' dan op de lokale factor. Dit lijkt een voorbeeld van de verbanden tussen kenmerken van lagere en hogere orde waarop Dean & Quinlan (2010) doelen in het citaat in paragraaf 1.

Kijken we vervolgens naar de relaties tussen annotatiekenmerken en tekstoordelen. Tabel 8 toont twee regressiemodellen. Het eerste model bevat alleen annotatiekenmerken en verklaart 45% van de variantie in oordelen; naast interpunctie en lexicale problemen zijn de andere drie voorspellers tekststructureel van aard. Dat klinkt aardig, maar alleen tekstlengte voorspelt ook al 52% variantie. Wanneer we aan een model met tekstlengte annotatiekenmerken toevoegen, ontstaat het krachtigste model, dat 65% variantie verklaart. In dat model zijn lexicale problemen niet meer opgenomen als voorspeller. We concluderen dat interpunctie en tekststructurele kenmerken de voorspelling op basis van alleen tekstlengte kunnen verbeteren.

De gegevens uit Tabel 6 en 8 leren ons iets over wijze waarop de beoordelaars van De Glopper & Prenger (2013) taalverzorgingsproblemen hebben meegewogen in het oordeel. Hoewel Tabel 6 toont dat spelling correleert met interpunctie en tekststructurele aspecten, blijkt in de regressie dat die laatste aspecten toch zwaarder wegen voor de beoordelaars. Dat is in overeenstemming met de beoordelingsinstructies die in par. 2 zijn besproken; het lijkt ook wenselijk, omdat spelling iets anders is dan tekstkwaliteit. Men zou nu kunnen betogen dat dit net zo goed geldt voor leestekens; maar dat aspect weegt wel degelijk mee in de tekstoordelen. Leestekenproblemen worden dus als hinderlijker ervaren door beoordelaars dan spellingproblemen.

Tabel 8 Twee regressiemodellen die tekstoordelen voorspellen

Model	Predictoren	B	SE	Beta	t	p
Alleen annotatiekenmerken						
1 R = .673; R ² = .453; SE = 10.87	Structurerend element	4,29	0,43	0,40	9,97	p < .001
	Interpunctieprobleem	-0,99	0,19	-0,22	-5,13	p < .001
	Formeel element	4,01	0,91	0,18	4,43	p < .001
	Coherentieprobleem	-1,83	0,45	-0,17	-4,08	p < .001
	Lexicaal probleem	-1,23	0,52	-0,10	-2,35	p < .05
Tekstlengte en annotatiekenmerken						
2 R = .804; R ² = .646; SE = 8.74	Tekstlengte	0,09	0,01	0,53	14,33	p < .001
	Structurerend element	2,88	0,36	0,27	7,99	p < .001
	Formeel element	3,10	0,73	0,14	4,25	p < .001
	Interpunctieprobleem	-0,44	0,16	-0,10	-2,79	p < .01
	Coherentieprobleem	-0,97	0,36	-0,09	-2,67	p < .01

7 Frequentie van annotaties per leerjaar en schoolniveau

Hoe vaak komen de verschillende verschijnselen nu voor in de praktijk? We beperken ons hier tot de vijf annotatiekenmerken uit Tabel 8, plus spelling.

Omdat de algemene kenmerken in Tabel 9 geteld zijn per 100 woorden, geven ze een helderder beeld van probleemfrequenties dan mogelijk was in Tabel 3 en 4. Uit de gemiddelden op de onderste regel van Tabel 9 blijkt om te beginnen dat een op de 30 woorden verkeerd gespeld is. Op elke 21 woorden doet zich een leestekenprobleem voor, op elke 52 woorden een lexicaal probleem en op elke 90 woorden een coherentieprobleem. Eigenlijk ligt het meer voor de hand om leesteken- en coherentieproblemen te standaardiseren op zinnen of deelzinnen dan op woorden. Omdat de zinsgrenzen minder betrouwbaar vast te stellen zijn (juist door de leestekenproblemen), werken we met de deelzinslengte, dat wil zeggen het aantal woorden per vervoegd werkwoord. Via T-Scan (Pander Maat 2017) hebben we vastgesteld dat die in dit corpus gemiddeld 9,5 woorden bedraagt. Dat betekent dat er ongeveer elke 2 deelzinnen een leestekenprobleem is, en elke 9 deelzinnen een coherentieprobleem. We zijn geneigd om de leestekenproblemen als de meest problematische categorie te beschouwen.

