

Taalproductie en -verwerking onderzoeken met de computer

Ter inleiding

ICT-toepassingen stellen taalonderzoekers nu in staat om lees- en schrijfprocessen, spreek- en luisterprocessen te registreren op een niveau van detail dat we hiervoor niet konden realiseren. Maar dat veronderstelt dat we wel goed weten hoe die nieuwe registratiemogelijkheden precies werken en wat er eigenlijk wordt geregistreerd. Welk soort onderzoeksvragen kunnen en willen we beantwoorden nu we deze nieuwe mogelijkheden tot onze beschikking hebben? Wat zijn de methodologische vragen en overwegingen die opgeroepen worden door deze nieuwe registratie- en analyse-instrumenten?

Wie processen van taalproductie en -verwerking wil onderzoeken, moet feitelijk drie stappen zetten.

1. Oproepen van het te bestuderen taalgedrag (*elicitatione*).

De lezer moet aanzet worden om iets te lezen, de schrijver moet opdracht krijgen om iets te schrijven, de luisteraar krijgt iets te horen, de spreker reageert op een of andere stimulus en zegt iets (terug). Het te bestuderen gedrag moet dus in gang gezet worden. Dat kan gebeuren in een zo natuurlijk mogelijke omgeving en context, waarbij de proefpersoon zich er misschien niet eens van bewust is dat zijn taalgedrag onderwerp van onderzoek is. Maar het kan ook plaatsvinden in een volkomen onnatuurlijke omgeving, waarbij de proefpersoon wordt uitgedost met een badmuts vol elektrodes of in de tunnel van een MRI-scanner ligt. Wat

Samenvatting

Dit themanummer richt zich op nieuwe toepassingen van ICT in de praktijk van het taalbeheersingsonderzoek. ICT biedt allerlei mogelijkheden om processen van taalproductie en -verwerking gedetailleerd te registreren, te visualiseren en te analyseren. Dit scala van mogelijkheden wordt hier getoond en met voorbeelden geïllustreerd. De artikelen beschrijven de volgende onderzoekstechnieken: hardopdenken, toetsregistratie, oogbewegingsregistratie en neuroimaging (fMRI en ERP). De auteurs beschrijven telkens een bepaalde methode en bespreken vervolgens de belangrijkste toepassingsmogelijkheden en beperkingen. Extra aandacht gaat daarbij uit naar methodologische aspecten van reactiviteit en triangulatie. De basis voor dit themanummer werd gelegd tijdens het VIOT-congres 2005 in Nijmegen waar we een symposium organiseerden rond 'Methodes en instrumenten voor observatie-onderzoek in schrijf- en leesonderzoek in digitale omgevingen'.

is de invloed van de elicitatie en de context waarin die plaatsvindt op de taalprocessen die we willen onderzoeken?

2. Opnemen en vastleggen van het taalgedrag (*registratie*).

Het proces dat onderwerp van onderzoek is, wordt op een of andere wijze vastgelegd, vaak door of met behulp van de computer. Soms is die registratie nauwelijks te beïnvloeden door de proefpersoon, bijvoorbeeld als de hersenactiviteit wordt geregistreerd die samen gaat met het luisteren naar taaluitingen. In andere gevallen is de invloed van de taalgebruikers op wat er geregistreerd wordt behoorlijk groot, bijvoorbeeld als ze hardopdenkend een tekst schrijven of een website gebruiken. Wat willen we registreren en hoe verhoudt de geregistreeerde werkelijkheid zich met de 'echte' processen die we willen onderzoeken? Hoe weten we of onze data een valide en betrouwbaar beeld geven van het proces dat we proberen te betrappen?

3. Analyseren en interpreteren van het geregistreeerde gedrag (*analyse*).

De onderzoeker zal proberen vast te stellen wat er nu precies gebeurt of gebeurd is tijdens de taalproductie en -verwerking. De registratie maakt het mogelijk het proces te herhalen, terug te spelen en eventueel zelfs de proefpersoon zelf er weer op te laten reageren. De 'real-time' observaties van de onderzoeker kunnen worden aangevuld als het geregistreeerde gedrag in alle rust en in alle detail kan worden bestudeerd. De codering en interpretatie zijn makkelijker door meerdere onafhankelijke beoordelaars uit te voeren. De ICT-toepassing zelf kan ook voorbewerkingen uitvoeren en zo een deel van het analysewerk vergemakkelijken. Kwantitatieve gegevens, zoals leestijden of cursorposities, kunnen gepresenteerd worden op een manier die (statistische) analyse vergemakkelijkt. Maar wat zijn de vragen waarop de analyse een antwoord moet geven en bieden onze instrumenten ons wel de gegevens die we nodig hebben om een antwoord te vinden? Elke vorm van analyse gaat onvermijdelijk gepaard met verlies van data aan de ene kant en toespitsing op wat de ontwerpers van het instrument belangrijke data vinden aan de andere kant. Wat kunnen we eigenlijk te weten komen met deze instrumenten en is dat ook wat we willen weten?

De artikelen in dit themanummer, de neerslag van een themasessie op het VIOT-congres 2005 aan de Radboud Universiteit Nijmegen, zijn gerangschikt op twee dimensies. We beginnen met wat wij veronderstellen dat de meest bekende onderzoeksmethode is: hardopdenkend gebruik maken van een website, geregistreerd met de computer. De reeks artikelen wordt gesloten met de minst bekende methode van registratie en analyse, neuroimaging. Misschien niet toevallig is er in de achtereenvolgende artikelen ook sprake van afnemende beïnvloeding. De mate waarin de geregistreeerde gegevens bewust beïnvloed kunnen worden door de proefpersoon zelf of door de elicitatie is bij hardopdenken veel groter dan bij oogbewegingregistratie of neuroimaging. Tegelijkertijd is, paradoxaal genoeg, bij die laatste methode de situatie en omgeving van de proefpersoon het minst natuurlijk en daardoor ook het meest nadrukkelijk aanwezig: de proefpersonen bevinden zich in een scanner tijdens het onderzoek naar hun taalgedrag. De mate van reactiviteit en de invloed daarvan op het te onderzoeken taalgedrag is eigenlijk alleen goed vast te stellen als we verschillende onderzoeksmethoden naast elkaar gebruiken om het proces te registreren en te analyseren (triangulatie). De mogelijkheden om bijvoorbeeld de reactiviteit van hardoplezen en commentaar geven te onderzoeken met behulp van oogbewegingregistratie wordt in het artikel van Lentz en zijn mede-auteurs gedemonstreerd.

De eerste bijdrage in dit themanummer richt zich op wat wij de laatste decennia als de ‘klasieke’ registratiewijze van lees- en schrijfprocessen zijn gaan beschouwen: de hardopdenkmethode. Maaïke van den Haak, Menno de Jong en Peter Jan Schellens verkennen in hun artikel *Hardopdenkprotocollen en gebruikersonderzoek: Volledigheid en reactiviteit van de synchrone hardopdenkmethode* de grenzen van die methode. De methode wordt in het door hen gerapporteerde onderzoek in de eerste plaats gebruikt als een manier om problemen op te sporen die gebruikers van een website ondervinden bij het uitvoeren van taken. De gebruikers worden aan het werk gezet met taken waarvoor zij tekst op de webpagina moeten lezen, verwerken en omzetten in acties (elicitatie). Die acties worden in dit geval geregistreerd met video, maar dat zou ook met een log-programma zoals Camtasia uitgevoerd kunnen worden. De registratie maakt het mogelijk het proces terug te spelen en de gebruiker om post-hoc-commentaar te vragen, in aanvulling op het hardopdenken tijdens de taakuitvoering.

In het artikel wordt nadrukkelijk het probleem van reactiviteit aan de orde gesteld. Wat is de invloed van de proefleider en diens interactie met de proefpersoon? Wat is de invloed van de specifieke website die gebruikt wordt om het taalgedrag op te roepen? Wat is de invloed van het verbaliseren zelf, als dit tijdens het feitelijk gebruik van de website plaatsvindt? Leidt de registratiemethode zelf, het hardopdenken, hier niet tot vermeende gebruiksproblemen, die dus artefacten van de toegepaste registratiemethode zijn? De vragen die Van den Haak en haar collega's aansnijden voor de hardopdenkmethode gelden in feite voor elke registratiemethode, dus ook voor de methoden die in de volgende artikelen in dit themanummer worden gepresenteerd.

Luuk van Waes en Mariëlle Leijten presenteren een tweede registratie-instrument, speciaal bedoeld om schrijfprocessen vast te leggen en te analyseren. Ze beschrijven het programma *Inputlog*, waarmee het gedrag wordt vastgelegd van schrijvers die gebruikmaken van de computer als schrijf- of dicteerinstrument. Elke toetsaanslag of elk ingesproken woord wordt geregistreerd en het geheel wordt voorbereid voor analyse. Ook hier kan het geregistreerde schrijfproces terugspeeld worden, zodat de schrijver in tweede instantie commentaar kan leveren op zijn eigen gedrag. De auteurs illustreren het gebruik van Inputlog met een studie van revisiegedrag, maar natuurlijk kan het programma ook dienen om de protocollen van hardopdenkende schrijvers of dicteerders te voorzien van een zeer gedetailleerde procesregistratie. Aan het eind van hun artikel doen Van Waes en Leijten de suggestie om een programma als Inputlog te combineren met oogbewegingsregistratie. Niet alleen kunnen we dan nog veel preciezer vaststellen hoe schrijvers de geproduceerde tekst nu eigenlijk herlezen tijdens het reviseren, het is tegelijkertijd een vorm van triangulatie. Daarbij kunnen we bijvoorbeeld de pauzetijden analyseren die Van Waes en Leijten meten tijdens de foutdetectie van de reviserende schrijver: kijkt de reviseerder daarin ook daadwerkelijk naar de fout, of is hij in feite ook nog met andere delen van de tekst bezig?

In het derde artikel richt Reinier Cozijn zich specifiek op oogbewegingsregistratie. Verschillende Nederlandse universiteiten hebben recent dergelijke apparatuur aangeschaft. Cozijns artikel bevat een duidelijke uitleg over wat oogbewegingsregistratie nu eigenlijk inhoudt. De oogbewegingen worden gebruikt als indicatie van (cognitieve) processen van taalverwerking en -productie. Elicitatie vindt plaats door experimenteel materiaal, bijvoorbeeld teksten, plaatjes of websites, via een beeldscherm aan te bieden. De nieuwste generaties apparatuur maken het mogelijk het bewegingspatroon en de blikrichting terug te tonen op het computerscherm, over het aangeboden experimenteel materiaal heen. Cozijn illustreert

de mogelijkheden van oogbewegingregistratie met een case study waarin hij het effect van tekststopmaak laat zien op leestijden, blikrichting en fixaties.

Ook in het vierde artikel staat oogbewegingregistratie als instrument centraal. Leo Lentz, Pim Mak en Henk Pander Maat laten zien hoe het ingezet kan worden voor gebruikers-onderzoek. Het onderzoek dat zij rapporteren ligt in het verlengde van wat Van den Haak en haar collega's rapporteerden (eerste artikel). Ook in hun geval voeren lezers hardopwerkend taken uit, in dit geval leestaken. Met behulp van de oogbewegingregistratie verkennen Lentz en zijn co-auteurs in hoeverre de taakuitvoering wordt beïnvloed door de opdracht hardop te lezen of naderhand commentaar te geven op de tekst. Oogbewegingen laten bijvoorbeeld zien of de proefpersonen die commentaar moesten geven, vaker teruggaan naar de gelezen tekst. Het artikel illustreert hoe een instrument als oogbewegingregistratie ingezet kan worden om invloeden van de elicitatie (in dit geval de opdracht om commentaar te geven) op het taalverwerkingsproces te observeren en te analyseren.

John Hoeks, Petra Hendriks en Gisela Redeker beschrijven in het laatste artikel, *Communicatie en het brein: Het gebruik van neuroimaging bij onderzoek naar dialogen*, de meest gebruikte neuroimaging methoden (PET, fMRI en ERP's), technieken die velen wellicht alleen kennen in de medische context. MRI-scans en andere scantechnieken worden ingezet om een beeld van hersenactiviteit te geven bij bepaalde taken. In het artikel van Hoeks en zijn collega's betreffen het talige taken. De techniek maakt het mogelijk om in te zoomen op de hersengebieden die actief zijn tijdens taalproductie of -verwerking. De mogelijkheden van deze techniek worden kritisch besproken en de ERP-methode wordt daarna geïllustreerd aan de hand van een experiment. De methode levert een visualisatie op van de verwerkingsactiviteit in de hersenen van de taalgebruiker: er wordt meer bepaald gekeken naar wat er gebeurt in de hersenen van een proefpersoon als die geconfronteerd wordt met taaluitingen die niet voldoen aan één van de maxims van Grice. Hoeks en zijn collega's maken in de discussieparagraaf van hun artikel duidelijk dat we nog maar aan het begin staan van het taalonderzoek met behulp van deze neuroimaging technieken. Tegelijkertijd stellen zij dat voor de interpretatie van de fijnmazige gegevens die hiermee verzameld worden, aanvulling met bijvoorbeeld oogbewegingregistratie onontbeerlijk is. Hiermee wordt de reeks artikelen in dit themanummer afgesloten, met een pleidooi voor de combinatie van onderzoeksinstrumenten en -methoden. Wij hopen dat het themanummer u op nieuwe ideeën brengt over de verschillende manieren waarop onderzoek naar taalproductie en -verwerking kan plaatsvinden en over de rol die ICT daarin kan spelen.

Hardopdenkprotocollen en gebruikersonderzoek

Volledigheid en reactiviteit van de synchrone hardopdenkmethode

1. Inleiding

De hardopdenkmethode is een dominante benadering geworden voor het registreren van gebruiksprocessen van communicatiemiddelen, met name websites en instructieve documenten. Het doel van een dergelijke registratie is vaak het pretesten van de communicatiemiddelen: het opsporen van mogelijke gebruikersproblemen en het op basis daarvan formuleren van revisiesuggesties. In grote lijnen komt de methode erop neer dat proefpersonen aan bepaalde taken worden gezet, met als extra opdracht om tijdens de taakuitvoering voortdurend hun gedachten te verbaliseren. De hardopdenkprotocollen, die bestaan uit een weergave van zowel de acties als de verbalisaties van proefpersonen, leveren gedetailleerde procesgegevens op over de wijze waarop gebruikers te werk gaan met het betreffende communicatiemiddel. Deze procesgegevens zorgen ervoor dat de onderzoeker precies te weten komt waar zich problemen voordoen, zodat deze vervolgens in een revisieronde verholpen kunnen worden.

Al sinds de vroege jaren '80 is de methode in de Verenigde Staten gebruikt om allerlei documenten te evalueren. Aanvankelijk kregen proefpersonen slechts de opdracht om een tekst hardop te lezen en hun gedachten daarbij te verwoorden (Flower, Hayes & Swarts, 1983; Swaney e.a., 1981). De

Samenvatting

De hardopdenkmethode heeft zich ontwikkeld tot een gangbare onderzoeksmethode voor uiteenlopend lees- en schrijfonderzoek. In dit artikel staat de validiteit van de hardopdenkmethode als pretestinstrument centraal. Discussies over de validiteit van hardopdenkprotocollen betreffen de reactiviteit van de methode en de volledigheid van de verbalisaties. De vraag wordt beantwoord in hoeverre de aard van het testobject van invloed is op de reactiviteit van de methode en de volledigheid van de protocollen. Daartoe zijn synchrone en retrospectieve hardopdenkprotocollen vergeleken voor twee typen internet-applicaties: een online bibliotheekcatalogus en een gemeentelijke website. In beide gevallen blijken synchrone hardopdenkprotocollen minder geverbaliseerde problemen te bevatten dan retrospectieve protocollen. De reactiviteit van de methode blijkt samen te hangen met het testobject: in de bibliotheekcatalogus leidde de opdracht om synchroon hardop te denken tot meer observeerbare fouten in de taakuitvoering, op de gemeentelijke website niet. Deze verschillen kunnen worden herleid tot verschillen in taken tussen de beide testobjecten.

begrijpelijkheid van de tekst stond in dergelijk onderzoek centraal. Daarna is er in toenemende mate gewerkt met specifieke taken die proefpersonen aan de hand van een instructieve tekst moesten verrichten (Schraver, 1987; Jansen & Steehouder, 1989). In die gevallen gaat het, naast begrijpelijkheid, ook om de vindbaarheid en de praktische toepassing van informatie. Dieli (1986) beschrijft en vergelijkt hardopdenkonderzoek mét (*user protocols*) en zonder (*reader protocols*) specifieke taken en concludeert dat beide vormen van hardopdenkonderzoek inderdaad verschillende soorten resultaten opleveren. Meer recent vestigen Lentz & Pander Maat (2003) opnieuw de aandacht op de mogelijke waarde van hardopdenkprotocollen zonder taken. Zij beschrijven enkele studies waarin de hardopdenkmethode zonder taken wordt vergeleken met twee andere pretestmethoden (de plus-en-minmethode en het computerprogramma Focus) en concluderen dat lezersprotocollen in ieder geval geschikter zijn om begripsproblemen op te sporen dan beide andere pretestmethoden. Over de hele linie heerst er grote tevredenheid over het gebruik van hardopdenkonderzoek voor het pretesten van communicatiemiddelen. Aangenomen wordt dat de geconstateerde problemen goede voorspellers zijn van de problemen die echte gebruikers in het dagelijks leven zullen gaan ondervinden.

De opkomst van het internet heeft de hardopdenkmethode als pretestinstrument alleen maar populairder gemaakt. Teksten die via het internet worden aangeboden, krijgen er bijna per definitie een zware selectiedimensie bij (het zoeken en navigeren op de website), die de aandacht voor veel andere soorten gebruikersproblemen lijkt te verdringen. Op natuurlijke wijze komen bij de evaluatie van websites twee afzonderlijke onderzoekstradities bij elkaar: die van tekstevaluatie (De Jong & Schellens, 1995) en usability testing (Nielsen, 1993; Dumas & Redish, 1999). Binnen de traditie van de usability testing golden hardopdenkprotocollen al van oudsher als de dominante evaluatiemethode. Een en ander heeft overigens wel geleid tot een onderbelichting van de tekstuele en visuele *content* van websites en een overmatige aandacht voor navigatieproblemen.

In de context van usability testing heeft de validiteit van hardopdenkprotocollen nauwelijks ter discussie gestaan. Integendeel, de methode wordt doorgaans als onbetwistbaar criterium gebruikt om andere evaluatiemethoden mee te vergelijken. Daarbij wordt sterk geïmpeerd op het standaardwerk van Ericsson & Simon (1993), waarin op basis van een groot aantal empirische studies wordt geconcludeerd dat hardopdenkgegevens, onder een aantal stringente voorwaarden, in principe valide zijn. Zo mag de taak zelf niet te moeilijk zijn voor de proefpersoon (anders zal deze stoppen met hardop denken), maar mag het anderzijds ook niet gaan om een min of meer geautomatiseerde taak. Ook wordt er onderscheid gemaakt tussen drie niveaus van verbaliseren. De eerste twee niveaus (het verklanken van informatie die al in dezelfde vorm in het korte termijn geheugen aanwezig is en gedachten die als enige transformatie vertaald moeten worden in verbale informatie) geven volgens Ericsson & Simon geen problemen. Alleen wanneer proefpersonen meerdere stappen moeten zetten om van hun gedachten naar verbalisaties te komen (dit wordt het derde niveau genoemd), levert hardopdenken volgens Ericsson & Simon geen valide gegevens meer op. Hierbij is te denken aan het filteren van de verbalisaties, het geven van verklaringen voor het eigen gedrag of het binnenhalen van kennis uit het lange termijn geheugen. Daarnaast geven Ericsson & Simon strenge gedragsregels voor de proefleider, die de proefpersoon er zonnodig wel aan moet herinneren hardop te denken ("keep talking"), maar op geen enkele andere manier mag intervensiëren in het proces.

De voornaamste discussie binnen het domein van de usability testing lijkt te gaan over de vraag in hoeverre het theoretisch raamwerk van Ericsson & Simon eigenlijk van toepassing is op de praktijk van usability testing. Wright & Monk (1991) deden een studie waarin ze een strikte uitvoering van de hardopdenkmethode (exact volgens de regels van Ericsson & Simon opgezet) probeerden te vergelijken met een vrijere variant. De vergelijking viel in het water doordat geen van hun proefleiders in de strikte conditie zich voldoende aan de richtlijnen hield. Tot vergelijkbare conclusies kwamen Boren & Ramey (2000) in een veldonderzoek naar de werkwijze van usability professionals: de gedragsregels van Ericsson & Simon worden niet nageleefd in de praktijk van usability testing. Boren & Ramey betogen dat dit niet per se slecht hoeft te zijn, omdat de doelstellingen van een usability test sterk afwijken van die van veel ander hardopdenkonderzoek. Niet het hele proces, maar de problemen die zich voordoen, zijn het centrale aandachtspunt. Ook is er doorgaans het nodige te observeren, waardoor de afhankelijkheid van de verbalisaties minder groot is. Als een alternatief voor de traditionele benadering stellen Boren & Ramey een “speech communication”-benadering voor, waarin de proefleider in sommige (beregelde) gevallen wel mag ingrijpen in het proces en ook meer interageert met de proefpersoon.

Buiten de context van usability testing is er doorlopend aandacht geweest voor de validiteit van hardopdenkprotocollen. Diverse studies hebben aangetoond dat het gebruik van de hardopdenkmethode in bepaalde contexten kan leiden tot vertekeningen. Proefpersonen kunnen hun taken beter of juist slechter uitvoeren dankzij de opdracht hardop te denken. Proefpersonen die slechter presteren, hebben te lijden onder een te zware cognitieve belasting. In een eerdere studie zagen we dat dergelijke effecten ook in een usability test kunnen optreden: bij de evaluatie van een online bibliotheekcatalogus stelden we vast dat hardopdenkende proefpersonen meer observeerbare fouten maakten en minder taken correct uitvoerden dan stilwerkende proefpersonen (Van den Haak, De Jong & Schellens, 2003). Ericsson & Simon (1993) veronderstellen dat proefpersonen minder gaan verbaliseren wanneer hun taak te moeilijk wordt. Dat kan zo zijn, maar er blijkt in de praktijk soms ook een negatief effect te zijn op de kwaliteit van de taakuitvoering. Proefpersonen die beter presteren, profiteren van een mogelijk faciliterende bijdrage van de hardopdenkopdracht: het feit dat ze gedwongen worden hardop te denken, zorgt ervoor dat ze zorgvuldiger of anderszins beter te werk gaan dan normaal. Dit lijkt bijvoorbeeld het geval in leesprocessen (Silvén & Vauras, 1992; Loxterman, Beck & McKeown, 1994; Kucan & Beck, 1997). Ook de manier waarop proefpersonen te werk gaan, kan door het hardopdenken worden beïnvloed. Janssen, Van Waes & Van den Bergh (1996) lieten bijvoorbeeld zien dat de opdracht om hardop te denken het schrijfproces (geoperationaliseerd als het pauzegedrag) van proefpersonen beïnvloedde, met name als het ging om een complexe schrijftaak. Mumma, Draguns & Seibel (1993) stelden vast dat hardopdenkende proefpersonen anders (efficiënter) te werk gingen bij het toekennen van persoonlijkheidskenmerken aan mensen dan proefpersonen die stil werkten. Knoblich & Rhenius (1995) vonden in hun onderzoek dat de opdracht om hardop te denken invloed had op de manier waarop proefpersonen een koelinstallatie bedienden.

Dergelijke studies roepen vragen op over de validiteit van de hardopdenkmethode. Deze lijkt in principe contextafhankelijk te zijn: in het ene geval leidt hardopdenken tot een betere taakuitvoering, in het andere geval tot een minder effectieve taakuitvoering en in verschillende situaties blijken proefpersonen anders te werk te gaan. We kunnen op grond van zulk onderzoek natuurlijk niet direct concluderen dat de validiteit van hardopdenkprotocollen als pretestmethode dus ook problematisch zal zijn, maar het onderzoek vestigt wel de aandacht

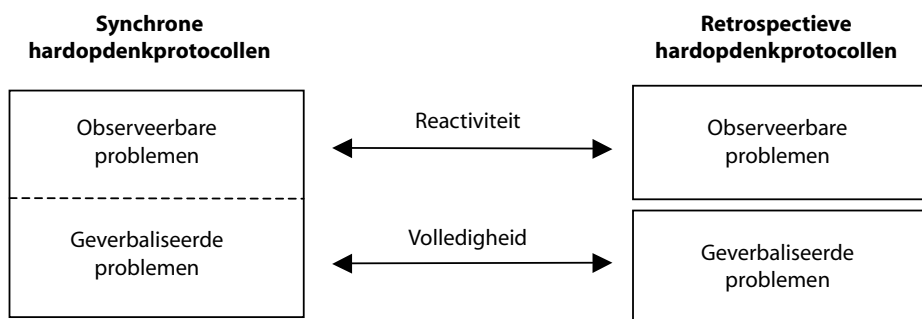
op de mogelijkheid dat er validiteitsproblemen optreden. Voor een usability test zouden alle hierboven beschreven afwijkingen een bedreiging vormen. Blijkbaar kunnen de resultaten van een hardopdenkttest een (positief dan wel negatief) vertekend totaalbeeld geven van de gebruiksvriendelijkheid van een communicatiemiddel. Het percentage correct uitgevoerde opdrachten hoeft niet te corresponderen met het succespercentage onder stilwerkende gebruikers. Ook kan het hardopdenken problemen in een communicatiemiddel veroorzaken en andere problemen verhullen. Usability professionals hoeven hier natuurlijk niets van te merken: de uitkomsten van een usability test zullen er in hun perceptie niet minder overtuigend om zijn.

Het al dan niet optreden van vertekeningen doordat proefpersonen hun gedachten moeten verbaliseren, lijkt voor een belangrijk deel samen te hangen met de (deel)taken die zij moeten uitvoeren. Bij elk empirisch tegenbewijs tegen de validiteit van de hardopdenkmethode verwijzen Ericsson & Simon (1993) ofwel naar onregelmatigheden in proefleidergedrag of hardopdenkinstructies ofwel naar onderzoeksdoelstellingen die de type 1- en type 2-verbalisaties te boven gaan. Algemene uitspraken over de validiteit van de hardopdenkmethode liggen onzes inziens niet voor de hand. In plaats daarvan zou een uitvoerige verkenning van de relatie tussen taaksoorten en de validiteit van hardopdenkprotocollen een essentiële stap zijn in de verdere ontwikkeling van de methode. De typologie van verbalisaties zoals Ericsson & Simon (1993) die gebruikten, biedt in dit opzicht nog te weinig houvast.

Ook binnen de context van usability testing lijkt een bezinning op de geschiktheid van de methode op allerhande deeltaken op zijn plaats. In de literatuur over human-computer interaction wordt de suggestie gewekt dat het testen van software en het testen van een website feitelijk op hetzelfde neerkomt, ondanks de grote verschillen tussen een regulier softwarepakket en een website. Hoewel hardopdenkprotocollen in principe op dezelfde manier kunnen worden gebruikt voor beide testobjecten, is het goed denkbaar dat het anders gesteld is met de validiteit van de methode. Ook in het onderzoek naar tekstevaluatie is voorsnog weinig rekening gehouden met de invloed van het testobject. In de overzichten die tot nu toe gemaakt zijn over het methodologische onderzoek naar tekstevaluatie (De Jong, 1998; De Jong & Schellens, 1995; 2000), zijn de inzichten uit de human-computer interaction onverkort meegenomen. Het is wederom de vraag of dat gerechtvaardigd is.

In dit artikel bespreken we twee studies waarin de resultaten van synchrone en retrospectieve hardopdenkprotocollen met elkaar vergeleken worden. Bij synchrone hardopdenkprotocollen moeten proefpersonen direct tijdens de taakuitvoering hun gedachten verbaliseren; bij retrospectieve hardopdenkprotocollen verrichten proefpersonen eerst in stilte de taken en verbaliseren ze hun gedachten achteraf, terwijl ze naar een video-opname van hun taakuitvoering kijken. De twee methoden werden al door Nielsen (1993) als concurrenten beschreven. Recent onderzoek waarin de verbalisaties van retrospectief hardopdenkende proefpersonen werden afgezet tegen de oogbewegingen tijdens de taakuitvoering, geeft nadere ondersteuning voor deze, wat bewerkelijker, vorm van hardopdenkonderzoek: de verbalisaties achteraf bleken goed overeen te komen met de oogbewegingen die de proefpersonen tijdens de taakuitvoering hadden gemaakt (Guan e.a., 2006). Een vergelijking van beide vormen van hardopdenken levert een beeld op van de reactiviteit van de synchrone hardopdenkmethode en de volledigheid van de verbalisaties van proefpersonen tijdens de taakuitvoering (zie Figuur 1). Bij synchrone hardopdenkprotocollen worden observeerbare en geverbaliseerde problemen tegelijkertijd verzameld; bij retrospectieve hardopdenkprotocollen zijn de twee soorten problemen verdeeld over twee ronden: de observeerbare problemen tijdens het stil

werken, de geverbaliseerde problemen tijdens het retrospectief hardopdenken. Een vergelijking van de observeerbare problemen betreft de reactiviteit van synchroon hardopdenken: de taakuitvoering wordt in de retrospectieve conditie immers in het geheel niet beïnvloed door een extra hardopdenktaak voor de proefpersonen. Een vergelijking van de geverbaliseerde problemen heeft betrekking op de volledigheid van synchrone hardopdenkprotocollen: de verbalisaties van de proefpersonen in de retrospectieve conditie staan immers niet onder druk van de belasting die de taakuitvoering op zich al met zich meebrengt.



Figuur 1: Schematisch overzicht van het onderzoek

Het gaat in dit onderzoek niet alleen om een vergelijking van de twee methoden, maar ook om een vergelijking van testobjecten. We kijken of het onderscheid tussen synchrone en retrospectieve hardopdenkprotocollen op dezelfde manier werkt voor een online bibliotheekcatalogus en een gemeentelijke website. De eerste studie betrof de evaluatie van de bibliotheekcatalogus van de Universiteit Utrecht (Van den Haak, De Jong & Schellens, 2004) en was een uitgebreide replicatie van een onderzoek dat we eerder in het *Tijdschrift voor Taalbeheersing* beschreven (Van den Haak, De Jong & Schellens, 2003). De tweede studie had betrekking op de website van de gemeente Haarlem (Van den Haak, De Jong & Schellens, 2007). In beide studies vergeleken we de twee typen hardopdenkprotocollen overigens ook nog met een derde variant, namelijk *constructive interaction* (waarbij proefpersonen in tweetallen aan de taken moesten werken), maar die laten we in dit artikel buiten beschouwing. Voordat we de opzet en de resultaten van ons onderzoek beschrijven, zullen we in de volgende paragraaf eerst ingaan op de kenmerkende verschillen die optreden tijdens het werken met de twee testobjecten. De verschillen worden gerelateerd aan verwachtingen ten aanzien van de reactiviteit van de hardopdenkmethode en de volledigheid van de verbalisaties in de protocollen.

2. Gebruikerstaken in de online catalogus en op de website

In de bibliotheekcatalogus en op de website gaven we de proefpersonen taken die kenmerkend waren voor het betreffende communicatiemiddel. In de online bibliotheekcatalogus moesten de proefpersonen in vijf opdrachten met diverse zoektechnieken werken om allerlei publicaties te vinden. Ze moesten zoeken op auteur en op onderwerp. Ze moesten combinaties van zoektermen gebruiken, selecteren op taal en op jaar van publicatie en zoek-

termen trunceren. In de gemeentelijke website moesten de proefpersonen in vijf scenario's in totaal twaalf deeltaken verrichten, die betrekking hadden op het kopen van een huis, het aanbieden van afval, het inschrijven bij de gemeente als inwoner, het betalen van gemeentelijke belastingen en het verkrijgen van een parkeervergunning. Bij een vergelijking van de taakuitvoering in beide studies komen we tot enkele kenmerkende verschillen.

Het werken met de online bibliotheekcatalogus kan getypeerd worden als complex maar overzichtelijk. Literatuur zoeken in een bibliotheekcatalogus is voor studenten een erkend lastige taak. In ons geval moesten de proefpersonen ook nog zoeken in een catalogus waar ze niet aan gewend waren (de proefpersonen studeerden aan de Universiteit Twente; de catalogus was van de Universiteit Utrecht). De complexiteit van de taak had vooral te maken met twee aspecten: (a) een gebrek aan voorkennis over informatiezoeken, en (b) problemen met de toepassing van de eigen voorkennis op een nieuw systeem. In beide gevallen moesten zij zich een mentaal model proberen te vormen van de opzet van de catalogus. Hoewel de vijf taken geheel afzonderlijk van elkaar konden worden verricht, is het goed denkbaar dat er bepaalde leereffecten zouden optreden. Dat laatste had te maken met de overzichtelijkheid: het aantal knoppen waaruit de proefpersonen moesten kiezen, was in wezen beperkt.

Het werken met de website van de gemeente Haarlem was in principe minder complex. Het ging om minder specialistische taken, die in principe aansloten op de dagelijkse leefwereld van de proefpersonen. Maar de taken waren aan de andere kant ook minder overzichtelijk. In plaats van één interface waarmee alle taken verricht moesten worden, moesten de proefpersonen navigeren over de hele site. Een belangrijke taak was steeds uit te vinden waar de gewenste informatie “verstopt” zat. Daarbij kregen de proefpersonen te maken met diverse menu's: op één webpagina stonden soms horizontaal en verticaal verschillende menu's waaruit gekozen moest worden en als er een keuze was gemaakt kwam op het lagere niveau vaak weer een ander menu tevoorschijn. Leereffecten waren in deze context zo goed als onmogelijk. Elk scenario stelde andere eisen aan de proefpersonen.

Het grootste verschil tussen beide testobjecten betrof het belang van leesactiviteiten in de taakuitvoering. Van lezen, browsen of skimmen was in de online catalogus nauwelijks sprake. Alleen als de proefpersonen ervoor kozen de online help te raadplegen, moesten ze een hoeveelheid tekst verwerken. Op de gemeentelijke website, daarentegen, waren lees-, browse- en skimactiviteiten juist erg dominant. Dat gold in de eerste plaats voor de *content* van de website, die uiteindelijk gelezen moest worden. Maar dergelijke activiteiten waren evenzeer van belang voor alle menu's, waarvan de linknamen meer waren dan lexicaal aanduidingen van in principe eenduidige begrippen. Linknamen als “Leven & wonen” en “Komen & gaan” lijken op kopjes in een reguliere tekst, waarbij er een voortdurende spanning is tussen informativiteit, inhoudelijke dekkendheid, bondigheid en aantrekkelijkheid.

Op basis van de complexiteit van de online bibliotheekcatalogus was het te verwachten dat hier gemakkelijk problemen met cognitieve belasting zouden optreden. Proefpersonen moeten graven in hun herinnering naar de functies die normaal gesproken beschikbaar zijn in een bibliotheekcatalogus en moeten deze functies relateren aan de knoppen op de interface. Het is goed denkbaar dat er bij een dergelijk testobject problemen optreden als de proefpersonen ook nog hardop moeten denken. Dat bleek in ieder geval in een vorige studie, waarin een andere online bibliotheekcatalogus centraal stond (Van den Haak, De Jong & Schellens, 2003). Het is ook voorstelbaar dat leereffecten in die situatie uitblijven die er bij stilwerkende proefpersonen wel zouden zijn. Dat leidde tot de hypothese:

- H1 In de online bibliotheekcatalogus maken proefpersonen in de synchrone conditie meer observeerbare fouten dan proefpersonen in de retrospectieve conditie.

Op basis van eerder onderzoek naar de invloed van hardopdenken op het leesproces verwachtten we bij de gemeentelijke website juist het tegenovergestelde (Silvén & Vauras, 1992; Loxterman, Beck & McKeown, 1994; Kucan & Beck, 1997). Verondersteld kan worden dat de opdracht om hardop te denken leidt tot een meer systematische werkwijze, waarbij moeilijk of niet te verbaliseren deelactiviteiten als browsen of skimmen worden vervangen door lezen. Daarnaast zou ook de vertraging van de taakuitvoering, die gepaard gaat met synchroon hardopdenken, kunnen leiden tot een minder impulsieve aanpak. Dergelijke mogelijke effecten van hardopdenken werden niet meegenomen in het overzicht van Ericsson (1988), die zich beperkte tot de effecten van het hardop lezen en het verbaliseren na elke zin in de tekst, en die op grond daarvan aanmerkelijk optimistischer was over de beperkte effecten van de hardopdenkmethode op tekstbegrip. Dat leidde tot de volgende hypothese:

- H2 Op de gemeentelijke website maken proefpersonen in de synchrone conditie minder observeerbare fouten dan proefpersonen in de retrospectieve conditie.

Wat betreft de verbalisaties verwachtten we geen interactie-effect tussen de hardopdenkvariant en het testobject. Hier verwachtten we dat de retrospectieve hardopdenkmethode in beide gevallen meer ruimte zou geven aan de proefpersonen om problemen te verbaliseren. Onze laatste hypothese was derhalve:

- H3 Zowel in de online bibliotheekcatalogus als op de gemeentelijke website verbaliseren proefpersonen in de synchrone conditie minder problemen dan proefpersonen in de retrospectieve conditie.

3. Onderzoeksopzet

In deze paragraaf geven we een beknopt overzicht van de opzet van het onderzoek. Een uitgebreidere weergave is te vinden in de afzonderlijke verslaglegging van beide studies (Van den Haak, De Jong & Schellens, 2004; 2007).

De online bibliotheekcatalogus van de Universiteit Utrecht werd getest met 40 studenten als proefpersonen: 20 in de synchrone hardopdenkconditie, en 20 in de retrospectieve hardopdenkconditie. De proefpersonen kregen vijf taken met de catalogus. In de synchrone conditie moesten ze tijdens de taakuitvoering hun gedachten verbaliseren. In de retrospectieve conditie bestond het onderzoek uit twee stappen: eerst moesten de proefpersonen stilwerkend de taken met de catalogus verrichten; vervolgens moesten ze hun gedachten verbaliseren terwijl ze naar een video-opname van hun taakuitvoering keken.

De website van de gemeente Haarlem werd eveneens met 40 studenten geëvalueerd, die verdeeld werden over de twee condities. In dit geval kregen de proefpersonen vijf scenario's voorgeschoteld, die elk bestonden uit twee tot vier taken. In totaal moesten ze twaalf vragen aan de hand van de website beantwoorden.

In beide gevallen richtte de analyse zich primair op de detectie van mogelijke gebruikersproblemen. Daarbij maakten we onderscheid tussen drie typen problemen. Het eerste type

bestond uit de *observeerbare* problemen: dat waren problemen waarbij de proefpersoon in de fout ging zonder dat dit leidde tot een vorm van verbalisatie. De detectie van deze problemen was gebaseerd op een vergelijking van het ideale handelingsverloop met de gebeurtenissen op het scherm. Het kon bijvoorbeeld zijn dat een proefpersoon een verkeerde zoekterm invulde of een verkeerde link aanklikte. Goed beschouwd zouden er voor dergelijke problemen helemaal geen hardopdenkgegevens nodig zijn. Het tweede type problemen bestond uit de *geverbaliseerde* problemen: dat waren problemen die niet zichtbaar waren op het scherm, maar die wel door een proefpersoon onder woorden werden gebracht. Het kon bijvoorbeeld zijn dat een proefpersoon duidelijk maakte te twijfelen tussen meerdere links of in verwarring te zijn geraakt. Het derde type bestond uit *gemengde* problemen: dat waren problemen waarbij de proefpersoon zowel een observeerbare fout had gemaakt als geverbaliseerd had. Dit is enigszins een restcategorie, omdat de relatie tussen verbalisaties en handelingen verschillende vormen kon aannemen. Het kon zijn dat een proefpersoon eerst een probleem verbaliseerde, waarna het in de daaropvolgende actie ook zou blijken. Maar het gebeurde ook dat een proefpersoon eerst een fout maakte en er vervolgens over begon te praten. Het kwam ook voor dat de foutieve handeling en de verbalisatie van het probleem op verschillende momenten in het proces plaatsvonden. Vanwege dit gemengde karakter hebben we besloten om deze laatste categorie in onze analyses buiten beschouwing te laten en ons te concentreren op de zuivere observeerbare en geverbaliseerde problemen.

4. Resultaten

De hoofdeffecten van het testobject en van de hardopdenkconditie en de interactie-effecten tussen beide zijn getoetst in een multivariate variantieanalyse met het aantal observeerbare problemen en het aantal geverbaliseerde problemen als afhankelijke variabelen. De resultaten van deze analyse zijn te vinden in Tabel 1 (gemiddelde scores en standaarddeviaties) en Tabel 2 (toetsing).

Een eerste bevinding is dat het aantal observeerbare problemen per proefpersoon significant verschilde tussen de twee testobjecten. In de online bibliotheekcatalogus werden, ongeacht de hardopdenkconditie, significant meer observeerbare fouten gemaakt dan in de gemeentelijke website. De η^2 van .17 duidt op een groot effect. Dit correspondeert met de aard van de twee testobjecten: bij het gebruik van een website moeten relatief meer mentale en minder fysieke handelingen worden verricht. Wel moet daarbij worden vastgesteld dat het aantal geverbaliseerde problemen niet verschilde tussen de beide testobjecten.

Tabel 1: Gemiddelde aantallen problemen per proefpersoon in de vier condities

	Observeerbare problemen		Geverbaliseerde problemen	
	Synchroon	Retrospectief	Synchroon	Retrospectief
Online bibliotheekcatalogus	5.5 (2.5)	3.1 (1.7)	1.7 (2.1)	3.4 (2.3)
Gemeentelijke website	1.8 (1.9)	2.8 (2.8)	2.0 (2.1)	3.4 (2.5)

Noot: Standaardafwijking (SD) tussen haakjes

Tabel 2: *Toetsing van de verschillen (Manova)*

	<i>F</i> -waarde	Vrijheidsgraden	Significantie	<i>Eta</i> ²
Overall-effect testobject	8.230	2.75	$p < .001$	.18
Overall-effect hardopdenkconditie	4.628	2.75	$p < .05$	.11
Overall interactie-effect	6.266	2.75	$p < .005$	.14
Effect testobject op:				
- observeerbare problemen	15.166	1.76	$p < .001$	.17
- geverbaliseerde problemen	.062	1.76	n.s.	
Effect hardopdenkconditie op:				
- observeerbare problemen	1.858	1.76	n.s.	
- geverbaliseerde problemen	9.296	1.76	$p < .005$	.11
Interactie-effect op:				
- observeerbare problemen	11.611	1.76	$p < .001$	.13
- geverbaliseerde problemen	.062	1.76	n.s.	

Vervolgens blijkt dat de hardopdenkconditie van invloed was op het aantal geverbaliseerde problemen. In de retrospectieve hardopdenkconditie werden meer problemen geverbaliseerd dan in de synchrone hardopdenkconditie. Wederom duidt de η^2 (.11) op een aanzienlijk effect. Met deze bevinding is de derde onderzoekshypothese, die betrekking had op de volledigheid van synchrone hardopdenkprotocollen, bevestigd. Er blijken achteraf problemen in het proces te zitten die tijdens de taakuitvoering niet geverbaliseerd worden door de proefpersonen. En deze constatering is onafhankelijk van de aard van het testobject. Onzes inziens even interessant is de constatering dat er geen verschil is tussen de condities in aantallen observeerbare problemen. Er zijn, in overeenstemming met onze bespreking van de literatuur, geen eenduidige uitspraken te doen over de reactiviteit van synchrone hardopdenkprotocollen.

Wanneer we kijken naar de interactie-effecten, blijkt dat het aantal observeerbare problemen samenhangt met het samenspel tussen testobject en hardopdenkconditie. Ook hier is de η^2 aanzienlijk (.13). In het geval van de online bibliotheekcatalogus lijkt het gebruik van de synchrone hardopdenkmethode problemen te veroorzaken bij de proefpersonen. Dit wordt bevestigd door een *t*-test: proefpersonen in de synchrone conditie hebben significant meer observeerbare problemen in de catalogus dan proefpersonen in de retrospectieve conditie ($t=3.601$, $df=38$, $p < .001$, Cohen's $d=1.14$). Daarmee is de eerste onderzoekshypothese bevestigd. In het geval van de gemeentelijke website is de tendens juist in de tegenovergestelde richting: het gebruik van de synchrone hardopdenkmethode lijkt problemen te voorkomen bij de proefpersonen. Dit verschil is echter niet significant (*t*-test, $t=-1.365$, $df=38$, $p=.180$). De tweede onderzoekshypothese is daarmee niet bevestigd, maar de onderliggende veronderstelling dat de reactiviteit van de synchrone hardopdenkmethode samenhangt met de aard van het testobject blijft wel overeind.

