

Schrijven met de stem

Bespreking van M. Leijten (2007). *Writing and Speech Recognition. Observing Error Correction Strategies of Professional Writers*. Utrecht: LOT. Proefschrift Universiteit Utrecht.

Het proefschrift van Mariëlle Leijten (verder ML) bestaat uit zeven artikelen, voorafgegaan door een korte inleiding. Daar wordt duidelijk dat met het gepresenteerde onderzoek een dubbel doel wordt gediend. Enerzijds wordt de invloed onderzocht van het gebruik van spraakherkennings-software op het schrijfproces. Wat gebeurt er als schrijvers hun tekst dicteren aan de computer die de gesproken tekst omzet in geschreven tekst die direct op het beeldscherm verschijnt? En hoe gaan schrijvers om met de fouten die de software daarbij maakt? Anderzijds wordt getracht om, gebruik makend van informatie die het gebruik van spraakherkennings-software oplevert, een nieuw licht op het schrijfproces te werpen. In de woorden van ML: "spraakherkenning legt *het vel* van schrijfprocessen bloot." Van de zeven artikelen waren er bij het verschijnen van het proefschrift vier al gepubliceerd, was er één aan een tijdschrift aangeboden en waren de overige twee nog in voorbereiding. Het eerste artikel, ofwel het tweede hoofdstuk is van de hand van ML alleen; de andere zijn coproducties waarbij steeds ook begeleider Luuk van Waes betrokken was.

Samenhang

In de inleiding verduidelijkt ML de organisatie van haar proefschrift door uit te leggen dat er vier secties zijn waarin afzonderlijke onderzoeksprojecten worden beschreven, en dat aan die projecten respectievelijk drie, twee, en twee keer één hoofdstuk is gewijd. De eerste sectie gaat over de invloed van spraakherkenning op adaptatie- en schrijfprocessen van schrijvers voor wie spraakherkennings-software een nieuw fenomeen is, in de tweede sectie worden correctiestrategieën in geïsoleerde zinnen beschreven, in de derde sectie gaat het over correctiestrategieën van professionele gebruikers van spraakherkennings-software, en in de laatste sectie bespreekt ML het (toets)registratieprogramma *Inputlog*. Met dit alles beoogt ML drie onderzoeksvragen te beantwoorden: die naar de invloed van spraakherkenning op het schrijfproces, die naar de invloed van fouten in tot dan toe geproduceerde tekst op de tekstproductie (bij het gebruik van spraakherkenning) en die naar de beste methode(s) om het schrijfproces te observeren en analyseren zonder dat proces te beïnvloeden.

Twee doelstellingen dus, drie hoofdvragen, vier secties en zeven hoofdstukken. ML maakt het haar lezers niet gemakkelijk om direct de rode draad te zien en die ook te blijven volgen. Dat twee van de hoofdstukken (5 en 6) over precies dezelfde data gaan, dat de studies in drie andere hoofdstukken (2, 3 en 4) elkaar voor een deel overlappen en dat het allerlaatste hoofdstuk een methodehoofdstuk is, dat alles maakt het er niet eenvoudiger voor wie een verslag verwacht van een project met een duidelijke kop en een staart: één rapport dus, waarin inleiding, theoretisch kader, methoden, resultaten en discussie op elkaar volgen. Daarmee is niet gezegd dat er tussen de diverse onderdelen in dit boek geen samenhang zou bestaan. Maar er wordt van de lezer wel een flinke inspanning verwacht om die samenhang te ontdekken. De uitvoerige Nederlandse samenvatting aan het eind van de verder Engelstalige dissertatie helpt daarbij gelukkig wel. Daar wordt niet alleen uiteengezet wat er in de diverse hoofdstukken aan de orde was, maar wordt ook per sectie uitgelegd waar het in het desbetreffende deelproject om ging. Ook wordt duidelijk gemaakt wat een lezer die de samenvatting tot het eind bewaart grotendeels zelf moet ontdekken: in de hoofdstukken 2, 3 en 4 gaat het steeds

om dezelfde dataset, verzameld bij twintig professionals (tien advocaten en tien academici) die beginnende gebruikers waren van spraakherkennings-software.