Bezien we vervolgens de verschillen tussen leerjaren en schoolniveaus. Tabel 10 laat zien dat de hoofdeffecten overal significant zijn, behalve het leerjaareffect bij coherentieproblemen. De hoofdeffecten worden echter behoorlijk genuanceerd door vier interacties. Ten eerste gedragen de

Tabel 9 Gemiddelden per leerjaar en schoolniveau voor zes tekstkenmerken

Leer-jaar	School-niveau	Spelling-probleem/ 100 w	Inter-punctie-probleem/ 100 w	Lexicaal-probleem/ 100 w	Coherentie-probleem/ 100 w	Formeel element/ brief	Structurerend element/ brief	Tekst-oordeel
1	Vmbo	5.16 (3.02)	7.13 (3.70)	2.71 (1.59)	2.01 (2.37)	1.62 (0.82)	1.45 (1.40)	90 (11.2)
	Havo	3.91 (2.31)	5.42 (2.98)	1.86 (0.91)	0.93 (1.10)	1.79 (0.41)	2.13 (1.33)	102 (7.8)
	Vwo	4.07 (3.07)	4.11 (2.84)	1.15 (1.07)	1.10 (1.40)	1.82 (0.48)	1.43 (1.23)	101 (13.8)
	Totaal	4.41 (2.87)	5.58 (3.42)	1.92 (1.39)	1.38 (1.79)	1.74 (0.61)	1.64 (1.34)	97 (12.5)
2	Vmbo	3.53 (1.94)	5.60 (2.84)	1.73 (1.14)	0.65 (0.86)	2.23 (0.62)	2.18 (1.22)	104 (10.9)
	Havo	2.80 (2.01)	5.22 (3.30)	1.69 (1.12)	1.18 (1.20)	2.34 (0.55)	1.85 (1.46)	98 (11.8)
	Vwo	2.57 (2.00)	3.39 (2.42)	1.25 (1.11)	0.84 (0.82)	2.32 (0.59)	2.23 (1.20)	107 (10.3)
	Totaal	2.93 (2.01)	4.71 (3.04)	1.55 (1.14)	0.91 (1.01)	2.30 (0.58)	2.07 (1.31)	103 (11.7)
3	Vmbo	4.12 (2.80)	6.05 (3.84)	1.82 (1.08)	2.11 (2.40)	2.03 (0.52)	1.93 (1.33)	99 (13.2)
	Havo	3.35 (2.38)	4.20 (2.46)	1.37 (0.93)	0.59 (0.74)	2.37 (0.76)	2.37 (1.24)	107 (11.9)
	Vwo	1.52 (1.07)	1.88 (1.43)	0.81 (0.57)	0.46 (0.75)	2.62 (0.57)	3.13 (1.15)	125 (9.8)
	Totaal	3.08 (2.51)	4.21 (3.35)	1.38 (0.99)	1.16 (1.80)	2.31 (0.66)	2.43 (1.34)	109 (16.2)
Totaal	Vmbo	4.17 (2.77)	6.15 (3.54)	1.99 (1.28)	1.63 (2.14)	2.00 (0.66)	1.90 (1.33)	98 (13.0)
	Havo	3.22 (2.23)	4.89 (2.98)	1.61 (1.02)	0.92 (1.06)	2.24 (0.65)	2.09 (1.37)	102 (11.8)
	Vwo	2.51 (2.24)	2.97 (2.38)	1.06 (0.94)	0.75 (0.99)	2.32 (0.63)	2.39 (1.35)	113 (14.9)
	Totaal	3.32 (2.48)	4.70 (3.29)	1.91 (1.42)	1.11 (1.55)	2.18 (0.66)	2.12 (1.36)	104 (14.6)

Gemiddelden en standaarddeviaties (lex.-gramm.=lexicaal-grammaticaal; /100 w = per 100 woorden)

Tabel 10 Resultaten van variantieanalyses met leerjaar en schoolniveau als onafhankelijke variabelen en zes tekstkenmerken als afhankelijke variabelen