5. *Discussie*

In de context van formatief evaluatieonderzoek wordt de hardopdenkmethode doorgaans gezien als een valide benadering, die veel van de nadelen van methoden op basis van zelf-rapportage (zoals de plus-en-minmethode) niet heeft. Een onmiskenbaar voordeel van de methode, althans wanneer er gewerkt wordt met taken, is dat we niet hoeven te vertrouwen op oordelen en inschattingen van de proefpersonen, maar dat we ook kunnen zien of ze daadwerkelijk met het communicatiemiddel uit de voeten kunnen. Het huidige onderzoek plaatst enkele vraagtekens bij de zekerheden die we meenden te hebben. Zo blijkt dat de problemen die proefpersonen tegenkomen in sommige gevallen worden veroorzaakt doordat zij tijdens de taakuitvoering hardop moeten denken. Deze bevinding correspondeert met de resultaten van een vorige studie, waarin synchrone en retrospectieve hardopdenkprotocollen werden vergeleken voor de evaluatie van een andere online bibliotheekcatalogus (Van den Haak, De Jong & Schellens, 2003). In die vorige studie vonden we zelfs dat ook het percentage correct uitgevoerde taken verschilde: synchroon hardopdenkende proefpersonen waren daarin significant minder succesvol dan retrospectief hardopdenkende proefpersonen. In het huidige onderzoek vinden we geen verschil in succespercentage, wat mogelijk is toe te schrijven aan het geringere aantal taken dat de proefpersonen moesten verrichten. Tot zover bevestigt het onderzoek de resultaten van de vorige studie. Maar uit de vergelijking met de gemeentelijke website blijkt in de huidige studie dat de reactiviteit van de synchrone hardopdenkmethode contextafhankelijk is. De opdracht om hardop te denken had op de website niet dezelfde uitwerking als in de bibliotheekcatalogus. De synchroon hardopdenkende proefpersonen presteerden weliswaar niet significant beter dan de retrospectief hardopdenkende proefpersonen, maar het is op basis van onze resultaten wel aannemelijk dat er testobjecten en opdrachten zullen zijn waarbij het hardopdenken faciliterend werkt voor de taakuitvoering.

Deze constatering vraagt om vervolgonderzoek dat moet leiden tot een raamwerk waarin een taaktypologie wordt verbonden aan assumpties over de reactiviteit van het hardopdenken. Daarvoor is een taakanalyse nodig waarin het gebruik van websites, interfaces en instructieve teksten zoals handleidingen of formulieren wordt uitgesplitst in eenduidige deeltaken. In zorgvuldig afgebakende experimenten kan vervolgens worden nagegaan welke invloed het hardopdenken heeft op de taakuitvoering. Dergelijk onderzoek zou, binnen het domein van de formatieve evaluatie maar mogelijk ook daarbuiten, substantieel bijdragen aan de theorievorming over de waarde en beperkingen van het hardopdenken als onderzoeksmethode.

Een andere bevinding is dat de synchrone hardopdenkmethode leidt tot minder geverbaliseerde problemen dan de retrospectieve hardopdenkmethode. Dit leidt tot vragen over de aard van synchrone en retrospectieve hardopdenkprotocollen. Een voor de hand liggende verklaring voor eventuele verschillen zou zijn dat de proefpersonen in de retrospectieve conditie meer los komen van de taakuitvoering. Toch lijkt het recente onderzoek van Guan e.a. (2006) dit niet te bevestigen: de retrospectieve verbalisaties betroffen passages waar de proefpersonen tijdens de taakuitvoering daadwerkelijk aandacht aan besteedden, zoals bleek uit hun eye-tracking data. Met andere woorden: ook in retrospectief hardopdenkonderzoek blijken proefpersonen behoorlijk getrouw hun eigen taakuitvoering te volgen. Gedetailleerd tekstanalytisch onderzoek naar de verschillen tussen verbalisaties in synchrone en retrospectieve hardopdenkprotocollen zou in dit opzicht erg waardevol zijn.

De bovengenoemde vragen hebben primair een theoretische inslag. Maar er zijn ook nog allerlei praktische vragen die evenzeer van belang lijken voor de validiteit van de methode. De discussie over proefleidergedrag die door Boren & Ramey (2000) is aangezwengeld, zal op basis van empirisch onderzoek tot een ontknoping moeten komen. Aan welke richtlijnen moet een proefleider zich precies houden in de “speech communication”-benadering en wat zijn de effecten van zulk proefleidergedrag? Ook de instructie die aan hardopdenkende proefpersonen wordt gegeven en de oefening of demonstratie die de proefpersonen aan het begin van een sessie krijgen voorgeschoteld, zijn uiterst relevant. Het verbaliseren van gedachten is niet iets wat alle proefpersonen intuïtief kunnen. De systematische ontwikkeling van een onderzoekscontext die proefpersonen optimaal voorbereidt op de hardopdenktaak lijkt praktisch gesproken de meest dringende kwestie waar onderzoek naar gedaan moet worden.

Dat de talrijke vragen met betrekking tot (synchrone en retrospectieve) hardopdenkprotocollen vooralsnog onbeantwoord zijn, is wellicht te wijten aan de robuustheid van de methode en aan het onmiddellijk herkenbare nut van de resultaten die deze oplevert. In de drie studies die we tot nu toe hebben verricht, waren er bijvoorbeeld geen systematische verschillen te ontdekken in de lijsten met problemen die op basis van synchrone en retrospectieve hardopdenkprotocollen werden gevonden. En de meeste problemen die we vonden leken zonder meer de moeite waard (wat ook bleek uit een analyse waarin experts de ernst en aannemelijkheid van de problemen moesten beoordelen). Deze beide kenmerken hebben mogelijk bijgedragen tot de mainstream kennis over usability testing die zich al een decennium lang nauwelijks ontwikkelt. In adviesliteratuur over formatief evaluatieonderzoek en op congressen wordt al jarenlang steeds dezelfde informatie over hardopdenkonderzoek herhaald. Om op dit gebied een stap verder te komen zal er meer gedetailleerd en systematisch onderzoek gedaan moeten worden naar de werking van de hardopdenkmethode.

Een belangrijke implicatie van het onderzoek dat we hier beschreven, is dat we voorzichtig moeten zijn om methodologische inzichten die met het ene testobject zijn verkregen, te vertalen naar andere onderzoeksobjecten. Het onderzoek naar interfaces en software vertoont grote verschillen met de praktijk van tekst- of website-evaluatie. Inzichten uit die hoek moeten derhalve met de nodige voorzichtigheid worden geïnterpreteerd. Denk bij deze inzichten niet alleen aan de werking van hardopdenkprotocollen, maar mogelijk ook aan heuristische evaluatie, cognitive walkthroughs en scenario-evaluatie (Nielsen & Mack, 1994) en aan de optimistische inschattingen over aantallen proefpersonen die nodig zouden zijn voor een betrouwbaar usability test resultaat (Nielsen, 1994; Virzi, 1992). Een bijkomende overweging is dat veel onderzoek uit deze hoek methodologisch niet sterk is (Gray & Salzman, 1998) en dat de verslaglegging ervan vaak zo bondig is dat er veel te raden overblijft over de precieze opzet van het onderzoek. We kunnen dergelijke studies beter als inspiratie zien voor eigen onderzoek naar tekst- en website-evaluatie dan als volwaardige en toepasbare bron van kennis.

Bibliografie

- Boren, M.T., & Ramey J. (2000).** Thinking aloud: Reconciling theory and practice. *IEEE Transactions on Professional Communication*, 43, 261-278.
- Dieli, M. (1986).** *Designing successful documents: An investigation of document evaluation methods*. Diss. Carnegie-Mellon University. Pittsburgh.
- Dumas, J.S., & Redish, J.C. (1993).** *A practical guide to usability testing*. Norwood, NJ: Ablex.
- Ericsson, K.A. (1988).** Concurrent verbal reports on text comprehension: A review. *Text*, 8, 295-325.
- Ericsson, K.A., & Simon, H.A. (1993).** *Protocol analysis. Verbal reports as data*. Cambridge, MS: MIT Press.
- Flower, L., Hayes, J.R., & Swarts, H. (1983).** Revising functional documents: the scenario principle. In: P.V. Anderson, R.J. Brockmann & C.R. Miller (Eds.), *New essays in technical and scientific communication. Research, theory and practice* (pp. 41-58). Farmingdale, NY: Baywood.
- Gray, W.D., & Salzman, M.C. (1998).** Damaged merchandise? A review of experiments that compare usability evaluation methods. *Human-Computer Interaction*, 13, 203-261.
- Guan, Z., Lee, S., Cuddihy, E., & Ramey, J. (2006).** The validity of the stimulated retrospective think-aloud method as measured by eye tracking. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems, Montréal, Québec, Canada* (pp. 1253-1262). New York: ACM Press.
- Haak, M. van den, Jong, M. de, & Schellens, P.J. (2003).** Hardopdenkprotocollen als pretestmethode: Synchroon en retrospectief hardopdenken vergeleken. *Tijdschrift voor Taalbeheersing*, 25, 236-252.
- Haak, M.J. van den, Jong, M.D.T. de, & Schellens, P.J. (2004).** Employing think-aloud protocols and constructive interaction to test the usability of online library catalogues: A methodological comparison. *Interacting with Computers*, 16, 1153-1170.
- Haak, M.J. van den, Jong, M.D.T. de, & Schellens, P.J. (2007).** Evaluation of a municipal Web site: Three variants of the think-aloud method compared. *Technical Communication*, 54 (te verschijnen).
- Jansen, C.J.M., & Steehouder, M.F. (1989).** *Taalverkeersproblemen tussen overheid en burger. Een onderzoek naar verbeteringsmogelijkheden van voorlichtingsteksten en formulieren*. Diss. Rijksuniversiteit Utrecht. 's-Gravenhage: SDU.
- Janssen D., Waes, L. van, & Bergh, H. van den (1996).** Effects of thinking aloud on writing processes. In: C.M. Levy (ed.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 233-250). Mahwah, NJ.: Lawrence Erlbaum.
- Jong, M. de (1998).** *Reader feedback in text design. Validity of the plus-minus method for the pretesting of public information brochures*. Dissertation University of Twente. Amsterdam/Atlanta, GA: Rodopi.
- Jong, M. de, & Schellens, P.J. (1995).** *Met het oog op de lezer. Pretestmethoden voor schriftelijk voorlichtingsmateriaal*. Amsterdam: Thesis.
- Jong, M. de, & Schellens, P.J. (2000).** Toward a document evaluation methodology. What does research tell us about the validity and reliability of evaluation methods? *IEEE Transactions on Professional Communication*, 43, 242-260.
- Knoblich, G., & Rhenius, D. (1995).** Zur Reaktivität Lauten Denkens beim komplexen Problemlösen [Reactivity of thinking aloud during complex problem solving]. *Zeitschrift für experimentelle und angewandte Psychologie*, XLII, 419-454.
- Kucan, L., & Beck, I.L. (1997).** Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, 67, 271-299.
- Lentz, L., & Pander Maat, H. (2003).** Waarom het lezersprotocol zo'n goede methode is om begripsproblemen op te sporen. *Tijdschrift voor Taalbeheersing*, 25, 202-220.
- Loxterman, J.A., Beck, I.L., & McKeown, M.G. (1994).** The effects of thinking aloud during reading on students' comprehension of more or less coherent text. *Reading Research Quarterly*, 29, 353-367.

- Mumma, G.H., Draguns, J.G., & Seibel, R. (1993).** Reactive affects of concurrent verbalization in person perception tasks. *European Journal of Social Psychology*, 23, 295-311.
- Nielsen, J. (1993).** *Usability engineering*. Boston, MA: Academic Press.
- Nielsen, J. (1994).** Estimating the number of subjects needed for a thinking aloud test. *International Journal of Human-Computer Studies*, 41, 385-397.
- Nielsen, J., & Mack, R.L. (1994).** *Usability inspection methods*. New York: John Wiley.
- Schrivver, K.A. (1987).** *Teaching writers to anticipate the reader's needs. Empirically based instruction*. Diss. Carnegie-Mellon University, Pittsburgh.
- Silvén, M., & Vauras, M. (1992).** Improving reading through thinking aloud. *Learning and Instruction*, 2, 69-88.
- Swaney, J.H., e.a. (1981).** *Editing for comprehension: improving the process through reading protocols*. Technical report no. 14. Document Design Project, Carnegie-Mellon University, Pittsburgh, PA.
- Virzi, R.A. (1992).** Refining the test phase of usability evaluation: How many subjects is enough? *Human Factors*, 34, 457-468.
- Wright, P.C., & Monk, A.F. (1991).** A cost-effective evaluation method for use by designers. *International Journal of Man-Machine Studies*, 35, 891-912.

Schrijfprocessen registreren met Inputlog

Een data-analyse van de interactie met de 'reeds geproduceerde tekst'

1. Inleiding

De laatste jaren is de computer niet alleen een schrijfinstrument maar ook steeds meer een onderzoeksinstrument dat we kunnen inzetten bij schrijfonderzoek. In dit artikel stellen we Inputlog voor. Inputlog is een computerprogramma dat onderzoekers in staat stelt digitale schrijfprocessen in detail te registreren en te analyseren. Spelman Miller & Sullivan (2006) beschrijven in de introductie van hun boek over toetsregistratie de functie van dit soort onderzoeksinstrumenten als volgt: *"As an observational tool, keystroke logging offers the opportunity to capture details of the activity of writing, not only for the purposes of the linguistic, textual and cognitive study of writing, but also for the broader applications concerning the development of language learning, literacy, and language pedagogy."* (p. 1).

Op dit moment bestaan er verschillende registratie-instrumenten met een vergelijkbaar concept. In essentie hebben ze met elkaar gemeen dat ze allemaal toetsaanslagen en muisbewegingen als data opslaan zodat die later beschikbaar zijn voor verwerking. Deze voortdurende registratie van inputgegevens infereert niet met het normale gebruik van de computer. Toetsregistratie-instrumenten vormen dus geen (technische) belasting tijdens het schrijven, waardoor er een ecologisch valide onderzoekscontext ontstaat. De geregistreerde data vormen de basis voor verdere analyses. Die analyses kunnen gedaan worden in de registratieprogramma's zelf, of kunnen geëxporteerd worden naar statistische programma's zoals SPSS of SAS. De dataregistratie en -verwerking kan veel sneller gebeuren dan handmatig ooit het geval was.

Samenvatting

Dankzij de ontwikkeling van toetsregistratieprogramma's is het mogelijk geworden om digitale schrijfprocessen nauwkeurig te registreren en te analyseren. Dit artikel beschrijft Inputlog, een registratieprogramma dat onderzoekers de mogelijkheid biedt om schrijfprocessen te loggen in Windowstoepassingen om die daarna vanuit een aantal invalshoeken te analyseren. We beschrijven de technische en functionele kenmerken van dit onderzoeksinstrument en illustreren een aantal toepassings-mogelijkheden aan de hand van data uit een experimenteel onderzoek. Het centrale thema in dit experiment is de interactie met de 'reeds geproduceerde tekst' ('text produced so far' - TPSF) bij het formuleren van nieuwe tekst. De resultaten tonen aan dat de foutgrootte in de TPSF een determinerende factor is voor de belasting van het werkgeheugen van de schrijver. Ook blijkt dat de aard van de fouten in de TPSF de manier bepaalt waarop de schrijver in interactie treedt met de bestaande tekst. Het artikel sluit af met een vooruitblik op nieuwe toepassingen van Inputlog die nog in ontwikkeling zijn.

Afhankelijk van de onderzoeksvraag kunnen onderzoekers verschillende aspecten van het (cognitieve) schrijfproces of het schrijfgedrag analyseren door bijvoorbeeld temporele data (zoals absolute tijd en pauzes) en procesdata (toetsaanslagen en muisbewegingen) te combineren en te interpreteren. Pauzegedrag is al geruime tijd een belangrijk onderzoeksthema in het domein van de gesproken taal. In het schrijfonderzoek heeft de rol van pauzes meer aandacht gekregen toen niet langer alleen schrijfproducten, maar ook schrijfprocessen object van onderzoek werden. Het onderzoek van Matsuhashi (1981) heeft daartoe de aanzet gegeven; voor een overzicht van het onderzoek op dit terrein de laatste 25 jaar, verwijzen we graag naar een artikel van Wengelin (2006). Pauzes vormen in de eerste plaats een interessante invalshoek om een beter beeld te krijgen van de mentale activiteit tijdens het schrijven. Elke (langere) pauze kunnen we immers opvatten als een onderbreking van een vloeiend proces, waarbij de schrijver tijd kan nemen voor drie mogelijke activiteiten (Goldman-Eisler, 1972): fysiologische activiteiten (bijvoorbeeld overgang van toetsenbord naar muis); cognitieve (bijvoorbeeld planning van een revisie) of communicatief-retorische (afstemming van de formulering op het doelpubliek). Die onderbrekingen vormen telkens indicatoren van een afwijkende (grotere) mentale belasting. Het aantal pauzes, de lengte van de pauzes, de distributie ervan en de (syntactische) plaats vormen daarom allemaal invalshoeken voor analyses die tot doel hebben het onderliggende cognitieve schrijfproces beter te begrijpen (Jansen et al., 1996; Spelman Miller, 2006). Bij de analyse van het pauzegedrag moeten we uiteraard altijd beseffen dat het enkel om een indirecte meting gaat van de cognitieve belasting en dat we dus altijd afhankelijk zijn van interpretaties en inferenties met alle onvolkomenheden vandie. Daarom is het vaak aan te raden om complementaire analyses en observatiemethodes in te zetten (zie andere bijdragen in dit nummer).

In deze bijdrage voor het themanummer over ‘onderzoeksinstrumenten’ voor schrijf- en leesonderzoek beschrijven we Inputlog: een (toets)registratieprogramma voor schrijfprocesonderzoek in Windowsomgevingen. Eerst gaan we in op de technische en functionele kenmerken van het programma. Daarna lichten we het gebruik van Inputlog toe aan de hand van een experiment waarbij foutdetectie tijdens het schrijfproces centraal stond¹. We sluiten af met de toekomstplannen voor Inputlog.

2. Kenmerken van Inputlog

Voor de ontwikkeling van Inputlog konden we terugvallen op de functionaliteit van twee bestaande toetsregistratieprogramma's: enerzijds JEdit and Trace-it (Severinsson Eklundh, 1994; Severinsson Eklundh & Kollberg, 1992; Severinsson Eklundh, 1996; Severinsson Eklundh & Kollberg, 1996; 2003; Kollberg, 1998; Spelman Miller & Sullivan, 2006), en anderzijds ScriptLog (Strömqvist & Malmsten, 1997). JEdit en Trace-it zijn enkel geschikt voor Macintosh computers. JEdit registreert bovendien alleen data in een zelfontwikkelde en beperkte tekstverwerker. ScriptLog registreert ook voornamelijk in een tekstverwerker die speciaal ontwikkeld is voor onderzoeksdoeleinden, nl. schrijfexperimenten met jonge kinderen op basis van plaatjes. Een belangrijk onderscheidend kenmerk van Trace-it is de uitgebreide interactieve revisiemodule² (Severinsson Eklundh, 1994) terwijl ScriptLog bijvoorbeeld als enige tot nu toe de mogelijkheid biedt om schrijfprocesdata te combineren met oogbewegingregistratie (zie ook Cozijn en Lentz, Mak & Pander Maat in dit nummer).

Zoals blijkt uit deze beschrijving, zijn de meeste toetsregistratie-instrumenten ontwikkeld voor een specifieke computeromgeving en/of functioneren ze niet in een Windows-omgeving³. Dit betekent dat ze niet gebruikt kunnen worden in een van de meest gebruikte schrijffomgevingen waar schrijvers gebruikmaken van commerciële tekstverwerkers zoals MS Word of WordPerfect. Dit is een van de belangrijkste redenen geweest om in 2003 te starten met de ontwikkeling van Inputlog. Een tweede reden om Inputlog te ontwikkelen is de opkomst van spraakherkenning als schrijfmodus voor tekstverwerking. De bestaande toetsregistratieprogramma's bieden namelijk niet de mogelijkheid om input van spraakherkenning te registreren. In vorig onderzoek (Leijten & Van Waes, 2005) hebben we de procesdata van teksten die met spraakherkenningssoftware geschreven waren, nog handmatig geanalyseerd. In het follow-up experiment willen we deze analyses via Inputlog uitvoeren om op die manier het zeer tijdsintensieve codeerwerk op te vangen.

Kort samengevat is het doel van Inputlog om analyses van schrijfprocesdata sneller en accurater te maken. De belangrijkste functies van dit registratieprogramma zijn:

- opslaan van data van schrijfsessies in een Windowsomgeving;
- genereren van databestanden als input voor verdere (statistische) analyse van tekst-, proces-, pauze-, revisie- en moduskenmerken van de schrijfsessie(s);
- integreren van loggingdata uit andere programma's;
- afspelen van geregistreerde schrijfprocessen met verschillende snelheden.

Inputlog is in eerste instantie ontwikkeld om schrijfprocessen in MS Word te registreren en te analyseren. Daarnaast registreert het echter ook toetsaanslagen en muisbewegingen in elk ander Windowsprogramma. Dit maakt Inputlog met andere woorden ook inzetbaar bij onderzoek naar bijvoorbeeld invullen van webformulieren of programmeren in eender welke Windowsgebaseerde programmeertaal. Het programma is beschikbaar via de website www.inputlog.net. Voor onderzoeksdoeleinden is het gratis op voorwaarde dat bij de publicatie van onderzoeksresultaten verwezen wordt naar de aangegeven basispublicatie. We stellen commentaar en suggesties van gebruikers sterk op prijs.

2.1 Technische beschrijving. In deze paragraaf beschrijven we Inputlog vanuit een technisch perspectief. Het centrale uitgangspunt van het programma is dat het alle inputdata registreert voordat ze naar scherm informatie worden omgezet. Concreet wil dit zeggen dat Inputlog het volgende registreert:

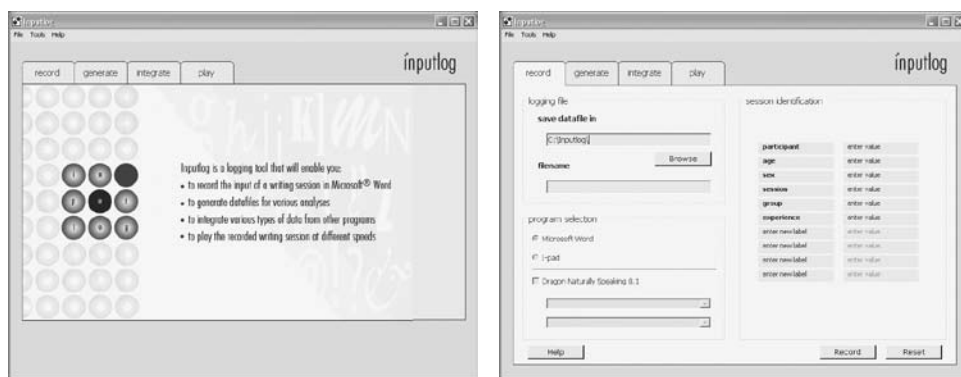
- scancodes van toetsen (bijvoorbeeld scancode '12' verwijst naar de letter 'e');
- muisbewegingen (kliks, bewegingen, schermlocaties);
- tijdsaanduidingen van alle 'input events'.

Al deze data worden bewaard in een zogenaamd IDF-bestand. IDF staat voor Inputlog Data File. In dit bronbestand wordt iedere 'event' of actie van het schrijfproces afzonderlijk bewaard. Dit bestand wordt achteraf geconverteerd⁴ naar verschillende leesbare output-bestanden; hierdoor worden de ruwe data geschikt gemaakt voor kwalitatieve en kwantitatieve analyses.

Het bronbestand kan op twee manieren gebruikt worden. Aan de ene kant voor de conversie van bestanden (*generate*). Aan de andere kant als input voor de afspeelmodule (*play*). Bij de conversie van bestanden worden de data in het bronbestand vertaald naar verschillende

analyses. Dit stelt onderzoekers in staat om pauze-, modus- en tekstdata op een geaggregeerd niveau te analyseren. De integratiemodule (*integrate*) maakt het bovendien mogelijk om loggingdata uit andere programma's (bijvoorbeeld Morae, zie paragraaf 4.3) te integreren met data van Inputlog. Op die manier is het mogelijk de data te verrijken met complementaire gegevens. De afspeelmodule ten slotte zorgt ervoor dat onderzoekers het schrijfproces precies zo kunnen afspelen zoals het verlopen is, inclusief alle pauzes en revisies. Stap voor stap kan zo de tekstontwikkeling opnieuw bekeken worden. In de toekomst zal deze afspeelmodule ook de intermediaire basis vormen voor de revisieanalyse.

2.2 Functionele beschrijving. De interface van Inputlog bestaat uit een openingsscherm en vier tabbladen die de belangrijkste functies van het programma aangeven: *record*, *generate*, *integrate* en *play* (zie Figuur 1). In deze paragraaf geven we een kort overzicht van die basisfuncties. Voor meer gedetailleerde informatie verwijzen we naar het helpbestand van het programma.



Figuur 1: Schermafdruk van het openingsscherm en het 'record' tabblad.

Door de *opnamefunctie* (record) te selecteren, kunnen onderzoekers een loggingssessie configureren en starten. De configuratie gebeurt door de sessie te koppelen aan een (bestands)naam en een aantal kenmerken. Inputlog biedt de mogelijkheid om zes voorgedefinieerde schrijverskenmerken met een sessie te verbinden (o.a. deelnemer, leeftijd, geslacht) en ook nog vier extra variabelen te bepalen. Omdat die variabelen aan elke dataregel toegevoegd worden bij het genereren van analysebestanden, is het eenvoudig om achteraf de loggingssessie te identificeren en eventueel verschillende datasets samen te brengen voor verdere analyse, bijvoorbeeld in SPSS.

De *generatiefunctie* (generate) biedt een overzicht van de beschikbare analyses. De functie maakt het mogelijk om een IDF-bestand te openen en op basis van dat bronbestand een aantal analyses te genereren. Dat hoeft niet onmiddellijk na de registratie te gebeuren, maar kan ook later. Inputlog 2.0 Beta biedt de mogelijkheid om de volgende analyses⁵ uit te voeren:

1. **ALGEMENE ANALYSE:** De output van de algemene analyse bestaat uit een XML-bestand met daarin een zeer gedetailleerde weergave van de logging van een schrijfsessie. Voor

elke input-actie (gegenereerd met toetsenbord, muis of spraaktechnologie) wordt een dataregel gecreëerd met daarin de identificatie van de sessie, een representatie van de input, tijdsinformatie, en – voor muisactiviteiten – de schermpositie. Het XML-bestand kan geconverteerd worden naar een Excel-bestand of ingelezen worden in bijvoorbeeld SPSS voor verdere analyse.

2. LINEAIRE TEKST: De lineaire tekstrepresentatie geeft de tekstevolutie actie per actie weer. Bij een deletie zien we bijvoorbeeld eerst het oorspronkelijke woord, dan de deletie-operatie en ten slotte de nieuwe input. Extra opties maken het mogelijk om de lineaire tekst in tijdsperiodes of -intervallen te verdelen. Ten slotte kunnen ook pauzes getoond worden in deze representatie. De onderzoeker bepaalt zelf de pauzedrempel die hij of zij wil hanteren.
3. STATISTISCHE ANALYSE: De statistische analyse genereert een overzicht met basisstatistieken die op een geaggregeerd niveau een aantal sessie- en proceskenmerken (o.a. geproduceerde woorden en zinnen) weergeven en ook het pauzegedrag en het modusgebruik (toetsenbord vs. spraak) beschrijven.
4. PAUZE-ANALYSE: In de pauze-analyse wordt een XML-bestand gegenereerd die elke periode waarin niet actief geschreven wordt weergeeft en van een aantal kenmerken voorziet (o.a. lengte en positie). De pauzedrempel kan standaard op bijvoorbeeld 1, 2 of 5 seconden gedefinieerd worden of vrij bepaald worden door de onderzoeker.
5. MODUSANALYSE: Voor schrijfprocessen waarin verschillende inputmodi gebruikt worden (toetsenbord, muis en/of spraaktechnologie) is het mogelijk een modusanalyse te genereren die aangeeft wat het aandeel is in de schrijfproductie van elk van die schrijfmodi.

De *integratiefunctie* biedt de mogelijkheid om de XML-output van andere logging- en observatieprogramma's te combineren met de logging van Inputlog. Inputlog 2.0 Beta voorziet in een integratie van logbestanden die we genereren in Dragon Naturally Speaking 8.1⁶. De integratie met logbestanden uit andere programma's staat op de agenda zodat de huidige analyses verder geëxpandeerd kunnen worden en er een nog ruimere interpretatiebasis ontstaat (zie paragraaf 4).

Ten slotte biedt Inputlog ook de mogelijkheid om een gelogde *schrijfsessie opnieuw af te spelen*. Het IDF-bestand wordt daarbij als input gebruikt. Het betreft dus geen video-opname⁷. Dit heeft onder meer tot gevolg dat de programma- en scherminstellingen van de beginsituatie bij het afspelen identiek moeten zijn aan degene die gebruikt werden bij de start van de gelogde sessie. Het programma biedt ook de mogelijkheid om pauzelengtes tot een bepaald minimum te reduceren of procentueel te verkorten om op die manier de weergave te versnellen.

3. Toepassing bij schrijfprocesonderzoek (TPSF)

Om aan te geven hoe Inputlog allerlei soorten schrijfprocesonderzoek kan ondersteunen, werken we in deze paragraaf een voorbeeld uit waarmee we een aantal van de mogelijkheden illustreren. Het betreft een onderzoek waarbij we het effect op het werkgeheugen nagingen van verschillende soorten fouten in de 'reeds geproduceerde tekst' ('*text produced so far*' – TPSF). Voor een algemeen verslag van het onderzoek verwijzen we naar Leijten, Ransdell & Van Waes (in voorbereiding). In dit artikel gaan we dieper in op een aantal resultaten

van het onderzoek en complementeren die met data die we tijdens het experiment verzameld hebben met Inputlog. Op die manier willen we aantonen hoe loggingdata kunnen bijdragen tot een scherpere interpretatie van algemene schrijfprocesdata.

Uiteraard zijn er binnen het schrijfonderzoek nog vele andere toepassingen mogelijk voor toetsregistratieprogramma's. In essentie gaat het steeds over toepassingen waarbij de onderzoekers vooral geïnteresseerd zijn in het verloop van het schrijfproces (plannings- en revisieprocessen) en minder in het product. Een combinatie van beide invalshoeken behoort uiteraard tot de mogelijkheden. Lindgren en Sullivan (2006) onderscheiden o.m. de volgende toepassingsgebieden: schrijven in de moedertaal en in vreemde talen, schrijfmoeilijkheden, beginnende en expertschrijvers, vertaalprocessen en valideringen van schrijfprocesmodellen. Ook als basis voor (gestimuleerde) retrospectieve protocollen biedt toetsregistratie zeker mogelijkheden omdat de play-functie de onderzoeker in staat stelt om de schrijver direct te 'confronteren' met de ontwikkeling van zijn of haar eigen schrijfproces. Dit levert een veel concretere input op voor een uitgesteld hardopdenkprotocol⁸ (zie ook Van den Haak, De Jong & Schellens in dit nummer).

Het TPSF-experiment dat we hieronder beschrijven, is een voorbeeld waarbij we stilstaan bij een specifiek deelproces van de schrijfactiviteit, namelijk de interactie van de schrijver met de 'reeds geproduceerde tekst' ('text produced so far' of TPSF) op het moment dat die nieuwe tekst moet formuleren.

3.1 Beschrijving experiment. Schrijfprocesmodellen tonen duidelijk aan dat bij de productie van nieuwe tekst de 'reeds geproduceerde tekst' (TPSF) een belangrijke rol speelt (zie o.a. herziene schrijfprocesmodel van Hayes 1996). Ook in ons onderzoek naar het gebruik van spraaktechnologie bij professionele schrijfprocessen kwam dat duidelijk naar voren (Leijten & Van Waes, 2005). Het blijkt namelijk dat schrijvers voortdurend in interactie treden met de TPSF, bijvoorbeeld om een zin of een alinea verder af te werken. Die interactie kan echter verstoord worden door allerlei fouten. In schrijfprocessen met toetsenbord en muis kunnen dat bijvoorbeeld typefouten zijn; in schrijfprocessen met spraaktechnologie kan dat foutieve herkenning door de spraakherkenningssoftware zijn. Als dergelijke fouten optreden, moeten schrijvers beslissen of ze in eerste instantie de fout corrigeren, dan wel of ze eerst de tekst die ze in hun hoofd hadden uitschrijven of dicteren. We gaan ervan uit dat die overweging leidt tot een extra belasting van het werkgeheugen. In het voorliggende onderzoek gingen we na of er een verband bestaat tussen verschillende soorten fouten en de belasting van het werkgeheugen.

Drie thema's stonden daarbij centraal:

1. het effect van de spraakconditie op de foutanalyse: in het experiment werd de TPSF in twee verschillende condities aangeboden, namelijk enerzijds enkel geschreven en anderzijds zowel gesproken (voorgelezen) als geschreven (latin square design waarbij dezelfde zinnen gekruist werden aangeboden met en zonder spraak vooraf);
2. het effect van de lexicaliteit: de foutmanipulatie resulteerde enerzijds in onbestaande woorden, anderzijds in (andere) bestaande woorden (bijvoorbeeld: 'drongen' vs. 'donrgen' en 'conferentie' vs. 'concurrentie');
3. het effect van de foutgrootte: de foutmanipulatie in de TPSF betrof kleine fouten (één of twee letters, bijvoorbeeld 'ze voeg' i.p.v. 'ze vroeg') versus grote fouten (bijvoorbeeld 'on-

der de brief' i.p.v. 'onder de brug' of zelfs 'ongedaan maken' ('ogen dan maken'). De grote fouten werden verzameld uit een procescorpus van teksten die met spraakherkenning werden geschreven. We hebben beide soorten fouten in beide condities (met en zonder voorafgaande spraak) aangeboden om op die manier na te gaan of de sterkere verstoring van het woordbeeld bij grote fouten leidt tot een andere benadering van de bestaande tekst. Ook wilden we nagaan of de aan- of afwezigheid van een spraakconditie daarop een invloed heeft.

In dit artikel behandelen we enkel het derde thema, met name het effect van de foutgrootte op het werkgeheugen en de interactie van de schrijver met de reeds bestaande tekst, de TPSF. We beschrijven eerst bondig de opzet van het experiment en presenteren dan de belangrijkste resultaten. Daarbij besteden we ook expliciet aandacht aan de technische aspecten van de registratie en de analyse.

Deelnemers

Zestig studenten namen deel aan het experiment. Alle studenten spraken Nederlands als moedertaal en waren tussen 18 en 22 jaar oud. Ze werden willekeurig verdeeld in vier groepen. De studenten ontvingen een gratis filmkaartje als vergoeding.

Design

De kern van het experiment bestond erin dat de proefpersonen deelzinnen (TPSF) moesten afmaken die hen werden voorgelegd. In sommige van die deelzinnen kwamen fouten voor uit een van de gedefinieerde foutcategorieën (zie hieronder). Om ervoor te zorgen dat de deelzinnen een TPSF-karakter kregen, werd eerst een context gecreëerd. Daarbij werden telkens twee deelzinnen gepresenteerd die causaal verbonden waren. De proefpersonen konden ze rustig lezen en dan beslissen om de context weg te klikken. Door de zinnen op die manier te presenteren werd de context in het kortetermijngeheugen opgeslagen als basis voor de productietaak. In het experiment werden ook enkele afleiders (*fillerzinnen*) gebruikt waarbij de deelzinnen temporeel verbonden waren. Die laatste zinnen werden niet gebruikt in de analyses.

Voorbeeld

- (1) context: Omdat hij mijn resultaten wilde bekijken, haalde hij het verslag uit mijn kast.
- (2) TPSF-correct: Hij haalde het verslag uit mijn kast, want ...
- (3) TPSF-incorrect: Hij haalde het verslag uit mij kast, want ...

In het experiment werd een gekruist 2x2 design toegepast. De studenten kregen elk twee reeksen zinnen en de TPSF-zinnen werden de ene keer alleen schriftelijk aangeboden en de andere keer eerst voorgelezen (spraakconditie) voordat ze op het scherm verschenen. Het aantal fouten werd in elke set op een vergelijkbare manier gevarieerd. In elke reeks werd ervoor gezorgd dat de correcte zinnen het meest voorkwamen, om ervoor te zorgen dat de proefpersonen niet te sterk op foutcorrectie geconditioneerd werden. Een resultaatcontrole van de experimentele data toont aan dat ongeveer 90% van de fouten door de studenten werd opgemerkt en correct werd verbeterd. De aanvulling van de zinnen bleek ook in zeer hoge mate overeen te stemmen met de gegeven context (gemiddeld slechts een afwijking van 1.3 karakters t.o.v. het theoretische optimum).

Om de belasting van het werkgeheugen te meten werd gebruikgemaakt van de zogenaamde ‘tweede-taaktechniek’ (Kellogg 1994, 1996, 2001a, 2001b; Levy & Ransdell 2001; Olive 2004). Daarbij krijgen de proefpersonen een pieptoon te horen waarop ze zo snel moeten reageren door op een speciale schakelaar te drukken. Korte reactietijden bij een tweede taak, wijzen op een relatief lage belasting van het werkgeheugen; langere tijden op een hogere belasting. De pieptoon werden semi-willekeurig aangeboden tijdens het lezen van de TPSF Om conditionering te voorkomen, werden sommige zinnen niet vergezeld van een pieptoon, of soms werd de toon zeer laat aangeboden (met name in afleidende zinnen, cf. materiaalbeschrijving).

Procedure

Het experiment werd afgenomen in een computerklas en de software voor de verschillende onderdelen van het experiment werd volledig geprogrammeerd in Visual Basic.NET. De deelnemers kregen eerst algemene informatie over het verloop van het experiment in gesproken en geschreven vorm (voorgelezen schermtekst). Het experiment bestond uit verschillende delen. Voor elk deel werd een specifieke instructie gepresenteerd gevolgd door enkele oefentaken.

Het experiment werd als volgt opgebouwd:

- (1) reactietest: procedure om de neutrale, mediane reactietijd per persoon te meten als basis voor de tweede-taaktechniek (cf. supra);
- (2) eerste lees- en schrijfsessie (gevolgd door een korte pauze);
- (3) tweede lees- en schrijfsessie met identieke opdracht maar in andere conditie (met of zonder spraak).

Voor de logging werd enerzijds gebruik gemaakt van het .NET-programma en van Inputlog. Algemene tijden en tussentijdse product-output na elke schrijftaak werden geregistreerd met het .NET programma; het verloop van het schrijfproces bij het formuleren en reviseren van de zinnen, werd in een IDF-bestand van Inputlog opgeslagen.

Materiaal

De hoofdtak van het experiment was een lees- en schrijftaak. Daarvoor kregen de proefpersonen 60 zinnen aangeboden in twee sets. In elke set fungeerden zes zinnen als testzinnen. In 24 van de 60 zinnen werd een fout gepresenteerd in het TPSF-gedeelte. Die fouten waren gebaseerd op een foutenanalyse uit een vorige studie waarbij professionele schrijvers die gebruik maakten van spraaktechnologie werden geobserveerd (Leijten & Van Waes, 2003; 2005).

De fouten werden gecategoriseerd in vier groepen (zie tabel 1). De foutenclassificatie is gebaseerd op combinaties van drie factoren:

- foutgrootte: meer of minder dan twee karakters verschillend t.o.v. het bedoelde woord (voorbeeld: ‘kast’ wordt resp. ‘kroost’ en ‘krast’)
- schrijfmodus: fouten zijn kenmerkend voor spraakherkenning dan wel typisch voor toetsenbord en muis (bijvoorbeeld: ‘concurrentie’ wordt ‘conferentie’ (spraak) vs. concurrent-nie (typefout));
- lexicaliteit: de fout resulteert al dan niet in een ander bestaand woord (zie bovenstaand voorbeeld: fouten bij spraakherkenning resulteren steeds in andere bestaande woorden).

Tabel 1: Foutenclassificatie

Categorie	Soort fout		
	Foutgrootte	Schrijfmodus	Lexicaliteit
SR groot	groot: > 2 karakters	alleen in spraakherkenning (SR)	bestaand
SR klein	klein: ≤ 2 karakters	alleen in spraakherkenning (SR)	bestaand
SR Toetsenbord klein	klein: ≤ 2 karakters	in spraakherkenning (SR) en in toetsenbord&muismodus	bestaand
Toetsenbord klein	klein: ≤ 2 karakters	Alleen toetsenbord&muis	niet-bestaand

Alle zinnen waren ongeveer even lang: de context varieerde tussen de 70 en de 90 karakters; de deelzinnen bestonden uit 31 tot 43 karakters. De manipulatie van de foutgrootte in het materiaal voor het experiment werd gepretest door 32 studenten. Die beoordeelden de ernst van de fout op een 7-puntschaal. Met een clusteranalyse werd de beoogde technische categorisering vergeleken met de gescoorde, subjectieve inschatting van de ernst van de fout. Items die niet overeenstemden met het uitgangspunt werden verwijderd.

Analyses

Zes afhankelijke variabelen werden geselecteerd uit de loggingbestanden:

(a) *Starttijd*: De tijd die verloopt tussen het moment dat het contextscherm gesloten wordt en de eerste muisklik om de TPSF af te maken dan wel te corrigeren (zie Figuur 2). Voor de items waarbij de TPSF eerst werd voorgelezen, werd de spreektijd van de starttijd afgetrokken.

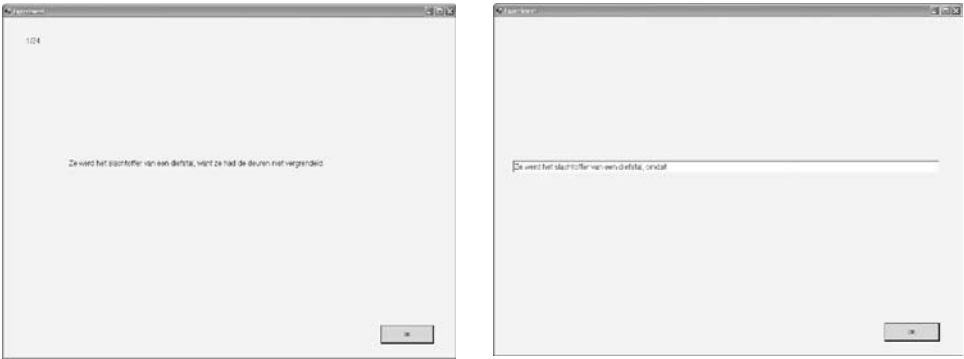
(b) *Cursorpositie*: Voor elk item werd bepaald of de cursor initieel gepositioneerd werd in het TPSF-gedeelte, dan wel of de proefpersonen de cursor achter de TPSF plaatsten om eerst de zin af te maken. In Inputlog zijn deze data bijvoorbeeld codeerbaar omdat in dit experiment het beginpunt van elke te vervolledigen zin op een vaste schermpositie geprogrammeerd was. Dat punt komt overeen met een vaste x-waarde in de geregistreeerde cursorpositie, i.c. 380. Waarden die kleiner zijn verwijzen naar een klik in de TPSF en vice versa⁹.

Voorbeeld van logging-output:

(1) Left Button: 00:00.15,109 (tijd) 274 (x-waarde) 345 (y-waarde) = in TPSF-zone

(2) Left Button: 00:00.18,329 (tijd) 390 (x-waarde) 339 (y-waarde) = in productiezone

In het experiment kozen in 89.2 % (SD: 15.6) van de gevallen de proefpersonen ervoor om eerst de zin af te maken. De individuele variatie schommelde tussen 50 % en 100 %.



Figuur 2: Schermafndruk van het context- en schrijfgedeelte van het TPSF-experiment.

(d) *Pauzetijd voor revisie:* De pauzetijd die de revisie in de TPSF voorafgaat, is een belangrijke indicator voor de cognitieve belasting. Om deze variabele te coderen maakten we gebruik van de lineaire representatie die Inputlog genereert. In die XML-output is het mogelijk om elke pauze voor een effectieve revisie te identificeren. In dit onderzoek hebben we de pauzedrempel bepaald op 500 ms. Als eerst de revisie in de TPSF plaatsheeft, valt de pauzetijd samen met de starttijd (=directe revisie); als de revisie volgt op de productie, dan vinden we de pauze terug op het einde van de afgemaakte zin en gaat ze een muis- of cursorbeweging vooraf (=uitgestelde revisie).