Beginnende gebruikers

De data die ML bij twintig beginnende gebruikers verzamelde, zijn in de eerste drie hoofdstukken die samen de eerste sectie vormen, steeds vanuit een ander perspectief benaderd: dat van de adaptatiestrategie, dat van de leerstijl en dat van de ervaring (met klassieke dicteerapparatuur). In hoofdstuk 2, waarin het gaat om twee advocaten met een verschillende leerstijl die voor het eerst met spraakherkenning aan de slag gaan, blijkt de divergeerder (Bart) gaandeweg zijn initiële schrijfgedrag aan te passen, terwijl de accommodator (Frederik) in hoge mate zijn vertrouwde schrijfstrategie handhaaft - ongeveer zoals ML van gebruikers met deze leerstijlen ook had verwacht.

Hoofdstuk 3 begint met een kwantitatieve analyse van de data van de helft van de participanten (de groep van tien die de langste teksten had geproduceerd) van wie er vijf ervaring hadden met dicteerapparatuur en vijf anderen die ervaring misten. Het hoofdstuk vervolgt met een kwalitatieve case study bij twee van hen: een advocaat mét (Frederik - dezelfde als in hoofdstuk 2) en een advocaat zonder ervaring met dicteerapparatuur (Steven). Uit de resultaten van beide analyses concludeert ML dat spraakherkenning een open schrijfomgeving creëert voor verschillende schrijfstijlen. Wie ervaring heeft met dicteerapparatuur houdt vast aan de gewoontes die men daarmee ontwikkeld heeft. Wie die ervaring niet heeft blijft zich gedragen als een gebruiker van gewone tekstverwerkings-software. Opvallend is dat daarbij geen relatie wordt gelegd met de leerstijl die in hoofdstuk 2 nog centraal stond. Daar was de suggestie dat er aan het gedrag van Frederik weinig veranderde omdat hij een accommodator was; hier heet het dat dat samenhangt met zijn ruime ervaring met dicteerapparatuur. Hoe de verschillende uitkomsten voor Frederik en Steven in hoofdstuk 3 zich laten verenigen met de korte mededeling eerder in dat hoofdstuk dat Frederik en Steven dezelfde leerstijl hadden, wordt niet duidelijk. Moet de lezer het zo begrijpen dat de invloed van leerstijl op een schrijfproces met spraakherkenning, minder groot is dan het effect van ervaring met andere schrijfhulpmiddelen?

In hoofdstuk 4 staan de correctiestrategieën voorop die door gebruikers van spraakherkenning worden gehanteerd. Gevolgd wordt hoe twee schrijvers, de een met en de ander zonder ervaring met dicteerapparatuur, omgaan met de fouten die ze in de tot dan toe geproduceerde tekst op het scherm zien verschijnen. De schrijver met dicteerervaring (dezelfde Frederik als in hoofdstuk 2 en hoofdstuk 3) stelt zijn correcties uit tot een latere fase, de schrijver zonder dicteerervaring (alweer Steven, die we al kennen uit hoofdstuk 2 en 3) anticipeert soms op fouten in de tekst en wisselt vaker van schrijfmodus om fouten te voorkomen. ML concludeert, net zoals ze dat al deed in hoofdstuk 3, dat spraakherkenning een schrijfomgeving creëert die open staat voor verschillende schrijfstijlen.

Experiment met studenten

In de tweede sectie staan ook correctiestrategieën voorop, maar nu gaat het om een project waarin die strategieën werden bestudeerd in een experimenteel gecontroleerde context. Nagegaan werd welke effecten er waren van verschillende soorten fouten (resultierend in bestaande maar onbedoelde woorden zoals *vary us* in plaats van *various*, dan wel in onbestaande woorden) in tot dusver geproduceerde tekst op de cognitieve belasting van de schrijftaak en op het revisiegedrag van de schrijvers. Aan het experiment werd deelgenomen door zestig studenten. Aan alle participanten werd

