

Catia Cucchiarini & Rintse van der Werf
Nederlandse Taalunie / Edia – Educatie Technologie
Contact: ccucchiarini@taalunie.org
rintse.vanderwerf@edia.nl

Taal- en spraaktechnologie voor het onderwijs in en van het Nederlands

1. Inleiding

De afgelopen jaren wordt in het onderwijs steeds meer de nadruk gelegd op het belang van maatwerk, diversiteit en differentiatie als elementen om de kwaliteit van het onderwijs te verhogen en om talentontwikkeling te stimuleren. Het bieden van maatwerk binnen de traditionele, klassikale aanpak is echter moeilijk. Daarom is er een groeiende behoefte aan ICT-leermiddelen die leren op maat aanbieden en aan motiveerende interactieve leeromgevingen die rekening houden met het niveau en de specifieke behoeftes van leeders.

ICT biedt mogelijkheden om aan die behoeftes te voldoen, maar als het gaat om taal leren en het ondersteunen daarvan, kunnen de mogelijkheden van ICT verder verrijkt worden met specifieke, op taal gerichte technologie: taal- en spraaktechnologie (TST). TST maakt het namelijk mogelijk om taalproducten – zowel geschreven als mondeling – door computers op een intelligente manier te laten verwerken, te beoordelen en eventueel van feedback te voorzien. TST heeft intussen een stadium bereikt waarin ze rijp en betrouwbaar genoeg wordt geacht om ingezet te worden in het onderwijs, mits er goed rekening wordt gehouden met de beperkingen die dergelijke technologie nog steeds heeft.

TST kan ingezet worden om het onderwijs in en van het Nederlands efficiënter te laten verlopen of om het aantrekkelijker en plezieriger te maken voor de leerlingen. Bijvoorbeeld voor vaardigheden waarbij er veel geoefend moet worden, kan TST gebruikt worden om leerlingen zo vaak als nodig te laten oefenen (het zogenaamde ‘kilometers maken’). Maar TST kan ook op een creatieve manier ingezet worden om interessantere en interactieve programma’s te ontwikkelen.

2. De rol van de Nederlandse Taalunie

De Taalunie wil zich inspannen om eraan bij te dragen dat het onderwijs in en van het Nederlands (basis- tot en met volwassenenonderwijs) in Nederland en Vlaanderen

optimaal gebruik kan maken van de huidige en toekomstige mogelijkheden van TST. Maatschappelijke processen zoals ‘individualisering’, ‘flexibilisering’ en ‘globalisering’ worden door de technologische ontwikkelingen nog verder versterkt. Samen met de vergrijzing leiden die ontwikkelingen tot een groeiende behoefte aan levenslang leren. Voor het onderwijs betekent dat dat er steeds meer vraag zal zijn naar interactieve leerprogramma’s, waarmee leerders in hun eigen omgeving en op hun eigen tempo kunnen leren.

In de veranderende context zullen multimediale leeromgevingen, digitale leermiddelen, virtual reality en serious games steeds belangrijker worden omdat die meer intensieve en motiverende leerervaringen kunnen bieden. Daarin zal TST een prominente rol spelen om de interactie met machines zo natuurlijk mogelijk te laten verlopen. De Taalunie wil ervoor zorgen dat die interactie met virtuele leeromgevingen niet in het Engels gebeurt, maar in het Nederlands.

Het bevorderen van TST voor het Nederlands gebeurt al via het door de Nederlandse en Vlaamse overheid gefinancierde STEVIN-programma, dat de Taalunie coördineert. In dat kader zijn verschillende TST-basisvoorzieningen gerealiseerd die relevant kunnen zijn voor het onderwijs in en van het Nederlands: corpora van gesproken en geschreven taal, software voor automatische spreekvaardigheidsoefening en uitspraakbeoordeling en software voor alfabetiseringsprogramma’s en voor spelling- en grammaticacontrole (ook specifiek voor dyslectische taalgebruikers). Een overzicht hiervan is te vinden in het nummer van het tijdschrift *DIXIT*, dat geheel gewijd is aan het STEVIN-programma en aan de inzet van TST in het onderwijs (http://www.notas.nl/dixit_release_pdfs/DIXIT_STEVIN_Onderwijs.pdf). In deze bijdrage zal een concreet voorbeeld worden gegeven van hoe taaltechnologie ingezet kan worden ten behoeve van het onderwijs: *de Slimme Nieuwslezer*.