Voorbeeld: het woord ‘concurrentie’ wordt vervangen door ‘conferentie’

(1) *Directe revisie: eerst foutcorrectie, dan vervolledigen van de zin*

tijd	Lineaire representatie	Toelichting
860000 - 899999	Left Button[947,710] Movement[340,342] {4000} Left Button[340,342] Move- ment[454,350] {672} BS BS BS BS BS BS BS BS BS {688}fer{766}entie{547} Movement[414,352] {844}het · thema · interesseert · mij. Movement[904,697] {1141} Left Button[904,697]	Klik op OK ([x,y] binnen zone ‘ok’-knop) positionering in TPSF-zone pauze van 4 sec voor revisie (=pauzetijd direct) verwijderen van ‘concurrentie’ typen van ‘ferentie’ beweging naar productie-zone afmaken van zin Klik op OK

(2) *Uitgestelde revisie: eerst vervolledigen van de zin, dan foutcorrectie*

Tijd	Lineaire representatie	Toelichting
1020000 – 1039999	Left Button[893,711] {3266} Movement[388,345] {547} Left Button[389,345] {687}het · thema · inte{672}resseeer BS BS rt · mij. {1203} Movement[280,346] Left Button[295,345] Move- ment[308,350] fe Movement[912,702] {3328} Left Button[912,702]	Klik op OK pauze van ruim 3 seconden positionering in productiezone afmaken van zin (correctie typefout) pauze van 1.2 sec voor revisie (=pauze- tijd uitgesteld) beweging naar TPSF-zone pauze en selectie van 'cur' overtypen met 'fe' Klik op OK

(e) *Productietijd (schrijven)*: De productietijd is de tijd tussen het afsluiten van het context-scherm en het afsluiten van het itemscherm met de 'ok'-knop. In die tijd lezen de proefpersonen de TPSF, vullen ze de zin aan, herlezen ze die en brengen ze eventueel correcties aan. Omdat de 'ok'-knop op een vaste plaats op het scherm aangeboden werd (rechts onderaan) is het mogelijk om de coördinaten van de muiskliks hieraan te koppelen. In ons geval betrof het x-waarden tussen 855 en 950 en y-waarden tussen 675 en 720.

(f) *Geproduceerde karakters*: In de tekstanalyse werd in het .NET-programma per item het aantal geproduceerde en gerealiseerde karakters geteld. Post-hoc werd ook een vergelijking gemaakt tussen het aantal karakters dat we in een correcte zin verwachten (op basis van de aangeboden contextzin) en het aantal karakters in de eindzin van de proefpersoon. Dit vormde een eerste indicatie van de kwaliteit van de taakuitvoering. Het verschil bedroeg amper 1.3 karakter.

(g) *Reactietijd*: De reactietijd geeft de tijd aan die verstrijkt tussen het aanbieden van de piepton en het moment waarop de drukknop werd ingedrukt (cf. beschrijving van de tweede-taaktechniek hierboven). De logging hiervan gebeurde via het .NET-programma.

(h) *Succesratio (correctie van de fouten in de TPSF)*: Voor elk item met een fout in de TPSF werd nagegaan of het correct werd verbeterd. De succesratio geeft het percentage zinnen aan dat correct werd aangepast. De deelnemers aan ons experiment corrigeerden gemiddeld 86.6 procent van de fouten (SD: 8.4). Alleen die zinnen die correct verbeterd werden, zijn in de verdere analyses opgenomen¹⁰.

3.1 Resultaten. Voordat we ingaan op de analyse van de foutgrootte, presenteren we in Tabel 2 kort enkele algemene resultaten voor de vergelijking tussen de correcte en de incorrecte items.

Tabel 2 geeft aan dat wanneer schrijvers een correcte TPSF-deelzin gepresenteerd krijgen, ze minder tijd nodig hebben om te beslissen hoe ze de zin gaan aanpassen dan wanneer er een deficiënte TPSF getoond wordt ($t(59) = -5.25, p < .001$). Ook de totale productietijd om een correcte TPSF-deelzin te vervolledigen is korter ($M = 14.34$ sec, $SD = 3.46$) in

vergelijking met de incorrecte items ($M = 17.17$ sec, $SD = 2.16$ $t(59) = -10.956$, $p < .001$). Dat verschil ligt natuurlijk voor de hand omdat er geen tijd nodig is om de fout in de TPSF te corrigeren in dat laatste geval. De extra productietijd houdt echter ook verband met de langere starttijd. Tijdens die pauze wordt in een aantal gevallen waarschijnlijk de strategie overwogen om al dan niet eerst de correctie uit te voeren (cf. infra).

Tabel 2: Gemiddelde waarden voor de correcte vs. de incorrecte items

	Correct		Incorrect		p
	Gemiddelde	SD	Gemiddelde	SD	
Starttijd(sec)	1.46	0.51	1.70	0.60	**
Cursorpositie (%)	98.43	6.15	80.31	2.16	**
Productietijd (sec)	14.34	3.46	17.17	4.35	**
Reactietijd (ms)	332	173	347	170	-

*significantieniveau <.05; ** significantieniveau <.01

De analyse van de cursorpositie bij de aanvang van de taak geeft aan dat de schrijvers in ongeveer 20% van de gevallen er de voorkeur aan gaven om eerst de fout te corrigeren en dan pas de zin te vervolledigen¹¹.

Tot slot geeft deze algemene analyse aan dat er geen verschil gemeten werd tussen de reactietijden bij het lezen van correcte en incorrecte items. Een meer specifieke spiegelanalyse waarbij alleen die items bekeken werden die zowel correct als incorrect aangeboden werden ($n=4$), resulteerde in een significant verschil in reactietijd ($t(59) = -2.284$, $p < .05$). Dit toont aan dat het werkgeheugen op een bepaalde manier door de interactie met een (deficiënte) TPSF beïnvloed wordt. De onderstaande analyse waarbij het gedrag bij kleine en grote fouten vergeleken wordt, vormt daar al een eerste verduidelijking van.

Zoals aangegeven, kunnen we in deze bijdrage niet ingaan op alle deelresultaten. We beperken ons hier tot de analyse van de verschillen tijdens de interactie tussen grote en kleine fouten in de TPSF. Concreet gaat het om de vergelijking van items waarbij de fout resulteert in een ander bestaand woord waarvan minder of meer dan twee karakters verschillend zijn.

Voorbeeld

- (1) kleine fout
- context: Omdat hij mijn resultaten wilde bekijken, haalde hij het verslag uit mijn kast.
- TPSF: Hij haalde het verslag uit **mij** kast, want ...
- (2) grote fout
- context: Omdat het thema mij interesseert, ga ik volgende week naar die conferentie.
- TPSF: Ik ga volgende week naar die **concurrentie**, want ...

Tabel 3 geeft aan dat de schrijvers er beter in slagen om grote fouten te corrigeren in de TPSF dan kleine fouten. De succesratio is significant hoger bij grote fouten dan bij kleine ($t(59) = 2.38, p < .05$). Kleine fouten worden sneller over het hoofd gezien, vooral in de spraakconditie waarbij de TPSF ook voorgelezen wordt. In dergelijke gevallen ontstaat er immers bij het beluisteren van de tekst al een mentale representatie in het geheugen die bij een snelle confrontatie met de TPSF op het scherm niet leidt tot een detectie van het kleine verschil.

Kleine fouten worden dus niet altijd gedetecteerd, maar als ze gedetecteerd worden creëren ze een lagere belasting voor het werkgeheugen. Uit de analyse van de reactietijd blijkt namelijk dat de schrijvers over het algemeen meer tijd nodig hebben om op de pieptoon te reageren bij grote fouten dan bij kleine fouten ($F(1,59) = 4.35, p < .05$). Dat bevestigt ons uitgangspunt. Grotere fouten vormen immers een grotere verstoring bij de afstemming tussen de mentale en de fysieke representatie en vergen daardoor meer van het werkgeheugen. Ook de diagnose is daardoor vaak complexer.

Ook de starttijden voor de items met een grote fout blijken in de TPSF langer zijn dan die voor de kleinere fouten (grote fouten: $M = 2.07, SD = 0.47$ vs. kleine fouten: $M = 1.48, SD = 0.54, F(1,59) = 36.39, p < .001$). Ook de interactie tussen foutgrootte en de spraakconditie is significant ($F(1,59) = 13.37, p < .01$). Grote fouten vergen dus iets meer initiële verwerkingstijd dan kleinere fouten, vooral als ze voorafgegaan niet worden voorgelezen.

Tabel 3: Gemiddelde waarden voor het effect van de foutgrootte in de TPSF

	Grote fouten (SR)		Kleine fouten (SR)		p
	Gemiddelde	SD	Gemiddelde	SD	
Succesratio (%)	92.50	16.14	84.17	19.51	*
Met spraak	92.50	22.22	85.00	26.52	-
Zonder spraak	92.50	22.22	83.34	23.77	*
Reactietijd (mil. sec)	372	255	303	161	*
Met spraak	332	344	270	172	-
Zonder spraak	423	309	345	233	**
Productietijd (sec)	20.48	6.40	15.61	4.39	**
Met spraak	20.83	7.97	15.32	5.05	**
Zonder spraak	20.12	8.22	15.89	5.25	**
Pauzetijd revisie (sec)	2.27	1.32	1.64	1.16	**
Direct	4.72	2.34	2.93	0.88	**
Uitgesteld	0.92	0.42	1.07	0.70	**

* significantieniveau $< .05$; ** significantieniveau $< .01$

Verder blijkt er geen verschil in de voorkeur om eerst de zin af te maken dan wel te corrigeren bij grote of kleine fouten. In beide gevallen geven de proefpersonen in vier van de vijf gevallen er de voorkeur aan om de tekstproductie voorrang te geven ($F(1,59) = 2.57, p = .115$). Er is wel een significant interactie-effect ($F(1,59) = 5.53, p < .05$), wat erop wijst dat het aanbieden van een gesproken TPSF het gedrag om eerst de zin af te maken nog versterkt. Dit wijst er waarschijnlijk op dat het aanbieden van de TPSF in gesproken vorm aanleiding

geeft tot een meer oppervlakkige, conceptuele verwerking van de geschreven tekst en dat een meer grondige lectuur van de gepresenteerde TPSF pas plaatsheeft nadat de zin volledig is afgemaakt.

In beide condities hebben de schrijvers meer tijd nodig om de zinnen met een grote fout correct af te werken dan die met een kleine fout ($F(1,59) = 115.95, p<.001$), hoewel het aantal te produceren karakters niet verschilde. De spraakconditie heeft geen invloed daarop.

De analyse van de gedetailleerde pauzetijden uit de logging van Inputlog stellen ons in staat dit beeld nog te verfijnen. De resultaten wijzen erop dat er significant langer gepauzeerd wordt voor een grotere fout dan wel voor een kleinere fout ($F(1,17) = .359, p<.01$). De spraakconditie beïnvloedt dit gedrag niet. Wel blijkt er een beduidend en versterkend verschil tussen het pauzegedrag bij directe en uitgestelde revisies (zie Tabel 3; interactieffect: $F(1,17) = .350, p<.01$). Bij directe correctie van de fout in de TPSF wordt de pauzetijd hoogstwaarschijnlijk verdeeld tussen het lezen van de TPSF, de detectie van de fout en mis-schien wordt in sommige gevallen ook al een mentale aanzet geformuleerd om de zin te vervullen (niet noodzakelijk in die volgorde). Dit verklaart waarschijnlijk de langere initiële pauzetijden. Bij uitgestelde revisie is de focus duidelijk op de foutdetectie, -diagnose en -correctie gericht. Bovendien blijkt uit het onderstaande voorbeeld (zie voorbeeld 1) dat in die gevallen de foutdetectie soms al gebeurd is tijdens de initiële starttijd, maar dat de correctie uitgesteld wordt. De schrijvers lijken in die gevallen voorrang te geven aan de tekstproductie, waarschijnlijk om het werkgeheugen niet te blijven belasten met de context-informatie die in het kortetermijngeheugen (KTG) opgeslagen is.

Voorbeeld van niet-directe revisie:

Lineaire representatie van Inputlog	Toelichting
Left Button[905,706] {2891} Movement[437,340] {563} het · t{750}hema · interesseert · mij. {531} Movement[286,348] {2156} Left Button[293,348] {1141} Movement[303,346] {703} BS BS BS fe{672} Movement[929,707] {1000} Left Button[928,706]	klik op OK startpauze van bijna 3 sec en positionering in productiezone vervullen van zin (pauze in 'thema') korte pauze voor muisbeweging naar in TPSF pauze van 2 sec en positionering van cursor pauze van 1 sec en kleine muisbeweging korte pauze en wijziging van 'eoneurentie' in 'conferentie' pauze van 1 sec en klik op OK

Dit voorbeeld geeft aan dat de detectie van de fout in de TPSF hoogstwaarschijnlijk al plaatsgevonden heeft tijdens de startpauze. Dit verklaart waarom een pauze van een halve seconde volstaat om de cursor in de buurt van de fout in de TPSF te positioneren. In andere voorbeelden is die pauzetijd zelfs kleiner dan 0.5 sec. De diagnose van de fout en het bepalen van een correctiestrategie (bijvoorbeeld hele woord schrappen, of toch voor letterssubstitutie kiezen) neemt beduidend meer tijd in beslag en verloopt gefaseerd.

In onze data zijn vier dominante profielen herkenbaar die de interactie met de deficiënte TPSF kenmerken. Op basis van onze observaties formuleren we de volgende interpretatie die we in replicatief vervolgonderzoek met oogbewegingsregistratie verder willen verifiëren:

- (1) korte initiële pauzes die leiden tot directe correctie: tijdens de initiële pauze richt de schrijver zich op de productie van de zin om de zin af te maken; pas daarna gaat de aandacht naar de correctie van eventuele fouten in de TPSF.
- (2) korte initiële pauzes die leiden tot uitgestelde correctie: tijdens de initiële pauze richt de schrijver zich op foutdetectie (evaluatief lezen); pas daarna gaat de aandacht naar het vervolledigen van de zin.
- (3) langere initiële pauzes die leiden tot directe correctie van de TPSF-fout: de schrijver richt zich op het vervolledigen van de zin tijdens de initiële pauze; in functie daarvan herleest hij de TPSF (oriënterend lezen) en wordt dan geconfronteerd/afgeleid door de fout, wat leidt tot correctie ervan.
- (4) langere initiële pauzes die leiden tot uitgestelde correctie: de schrijver doorloopt (waarschijnlijk) eerst een vergelijkbaar proces als in het vorige profiel, maar beslist dan om de voorrang te geven aan het vervolledigen van de zin en pas daarna de TPSF-fout te corrigeren; de detectie wordt opgeslagen in het kortetermijngeheugen (KTG) om daarna snel de fout te kunnen lokaliseren.

Dat die laatste strategie soms tot verlies van informatie kan leiden, toont het onderstaande voorbeeld misschien aan. (Door het ontbreken van oogbewegingen kunnen we natuurlijk niet met zekerheid vaststellen of de fout bij de start al gedetecteerd is.) Ook hier hebben we een relatief lange starttijd en geen directe correctie. Een aarzeling binnen een woord en een typefout, leiden de schrijver tijdens de productie af, wat leidt tot een lange (her?)oriëntering op de TPSF (zie voorbeeld 2)

Voorbeeld van uitgestelde revisie met extra foutcorrectie:

Lineaire representatie van Inputlog	Toelichting
Button[907,693] Movement[393,346] {3468} Left Button[393,345] {687}het · thema{796} · {907}inter{594}esser {640} BS et BS rt · mij {3281} · {859} Movement[297,348] {2515} Left Button[296,348] Left Button[297,348] {1094} conferentie {1531} Movement[916,702] Left Button[916,700]	klik op OK en directe beweging naar productiezone startpauze van 3.5 sec en bevestiging positionering vervolledigen van zin (pauze tussen en in woorden) korte pauze, dubbele correctie van typefout en afwerken zin pauze voor en na het punt en positionering in TPSF pauze van 2.5 selectie van TPSF-fout (volledige woord door dubbelklik) pauze van 1 sec en correctie van fout pauze van 1.5 sec klik op OK

Bovenstaande analyses tonen aan dat de nauwkeurige registratie van schrijfprocesdata een belangrijke meerwaarde kan vormen bij de analyse van schrijfprocessen, zowel in ecologische omgevingen als in experimentele contexten. Loggingdata bieden de onderzoeker immers de mogelijkheid om – in veel gevallen na een zeer gerichte selectie uit de veelheid van data – aspecten van het schrijfproces scherper in beeld te krijgen en zuiverder te interpreteren.

4. Toekomstplannen

Inputlog is een programma dat nog volop in ontwikkeling is. In de eerste plaats streven we ernaar om de huidige functionaliteit te stabiliseren en te optimaliseren. Zo zijn de gebruikte algoritmes voor de bepaling van woord- en zinsgrenzen nog erg beperkt en dus onvoldoende onnauwkeurig. Een extra complicatie daarbij is dat de algoritmes ook geldig moeten zijn voor procesdata. Moet je bijvoorbeeld een pauze voor een punt anders interpreteren dan een pauze na een punt? Uiteraard vergt de compatibiliteit met nieuwe versies van Windows en Office ook voortdurende aandacht.

Een andere invalshoek bij de verdere ontwikkeling is erop gericht de huidige functionaliteit verder uit te breiden. Daarbij gaat onze aandacht voornamelijk naar de volgende modules:

1. revisiemodule
2. progressiemodule
3. integratiemodule (Morae / Eyelink / I-Pad)

4.1 Revisiemodule. Revisie vormt een cruciale invalshoek voor de analyse van (cognitieve) schrijfprocessen. Een (semi-)automatische analyse van het revisieproces zou dan ook een belangrijk hulpmiddel kunnen vormen voor schrijfondzoekers. In de revisiemodule streven we ernaar om elke revisie op een aantal verschillende dimensies te beschrijven, o.a. tijd, operatie, niveau, locatie en positie t.o.v. de laatste schrijfactie ('point of inscription'). Een eerste aanzet voor een dergelijke analyse is ondertussen geprogrammeerd, maar de complexiteit van deze analyse leidt ertoe dat het erg tijdrovend is om tot een betrouwbaar en stabiel eindresultaat te komen.

Om revisies te detecteren hebben we een aantal algoritmes ontwikkeld in combinatie met regelsets. Aan de hand daarvan identificeert de revisieanalyse in eerste instantie zogenaamde *critical events* die kunnen wijzen op een revisie of een revisiecluster. Vervolgens evalueert de analyse elke mogelijke revisiekandidaat door het procesverloop direct erna te monitoren. Inputlog bevriest m.a.w. de situatie direct voor en direct na een revisie door muisbewegingen, selecties, cursorhandelingen en specifieke functies te isoleren en als afwijkend te interpreteren t.o.v. productiehandelingen die erop gericht zijn nieuwe tekst te produceren.

Onderstaand voorbeeld beschrijft bondig de werkwijze en geeft aan hoe verschillende technische operaties kunnen leiden tot een vergelijkbare revisie, i.c. de substitutie van een woord op het einde van een zin. Het woord 'gebracht' door 'geleid' (S-notatie¹²).

- (1) Toen de prinses uit haar bad kwam, werd ze naar de logeerkamer [gebracht]¹. |₁ {geleid}

Een eerste mogelijkheid is dat de schrijver de revisie uitvoert door de cursor een positie naar links te verplaatsen met de pijltjestoets en vervolgens het woord 'gebracht' verwijdert met de backspace-toets (8x). Door het vervangende woord 'geleid' te typen¹³ en de cursor weer naar rechts te plaatsen, sluit de schrijver de revisie af (zie regel 1, Figuur 3). Een andere mogelijkheid om deze substitutie uit te voeren is bijvoorbeeld het te vervangen woord dubbelklikken, de selectie verwijderen met de delete-toets en dan het nieuwe woord typen. De revisie kan in dat geval afgerond worden door de cursor met de muis te verplaatsen naar het einde van de tekst (zie regel 2, Figuur 3).

	begin (beweging)	selectie/positionering	deletie	einde (beweging)
1	-	cursorbeweging links	backspace	cursorbeweging rechts
2	muisbeweging links	dubbele muisklik links	delete	muisbeweging rechts
n	...	...	...	...

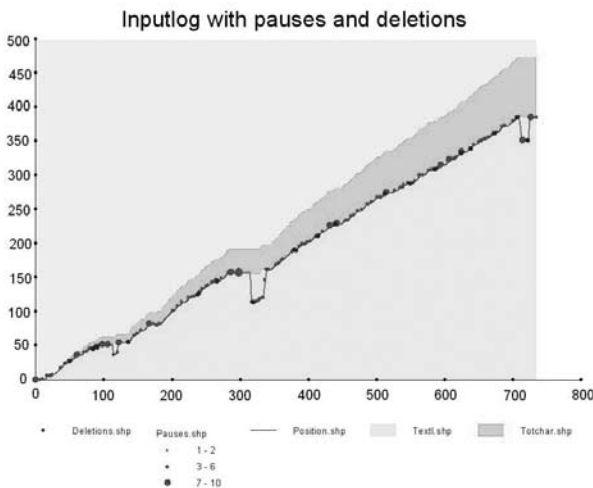
Figuur 3: Voorbeeld van revisieregels.

Uiteraard zijn er nog veel meer mogelijkheden om deze operatie uit te voeren. Het is onze bedoeling om de set van een vijftigtal regels die we nu voor deze operatie hebben opgebouwd, verder te testen en uit te breiden, ook voor andere revisietypes.

4.2 Progressie-analyse. Op dit moment representeert Inputlog de ontwikkeling van de tekst in een lineair tekstbestand. Om de tekstproductie ook visueel te kunnen voorstellen, willen we het programma uitbreiden met twee grafische voorstellingen van de tekstontwikkeling: een statische en een dynamische. In de statische progressie-representatie is het de bedoeling om het aantal geproduceerde karakters op een tijdlijn te plaatsen waarbij we uiteraard rekening houden met de tekens die in het verloop van het schrijfproces verwijderd worden. Die representatie kan de basis vormen voor een progressie-analyse zoals die bijvoorbeeld beschreven is door Perrin (2003; 2006).

De statische weergave zouden we willen aanvullen met een dynamische, interactieve weergave van de ontwikkeling van het schrijfproces. Daarbij laten we ons inspireren door het onderzoek van Lindgren (2006). Zij gebruikt Geografische Informatiesystemen (GIS) om het verloop van het schrijfproces te visualiseren en toegankelijker te maken voor verdere analyse. GIS stelt de onderzoeker in staat om via een grafische representatie van het schrijfproces te interageren met de onderliggende data. Door representatieve variabelen te selecteren (bijvoorbeeld langere pauzetijden) wordt een dynamische (kwalitatieve) analyse van schrijfprocesdata een heel stuk eenvoudiger en inzichtelijker.

Figuur 4 toont een voorbeeld van een GIS-representatie. De x-as vormt een tijdlijn; de y-as stelt het aantal karakters voor dat geproduceerd is (bovenste lijn), cq. overgebleven is in de



Figuur 4: Grafische representatie van schrijfdata in een GIS-applicatie.

tekst (onderste lijn). Als die laatste lijn een daling vertoont, wijst dit erop dat er tekst verwijderd werd. De stippellijn stelt de schrijf(in)activiteit voor: de grootte van de stippen geeft de lengte van de pauzes aan. Dit voorbeeld is opgebouwd op basis van een manueel aangepaste dataset uit Inputlog. Om een dergelijke figuur automatisch te genereren moet eerst de revisie-analyse afgewerkt worden, omdat het anders niet mogelijk is om de tekstevolutie correct te interpreteren.

4.3 Integratiemodule. De logdata van Inputlog kunnen op verschillende manieren gecomplementeerd worden. De toevoeging van spraakdata uit het dicteerprogramma Dragon Naturally Speaking is daar een voorbeeld van. Op dit moment verkennen we nog drie andere uitbreidingen:

1. *Morae*: een programma voor gebruikersonderzoek;
2. *Eyelink*: een programma voor de registratie van oogbewegingen;
3. *I-Pad*: een experimentele tekstverwerker.

Inputlog is in de eerste plaats een instrument dat micro-analytische analyses van schrijfprocessen ondersteunt. De combinatie met meer macro-analytische data opent echter nieuwe perspectieven. Daarbij denken we in de eerste plaats aan de integratie van logdata uit het programma *Morae* (Techsmith). Dit is een programma voor gebruikersonderzoek dat o.m. videodata genereert, online codering mogelijk maakt en ook schermwisselingen tussen programma's registreert (bijvoorbeeld url-adressen van websites die geraadpleegd werden). Omdat de logdata ook een tijdsaanduiding meekrijgen is integratie met Inputlog via een intermediair XML-formaat mogelijk. Voor de analyse van schrijfprocessen in meer geïntegreerde schrijfomgevingen biedt die extra informatie zeker mogelijkheden. We denken hierbij bijvoorbeeld aan de analyse van schrijfp opdrachten waarbij gebruik gemaakt wordt van verschillende documenten en/of websites. Omdat *Morae* bijvoorbeeld ook url-adressen bewaart, is het mogelijk de inputlogdata te verdelen over de verschillende toepassingen (cq. webpagina's) om zo een beter beeld te krijgen van het gebruik van verschillende bronnen.

Daarnaast verkennen we de mogelijkheden om de data van oogbewegingsregistraties te integreren met Inputlogdata (zie ook Cozijn en Lentz, Mak & Pander Maat in dit nummer). Het logbestand van *Eyelink* (SR Research) bijvoorbeeld biedt daartoe een aantal mogelijkheden. Een dergelijke integratie opent vooral nieuwe perspectieven voor onderzoek waarbij ook de interactie tussen lees- en schrijfprocessen een belangrijke component is. Een geïntegreerd XML-bestand zou immers de mogelijkheid bieden om de micro-analyses aan te vullen met gedetailleerde en geaggregeerde gegevens over de oogbewegingen. We denken daarbij vooral aan de beschrijving van bijvoorbeeld oogbewegingen en -fixaties tijdens schrijfpauzes. Dit zou ons in staat moeten stellen om het leesgedrag tijdens inactieve periodes beter te interpreteren. Als basis voor deze uitbreiding van het programma willen we de data gebruiken van de replicatie van het hierboven beschreven experiment. We hebben namelijk in een vergelijkbare setting het experiment opnieuw uitgevoerd en daarbij gebruikt gemaakt van een eyetracker¹⁴. In dit experiment werden de fouten in de 'reeds geproduceerde tekst' als kijkzones gemarkeerd om zo een betere interpretatiebasis te creëren voor de interactie van de schrijver met de TPSF-fouten in de verschillende fases van de tekstproductie.

Ten slotte is er ook een integratie gepland met loggingdata van een experimentele tekstverwerker, *I-Pad*. Dit is een eigen tekstverwerker met beperkte functies die ingezet kan worden voor schrijfonderzoek waar de ecologische aspecten van de schrijfomgeving – en

meer bepaald de tekstverwerker – niet belangrijk zijn. Het grote voordeel van een dergelijke integratie is dat het op die manier ook mogelijk is om voor elk getypt karakter nauwkeurige schermposities te bepalen, zowel absoluut (x/y-positie op het scherm) als relatief (aantal karakters t.o.v. het einde van de alinea of de tekst). Die informatie opent weer extra mogelijkheden voor aanvullende analyses waarbij het bijvoorbeeld van belang is om na te gaan in welke mate schrijvers hun tekst recursief opbouwen. Ook in combinatie met oogbewegingsregistraties biedt deze heel precieze benadering veel voordelen voor een nauwkeurige interpretatie van het leesgedrag.

5. Conclusie

Toetsregistratieprogramma's maken het mogelijk om zeer gedetailleerd schrijfprocessen te registreren, te registreren en te analyseren. De toepassingen van deze programma's bieden ruime mogelijkheden voor linguïstisch, tekstueel en cognitief georiënteerd schrijfprocesonderzoek. Maar ook ruimere toepassingen in het domein van schrijfontwikkeling, schrijfstoornissen of schrijfdidactiek behoren tot de mogelijkheden. In dit artikel hebben we de technische en functionele kenmerken van Inputlog beschreven. Het grote verschil met andere registratieprogramma's is dat Inputlog ontwikkeld is om schrijfprocessen in Word te loggen (Windowsomgeving) en dat niet alleen toets- en muisinput geregistreerd wordt, maar dat ook de spraakinput van dicteerprogramma's (Dragon Naturally Speaking) in de logging verwerkt wordt. Voor meer details verwijzen we graag naar de website van het programma: www.inputlog.net. Op die locatie wordt het programma ook beschikbaar gesteld voor onderzoekers.

De toepassing die we in dit artikel bespreken, maakt duidelijk wat de meerwaarde van schrijfprocesobservatie is voor de analyse van schrijfprocessen. De algemene analyse – en vooral de pauze-analyse in combinatie met de lineaire output – die Inputlog genereert, biedt een zeer bruikbare basis om ook meer theoretische onderzoeksvragen te beantwoorden. Zo hebben we in dit artikel voorbeelden gegeven van de verschillende strategieën die schrijvers ontwikkelen om te interageren met (deficiënte) representaties van de 'reeds geproduceerde tekst' (TPSF). De analyse van de data van het experimentele onderzoek geven onder meer aan dat de aard van de fouten in de TPSF bepalend kunnen zijn voor de manier waarop een schrijver in interactie treedt met de bestaande tekst. Uit de lineaire weergave van de gelogde data leiden we ook af dat het pauzegedrag bij het begin van de interactie met de TPSF en vóór de (uitgestelde) revisie een specifiek profiel vertoont. Die profielen willen we in verder onderzoek scherper omschrijven.

In vervolgonderzoek hopen we ook verslag te kunnen uitbrengen over de ontwikkelingen van nieuwe toepassingen en analyses in Inputlog. Vooral de revisie-, de progressie-analyse en de integratie met Eyelink (oogbewegingsregistratie), Morae (usability programma) en I-Pad (een experimentele tekstverwerker gekoppeld aan Inputlog) moeten extra mogelijkheden creëren voor een nog meer gedetailleerde registratie van schrijfprocesdata en een gerichtere analyse ervan.

- * Hierbij bedanken we Wesley Cabus, Ahmed Essahli, Bart van de Velde, Mathia van de Poel en Wim Claessens voor hun uitstekende programmeerwerk van Inputlog. Bart van de Velde schreef ook het Visual Basic.NET programma voor het TPSF-experiment.
- Graag danken we ook Thea van der Geest, Maaike van den Haak en Menno de Jong voor hun kritische opmerkingen bij een eerste versie van deze bijdrage.
1. Dit onderzoek is gefinancierd in het kader van de Bijzonder Onderzoeksfinitiering (BOF) van de Universiteit Antwerpen (project 2002-2006).
 2. De revisiemodule in Trace-It (Mac) maakt het mogelijk om een schrijffproces opnieuw af te spelen van revisie tot revisie, zowel voorwaarts als achterwaarts. Een dergelijke interface (en analysemodule) maakt het veel eenvoudiger om de precieze tekstontwikkeling en de rol van revisies daarin te analyseren.
 3. In juni 2006 heeft Noldus ook een toetsregistratieprogramma uitgebracht, uLog, als aanvulling op hun Observer software. Ook dit programma werkt in een Windowsomgeving.
 4. Inputlog 2.0 Beta is geprogrammeerd in Visual C++ en de interface is geprogrammeerd in Visual Basic.NET. De conversie van de IDF-bestanden naar (tussentiggende) XML-bestanden verloopt via parsing waarbij we gebruik maken van Flex en Bison. De snelheid en het geheugengebruik van het programma – en in het bijzonder die van de gebruikte *hooks* – is erg belangrijk omdat het residente registratieprogramma niet mag interfereren met het normale gebruik van de tekstverwerker. Voor het registreren en afspelen van het schrijffproces gebruikt Inputlog twee Windows-hooks: *Journalrecord* en *Journalplayback*. *Journalrecord* registreert iedere toetsaanslag, muisbeweging en -klik telkens in combinatie met de gerelateerde tijds aanduidingen. De *Journalplayback* ondersteunt de afspelmodule. Elk deel van Inputlog is geprogrammeerd als een verschillende klasse. Dit stelt ons in staat om gemakkelijk updates uit te voeren, huidige analyses te debuggen en om de functionaliteit van het programma verder uit te breiden.
 5. Op dit moment zijn nog een aantal aanvullende analyses in ontwikkeling (zie toekomstplannen).
 6. Deze logging van de spraakinput via Dragon Naturally Speaking kwam tot stand in nauwe samenwerking met Nuance (vroeger Scansoft). We danken in het bijzonder Guido Gallopyn, Stijn van Even, Neil Grant en Hans Geuns (Boston) voor hun grote bereidheid om samen te werken in dit project.
 7. Wie de afspelfunctie wil gebruiken als basis voor bijvoorbeeld een retrospectief protocol, raden we aan om de registratie via Inputlog te combineren met een online videoregistratie (bijvoorbeeld Camtasia van Techsmith; www.techsmith.com). Een combinatie van observatie-instrumenten levert in principe geen systeemconflicten op.
 8. Ook de programma's Camtasia en Morae van Techsmith (www.techsmith.com) kunnen zeer bruikbaar zijn voor dergelijke toepassingen. Deze programma's maken een online videoregistratie (in bijvoorbeeld AVI-formaat) door voortdurend schermfoto's te bewaren en die achter elkaar te plaatsen.
 9. Met de functie conditionele opmaak in Excel is het bijvoorbeeld mogelijk om die waarden te markeren in een bestand om ze sneller toegankelijk te maken voor codering.
 10. Voor zinnen die niet verbeterd werden, gaan we er m.a.w. vanuit dat de fout niet werd opgemerkt. Het gedrag van de proefpersoon in kwestie voor dergelijke zinnen is in die gevallen in feite vergelijkbaar met het leesgedrag van correcte zinnen. Uiteraard is het ook interessant om in post-hoonderzoek na te gaan welke fouten het makkelijkst over het hoofd worden gezien.
 11. We moeten er rekening mee houden dat deze analyse van de initiële cursorpositie slechts een indicatie geeft van het correctiegedrag. Een schrijver kan immers aanvankelijk beslissen om de cursor in het productiegedeelte te positioneren om zo een zin te vervolledigen, maar dan toch beslissen om de fout te corrigeren. Het feit dat het percentage voor de cursorpositie bij correcte zinnen niet gelijk is aan 100%, is waarschijnlijk te wijten aan een vergelijkbare ruisfactor in deze meting.
 12. Zie Kollberg (1998) voor een beschrijving van deze notatiemethode.
 13. Om de revisieregels zo eenvoudig mogelijk te houden, wordt nieuw gegenereerde tekst niet opgenomen in de regels.

14. Het experiment werd gecoördineerd door Thomas Quinlan, post-doconderzoeker aan de Universiteit Antwerpen. We danken de collega's van de Universiteit van Tilburg (Bedrijfscommunicatie en Digitale Media) voor de gastvrijheid om het onderzoek in hun laboratorium uit te voeren.

Bibliografie

- Degenhardt, M. (2005).** Camtasia and Catmovie: Two digital tools for observing, documenting and analysing writing processes of university students. In L. Van Waes, M. Leijten and C. Neuwirth (Eds.), *Writing and Digital Media* (pp. 180-186). Oxford: Elsevier.
- Goldman-Eisler, F. (1972).** Pauses, clauses, sentences. *Language and Speech*, 15, 103-113.
- Janssen, D., Waes, L. van, & Bergh, H. van den (1996).** Effects of thinking aloud on writing processes. In M. Levy & S. Ransdell (edit.), *The science of writing: Theories, methods, individual differences, and applications* (pp. 233-250). Mahwah, N.J., Lawrence Erlbaum.
- Kellogg, R.T. (1994).** *The psychology of writing*. New York: Oxford University Press.
- Kellogg, R.T. (1996).** A model of working memory in writing. In C.M. Levy & S. Ransdell (Eds.). *The science of writing: Theories, methods, individual differences and applications*. Hillsdale, NJ: Erlbaum.
- Kellogg, R.T. (2001a).** Competition for working memory among writing processes. *American Journal of Psychology*, 114, 175-191.
- Kellogg, R. T. (2001b).** Long-term working memory in text production. *Memory & Cognition*, 29, 43-52.
- Kollberg, P. (1998).** *S-Notation - a computer based method for studying and representing text composition*. Stockholm: Stockholm University.
- Leijten, M., & Waes, L. van (2003).** The writing processes and learning strategies of initial users of speech recognition: A case study on the adaptation process of two professional writers. Antwerp: Faculty of Applied Economics, University of Antwerp (Research paper, beschikbaar op <http://www.ua.ac.be/luuk.vanwaes/publications.html>).
- Leijten, M., & Waes, L. van (2005).** Writing with speech recognition: The adaptation process of professional writers with and without dictating experience. *Interacting with Computers*, 17(6), 736-772.
- Leijten, M., Ransdell, S., & Waes, L. van (in voorbereiding).** *Working Memory and Error Analysis in Writing: Strategy decisions based on error span, lexicality, and mode of writing*.
- Levy, C. M., & Ransdell, S. (2001).** Writing with concurrent memory loads. In T. Olive and C. M. Levy (Eds.) *Contemporary Tools and Techniques for Studying Writing*, (pp. 9-29). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Lindgren, E. (2006, to be published).** GIS for writing: applying geographic information system techniques to data-mine writing's cognitive processes. In D. Galbraith, M. Torrance and L. Van Waes, *Writing and Cognition*. Oxford: Elsevier.
- Lindgren, E., & Sullivan K.P.H. (2002).** The LS graph: A methodology for visualising writing revision. *Language Learning*, 52(3), 565-595.
- Lindgren, E., & Sullivan, K.P.H. (2006).** Key-stroke Logging: An Introduction. In K.P.H. Sullivan & E. Lindgren, *Computer Key-Stroke Logging and Writing: Methods and Applications* (pp. 1-10). Elsevier Science.
- Matsuhashi (1981).** Pausing and Planning: The Tempo of Written Discourse Production. *Research in the Teaching of English*, 15(2), 113-134.
- Olive, T. (2004).** Working memory in writing: Empirical evidences from the dual-task technique. *European Psychologist*, 9(1), 32-42
- Perrin, D. (2003).** Progression analysis: Investigating writing strategies at the workplace. *Journal of Pragmatics*, 35(6), 907-921.

- Perrin, D. (2006).** Progression Analysis: An Ethnographic, Computer-Based Multi- Method Approach to Investigate Natural Writing Processes. In L. Van Waes, M. Leijten and C. Neuwirth, *Writing and Digital Media* (pp. 173-179). Oxford: Elsevier Science.
- Severinsson Eklundh, K.S. (1994).** Linear and Non-linear strategies in computer-based writing. *Computers and Com-position*, 11, 203-216.
- Severinsson Eklundh, K.S., & Kollberg, P. (1992).** *Translating keystroke records into a general notation for the writing process*. Royal Institute of Technology, Stockholm: IPLab, Department of Numerical Analysis and Computing Science.
- Severinsson Eklundh, K.S., & Kollberg, P. (1996).** Computer tools for tracing the writing process: From keystroke records to S-notation. In G. Rijlaarsdam, H. van den Bergh & M. Couzijn, Theories, models and methodology in writing research, pp.526-541. Amsterdam: Amsterdam University Press.
- Severinsson Eklundh, K.S., & Kollberg, P. (2003).** Emerging discourse structure: Computer-assisted episode analysis as a window to global revision in university students' writing. *Journal of Pragmatics*, 35(4), 869-891.
- Spelman Miller, K. (2006).** Pausing, Productivity and the Processing of Topic in Online Writing. In K.P.H. Sullivan & E. Lindgren, *Computer Key-Stroke Logging and Writing: Methods and Applications*. Elsevier Science.
- Strömqvist, S., & Malmsten, L. (1997).** *ScriptLog Pro User's manual*. Göteborg University, Dept of Linguistics.
- Waes, L. van, & Leijten, M. (2006).** Logging writing processes with Inputlog. In L. Van Waes, M. Leijten & C. Neuwirth (Eds.) *Writing and Digital Media* (pp. 158-166). Oxford: Elsevier Science.
- Wengelin, A. (2006).** Examining Pauses in Writing: Theory, Methods and Empirical Data. In K.P.H. Sullivan & E. Lindgren, *Computer Key-Stroke Logging and Writing: Methods and Applications*. Oxford: Elsevier Science.

Het gebruik van oogbewegingen in leesonderzoek

1. Inleiding

Er is in de gedragswetenschap een toenemende aandacht voor oogbewegingsregistratie als methode voor het doen van onderzoek. Die methode leent zich bij uitstek om de cognitieve aspecten van informatieverwerking en -productie bij mensen in kaart te brengen. Oogbewegingsregistratie wordt bijvoorbeeld gebruikt in taalverwerkings- en taalproductieonderzoek (zie ook Lentz, Mak & Pander Maat in dit nummer), usability-onderzoek, mens-machine interactie en industrieel ontwerp (zie voor een overzicht Jakob & Karn, 2003). Andere toepassingen van de oogbewegingsregistratie zijn te vinden in de medische wetenschap (bv. diagnostiek en computeraansturing voor mindervaliden) en sinds kort ook in de computerspelindustrie (zie bv. Kenny *et al.*, 2005).

Bij taalbeheersers lijkt er nog een aarzeling te bestaan om deze techniek toe te passen, omdat hij complex en bewerkelijk is. Gelukkig is daar de laatste jaren veel in verbeterd zowel op hardware- als op softwaregebied. De verbetering van de techniek heeft niet alleen betrekking op de nauwkeurigheid en de hanteerbaarheid van de apparatuur maar ook op de analyseerbaarheid van de data. In dit artikel wordt ingegaan op het gebruik van de oogbewegingsregistratie in leesonderzoek en in het bijzonder op het analyseren van oogbewegingsdata. Hieronder volgen een schets van de oogbewegingspatronen tijdens het lezen en een bespreking van het probleem van het afleiden van betekenisvolle afhankelijkke maten van onderzoek (leestijden). De toepassing van de techniek en de analyse van oogbewegingen wordt uitgelegd aan de hand van een casus en het artikel sluit af met een korte discussie.

Samenvatting

Aan de hand van een leesonderzoek naar de invloed van vetgedrukte kernwoorden en opsommingen op de verwerking van korte verklarende teksten wordt ingegaan op het gebruik van oogbewegingsregistratie voor het doen van leesonderzoek. De aandacht gaat uit naar de wijze waarop betekenisvolle leestijden uit oogbewegingen kunnen worden afgeleid. In dat bestek wordt het programma *Fixation* besproken dat de onderzoeker in staat stelt op vrij eenvoudige wijze oogbewegingspatronen te analyseren en klaar te maken voor statistische analyse.

2. Oogbewegingen en lezen

De populariteit van oogbewegingsregistratie voor het doen van leesonderzoek schuilt in het gegeven dat oogbewegingen van nature voorkomen tijdens het lezen. De meting verstoort het normale leesgedrag niet en de deelnemers aan een leesexperiment hoeven geen

extra taak uit te voeren. Oogbewegingen worden al langere tijd gebruikt voor het doen van onderzoek naar taalprocessen (zie voor een overzicht Rayner, 1998; Rayner & Pollatsek, 1989). De eerste onderzoeken naar oogbewegingen waren gericht op het vaststellen van de specifieke oogbewegingspatronen die plaatsvinden tijdens het lezen. Uit dat onderzoek valt het volgende beeld te destilleren. Wanneer iemand een tekst leest, springen de ogen van één positie in de tekst naar de volgende. De gemiddelde lengte van zo'n sprong, ook wel *saccade* genoemd, is 8 lettertekens (met een ondergrens van 1 en een bovengrens van 15 lettertekens) die gemiddeld 30 milliseconden duurt. Een uitzondering vormt de sprong van het einde van een regel naar het begin van de volgende regel (return-sweep) die uiteraard langer is en langer duurt (60 tot 80 milliseconden). Tijdens een sprong nemen de ogen geen visuele informatie op (Matin, 1974). Visuele informatie wordt opgenomen in de periode van relatieve rust tussen twee sprongen in. Deze periode wordt een *fixatie* genoemd. Lezers hebben fixaties van gemiddeld 200 tot 250 milliseconden. Bij normale fontgrootte en leesafstand heeft de lezer tijdens een fixatie een blikveld van ongeveer 3 lettertekens naar links en 15 lettertekens naar rechts (de richting van dit blikveld is natuurlijk taalafhankelijk). In het centrum van het blikveld, de fovea, is de waarneming het scherpst. Dit gebied van grote scherpste reikt meestal niet verder dan 7 lettertekens naar rechts. Lezers lezen niet voortdurend vooruit. Gemiddeld genomen springen lezers in 15% van de gevallen terug in de tekst. Zo'n sprong terug wordt ook wel *regressie* genoemd.