	Spelling-probleem	Inter-punctie-probleem	Lex.-gramm. Problem	Coherentie-probleem	Formeel element	Structurerend element	Tekst-Oordeel
F leerjaar [2,362]	11.56***, $\eta_p^2 = .060$	6.89**, $\eta_p^2 = .037$	7.40**, $\eta_p^2 = .039$	n.s.	28.75***, $\eta_p^2 = .137$	10.54***, $\eta_p^2 = .055$	34.15***, $\eta_p^2 = .160$
F niveau [2,362]	13.37***, $\eta_p^2 = .069$	32.92***, $\eta_p^2 = .154$	26.58***, $\eta_p^2 = .128$	10.19***, $\eta_p^2 = .053$	7.32**, $\eta_p^2 = .039$	3.04*, $\eta_p^2 = .029$	41.36***, $\eta_p^2 = .188$
F LxN [4,362]	2.60*, $\eta_p^2 = .028$	n.s.	2.04*, $\eta_p^2 = .026$	7.51***, $\eta_p^2 = .077$	n.s.	4.90**, $\eta_p^2 = .051$	16.60***, $\eta_p^2 = .156$

F^{LxN} = F-waarde voor de interactie tussen leerjaar en niveau; * = $p < .05$; ** = $p < .01$; *** = $p < .001$; η_p^2 = partial eta-squared

havo-klassen zich per leerjaar verschillend; zo scoort havo-1 op spelling en coherentie even goed als vwo-1, en op structurerende elementen zelfs het hoogst; maar havo-2 heeft dan weer evenveel lexicale problemen als vmbo-2, en minder structurerende elementen. De afwijkende posities van havo-1 en havo-2 zijn op dezelfde manier zichtbaar in de tekstoorden. Ten

tweede valt op dat vmbo-2 in een aantal opzichten beter scoort dan vmbo-3. Kortom, het beeld is grillig.

In het algemeen is de binnengroepsvariantie groot en zijn de effecten dus klein (de η^2 is maximaal .19). Bij interpunctie en lexicale problemen zijn de effecten van schoolniveau veel groter dan die van leerjaar, terwijl dat bij formele briefelementen andersom is. Omdat tekstoordelen gevoelig zijn voor meerdere elementen, is het begrijpelijk dat zij voor zowel leerjaar als schoolniveau een redelijke effectgrootte laten zien (resp. .16 en .19). Wij hebben geen informatie over de aandacht die op de deelnemende scholen besteed is aan de verschillende tekstkenmerken, zodat het moeilijk is om deze verschillen te interpreteren.

8 Conclusies en discussie

In dit artikel hebben we USALT geïntroduceerd, een systeem voor de handmatige annotatie van tekstkwaliteitsaspecten in schrijfproducten. USALT maakt onderscheid tussen algemene tekstkenmerken (spelling, interpunctie, formulering en coherentie) en genrespecifieke kenmerken (zoals in dit geval formele briefelementen, structurerende elementen en beleefdheidselementen). USALT onderscheidt drie annotatiedimensies: geannoteerde eenheid, probleemtype en geannoteerde woordsoort. Door de dimensies waar mogelijk te kruisen, wordt een behoorlijke mate van detail per annotatie bereikt. Naast correctheid geven de annotaties een beeld van de tekstkwaliteitsdimensies begrijpelijkheid en gepastheid (zowel stilistisch als wat betreft genrenormen). Het systeem is toegepast op 371 teksten uit de eerste drie klassen van het voortgezet onderwijs; de schrijfpdracht vroeg om een brief aan een Zweedse leeftijdgenoot over 'typisch Nederlandse' verschijnselen. Na training leidt USALT tot betrouwbare annotaties. Binnen een jaar zullen USALT-annotaties van ruim 800 schrijfproducten afkomstig van twee schrijftaken beschikbaar worden gesteld aan onderzoekers.