gevraagd deelzinnen af te maken waar een fout in zat. Daarbij kon het gaan om een woord dat al bestond (een typische fout veroorzaakt door imperfecte spraakherkenningssoftware zoals *eye* in plaats van *l*) of een woord dat niet bestond (een typisch gevolg van een tikfout bij het gebruik van klassieke tekstverwerkingssoftware zoals *streert* in plaats van *street*). In de helft van de gevallen waren de zinnen alleen op een scherm te lezen; voor dertig participanten begon het experiment met deze wijze van presentatie. In de andere gevallen waren de zinnen niet alleen te lezen maar ook te horen; voor de andere dertig participanten begon het experiment met deze wijze van presentatie. Uit de analyses bleek – niet erg verrassend – dat fouten die in bestaande woorden resulteren, moeilijker te identificeren zijn dan fouten die niet-bestaande woorden opleveren.

Een opmerkelijker uitkomst was dat de presentatiewijze van de tot dusver geproduceerde tekst van invloed was op de strategie die schrijvers hanteerden. Wanneer de al geproduceerde zinnen behalve te lezen ook te horen waren, gaven de schrijvers er vaker de voorkeur aan om door te gaan met tekstproductie in plaats van met foutcorrectie, dan wanneer ze de tekst alleen konden lezen. De verklaring van ML is dat schrijvers hun strategie aanpassen bij de mate van zekerheid die ze hebben dat wat er op het scherm staat overeenstemt met, dan wel afwijkt van wat de bedoeling was. Als ze de tot dusver geproduceerde tekst behalve te lezen ook te horen krijgen, is er weinig twijfel mogelijk over wat er wel en niet correct op het scherm is verschenen, en zijn schrijvers er kennelijk geruster op dat ze de fouten die ze zien ook later wel kunnen corrigeren. Buiten bespreking blijft de vraag naar de praktische waarde van dit inzicht. Is het bij spraakherkenningssoftware gebruikelijk dat de schrijver bij de controle van zijn tekst niet alleen naleest wat hij tot nu toe genoteerd heeft, maar daar ook nog eens naar luistert? Ook wordt niet duidelijk hoe de keuze voor studenten als participanten in dit onderzoek zich verhoudt tot de aanduiding 'professional writers' in de ondertitel van het proefschrift.

Over het experiment wordt twee keer gerapporteerd. In hoofdstuk 5 zijn de data bestudeerd met uni-level statistische analyses; in hoofdstuk 6 zijn multi-level analyses gebruikt, waarbij meer recht werd gedaan aan de hiërarchische structuur van de data waarin de aangeboden zinnen waren 'genest' in de participanten. De resultaten van de analyses in hoofdstuk 5 en 6 verschilden niet. Dat impliceert niet dat ML ongelukkig is met de uitkomsten in hoofdstuk 6. De toename in statistische power en daarmee van de waarschijnlijkheid dat de conclusies uit de analyses valide zijn, rechtvaardigde in haar visie de extra inspanningen die de multi-level analyses met zich mee brachten.

Gevorderde gebruikers

In hoofdstuk 7, tevens de derde sectie van het proefschrift, staan tien professionele schrijvers centraal die ervaring hadden in het gebruik van spraakherkenningssoftware. De participanten werkten bij de VRT als, meestal academisch opgeleide, ondertitelaars van anderstalige televisieprogramma's. In dit onderzoek werden ze geobserveerd bij het schrijven van een zakelijk rapport over welzijn op de werkplek. De nadruk in dit deelproject lag op de strategieën die de schrijvers hanteerden om fouten te herstellen. Gebruik werd gemaakt van drie elkaar aanvullende onderzoeksmethoden: analyse van het eindproduct, analyse van het schrijfproces met een toetsregistratieprogramma en een spraakherkenningsprogramma, en protocolanalyse door de participanten zelf - in retrospectieve interviews. Er kwamen drie verschillende patronen van foutcorrectie naar voren. Er zijn schrijvers die proberen alle fouten direct te herstellen, en zo een eerste tekstversie te produceren die zo correct mogelijk is. Andere schrijvers lossen meer dan de helft van de problemen meteen op, maar stellen de andere verbeteringen (niet alleen de oplossing van technische problemen maar ook inhoudelijke revisies) uit tot een later

moment. En de derde groep stelt het merendeel van de correcties, met name van technische problemen, uit tot een tweede schrijfronde.