Na de beginperiode waarin is komen vast te staan welke fysieke factoren de visuele waarneming van tekst beïnvloeden, is de aandacht in het oogbewegingsonderzoek verschoven naar de cognitieve factoren die het kijkpatroon sturen. Aan de cognitieve interpretatie van oogbewegingen liggen twee uitgangspunten ten grondslag: de *immediacy assumption* en de *eye-mind assumption* (Just & Carpenter, 1980). De eerste assumptie stelt dat alle verwerkingsprocessen die worden uitgevoerd op een woord starten zodra het woord wordt bekeken. De tweede geeft aan dat de ogen op een woord gericht blijven zolang de verwerking van het woord duurt. Ofschoon voor beide assumpties evidentie is gevonden, zijn ze toch niet sluitend. Zo is uit onderzoek gebleken dat niet alle woorden in een tekst worden bekeken. Korte, hoogfrequente en/of zeer voorspelbare woorden, zoals functiewoorden, worden vaak overgeslagen. Dat wil niet zeggen dat ze niet worden verwerkt. Er is aangetoond dat ze worden waargenomen en verwerkt gedurende het kijken naar de woorden die eraan voorafgaan. Dit betekent dat de kijktijd op een woord niet noodzakelijkerwijs een afspiegeling is van de verwerking van alleen het bekeken woord. Een tweede probleem is dat de verwerking van een woord soms doorgaat, terwijl de ogen al verder springen in de tekst. Dit is bijvoorbeeld aangetoond in een studie naar de verwerking van laagfrequente woorden (zie Rayner & Duffy, 1986).

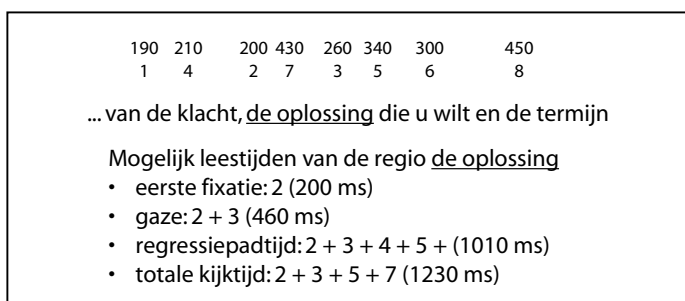
Ondanks deze beperkingen heeft de studie van oogbewegingen veel informatie opgeleverd over de verwerking van geschreven taal en het onderzoek naar de invloed van lexicale, syntactische, semantische en tekstuele factoren op de verwerking van tekst op een hoger plan getild. De methode is met name zo geschikt omdat ze inzicht verschaft in de temporele eigenschappen van het leesproces. De ogen verraden niet alleen hoe lang de verwerking van een stuk tekst duurt maar ook wanneer het voor het eerst wordt waargenomen en wanneer er sprake is van een herlezing. Zoals hierboven gesteld, springen lezers regelmatig terug in een tekst en herlezen ze tekstgedeeltes die ze in een eerder stadium al hadden verwerkt. Deze herlezingen kunnen indicatief zijn voor leesproblemen, maar dat hoeft niet. Indien lezers op een probleem in de tekst stuiten, kunnen ze in principe drie dingen doen: ze pau-

zeren op de plaats waar het probleem optreedt, ze springen terug in de tekst of ze lezen door en hopen verderop in de tekst het probleem op te lossen. Lukt dat laatste niet, dan wordt de gehele zin veelal nog eens gelezen (Frazier & Rayner, 1982; Liversedge *et al.*, 1998).

Regressies en fixatietijden zijn dus indicatief voor de cognitieve processen die zich afspelen tijdens het lezen. Waar de interpretatie van regressies redelijk rechttoe rechtaan is, is de interpretatie van fixatietijden min of meer problematisch. Want hoe moet de leestijd van een woord bepaald worden als het meerdere malen is gefixeerd, of hoe moet de leestijd bepaald worden van grotere teksteenheden zoals een constituent, een deelzin, een zin, of een alinea? En wat als de onderzoeker geïnteresseerd is in vroeg optredende talige processen, zoals lexicale toegang of woordfrequentie-effecten, en wat als de interesse uitgaat naar laat optredende processen, zoals zins- of tekstintegratie? Het mag wel duidelijk zijn dat deze vraag niet eenvoudig te beantwoorden is. De vaststelling van geschikte maten voor leestijden is dan ook lang een belangrijk thema geweest in het leesonderzoek waarbij oogbewegingen geregistreerd worden.

3. Het afleiden van leestijden

Zoals hierboven is uiteengezet, bestaan oogbewegingen uit fixaties en saccades. Ofschoon fixaties aangeven wat er verwerkt wordt en fixatietijden iets zeggen over hoe lang iets verwerkt wordt, zijn ze toch niet te gebruiken als leestijden in leesonderzoek. Woorden kunnen namelijk meerdere malen worden gefixeerd en soms is de onderzoeker geïnteresseerd in de verwerking van een regio die bestaat uit meerdere woorden waarop vaak meerdere malen gefixeerd wordt. Enkelvoudige fixaties schieten dan tekort om de verwerkingstijd weer te geven. Dat het bepalen van een leestijd op basis van fixaties niet triviaal is, kan aan de hand van een voorbeeld geïllustreerd worden. Figuur 1 geeft een hypothetisch (maar wel realistisch) fixatiepatroon weer boven een kort stukje tekst. De fixaties zijn in kijkvolgorde genummerd en boven de woorden geplaatst. Boven elke fixatie staat de tijd in milliseconden. De onderzoeker is geïnteresseerd in de verwerking van de regio 'de oplossing' (in het voorbeeld onderstreept).



Figuur 1: Voorbeeld van acht fixaties (in de volgorde 1 t/m 8) met daarboven hun tijden (in milliseconden) op een tekstgedeelte met daaronder enkele mogelijke leestijden van een tekstregio ('de oplossing', onderstreept).

Het kijkpatroon in het voorbeeld verloopt van fixatie 1 ('de') naar 2 ('de oplossing') naar 3 ('de oplossing') naar 4 ('klacht'), etc. De vraag is nu: wat is de leestijd van 'de oplossing' in

het voorbeeld? Indien de onderzoeker geïnteresseerd is in zeer vroege verwerkingsprocessen (bv. lexicaal), dan zou volstaan kunnen worden met de duur van de eerste fixatie op de regio (fixatie 2). De duur van eerste verwerking zou daarmee op 200 milliseconden komen. Deze fixatie wordt echter gevolgd door nog een fixatie op dezelfde regio (fixatie 3), zodat men ook deze tot de eerste verwerking zou kunnen rekenen. De duur van de eerste verwerking is dan 460 milliseconden. Deze manier om verwerkingstijden te berekenen is als eerste gebruikt door Just en Carpenter (1980). Zij gaven daaraan de naam *gaze*, die moeilijk in het Nederlands te vertalen is. De *gaze* is gedefinieerd als de som van alle fixatietijden op een regio tot de regio in voor- of achterwaartse richting wordt verlaten.

Indien de onderzoeker echter geïnteresseerd is in latere verwerkingsprocessen, (syntactisch/semantisch/tekstueel) zouden wellicht fixaties 4 en 5 ook meegeteld kunnen worden. De redenering om fixaties mee te tellen die deel uitmaken van een regressie vanaf een regio, is gebaseerd op de veronderstelling dat de lezer kennelijk op grond van de verwerking van de regio besloot om terug te springen in de tekst. De verwerking van de regio is dus nog niet af en processen die plaatsvinden tijdens de regressie zouden meegeteld moeten worden in een maat van eerste verwerking. Volgens deze redenering is de eerste verwerking van een regio pas voltooid als hij in voorwaartse richting wordt verlaten en de lezer dus *nieuwe* informatie tot zich neemt (bij fixatie 6). Deze definitie van de leestijd van de regio resulteert in een leestijd van 1010 milliseconden. Weer een andere benadering van de leestijd is om niet de regressies mee te rekenen in de leestijden van de regio's waarvandaan ze starten, maar om simpelweg alle fixatietijden op een regio op te tellen, ongeacht waar ze vandaan komen. Volgens deze visie zouden fixaties 2, 3, 5 en 7 moeten worden opgeteld om de totale verwerkingstijd van de regio te verkrijgen. De leestijd is dan 1230 milliseconden.

Hopelijk maakt dit voorbeeld niet alleen duidelijk dat de berekening van leestijden op basis van fixaties complex is, maar ook dat het bepalen van een afhankelijke maat van verwerking een beredeneerde keuze van de onderzoeker moet zijn. De keuze van de afhankelijke maat leidt in het hier gegeven voorbeeld tot leestijden van eenzelfde regio die lopen van 200 tot 1230 milliseconden, hetgeen natuurlijk een enorm verschil is. Het probleem van de leestijdenberekening heeft in de afgelopen decennia niet alleen geleid tot een veelvoud aan afhankelijke maten maar ook tot veel verwarring, omdat dezelfde maten terugkeerden in de literatuur onder andere (vaak zelfverzonnen) namen (Murray, 2000). Onder onderzoekers die zich bezighouden met leesonderzoek is er nu enigszins consensus bereikt over de volgende maten: *first-fixation duration*, *gaze*, *total-pass reading time* en *total reading time* (respectievelijk *eerste fixatie*, *gaze*, *regressiepadtijd* en *totale kijktijd* in Figuur 1). Voorgesteld is om in oogbewegingsleesonderzoek in ieder geval de regressiepadtijden en het aantal regressies te rapporteren (zie voor een uitgebreide discussie Cozijn *et al.*, 2003).

Zoals gesteld in paragraaf 2 nemen de ogen tijdens een saccade geen visuele informatie op. Dit verschijnsel heeft ertoe geleid dat leesonderzoekers de duur van saccades buiten beschouwing lieten wanneer ze leestijden berekenden. Dit is onjuist. Dat er tijdens saccades geen nieuwe informatie wordt opgenomen betekent nog niet dat er geen informatie wordt verwerkt. Irwin (1998) heeft aangetoond dat weliswaar sommige visuele cognitieve processen worden gehinderd door een saccade, maar dat taalverwerking ongehinderd doorloopt. Daarom moeten, anders dan in het bovenstaande voorbeeld wordt gesuggereerd, saccadetijden meegeteld worden in de berekening van afhankelijke maten van taalverwerking. Het effect van het meetellen van saccadetijden is uiteraard klein maar telt op bij het effect van de eraan gerelateerde fixatietijden (Vonk & Cozijn, 2003).

De complexiteit van het berekenen van leestijden op basis van oogbewegingen zou menig onderzoeker afschrikken als er inmiddels niet makkelijk te gebruiken programma's zouden zijn om het werk te vereenvoudigen. Aan de hand van een leesonderzoek als casus zal het gebruik van één zo'n programma, namelijk het onder Microsoft Windows werkende programma *Fixation*¹, worden besproken.

4. De casus

De toepassing van het programma om oogbewegingen over tekst te analyseren, *Fixation*, wordt geïllustreerd aan de hand van een leesonderzoek naar de invloed van layout-kenmerken op het begrijpen van tekst. Na een korte inleiding op het onderzoek volgt een wat uitgebreidere beschrijving van de methode, gevolgd door de resultaten en de conclusie. Voor de doeleinden van dit artikel wordt slechts een klein gedeelte van de resultaten besproken.

In de usability-literatuur (zie Nielsen, 2000 e.a.) en de typografie-literatuur (zie Waller, 1980 e.a.) wordt geadviseerd om belangrijke termen in de tekst vet te maken en om tekstuele informatie in puntsgewijze opsommingen te plaatsen. De idee is dat deze layout-kenmerken fungeren als aandachttrekkers en dat ze de begrijpelijkheid en de leesbaarheid van teksten bevorderen. Naar de invloed van vetgedrukte woorden en opsommingen op het leesproces zelf is weinig onderzoek verricht. In een oogbewegingsexperiment is de invloed onderzocht van vetgedrukte kernwoorden en van puntsgewijze opsommingen op de verwerking van twintig korte, verklarende teksten over consumentenrecht. De verwachtingen waren dat het vetdrukken van kernwoorden en het plaatsen van informatie in een puntsgewijze opsomming zouden leiden tot diepere verwerking en dus tot langere leestijden (Reynolds & Shirey, 1988).

4.1 Materiaal. Als materiaal voor deze studie zijn 20 teksten gebruikt over consumentenrecht afkomstig van het Ministerie van Economische Zaken. De oorspronkelijke teksten, die waren geschreven om op een website te plaatsen², bevatten zelf al opsommingen en vetgedrukte woorden. Om vast te stellen of de juiste woorden vetgedrukt waren is een vooronderzoek onder 72 proefpersonen uitgevoerd waarin is vastgesteld welke woorden als kernwoorden van de teksten werden aangemerkt. De proefpersonen konden maximaal zeven begrippen aanmerken. Van deze zeven begrippen werden de vijf begrippen genomen die door negen of meer proefpersonen waren geselecteerd. Deze procedure resulteerde in vijf kernwoorden per tekst die door gemiddeld 23% van de proefpersonen als belangrijk waren aangemerkt.

De opmaak van de teksten werd voor het experiment aangepast door ze te voorzien van een titel, een inleiding, een opsomming en een uitleiding. De vijf vetgedrukte kernwoorden konden voorkomen in de opsomming en de uitleiding. Van de teksten werden vier versies gemaakt door de aan- of afwezigheid van een puntsgewijze opsomming en de aan- of afwezigheid van vetgedrukte kernwoorden te variëren. Wanneer de opsomming niet puntsgewijs werd aangeboden, waren de leden van de opsomming aaneengeschreven met behulp van komma's en het voegwoord *en*. Een voorbeeld van een tekst met een puntsgewijze opsomming en vetgedrukte kernwoorden is gegeven in Figuur 2.



Figuur 2: Voorbeeld van een experimentele tekst in de versie met een puntsgewijze opsomming en vetgedrukte kernwoorden (afbeelding van 800 x 600 pixels).

De teksten werden in het leesexperiment aangeboden met behulp van de bij de eyetracker meegeleverde software. Voor deze software was het noodzakelijk de teksten om te zetten in afbeeldingen (grootte: 800 x 600 pixels). Omdat het meten van oogbewegingen de betrouwbaarste resultaten geeft in het midden van het scherm, waren de teksten centraal op het scherm geplaatst met een linkermarge van 90 en een topmarge van 95 pixels. De teksten hadden een lichte achtergrond en waren opgemaakt als webteksten in het lettertype Verdana met puntgrootte 10. Het omzetten van de teksten in plaatjes gebeurde met behulp van het programma TSG³. Dit programma maakte van elke tekst tevens een object-bestand, waarin voor elk woord van de tekst de grafische coördinaten zijn opgeslagen. Dit bestand is nodig voor de analyse van de oogbewegingen.

4.2 Instrumentatie. Naast de leestijden van de vetgedrukte woorden en de opsommingen werd het begrip van een tekst gemeten door middel van beweringen. Na het lezen van een tekst werd een bewering over de tekst aangeboden die beoordeeld moest worden op zijn juistheid. Voor elke tekst werden twee beweringen gemaakt, een juiste en een onjuiste. De juiste bewering over de voorbeeldtekst van Figuur 2 luidde: *Een klacht kunt u het beste schriftelijk verwoorden*. De onjuiste versie luidde: *Door het schrijven van een klachtenbrief maakt u de verkoper duidelijk dat u een discussie aan wilt gaan*. Ook van de beweringen werden plaatjes gemaakt met een afmeting van 800 bij 600 pixels. Ze hadden dezelfde achtergrond, hetzelfde lettertype en dezelfde linkermarge als de teksten maar hadden een topmarge van 390 pixels, ongeveer ter hoogte van de plaats waar de laatste regel van de teksten stond. Deze hoogte is gekozen om ervoor te zorgen dat proefpersonen terstond na het lezen van de laatste regel van een tekst verder konden gaan met het beoordelen van de bewering.

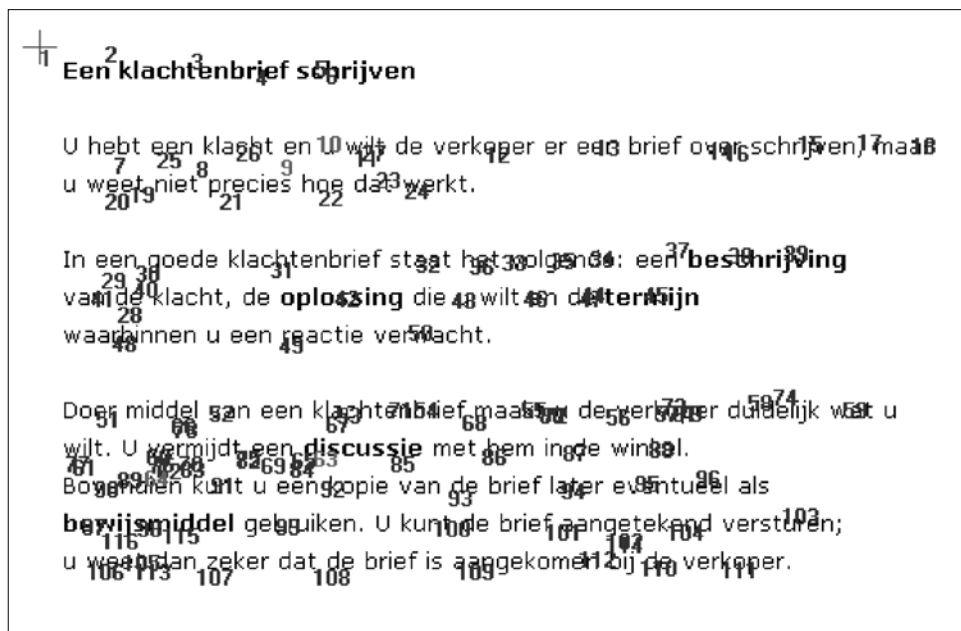
4.3 Design. De kruising van de variabelen Vet en Opsomming leverde vier condities op: vet/puntsgewijze opsomming, vet/niet-puntsgewijze opsomming, niet-vet/puntsgewijze opsomming en niet-vet/niet-puntsgewijze opsomming. De vier condities werden semi-ran-

dom verdeeld over de 20 teksten, hetgeen resulteerde in vijf teksten per conditie. Omdat een tekst slechts eenmaal aan een proefpersoon kon worden aangeboden, werden er vier lijsten van 20 teksten gemaakt, waarbij elke tekst op elke lijst in een andere conditie stond. Elke proefpersoon kreeg willekeurig een lijst toegewezen. Ook de verdeling van juiste en onjuiste beweringen was semi-random. Elke proefpersoon zag 10 teksten met een juiste en 10 teksten met een onjuiste bewering.

4.4 Proefpersonen. Aan het experiment namen 33 studenten van de Universiteit van Tilburg deel, 16 vrouwen en 17 mannen met een gemiddelde leeftijd van 22,4 jaar. De proefpersonen droegen geen bril of contactlenzen (brillen en contactlenzen kunnen de betrouwbaarheid van de meting negatief beïnvloeden).

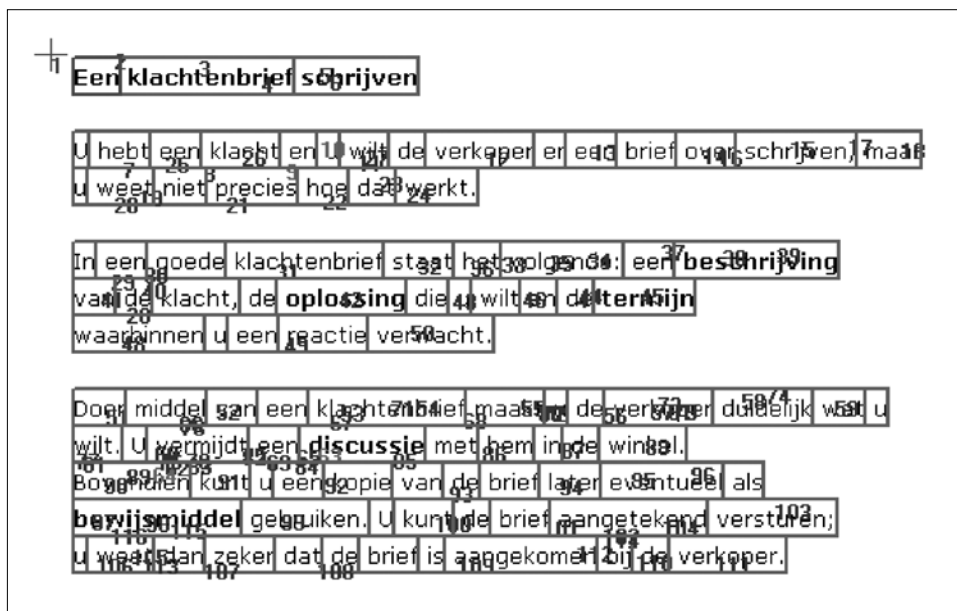
4.5 Apparatuur. De oogbewegingen werden in een onderzoekslaboratorium geregistreerd met behulp van een SR Research Eyelink II eyetracker⁴. De software behorende bij de eyetracker (een in C geschreven bibliotheek van routines) werd gebruikt om de teksten aan te bieden. De teksten werden getoond op een 19-inch monitor met een schermresolutie van 800 bij 600 pixels. Proefpersonen zaten op ongeveer 70 cm van het beeldscherm. Er werd gemeten op het rechteroog (de meting op slechts één oog bleek voldoende betrouwbaar) met een frequentie van 250 Hz (de standaard snelheid van de eyetracker).

4.6 Procedure. De proefpersonen werd verteld dat het onderzoek ging over hoe mensen teksten lezen. Ze waren niet op de hoogte van het eigenlijke doel van het experiment. Na het instellen van de apparatuur en het kalibreren van de proefpersoon (kalibreren is het tot stand brengen van een correspondentie tussen de metingen op het oog en de kijkposities op het scherm; zie Lentz, Mak en Pander Maat in dit nummer voor een uitgebreide beschrijving) werd een oefenblok bestaande uit drie teksten getoond. Het eigenlijke experiment startte na het oefenblok. De procedure van de aanbidding van een tekst was als volgt. Voorafgaand aan een tekst werd een scherm getoond met een sterretje. Het sterretje stond precies op de plaats waar de titel van een tekst begon. De proefpersoon moest strak naar het sterretje kijken en op de enter-toets drukken. Het sterretje verdween dan en op de plaats van het sterretje verscheen een tekst. De instructie luidde om de tekst goed en snel door te lezen en wanneer hij begrepen was meteen op de enter-toets te drukken. De tekst verdween en een bewering over de tekst verscheen op de plaats waar de tekst geëindigd was. De opdracht was om de bewering te beoordelen op zijn juistheid op basis van de tekst die net was gelezen. Het oordeel kon worden gegeven door op de 'z'-toets (onwaar) of de '/'-toets (waar) te drukken. Na het drukken verdween de bewering van het scherm en verscheen er een kalibratiepunt midden op het scherm. Dit punt diende om een eventuele lineaire verschuiving in de meting te corrigeren. Na het kalibratiepunt verscheen weer het sterretje op het scherm en herhaalde de procedure zich. Het experiment duurde ongeveer 30 minuten.



Figuur 3: Voorbeeld van een experimentele tekst met daarover geplaatst de genummerde fixaties van een proefpersoon (het kruisje bij fixatie nummer 1 geeft de positie van de cursor in het programma weer).

4.7 Verwerking van de gegevens. De oogbewegingsgegevens zijn geanalyseerd met behulp van het programma *Fixation*. Het programma leest de oogbewegingsdata van een proefpersoon in ascii-formaat⁵ in en plaatst de fixaties over de plaatjes die zijn aangeboden in het experiment (zie Figuur 3). Om de fixaties te kunnen koppelen aan de woorden van de tekst heeft het programma de grafische coördinaten van de woorden nodig. Zoals eerder gezegd zijn deze coördinaten opgeslagen in zogenaamde object-bestanden tijdens de vervaardiging van het materiaal. De objecten zijn zo gecodeerd dat ze een eventuele spatie voorafgaand aan een woord bij het woord trekken. De reden daarvoor is dat lezers van links naar rechts lezen en een blikveld hebben dat naar rechts groter is dan naar links (zie paragraaf 2). Wanneer dus een fixatie op een spatie vóór een woord valt, kan ervan uitgegaan worden dat het woord rechts van de spatie gelezen wordt. Het programma koppelt automatisch alle fixaties aan de objecten waarbinnen ze vallen, waarbij de linker- en bovengrens tot een object gerekend worden maar de rechter- en de ondergrens niet. Voor het opnemen van de linkergrens geldt dezelfde redering als voor het opnemen van spaties. Dat de bovengrens tot het woord eronder gerekend wordt, komt omdat lezers vaker boven de regel lezen dan eronder. De automatische koppeling van fixaties aan objecten bespaart de onderzoeker weliswaar veel werk, maar toch zullen alle koppelingen handmatig gecontroleerd moeten worden. Het programma *Fixation* helpt de onderzoeker ook daarbij. Met behulp van de muis kunnen verkeerd of niet gekoppelde fixaties eenvoudig gecorrigeerd worden door ze op te pakken en naar het juiste object te slepen. Een voorbeeld van een niet-gekoppelde fixatie is fixatie nummer 1 in Figuur 4. Deze fixatie valt duidelijk binnen geen enkel object. Met behulp van het programma is de fixatie handmatig gekoppeld aan het eerste object op de pagina, namelijk het object waarbinnen het woord 'Een' valt.



Figuur 4: Voorbeeld van een experimentele tekst met daarover geplaatst de genummerde fixaties van een proefpersoon en de grafische objectgegevens van de woorden van de tekst. Fixatie nummer 1 die ogenschijnlijk niet in een object valt, is handmatig toegekend aan het eerste object op de pagina dat codeert voor het woord 'Een'.

Op dezelfde wijze zijn alle teksten van alle proefpersonen gecontroleerd en gecorrigeerd. Het programma *Fixation* laat onmiddellijk zien of een meting op een tekst gelukt of mislukt is en staat op eenvoudige wijze toe fixatietoewijzingen te veranderen of zelfs fixaties van analyse uit te sluiten.

Zoals in paragraaf 3 is gesteld, moeten uit de fixaties leestijden worden afgeleid. Daarvoor is het noodzakelijk dat binnen de teksten de regio's zijn gedefinieerd waarover de leestijden moeten worden berekend. In dit experiment waren die regio's de (al dan niet vetgedrukte) kernwoorden, de opsomming en de uitleiding. Het coderen van regio's is gedaan in de object-bestanden door de objecten die tot dezelfde regio behoorden hetzelfde codenummer te geven. De codering in regio's wordt voor alle teksten op gelijke wijze gedaan, zodat bijvoorbeeld regio 2 in alle teksten codeert voor het eerste (al dan niet) vetgedrukte kernwoord. Met behulp van het programma zijn vervolgens per regio de volgende leestijden gegenereerd: *eerste fixatie*, *gaze*, *regressiepadtijd* en *totale kijktijd*. Deze tijden zijn berekend met inbegrip van de duur van de saccades (zie paragraaf 3 voor een uitleg). Bij de leestijden is aanvullende informatie opgeslagen over het aantal fixaties waaruit een leestijd is opgebouwd, of de leestijd wordt afgesloten met een sprong vooruit of een sprong terug in de tekst, of er in de leestijd een afgewezen fixatie zit en of een leestijd in het begin, het midden of aan het einde een knippering bevat. Met 'knippering' wordt bedoeld dat de lezer met de ogen knippert. Tijdens een knippering is er per definitie geen geldige oogmeting (wanneer de ogen dicht zijn, wordt er niets geregistreerd). Dat wil echter niet zeggen dat alle knipperingen van de analyse moeten worden uitgesloten. Wanneer een lezer tijdens het kijken naar een en dezelfde regio knippert, kan de tijd van de knippering meegenomen worden in de bereke-

ning van de leestijd. Wanneer echter een knippering plaatsvindt aan het begin of het einde van het kijken naar een regio, is de leestijd van deze regio in principe ongeldig, want dan is niet met zekerheid te zeggen wanneer het lezen van deze regio startte respectievelijk eindigde. De leestijden zijn weggeschreven in ascii-formaat en ingelezen in SPSS voor statistische analyse.

4.8 Resultaten. Binnen de context van dit artikel is ervoor gekozen om niet alle resultaten van het onderzoek maar slechts een gedeelte te laten zien. Hieronder staan de resultaten van de leestijden van alleen die kernwoorden die in de opsomming stonden (40 kernwoorden in 20 teksten).

In het onderzoek ging de belangstelling uit naar de invloed van vetgedruktheid op de verwerking van kernwoorden. De verwachting was dat een kernwoord meer aandacht zou krijgen en dieper verwerkt zou worden als het vetgedrukt is dan wanneer het niet vetgedrukt is. Twee afhankelijke maten die de diepere verwerking weerspiegelen zijn de *gaze* en de *regressiepadtijd* (zie paragraaf 3). Het verschil tussen de twee maten is dat de *regressiepadtijd* in tegenstelling tot de *gaze* rekening houdt met regressies. Ter illustratie worden hieronder de resultaten van deze twee maten besproken en bediscussieerd.

In oogbewegingsonderzoek is er altijd redelijk veel uitval van data. Daarvoor zijn drie oorzaken aan te wijzen. In de eerste plaats slaan lezers soms regio's in een tekst over en zijn er voor die regio's dus geen metingen. In de tweede plaats vallen er metingen af, omdat lezers met hun ogen knipperen (zie paragraaf 4.7). En in de derde plaats vallen metingen met extreme waarden af, zoals dat in tijdmetingsonderzoek gebruikelijk is. Het is daarom belangrijk om de data op te schonen alvorens over te gaan tot statistische toetsing.

In totaal waren er 1320 metingen mogelijk (33 proefpersonen maal 40 kernwoorden). In 15,9% van de gevallen was een kernwoord overgeslagen. Voor de gazetijden gold dat in 3,3% van de gevallen de gazetijd begon of eindigde met een knippering en in 0,9% van de gevallen dat de tijd drie of meer standaarddeviaties afweek van het proefpersonen- en itemgemiddelde per conditie. Bij de regressiepadtijden begon of eindigde 3,2% van de gevallen met een knippering en viel 2,0% uit vanwege extreme waarden. De resterende gaze- en regressiepadtijden werden onderworpen aan univariate variantieanalyses over proefpersonen (F1) en over teksten (F2) met de factoren Vet (wel/niet vetgedrukt) en Opsomming (wel/niet puntsgewijs). De gemiddelde gaze- en regressiepadtijden en de bijbehorende percentages regressies staan in Tabel 1.

In de analyses van de gazetijden werden geen effecten gevonden: alle F 's waren kleiner dan 1. De analyses van de regressiepadtijden lieten echter een interactie zien tussen Vet en Opsomming: $F1(1,116) = 7,19$; $MSE = 132140$; $p < 0,01$; $\eta^2 = 0,058$ en $F2(1,40) = 5,17$; $MSE = 69718$; $p < 0,05$; $\eta^2 = 0,114$. Wanneer er een puntsgewijze opsomming was, waren de regressiepadtijden van de kernwoorden langer als ze vetgedrukt waren dan wanneer ze niet vetgedrukt waren, maar als er geen puntsgewijze opsomming was, was het precies andersom: de regressiepadtijden waren korter als de kernwoorden vetgedrukt waren dan wanneer ze niet vetgedrukt waren. Er waren verder geen effecten van Vet en Opsomming in deze analyses.

Tabel 1: De gemiddelde gaze- en regressiepadtijden (ms) en het aantal regressies (percentages) van de kernwoorden in de opsomming als functie van Vet (wel/niet vetgedrukt) en Opsomming (wel/niet puntsgewijs).

		kernwoorden in de opsomming	
	opsomming	vet	niet vet
gazetijden	puntsgewijs	319	326
	niet puntsgewijs	310	320
regressiepadtijden	puntsgewijs	574	471
	niet puntsgewijs	409	645
regressies	puntsgewijs	29,0%	19,5%
	niet puntsgewijs	27,7%	25,5%

4.9 Discussie. De interactie tussen Vet en Opsomming in de analyses van de regressiepadtijden suggereert dat de twee layout-factoren elkaar beïnvloeden en dus niet onafhankelijk zijn. Een mogelijke interpretatie is dat als vetgedrukte woorden en puntsgewijze opsommingen tegelijkertijd en op dezelfde plaats in de tekst gebruikt worden, ze elkaar storen. Een volledige bespreking van deze interactie valt echter buiten het bestek van dit artikel.

Een interessante en hier meer relevante bevinding is dat de gazetijden geen interactie lieten zien. Een mogelijke verklaring daarvoor schuilt in het relatief hoge aantal regressies en de verdeling van de regressies over de verschillende condities. Als er veel regressies zijn vanuit een regio, zijn de gazetijden van die regio korter. De reden daarvoor is dat als lezers besluiten om terug te springen in de tekst, ze dat besluit vaak al heel snel bij de eerste fixatie maken (Vonk & Cozijn, 2003). De gazetijden geven dan een onderschatting van de werkelijke verwerkingstijd, gegeven dat de tijd die doorgebracht wordt tijdens de regressie bijdraagt aan de verwerkingstijd van de regio. Het relatief hoge aantal regressies (gemiddeld 25,4% tegen 15% bij normaal lezen, zie paragraaf 2) maakt de gazetijden als maat voor verwerking daarom minder betrouwbaar. De scheve verdeling van regressies in Tabel 1 doet verder vermoeden dat de onderschatting van de gazetijden groter is in de conditie puntsgewijs/vet ten opzichte van niet-puntsgewijs/vet (29,0% versus 27,7%) en niet-puntsgewijs/niet-vet ten opzichte van puntsgewijs/niet-vet (25,5% versus 19,5%) en dat daardoor de interactie in de gazetijdenanalyses niet zichtbaar werd.

De verschillende resultaten met de gazetijden en de regressiepadtijden laten zien dat het belangrijk is om de aantallen regressies mee te laten wegen in de keuze voor een van deze maten.

5. Conclusie

Het mag wel duidelijk zijn dat de hier gepresenteerde casus een onvolledig beeld geeft van het uitgevoerde onderzoek. Het doel van dit artikel is echter om duidelijk te maken dat het inzetten van oogbewegingsregistratie in leesonderzoek binnen ieders bereik ligt. De hardware wordt steeds eenvoudiger te bedienen en in te stellen. Dat geldt bijvoorbeeld voor de

eyetracker die hier is besproken (de EyeLink II van SR Research) maar in nog hogere mate voor bijvoorbeeld de nieuwe eyetracker van Tobii⁶. Het enige waar de taalverwerkingswetenschapper op moet letten is dat de apparatuur een adequate temporele resolutie heeft, d.w.z. dat hij zeer korte tijden kan meten. In leesonderzoek zijn effecten van 20 milliseconden geen uitzondering. Een apparaat met een temporele resolutie van 60 Hz, dat dus elke 16,6 milliseconden een meting doet, maakt het welhaast onmogelijk een dergelijk verschil aan het licht te brengen, aangezien de grootte van het gezochte effect min of meer gelijk staat aan slechts één enkele meetwaarde. Ook de software is geen belemmering meer om oogbewegingsonderzoek te doen. Software zoals *Fixation* heeft een enorm bedieningsgemak en een grote flexibiliteit. Het is geen enkel probleem om in leesonderzoek de leestijden te analyseren van woordsegmenten, afzonderlijke woorden, groepen van woorden of alinea's. En dat kan met verschillende afhankelijke maten van verwerking. In het artikel is duidelijk gemaakt dat de onderzoeker hieraan veel aandacht zal moeten besteden. Daarnaast biedt het programma *Fixation* de mogelijkheid om de verwerking van beeldmateriaal te analyseren. Het genereert daartoe de tijden van de stappen in een kijkpad. Ook experimenten binnen het Visual World Paradigma kunnen worden geanalyseerd. Dit paradigma, dat binnen het taalverwerkingsonderzoek een grote vlucht heeft genomen, maakt gebruik van de eigenaardigheid van mensen om te kijken naar de afbeeldingen van de entiteiten waarover de taaluiting gaat die ze beluisteren. Zo kan simpelweg gezien worden welke entiteit in een zin of tekst in focus is, terwijl een luisteraar taal verwerkt.

Tot slot kan nog gezegd worden dat de oogbewegingsregistratie op een toenemend aantal onderzoeksterreinen wordt ingezet. Zo wordt de techniek steeds meer gebruikt in combinatie met neurimaging-technieken om convergerende resultaten te verkrijgen (zie ook Hoeks, Hendriks & Redeker in dit nummer). De verwachting is dat de techniek verder gebruikt zal worden als pretestinstrument (zie bijvoorbeeld Lentz, Mak & Pander Maat en Van den Haak, De Jong & Schellens in dit nummer), om schrijfprocessen te bestuderen (zie bijvoorbeeld Van Waes & Leijten in dit nummer) en om vast te stellen hoe mensen multimodale informatie verwerken zoals hypertext, beeld en tekst, video en tekst en allerlei combinaties met spraak. Programma's zoals *Fixation* zullen in deze ontwikkelingen meegaan.

Noten

- * Ik wil scriptiestudente Hellen Megens (doctoraalopleiding Tekst en Communicatie aan de Faculteit Communicatie en Cultuur van de Universiteit van Tilburg) bedanken voor het uitvoeren van het leesonderzoek dat als casus dient in dit artikel.
- 1 Informatie over het programma *Fixation* is verkrijgbaar bij de auteur (r.cozijn@uvt.nl).
- 2 Zie www.staiksterk.nl.
- 3 Informatie over het programma *tsg.exe* is verkrijgbaar bij de auteur.
- 4 Zie www.eyelinkinfo.com.
- 5 De ruwe data van de EyeLink II eyetracker zijn binair opgeslagen in zogenaamde EDF-bestanden. Het omzetten naar ascii-formaat gebeurt met het programma *edf2asc.exe* dat bij de apparatuur is meegeleverd. Het resultaat zijn ascii-bestanden met de extensie ASC.
- 6 Zie www.tobii.se.

Bibliografie

- Cozijn, R., Vonk, W., & Noordman, L. G. M. (2003). Afleidingen uit oogbewegingen: De invloed van het connectief 'omdat' op het maken van causale inferenties. *Gramma/TTT*, 9, 141-156.
- Frazier, L., & Rayner, K. (1982). Making and correcting errors during sentence comprehension: Eye movements in the analysis of structurally ambiguous sentences. *Cognitive Psychology*, 14, 178-210.
- Irwin, D. E. (1998). Lexical processing during saccadic eye movements. *Cognitive Psychology*, 36, 1-27.
- Jakob, R. J. K., & Karn, S. K. (2003). Eye tracking in human-computer interaction and usability research. In J. Hyönä, R. Radach & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 291-312). Oxford: Elsevier.
- Just, M. A., & Carpenter, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, 87, 329-354.
- Kenny, A., Koesling, H., Delaney, D., McLoone, S., & Ward, T. (2005). *A preliminary investigation into eye gaze data in a first person shooter game*. Paper presented at the Proceedings 19th European Conference on Modeling and Simulation, Riga, Latvia.
- Liversedge, S. P., Paterson, K. B., & Pickering, M. J. (1998). Eye movements and measures of reading time. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 55-76). Oxford: Elsevier.
- Matin, E. (1974). Saccadic suppression: A review. *Psychological Bulletin*, 81, 899-917.
- Murray, W. S. (2000). Sentence processing: Issues and measures. In A. Kennedy, R. Radach, D. Heller & J. Pynte (Eds.), *Reading as a perceptual process* (pp. 649-664). Oxford: Elsevier.
- Nielsen, J. (2000). *Designing web usability*. Indianapolis: New Riders Publishing.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 12, 372-422.
- Rayner, K., & Duffy, S. A. (1986). Lexical complexity and fixation times in reading: Effects of word frequency, verb complexity, and lexical ambiguity. *Memory & Cognition*, 14, 191-201.
- Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Englewood Cliffs: Prentice Hall.
- Reynolds, R. E., & Shirey, L. L. (1988). The role of attention in studying and learning. In C. E. Weinstein, E. T. Goetz & P. A. Alexander (Eds.), *Learning and study strategies* (pp. 77-100). New York: Academic Press.
- Vonk, W., & Cozijn, R. (2003). On the treatment of saccades and regressions in eye movement measures of reading time. In J. Hyönä, R. Radach & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 291-312). Oxford: Elsevier.
- Waller, R. H. W. (Ed.). (1980). *Graphic aspects of complex texts: Typography as macropunctuation* (Vol. 13). New York: Plenum Press and NATO Scientific Affairs Division.

Oogbewegingsregistratie en gebruikersonderzoek

1. Inleiding

Er zijn allerlei manieren om teksten te pretesten op kwaliteit. Zo kan onderscheid worden gemaakt tussen synchrone en retrospectieve methoden, en tussen methoden die de nadruk leggen op tekstbeoordeling en op het daadwerkelijk gebruiken van de tekst. In eerder onderzoek hebben we vastgesteld dat het *lezersprotocol* een sterke methode is om begripsproblemen te identificeren (Pander Maat & Lentz, 2003). Het lezersprotocol is een variant van de hardopdenkmethode waarbij lezers tijdens het hardoplezen van een tekst commentaar leveren. Het sterke punt van de methode lijkt er vooral in te liggen dat lezers het verbaliseren van problemen nauwelijks hoeven *uit te stellen*, anders dan bijvoorbeeld bij de plus-en-minmethode het geval is. Pander Maat & Lentz concluderen dat deze methode veel meer probleemsignaleringen oplevert bij zin-voor-zin commentaar geven dan bij alinea-voor-alinea commentaar geven.

Het is mooi dat het hardopdenken veel informatie levert over begripsproblemen, maar het is wel de vraag of de opdracht om commentaar te leveren het leesproces beïnvloedt. Leest een lezer anders wanneer hij weet dat er straks commentaar moet worden gegeven? We hebben op dit punt geen hypothesen, omdat we geen eerder onderzoek naar het effect van commentaartaken op het leesproces kennen.

In Pander Maat en Lentz (2003) bleek het voor de hoeveelheid commentaar niet uit te maken of de tekst hardop dan wel stil gelezen werd. Wij vroegen ons af welke invloed hardoplezen op het leesproces zou kunnen hebben. Het meest voor de hand liggende effect is natuurlijk dat hardoplezen het leesproces vertraagt (zie Rayner, 1998). Maar het is ook denkbaar dat bij hardoplezen de lezer in veel hogere mate lineair door de tekst heen gaat

Samenvatting

Welke mogelijkheden biedt oogbewegingsregistratie ons vakgebied om op het spoor te komen van problemen die lezers ervaren met teksten? In een experiment is aan proefpersonen een tekst voorgelegd van een financiële bijsluiter. Deze werd in verschillende condities gelezen: wel of geen commentaar geven en wel of niet hardoplezen. Na afloop werd een begripstoets afgenomen op een aantal moeilijke woorden in de tekst. De resultaten maken duidelijk dat zowel de opdracht hardop te lezen als de opdracht om na het lezen commentaar te geven gevolgen heeft voor het leesproces zelf in vergelijking met proefpersonen die de tekst stillezen. Voor het opsporen van lexicale begripsproblemen biedt de gebruikte techniek nog weinig steun. In een slotbeschouwing wordt gereflecteerd op de problemen en de mogelijkheden van oogbewegingsregistratie voor evaluatie-onderzoek.

dan bij stillezen het geval is. Omdat hij gedwongen is het ‘voorlezen’ voort te zetten, kan hij immers niet vrijelijk terugkijken naar eerdere tekstgedeelten, ook wanneer hij dat normaal gesproken wel zou doen. Dit soort vragen kan eigenlijk alleen worden beantwoord wanneer het leesproces rechtstreeks geregistreerd wordt. Daarom hebben we in dit onderzoek met oogbewegingsregistratie gewerkt.

Maar onze interesse in oogbewegingsregistratie ging verder dan de hierboven genoemde vragen. We hebben de methode niet alleen gebruikt bij het onderzoek naar het effect van commentaar geven en hardoplezen op het leesproces, we zijn ook geïnteresseerd in de mogelijkheid om oogbewegingsregistratie *zelf* als pretestmethode te gebruiken, met name bij het opsporen van begripsproblemen. Onze redenering is dat deze problemen vaak zichtbaar kunnen zijn in verstoringen van het leesproces. Wanneer bijvoorbeeld een woord extreem lang gelezen wordt, of er wordt later nog eens naar teruggekeken, of er wordt aan het eind van de zin vrij lang gewacht, dan zou dat kunnen duiden op begripsproblemen. Rayner (1998) geeft allerlei onderzoek weer waaruit blijkt dat met name herlezen gerelateerd is aan problemen bij het opbouwen van mentale representaties van de tekstinhoud. Daarnaast is bekend dat minder frequente woorden langere fixaties krijgen; en moeilijke woorden behoren vaak tot de minder frequente woorden.

Daarom hebben we in dit onderzoek verkend in hoeverre oogbewegingen indicaties voor begripsproblemen kunnen bevatten. Het zou natuurlijk heel mooi zijn als begripsproblemen konden worden opgespoord met een methode die niet of nauwelijks ingrijpt in het leesproces, in die zin dat de lezers niets ‘extra’s’ hoeven te doen naast het lezen van de tekst. We hebben ons hier beperkt tot een specifieke categorie begripsproblemen, namelijk problemen met moeilijke woorden.

Samenvattend zijn er drie onderzoeksvragen:

- Welke invloed heeft de instructie om commentaar te leveren op het leesproces van een zakelijke tekst?
- Welke invloed heeft de instructie om een tekst hardop (in plaats van stil) te lezen op het leesproces?
- Kan registratie van oogbewegingen indicaties opleveren voor lexicale begripsproblemen bij het lezen van zakelijke teksten?