3. Lesmateriaal op maat door de inzet van taaltechnologie

Een voorbeeld van hoe taaltechnologie kan worden ingezet in het onderwijs is afkomstig uit het NWO PROO onderzoek ‘MASLA’, dat van 2003 tot 2008 is uitgevoerd aan de Universiteit van Tilburg. In dat project is onderzoek gedaan naar de mogelijkheden voor het op maat aanbieden van digitaal lesmateriaal en naar de leereffecten die daarbij optreden. ‘Op maat’ betekent hier aansluitend bij kenmerken van de leerder, zoals taalniveau en leerstijl, maar ook interesses, voorkeuren en kenmerken zoals eerste taal en onderwijsachtergrond. Maatwerk leveren in het onderwijs is voor docenten een selectieproces dat informatie vereist over (1) de leerder en over (2) het beschikbare lesmateriaal. Voor auteurs van lesmateriaal betekent maatwerk het ontwikkelen van alternatieve leerroutes of aanvullend, extra lesmateriaal dat geschikt is voor bepaalde leerders. Zowel de ontwikkeling van geschikt lesmateriaal als de selectie van het juiste materiaal voor een leerder zijn veeleisende, tijdsintensieve taken. Voor beide taken is in het MASLA-project onderzocht hoe technologie kan helpen. Er zijn prototype toe-

passingen ontwikkeld voor het automatisch genereren van lesmateriaal en er zijn regels geïmplementeerd voor het automatisch selecteren van lesmateriaal, net boven het taalniveau van de leerders. Het leereffect is steeds gemeten door pretest-posttest onderzoek.

4. Het genereren en selecteren van lesmateriaal

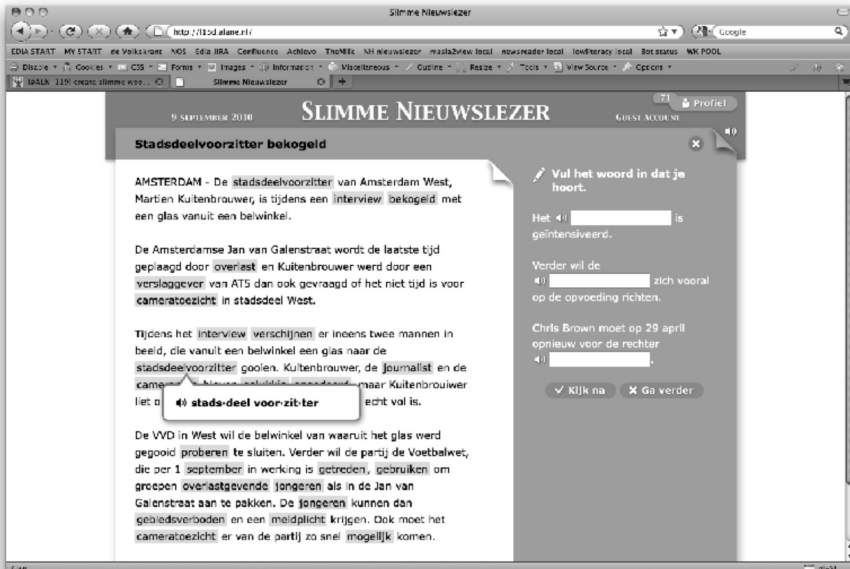
Bij het ontwikkelen van lesmateriaal voor taalmethodes vormen tekstcorpora al vele decennia een belangrijke leidraad voor auteurs, bijvoorbeeld in de selectie van woorden voor een methode. Dergelijke corpora zijn in toenemende mate digitaal en online beschikbaar, waardoor ze ingezet kunnen worden in de automatische talige analyse van teksten. Voorbeelden zijn automatische lemmatisering en 'Part of Speech tagging' (POS tagging) van teksten. Die basisinformatie kan daarna worden gebruikt voor meer geavanceerde analyses zoals 'keyword detectie' en 'tekstclustering'. Door de inzet van die technologie wordt (talige) informatie verzameld die belangrijk is voor het genereren van lesmateriaal. Wat is het onderwerp van een tekst? Waar zitten syntactisch complexe woorden? Welke woorden zijn laagfrequent en welke hoogfrequent? Wat zijn belangrijke eigennamen? Als die kenmerken van lesmateriaal bekend zijn, kan op de juiste plekken hulp en ondersteuning worden toegevoegd. Betekenisinformatie bij belangrijke keywords kan helpen bij tekstbegrip (Hootsen et al. 2007), een voorleesfunctie (spraaktechnologie) helpt laaggeletterden bij het lezen van syntactisch complexe woorden, morfologische regels helpen bij het afleiden van de betekenis van onbekende woorden... (Milliano et al. 2008). Daarnaast stelt taaltechnologie ons in staat om voor een willekeurige tekst taken en opdrachten te genereren die het leren ondersteunen. Een voorbeeld zijn oefeningen, waarbij op basis van de resultaten van POS tagging alle werkwoorden uit een zin worden geselecteerd, die de leerder vervolgens correct moet vervoegen.