De brede scope van de USALT-annotaties leidt tot een gedetailleerd beeld van tekstkwaliteitsproblemen en hun frequentie. Bij de algemene problemen bleken leestekenproblemen het meest frequent te zijn, met name ontbrekende komma's en punten. Daarnaast zijn er natuurlijk spellingproblemen, alsmede allerlei formuleringsproblemen: lexicale problemen, problemen met woordgroepen en zinnen, en verschillende soorten coherentieproblemen. (Bij 'lexicale' problemen gaat het overigens vaak om problemen rondom de correctheid van woordreeksen, dus eigenlijk om constructie- en collocatieproblemen.) Bij de genrespecifieke elementen

blijkt bijvoorbeeld dat het merendeel der brieven een aanhef en groet bevat, maar slechts een minderheid een dagtekening. Driekwart van de brieven bevat een zin die als afsluiting te herkennen is, twee derde van de brieven een inleidende zin.

Vervolgens zijn relaties onderzocht tussen de verschillende tekstkwaliteitsaspecten, en tussen tekstkwaliteiten en oordelen over de teksten. Wat betreft de onderlinge relaties valt vooral op dat spellingproblemen meer samenhangen met genregebonden kenmerken dan met de andere algemene kenmerken. Is dat wellicht omdat het in beide gevallen gaat om pure schrijftaalconventies? Wij menen van niet. Immers, wij beschouwen de afwezigheid van inleidende en afsluitende zinnen een signaal voor communicatieve kwaliteit: wie met de deur in huis valt en plotsklaps eindigt, heeft weinig aandacht voor de ontvanger van een tekst. Zeker voor deze structurerende elementen constateren we dus een link tussen tekstkenmerken van lagere en hogere orde.

Wat betreft de relatie tussen tekstkenmerken en tekstoordelen blijkt dat met alleen de geannoteerde tekstkwaliteiten 45% van de variantie in tekstoordelen verklaard kan worden. Wanneer we ook tekstlengte als voorspeller gebruiken, verklaren vier tekstkwaliteiten 13% extra oordeelsvariantie bovenop de 52% variantie verklaard door tekstlengte: structurende tekstelementen, formele brievelementen, interpunctieproblemen en coherentieproblemen.

Het belang van tekstlengte als voorspeller van tekstoordelen is in onze data overigens volkomen begrijpelijk, om meerdere redenen. Ten eerste zijn er directe links tussen tekstlengte en communicatieve effectiviteit. Een korte tekst is immers onvoldoende om een buitenlandse lezer een concreet beeld te geven van de Nederlandse cultuur, omdat hij simpelweg te weinig informatie bevat. Bovendien maakt het verwerken van genre-elementen teksten langer; veelal zullen de teksten waarin die elementen ontbreken korter zijn. Kortom, in de woorden van Chodorow & Burstein (2004, 8): “many of the features of highly scored essays are difficult to embed in a shorter essay”. En inderdaad blijkt uit de data dat de teksten van minder dan 120 woorden ($n=43$) een lage gemiddelde score hebben van 84. Opvallender is dat het verband tussen tekstlengte en tekstoordeel behouden blijft bij hogere tekstlengtes, tot boven de 300 woorden. We zijn geneigd dat toe te schrijven aan een indirecte link tussen tekstlengte en tekstkwaliteit: betere schrijvers produceren makkelijker teksten van enige omvang; zij hebben een hogere ‘writing fluency’ (Chenoweth & Hayes 2001). De langere teksten zijn dus veelal afkomstig van betere schrijvers, zodat lengte een indicatie kan zijn van tekstkwaliteitsaspecten die nog niet specifiek gemeten kunnen worden.

Wat leert onze analyse over de wijze waarop tekstoordelen tot stand komen? Om te beginnen valt op dat spelling en grammatica niet de hoofdrol spelen. Dat is in die zin geruststellend dat de beoordelaars mee zijn gegaan in de nadruk die het beoordelingsvoorschrift en de anker teksten leggen op tekststructuur. Dat wil niet zeggen dat de beoordeling geheel op diepgaande tekstanalyse berust. Twee van de aspecten die meewegen in het oordeel (interpunctie en formele briefelementen) zijn in één oogopslag te beoordelen; maar de andere twee (structureerende tekstelementen en coherentieproblemen) vergen wel degelijk interpretatie van de tekst.