Het onderzoek naar de eerste twee vragen is experimenteel te noemen, dat naar de derde vraag exploratief.

2. Methode

Om de eerste twee vragen te beantwoorden is een experiment opgezet met als onafhankelijke variabelen *commentaar geven* en *manier van lezen*. Aldus ontstonden vier condities:

- stillezen en geen commentaar geven;
- hardoplezen en geen commentaar geven;
- stillezen en commentaar geven;
- hardoplezen en commentaar geven.

Ten behoeve van de verkennende onderzoeksvraag naar indicaties voor leesproblemen vulden alle proefpersonen na afloop van de leesactiviteit een digitale vragenlijst in waarin gevraagd werd naar de betekenis van tien woorden die in de tekst voorkwamen. Op die manier wilden we nagaan of er op de een of andere manier een relatie te vinden was tussen het niet begrijpen van specifieke woorden en het leesgedrag in passages waarin die woorden voorkwamen.

2.1 Proefopstelling. De proefopstelling was als volgt. In het laboratorium van het UiL-OTS is een cabine waarin een eyetracker is opgesteld. Figuur 1 biedt een weergave van de opstelling.



Figuur 1: Proefopstelling met de eye-tracker.

Deze opstelling bestaat uit de volgende componenten. Voor de registratie van oogbewegingen wordt gebruikgemaakt van de SMI Eyelink I. Dit is een systeem met een hoofddset waar twee camera's aan gemonteerd zijn, die zodanig worden ingesteld dat ze de bewegingen van de pupillen kunnen registreren. Dit systeem maakt gebruik van infrarood-technologie en meet met een frequentie van 250 Hz, binnen een bereik van dertig graden horizontaal en twintig graden verticaal, de pupilpositie met een nauwkeurigheid van 0.5 tot 1 graad. Twee computers maken deel uit van het systeem: op de ene pc draait de SMI eye-tracker en worden dus de data geregistreerd; op de andere pc draait het stimulusmateriaal waar de proefpersoon mee geconfronteerd wordt. Aan deze pc zijn twee monitors gekoppeld, een voor de proefpersoon zelf en een in de ruimte van de proefleider. Voor dit specifieke experiment maakten we gebruik van het softwareprogramma *Gazetracker* waarmee het mogelijk is onder Windows stimulusmateriaal aan te bieden en data te registreren.

2.2 Materiaal. Het materiaal bestaat uit een financiële bijsluiter *Rendemix* zoals we die op het Internet hebben aangetroffen op de site van de Rabobank. Sinds 2002 zijn banken verplicht voor hun financiële producten een bijsluiter te maken die het publiek in staat moet

stellen een keuze te maken tussen verschillende producten en de risico's goed in te schatten. De Autoriteit Financiële Markten heeft regels opgesteld waar deze bijsluiters aan moeten voldoen.

De *Rendemix* bijsluiters is eerder object van onderzoek geweest (Kippersluis, 2005) met als doel na te gaan of er problemen waren met moeilijke woorden in de tekst. Daartoe werd een digitale versie gemaakt waarin proefpersonen de mogelijkheid kregen woorden die zij niet begrepen aan te klikken, waarna uitleg op het scherm verscheen. Aldus werd een indicatie verkregen van de woorden waar proefpersonen moeite mee hadden. Deze tekst is enigszins herzien, om zo op elk scherm voor de proefpersoon een afgeronde passage te kunnen presenteren. Aldus ontstond een versie van de bijsluiters die in tien schermen gepresenteerd werd. Figuur 2 biedt een weergave van de uiteindelijke presentatievorm.



Figuur 2: Weergave van een scherm in de proefopstelling.

Vanwege de gewenste nauwkeurigheid van meting werd gekozen voor een regelafstand die groter is dan normaal. En om de verhoudingen in balans te houden, werd ook gekozen voor een wat groter lettertype. Per scherm werden aldus maximaal negen regels tekst aangeboden met een regelafstand van 1,5 centimeter (36 pt.) en een lettergrootte van 16,5 pt. Aldus ontstond een scherm dat zeker niet gebruikelijk is voor de financiële bijsluiters op het web. Hier hebben we dus een concessie moeten doen ten gunste van de kwaliteit van de meting. Verderop zal blijken dat we hierin misschien nog niet ver genoeg gegaan zijn.

De interface van de proefopstelling zag er als volgt uit. In een beginscherm stond de toewijzing aan een van de condities (hardop of stil werken en wel of geen commentaar geven op elk scherm) en een link naar het eerste scherm. De interface verschilde per conditie, omdat de commentaargevers onder elk scherm een link *commentaar* aantroffen, die zij aan moesten klikken alvorens commentaar te geven. Daarna verscheen precies hetzelfde scherm met rechtsonder in rood de opdracht: *geef commentaar*. Daaronder stond een link *Verder*, die leidde

naar de volgende passage van de bijsluiters. Op deze manier konden we de registratie van de oogbewegingen tijdens de eerste lezing en die tijdens het commentaar geven van elkaar scheiden. Proefpersonen werden geacht hun commentaar uit te stellen tot na afronding van het lezen van de passage. De interface dwong dus de volgende volgorde af: lezen – klikken op *commentaar* – commentaar geven op de passage – klikken op *verder* – lezen van de volgende passage. De interface van de hardoplezers verschilde niet van die van de stillezers.

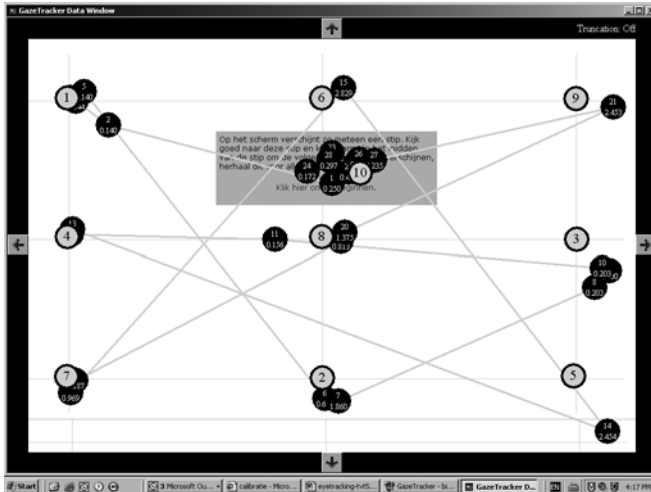
2.3 Proefpersonen. De proefpersonen werden geworven via een site van het UiL-OTS waar vooral studenten op afkomen. Zij konden zich daar digitaal inschrijven voor een sessie en ontvingen na afloop een vergoeding van 5 euro. Aan de proefpersonen werden enkele specifieke voorwaarden gesteld die voortvloeien uit de technologie. Zo waren brildragers bij voorbaat uitgesloten omdat de registratie bij hen tot onvoldoende resultaten leidt. Proefpersonen die mascara droegen, moesten dit verwijderen voordat de sessie kon beginnen, omdat mascara tot gevolg heeft dat het systeem de pupil niet goed kan detecteren. Daartoe was materiaal in het laboratorium aanwezig. Proefpersonen mogen niet langer dan 30 minuten over de taak doen, omdat dan langzamerhand de ogen uitdrogen, wat tot extra knipperen leidt, hetgeen ook weer leidt tot verhoogde onnauwkeurigheid. Onze taak kostte de proefpersonen vijf tot tien minuten. In totaal namen 26 proefpersonen deel aan dit onderzoek. Na controle van de data vielen er twee af, omdat de nauwkeurigheid van de registratie te gering bleek te zijn.

2.4 Procedure. De procedure was als volgt. Na binnenkomst werd de proefpersonen kort uitgelegd dat het om een onderzoek ging naar de kwaliteit van een financiële bijsluiters, waar hun feedback voor nodig was. Daarna werd de proefpersonen gevraagd goed te gaan zitten, zodanig dat de muis goed bereikbaar was en het scherm goed zichtbaar was. Als ze eenmaal goed zaten, werd de hoofdset opgezet, werden de camera's goed op de pupil gericht, en werd gecontroleerd of de hoofdset goed stevig zat. Ook dit is van groot belang voor de kwaliteit van de meting. Onverwachte bewegingen met het hoofd kunnen namelijk leiden tot een verschuiving van de hoofdset en een slechtere kwaliteit van de meting.

Daarna startte de eerste kalibratie. De proefpersoon werd gevraagd zich goed te concentreren op een zwarte stip die op negen plaatsen in het scherm verscheen en daar zichtbaar bleef totdat de proefpersoon goed fixeerde. De proefleider-pc registreerde de positie van de pupil en gaf een score in termen van *failed*, *poor* of *good calibration*. In het eerste geval werd deze procedure herhaald. Vervolgens werd de Gazetracker software geactiveerd en doorliep de proefpersoon een procedure die vergelijkbaar was met de kalibratie. Hij of zij moest telkens met de muis op de juist verschenen stip klikken, waarna een volgende stip verscheen. Deze procedure werd aan het eind van de sessie nogmaals uitgevoerd. Op die manier konden we controleren in hoeverre de data aan het begin en het eind van de sessie met dezelfde mate van nauwkeurigheid geregistreerd waren. Bovendien konden we op basis van deze procedure eventuele afwijkingen in de metingen achteraf corrigeren.

Na het lezen en eventueel commentaar geven werd de opname van geluid en oogbewegingen gestopt. De hoofdset werd afgenomen en vervolgens werd de korte digitale vragenlijst op het scherm getoond met tien meerkeuzevragen over moeilijke woorden. Na elke vraag kwam een vraag waarin op een vijfpuntsschaal *feeling of knowing* werd bevraagd. De vraag luidde: *In hoeverre bent u zeker van het antwoord dat u gaf.*

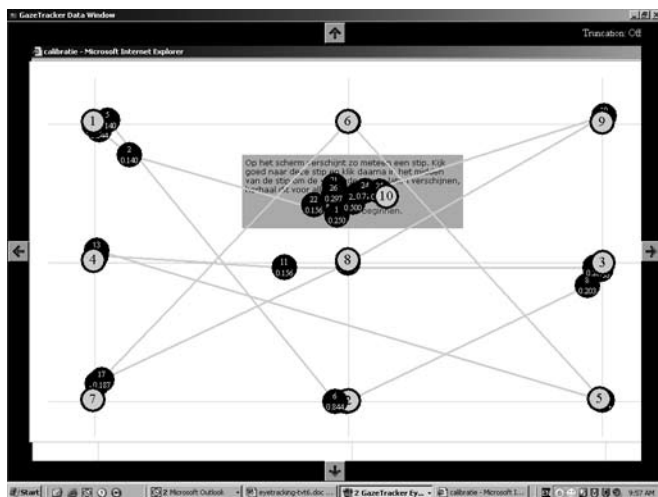
2.5 Analyse. De analyse van de data begon voor elke proefpersoon met een controle van de betrouwbaarheid van de meting. Daartoe werden de door ons ingevoegde twee extra kalibraties gebruikt. Figuur 3 biedt een weergave van zo'n kalibratie.



Figuur 3: Weergave van de kalibratie in Gazetracker.

In de figuur zien we de negen stippen die na elkaar op het scherm verschijnen met daarop (in groen) weergegeven een rondje dat aangeeft dat daar geklikt is door de proefpersoon. De nummers geven de volgorde van de kliks aan. De zwarte rondjes zijn weergaven van de fixaties van de proefpersoon. In elk rondje staat een nummer dat weergeeft de hoeveelste fixatie het betreft en een getal dat de duur van de fixatie weergeeft. Met behulp van deze output kunnen we nagaan in hoeverre *Gazetracker* een fixatie heeft geregistreerd op het punt waar geklikt werd. De applicatie is door ons zo gebouwd dat de proefpersoon goed moet mikken in het midden van het rondje. Dat kan niet zonder een duidelijke fixatie. Wanneer de fixatie dus naast de klik getoond wordt, is er sprake van een meetfout. Zo zien we in Figuur 3 dat de meting in de linker bovenhoek precies op de juiste plaats is, terwijl de meting in de rechter benedenhoek naar boven gecorrigeerd zou moeten worden, omdat de fixatie enigszins onder het te fixeren punt is weergegeven. Aan dit scherm is goed te zien dat de fout niet op het hele scherm gelijk is, en dat de mate van correctie dus afhankelijk zou moeten zijn van de positie op het scherm.

Nu biedt de software van *Gazetracker* de mogelijkheid om de registratie achteraf te corrigeren met een functie *Shift data* (zie ook de bespreking van *Fixation* door Cozijn in dit nummer). De proefleider verplaatst dan het fixatiepunt met de muis – in het geval van Figuur 3 – een klein stukje naar boven. Wanneer dit voor alle negen punten gedaan wordt, herberekent *Gazetracker* met een algoritmische functie de data voor de gecorrigeerde registratie. Figuur 4 toont het resultaat van de correctie voor het kalibratiescherm. De correctie wordt vervolgens ook toegepast op de data van de schermen met tekst.



Figuur 4: Kalibratiescherm na correctie van de data.

Wij hebben op deze manier alle data gecontroleerd en gecorrigeerd door voor de eerste vijf schermen uit te gaan van de eerste kalibratie en voor de laatste vijf schermen uit te gaan van de laatste kalibratie. Hoewel dit een prachtige functie is, gegeven de onnauwkeurigheid waar we nu eenmaal mee moeten leven, is deze toepassing niet zonder risico's. Wanneer we de data van de schermen 6 tot en met 10 namelijk corrigeren op basis van de kalibratie die plaatsvond na het tiende scherm, veronderstellen we dat de onnauwkeurigheid die we aldaar corrigeren ook van toepassing is op eerdere schermen. We corrigeren immers achterwaarts, zonder precies te weten op welk moment de toenemende onnauwkeurigheid optrad. Gesteld dat het probleem zich pas voordeed in het zevende scherm, dan heeft deze operatie voor het zesde scherm eerder een verslechtering dan een verbetering tot gevolg. Idealiter vindt na elk scherm een kalibratie plaats, maar daarmee zouden we het leesgedrag wel heel erg verstoren. Daarom beperkten we ons tot een extra kalibratie aan het begin en het eind van het leesproces, wetend dat in het algemeen de waarde van de correctie daalt naarmate je verder van de kalibratiepagina komt.

Ten behoeve van de analyse hanteerden we de volgende afhankelijke variabelen:

- de totale leestijd van de gehele tekst;
- de terugleestijd (de tijd die wordt besteed aan het lezen van reeds eerder gelezen regels in de tekst);
- de vooruitleestijd (de leestijd minus de terugleestijd);
- het aantal keren teruglezen.

3. Resultaten

De eerste stap bij de analyse van de data was het exploratief bekijken van de oogbewegingspatronen. Een belangrijke constatering hierbij was dat de data ook na correctie niet voldoende betrouwbaar waren om op woordniveau te kunnen vaststellen waar fixaties hadden

plaatsgevonden. Er kon daarom geen analyse van de leestijden per woord worden uitgevoerd. In de volgende paragraaf gaan we dieper op dit punt in.

Bij het analyseren van de effecten van de onafhankelijke variabelen op het leesgedrag werden vier maten gebruikt. De eerste maat was de *totale leestijd* op de pagina. Daarnaast waren we geïnteresseerd in hoe eventuele effecten van de onafhankelijke variabelen tot uitdrukking zouden komen in het leespatroon. De verwachting daarbij was dat het effect van hardop lezen zou zijn dat proefpersonen meer 'lineair' zouden lezen, d.w.z. dat ze minder vaak dan bij stillezen terug zouden springen naar eerder gelezen gedeelten van de tekst. Een andere verwachting was dat commentaar geven juist zou leiden tot meer terugspringen in de tekst dan wanneer geen commentaar hoefde te worden gegeven.

Vanwege het feit dat de fixatiegegevens niet betrouwbaar genoeg waren om op woordniveau de fixaties toe te kennen, werd er voor de analyse van het leespatroon uitgegaan van de leestijden per regel. Hierbij werd onderscheid gemaakt tussen de *vooruitleestijd* en de *terugleestijd*. Om deze twee maten te definiëren werd bijgehouden tot welke regel in de tekst een proefpersoon was gevorderd. Alle fixaties op deze regel werden gerekend bij de vooruitleestijd, fixaties op eerdere regels werden gerekend bij de terugleestijd. We hebben hier voor deze vrij grove maat van terugleestijd gekozen omdat we, vanwege het gebrek aan precisie van de metingen niet konden vaststellen of een fixatie die links van een voorafgaande fixatie terechtkwam, een sprong terug was naar een eerder woord (en het bij terugleestijd gerekend moest worden) of een tweede fixatie was op hetzelfde woord (en het bij vooruitleestijd gerekend moest worden).

Naast de terugleestijd hebben we ook gekeken naar het aantal keren dat de lezer tijdens het lezen van de tekst *terugsprong* naar eerdere regels.

Op deze data werden variantie-analyses uitgevoerd. De resultaten van de verschillende afhankelijke maten zijn weergegeven in Tabel 1. De variantie-analyse over de totale leestijd liet hoofdeffecten zien van commentaar ($F(1,20) = 8.66, p < .01$) en van manier van lezen ($F(1,20) = 11.87, p < .005$). De proefpersonen deden langer over het lezen van een pagina wanneer ze na het lezen commentaar moesten geven dan wanneer ze geen commentaar hoefden te geven, en ze deden langer over het lezen van een pagina wanneer ze hardop lasen dan wanneer ze stil lasen. Er was geen interactie tussen beide onafhankelijke variabelen.

In de vooruitleestijd was alleen een hoofdeffect te zien van manier van lezen ($F(1,20) = 9.30, p < .01$). Wanneer proefpersonen hardop lasen was de vooruitleestijd langer dan wanneer ze stil lasen. In de terugleestijd was alleen een hoofdeffect te zien van commentaar ($F(1,20) = 8.32, p < .01$). In de conditie met commentaar besteedden de proefpersonen meer tijd aan het herlezen van eerdere delen van de tekst dan in de conditie zonder commentaar. In het aantal keren teruglezen was hetzelfde patroon te zien, dit effect was echter niet significant ($F(1,20) = 3.01, p = .09$).

Er waren grote verschillen in de aantallen commentaren die de proefpersonen gaven. De aantallen commentaren op de tien pagina's liepen uiteen van in totaal 3 tot 20, met een gemiddelde van 8.1. Er was geen verschil in aantal commentaren tussen de hardoplezers en de stillezers. Ook was er geen correlatie tussen het aantal commentaren en de leestijdmaten.

Tabel 1: Totale leestijd, Vooruitleestijd en Terugleestijd per conditie (in seconden per gelezen scherm) en Aantal keren terug lezen per scherm

		Totaal	Vooruit	Terug	Aantal keren terug
Stillezen	Geen commentaar	14.2	10.3	4.0	1.4
	Wel commentaar	19.3	11.4	7.9	3.4
Hardoplezen	Geen commentaar	20.1	15.6	4.6	2.5
	Wel commentaar	25.1	16.2	8.9	3.3

4. Discussie

Het experiment had twee doelen. Het eerste doel was te onderzoeken welke invloed het hardoplezen van een tekst of het geven van commentaar na het lezen van een tekst hebben op het leespatroon. Uit het experiment werd duidelijk dat beide taken een effect hebben op het leespatroon, alleen wel een verschillend effect. Hardoplezen had alleen een effect op het *leestempo*. Hardoplezers lazen langzamer dan stillezers. In tegenstelling tot wat we verwachtten hadden hardoplezers niet meer dan stillezers de neiging om 'lineair te lezen'. Hardoplezers keken niet minder vaak of minder lang dan stillezers terug naar eerdere delen van de tekst. Commentaar geven had juist alleen een effect op het *teruglezen* in de tekst. De vooruitleestijd van commentaargevers was niet langer dan die van proefpersonen die geen commentaar moesten geven, wel besteedden de commentaargevers meer tijd aan reeds eerder gelezen delen van de tekst. Let wel, dit terugkijken trad niet op tijdens het commentaar geven, maar reeds in het eerste deel van het leesproces.

Wat betekent dit resultaat nu voor de validiteit van pretesten waarin proefpersonen hardop moeten lezen of commentaar moeten geven op teksten? Uit het feit dat proefpersonen in de commentaarconditie meer en vaker teruglezen mogen we afleiden dat zij intensiever lezen, en vermoedelijk een hogere ambitie hebben om de tekst te begrijpen dan mensen die niet deelnemen aan een pretest en dezelfde tekst gewoon thuis lezen. Dat kan ertoe leiden dat we in een pretest problemen op het spoor komen die zich in de praktijk niet voordoen, omdat mensen buiten de pretest met een lagere ambitie lezen. Zij lezen als het ware over de problemen heen. Maar het omgekeerde is ook denkbaar: in de praktijk zullen mensen meer problemen ervaren, omdat zij minder cognitieve energie steken in de constructie van een mentale representatie. Er lijkt dus een validiteitsbedreigende factor in het geding te zijn, maar we weten niet precies wat het effect daarvan is. Er is echter ook een derde redenering denkbaar die de validiteitsbedreigende factor nuanceert. De proefpersonen in dit onderzoek hadden zelf geen initiatief genomen om een tekst te lezen over een beleggingsproduct. Zij waren niet echt met beleggen bezig. Vermoedelijk zullen lezers die overwegen in te stappen in het beschreven fonds een hogere ambitie hebben om de tekst te lezen dan de proefpersonen die deelnamen aan de pretest. En daarmee zou hun ambitieniveau goed kunnen lijken op dat van de proefpersonen aan wie gevraagd werd commentaar op de tekst te leveren. Het lijkt ons wenselijk dat in de toekomst meer onderzoek gedaan wordt met oogbewegingsregistratie naar leespatronen van mensen die in verschillende pretestcondities dezelfde tekst lezen.

Een daarmee verbonden discussiepunt is de vraag of we het leesgedrag zoals dat thuis plaatsvindt als norm moeten stellen. Gesteld dat lezers daar nauwelijks problemen ervaren, terwijl we als experts toch sterke vermoedens hebben dat lezers niet weten wat woorden als *benchmark* en *bewaarloon* betekenen, welk criterium leggen we dan aan voor “adequaat begrip”? De norm die voortvloeit uit een functionele analyse (mensen kunnen alleen begrijpen wat het doel van het fonds is als zij weten wat een *benchmark* is) of het meer descriptieve criterium dat het gros van de lezers het niet nodig vindt om dat te begrijpen als zij overwegen geld in een beleggingsfonds te storten?

Tijdens de afname van de test hadden we de indruk dat de proefpersonen niet erg gul waren met commentaar. Gemiddeld produceerden zij in totaal acht commentaren met een probleemdetectie over de tien getoonde schermen. Dat is dus minder dan één commentaar per scherm. Als we dat vergelijken met commentaren die geproduceerd werden bij teksten over veilig vrijen en de gevaren van alcohol, zoals door De Jong (1998) verzameld, dan is er echter nauwelijks verschil. De teksten zijn ongeveer even lang en de gemiddelde hoeveelheid commentaar is bij De Jong tussen de 6 en de 9 commentaren per proefpersoon. Het is denkbaar dat onze indruk een gevolg is van de opstelling in het laboratorium. De proefleider en de proefpersoon bevinden zich in een aparte ruimte, verbonden door een deur (die we meestal open lieten staan) en een glazen ruit. De proefpersoon moest vooral stil zitten en naar het scherm kijken. Die twee factoren – de ruimtelijke opstelling en de gerichtheid op een beeldscherm – hadden tot gevolg dat er nauwelijks interactie was tijdens de sessie tussen proefleider en proefpersoon. Dat was uitdrukkelijk ook de opzet, omdat elke interactie het risico inhoudt dat de proefpersoon het hoofd ineens naar de proefleider wendt, waardoor de hoofdset verschuift en de data verloren gaan. Hoewel de hoeveelheid geproduceerd commentaar dus misschien niet eens tegenvalt, trekken we toch de conclusie dat voor een pretest interactie wel degelijk wenselijk is, en oogbewegingsregistratie alleen een optie is als dat zonder hoofdset gebeurt. Het voornaamste argument voor die conclusie is het gebrek aan diagnostische kwaliteit van veel commentaar. Proefpersonen merken bijvoorbeeld op dat de tweede zin in het fragment best moeilijk is en klikken dan door naar het volgende scherm. Een proefleider heeft dan behoefte om een vraag te stellen over de aard van die moeilijkheid. Staan er moeilijke woorden in die zin of is de constructie lastig?

Het derde doel van het experiment was inzicht te krijgen in de vraag in hoeverre oogbewegingsregistratie kan helpen bij het detecteren van moeilijke woorden in de tekst. Helaas waren de data niet nauwkeurig genoeg om hierover uitspraken te kunnen doen. Het gebrek aan nauwkeurigheid in de data heeft te maken met het feit dat we gebruik maakten van een webomgeving, waarin de proefpersonen met behulp van muiskliks door de pagina's moesten navigeren. Het eyelink systeem is te gevoelig voor de bewegingen van het hoofd die hiermee gepaard gaan, zodat het niet mogelijk bleek om op woordniveau te meten. Om wel de juiste nauwkeurigheid in de metingen te bereiken hebben we twee opties: ofwel de proefpersonen minder laten bewegen en dus geen muis meer te laten gebruiken, ofwel een eye-tracking systeem te gebruiken dat minder gevoelig is voor de bewegingen die gepaard gaan met het gebruik van muis en toetsenbord.

Het is mogelijk om de bewegingen van de proefpersoon te beperken, door bijvoorbeeld in plaats van de muis een ander display met slechts twee toetsen of een knop op het toetsenbord te gebruiken. In het huidige experiment was dat mogelijk geweest, maar wanneer het

normale gedrag van mensen tijdens het surfen op internet moet worden vastgelegd, is dat een te grote beperking, zeker wanneer de proefpersoon ook de mogelijkheid moet hebben om te typen tijdens de meting van de oogbewegingen. Onze conclusie is dat in die gevallen een systeem moet worden gebruikt dat een zekere mate van beweging van de proefpersoon toestaat.

5. Hoe verder?

Wat kunnen we nu verwachten van deze techniek en wat zijn onze plannen voor de toekomst om met oogbewegingsregistratie problemen met teksten op te sporen? We hebben de hoop nog niet opgegeven dat lexicale begripsproblemen met oogbewegingsregistratie opgespoord kunnen worden. Daartoe zijn we inmiddels overgestapt op een nieuw systeem, ontwikkeld door Tobii, waarmee proefpersonen alle vrijheid hebben om te bewegen terwijl de kalibratie heel nauwkeurig is. In feite hebben proefpersonen met dit systeem nauwelijks meer in de gaten dat hun oogbewegingen geregistreerd worden. Zij kunnen zelfs opstaan, weglopen en weer terugkeren naar de computer, waarna automatisch de blik weer gevangen wordt. Het nadeel van dit systeem is dat de frequentie daalt van 250 Hz in het tot nu toe gehanteerde systeem naar 50 Hz, waardoor een verlies in nauwkeurigheid ontstaat van de meting van fixatietijden en saccades. Met dit systeem zullen wij het hier gerapporteerde onderzoek repliceren om na te gaan of langere leestijden indicaties zijn voor lexicale begripsproblemen. Uit Fins onderzoek is steun gekomen voor dat idee. Hyrskykari e.a. (2003) rapporteren over een proactief systeem dat in staat is om op grond van oogbewegingen te voorspellen dat een lezer een begripsprobleem ervaart, waarna een toelichting op de betekenis van dat woord automatisch in beeld verschijnt. Dit systeem is ontwikkeld in de context van het lezen van teksten in een vreemde taal; die toelichting is dus in feite een vertaling van de term. Het is natuurlijk goed denkbaar dat zoiets ook voor teksten in de moedertaal ontwikkeld wordt.

Voor het opsporen van andere problemen met teksten dan lexicale problemen achten wij oogbewegingsregistratie zeker ook kansrijk. In eerste instantie denken we dan aan problemen met de structuur van de tekst. Voorspelbaar is dat goed gestructureerde teksten de lezer minder noodzaken heen en weer te springen in de tekst, en dat zoekopdrachten in goed gestructureerde teksten gerichter en sneller uitgevoerd kunnen worden dan in slecht gestructureerde teksten. De frequentie van teruglezen en de duur van de terugleestijd zouden dus goede indicatoren kunnen zijn voor problemen met de structuur van een document. Er zijn ook plannen om oogbewegingsregistratie in te zetten om indicaties op te sporen voor persuasieve effecten van teksten. Recent uitgevoerd onderzoek van Kamalski et al. (in voorbereiding) wijst erop dat proefpersonen gevoelig zijn voor signalen in de tekst die wijzen op een persuasief doel, hetgeen zich vertaalt in een langere leestijd.

Voor de evaluatie van websites lijkt oogbewegingsregistratie een zeer aantrekkelijke methode. In bescheiden mate hebben we hier ervaring mee opgedaan. De grote kracht van deze methode is dat we precies na kunnen gaan of bezoekers van een site een bepaalde cruciale link voor verdere navigatie negeren omdat zij hem niet waargenomen hebben, of omdat ze hem verkeerd geïnterpreteerd hebben. We kunnen zien of een bezoeker waarneemt dat

het menu voor hoofdnavigatie van een site ergens onderweg veranderd is in een domein-specifiek menu. Nauwkeurig kan vastgesteld worden of de banners op een site wel of niet bekeken worden. Aansluitend bij Cooke (2005) concluderen we dat het “lezen” op een site in hoge mate ook “kijken” is, daarom leent oogbewegingsregistratie zich goed om in een evaluatie-onderzoek problemen op het spoor te komen die bezoekers van websites ervaren.

Bibliografie

- Cooke, L. (2005).** Eye Tracking: How it works and how it relates to usability. *Technical Communication*, 52(4), 456–463.
- Hyrskykari, A., Majaranta, P., & Riih  K. (2003).** Proactive response to eye-movements. In M. Rautenberg e.a. (eds.), *Human-computer interaction – Interact ’03*. IOS Press – IFIP, 129–136.
- Jong, M. de (1998).** *Reader feedback in text design: Validity of the plus-minus method for the pretesting of public information brochures*. Amsterdam/Atlanta: Rodopi.
- Kamalski, J., Lentz, L., Sanders, T., & Zwaan R. (in voorbereiding).** The fore-warning effect of coherence markers in persuasion: Off-line and on-line evidence.
- Kippersluis, B. van (2005).** *De financiële bijsluiter: Experimenteel onderzoek naar het begrip van jargon*. Eindwerkstuk Bachelor. Universiteit Utrecht.
- Pander Maat, H., & Lentz, L. (2003).** Waarom het lezersprotocol zo’n goede pretestmethode is. *Tijdschrift voor Taalbeheersing*, 25(3), 202–220.
- Rayner, K. (1998).** Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372–422.

Communicatie en het brein

Het gebruik van neuroimaging bij onderzoek naar dialogen

1. Inleiding

Bij hun deelname aan een gesprek gaan spreker en luisteraar uit van bepaalde veronderstellingen. Deze veronderstellingen zijn door Grice (1975) geformuleerd in de vorm van vier 'maximes', die als volgt kunnen worden beschreven:

- 1. KWALITEIT: Wees waarheidsgetrouw!
- HOEVEELHEID: Wees niet informatiever, maar ook niet minder informatief dan vereist!
- RELEVANTIE: Wees relevant!
- STIJL: Wees helder!

Dit 'model' van Grice vormde de basis voor een overvloed aan onderzoeken waarin de maximes op hun geldigheid en bruikbaarheid werden onderzocht. Mede omdat deze stelregels voor communicatie in de gebiedende wijs zijn geformuleerd is lang gedacht (en wordt verrassend genoeg nog steeds gedacht), dat de maximes een soort van overgedetailleerd, puriteins recept vormen voor een geslaagde conversatie. Campbell (2001, p. 256), bijvoorbeeld, verzucht: "Would we want to have dinner with such a person, such an impeccably polite maxim observer?" Een meer vruchtbare benadering is echter de maximes op te vatten als een standaardset van veronderstellingen, met name van de luisteraar over de spreker, waarvan alle deelnemers aan het gesprek zich bewust zijn (Horn, 2004). Het Maxime van Hoeveelheid, bijvoorbeeld, beschrijft hoe de luisteraar er in gewone dagelijkse dialogen van uitgaat dat de spreker inderdaad niet meer, maar ook niet minder zegt dan nood-

Samenvatting

Zowel spreker als luisteraar gaan bij hun deelname aan een gesprek uit van bepaalde veronderstellingen (Grice, 1975). De spreker, bijvoorbeeld, wordt geacht niet meer, maar ook niet minder mee te delen dan noodzakelijk. In het huidige onderzoek wordt bekeken hoe de luisteraar (of beter gezegd: haar brein) reageert wanneer deze pragmatische veronderstelling geschonden wordt. Dit soort van vragen kan beantwoord worden met behulp van zogeheten 'neuroimaging' methoden, waarbij de werking van de hersenen in beeld kan worden gebracht. De meest gebruikte neuroimaging methoden (PET, fMRI en ERP's) worden kritisch besproken en de ERP-methode wordt daarna geïllustreerd aan de hand van een experiment. In dit experiment kregen proefpersonen visueel vraag-antwoord paren aangeboden waarin het Griceaanse 'Maxime van Hoeveelheid' werd geschonden. De pragmatische schendingen bleken aanleiding te geven tot een specifieke ERP-component, die voor zover wij weten niet eerder is gevonden. Daarnaast werden andere effecten gevonden die duiden op thematische en syntactische heranalyse van de aangeboden taaluiting. De schending van het Griceaanse maxime wordt dus onmiddellijk signaleerd, en leidt tot aanzienlijke verwerkingsproblemen.

zakelijk is. In het huidige onderzoek hebben we er, om praktische redenen, voor gekozen om de onderzochte dialogen visueel aan de proefpersonen aan te bieden. Ons experiment laat dus zien hoe een taalgebruiker omgaat met schendingen van het Maxime van Hoeveelheid in een uitgeschreven dialoog. Bekijk bijvoorbeeld de fictieve mini-dialoog in (2).

2. V: Wat deed Jan?

A: Hij zoende Annet.

In dit gesprekje geeft A een antwoord dat pragmatisch gezien correct is: A geeft, voor zover we hier kunnen overzien, precies genoeg informatie: het antwoord gaat over 'Jan' en beschrijft een handeling die 'Jan' uitvoerde. Op zich zou de vragensteller dus tevreden moeten zijn. Dat geldt ook voor het vraag-antwoord paar in (3).

3. V: Wat deden Jan en Piet?

A: Jan zoende Annet en Piet zoende Henk.

Ook hier krijgt de vragensteller keurig antwoord op de vraag wat 'Jan' en 'Piet' gedaan hebben. Dit is echter niet het geval in de nu volgende dialoog (4), en dat is juist de reden waarom we dit gesprek als basis nemen voor ons onderzoek.

4. V: Wat deden Jan en Piet?

A: Jan zoende Annet en Piet in de tuin.

We zien hier dat zowel 'Jan' als 'Piet' voorkomen in het antwoord op de vraag, maar dat er ook cruciale informatie ontbreekt, te weten het antwoord op de deelvraag "Wat deed Piet?". Door de luisteraar deze informatie te onthouden, schendt de spreker het Maxime van Hoeveelheid.

Het is misschien goed om te vermelden dat dit niet de enige manier is om het onderhavige maxime te schenden. Bijvoorbeeld, iemand kan op de vraag hoeveel kinderen zij heeft het antwoord "twee" geven, terwijl ze er in werkelijkheid drie heeft, of onterecht zeggen dat het "warm" is, terwijl het snikheet is. Het verschijnsel waar we in (4) mee te maken hebben bevindt zich in een enigszins ander domein en is gerelateerd aan het pragmatische concept 'topic-structuur'. Een 'topic' kan omschreven worden als de entiteit waarover de zin informatie verschaft (Lambrecht, 1994). Door de vraag in (3) en (4) worden er twee entiteiten 'Jan' en 'Piet' geïntroduceerd en tot zeer waarschijnlijke topics voor de volgende zin gemaakt. Anders gezegd: de vraag maakt het zeer waarschijnlijk dat 'Jan' en 'Piet' het zogenaamde 'tweeledig topic' van de volgende zin zullen zijn. Wanneer 'Jan' en 'Piet' samen iets hadden gedaan, zou het antwoord kunnen zijn "Zij deden X". Aan de andere kant zou het antwoord ook een constructie kunnen zijn met zogeheten 'contrastieve' topics, waarbij elk van de jongens afzonderlijk iets doet: "Jan deed X en Piet deed Y." In dialoog (4) wordt dus het Maxime van Hoeveelheid geschonden door het ontbreken van de verwachte topic-structuur: de zin zou een tweeledig topic moeten hebben ('Jan en Piet'), maar heeft slechts een enkelvoudig topic ('Jan').

Nu weten we van eerder onderzoek dat topic-structuur een belangrijke rol speelt in het incrementeel verwerken van taal (Hoeks, 1999; Hoeks, Hendriks, Vonk, Hagoort, & Brown, 2006; Hoeks, Vonk, & Schriefers, 2002). Dit is met name het geval wanneer taaluitingen syntactische ambiguïteiten bevatten, zoals bij de antwoorden van voorbeelden (3) en (4). Doordat taalverwerking incrementeel is, en niet opgeschort wordt tot het einde van de zin,

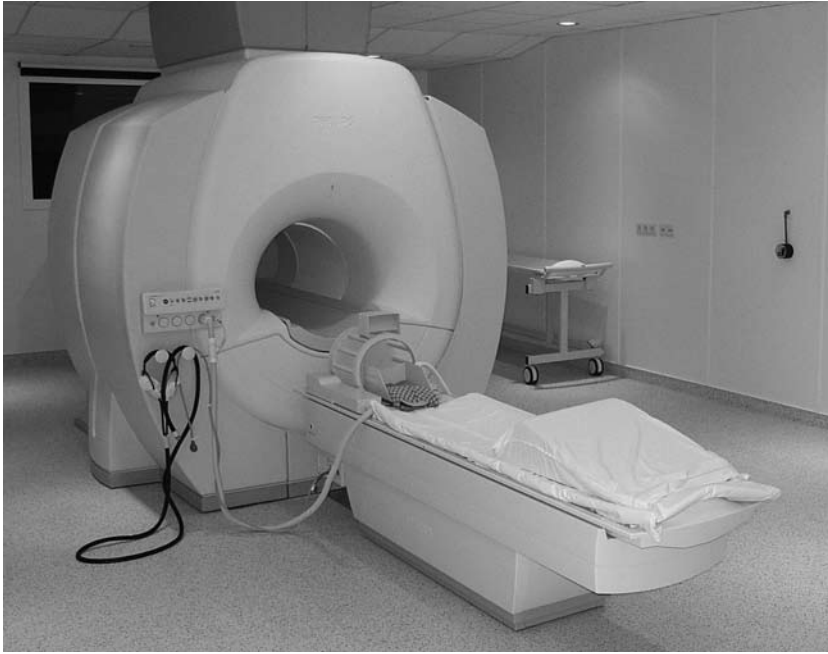
zal de taalverwerker soms voor lastige keuzes komen te staan. Als de luisteraar of lezer van het zinsfragment “Jan zoende Annet en Piet ..” bij de naam ‘Piet’ is aangekomen, zijn er twee mogelijkheden. Ofwel ‘Piet’ is het topic van een gecoördineerde zin, zoals in het antwoord bij voorbeeld (3), ofwel ‘Piet’ is onderdeel van het groepje mensen dat door topic ‘Jan’ wordt gezond, als in het antwoord van (4). Onderzoek heeft aangetoond dat taalgebruikers normaliter een voorkeur hebben voor de structuur in (4), die ook wel ‘NP-coördinatie’ wordt genoemd. In het hier besproken experiment hebben alle te onderzoeken dialogen deze geprefereerde syntactische structuur. De vraag die in dit onderzoek centraal staat, is nu: Wat is het effect op de taalgebruiker van een schending van het Maxime van Hoeveelheid, teweeggebracht door het ontbreken van de verwachte topic-structuur?

2. Vensters op het brein

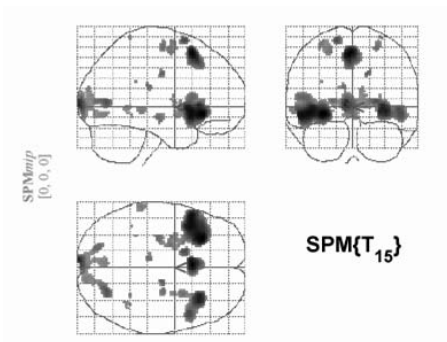
Om te achterhalen hoe een taalgebruiker reageert op een pragmatische schending, zou je het natuurlijk gewoon kunnen vragen. Maar een taalgebruiker kan dit soort vragen alleen beantwoorden als ze toegang heeft tot de cognitieve processen die aan de reactie ten grondslag liggen. Het probleem is dat deze processen meestal onbewust verlopen en daarom is het bijzonder prettig dat er methoden zijn om (reflecties van) cognitieve activiteit waar te nemen zonder dat de proefpersoon daar uitspraken over hoeft te doen. Bedoelde methoden worden ‘neuroimaging’ methoden genoemd omdat ze afbeeldingen voortbrengen van het zenuwstelsel in actie. Ze kunnen in twee klassen verdeeld worden: methoden gebaseerd op veranderingen in de locale doorbloeding van de hersenen (zie sectie ‘Bloed’), en methoden gebaseerd op veranderingen in de elektrische hersenpotentialen (zie sectie ‘Stroom’).

2.1 Bloed. Wanneer een groepje hersencellen actief is tijdens informatieverwerking, bijvoorbeeld tijdens het stellen van een vraag of bij het begrijpen van een antwoord, nemen die cellen zuurstof op uit het bloed, en ontstaat er op korte termijn een (relatief) zuurstoftekort. De hersenen reageren hier onmiddellijk op door de dichtbijzijnde bloedvaten iets te verwijden waardoor er een grote hoeveelheid zuurstofrijk bloed wordt aangevoerd. Deze toestroom van zuurstofrijk bloed kan worden gemeten en vormt de basis voor het zichtbaar maken van actieve hersengebieden. Let wel, het gaat hier dus om een relatief indirecte maat: niet de hersenactiviteit zelf wordt gemeten, maar de reactie van het bloedvatstelsel op die activiteit. De op dit moment meest gebruikte methode om dit soort van veranderingen te meten is fMRI, ofwel functional Magnetic Resonance Imaging. Dit gaat als volgt: het hoofd van een proefpersoon wordt in een zeer krachtig magnetisch veld gebracht (meestal 1.5 of 3 Tesla). Door dit magnetische veld nemen de waterstofatomen in de hersenen een bepaalde configuratie aan. Wanneer dan een radiopuls (als bij een radio-zendstation) in de juiste frequentie (in “resonantie”) op de hersenen wordt afgestuurd, nemen de waterstofatomen die energie op. Stopt de radiopuls, dan schieten de waterstofatomen weer terug in hun basistoestand en geven daarbij de toegevoerde energie weer af. Dit afgifteproces, waarbij de energie weer in de vorm van radiopulsen wordt afgestaan, is verschillend voor zuurstofrijk en zuurstofarm bloed. Met fMRI worden deze verschillen gedetecteerd, en kan dus zichtbaar gemaakt worden waar de ‘geactiveerde’ hersengebieden zitten. Naast het maken van dit soort ‘functionele’ beelden is MRI ook zeer geschikt voor het maken van ‘fysieke’ beelden, zoals van de verschillende anatomische structuren in de hersenen, omdat al deze structuren

verschillen in hoe ze de opgenomen energie weer afgeven. In Figuur 1 is een MR-scanner te zien; Figuur 2 geeft een voorbeeld van de uitkomst van een fMRI-experiment (Hoeks, Pijnacker, & Stowe, in voorbereiding).



Figuur 1: De Philips Intera 3T MR scanner van het BCN NeuroImaging Center, Rijksuniversiteit Groningen (foto: Anita Kuiper).



Figuur 2: Voorbeeld van de uitkomst van een fMRI experiment (Hoeks, Pijnacker, & Stowe, in voorbereiding). In dit plaatje zijn de hersengebieden te zien die actief zijn tijdens het verwerken van grammaticaal correcte zinnen. De drie plaatjes in dit figuur laten de geactiveerde gebieden zien alsof de hersenen van glas zijn: je ziet in de plaatjes steeds dezelfde activaties, alleen vanuit verschillend perspectief. Het plaatje linksboven geeft het brein weer alsof je er van rechts tegenaan kijkt (links is achterkant van het brein, rechts is de voorkant van het brein), boven rechts is het achteraanzicht van de hersenen, het onderste plaatje laat de activaties van bovenaf zien (links is achterkant van de hersenen). SPM is de naam van het software pakket waarmee de analyses worden uitgevoerd.