Op het moment dat grote hoeveelheden lesmateriaal beschikbaar zijn, wordt de selectie van het juiste materiaal voor de leerder een uitdagende taak. Om het meest effectief te leren, moet materiaal worden geselecteerd dat net boven het niveau van de leerder ligt (i+1) en dat de leerder helpt bij het bereiken van een zeker leerdoel (Krashen 1987). Op het moment dat het taalniveau en de leerdoelen in formele modellen zijn gevat, kan het selectieproces worden geautomatiseerd. Kortom: de inzet van taaltechnologie biedt mogelijkheden om grote hoeveelheden lesmateriaal te verzamelen, te verrijken, te genereren en te selecteren.

5. Een voorbeeld: *De Slimme Nieuwslezer*

De uitkomsten van het MASLA-onderzoek zijn door Edia gebruikt bij het ontwikkelen van een toepassing die automatisch lesmateriaal maakt van alle ruim 1200 nieuws-

berichten die dagelijks online worden gepubliceerd door de belangrijkste dagbladen. Uit die 1200 berichten wordt vervolgens voor elke individuele leerder een selectie gemaakt van berichten die het best passen bij de kenmerken van die leerder. De onderdelen uit het bericht boven het niveau van de leerling worden verrijkt met extra hulp, zoals informatie over de betekenis, over woordopbouw en/of door vertalingen.



Figuur 1: De Slimme Nieuwslezer.

Tijdens de presentatie zal de werking en het gebruik van *De Slimme Nieuwslezer* worden gedemonstreerd. De nieuwslezer is een voorbeeld van hoe, door de inzet van taaltechnologie, nieuwe vormen van maatwerk kunnen worden gerealiseerd: kenmerken van de leerder zijn leidend voor het selecteren en – indien gewenst – het aanpassen of genereren van lesmateriaal (Van der Werf & Vermeer 2008).

Referenties

- Hootsen, G., R. van der Werf & A. Vermeer (2007). "E-learning op maat: automatische geïndividualiseerde materiaalselectie in het tweede-taalonderwijs". In: *Toegepaste Taalwetenschap in Artikelen*, vol. 78, p. 119-130.
- Krashen, S.D. (1987). *Principles and practice in second language acquisition*. Englewoods Cliffs, N.J. [etc.]: Prentice-Hall international.

Van der Werf, R., G. Hootsen & A. Vermeer (2008). "Automated user-centred task selection and input modification in language learning". In: *ITL Review of Applied Linguistics*, jg. 2008, nr. 155, p. 1-22.

Van der Werf, R. & A. Vermeer (2008). "Online kranten, een model van de (tweede-) taalleerder en het genereren van adaptief lesmateriaal". In: *Toegepaste Taalwetenschap in Artikelen*, vol. 80, p. 49-61.

Ronde 2

Dries Tanghe & Philip van Oosten
Hogeschool Gent
Contact: dries.tanghe@hogent.be
philip.vanoosten@hogent.be

Op zoek naar een methode om de leesbaarheid van teksten automatisch te beoordelen

1. Achtergrond: leesbaarheidsonderzoek en leesbaarheidsformules

Het leesbaarheidsonderzoek is ontstaan in de Verenigde Staten in de jaren 1920. De toenemende instroom van immigranten in de scholen zorgde ervoor dat leerlingen steeds vaker met substantieel verschillende leesvaardigheidsniveaus in dezelfde klasgroepen werden ingedeeld. Daardoor werd de vraag naar een manier om tekstmateriaal te kunnen afstemmen op de leesvaardigheid van het individu groter. Leesbaarheidsformules zijn ontstaan als een antwoord op die vraag. Op basis van een aantal kenmerken van een tekst, zoals de gemiddelde zinslengte of het gemiddelde aantal lettergrepen per woord, pretenderen ze een objectieve meting te kunnen geven van de leesbaarheid van een tekst. Voorbeelden zijn de 'Flesch Reading Ease' formule (Flesch 1948) en de 'Flesch-Kincaid Grade Level' formule (Kincaid 1975), die tot op vandaag druk gebruikt worden over de hele wereld. Dat blijkt onder andere uit het feit dat ze standaard zijn ingebouwd in de meeste versies van Microsoft Word¹.

De meeste leesbaarheidsformules werden in de Verenigde Staten ontwikkeld en zijn dan ook eerst en vooral ontworpen voor de beoordeling van Engelstalige teksten. In ons taalgebied paste Brouwer de 'Flesch Reading Ease' formule aan aan het Nederlands, wat resulteerde in de AVI-leesniveaus (Brouwer 1963). Die AVI-niveaus, die op een groot aantal kinder- en jeugdboeken staan aangegeven, zijn bedoeld als leidraad bij de zoektocht naar lesmateriaal dat past bij een bepaalde leesvaardigheid. Na experimenteel onderzoek van Staphorsius werden de AVI-niveaus vernieuwd en wer-