Ten slotte zijn relaties geanalyseerd tussen tekstkenmerken en tekstoordelen enerzijds en leerjaar en schoolniveau anderzijds. Naast globale verbeteringen in de verwachte richting stuiten we daarbij op nogal wat interacties tussen leerjaar en schoolniveau, die te maken hebben met grilligheden in de steekproef. Niet alle scholen hebben dezelfde leerjaren en schoolniveaus bijgedragen. Verder is het aantal waarnemingen per cel bescheiden.

We concluderen dat we een werkbaar annotatiesysteem hebben ontwikkeld, dat inzicht biedt in de aard en de frequentie van tekstkwaliteitsproblemen, helpt bij het inzichtelijk maken van oordelen over tekstkwaliteit, en licht werpt op de relaties tussen verschillende aspecten van tekstkwaliteit. In vervolgonderzoek naar de Annika-brieven denken we vooral aan een verbeterde annotatie van topicovergangen en hun markeringen. Verder zullen we naast deze descriptieve taak ook argumentatieve teksten van dezelfde schrijvers annoteren. Daarbij komen andere tekstkwaliteiten aan bod, zoals de breedte en de diepte van de argumentatie. De analyse van topics en argumentatie stelt ons wellicht in staat om meer te begrijpen van de tekststructurele en inhoudelijke tekstkwaliteiten, en zo wellicht ook het tekstlengte-effect op tekstoordelen beter te doorzien. Verder maken we USALT zo geschikt voor meerdere schrijftaken, waardoor een breder inzicht ontstaat in de relatie tussen algemene en genrespecifieke tekstkwaliteiten. Daarnaast kunnen we op termijn met de gecorrigeerde teksten nagaan wat een automatische tekstanalyse kan toevoegen aan ons inzicht in de relatie tussen tekstkenmerken en tekstkwaliteit.

Wat is ten slotte de praktische relevantie van dit werk? Ons annotatiesysteem is zelf niet bedoeld voor gebruik in de onderwijspraktijk. Maar annotatie-onderzoek geeft een beeld van de meest voorkomende problemen voor verschillende soorten schrijftaken, en op verschillende niveaus van schrijfvaardigheid; en zo'n beeld lijkt ons heel nuttig voor schrijfdocenten en -didactici, zo niet onmisbaar.

Referenties

- Bazerman, C. (2015). What do sociocultural studies of writing tell us about learning to write. In C. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (Vol. 2, pp. 11-23). New York: Guilford.
- Bochove, A. van, & Vermeer, A. (2015). Schrijvers in spe? Stilistische kenmerken in verhalen van groep 5 tot en met 8. *Tijdschrift voor Taalbeheersing* 38(3), 305-336.
- Boekelder, A., Gloppe, K. de, & Wijk, C. van (1991). Tekstanalytisch onderzoek van een functionele schrijfpdracht: de inhoud, opbouw en beoordeling van excuusbrieven. *Toegepaste Taalwetenschap in Artikelen*, 40, 23-34.
- Bouwer, R., Koster, M. Van der Bergh, H. (2016). Benchmark rating procedure, best of both worlds? Comparing procedures to rate text quality in a reliable and valid manner. Manuscript submitted for publication.
- Chenoweth, N.A., & Hayes, J.R. (2001). Fluency in writing: Generating text in L1 and L2. *Written communication*, 18(1), 80-98.
- Chodorow, M., & Burstein, J. (2004). Beyond essay length: evaluating e-rater's performance on toefl® essays. *ETS Research Report Series*, 2004(1).
- Christie, F. (2010). The ontogenesis of writing in childhood and adolescence. In D. Wyse, R. Andrews and J. Hoffman (Eds.) *The Routledge international handbook of English, language and literacy teaching* (pp. 146-58). Oxford: Routledge.
- Deane, P., & Quinlan, T. (2010). What automated analyses of corpora can tell us about students' writing skills. *Journal of Writing Research* 2, 151-177.
- Díaz-Negrillo, A., & Domínguez, J.F. (2006). Error tagging systems for learner corpora. *Revista española de lingüística aplicada*, (19), 83-102.
- Fayol, M. (2016). From language to text: The development and learning of translation. In C. A. MacArthur, Graham, S., & Fitzgerald, J. (Eds.), *Handbook of writing research* (pp. 130-143). New York, NY: Guilford.
- Gloppe, K. de. (1984). Opstelkenmerken en opstelbeoordelingen. *Tijdschrift voor Taalbeheersing*, 6, 176-189.
- Gloppe, K. de, & Prenger, J. (2013). Schrijfmeters maken. In A. Mottart & S. Vanhooren (Red.), *Zevenentwintigste conferentie Onderwijs Nederlands* (pp. 86-89). Gent: Academia Press.
- Hawkins, J.A., & Buttery, P. (2010). Criterial features in learner corpora: Theory and illustrations. *English Profile Journal*, 1(1), e5.
- Jisa, H., Reilly, J.S., Verhoeven, L., Baruch, E., & Rosado, E. (2002). Passive voice constructions in written texts: A cross-linguistic developmental study. *Written Language & Literacy*, 5(2), 163-181.
- MacArthur, C.A., & Graham, S. (2016). Writing research from a cognitive perspective. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 24-40). New York, NY: Guilford.
- Myhill, D. (2008). Towards a linguistic model of sentence development in writing. *Language and Education*, 22(5), 271-288.
- Nicholls, D. (2003). The Cambridge Learner Corpus: Error coding and analysis for lexicography and ELT. In *Proceedings of the Corpus Linguistics 2003 conference* (Vol. 16, pp. 572-581).
- Pander Maat, H., Raaijmakers, K., Veerbeek, J., & Vermeulen, D. (2017). *Handleiding Utrechts systeem voor annotatie van leerlingstekst*. Interne publicatie Universiteit Utrecht.
- Pollmann, E., Prenger, J., & Gloppe, K. D. (2012). Het beoordelen van leerlingteksten met behulp van een schaalmodel. *Levende Talen Tijdschrift*, 13(3), 15-24.