Voordat fMRI in zwang kwam, is er veel gebruik gemaakt van PET (Positron Emission Tomography) om de hersenen in vivo te bestuderen (Stowe et al., 1994). Net als fMRI is PET gebaseerd op locale veranderingen in doorbloeding in de hersenen. Maar waar fMRI gebruik maakt van magnetische eigenschappen van het bloed, wordt bij PET iets aan het bloed toegevoegd waardoor veranderingen in bloedtoevoer direct traceerbaar zijn. Deze toevoeging is een zeer licht radioactieve stof, meestal een radioactieve variant van water. Radioactieve stoffen zijn instabiel, en vervallen continu naar de meer stabiele, niet-radioactieve variant. Bij dit verval worden kleine deeltjes uitgezonden, waaronder zogeheten “positronen” (met een positieve lading), die reageren met het eerste het beste “electron” (identiek aan een positron, maar met een negatieve lading) dat ze op hun pad tegenkomen. Bij elk van zulke reacties (waarbij de massa van de deeltjes volledig wordt omgezet in energie) komen precies twee fotonen vrij die een heel specifieke baan beschrijven. Wanneer deze fotonen door de PET-camera worden gedetecteerd, kan men op basis van het tijdstip en de plaats in de camera waar ze zijn waargenomen met grote precisie reconstrueren uit welk deel van de hersenen ze afkomstig zijn.

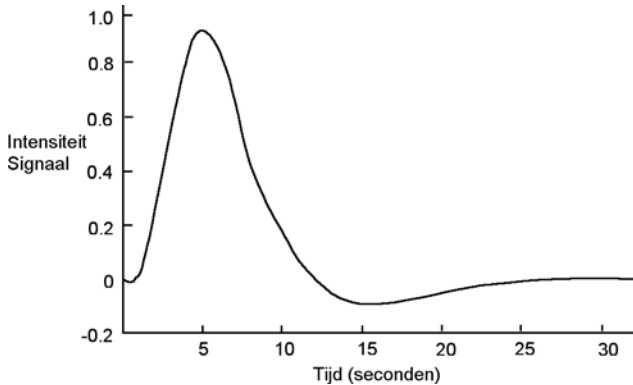
Een belangrijk nadeel van PET is dat het aantal mogelijke experimentele designs zeer beperkt is: eigenlijk is alleen een ‘geblokte’ opzet mogelijk, omdat een enkele meting bij PET relatief lang duurt (vanaf 40 seconden). Tijdens elk van deze blokken komen de stimuli steeds in dezelfde conditie voor, om zo het brein in een ‘steady state’ te brengen. In het voorbeeld van onderzoek naar dialogen zou dat dus betekenen dat er voor één van een serie metingen 40 seconden lang alleen incongruente, of alleen congruente dialogen worden aangeboden. Dit is ongewenst omdat het tot gevolg kan hebben dat proefpersonen tijdens het experiment de bedoeling gaan proberen te raden van het onderzoek, of, bewust of onbewust, strategieën gaan ontwikkelen die afwijken van hun normale taalgedrag. Daarnaast is PET-onderzoek enigszins belastend voor de proefpersoon vanwege de radioactiviteit van de stof die wordt ingespoten, maar ook vanwege de injectie zelf, die een (minimaal) risico op ontsteking van de ingespoten ader met zich meebrengt.

fMRI heeft deze nadelen niet: er hoeft niets te worden ingespoten, en gemengde onderzoeksopzettingen waarbij de condities elkaar voortdurend afwisselen (zogeheten Event-Related fMRI), zijn inmiddels de standaard manier van onderzoek geworden. Verder zijn er geen nadelige effecten bekend van kortstondige blootstelling aan sterke magnetische velden, zoals bij fMRI. Wat wel een probleem kan zijn, met name voor stimuli die auditief worden aangeboden, is het enorme lawaai dat de MR-scanner maakt tijdens het data-verzamelen: doordat elk (virtueel) plakje van de hersenen zo snel mogelijk moet worden gemeten, worden bepaalde mechanische elementen in de scanner razendsnel bewogen, hetgeen met groot lawaai gepaard gaat (110 – 130 decibel, dus iets meer dan een live rockconcert). Ondanks de verplichte oordopjes is dat geluid heel goed waarneembaar. Verder is het zo dat de proefpersoon gedurende het experiment heel stil moet blijven liggen om de metingen niet te verstoren.

Behalve deze relatief praktische problemen zijn er ook wat meer theoretische tekortkomingen van PET en fMRI. Eén daarvan is het al eerder genoemde punt dat de neurale activiteit in de hersenen niet direct wordt gemeten, maar indirect, door te kijken naar veranderingen in de doorbloeding die het *gevolg* zijn van hersenactiviteit. Dit betekent dat, hoewel de localisatie van de verandering in bloedtoevoer heel nauwkeurig vast te stellen is, we niet precies weten waar de actieve groepen van hersencellen nu zitten: de extra toevoer van zuurstofrijk bloed is niet altijd het grootst precies op de plaats van het groepje actieve cellen, maar vaker iets verderop,

aan de ‘afvoerkant’ van het bloedvat. Met andere woorden, we moeten ons er rekenschap van geven dat we feitelijk plaatjes zien van bloedvaten en niet van hersencellen.

Verder is het zo dat er soms grote verschillen zijn tussen mensen wat de vorm en de functionele indeling van hun hersenen aangaat. Omdat bij de meeste onderzoeken sprake is van groepsgemiddelden, is er geavanceerde software ontwikkeld om alle individuele hersenen ‘op elkaar te passen’ en de activatieniveaus te middelen. Deze programmatuur kan echter niet voorkomen dat er door die individuele verschillen veel variabiliteit in deze gemiddelden zit. Een ander probleem is dat, waar hersenprocessen veranderen in intervals van tientallen tot enkele honderden milliseconden (1 ms is een duizendste van een seconde), de metingen met fMRI relatief traag verlopen, in termen van meerdere seconden (om nog maar niet te spreken over PET, waarin gewerkt wordt met metingen van 40 seconden!). Dit is zo vanwege de traagheid van het bloedtoevoersysteem: pas na een seconde of zes is het te meten MR-sigitaal maximaal (zie Figuur 3). Ten slotte moet nog gemeld worden dat bepaalde delen van de hersenen, zoals gedeeltes van de frontale en temporale kwabben, moeilijk te meten zijn met fMRI, omdat ze aan schedelholtes grenzen, hetgeen bij fMRI meet-artefacten en signaalverlies kan veroorzaken.



Figuur 3: Schematische weergave van de intensiteit van het MR-sigitaal na het aanbieden van een stimulus

Tegenover de genoemde problemen van deze soort van neuroimaging staan natuurlijk ook grote voordelen, met name bij fMRI. Het belangrijkste is dat de techniek het mogelijk maakt om in te zoomen op hersengebieden die actief zijn tijdens het proces waarin de onderzoeker geïnteresseerd is. En door na te gaan in welke eerdere onderzoeken deze gebieden zijn gevonden, in welke taken en in welke condities, kan vastgesteld worden op welke manier de hersengebieden het veronderstelde proces ondersteunen. Deze tak van wetenschap werkt dus op twee vlakken tegelijkertijd: op basis van het veronderstelde proces kan een functie of een serie van mogelijke functies aan een hersengebied worden toegekend, aan de andere kant zijn onderzoekers vaak hoogst benieuwd wat er in een bepaalde conditie gebeurt, en kan op basis van wat reeds bekend is uit eerdere onderzoeken een hypothese gevormd worden over welke cognitieve processen door de experimentele manipulatie worden opgeroepen. Hoe meer onderzoek er plaatsvindt naar de functie van hersengebieden, des te nauwkeuriger zal op een gegeven moment het verband tussen locatie en functie kunnen worden vastgesteld. Voorlopig echter zijn de hersenen, met name waar het de hogere mentale

processen aangaat, nog grotendeels een terra incognita, hetgeen neuroimaging tot een zeer interessante en spannende onderneming maakt. Dat geldt overigens evenzeer voor de methoden die op basis van een ander kenmerk van neurale activiteit werken: stroom!

2.2 Stroom. Wanneer groepjes hersencellen, of ‘neuronen’, actief zijn tijdens het verwerken van informatie, zal deze activiteit in eerste instantie, voordat er ook maar sprake is van veranderingen in de bloedtoevoer, leiden tot elektrische potentiaalverschillen in de onmiddellijke omgeving van die cellen. Deze ‘primaire’ potentiaalverschillen zullen zich verplaatsen door de hersenmassa, het hersenvocht en de schedel, en uiteindelijk de hoofdhuid bereiken, waar ze inmiddels ‘secundair’ genoemd worden. De schedel is overigens bijzonder slecht in het geleiden van stroom en werkt als een soort isolatielaag waardoor de primaire potentiaalverschillen als het ware worden uitgesmeerd. Niet alle informatieverwerkingsprocessen zullen tot secundaire potentiaalverschillen leiden: alleen als er een grote hoeveelheid neuronon (1000 tot 10.000 is de schatting) die parallel aan elkaar liggen, met dezelfde frequentie actief zijn, ontstaan er primaire potentiaalverschillen die mogelijk detecteerbaar zijn op de scalp. Dit betekent dat er hersengebieden zijn die nooit tot secundaire potentiaalverschillen aanleiding geven, ook al zijn ze belangrijk voor een bepaald cognitief proces. Dit geldt met name voor dieperliggende hersenstructuren. De potentiaalverschillen die wel op de hoofdhuid aankomen zijn heel klein (we spreken hier in termen van micro-volts, ofwel miljoenste van een volt), maar kunnen desalniettemin gemeten worden. Dit wordt gedaan door bovenop de hoofdhuid metalen elektrodes te plaatsen, de weerstand tussen deze elektrodes en de hoofdhuid te minimaliseren met een zoute pasta, en de gemeten potentiaalverschillen aanzienlijk te versterken. Het gaat hier uitdrukkelijk om verschillen: de elektrische potentiaal van elke elektrode wordt vergeleken met die van zogeheten referentie-elektrodes, die gewoonlijk achter de oren of op de oorlel worden aangebracht. Wanneer de potentiaal van de gemeten elektrode hoger is dan die van de referentie-elektrode is er sprake van een positieve potentiaal en wanneer de gemeten potentiaal lager is dan de referentie-waarde is er sprake van een negatieve potentiaal. Deze polariteit (d.w.z. het positief of negatief zijn) is, zoals we zo meteen zullen zien, een belangrijk kenmerk van hersenpotentialen.

Natuurlijk zijn de hersenen met meer dingen bezig dan alleen taalverwerking: een groot aantal mentale en fysiologische processen wordt door de hersenen aangestuurd en bewaakt. Daardoor is het EEG (ElectroEncefaloGram), zoals het samenstel van continue, ritmische potentiaalveranderingen wordt genoemd, op zichzelf moeilijk te interpreteren; het vormt immers een reflectie van al deze verschillende processen. Als echter bekend is wanneer precies een bepaalde stimulus wordt aangeboden, kan de hersenrespons op die specifieke stimulus worden vastgesteld. Om de reactie op elke stimulus te ontdoen van sporen van spontane hersenactiviteit, zijn er veel herhalingen van de stimulus nodig (‘trials’); door het relatief grote aantal stimuli (minimaal 25, sommigen zeggen zelfs 40, per te meten conditie) te middelen, zullen alleen die componenten van het EEG overblijven die systematisch aan de stimulus verbonden zijn; de bijdrage van andere mentale en fysiologische processen aan het EEG die dus niet aan de presentatie van de stimulus verbonden zijn wordt uitgemiddeld. Wat we overhouden zijn ‘stimulus-gebonden’ potentialen ofwel ERP’s (Event-Related brain Potentials). Dus door het EEG over een groot aantal trials te middelen, komen we tot een benadering van de reactie van het brein op dit specifieke soort van stimulus.

Het is belangrijk om op te merken dat de verdeling van de secundaire potentialen vaak volstrekt anders is dan die van de primaire voltages, doordat de weefsels in het hoofd verschillen

in hoe ze stroom geleiden. Met name het bot in de schedel zorgt voor het uitsmeren van primaire potentiaalverschillen. Bovendien is het zo dat de hersenen niet homogeen zijn, maar aan de oppervlakte ('cortex') juist bestaan uit talloze windingen ('gyri') en groeven ('sulci'). Omdat één en hetzelfde secundaire ERP-patroon zoals gemeten op de schedel op meerdere manieren gerealiseerd kan zijn in termen van primaire potentialen, is het onmogelijk om vanuit de geobserveerde schedelverdeling van een bepaald effect terug te redeneren naar de bron van dat effect. Dit wordt het 'inverse-probleem' genoemd. Dit probleem wordt verergerd doordat het aantal primaire bronnen voor een secundair verschijnsel niet uit het gemeten ERP-patroon kan worden afgeleid, en ook vanwege het eerder genoemde feit dat het hoofd geen perfecte bol is met een perfect homogene inhoud. Hierdoor zijn ERP's nauwelijks geschikt voor de localisatie van actieve hersengebieden, zeker niet vergeleken met bijvoorbeeld PET of fMRI. Desalniettemin is er software beschikbaar die, gegeven een aantal assumpties over vorm van de hersenen en een schatting van het aantal betrokken hersengebieden, hypothesen oplevert over de oorsprong van de gemeten ERP-effecten (zogenoeten 'bronanalyse').

De ERP's die na middeling kunnen worden vastgesteld bestaan uit een serie pieken en dalen die benoemd worden op basis van de polariteit (dus P voor positief en N voor negatief), de volgorde waarin ze optreden (bijv. P1, P2, P3, etc.) of ook het moment waarop ze maximaal zijn, gemeten na de aanbieding van de stimulus, zoals de P300, een positieve piek die maximaal is rond 300 milliseconden nadat een stimulus is aangeboden. In de laatste vijftig jaar is er een aantal componenten geïdentificeerd die met name van belang zijn voor onderzoek naar taal. Wat hierbij opgemerkt moet worden is dat het hier niet specifiek gaat om de piek of het dal in het ERP patroon, maar vooral om de modulatie van de desbetreffende piek of het desbetreffende dal: het gaat dus met name om ERP-effecten, ook wel verschuivingen (of 'shifts') genoemd, waarbij minimaal twee condities worden vergeleken: vaak een normale controleconditie en een conditie met een taalkundige manipulatie (bijv. een syntactische, semantische of pragmatische schending).

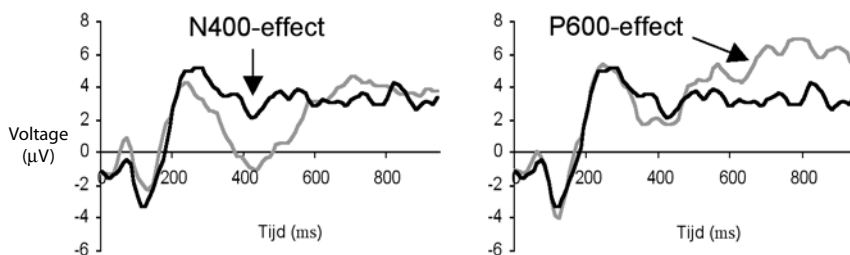
Hieronder bespreken we een selectie van de belangrijkste taalgerelateerde ERP-effecten, te weten de "ELAN", de "LAN", de "N400" en de "P600"; voor een complete lijst wordt de geïnteresseerde lezer verwezen naar een zeer recent overzichtsartikel van Kutas, Van Petten en Kluender (te verschijnen; zie ook Van Petten, in druk, voor een beschrijving van 'MEG' (Magneto-EncefaloGrafie) een methode die nauw verwant is aan EEG-meting).

ELAN

De ELAN of Early Left Anterior Negativity is, zoals de naam al zegt, een vroege verschuiving in de ERP's die maximaal is aan de linkerkant van de schedel, met name op frontale elektrodes (d.w.z., aan de voorkant van het hoofd, bijv. F7, zie Figuur 5). Dit effect treedt op tussen 100 en 300 ms na het kritische woord, en wordt gevonden wanneer in een taaluiting een woord van een bepaalde categorie wordt verwacht, maar een woord van een andere categorie wordt aangeboden (Friederici, 1995; Neville, Nicol, Barss, Forster, & Garrett, 1991). Friederici, Pfeifer, en Hahne (1993), bijvoorbeeld, vonden een ELAN in "Der Freund wurde im besucht" (letterlijk: 'De vriend werd in_het bezocht'), waarin een zelfstandig naamwoord wordt verwacht na "im", maar waar een werkwoord wordt gepresenteerd (controle-zin: "Der Freund wurde besucht"; lett., 'De vriend werd bezocht'). Om nog onduidelijke redenen is niet in alle onderzoeken bij dit soort van schendingen een ELAN gevonden.

LAN

Net als de ELAN is de LAN (Left Anterior Negativity) een negatieve verschuiving in het ERP-signaal dat optreedt in reactie op syntactische schendingen. De LAN is vooral zichtbaar op (links) frontale elektrodes, zo'n 400 ms na het aanbieden van de kritische stimulus. Gunter, Stowe en Mulder (1997) vonden (onder andere) een LAN als respons op een morfo-syntactische schending in zinnen als “De kleine drenkeling werd door de held redden”, vergeleken met de correcte versie “De kleine drenkeling werd door de held gered”. Hier gaat het dus niet, als bij de ELAN, om een schending van de syntactische categorie, maar om een schending van de correcte vorm van het woord.



Figuur 4: Voorbeelden van de bekendste ERP-effecten: het N400-effect (linkerpaneel) en het P600-effect (rechterpaneel): de grijze lijn is steeds de conditie met de schending; de zwarte lijn is de normale controle-conditie; het gaat hier om de electrode Pz.

N400

De N400 is een negatieve verschuiving die optreedt wanneer in een taaluiting waarin een semantische schending voorkomt, zoals “Het brood werd door de bakkers bedreigd”, wordt vergeleken met een semantisch correcte controlezin zoals “Het brood werd door de bakkers gebakken”. Zie Figuur 4, linkerpaneel, voor een grafische weergave van deze vergelijking (bron: Hoeks & Stowe, in voorbereiding). De negatieve verschuiving ten opzichte van de normale zin is maximaal rond 400 ms na aanbieding van het kritische woord (hier: “bedreigd”) en is met name geprononceerd op elektrodes in het midden en aan de achterkant van het hoofd (Kutas & Hillyard, 1980). In sommige omstandigheden wordt er geen N400 gevonden, ook al is de taaluiting semantisch anomaal, zoals in “Het brood heeft de bakkers gebakken”; dit wordt ook wel een ‘semantische illusie’ genoemd (Hoeks, Stowe, & Doedens, 2004). De N400 komt sterk overeen met de LAN, maar verschilt in de plaats op het hoofd waar het effect maximaal is: zoals gezegd is de N400 het grootst in het midden van het hoofd en achterop, terwijl de LAN juist aan de voorkant (‘frontaal’) de grootste uitslag heeft.

P600

De P600 is een positieve verschuiving die optreedt wanneer een taaluiting waarin een syntactische schending voorkomt, zoals “Het verwende kind gooien het speelgoed op de grond”, wordt vergeleken met een syntactisch correcte controlezin zoals “Het verwende kind gooit het speelgoed op de grond” (Hagoort, Brown, & Groothusen, 1993). De positieve verschuiving die door de schending wordt opgeroepen is maximaal rond 600 ms na aanbieding van het kritische woord (hier: “gooien”) en is net als de N400 met name zichtbaar

op elektrodes nabij het midden en de achterkant van het hoofd. Over het algemeen wordt aangenomen dat de P600 eerder een reflectie is van syntactische heranalyse dan van syntactische foutdetectie. Daarom kan uit het feit dat de N400 rond 400 ms maximaal is en de P600 pas rond 600 ms niet worden afgeleid dat semantische verwerking aan alle soorten van syntactische verwerking voorafgaat (zoals het vaststellen van de syntactische structuur van een taaluiting). Zie Figuur 4 (rechterpaneel) voor een voorbeeld van de P600 (bron: Hoeks & Stowe, in voorbereiding).

ERP's kunnen de onderzoeker dus waardevolle aanwijzingen geven over wat er gebeurt in de hersenen van de taalgebruiker: ze laten zien of er al dan niet een reflectie is van de experimentele manipulatie en vaak ook wat voor soort proces door de manipulatie is opgeroepen. Met name tijdsinterval, grootte en schedelverdeling van het effect zijn belangrijke parameters om te bepalen met wat voor ERP-effect men van doen heeft (zie Van Berkum, 2004, voor een uiteenzetting van welke conclusies men wel en niet kan trekken op basis van ERP-data). In Tabel 1 worden de belangrijkste kenmerken van ERP-registratie, PET en fMRI nog eens bij elkaar gezet.

Tabel 1: Kenmerken van de verschillende neuroimaging methoden

	<i>fMRI</i>	<i>PET</i>	<i>ERP's</i>
<i>Signaal intensiteit hangt af van</i>	concentratie van zuurstofrijk bloed	absolute hoeveelheid bloed	potentiaalverschillen op schedel door neuronale activiteit
<i>Nauwkeurigheid (Tijd)</i>	ca. 6 s	ca. 40 s	ca. 1-10 ms
<i>Nauwkeurigheid (Plaats)</i>	ca. 1-4 mm	ca. 5-10 mm	onbekend
<i>Experimenteel Design</i>	geblokt en event-related	geblokt	geblokt en event-related
<i>Meetfouten</i>	- daadwerkelijk actief hersengebied kan tientallen mm verwijderd zijn van gemeten activiteit door kenmerken van bloedvatstelsel - beweging van de proefpersoon tijdens de sessie leidt tot ruis - bepaalde delen frontale en temporale kwab moeilijk te meten door effecten van naastgelegen schedelholten - ruis door verschillen in vorm en functie van hersengebieden bij verschillende proefpersonen	- daadwerkelijk actief hersengebied kan tientallen mm verwijderd zijn van gemeten activiteit door kenmerken van bloedvatstelsel - ruis door verschillen in vorm en functie van hersengebieden bij verschillende proefpersonen	- niet alle hersenactiviteit leidt tot (op de schedel) meetbare potentiaalverschillen - oogbewegingen en knippering leiden tot ruis - ruis door verschillen in vorm en functie van hersengebieden bij verschillende proefpersonen
<i>Ongemak voor Proefpersoon</i>	- moet stilliggen in een 'tunnel' - krijgt gehoorbeschermers tegen lawaai	- krijgt radio-actief materiaal ingespoten - mag niet te veel bewegen	- mag niet met de ogen knippen of de ogen bewegen tijdens bepaalde gedeeltes van de meetsessie - krijgt elektroden op het hoofd en op het gezicht

In het onderstaande experiment gebruiken we ERP-metingen om vast te stellen hoe lezers reageren op pragmatische schendingen van het Griceaanse Maxime van Hoeveelheid: vinden we effecten die gerelateerd zijn aan verschillende aspecten van syntactische verwerking (ELAN, LAN, P600-effect), of heeft de schending een duidelijk effect op de semantische verwerking (N400-effect)? In ieder geval zijn er tot op heden geen ERP-effecten gevonden die specifiek zijn voor pragmatische verwerking (Coulson, 2004).

3. Experiment

3.1 Proefpersonen. Aan het experiment deden 18 studenten (waarvan 6 mannen) van de Rijksuniversiteit Groningen mee. De deelnemers waren tussen de 18 en de 29 jaar, met een gemiddelde leeftijd van 20 jaar. De meesten kregen betaald voor deelname, een aantal kreeg studiepunten. Alle proefpersonen waren rechtshandig¹ en hadden Nederlands als moedertaal.

3.2 Materiaal. Het materiaal voor het experiment werd geconstrueerd op basis van 60 zinnen die een NP-coördinatie bevatten (ontleend aan Hoeks, 1999), zie (5) voor een voorbeeld:

5. De burgemeester prees het raadslid en de wethouder uitbundig.

Op het moment dat de lezer het woord ‘wethouder’ tegenkomt, dus nog zonder het laatste woord te hebben gezien, zijn er twee interpretaties mogelijk: enerzijds NP-coördinatie, als in (6) en anderzijds zinscoördinatie, als in (7).

6. De burgemeester prees [het raadslid en de wethouder].

7. [De burgemeester prees het raadslid] en [de wethouder (lachte)].

Zoals hierboven vermeld, is bekend dat lezers een voorkeur hebben voor NP-coördinatie (bijv., Hoeks, 1999; Hoeks et al., 2006). Dit betekent dat in zin (5) het bijwoord in laatste positie (‘uitbundig’) op zich geen verwerkingsproblemen zou moeten opleveren, het desambigueert de zin namelijk naar NP-coördinatie, hetgeen de voorkeursinterpretatie is.

Op basis van elk van deze 60 zinnen werden drie dialogen gemaakt. In de eerste versie van de dialoog werd de zin voorafgegaan door een neutrale vraag: “Wat gebeurde er?”. In de tweede versie werd de vraag zodanig geconstrueerd dat die zeer specifieke verwachtingen op zou roepen omtrent de topic-structuur van de daaropvolgende zin: “Wat deden de burgemeester en de wethouder?”. Door de vraag op deze manier te stellen worden zowel ‘de burgemeester’ als ‘de wethouder’ zeer waarschijnlijke topics in het antwoord op de vraag (zoals ook aangetoond in Hoeks et al., 2002). De derde versie is hier niet relevant, maar wordt voor de compleetheit toch in het rijtje hieronder weergegeven. Samengevat hebben we dus drie condities, waarvan alleen de eerste twee van belang zijn.

8. Congruente conditie

V: Wat gebeurde er?

A: De burgemeester prees het raadslid en de wethouder uitbundig.

9. Incongruente conditie

V: Wat deden de burgemeester en de wethouder?

A: De burgemeester prees het raadslid en de wethouder uitbundig.

10. Extra conditie (wordt verder niet besproken)

V: Wat gebeurde er?

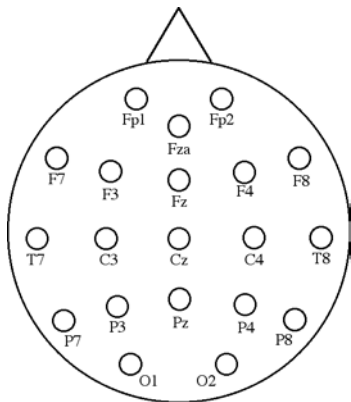
A: Na de vergadering prees de burgemeester de wethouder uitbundig.

Merk op dat het bijwoord ('uitbundig') hier een cruciale rol speelt. Het maakt immers duidelijk dat de antwoordzin slechts één topic heeft ('de burgemeester'), en niet twee, zoals verwacht in de incongruente conditie (9). De NP 'de wethouder' blijkt geen topic te zijn in de antwoordzin, wat een schending van het Maxime van Hoeveelheid inhoudt.

Naast de experimentele dialogen zijn er ook zogeheten 'fillers' of 'opvulstimuli' die moeten verhinderen dat proefpersonen tijdens het experiment ongewenste strategieën ontwikkelen of de bedoeling van het onderzoek gaan raden. Zo zijn er 40 dialogen waarin zinnen met zinscoördinaties voorkomen om een zeker tegenwicht te geven aan de overvloed aan dialogen met NP-coördinaties (60 in totaal). Daarnaast zijn er ook 100 ongerelateerde fillers die bij een apart onderzoek horen.² Ten slotte is er een oefenset van 40 dialogen gemaakt om de proefpersonen te laten wennen aan de materialen.

3.3 Opzet. Er werden drie lijsten van stimuli gemaakt. In elk van die lijsten waren er 20 stimuli voor elk van de drie experimentele condities (d.w.z., 20 congruent, 20 incongruent en 20 extra), en elke lijst bevatte slechts één versie van een bepaalde stimulus. De volgorde van experimentele en filler-dialogen (zie paragraaf 3.2, Materiaal) was hetzelfde voor alle lijsten en werd door de experimentator quasi-willekeurig vastgesteld (maximaal drie experimentele stimuli na elkaar, maar nooit twee of meer opeenvolgende stimuli in dezelfde conditie). Elke lijst werd aan een gelijk aantal proefpersonen aangeboden (hier: 6) en elke proefpersoon kreeg slechts één lijst te zien. Elke lijst bevatte 200 stimuli (60 experimentele stimuli, 40 fillers met zinscoördinatie en 100 stimuli van het ongerelateerde experiment).

3.4 Procedure. Tijdens de voorbereidingsfase werden bij een proefpersoon de elektrodes aangebracht middels een elastische 'cap' met 20 elektrodes. Zie Figuur 5 voor een grafische weergave van de positie van deze elektrodes. Tijdens het experiment zaten de proefpersonen op een comfortabele stoel achter een computerscherm (op ongeveer 60 cm afstand) in een cabine die afgeschermd was voor geluid en elektrische signalen. Alle dialogen werden woord-voor-woord, visueel gepresenteerd in het midden van het scherm. Een woord was 243 ms zichtbaar en werd gevolgd door 243 ms leeg scherm. Tussen de vraagcontext (bijv. "Wat gebeurde er?") en het daaropvolgende antwoord zat 729 ms. Elke dialoog werd voorafgegaan door een sterretje (*) dat één seconde zichtbaar was en dat aangaf dat een nieuwe dialoog zou beginnen. Aan het einde van de zin werd het woord 'knipper' vertoond (gedurende drie sec.) dat aangaf dat de proefpersoon met de ogen kon knipperen; proefpersonen werd gevraagd om niet te knipperen gedurende de presentatie van de dialogen, omdat dit de metingen ernstig zou kunnen verstoren. Om de proefpersonen aandachtig te laten lezen werden er verspreid over het experiment 35 vragen gesteld over de inhoud van de voorafgaande dialoog (ongeveer één vraag per zes dialogen). Deze vraag kon met 'ja' of 'nee' beantwoord worden met behulp van twee knoppen op een knoppenkast. Elke vraag werd



Figuur 5: Plaatsing van de elektrodes op het hoofd; het driehoekje aan de bovenkant stelt de neus van de proefpersoon voor.

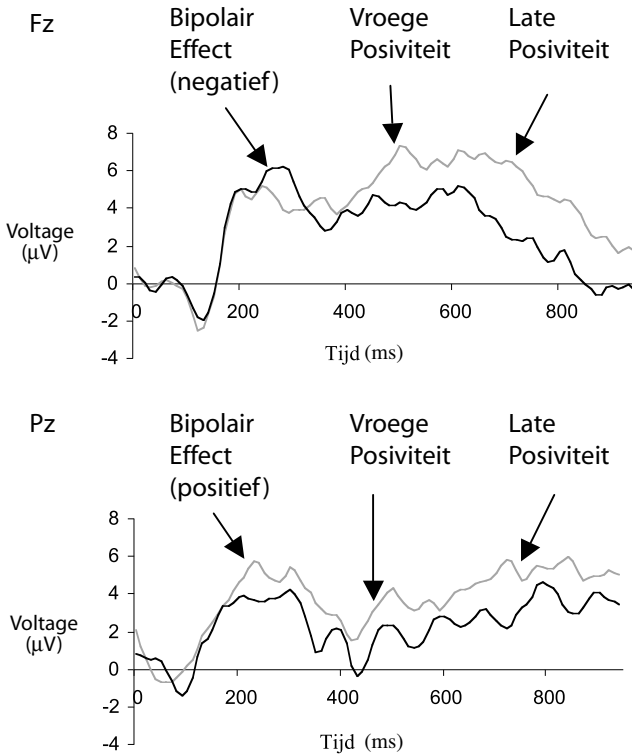
gevolgd door een filler-dialogoog, om te voorkomen dat mogelijke problemen bij het beantwoorden van de vragen de verwerking van de experimentele dialogen zouden beïnvloeden. Gedurende het experiment kreeg de proefpersoon gelegenheid om te pauzeren na elk blok van 50 dialogen. Het totale experiment duurde ongeveer een uur en drie kwartier, inclusief voorbereiding.

3.5 Resultaten. Gemiddeld werd 85% ($SD = 5.6$) van alle inhoudsvragen (zie paragraaf 3.4, Procedure) correct beantwoord, hetgeen aangeeft dat de proefpersonen met volle aandacht aan het experiment hebben meegedaan. Op basis van visuele inspectie van de data zijn drie tijdsintervallen uitgekozen voor verdere analyse: 1) 180–320 ms na aanbidding van het kritische woord (i.e., het bijwoord); 2) 350–550 ms na het kritische woord; 3) 550–750 ms na het kritische woord.³ Over elk van de drie gekozen tijdsintervallen is een gemiddeld voltage berekend voor elke conditie van elke proefpersoon op elke elektrode. Ten behoeve van de statistische analyse zijn de 20 elektrodes daarna in drie sets verdeeld: 1) de elektrodes Fp1, Fza en Fp2 voor de ‘prefrontale’ analyse; 2) de elektrodes O1 en O2 voor de ‘occipitale analyse’; en ten slotte 3) de 15 overige elektrodes voor de belangrijkste analyse, die we de ‘basisanalyse’ zullen noemen (zie Figuur 5).

In elk van deze drie sets van elektrodes wordt niet alleen gezocht naar een effect van de inhoudelijke factor Conditie (incongruent vs. congruent), maar wordt ook gekeken naar de topografie oftewel ‘schedelverdeling’ van dat effect: waar is het het grootst, waar minder groot, en waar is het afwezig? In de basisanalyse wordt de plaats van de elektrodes geanalyseerd met behulp van twee topografische factoren, te weten Anterioriteit (met de niveaus: frontaal (F7, F3, Fz, F4 en F8) vs. centraal (T7, C3, Cz, C4, T8) vs. posterieur (P7, P3, Pz, P4, en P8)) en Lateraliteit (met de niveaus: uiterst links (F7, T7 en P7) vs. links (F3, C3 en P3) vs. midden (Fz, Cz, en Pz) vs. rechts (F4, C4 en P4) vs. uiterst rechts (F8, T8 en P8)).

Bij de analyses op prefrontale elektrodes heeft de factor Lateraliteit drie niveaus (links (Fp1) vs. midden (Fza) vs. rechts (Fp2)); bij de occipitale elektrodes heeft Lateraliteit twee niveaus (links (O1) vs. rechts (O2)); de factor Anterioriteit is bij deze laatste twee analyses niet van toepassing, omdat in elk van deze twee sets de elektrodes niet in anterioriteit verschillen. Omdat alleen effecten waarin de factor Conditie voorkomt iets zeggen over onze pragmatische manipulatie, zullen andere (puur topografische) effecten hier niet verder wor-

den besproken. In Figuur 6 worden de belangrijkste effecten getoond op twee elektroden (Fz en Pz).



Figuur 6: ERP-golfs patronen voor incongruente (grijs) en congruente dialogen (zwart) op een frontale (Fz) en een posterieure elektrode (Pz).

Interval 1 (180-320 ms post-onset). Het hoofdeffect van Conditie en het interactie-effect tussen Conditie en Lateraliteit waren niet significant (F -waarden < 1). De interactie Conditie $\times$ Anterioriteit was dat wel ($F(2,30) = 5.34, p < .05$), maar werd gekwalificeerd door de significante drieweg-interactie tussen Conditie, Anterioriteit en Lateraliteit ($F(8,120) = 2.22, p < .05$). Vervolganalyses lieten zien dat er interacties waren tussen Conditie en Anterioriteit op elk niveau van Lateraliteit, met uitzondering van de uiterst rechterkant (uiterst links: $F(2,30) = 3.20, p = .069$; links: $F(2,30) = 7.36, p < .01$; midden: $F(2,30) = 8.72, p < .01$; rechts: $F(2,30) = 3.18, p = .069$; uiterst rechts $F < 1$). In elk van deze interacties was er sprake van een negativiteit (incongruent negatiever dan congruent) op de frontale elektrodes en een positiviteit (incongruent positiever dan congruent) op de posterieure elektrodes. De centrale elektrodes zaten daar tussenin. Zie Tabel 2 voor een weergave van de effectgroottes en de bijbehorende significanties.

Tabel 2: Effectgroottes (incongruent minus congruent, in μV) voor frontale, centrale en posterieure elektrodes op elk niveau van Lateraliteit. In grijs gemarkeerde cellen is het effect van Conditie (marginaal) significant ($p < .10$).

	Uiterst links	Links	Midden	Rechts	Uiterst rechts
Frontaal	-0.8	-0.7	-0.9	-0.5	-0.4
Centraal	0.0	0.4	0.5	0.2	0.2
Posterieur	0.9	1.1	1.2	0.6	0.1

De analyse op de prefrontale elektrodes leverde geen significante resultaten op; de analyse op de occipitale elektrodes liet een significant hoofdeffect van Conditie zien ($F(1,15) = 5.53$, $p < .01$), waarbij de incongruente conditie positiever was dan de congruente conditie (een verschil van $0.81 \mu V$); er was geen interactie met Lateraliteit ($p > .25$). Er is dus sprake van een vroege, bipolaire component, bestaande uit een negativiteit aan de voorkant van het hoofd, gekoppeld aan een positiviteit aan de achterkant, die met name zichtbaar was op de elektrodes in het midden en links op de schedel, maar afwezig aan de uiterste rechterkant.

Interval 2 (350-550 ms post-onset). In de basisanalyse was er een significant hoofdeffect van Conditie ($F(1,15) = 5.95$, $p < .05$), met een grotere positiviteit voor de incongruente conditie vergeleken met de congruente conditie (een verschil van $1.27 \mu V$). Er was geen interactie met de topografische factoren Anterioriteit en Lateraliteit (F -waarden < 1). In de analyse op de prefrontale elektrodes was ook alleen het hoofdeffect van Conditie significant (een verschil van $1.82 \mu V$; $F(1,15) = 7.61$, $p < .05$). Er waren geen significante effecten in de analyse op de occipitale elektrodes (alle p -waarden $> .19$). De pragmatische incongruentie veroorzaakte in dit interval dus een significante positiviteit, die overal op het hoofd te zien was, behalve aan de achterkant.

Interval 3 (550-750 ms post-onset). Wat betreft het derde interval werd er een hoofdeffect gevonden van Conditie ($F(1,15) = 7.99$, $p < .05$), met een grotere positiviteit voor de incongruente conditie vergeleken met de congruente conditie (een verschil van $1.89 \mu V$). Er was geen interactie met Anterioriteit ($F < 1$), maar de interactie met Lateraliteit was marginaal significant ($F(4,60) = 2.22$, $p = .10$). Deze interactie werd echter gekwalificeerd door een significante drieweg-interactie tussen Conditie, Anterioriteit en Lateraliteit ($F(8,120) = 2.13$, $p = .05$). Die interactie werd veroorzaakt doordat het effect van Conditie (hier: incongruent positiever dan congruent) met name groot is op de linkse en uiterst linkse elektrodes; in het midden en rechts is de positiviteit vooral op frontale elektrodes te zien, en uiterst rechts is het effect van Conditie zelfs afwezig. Zie Tabel 3 voor een weergave van effectgroottes en significanties.

Tabel 3: Effectgroottes (incongruent minus congruent, in μV) voor frontale, centrale en posterieure elektrodes op elk niveau van Lateraliteit. In grijs gemarkeerde cellen is het effect van Conditie (marginaal) significant ($p < .10$).

	Uiterst links	Links	Midden	Rechts	Uiterst rechts
Frontaal	2.1	3.2	2.6	2.3	1.1
Centraal	2.3	2.1	1.2	1.4	1.3
Posterieur	2.3	2.7	1.9	1.3	0.5

In de prefrontale analyse was alleen het hoofdeffect van Conditie significant: de incongruente conditie was positiever dan de congruente (een verschil van $2.82 \mu\text{V}$; $F(1,15) = 11.65$, $p < .005$). Uit de analyses van de occipitale elektrodes bleek ten slotte dat incongruentie een positief verschil veroorzaakte aan de linkerkant (O1: $0.54 \mu\text{V}$) en een negatief verschil aan de rechterkant (O2: $-0.16 \mu\text{V}$). Deze interactie was marginaal significant ($F(1,15) = 3.64$, $p = .076$). Dus ook in dit laatste interval was de incongruente conditie positiever dan de congruente conditie. Deze positiviteit was het grootste aan de linkerkant van het hoofd, en was met name zichtbaar op frontale en prefrontale elektrodes.

Samengevat kunnen we stellen dat de schending van pragmatische verwachtingen in deze mini-dialogen een zeer duidelijke weerslag op het ERP-golfpatroon had, en wel op drie verschillende momenten. Allereerst was er een relatief vroeg effect te zien dat al na 180 ms na het verschijnen van het kritische woord optrad. Deze vroege respons was een *bipolair effect* – een frontale negativiteit gekoppeld aan een posterieure positiviteit – dat met name zichtbaar was op de elektrodes in het midden en links op de schedel (het was afwezig aan de uiterste rechterzijde). Dit bipolaire effect werd gevolgd door een tweede effect: een positiviteit die algemeen op de schedel te zien was (behalve op de occipitale elektrodes). Dit effect was te zien tussen 350 en 550 ms na het kritische woord. Deze *vroege positiviteit* mondde uiteindelijk uit in een *late positiviteit* (550–750 ms post-onset) die het meest geprononceerd was aan de linkerzijde van de schedel, en die met name groot was op de frontale en de prefrontale elektrodes.

3.6 Discussie. We hebben gezien dat het Griceaanse Maxime van Hoeveelheid van groot belang is voor de verwerking van taaluitingen. Nog geen 180 ms nadat duidelijk is geworden dat het Maxime is geschonden is er al een significante verschuiving in het ERP-sig-naal te zien van de incongruente conditie ten opzichte van de congruente conditie. Op frontale elektrodes (d.w.z., aan de voorzijde van de schedel), is deze verschuiving negatief, op meer posterieure elektrodes (aan de achterkant) is de verschuiving positief. Dit vroege bipolaire effect in het ERP-patroon – dat volgens ons nog niet eerder gevonden is – is hoogstwaarschijnlijk een weerslag van de detectie van een pragmatische fout. De taalgebruiker verwacht dat de zojuist gepresenteerde entiteit (hier, ‘de wethouder’) een actie uit gaat voeren. Zodra het desambiguerende bijwoord wordt gepresenteerd blijkt dat deze verwachting niet klopt, hetgeen zich dan manifesteert als een pragmatische ‘foutmelding’. Men zou kunnen tegenwerpen dat dit niet een puur pragmatisch effect is, maar juist een syntactisch effect dat wordt geïnstigeerd door de pragmatische manipulatie. Immers, de taalgebruiker verwacht naar alle waarschijnlijkheid een werkwoord (als in: “... en de wethouder lachte”), maar krijgt een woord van een geheel andere grammaticale klasse, namelijk een bijwoord (bijv., “uitbundig”). Echter, in dat geval zouden we een effect verwachten dat optreedt bij schendingen van syntactische categorie, namelijk de ELAN (Early Left Anterior Negativity; Friederici, 1995). In het huidige onderzoek vinden we echter naast de negativiteit ook duidelijk een positiviteit. Het lijkt er dus op dat we een uniek, pragmatisch effect hebben geïdentificeerd dat optreedt wanneer pragmatische verwachtingen, in het huidige onderzoek gecreëerd op basis van het Maxime van Hoeveelheid, geschonden worden.

Na het vroege bipolaire effect vinden we een relatief vroege positiviteit in het tijdsinterval waar normaliter de N400 voorkomt (350–550 ms post-onset). Deze positiviteit komt op bijna alle elektrodes voor behalve helemaal achter op de schedel, bij de occipitale elektrodes.

Dit effect vertoont grote overeenkomsten met een vroege positiviteit zoals gevonden door Bornkessel, Schleewsky en Friederici (2002). Volgens Bornkessel et al. was deze positiviteit een weerspiegeling van ‘thematische’ heranalyse, waarbij de thematische rol die in eerste instantie aan een discourse-entiteit was toegekend, uiteindelijk niet bleek te passen. In het huidige onderzoek zou zeker ook een thematische heranalyse op kunnen treden omdat de ambigue NP (hier: ‘de wethouder’) pragmatisch gezien een AGENS (de entiteit die een handeling uitvoert) zou moeten zijn, maar uiteindelijk een PATIENS (de entiteit die een handeling ondergaat) blijkt te zijn. Het zou dus deze thematische verwarring kunnen zijn geweest die tot de vroege positiviteit in het tweede tijdsinterval heeft geleid.

Ten slotte vinden we na het vroege bipolaire en het daaropvolgende positieve effect een grote positieve verschuiving voor de incongruente dialogen dat gezien het tijdsverloop sterk aan een P600 doet denken. De schedelverdeling wijkt enigszins af van die van de ‘standaard’ P600: in plaats van een brede verdeling op de achterkant van de schedel vinden we hier een enigszins naar links verschoven positiviteit die met name op de frontale en prefrontale elektrodes erg sterk is. Toch is het waarschijnlijk dat we hier met een lid van de P600-familie te maken hebben: de P600 is namelijk een universele reactie op een syntactische en een semantische anomalie, dus waarom niet op een pragmatische anomalie? De taalgebruiker doet haar uiterste best om alle taaluitingen te begrijpen en investeert over het algemeen veel om tot een coherente betekenis van een uiting te komen, hetgeen mogelijk tot uiting komt in de duur en intensiteit van de P600 (Hoeks et al., 2004). De reden voor deze ‘investering’ is dat de lezer/luisteraar ervan uitgaat dat de schrijver/spreker een rationeel wezen is dat waarheidsgetrouwe uitspraken doet, zoals ook door Grice (1975) beschreven. Wanneer taalgebruikers dus met taaluitingen worden geconfronteerd die niet helemaal kloppen, zullen ze alles doen om tot een coherente representatie van hetgeen gezegd wordt te komen, met als laatste optie de syntactische heranalyse die zich manifesteert als een P600. Dit zou voor schrijvers en sprekers een les moeten zijn: schendingen van Griceaanse maxims kosten uw lezers en toehoorders meer cognitieve energie dan hen lief is.