In hoofdstuk 8, ofwel de vierde sectie, wordt ingezoomd op een van de instrumenten die in het onderzoek uit hoofdstuk 7 werden ingezet om het schrijfproces in beeld te krijgen: het mede door ML ontwikkelde toetsregistratieprogramma *Inputlog*. Met *Inputlog* kunnen schrijfprocessen in Windows-toepassingen worden geregistreerd, en vervolgens vanuit diverse invalshoeken worden geanalyseerd. Zo kunnen er databestanden worden gegenereerd als input voor verdere analyse van tekst-, proces-, pause- en revisiekenmerken van een schrijfsessie, kunnen logging-data uit andere programma's worden geïntegreerd en kunnen de geregistreerde schrijfprocessen met verschillende snelheden worden afgespeeld.

Nogmaals: samenhang

Tot voor kort kwam het in de geesteswetenschappen niet vaak voor dat een proefschrift was opgebouwd uit een aantal artikelen die ook los waren of nog werden gepubliceerd. Van die optie, overgenomen uit de natuur- en de gedragswetenschappen, wordt in de humaniora eigenlijk pas de laatste decennia gebruik gemaakt. Er zijn ook duidelijke voordelen aan verbonden. De uitkomsten van het onderzoek worden wijder verspreid dan wanneer ze alleen in de dissertatie zouden worden gepresenteerd, en ze worden ook eerder gepubliceerd, wat hun actualiteitswaarde verhoogt. De artikelen zijn al eens op een goed niveau gereviewd voordat ze in het proefschrift terecht komen, wat ten goede komt aan het vertrouwen dat promovendus en promotor mogen hebben in de kwaliteit van het onderzoek. En de artikelen kunnen ook op naam van meer dan één auteur zijn gesteld, waarmee meer recht kan worden gedaan aan het werk van de promotor dan wanneer er een dissertatie verschijnt met noodzakelijkerwijs alleen de naam van de promovendus op het titelblad.

Maar met een proefschrift dat uit artikelen bestaat gaan ook risico's gepaard. Daar vormt de dissertatie van ML in een aantal opzichten een illustratie van. Zonder daarvoor duidelijk te zijn gewaarschuwd krijgt de lezer in de eerste sectie drie analyses voorgeschoteld van het schrijfproces van deels dezelfde schrijvers, en worden in de volgende twee hoofdstukken de data van een en hetzelfde experiment twee keer geanalyseerd. Zeker in het laatste geval is dat één keer te veel. Nadat het inzicht was ontstaan dat een multi-variate analyse de voorkeur verdiende boven een uni-variate analyse, had met de presentatie van de uitkomsten van die laatste analyse kunnen, nee moeten worden volstaan. Ook bij andere hoofdstukken dringt de vraag zich op of en hoe ze voortbouwen op wat er direct aan vooraf ging. Te vaak blijft onduidelijk wat de uitkomsten in het ene hoofdstuk voor consequenties hebben voor de vraagstelling en de aanpak waarvoor wordt gekozen in het volgende hoofdstuk. En wie aan het eind van het boek nog eens terugkijkt naar de hoofdvragen die in de inleiding worden genoemd (over de invloed van spraakherkenning op het schrijfproces, de invloed van eerdere fouten, en over de beste methode om het schrijfproces te observeren), die moet constateren dat de antwoorden op die vragen daar niet te vinden zijn.

Alles overziend: de afzonderlijke hoofdstukken zijn als rapportages van de studies die erin worden besproken zeker geslaagd, en het niveau dat wordt bereikt rechtvaardigt zonder twijfel de doctorsgraad die de auteur ermee verdiend heeft. Maar een goed boek is dit proefschrift niet geworden. Daar komt meer voor kijken dan een kaft om een aantal onderzoeksartikelen.

Carel Jansen (hoogleraar Bedrijfscommunicatie aan de Radboud Universiteit Nijmegen)