- Prior, P. (2006). A sociocultural theory of writing. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 54-66). New York, NY: Guilford.
- Rijlaarsdam, G., & Van den Bergh, H. (2006). Writing process theory. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 41-53). New York, NY: Guilford.
- Sanders, T., & Schilperoord, J. (2006). Text structure as a window on the cognition of writing. In C. A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 386-402). New York, NY: Guilford.
- Tolchinsky, L. (2016). From text to language and back: The emergence of written language. In C.A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 144-159). New York, NY: Guilford.
- Torrance, M., & Galbraith, D. (2006). The processing demands of writing. In C.A. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 67-80). New York, NY: Guilford.
- Upton, T.A., & Cohen, M.A. (2009). An approach to corpus-based discourse analysis: The move analysis as example. *Discourse Studies*, 11(5), 585-605.
- VandeKerckhove, R., & Sandra, D. (2015). De potentiële impact van informele online communicatie op de spellingpraktijk van Vlaamse tieners in schoolcontext. *Tijdschrift voor Taalbeheersing* 38(3), 201-234.
- Verheijen, L. (2013). The effects of text messaging and instant messaging on literacy. *English studies*, 94(5), 582-602.
- Verhoeven, L., Aparici, M., Cahana-Amitay, D., Hell, J. V., Kriz, S., & Vigué-Simon, A. (2002). Clause packaging in writing and speech: A cross-linguistic developmental analysis. *Written Language & Literacy*, 5(2), 135-161.

Over de auteurs

Henk Pander Maat is senior-onderzoeker bij de disciplinegroep Taal en Communicatie aan de Universiteit Utrecht. Zijn onderzoek gaat over documentontwerp en stilistiek.

Kay Raaijmakers werkte als onderzoeksassistent aan de Universiteit Utrecht en is nu Interimspecialist Actuarieel bij Talent & Pro.

Dennis Vermeulen studeert na een master Media Studies nu Aardwetenschappen aan de Universiteit Utrecht.

Kees de Gopper is hoogleraar Taalbeheersing van het Nederlands aan de Rijksuniversiteit Groningen. Hij doet onderzoek naar allerlei aspecten van taalvaardigheidsdidactiek.

Bijlage 1 De schrijfpdracht

Typisch Nederlands!

STEL JE VOOR

Volgende maand gaat Annika, een meisje uit Zweden, met haar ouders naar Nederland verhuizen. Annika heeft in Zweden net een taal cursus Nederlands gevolgd en beheerst het Nederlands goed. Annika is nog nooit in Nederland geweest en nu vraagt ze zich af wat er allemaal typisch Nederlands is. Daarom besluit je haar een brief te sturen. Je wilt niet over tulpen en molens schrijven, maar over andere typisch Nederlandse zaken.