4. Is neuroimaging de moeite waard?

Neuroimaging methoden hebben ontegenzeggelijk een aantal praktische en theoretische tekortkomingen. Bij ERP-registratie mag een proefpersoon niet met de ogen knippen, moet ze rechtuit kijken om oogbewegingsartefacten te voorkomen, en krijgt ze geschreven taal vaak woord-voor-woord in het midden van een computerscherm aangeboden, hetgeen niet echt ‘natuurlijk’ genoemd kan worden. Bovendien zijn ERP’s nauwelijks geschikt voor de localisatie van de bron van cognitieve processen, komt niet alle hersenactiviteit in het EEG terecht, en is het correct interpreteren van effecten vaak moeilijk. Bij fMRI moet de proefpersoon lange tijd liefst helemaal stilliggen terwijl ze, omgeven door lawaai, met haar hoofd in een soort tunnel ligt. Verder is fMRI relatief ongevoelig voor het tijdsverloop van cognitieve processen, komt het gemeten signaal niet van actieve zenuwcellen maar van bloedvaten, zijn er grote verschillen tussen mensen in vorm en functionele indeling van hun hersenen, en wordt fMRI bovendien geplaagd door meetartefacten bij frontale en temporale hersengebieden.

Voor beide methoden, maar het meest voor ERP’s, geldt dat er veel herhalingen van stimuli in dezelfde conditie nodig zijn om een helder signaal te verkrijgen. Het effect van

deze herhaling op de proefpersonen, en uiteindelijk op de uitkomsten van het onderzoek is onbekend. Bovendien is het zo dat onder andere de eis om veel stimuli te gebruiken het doen van dit soort experimenten behoorlijk kostbaar maakt, met name als het gaat om experimentele stimuli die het woord- en het zinsniveau overstijgen, zoals in onderzoek naar pragmatische en discourse-gerelateerde verschijnselen. Het construeren en valideren van materiaal bestaand uit kleine verhaaltjes kan gemakkelijk een half jaar in beslag nemen (Van Berkum, 2004). En niet alleen het voorbereiden van een experiment is een tijdrovende bezigheid. Ook in het verkrijgen van de data (het 'draaien' van het experiment), en in de uiteindelijke analyse van de verkregen data kan behoorlijk veel tijd gaan zitten. Dan rijst de vraag: is het doen van dit soort onderzoek, met deze methoden, de moeite wel waard?

De uitkomst van het experiment dat hierboven beschreven is, geeft eigenlijk al een antwoord op die vraag. Immers, het laat zien hoe snel een pragmatische fout wordt gedetecteerd, en geeft ook een indicatie van welke processen op deze detectie volgen. De interpretatie van de gevonden ERP-effecten ligt niet vast, en kan veranderen naarmate we meer kennis vergaren over de relatie tussen ERP-effecten (of actieve hersengebieden) enerzijds en cognitieve verschijnselen anderzijds. Maar het is wel zo dat deze fijnmazigheid van resultaten niet kan worden bereikt in paradigma's als 'drukknop-lezen' of 'oogbewegingsregistratie'. De resultaten van dit soort experimenten hebben slechts één dimensie: die van de tijd. Een proces duurt kort of het duurt lang. Het is natuurlijk wel zo dat in oogbewegingsexperimenten ook andere maten te berekenen zijn, zoals het aantal regressies (waarbij de ogen terugspringen naar eerder gelezen gedeeltes van de tekst) of het aantal keren dat een lezer een woord of een tekstgedeelte overslaat (zie Cozijn en Lentz, Mak & Pander Maat in dit nummer). Wat voor soort cognitieve processen echter aan verschillen in tijdsduur (of regressies) ten grondslag liggen, en in welke volgorde deze processen plaatsvinden is vrijwel onmogelijk vast te stellen. Desondanks zijn dit soort experimenten cruciaal voor het begrijpen en het juist interpreteren van neuroimaging onderzoek. Met name in ERP-onderzoek is het niet mogelijk om puur op basis van de gevonden patronen af te leiden of een bepaalde conditie nu moeilijker is dan een andere, of andersom, hetgeen tot een volstrekt verkeerde interpretatie kan leiden (zie Hoeks, Stowe, & Wunderink, 2004, waarin aangetoond wordt dat de negatieve verschuiving die door Münte, Schiltz, & Kutas, 1998, werd gevonden geen moeilijkheid, maar juist gemak van verwerking representeert). Ook in fMRI onderzoek is het niet altijd duidelijk welke stimuli nu moeilijker zijn en op welke dimensie. Het is dus essentieel om neuroimaging en reactietijd / oogbewegingsonderzoek te combineren. Dat is ook de reden dat we op dit moment bezig zijn het huidige experiment te herhalen met reactietijdmetingen.

Noten

1. Het is bekend dat de organisatie van de taalgebieden in de hersenen anders is voor rechtshandigen dan voor linkshandigen. Terwijl bij rechtshandigen in overgrote meerderheid de linkerhersen helft in taal is gespecialiseerd, is dat bij linkshandigen vaak andersom of is er minder specialisatie. Om te voorkomen dat deze verschillen teveel ruis veroorzaken, worden uitsluitend rechtshandige proefpersonen geworven voor dit soort van experimenten.
2. De filler-zinnen die bij het ongerelateerde experiment hoorden, waren zogeheten 'aanhechtingsambiguiteiten', zoals te zien in de volgende zin "Ik zag de mentor van de jongens die met het toneelstuk had meegedaan." In

deze zin is het tot aan het de werkwoordsvorm 'had' voor de lezer onduidelijk of de relatieve bijzin 'die met het toneelstuk ..' aan 'de mentor' of aan 'de jongens' aangehecht zou moeten worden, met andere woorden, de lezer weet een tijdlang niet of de mentor iets met het toneelstuk te maken had, of dat het juist de jongens waren. De condities in dit 'filler-experiment' verschilden in of het desambiguerend werkwoord enkelvoud ('had') of meervoud was ('hadden') en of de ambigue zinnen in de context van twee mentoren of van twee groepen jongens gepresenteerd werden.

3. De keuze van de intervallen waarover ERP-effecten worden berekend, is vaak enigszins arbitrair. In dit onderzoek is op basis van visuele inspectie vastgesteld a) welke verschillen aanwezig waren tussen de ERP-golfvormen behorend bij de twee condities, en b) in welke intervallen deze verschillen relatief consistent waren in termen van polariteit: als op één elektrode twee condities zowel negatieve als positieve verschillen vertonen in één en hetzelfde interval wordt dit niet als effect gezien.

Bibliografie

- Berkum, J. J. A. van (2004).** Sentence comprehension in a wider discourse: can we use ERPs to keep track of things? In: M. Carreiras & C. Clifton, Jr. (Eds.), *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond* (pp. 229–270). New York: Psychology Press.
- Bornkessel, I., Schlesewsky, M., & Friederici, A. D. (2002).** Beyond syntax: Language-related positivities reflect the revision of hierarchies. *NeuroReport*, 13, 361–364.
- Campbell, J. (2001).** *The liar's tale*. New York, NY: Norton.
- Coulson, S. (2004).** Electrophysiology and pragmatic language comprehension. In I. Noveck & D. Sperber (Eds.), *Experimental pragmatics* (pp. 187–206). San Diego, CA: Palgrave MacMillan.
- Friederici, A. D. (1995).** The time course of syntactic activation during language processing: A model based on neuro-psychological and neurophysiological data. *Brain and Language*, 50, 259–281.
- Friederici, A. D., Pfeifer, E., & Hahne, A. (1993).** Event-related brain potentials during natural speech processing: Effects of semantic, morphological and syntactic violations. *Cognitive Brain Research*, 1, 183–192.
- Grice, H. P. (1975).** Logic and Conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and Semantics, Volume 3: Speech Acts* (pp. 41–58). New York: Academic Press.
- Gunter, T. C., Stowe, L. A., & Mulder, G. (1997).** When syntax meets semantics, *Psychophysiology*, 34, 660–676.
- Hagoort, P., Brown, C. M., & Groothusen, J. (1993).** The syntactic positive shift as an ERP measure of syntactic processing. *Language and Cognitive Processes*, 8, 439–484.
- Hoeks, J. C. J. (1999).** *The processing of coordination: semantic and pragmatic constraints on ambiguity resolution*. Proefschrift, Universiteit Nijmegen.
- Hoeks, J. C. J., Hendriks, P., Vonk, W., Brown, C. M., & Hagoort, P. (2006).** Processing the NP- versus S-coordination ambiguity: thematic information does not completely eliminate processing difficulty. *The Quarterly Journal of Experimental Psychology*, 59, 1581–1599.
- Hoeks, J. C. J., Pijnacker, J., & Stowe, L. A. (in voorbereiding).** *The neural substrate of the semantic illusion: an event-related fMRI study*. Manuscript in voorbereiding, Rijksuniversiteit Groningen.
- Hoeks, J. C. J., & Stowe, L. A. (in voorbereiding).** *The time-course of facilitatory and inhibitory processes in sentence comprehension*. Manuscript in voorbereiding, Rijksuniversiteit Groningen.
- Hoeks, J. C. J., Stowe, L. A., & Doedens, L. G. (2004).** Seeing words in context: the interaction of lexical and sentence level information during reading. *Cognitive Brain Research*, 19, 59–73.
- Hoeks, J. C. J., Stowe, L. A., & Wunderink, C. (2004).** Time is of the essence: Processing temporal connectives during reading. In K.D. Forbus, D. Gentner, & T. Regier (Eds.), *Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society*. Mahwah, NJ: Erlbaum.

- Hoeks, J. C. J., Vonk, W., & Schriefers, H. (2002).** Processing coordinated structures in context: The effect of topic-structure on ambiguity resolution. *Journal of Memory and Language*, 46, 99-119.
- Horn, L. R. (2004).** Implicature. In L. R. Horn & G. Ward (Eds.), *The handbook of pragmatics* (pp. 3-28). Malden, MA: Blackwell.
- Kutas, M., & Hillyard, S. A. (1980).** Reading senseless sentences: brain potentials reflect semantic incongruity. *Science*, 207, 203-205.
- Kutas, M., Van Petten, C., & Kluender, R. (te verschijnen).** Psycholinguistics electrified II (1994-2005). In M. A. Gernsbacher & M. Traxler (Eds.), *Handbook of psycholinguistics* (2nd Edition). New York, NY: Elsevier.
- Lambrecht, K. (1994).** *Information structure and sentence form: Topic, focus, and the mental representation of discourse referents*. Cambridge, MA: University Press.
- Münte, T. F., Schiltz, K., & Kutas, M. (1998).** When temporal terms belie conceptual order: An electrophysiological analysis. *Nature*, 395, 71-73.
- Neville, H., Nicol, J. L., Barss, A., Forster, K. I., & Garrett, M. F. (1991).** Syntactically based sentence processing classes: Evidence from event-related brain potentials. *Journal of Cognitive Neuroscience*, 3, 151-165.
- Stowe, L. A. & Wijers, A. A. (1998).** Onderzoek naar de verwerking van taal met behulp van Event-Related Potentials. *Stem- Spraak- en Taalpathologie*, 7, 3-22.
- Stowe, L. A., Wijers, A. A., Willemsen, A., Reuland, E., Paans, A. M. J., & Vaalburg, W. (1994).** PET studies of language: an assessment of the reliability of the technique. *Journal of Psycholinguistic Research*, 23, 499-527.
- Van Petten, C. (in druk).** Neural localization of semantic context effects in electromagnetic and hemodynamic studies. *Brain & Language*.

Boekbeoordelingen

Van Enschoot, Renske (2006). *Retoriek in reclame; waardering voor schema's en tropen in tekst en beeld*. Dissertatie Faculteit Letteren, Katholieke Universiteit Nijmegen.
Promotores prof. dr. J.A.L. Hoeken en prof. dr. C.J.M. Jansen

Inleiding

Studenten vinden reclame leuk zoals iedereen die wel eens een cursus 'persuasieve communicatie' geeft, uit eigen ervaring weet. Het goede nieuws voor docenten en onderzoekers in de communicatie- en taalwetenschap is dat zij reclame nu zonder gewetenswroeging óók leuk kunnen vinden, en wel vanwege de sterk groeiende belangstelling voor (en dito inzicht in) één der oudste bewegingen der wereld: *retorische figuraties* in reclameboodschappen. Die belangstelling is juist zo gegroeid door de 'ontdekking' van de *visuele* retorica. Daar waar retorische vormen als rijm, paradox, metafoor en woordspeling (gedenken wij Lodewicks *Litteraire kunst*) exclusief leken voor het verbale domein, maken de recente ontwikkelingen in de cognitieve taalkunde, de communicatiewetenschap en de advertentie- en marketingwetenschap duidelijk dat veel van de ogenschijnlijk ongrijpbare 'beeldcultuur' die ons omringt op inzichtelijke wijze in kaart gebracht kan worden wanneer daarop de aloude *elocutio*-inzichten en -schema's losgelaten worden. Dat 'Heerlijk helder Heineken' begrepen kan worden als een alliteratie, dat wisten we

al wel, maar dat we de rol van een prent van Marilyn Monroe in een reclame voor Grolsch blond bier beter op waarde kunnen schatten op basis van de templaar van de retorische *simile*, dat is toch nieuws, en wetenschappelijk nog verantwoord ook.

Renske van Enschoot (VE vanaf nu) promoveerde op 27 april jongst leden in Nijmegen op haar proefschrift 'Retoriek in reclame', en daarmee levert zij, na het pionierswerk van Charles Forceville (1997), in Nederland de eerste serieuze proeve van wetenschappelijke bekwaamheid af op dit fascinerende onderzoeksterrein. 'Bekwaamheid', welzeker, want *Retoriek in reclame* is een mooi, informatief en tot denken stemmend proefschrift geworden.

Inhoud

Retoriek in reclame bestaat uit 6 hoofdstukken. Hoofdstuk 1 biedt een overzicht van het vakgebied waarin VE kwesties behandelt als de definiëring en de (vaak problematische) categorisering van retorische vormen. In navolging van McQuarrie & Mick (1996) definieert zij retorische vormen als 'kunstzinnig afwijkende vormen' waarmee een bepaalde basispropositie in een advertentie vorm wordt gegeven.

Verder biedt dit hoofdstuk een mooi overzicht van het empirische onderzoek naar de effecten van retorische vormen op boodschap-waardering en de gepercipieerde complexiteit van retorische vormen, de twee centrale afhankelijke variabelen in VEs ex-

perimentaties. Vooral het werk van de disciplinaire 'voorlieden' Barbara Phillips, Edward McQuarrie en David Mick komt hierin ruimschoots aan bod.

Ook plaatst VE het gebruik en begrip van retorische vormen (in reclame-uitingen) binnen de Relevantietheorie van Sperber en Wilson. Dat laatste is informatief, omdat het gebruik van retorische vormen een paradox impliceert. Sperber en Wilson noemen een uiting relevant wanneer die de ontvanger ervan een bepaald (gunstig) effect oplevert, en dat bij voorkeur tegen een *minimale* (cognitieve) verwerkingsoverhead. Die laatste aanname lijkt een zender die een retorische vorm gebruikt om de boodschap te verpakken, te schenden omdat zulke vormen juist vaak nopen tot een zeker verwerkingssurplus in vergelijking met 'niet-retorisch' verpakte boodschappen. De ontvanger moet zich een zekere moeite getroosten om aan een interpretatie te komen waar z/hij mee uit de voeten kan. De uiting 'Bij de Spar bent u goedkoper uit' kost minder dan 'Kopen bij de Spar is sparen bij de koop', dus wie zijn boodschap op die laatste wijze vormgeeft, riskeert afbreuk aan relevantie.

Dus waarom dan toch retorica? Die extra inspanning wordt, zo is het idee, goed gemaakt doordat de verwerking van retorische vormen een zogenaamd 'affectief' effect oplevert (je vindt het 'leuk' of 'aardig' het 'sommetje' op te lossen), en juist dát is het effect waarop de zender mikt die zich van retorische vormen bedient. De affectieve response zou zich op gunstige wijze verhouden met het (uiteindelijk nagestreefde) persuasieve effect. Het is deze aanname die centraal staat in (veel van) het onderzoek naar retorische vormen (in beeld en taal) in (persuasieve) communicatie, en zo ook in VE's onderzoek. Uiteindelijk wil ze met haar onderzoek daarom inzicht bieden in de (verschillen in) waardering en gepercipieerde complexiteit van retorische vormen

als *schema's* (de koepelterm voor op herhaling gebaseerde stijlfiguren als rijm en alliteratie) en *tropen* (de koepeltermen voor 'afwijkende' vormen als paradox en ironie), en die vormen dan geopponeerd aan 'niet-retorische' boodschappen.

Hoofdstuk 2 biedt een inventaris van retorische vormen (verbaal en visueel) in een door VE aangelegd corpus advertenties. Het optreden van die vormen wordt hierin ondermeer in relatie gebracht met voor de advertentiwetenschap relevante parameters als 'betrokkenheid', 'aankoopmotivatie' en 'producttype'. Tevens rapporteert VE in dit hoofdstuk onderzoek naar de beoordeelde complexiteit van de diverse vormen (verbaal schema/troop; visueel schema/troop). Visuele tropen blijken complexer te zijn dan visuele schema's. Een soortgelijke uitkomst wordt bij de verbale vormen gevonden, zij het dat de contrasten daar minder aangezet zijn.

Voor al in het derde tot en met vijfde hoofdstuk is veel interessants en waardevols aan experimentele resultaten te vinden. In hoofdstuk 3 gaat VE experimenteel na of er steun gevonden kan worden voor het idee dat vooral tropen complex gevonden worden, in tegenstelling tot advertenties met schema's en advertenties zonder retorische vormen. Tevens onderzoekt ze hier de vraag of de advertenties, dat wil zeggen hun 'basisproposities' ook daadwerkelijk begrepen worden door respondenten, en hoe verre dat hun oordelen over complexiteit en hun waardering beïnvloedt. Dat alleen al maakt dit hoofdstuk tot een aanwinst voor de retorische vakliteratuur, want de begripskwestie wordt in het meeste (experimentele) effectonderzoek simpelweg genegeerd (zie Schilperoord & Maes 2003, bijvoorbeeld).

Voor het bijeenbrengen van het stimulusmateriaal wordt de centrale variabele 'retorische vorm' (niveaus: *geen*, *schema* of *troop*) onafhankelijk gehanteerd voor zo-

wel talige als visuele vormen. Dat levert VE uiteindelijk 9 categorieën op. De metingen betreffen de factoren: *gepercipieerde complexiteit*, *waardering*, *gepercipieerd begrip* en *productwaardering*. Ik doe een (kleine) greep uit de resultaten. Bij de verbale vormen betreffen de gevonden effecten meestal de niet-retorische vormen. Die worden minder complex gevonden, en ook meer respondenten hebben de boodschap correct begrepen wanneer die geen retorische vorm bevat. Voor waardering vindt VE hier geen verschil, en op geen enkele meting vindt ze een verschil tussen schema's en tropen.

De visuele vormen geven een complexer beeld te zien. Allereerst worden de visuele schema's het meest complex beoordeeld, dus, tegengesteld aan de verwachting: complexer ook dan de tropen. Tropen worden op hun beurt weer complexer beoordeeld dan de niet-retorische vormen'. Tropen worden wel het hoogst gewaardeerd, maar alleen hoger dan de niet-retorische vormen – dus tussen tropen en schema's vindt VE hier geen verschil.

Interessant(er) zijn (ook) de interacties, die overigens alleen worden gevonden voor de gepercipieerde complexiteit. Vooral de complexiteit van advertenties met een visuele troep gedraagt zich grillig, en wordt het hoogst beoordeeld wanneer de *tekst* een schema bevat. Ook het omgekeerde lijkt het geval: wanneer het beeld schematisch is, beoordelen de proefpersonen de complexiteit juist van tropen (maar ook van niet-retorische vormen') hoog.

Complex is ook het beeld dat de analyses te zien geven wanneer VE daar de factor 'gepercipieerd begrip' bij betreft. Alweer vallen de mengvormen (beeld-schema & tekst-troep, of omgekeerd) op, in die zin dat juist deze combinaties tot de laagste begripsscores leiden. In termen van hoofdeffecten, springen de visuele schema's (!) eruit op dit punt. Ruim een kwart van de respondenten bleek de betekenis van die

vormen niet te hebben achterhaald.

Wat is nu het effect van dat al dan niet begrijpen van de boodschap? VE aggregiert haar metingen hier over drie begripsgroepen: *wel* begrepen, *deels* begrepen en *niet* begrepen. Uit de analyses blijkt dat de proefpersonen die de boodschap wel hebben begrepen de boodschappen minder complex vinden en hoger waarderen dan de niet-begrijpers. Bovendien blijken zij het geadverteerde product hoger te waarderen. Vanwege dit verband tussen begrijpen en waarderen, en omdat juist retorische schema's het minst correct begrepen worden, zou je uitgaande van de effectiviteit van de boodschap het gebruik van die schema's dus moeten afraden.

De resultaten bieden volgens VE voldoende aanleiding om met behulp van een aantal interviews met respondenten meer inzicht te verkrijgen in de perceptie en het begrip van schema's versus tropen. VE wil zo vooral meer te weten komen over de oorzaak van de hoge gepercipieerde complexiteit van visuele schema's. Een mogelijke verklaring is dat respondenten op zoek gaan naar betekenis(-sen) van visuele schema's die 'er niet is/zijn' (want schema's vat VE op als betekenisloze retorische templates, waarover straks meer), en dat zou dan de verwarring, de gepercipieerde complexiteit en het onbegrip kunnen verklaren.

De stimuli die retorische schema's bevatten werden geselecteerd, en in een aantal diepte-interviews aan respondenten voorgelegd. Het is een exercitie als deze die je (weer eens) attent maakt op de tekortkomingen van (grootscheeps) experimenteel onderzoek. De resultaten van de diepte-interviews wijzen namelijk uit dat de visuele schema-advertenties om andere dan de hier relevante kenmerken tot problemen leidden (bijvoorbeeld: de hoeveelheid beeldelementen, of de onduidelijke productpositionering in de advertentie).

Hoofdstuk 4 borduurt verder op de

problematiek van schema's versus tropen, maar nu door in een experiment op meer gecontroleerde wijze de variabele 'visuele retorische vorm' te onderzoeken. Daartoe manipuleert VE de diverse vormen nu zelf, in plaats van die uit haar corpus op te diepen. Daarbij gaat ze uit van bestaande advertenties die een visuele troep bevatten. Die troep wordt er voor de condities 'schema' en 'niet-retorische vorm' eerst uitgesloopt, en vervangen door een visueel schema of een niet-retorisch beeld. Zo is er een advertentie voor PiM-koekjes met grapefruitsmaak, waarin dat koekje (in bovenaanzicht) de ene helft vormt van een doormidden gesneden grapefruit. Ook de andere helft, een 'gewone' halve grapefruit, is te zien, samen met het mes waarmee de vrucht gekliefd is. Om de troep tot retorisch schema om te vormen, verwijdert VE de 'echte' halve grapefruit, en brengt ze vervolgens een aantal exemplaren van het koekje aan in de voorstelling. Hiermee is dus de schematische. Om de troep ook tot een niet-retorische vorm te bewerken handhaaft VE nog maar 1 exemplaar van het koekje in het beeld. Volgens deze procedure zijn zes advertenties bewerkt. De tekst wordt daarbij constant gehouden.

VE laat de diverse varianten weer beoordelen op de maten 'afwijking', 'complexiteit', 'waardering advertentie' en 'waardering product'. De resultaten wijzen uit dat de tropen als meer afwijkend, en complexer worden beoordeeld dan de schema's en de niet-retorische vormen. Ook op 'waardering advertentie' scoren de tropen het hoogst. De retorische vormen bleken geen effect te sorteren op de factor 'waardering voor het product'. De conclusie van het experiment luidt dat vooral tropen de retorische vorm is die verantwoordelijk is voor de effecten. Het valt dus vooral op is dat VE's proefpersonen geen verschil maken tussen retorische schema's en niet-retorische vormen.

In hoofdstuk 5 tot slot, onderzoekt VE de waardering voor de zogenaamde 'verbo-pictorale' vormen van retoriek: vormen van retorica die ontstaan door een bepaalde combinatie van tekst en beeldelementen. Een verbo-pictoraal schema ontstaat naar haar oordeel wanneer tekst en beeld 'verweven' zijn, maar wel samen een coherente boodschap vormen. Zo wordt het =-teken in de tekst 'Onze ontbijtkoek = er in vijf smaken' in een AH-advertentie door twee plakjes ontbijtkoek weergegeven. Van een verbo-pictorale troep is in VEs optiek sprake wanneer tekst en beeld een (ogenschijnlijke) tegenstelling bevatten (een paradox?). Zo verschijnt in een advertentie voor een zeker merk bananensap de tekst 'Dorst?' boven een (prominent) in beeld gebrachte banaan. Uit dit experiment komt naar voren dat tropen en schema's meer afwijkend beoordeeld worden dan niet-retorische vormen; dat tropen complexer beoordeeld worden dan de overige twee varianten, maar dat juist schema's hoger gewaardeerd worden. De productwaardering gaf eenzelfde beeld te zien.

In het slothoofdstuk worden de resultaten op een rij gezet, en teruggekoppeld naar de theorie.

Bespreking

Je kunt niet anders dan bewondering hebben voor de precisie en nauwgezetheid waarmee VE experimenteert, vooral als bedacht wordt dat het manipuleren van beeldmateriaal, of het bijeenbrengen van stimuli uit een corpus volgens de retorische criteria, een uitermate lastige zaak is. Wanneer je van een advertentie die een retorische troep bevat een schematische advertentie moet construeren, is dat wel even een eind lastiger dan wanneer bijvoorbeeld een tekst met veel lijdende vormen, of met veel connectieven omgewerkt moet worden naar een tekst met veel actieve vormen, of naar een tekst met weinig connectieven (en

zelfs dat is al geen sinecure). De bedachtzame wijze waarop VE bij deze arbeid het in hoofdstuk 1 ontwikkelde theoretische model van retorische vormen hanteert, dwingt veel respect af. Juist daarom leveren haar experimenten waar inzicht in en ware kennis van de effecten van die vormen op.

Maar ik ontkom toch niet aan de adversatieve wending die de conventie van het genre 'proefschriftbespreking' op deze plaats dicteert. Zo heb ik moeite met VE's operationalisering in hoofdstuk 5 van de diverse (combinatorische) vormen van tekst-beeld retoriek. Ik zie wel dat er iets aan de hand is met de koekplakjes in 'Onze ontbijtkoek = er in vijf smaken', en met de tekst 'Dorst?' boven een afgebeelde banaan, maar ik zie niet wat daar schematisch of tropisch aan is. Hoewel VE op dit punt stelt uit te gaan van de VPM-test van Forceville (1997), zijn de tekst- en beeldcombinaties in haar materiaal niet te reconstrueren als de twee termen die een retorische figuur met elkaar verbindt.

Een ander punt betreft VE's begrip 'retorisch schema'. Allereerst is het onderscheid tussen tropen en schema's in haar onderzoek nogal grofmazig. VE heeft niet echt gekozen voor een bepaald classificatiesysteem. Enerzijds houdt ze vast aan 'de klassieke' tweedeling tussen tropen en schema's, maar anderzijds situeert ze de problematiek wel binnen het recente classificatiesysteem dat in Phillips & McQuarrie (2004) wordt voorgesteld. Gevolg is dat haar tropen, bijvoorbeeld die ze gebruikt in hoofdstuk 4, nogal wat variatie vertonen. Zo zou je haar afbeelding 4.6 voor Deoleen een geval van 'juxtapositie' kunnen noemen, en haar PiM-advertentie een geval van 'fusie'. Phillips en McQuarrie voorspellen juist verschillende retorische effecten van zulke vormen, maar door VE's aanpak worden die nu min of meer één hoop gegooid.

Het begrip 'retorisch schema' levert trouwens meer problemen op, zowel in em-

pirisch als theoretisch opzicht. In het eerste en tweede experiment blijken proefpersonen weinig verschil te zien tussen niet-retorische vormen en retorische schema's, of juist tussen tropen en schema's. Een sleutelopmerking dienaangaande is te vinden op p. 149, waar VE naar aanleiding van het uitblijven van de verwachte resultaten in experiment 2, schrijft:

'Het blijkt echter moeilijk te zijn om het concept 'visueel schema' goed te operationaliseren, om vormgevingselementen te vinden die gepercipieerd worden als kunstzinnig en betekenisarm zijn'.

Zij maakt hier veel werk van de noties 'kunstzinnig' en 'betekenisarm'. 'Betekenisarm' moet hier in contrastieve zin begrepen worden – schema's zijn betekenisarm, en tropen betekenisrijk. Bovendien moeten schema's als 'kunstzinnige' afwijkingen gepercipieerd worden.

Ik heb het idee dat beide uitgangspunten op een misverstand berusten. Naar mijn oordeel zijn *alle* retorische vormen 'betekenisarm', die zin dat de afwijking die ze impliceren altijd een *formele* afwijking is. De templatens waarop zulke uiteenlopende figuren als *rijm* of *metaforiek* zijn gebaseerd, zijn te beschrijven als procedurele instructies waarmee die figuren in taal of in beeld kunnen worden gerealiseerd ter *ondersteuning* van een bepaalde basispropositie, en dus niet al dan niet als *dragers* daarvan. Omdat VE uitgaat van het idee dat alleen schema's betekenisarm zijn, reduceert ze die figuur uiteindelijk tot het simpelweg herhalen van een bepaald beeldobject. Dat leidt ertoe dat de proefpersonen in experiment 2 de mogelijkheid ontnomen wordt om te bepalen welke basispropositie nu ondersteund wordt door het afbeelden van – bijvoorbeeld – meerdere PiM-koekjes.

Verder is de term 'kunstzinnig' een vertaling van VE van de term 'artful' in de de-

finitie van McQuarrie & Mick (1996), en niet van haar alleen. Ook Schilperoord & Maes (2003) vertalen het begrip zo. Artful betekent echter eerder zoiets als ‘handig’ of ‘gewiekst’, en dat impliceert in dit geval meer dan alleen maar wat semantische haarkloverij. Juist vanwege VE’s criterium ‘kunstzinnigheid’ is haar voorstel tot verbreding van de notie ‘retorisch schema’ naar mijn idee eerder een uitholling van het begrip (zie p. 150). Maar omdat een uitvoerige bespreking van dit discussiepunt de bemeten ruimte in deze bespreking te buiten gaat, wil ik er graag in een aparte tekst op terugkomen.

Een laatste punt dat ik wil noemen betreft de *persuasieve* effecten van retorische figuren in reclames. VE presenteert het als een beperking van haar onderzoek dat ze erin de veronderstelde relatie tussen enerzijds ‘waardering voor de boodschap’ en anderzijds ‘waardering voor het product’ en ‘koopintentie’ niet empirisch heeft onderzocht. De afhankelijke variabele ‘waardering voor het product’ komt in haar experimenten zijdelings aan de orde (en wel in de vorm van een vraag naar ‘waardering voor het product’). Die inperking is in alle opzichten verantwoord en begrijpelijk: men kan nu eenmaal niet alles doen. Bovendien gaat het VE, zoals aangegeven, vooral om de factoren ‘(gepercipieerde) complexiteit’ en ‘waardering’, en daarbij gaat ze uit van het idee van het ‘affectieve’ effect, en de aangenomen relatie met persuasie ‘pur sang’ volgens de ‘experientiële verwerkingsstrategie’. Advertenties met retorische figuren vestigen de aandacht van de kijker/lezer op het verwerkingsproces zelf, en wanneer dat leidt tot ‘positieve’ affecten (je vindt de advertentie ‘leuk’ o.i.d.), is het beoogde persuasieve effect daarmee gediend.

Als gezegd, ik heb geen enkel bezwaar tegen deze redenering – sterker nog, in het arti-

kel dat ik in dit tijdschrift samen met Fons Maes heb gepubliceerd, wordt van precies diezelfde redenering uitgegaan: positieve gevoelens over de vorm worden overgedragen op de inhoud – als de verpakking deugt, dan ook de inhoud (zie Schilperoord & Maes 2003). Dat ik het punt toch noem, heeft te maken met het feit dat we onlangs in Tilburg experimentele gegevens verzameld hebben die vraagtekens lijken te zetten bij het idee van de experientiële verwerkingsstrategie. Hoe steekhoudend en overtuigend de redenering ‘waardering voor vorm leidt tot waardering voor inhoud en dat op zijn beurt weer tot het overtuigd raken van de inhoud’ ook moge wezen, de vergaarde data roepen de vraag op of die relatie er ook werkelijk zo is. De twee experimenten betroffen replicaties van onderzoek door Giora *et al.* (2004) naar zogenaamde ‘optimale innovaties’, woordspelingen (retorische vormen dus, zij het verbale) van het type die ook VE zelf heeft onderzocht (zie Van Mulken, Van Enschot- van Dijk & Hoeken 2005). Onze proefpersonen percipieerden bij zulke retorische vormen de afwijkingen, ze waardeerden ze ook, en ook konden ze waardering opbrengen voor het ontwerp ervan, maar op het punt van overtuigingskracht lieten ze het afweten. Ik weet (nog) niet wat de scope is van resultaten als deze, maar wellicht dat we hier van doen hebben met de ‘doe-maar-gewoon-dan-doe-je-al-geek-genoege’ geneigdheid die nogal eens tot ons nationale volkseigen gerekend wordt. Duidelijk is hoe dan ook dat daarmee de veronderstelde verbanden tussen ‘affect’ en persuasie (zie VEs proefschrift, hoofdstuk 1 en 6) op scherp gezet worden, en dat verder onderzoek hier dringend gewenst is.

Maar overigens: hulde voor Van Enschot’s werk, en van harte aanbevolen!

Joost Schilperoord

Bibliografie

- Forceville, C.** (1997). *Pictorial metaphor in advertising*. London [etc.], Routledge
- Giora, R., et al.** (2004). Weapons of mass distraction: Optimal innovation and pleasure ratings. *Metaphor and symbol*, 19, 115-141.
- McQuarrie, E.F. & D.G. Mick** (1996). Figures of rhetoric in advertising language. *Journal of Consumer Research*, 23, 424-438.
- Mulken, M. van, R. Van Enschoot-Van Dijk, H. Hoeken** (2005). Puns, relevance and appreciation in advertisements. *Journal of Pragmatics*, 37, 707-721
- Phillips, B.J. & E.F. McQuarrie** (2004). Beyond visual metaphor: A new typology of visual rhetoric in advertising. *Marketing theory*, 4, 113-136.
- Schilperoord, J. & A. Maes** (2003). Overtuigen met visuele en verbale retoriek. *Tijdschrift voor Taalbeheersing*, 25, 119-141.

Bekker, Birgit (2006). *De feiten verdraaid. Over tekstvolgorde, talige markering en sprekerbetrokkenheid*. Dissertatie Universiteit van Tilburg. Promotores prof. dr. L.G.M. Noordman en prof. dr. W.P.M.S. Spooren.

Begin dit jaar verdedigde Birgit Bekker haar proefschrift *De feiten verdraaid* aan de Universiteit van Tilburg. De dissertatie past in de Tilburgse traditie van tekstlinguïstische studies maar draagt een sterk taalkundig accent, waarin Bekkers achtergrond als taalwetenschapster duidelijk zichtbaar is. De kwestie die Bekker in haar studie aan de orde stelt is die van de volgorde waarin tekstelementen gepresenteerd worden: als deze volgorde afwijkt van de werkelijke volgorde, helpt de talige vorm ons dan dat te ontdekken?

Stel een tweetal gebeurtenissen: Jan geeft Max een duw, en tengevolge van die duw valt Max. In de sequentie *Max viel, toen/omdat Jan hem duwde* is de iconische volgorde – de volgorde waarin de gebeurtenissen zich in werkelijkheid afspeelden – omgekeerd

in een achterwaartse presentatie. De iconische, voorwaartse ordening wordt door de lezer verwacht (o.m. Noordman en Vonk 1998); afwijking van die standaardpresentatie verklaart de talige markering van de achterwaartse ordening met het temporele of causale connectief *toen* of *omdat*.

Achterwaartse volgorde is ook mogelijk zonder talige markering, zoals in de sequentie “Max viel. Jan duwde hem.” Caenepeel (1993) betoogt echter dat een dergelijke ongemarkeerde sequentie niet goed mogelijk is in een fictieve narratieve context, waarin de gebeurtenissen door de verteller worden geordend ten opzichte van elkaar; wel interpreteerbaar is ongemarkeerde achterwaartse ordening in een context waarin alle gebeurtenissen worden geordend ten opzichte van het hier-en-nu van de schrijver, zoals in een nieuwsbericht.

Bekker stelt dat de schrijver zijn positie ten opzichte van de achterwaarts gepresenteerde uiting ook kan weergeven door middel van sprekerbetrokkenheid, zoals temporele betrokkenheid in de onvoltooid tegenwoordige tijd: *Max is gevallen. Jan heeft hem geduwd*; of door mentale sprekerbetrokkenheid *Max viel. Volgens mij/Volgens Piet duwde Jan hem*; dergelijke vormen van sprekerbetrokkenheid vergemakkelijken een achterwaartse lezing in alle contexten.

Vraagstelling van de studie

De centrale vraag die Bekker aan het eind van het eerste hoofdstuk van haar dissertatie aan de orde stelt is nu, of dergelijke sprekerbetrokkenheid-elementen in afwezigheid van expliciete volgordemarkering de rol van achterwaarts volgordesignaal overnemen. Deze formulering suggereert, dat Bekker in geval van achterwaartse volgorde, de aanwezigheid van een temporeel of causaal signaal als *default*-situatie beschouwt, en de aanwezigheid van elementen van sprekerbetrokkenheid als gemarkeerd. Deze

opvatting komt terug in hoofdstuk 6, p. 202. Op p. 205 concludeert Bekker dat inderdaad achterwaartse ordening zonder volgordesignaal bij voorkeur plaatsvindt in een context waarin sprake is van sprekerbetrokkenheid. Bij afwezigheid van een volgordesignaal, zo concludeert zij, lijken we elementen van sprekerbetrokkenheid te mogen opvatten als alternatief signaal van achterwaartse volgorde.

De dissertatie als geheel vormt aldus een uiteindelijk bevestigend antwoord op een interessante maar tegelijk enigszins geïsoleerd blijvende vraag; wel een antwoord dat uiterst zorgvuldig, methodologisch uitgewerkt en empirisch onderbouwd is. In deze kwalificatie liggen zowel mijn waardering voor als kritiek op deze dissertatie besloten. Hieronder geef ik kort de inhoud van de studie weer. Vervolgens beargumenteer ik mijn oordeel met de nadere bespreking van een aantal hoofdstukken.

Opzet en inhoud van de studie

Het eigenlijke onderzoek bestond uit vijf fasen.

1. Het definiëren van het begrip sprekerbetrokkenheid. Hoofdstuk 2 is besteed aan het uitwerken van deze notie, waarbij gebruik wordt gemaakt van de begrippen *perspectief* en *subjectiviteit*. Ook wordt in dit hoofdstuk een tweetal testen geïntroduceerd als methode om het perspectief van een uiting te bepalen.
2. Het samenstellen van een corpus. Het eerste deel van hoofdstuk 3 geeft een toelichting op de keuze voor het empirisch tekstmateriaal. Een corpus werd samengesteld uit nieuwsteksten en opiniërende teksten uit de Volkskrant en de NRC. Ruim 700 fragmenten werden geselecteerd die hetzij voorwaartse hetzij achterwaartse temporele volgorde vertonen (paragraaf 3.2).

3. Het ontwikkelen van een analysemodel. Het tweede deel van hoofdstuk 3 is hieraan gewijd: paragraaf 3.3 beschrijft aan de hand van welke criteria de talige markering en de sprekerbetrokkenheid van de fragmenten werden gecategoriseerd, paragraaf 3.4 beschrijft op welke wijze de betrouwbaarheid van de beoordelingen werd gewaarborgd en getoetst.
4. De eigenlijke corpusanalyse. Hoofdstuk 4 geeft een bespreking en statistische toetsing van de resultaten van de corpusanalyse. Eerst worden de beiden teksttypen – nieuwstekst en opiniërende tekst – vergeleken wat betreft sprekerbetrokkenheid in paragraaf 4.2. De kenmerken van achterwaartse fragmenten worden beschreven in paragraaf 4.3, waarna het verband tussen achterwaartse betrokkenheid, markering en sprekerbetrokkenheid aan de orde komt in 4.4. Zo ook de voorwaarts geordende fragmenten: deze worden wat betreft hun kenmerken beschreven in 4.5, waarna het verband met markering en sprekerbetrokkenheid in 4.6 wordt besproken.
5. Een nadere uitwerking van een deel van de analyse: de talige markering van temporele ordening. In het vijfde hoofdstuk staat de voltooid verleden tijd centraal als potentiële markering van subjectieve sprekerbetrokkenheid. Deze analyse behelst feitelijk een tweede onderzoek dat bestaat uit een aantal subfasen, te weten:
 - a. Analyse van de voltooid verleden tijd in het kader van een theorie over temporele discourse structuur (eerste deel van paragraaf 5.2).
 - b. Modelleren van drie lezingen van de voltooid verleden tijd op basis van de plaatsing van het temporele evaluatiepunt. Hypothesen worden geformuleerd over de samenhang met subjectieve lezing van de voltooid verleden tijd (tweede deel van paragraaf 5.2).

- c. Opzet, uitvoering en resultaat van een corpusanalyse dat deze hypothesen toetst (paragraaf 5.3 en 5.4).

Temporeel-semantiche analyse van voltooid verleden tijd

De meest originele bijdragen in deze dissertatie zijn in mijn ogen de testen waarmee het perspectief kan worden vastgesteld in hoofdstuk 2 en de temporeel-semantiche analyse-methode in hoofdstuk 5. Vanuit theoretisch oogpunt zijn er zeker vragen te stellen bij de perspectieftesten – en dat zal ik verderop in deze bespreking dan ook doen – als methode zijn ze vindendrijf en toepasbaar. Ik begin mijn bespreking met hoofdstuk 5 omdat dit wat minder hecht verbonden is met de kernvraag dan de overige hoofdstukken. Dit hoofdstuk, gebaseerd op een artikel samen met Leonoor Oversteegen in *Linguistics* (Oversteegen en Bekker 2002), brengt het concept van het vrije evaluatiepunt in bij de representatie van de voltooid tegenwoordige tijd (VVT). Deze modellering is een toepassing en uitwerking van Oversteegens (1989) *Tivo Track Theory*, waarin de *event line* onderscheiden wordt van de sprekergeoriënteerde temporele lijn. De VVT vraagt, afhankelijk van de interpretatie, twee of drie evaluatiepunten op de temporele lijn om de door de spreker beschreven situatie in te bedden. In sommige teksttypen, waarin de temporele structuur wordt bepaald door de ordening van situaties ten opzichte van elkaar, zijn de evaluatiepunten temporeel gemotiveerd; dit geldt met name in narratieve fictie. Bekker geeft als voorbeeld: *Toen An en Jo trouwden, hadden ze de hele straat uitgenodigd*. Het uitnodigen wordt door de VVT temporeel geordend voorafgaand aan het trouwen en het resultaat (“resultatieve toestand”) van het uitnodigen valt samen met trouwen (p. 181). In andere teksttypen, zoals journalistieke nieuwsteksten waarin situaties primair worden geordend ten opzichte van

het hier-en-nu van de schrijver, komt het voor dat de het evaluatiepunt van de VVT niet temporeel gemotiveerd is. Bekker geeft als voorbeeld: *Feyenoord won gisteren met 3-0 van Ajax. Opstelden had op zijn kop gestaan*. In het laatste geval van VVT, zo analyseert Bekker, is het denkbeeldig moment van temporele evaluatie niet te bepalen: het is vrij (p. 183).

Aan de hand van een corpusanalyse van narratieve en journalistieke teksten laat Bekker zien dat het vrije evaluatiepunt van de VVT een markering vormt van de subjectieve lezing van de zin. In beide teksttypen is de temporeel vooropgeplaatste uiting met VVT te lezen als een waarneming of herinnering door de ogen van een personage. Ik vind dit zeer interessant werk: theoretisch gemotiveerd, empirisch onderbouwd, en richtinggevend naar verder functioneel genreonderzoek, met bijvoorbeeld als vragen: duidt het gebruik van deze markering op een element van narratieve fictionaliteit in nieuwsteksten? Met welk beoogd effect wordt deze markering toegepast en hoe waarderen lezers hem?

Definitie van sprekerbetrokkenheid

Hoofdstuk 2 is gewijd aan perspectief en subjectiviteit, begrippen die Bekker nodig heeft om sprekerbetrokkenheid te definiëren: deze term wordt “overkoepelend gebruikt om te verwijzen naar tekstelementen die een deel van hun betekenis ontleen aan de relatie met een persoonlijk standpunt” (p. 20). Een flink aantal van de uiteenlopende modellen van perspectief en subjectiviteit die in de literatuur voorhanden zijn bespreekt ze op een, zoals ze op diezelfde pagina erkent, tamelijk eclectische wijze, niet met het theoretisch doel de concepten nader te verklaren en modelmatig te relateren, doch met het pragmatistisch doel om met hulp van deze concepten een werkbaar begrip te krijgen voor haar centrale vraagstelling – de relatie tussen

sprekerbetrokkenheid en achterwaartse presentatie van informatie. Het resultaat is dan ook een praktisch werkschema (p. 54) van de diverse tekstuele invalshoeken waar sprekerbetrokkenheid zich potentieel manifesteert. Bekker onderscheidt:

1. de vormkenmerken van tekstsegmenten – betreft de vertelstijl, deixis, grammaticaal aspect en aspectuele klasse, mentale activiteit en communicatieve handelingen, modaliteit en evaluatieve expressies;
2. de lezingen van tekstsegmenten – vanuit het nul-, schrijver- of ander-perspectief (waarover hierna meer);
3. de interpretatie van coherentie tussen segmenten – waar sprekerbetrokkenheid zich kan manifesteren in subjectgebonden coherentierelaties (o.a. Sweetser 1990 onderscheidt epistemische en *speech act* relaties; Mann en Thompson 1988 spreken van presentationele relaties; Sanders, Spooren en Noordman 1992 onderscheiden pragmatische van semantische relaties).

Zoals gezegd vormen de testen waarmee de perspectieffezingen van uitingen vastgesteld kunnen worden een originele bijdrage. De testen worden bepaald door de wijze waarop Bekker het begrip *subjectiviteit* hanteert. Zowel bij presentatie van het standpunt van de schrijver als bij een ander persoonlijk standpunt moet de lezer de verantwoordelijkheid toeschrijven aan een persoonlijke instantie, en daarom is volgens Bekker in beide gevallen sprake van subjectieve informatie. Om te ontkomen aan de door haar omschrijving opgeroepen consequentie dat alle in tekst gepresenteerde informatie subjectief is, introduceert Bekker het begrip *nul-perspectief* naast het *schrijver-perspectief* en het *ander-perspectief*. Het nul-perspectief is dan de objectieve lezing van een zin, waarin de geldigheid van gepresenteerde informatie niet beperkt moet worden tot het men-

tale domein van een persoon; ze verwijst hierbij naar het “technisch basisstandpunt” van de schrijver (p. 40).

Wat is nul-perspectief?

Het nulperspectief komt overeen met een objectieve lezing van de zin waarin geen persoonlijk standpunt tot uitdrukking komt; bv. *Het heeft geregend*; het schrijverperspectief komt overeen met een subjectieve lezing van de zin vanuit het perspectief van degene die de tekst uitspreekt of schrijft, bv. *Het moet geregend hebben* (subjectiviteit uitgedrukt door een epistemische modaal, zie J. Sanders en Spooren 1996); en in het ander-perspectief is een ander subject dan de spreker verantwoordelijk voor het perspectief, bv. *Volgens Jan heeft het geregend*.

Het verschil tussen de perspectieven wordt verhelderd door de parafrase- en attributietest (p. 43 e.v.), waarbij interpreteerbare toevoeging van *dat denkt hij/zij/men* duidt op ander-perspectief, van *dat denk ik* op schrijverperspectief en van *dat is een objectief gegeven* op nul-perspectief. De ontkenbaarheidstest (p. 41) is er om onderscheid te maken tussen enerzijds nulperspectief en anderzijds schrijver- of ander-perspectief. Deze test gaat ervan uit dat een uiting niet tegelijk geldig en ongeldig kan zijn binnen één en hetzelfde domein van interpretatie; interpreteerbare toevoeging vanuit het nul-perspectief van *maar dat is/was niet zo/wat er gebeurt/gebeurde* leidt tot schrijver- of ander-perspectief.

Bekkers voorstel voor het nul-perspectief roept bij mij meer vragen op, dan hij beantwoordt. Ik wil dit illustreren aan de hand van een voorbeeld van Sweetser 1990, door Bekker geciteerd (p. 48): *John loved her, because he came back*. De eerste zin van dat voorbeeld, *Jan hield van haar*, lijkt op grond van de ontkenbaarheidstest een geval van nul-perspectief. Hoewel *houden van* geen objectief vaststelbaar gegeven is, kan de uiting niet ontkend worden zonder dat dit tot

inconsistentie leidt: *Jan hield van haar, maar dat is/was niet zo* is niet interpreteerbaar. De uiting is ook niet subjectief vanuit de schrijver: eerder vanuit de ander, namelijk Jan, wiens innerlijke toestand met *houden van* wordt beschreven.

Wanneer nu wordt vervolgd met *want* hij kwam terug, dan classificeert Sweetser zo'n sequentie als behorend tot het epistemische domein van de redenering. Bekker parafraseert de sequentie als *Ik denk dat Jan van haar hield, dat denk ik op basis van mijn kennis van het feit dat hij terugkwam*. Echter, als dit de juiste parafrase is, dan moet *Jan hield van haar* geïnterpreteerd worden als schrijver-perspectief, bepaald door het gebruik van het connectief *want*: een schrijver-perspectief met terugwerkende kracht. De consequentie is dat categorising van nul-perspectief problematisch is. Elders heb ik betoogd dat het perspectief van de schrijver *hier-en-nu* in elke uiting aanwezig is en dan meer of minder subjectief is (afhankelijk van de mate van *foregrounding*, zie Langacker 1990), tenzij het perspectief expliciet aan een ander subject is gebonden. Die ander kan natuurlijk ook de schrijver zelf, *niet-hier-en-nu* zijn (J. Sanders 1994; J. Sanders en Redeker 1996). De ontkenbaarheidstest werkt dan ook prima: *Ik dacht dat Jan van haar hield, maar dat is niet zo*.

Ik concludeer dat de besproken literatuur – hoewel inmiddels misschien enigszins gedateerd – en het resulterende schema een compleet en bruikbaar overzicht bieden van de talrijke manieren waarop subjectieve betrokkenheid in tekst manifest kan worden; dit overzicht kan zonder meer aanbevolen worden als kennismaking met de vele tekstuele verschijningsvormen van subjectiviteit en perspectief in tekst. De bijbehorende testen om perspectief vast te stellen zijn ook in andere contexten toepasbaar. Vanuit theoretisch oogpunt zijn Bekkers perspectiefdefinities echter niet helemaal bevredigend, hetgeen mijns

inziens wordt veroorzaakt door de node-loos ingewikkelde omweg die zij maakt om tot het nul-perspectief te geraken. Bekker laat ons gissen naar de precieze beschrijving en functie ervan: moet het gezien worden als een uitzonderlijke situatie, semantisch en temporeel bepaald? Het is bijvoorbeeld voorstelbaar dat handelingen worden gezien als invulling van nul-perspectief: *Jan roept haar*, en/of uiterlijke toestanden in het hier-en-nu: *Het regent*. Het is zelfs mogelijk dat het nul-perspectief misschien een restcategorie blijkt, die overblijft als er geen enkele sprekerbetrokkenheid is, ook niet achteraf.

Fundering van de probleemstelling

De laatste opmerkingen raken aan een meer fundamentele kwestie die aan de orde is bij de centrale probleemstelling in deze dissertatie. Bekker geeft zelf aan in paragraaf 2.5 dat coherentierelaties subjectiviteit kunnen uitdrukken. Dat betekent, dat een volgorde-markering als *omdat* of *want* ook een uiting van sprekerbetrokkenheid is: de spreker drukt immers met het causale connectief een redeneerrelatie tussen twee uitingen uit. Dergelijke connectieven worden in cognitief-linguïstische theorieën als subjectief gecategoriseerd, omdat ze een interpretatieve inbreng van de spreker zelf markeren (zie bijvoorbeeld T. Sanders en Spooren, te verschijnen). Dergelijke subjectieve markeringen, het woord zegt het al, zijn dus niet ongemarkeerd. Deze kwestie sluit direct aan bij de eerder gemaakte opmerkingen bij het mijn inziens moeilijk houdbare verschil tussen nul- en schrijverperspectief. Als er ook binnen het nul-perspectief subjectiviteit mogelijk is, wat betekent dan die subjectiviteit ten opzichte van die in het schrijver- en ander-perspectief en welke relatie heeft die subjectiviteit met die welke wordt uitgedrukt door temporele ordening?

Nu stuiten we op wat ik in het begin heb genoemd: het enigszins geïsoleerd blijven

van de centrale probleemstelling. Wat in de eerste twee hoofdstukken opvalt is dat Bekker haar kwestie in beperkte mate expliciet inbedt in bestaande (cognitief)linguïstische theorieën. De geraadpleegde bronnen worden tamelijk los van elkaar in aparte subparagraafjes besproken en de bespreking mondt vrij plotseling uit in de centrale probleemstelling. Waaróm zouden schrijvers in het ene geval een volgordemarkering boven hogere sprekerbetrokkenheid verkiezen en in het andere geval het omgekeerde, en in weer andere gevallen beide of geen van beide? Met andere woorden: waarom zegt men soms: *Max viel, omdat/toen Jan hem duwde*, in andere gevallen: *Max viel, ik zag/Piet dacht dat Jan hem duwde*, en in weer andere gevallen: *Max viel. Jan duwde hem/heeft hem geduwd*. Is het inderdaad zo, dat de eerste keuze met *omdat* of *toen* als *default* te beschouwen is voor de gemarkeerde keuze met sprekerbetrokkenheid, en zo ja op grond van welk model? Kortom, het probleem dat Bekker aan de orde stelt is op zichzelf belangwekkend, maar zij had misschien nog duidelijker kunnen maken wat de oplossing van het geponeerde probleem zou kunnen bijdragen aan een meer overkoepelende theorie van tekstrepresentatie en taalgebruik.

Corpusanalyse van journalistieke teksten

Hoofdstuk 3 gaat nader in op de opzet en uitvoering van de corpusanalyse. Bekker verantwoordt op p. 67 de keuze voor de twee journalistieke tekstgenres (nieuwstekst en opiniërende tekst) vanuit de hier-en-nu positie die de schrijver in beide gevallen inneemt en die de achterwaartse volgorde mogelijk maakt. Terloops wijst Bekker hier ook op mogelijke verschillen in markering van achterwaartse volgorde die vergelijking van de twee genres zal kunnen opleveren. Aanvankelijk blijft wat onderbelicht, welke rol de keuze voor de twee genres in het corpus heeft gespeeld: was dit om te controleren voor eventuele genre-effecten van

één van beide genres op zichzelf, of was er (ook) een hypothese die ter zake van de centrale probleemstelling juist een specifiek verschil voorspelt, en zo ja, op grond van welk model? Immers, in hoofdstuk 4 wordt bij de rapportage van resultaten steeds nadrukkelijk onderscheid gemaakt tussen gegevens verkregen uit beide teksttypen. Bekker pleit aan het slot van hoofdstuk 4 op grond van de soms uiteenlopende resultaten “voor verder onderzoek naar vormen en interpretatiekenmerken en naar de patronen van tekstuele samenhang in diverse teksttypen” (p. 155). Er zijn op grond van de verschillende functies van de beide genres (informerend versus persuasief) wellicht nadere verwachtingen te formuleren over volgordemarkeringen en subjectieve lezingen van achterwaartse volgorde.

Doorwerking

Het afsluitende hoofdstuk 6 vat de twee onderzoeken (naar markering van achterwaartse ordening en naar de VVT) afzonderlijk samen in paragraaf 6.1 en 6.2 en sluit af met enkele overwegingen over de rol van het teksttype en sprekerbetrokkenheid in paragraaf 6.3. Evenals bij de inleidende hoofdstukken valt hier op dat de theoretische verankering van de studie beperkt blijft. Sommige mogelijkheden hier toe worden wel aangeroerd, maar Bekker blijft hier wat beknopt en opsommend.

Er zijn verschillende routes waarlangs de gevonden resultaten met verdere theorievorming over verwerkings-, productie- en functieaspecten kunnen worden verbonden. De resultaten geven aanleiding tot interessante vragen naar het lezen van tekstfragmenten in achterwaartse iconiciteitsvolgorde met en zonder volgordemarkering, met en zonder sprekerbetrokkenheid; naar de betekenis van markering van temporele ordening in relatie tot tekstgenre, fictionaliteit, mate van argumentatie, stijlvariatie en moeilijkheidsgraad van de tekst.

De waarde van Bekkers analyse ligt in de sterke empirische basis. Zij maakt uitstekend duidelijk hoe zij te werk is gegaan bij de categorisering en beoordeling van haar natuurlijke corpusmateriaal. De tekstanalytische resultaten en statistische toetsingen zijn helder en uitvoerig gerapporteerd. De wijze waarop Bekker de betrouwbaarheid van de resultaten voorziet van fundering door statistische vergelijking van de drie beoordelaars is voorbeeldig en zou goed bewerkt kunnen worden tot een methodologisch artikel voor collega-tekstwetenschappers. Als Bekkers data op de een of andere wijze toegankelijk zijn, kunnen tekstwetenschappers die zich bezighouden met subjectiviteit en temporele ordening in verschillende teksttypen met vertrouwen voortbouwen op de omvangrijke data-analyse.

José Sanders

Bibliografie

- Caenepeel, M.** 1993. *Aspect and text structure*. Deliverable R1.3.1, Part 1, Esprit basic Research Project 6665 DANDELION. University of Edinburgh/CLS/CEC
- Langacker, R.** 1990. Subjectification. *Cognitive Linguistics* 1, 5–38.
- Mann, W. & S. Thompson** 1988. Rhetorical Structure Theory: Toward a functional theory of text organization. *Text* 8, 243–281

- Noordman, L. & W. Vonk** 1998. Memory-based processing of causal relations. *Discourse Processes* 25, 191–212.
- Oversteegen, E.** 1989. *Tracking time: A proposal for the representation of temporal expressions in Dutch*. Dissertatie, Utrecht: Universiteit Utrecht.
- Oversteegen, E. & B. Bekker**, 2002. Computing perspective: the pluperfect in Dutch. *Linguistics* 40, 111–161.
- Sanders, J.** 1994. *Perspective in narrative discourse*. Dissertatie, Tilburg: KUB.
- Sanders, J. & G. Redeker** 1996. Perspective and the representation of speech and thought in narrative discourse. In: G. Fauconnier & E. Sweetser (eds.), *Spaces, worlds and grammars* Chicago: University of Chicago Press, 290–317.
- Sanders, T., W. Spooren & L. Noordman** 1992. Toward a taxonomy of coherence relations. *Discourse Processes* 15, 1–35.
- Sanders, T., & W. Spooren** (te verschijnen). Discourse and text structure. In: D. Geeraerts & H. Cuykens (eds.) *Handbook of cognitive linguistics*, Oxford: Oxford University Press.
- Sanders, J. & W. Spooren** 1996. Subjectivity and certainty in epistemic modification. *Cognitive Linguistics*, 7 (3), 241–264.
- Sweetser, E.** 1990. *From etymology to pragmatics. Metaphorical and cultural aspects of semantic structure*. Cambridge: Cambridge University Press.

Signaleringen

Jan Renkema (2006), *Geef de spelling wat speling.* Den Haag: Sdu Uitgevers. ISBN: 9012114721. Prijs: € 5,95 (88 pp.)

Ter gelegenheid van het feit dat er van zijn vermaarde *Schrijfwijzer* 400.000 exemplaren verkocht zijn, kreeg Renkema van de Sdu rond Kerstmis 2005 het verzoek een Pamflet te schrijven. Renkema koos voor de spelling-2005 als onderwerp. Zoals de term 'pamflet' al aangeeft, is hier geen sprake van een wetenschappelijk werk (noten en een bibliografie ontbreken dan ook), maar van een essayachtige beschouwing over de veranderingen in de spelling, over het waarom daarvan en vooral over de vraag hoe we het beste met die verandering, hoe beperkt ook, kunnen omgaan.

De kern van zijn betoog verwoordt Renkema reeds op p. 8: 'Laten we dus deze spelling volgen, ondanks alle kritiek, maar wel met enige speling.' Hij schaart zich dus niet achter de actie van een aantal Nederlandse dag- en weekbladen om de spelling-2005 te boycotten. Op de vraag of een spelingsysteem speling verdraagt en hoeveel gaat de auteur niet fundamenteel in, wat binnen het kader van een boekje als dit ook niet erg voor de hand ligt. Interessant blijft die vraag overigens wel. Wel laat hij aan de hand van enkele concrete voorbeelden zien dat er in de spelling, zoals ook op andere terreinen van de taal, vaak sprake is van een glijdende schaal. Zo zal men op dit moment kunnen schrijven over een *Wilderskapsel*,

maar als die coiffure mode zou worden, zou die schrijfwijze allengs veranderen in *wilderskapsel*. Zo schrijven we *pianospelen* naast *klavecimbel spelen*, maar als dat laatste even gewoon zou worden als het eerste, dan zou je misschien *klavecimbelspelen* gaan schrijven. Op deze en dergelijke terreinen van de spelling, zou enige vrijheid gewenst zijn.

In het eerste hoofdstuk beantwoordt Renkema twaalf gangbare vragen over spelling. Het betreft hier vragen als: 'Het spelingsysteem is zo lastig. Wordt het niet eenvoudiger wanneer we gewoon opschrijven wat we horen?', 'Zijn wij in Nederland en Vlaanderen nu zo bijzonder? Wat kunnen wij leren van andere landen?' en 'Wat zijn nu de kenmerken van al die kleine veranderingen?' De antwoorden van de auteur komen respectievelijk op het volgend neer: 'Nee, daarmee wordt het niet eenvoudiger', 'Als je de spelling niet door de overheid laat regelen, ben je beter af' en 'De spelling-2005 is conserverend van aard, en vermindert het aantal hoofdletters en liggende streepjes'. Met het eerste en derde antwoord ben ik het eens, met het middelste niet. Aan het slot van deze signalering zal ik dat laatste toelichten.

In het tweede hoofdstuk, 'De nieuwe spelling: Betuwelijn van het taalverkeer' gaat Renkema in op de rol die verschillende actoren spelen bij de verandering die volgens hem bijna niemand wilde. Als kernen van het probleem schetst hij de onmogelijke opdrachten die de spellingcommissies krijgen

(ook de spelling-1995 komt hier zijdelings aan de orde), de hoge mate van geheimhouding totdat het station bijna gepasseerd is, de ondeskundige reacties in de pers en de veel te late reacties van belangengroepen.

In hoofdstuk 3 gaat hij in op de tussen-n-beregeling uit 1995. Renkema acht die te ingewikkeld, zelfs nog ingewikkelder dan de regeling uit 1954. Daarom stelt hij aan degenen die niet gehouden zijn tot het volgen van de officiële spelling (dus ieder die geen scholier of ambtenaar is) de volgende regel ter overweging voor: ‘Schrijf alleen een tussen-n (na een e), wanneer uit een omschrijving blijkt dat bij het eerste deel alleen het meervoud (op –en) is bedoeld.’ Deze regel lijkt op de oude regel betreffende het noodzakelijke meervoud, maar wijkt daarvan af omdat het hier niet gaat om wat in principe mogelijk is, maar wat concreet is bedoeld. Voorbeeld: in het Groene Boekje van 1954 krijgt *bessestruik* geen *n*, omdat ook een struik waar geen enkele bes aan zit (bijvoorbeeld ’s winters) een bessenstruik blijft. Bij Renkema’s regel moet je een omschrijving proberen: Struik waaraan een bes / bessen [doorhalen wat niet verlangd wordt] zit(ten) / kan(kunnen) zitten. De tweede optie ligt het meest voor de hand. Dus: *bessenstruik*. Als beide mogelijkheden reëel zijn, zoals bij *pannenkoek*, vervalt de *n*. De output van deze regel zal sterk lijken op de situatie van voor 1995, met dien verstande dat ook woorden als *herenfiets* en *boerenknecht* hun *n* zullen verliezen.

Ik vind dit voorstel, hoewel interessant, teveel afwijkingen opleveren ten opzichte van de officiële spelling. Kun je nog wel zeggen dat je met deze spelling die spelling nog volgt? Persoonlijk zou ik meer zien in een vereenvoudiging van de huidige regel voor de woorden die eindigen op –e. Als die ook een tussen-n krijgen als ze een meervoud op –en hebben (los van de vraag of ze ook een meervoud op –s hebben), dan zijn er een hoop problemen opgelost. Je

krijgt dan *gedachtenwisseling* en *secondenwijzer*. Daarnaast zou men nog meer moeten wieden in de uitzonderingen en gewoon schrijven *zonnenschijn* en *apenzat*.

In hoofdstuk 4 behandelt Renkema 26 voorstellen om af te wijken van de spelling-2005. Met veel van zijn voorstellen, zoals *Middeleeuwen* i.p.v. *middeleeuwen*, *quasi-authentiek* i.p.v. *quasiauthentiek*, *reïntegratie* i.p.v. *re-integratie*, *sociaal-psychologisch* in plaats van *sociaalpsychologisch* ben ik het van harte eens. Ik ben vast van plan vooralsnog de linkervormen te blijven gebruiken. Het betreft hier slechts veranderde toepassingen van bestaande regels en aanpassing van regels die geen centrale rol spelen.

Tot slot formuleert Renkema in zijn laatste hoofdstuk ‘Rust aan het spellingfront...’ een aantal desiderata. Kernen daarvan zijn: verander voorlopig niet, ga soepel om met afwijkingen en besteedt de spelling uit aan een academie waar niet alleen taalkundigen in zitting hebben, maar ook andere professionele taalgebruikers, die openbare adviezen uitbrengt.

Ik heb Renkema’s pamflet vlot en met plezier gelezen. Nu eens geen ondeskundige opmerkingen en afgezaagde voorbeelden (*ideeëloos*) waar de kranten in grossieren en ook niet de arrogante en neerbuigende toon die veel leden van de spellingcommissie hanteren. Veel van zijn bezwaren tegen de spelling-2005 deel ik, zoals ik in dit tijdschrift reeds eerder heb laten blijken. Weer andere bezwaren deel ik niet.

Op twee punten heb ik wat fundamentele kritiek. Het ene heb ik reeds genoemd. Renkema’s alternatieve regel voor het schrijven van de tussen-n leidt naar mijn smaak tot te veel afwijkingen van de officiële spelling. Ik heb bij het alternatieve gebruik van hoofdletters en liggende streepjes al een licht gevoel van burgerlijke ongehoorzaamheid. Maar ik besef ook dat voor anderen de kern van de spelling kleiner kan zijn en de variabele marge groter. Daarover

zou een fundamentele discussie interessant zijn.

Het andere punt waar ik een andere visie op heb is de rol van de overheid. Renkema stelt op pp. 18-19 dat Engeland en Frankrijk geen grote spellingproblemen hebben, omdat die wordt overgelaten aan de woordenboekenmakers, terwijl in het Duitse taalgebied, waar de overheid de spelling vaststelt, nu verschillende deelstaten een verschillende koers varen. Die problemen in het Duitse taalgebied vloeien echter niet zozeer voort uit het feit dat de overheid de spelling vaststelt, als wel uit het feit dat alle deelstaten en Zwitserland en Oostenrijk vetorecht hebben. Daardoor heeft het lang geduurd voor er iets veranderde (een eeuw), maar – wat erger is – daardoor kunnen ook landen en deelstaten kennelijk weer terugkrabbelen. Overigens bestaan er binnen het Engelse taalgebied ook verschillen tussen Groot-Brittannië en de Verenigde Staten (bijv. *behaviour* vs. *behavior*), samenhangend met het feit dat er verschillende gezaghebbende woordenboeken bestaan (o.a. de OED en Webster). In Frankrijk ligt de situatie weer anders.

Kortom, als je de uniformiteit van een spelling die in meerdere landen of autonome gebieden gebruikt wordt, blijvend wilt garanderen, zul je mijns inziens juist een Taalunie-achtige instelling nodig hebben. Het Duitse en het Engelse taalgebied laten zien hoe het gaat als zo'n overkoepelend orgaan ontbreekt. Dit laat overigens onverlet dat de Taalunie de spelling voorlopig het beste met rust kan laten en verder kan uitbesteden aan een academie zoals door Renkema gewenst.

Gerard Verhoeven

Grubben, Ed en Plattel, Ron (2006). *Klantgericht corresponderen. Handleiding bij het schrijven van enthousiasmerende en overtuigende brieven*. Den Haag: Academic Service. ISBN 9052615632. Prijs: € 21,50 (164 pp.)

“Afgezien van enkele literatoren ‘die toch niets beters te doen hebben’, zijn wij Nederlanders geen liefhebbers van corresponderen of de ‘epistolaire kunst’”. Tot deze conclusie komen Ed Grubben en Ron Plattel, communicatieadviseurs, naar aanleiding van hun ervaringen in de praktijk. Toch versturen Nederlandse organisaties wekelijks miljoenen brieven en wordt de brief vaak maar al te graag als het visitekaartje van het bedrijf gezien. Maar dat ‘visitekaartje’, merken Grubben en Plattel op, ziet er helaas nogal eens beduimd uit.

In hun boek *Klantgericht corresponderen*, dat in april 2006 in een herziene druk verscheen (eerste druk in 1996), willen Grubben en Plattel medewerkers van organisaties technieken aanreiken om ‘zelfstandig brieven te schrijven’, in ‘een soepele, klantvriendelijke stijl’, met ‘creatieve oplossingen voor lastige kwesties’ en een ‘heldere, overzichtelijke structuur’. Zodat medewerkers brieven kunnen schrijven die ‘elke toets der kritiek kunnen doorstaan’. *Klantgericht corresponderen* is dan ook geen ‘brievenboek’ met voorbeeldbrieven geworden, hoewel aan bepaalde briefsoorten die in het bedrijfsleven veel voorkomen, zoals offertes, klachtbehandelingsbrieven en betalingsverzoeken, wel specifiek aandacht is besteed.

Om de schriftelijke communicatie tussen organisatie en cliënt ‘soepel’ en ‘klantvriendelijk’ te laten verlopen, moet volgens de auteurs eerst en vooral ‘strategisch’, ‘klantgericht’, en ‘correct en verzorgd’ worden gecommuniceerd. De eerste helft van het boek is voor de behandeling van deze drie kwesties ingeruimd; nadat in hoofdstuk twee zeven factoren voor ‘een succesvolle correspondentie’ zijn benoemd, worden een

aantal van deze 'succesfactoren' in de hoofdstukken drie tot en met vijf verder uitwerkt.

Om 'strategisch' te kunnen communiceren is het volgens Grubben en Plattel bijvoorbeeld van belang dat de schrijver zich bewust is van de 'context' van de brief: 'welke potentiële lezers behoren tot de doelgroep van de brief', 'wat is de bagage van die lezers', 'welke behoeften heeft de klant'? Maar ook: 'wat moet de lezer na het lezen weten', 'wat moet hij vinden', 'wat moet hij doen'? Pas wanneer deze en andere relevante vragen bevredigend zijn beantwoord, eventueel met behulp van een van de checklisten in het boek, kan met het daadwerkelijke schrijven worden gestart.

Om alle gestelde doelen zo goed mogelijk te bereiken, is volgens de auteurs ook een 'klantgerichte' communicatie vereist. Deze klantgerichtheid bestaat er onder meer uit dat de brief 'toegankelijk' en 'plezierig' is geschreven, 'met een heldere structuur', 'zonder clichés in opening en slot', 'in positieve bewoordingen' en 'in gewone mensentaal'. Maar ook dat de brief 'overtuigend' is, in die zin dat alle voor de hand liggende lezersvragen beantwoord worden en dat die antwoorden op een vriendelijke en zakelijke toon worden gepresenteerd. Hoe de schrijver dit kan aanpakken wordt steeds kort aan de hand van sprekende voorbeelden geïllustreerd.

Last but not least is het volgens Grubben en Plattel ook zaak dat een brief is opgesteld in 'correct en verzorgd' Nederlands. 'Kromme zinnen laten een slordige indruk achter, die nog wel eens naar de kwaliteit van de dienstverlening wordt vertaald'. In zestien pagina's passeren de belangrijkste struikelblokken de revue. Veelvoorkomende taalkwesties worden uitgelegd, de belangrijkste regels voor het gebruik van leestekens op een rij gezet, gangbare conventies voor aanhef en afsluiting van een brief besproken: 'de etiquette is taboe, de presentatie informeel'.

Voor de spellingregels, ook een niet-onbelangrijk struikelblok voor de gemiddelde schrijver, verwijzen de auteurs vooral naar de leidraad bij de WNT.

Nadat de minimumvoorwaarden voor het welslagen van een 'soepele' en 'klantvriendelijke' schriftelijke communicatie zo in de eerste vijf hoofdstukken uiteen zijn gezet, spitst de aandacht zich in de hoofdstukken 6 tot en met 9 op enkele veelvoorkomende briefsoorten in het bedrijfsleven toe. Daarbij verliezen de auteurs de drieslag 'strategisch, klantgericht, correct en verzorgd' niet uit het oog. Zij stellen bijvoorbeeld voor de klant ook vriendelijk te *bedanken* als de opdracht naar de concurrent is gegaan. "Een verloren opdracht is nog geen verloren klant!", houden zij de lezer voor. Voor de medewerker die ondanks alle goede bedoelingen toch genoodzaakt is met standaardteksten te werken, worden in hoofdstuk 10 verschillende aandachtspunten genoemd. Tot slot wordt in het elfde hoofdstuk afgesloten met vijf pagina's klantgerichte 'tips voor e-mail en fax'.

Grubben en Plattel wilden met dit boek 'niet alleen de effectiviteit, maar ook de efficiency van de schrijfinspanning' van medewerkers in het bedrijfsleven vergroten. Of dat is gelukt, is niet op voorhand te zeggen, maar wellicht biedt deze Praktijkgids hier inderdaad mogelijkheden voor.

Marian Pijnenburg

Jaspers, Joep en Weeren Braaksma, Marion van (2005).

Professionals & Schrijfwerk. De focusgerichte tekst als succesfactor in de professionele beroepspraktijk. Den Haag: Sdu Uitgevers. ISBN 90 5261 428 8. Prijs € 29,95 (215 pp.).

De rode draad in *Professionals & Schrijfwerk* is de focusgerichte tekst die de hoger opgeleide lezer aan de hand van dit boek leert

schrijven: focus in plaats van doel om de lezer op zijn eigen niveau aan te spreken en hem te stimuleren om zijn schrijfkwaliteiten verder te ontwikkelen.

Het boek maakt deel uit van de serie *Professionals & hun praktijk* die Academic Service (een imprint van Sdu Uitgevers) uitgeeft: een managementserie door en voor professionals uit de zakelijke dienstverlening. Dit deel uit de serie is geschreven door Joep Jaspers, senior adviseur en trainer schriftelijke bedrijfscommunicatie voor Vergouwen Overduin, en Marion van Weeren Braaksmā, communicatiespecialist, tekstschrijver en schrijfcoach. De auteurs richten zich op medewerkers die beroepshalve veel schrijven, zoals beleidsmedewerkers, accountants, juristen, ingenieurs en andere vakexperts.

Dit is te zien aan de opzet van dit zelfstudieboek annex naslagwerk. De lezer maakt kennis met het belang van een duidelijke doelstelling en een voorlopige structuur als startpunt voor de opzet van een tekst. In de vijf hoofdstukken die volgen, bespreken de auteurs een groot aantal tekstgenres. Zo staat in hoofdstuk 2 de opdrachtreportage centraal, waarin de schrijver rapporteert over onderzoek of beleid. In hoofdstuk 3 worden teksten besproken die je schrijft om jezelf als expert te profileren: artikelen, boeken, lezingen en speeches, terwijl in hoofdstuk 4 de werving van klanten centraal staat. Hierin worden niet alleen handreikingen gegeven voor het schrijven van een effectieve offerte, maar komen ook andere werende teksten aan bod, zoals webteksten, brochures, nieuwsbrieven, *fact sheets*, het cv en de profielschets. Het vijfde hoofdstuk hangt inhoudelijk nauw samen met hoofdstuk 2. Het draait om het schrijven van plannen en verslagen, variërend van ondernemingsplannen en jaarverslagen tot teamplannen en persoonlijke ontwikkelingsplannen. In hoofdstuk 6 komt persoonlijke communicatie via e-mails, brieven, faxen, (internet)formulieren, memo's en korte no-

tities aan de orde. In de bijlage staan negentien 'basisrecepten' ter vergroting van de algemene schrijfvaardigheid, bijvoorbeeld voor het schrijven van een inleiding, voor het adviseren met de causale keten, en voor het vragen, geven en ontvangen van tekstcommentaar.

Bij de uitwerking van *Professionals & Schrijfwerk* is goed rekening gehouden met de doelgroep en het doel van het boek. De auteurs geven niet alleen beknopt de genreconventies weer van teksten waar de lezer mee te maken kan krijgen in zijn beroepspraktijk, maar verlevendigen de tekst met praktijkvoorbeelden uit het werkveld van de lezer. Daarnaast besluit elk hoofdstuk met een kort interview waarin een 'ervaringsdeskundige' vertelt over zijn schrijfervaring en in de vorm van praktische tips de lezer een hart onder de riem steekt: 'Begin met schrijven. Blijf je niet inlezen. De echte wijsheid zit in het doen.' Ook het register is afgestemd op de hoger opgeleide lezer. Deze wordt gevousvoyeerd en regelmatig benaderd in managementtaal: 'Als u gevraagd wordt een beschouwing te schrijven, gebeurt dat doorgaans omdat u geldt als expert op een specifiek gebied. Neemt u zelf het initiatief, dan positioneert u zich tegenover uw lezers en de redactie ten minste impliciet als expert. Stel van tevoren vast of u dat wel wilt, ook met het oog op uw personal brand.'

Of deze ietwat gekunstelde stijl schrijven sexyer maakt? Ik weet het niet. De adviezen in het boek ondersteunen de schrijvende professional in ieder geval bij het uitvoeren van hun schrijftaken. Vooral de tekststructuren van de verschillende genres stimuleren om met een schrijfopdracht aan de slag te gaan. Want zoals de auteurs in de epiloog terecht opmerken over teksten die je beroepsmatig produceert: 'Schrijven is werken, dan volgt de inspiratie vanzelf.'

Everdien Rietstap

Nieuws uit het vakgebied

Oraties

Op vrijdag 8 september heeft prof. dr. J.A.E. Bons, benoemd tot bijzonder hoogleraar Geschiedenis van de retorica, leerstoelgroep Taalbeheersing, argumentatietheorie en retorica, Universiteit van Amsterdam, zijn oratie gehouden getiteld *Gemeenplaatsen in de grafrede van Pericles*.

Benoemingen

Vanaf 1 september 2006 is Leah Polcar als universitair docent aangesteld bij de leerstoelgroep Taalbeheersing, argumentatietheorie en retorica van de Universiteit van Amsterdam. Marcin Lewiński is als Aio aangesteld bij dezelfde leerstoelgroep. Zijn dissertatieonderzoek gaat over strategisch manoeuvreren in internetforums.

Promoties

Sanne van 't Hof promoveerde op vrijdag 6 oktober aan de Universiteit Utrecht op

het proefschrift *From text to talk. Answers and their uptake in standardised survey interviews*. Promotoren waren prof.dr.mr. P.J. van den Hoven en prof.dr. W.P. Drew; co-promotor was Dr. A.J. Koole.

Congressen

Ter gelegenheid van het 25-jarig bestaan organiseert de leerstoelgroep Taalbeheersing van de Universiteit van Utrecht op 9 en 10 november de conferentie *Discourse, Cognition & Communication*. Centraal staan de aspecten 'multidisciplinaireit' en de 'relatie tussen fundamenteen en toepassing'. Informatie is te vinden op: <http://www.let.uu.nl/dcc/>

Abstracts

Volume 28

no. 3 2006

SPECIAL ISSUE: Luuk van Waes, Mariëlle Leijten and Thea van der Geest (guest editors). *Methods and Instruments for Reading and Writing Research*

Thea van der Geest, Mariëlle Leijten and Luuk van Waes

University of Twente, The Netherlands /
University of Antwerp, Belgium

Researching language production and processing with the computer

ABSTRACT: This special issue discusses newly developed applications of ICT which can be used in writing and reading research. These high tech instruments and methods open new perspectives for language researchers. The various articles describe different research techniques: thinking aloud, keystroke logging, eye tracking and neuroimaging. The ICT applications enable researchers to record, visualize and analyze language data in great detail. The possibilities and limitations of these methods and instruments are discussed and illustrated with exemplary studies. Special attention is given to the methodological issues of reactivity and triangulation.

KEYWORDS: research methods, research instruments, language production, language processing, registration methods, observation methods, visualization methods, reading research, writing research, reactivity

Maaïke van den Haak, Menno de Jong and Peter Jan Schellens

University of Twente / Radboud University Nijmegen, The Netherlands

Think-aloud protocols and usability testing: Completeness and reactivity of concurrent think-aloud protocols

ABSTRACT: The think-aloud method has evolved into a widely used tool for a range of studies of reading and writing processes. This paper focuses on the validity of the think-aloud method as a formative evaluation tool. The validity of the method is discussed in terms of reactivity and the degree of completeness of the verbalizations. The paper explores to what extent the nature of the test object may affect reactivity and completeness. To this end, concurrent and retrospective think-aloud protocols were compared for two types of internet applications: an online library catalogue and a municipal web site. Results showed that for both applications the concurrent think-aloud protocols revealed fewer verbalized problems than the retrospective think-aloud protocols. The nature of the test objects did indeed affect the reactivity of the method: the instruction to think aloud while working with the test object led to more observable problems in the task performance of the library catalogue, but not in the task performance of the municipal web site. This result can be explained by the

fact that the two test objects involve different kinds of task performance.

KEYWORDS: concurrent and retrospective think-aloud protocols, validity, reactivity, formative evaluation.

Luuk Van Waes and Mariëlle Leijten
University of Antwerp, Belgium

Observing writing processes with inputlog: A data analysis of the interaction with the text produced so far

ABSTRACT: The use of computers as writing instruments has not only had a profound effect on the writing practice and the attitudes towards writing, it has also created new possibilities for writing research. In this paper we describe Inputlog, a Windows-based logging program. It enables researchers (i) to collect writing process data from different input sources (keyboard, mouse and speech), (ii) to analyze these data from different perspectives (product, process, pausing behavior) and (iii) to integrate them with data logged in other programs. A play module (iv) offers the possibility to review a writing session. In this article we describe the main technical and functional characteristics of Inputlog. To illustrate its possibilities we describe the data from an experiment in which we asked participants to detect and correct discrepancies between the 'text produced so far' (TPSF) and their mental representation of what the text should be. Text production is shown to be influenced most by error span – whether the error spans less or more than two characters – and by mode of writing – whether text has been generated by speech recognition or keyboard. This research also shows how working memory load during error analysis affects strategy choice and success as measured by a series of online measures of text

production. We conclude the paper with a preview of the plans for further developments (<http://www.inputlog.net>).

KEYWORDS: Inputlog, keystroke logging, registration tool, on-line writing processes, writing modes, pause analysis, cognitive processes, error analysis, speech recognition, text produced so far (TPSF), working memory.

Reinier Cozijn
Tilburg University, The Netherlands

Analyzing eye movements in reading research

ABSTRACT: The use of eye movements in reading research is explained in a case study on the influence of bold type font and bullet lists on the comprehension of short explanatory texts. The article focuses on how meaningful reading times can be calculated from eye movements and discusses the computer program *Fixation* which allows for easy analysis of eye movements and data preparation for statistical testing.

KEYWORDS: reading, reading research, eye movements, reading times, measures, software, fixation

Leo Lentz, Pim Mak and Henk Pandder Maat
Utrecht University, The Netherlands

Using eye tracking in usability research

ABSTRACT: Does eye tracking help us to detect problems readers experience with documents? In an experimental setting, respondents read an information leaflet about a financial product. Two independent

variables were used: participants read the leaflet either silently or aloud, and they either did or did not give comments on the document after reading the text. All respondents did a performance test on the comprehension of ten lexical items. The results showed that reading-aloud and the assignment to comment on the document both have consequences for the reading process. The technique used in the experiment did not allow us to detect lexical problems.

KEYWORDS: eye tracking, web usability, think-aloud protocols, text comprehension, comprehension of lexical items

John Hoeks, Petra Hendriks & Gisela Redeker

University of Groningen, the Netherlands

Communication and the brain: using neuroimaging to investigate dialogues

Speakers and listeners involved in a conversation seem to adhere to specific pragmatic assumptions (Grice, 1975). Speakers,

for instance, are supposed to say as much as is necessary, not more, but also not less. In the current study, we investigate how listeners (or better: their brains) react when this pragmatic rule is violated. This sort of question can be answered by making use of what are called 'neuroimaging' methods. By means of neuroimaging the workings of the active brain can be visualized. We present a critical review of the most commonly used neuroimaging methods (PET, fMRI, and ERPs), and we describe an experiment to illustrate the use of ERP registration. In that experiment, participants were presented with written question-answer pairs in which the Gricean Maxim of Quantity was violated. This violation was found to be associated with a specific ERP component that, to our knowledge, has not been found before. In addition, other effects were found that point to thematic and syntactic reanalysis of the critical sentence. We conclude that the violation of the Gricean Maxim of Quantity is signaled immediately, and leads to considerable processing difficulty.

KEYWORDS: psycholinguistics, pragmatics, neuroimaging, PET, fMRI, ERPs

Medewerkers aan dit nummer

R. Cozijn, universitair docent, Faculteit Communicatie en Informatiewetenschappen, Universiteit van Tilburg, Postbus 90153, 5000 LE Tilburg

Th. M. van der Geest, universitair hoofddocent Technische en Professionele Communicatie, Universiteit Twente, Communicatiewetenschap, Postbus 217, 7500 AE Enschede

M. van den Haak, promovenda Communicatiewetenschap Universiteit Twente en docente Engelse taalkunde, Vrije Universiteit, De Boelelaan 1105, 1081 HV, Amsterdam

P. Hendriks, universitair hoofddocent Nederlandse Taal en Cultuur, Rijksuniversiteit Groningen, Oude Kijk in't Jatstraat 26, 9712 EK Groningen

J.C.J. Hoeks, universitair docent Communicatie- en Informatiewetenschappen, Rijksuniversiteit Groningen, Oude Kijk in't Jatstraat 26, 9712 EK Groningen

M. de Jong, universitair hoofddocent, afdeling Technische en Professionele Communicatie, Universiteit Twente, Postbus 217, 7500 AE Enschede

M. Leijten, promovenda Taalbeheersing en mandaatassistente Zakelijke en Technische Communicatie, Faculteit Toegepaste Economische Wetenschappen, Departement Management, Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerpen

L.R. Lentz, senior docent-onderzoeker Taalbeheersing, Universiteit Utrecht, Trans 10, 3512 JK Utrecht

W.M. Mak, universitair docent Taalbeheersing, Universiteit Utrecht, Trans 10, 3512 JK Utrecht

H. Pander Maat, senior-docent-onderzoeker Taalbeheersing, Universiteit Utrecht, Trans 10, 3512 JK Utrecht

M.P.F. Pijnenburg, toegevoegd docent leerstoelgroep Taalbeheersing, argumentatietheorie en retorica, UvA, Spuistraat 134, 1012 VB Amsterdam

G. Redeker, hoogleraar Communicatiekunde, Rijksuniversiteit Groningen, Oude Kijk in't Jatstraat 26, 9712 EK Groningen

E. Rietstap, toegevoegd docent leerstoelgroep Taalbeheersing, argumentatietheorie en retorica, UvA, Spuistraat 134, 1012 VB Amsterdam

J. Sanders, universitair docent Taal en Communicatie, VU, De Boelelaan 1105, 1081 HV Amsterdam

P.J. Schellens, hoogleraar Taalbeheersing, Faculteit der Letteren, Nederlandse Taal en Cultuur, Radboud Universiteit Nijmegen, Postbus 9103, 6500 HD Nijmegen

J. Schilperoord, universitair docent Tekstwetenschap, Faculteit der Letteren, Universiteit van Tilburg, Postbus 90153, 5000 LE Tilburg

G. Verhoeven, adviseur en trainer op het gebied van taal en communicatie, Laan van Meerdervoort 339, 2563 AN Den Haag

L. Van Waes, hoogleraar Zakelijke en Technische communicatie, Faculteit Toegepaste Economische Wetenschappen, Departement Management, Universiteit Antwerpen, Prinsstraat 13, 2000 Antwerpen