WAT MOET JE DOEN?

Schrijf een brief aan Annika waarin je typisch Nederlandse dingen beschrijft.

AANWIJZINGEN

Beschrijf in je brief in ieder geval 4 typisch Nederlandse dingen (die bijvoorbeeld te maken hebben met: sport, lekkernijen, het landschap, ...).
 Weet je nog meer typisch Nederlandse dingen die Annika zou moeten weten voor ze in Nederland komt wonen? Schrijf daar dan ook over.
 Beschrijf alles zo duidelijk dat Annika zich goed kan voorstellen waar jij haar over vertelt. Zorg ervoor dat je tekst eruit ziet als een brief.

Bijlage 2 Voorbeeld van geannoteerde tekst

Tekst met annotatie	Toelichting op afkortingen
[<plaats>, 26 november 2012]PD	PD: plaats en datum
[Beste Annika,]AH	AH: aanhef
[Gisteren is aan mij verteld dat ik [aan/Ø]W*W*VZ jou een brief moest schrijven over typisch Nederlandse dingen[./:] L*! leuk!]AL [In deze brief zal ik dus proberen wat over Nederland te vertellen.]AK [Hopelijk zul je alles begrijpen!]IL	W*W*VZ: eenheid=Woord, probleem=Weglaten, woordsoort=Voorzetsel L*!: eenheid=Leesteken, probleem=Incorrect AL: aanleiding voor de tekst AK: aankondiging van tekstinhoud IL: blij van inleven in de lezer
Nederland is een plat land, zonder reliëf. [De wind waait er het hele jaar door en met regen zijn wij Nederlanders zeer vertrouwd.]VMT*O Dat het Nederlandse weer niet vaak aangenaam is, is een minpuntje, [maar]VMT*I je zult in niet veel landen zulke gezellige en knusse markten vinden zoals wij die hebben! Zeker rond de Kerst is het daar erg aangenaam.	VMT*O: eenheid=Verbale TopicMarkering; probleem=Ontbreekt VMT*I: eenheid=Verbale TopicMarkering; probleem=Incorrect

<p>[In ons land hebben we, [vast ook zoals in jouw land]]IL, bepaalde lekkernijen, zoals stroopwafels, pannenkoeken en haring!]VMT*O Bijna elke Nederlander [vind/vindt]SPE*I*W dit wel lekker. Bovendien eten wij [bij/met]W*I*VZ Sinterklaas (een feest in december) pepernoten. Dat zijn een soort kleine koekjes waaraan je je, als je niet oppast, helemaal [miselijk/misselijk]SPE*I*A kunt eten!</p>	<p>IL: blijk van inleven in de lezer VMT*O: eenheid=Verbale TopicMarkering; probleem=Ontbreekt SPE*I*W: eenheid=Spelling, probleem=Incorrect, woordsoort=Werkwoord W*I*VZ: eenheid=Woord, probleem=Incorrect, woordsoort=Voorzetsel SPE*I*A: eenheid=Spelling,</p>
<p>[Naast veel eten]VMT kunnen [Nederlandes/ Nederlanders]SPE*I*N ook goed sporten en er met name goed naar kijken. Nu heb ik het over voetbal, dé sport in Nederland. Als je met een Nederlander een gesprekje wilt houden, kun je het daar altijd over hebben!</p>	<p>probleem=Incorrect, woordsoort=Adjectief VMT=eenheid=Verbale TopicMarkering SPE*I*N: eenheid=Spelling, probleem=Incorrect, woordsoort=zelfstandig naamwoord</p>
<p>[[Ik hoop dat je alles in mijn brief hebt begrepen]]IL en ik wens jou en je ouders een [hele/heel]FA*I*A goede tijd in Nederland toe!] AFZ</p>	<p>IL: blijk van inleven in de lezer FA*I*A: eenheid is Flexie/Affix; probleem=Incorrect; woordsoort= Adjectief AFZ: afsluitzin</p>
<p>[Hartelijke groeten, <voornaam>]G</p>	<p>G: groet</p>

