

DIXIT

TIJDSCHRIFT OVER TOEGEPASTE TAAL- EN SPRAAKTECHNOLOGIE

Taal in Bedrijf

Congresverslag toegepaste taal- en spraaktechnologie

Innovatiesubsidies werken averechts

Ofwel: waarom innovatie en subsidie niet samengaan



‘Dat witte paard daar’

*Over communicatie
tussen mens en
machine*

VAN DE MAKERS VAN VAN DALE XIII

van Dale

XIV



NU IN DE
BOEKHANDEL

“Opnieuw een meesterwerk ...
een literaire standaard voor jong en oud.”

Thomas Rosenboom

“Van Dale, de enige vorm
van vandalisme die ik toejuich.”

Harry Mulisch

“Van Dale is voor mij
de moeder van alle woordenboeken.”

Daphne Deckers

IN DRIE DELEN MET CD-ROM

DOOR DRS. TON DEN BOON EN PROF. DR. DIRK GEERAERTS
ETYMOLOGIE DOOR DR. NICOLINE VAN DER SIJS

VAN DALE LEXICOGRAFIE UTRECHT . ANTWERPEN

www.VanDaleXIV.nl

van Dale

INHOUD

Taal in Bedrijf

Op dinsdag 22 november was het zover: de eerste Taal in Bedrijf, de opvolger van Team voor Taal. Dankzij een grote subsidie van de Nederlandse Taalunie, was het mogelijk er een waar spektakel van te maken. Het doel van dit congres was primair om potentiële afnemers van Taal en Spraaktechnologie met leveranciers en ontwikkelaars in contact te brengen. Dat is nodig, want er bestaat nog steeds een hoge mate aan koudwatervrees op het gebied van TST-toepassingen. Dat wilde men wegnemen door vooral échte, dus commerciële gebruikers van TST-applicaties aan het woord te laten. Maar natuurlijk was er ook de nodige ruimte voor netwerken en andere vormen van divertissement. Arjan van Hessen doet verslag.

4

Innovatiesubsidies werken averechts

Innovatie en subsidie gaan niet samen. Weliswaar is extra geld altijd welkom, maar in de praktijk is de weg er naar toe meer dan een barrière. Sterker: het hindert de bedrijfsvoering. Norbert Mergen, ex-ondernemer, gaat voor de gelegenheid weer eens op de stoel van de ondernemer zitten.

7

'Dat witte paard daar'

In de ontwikkeling naar meer robuuste en efficiënte mens-computer systemen speelt het gebruik van meer 'modaliteiten' dan alleen taal, een grote rol. De verwachting is dat de communicatie tussen mens en computer veel sneller en beter verloopt wanneer meer communicatieve middelen dan alleen taal gebruikt worden, omdat mensen dat onderling ook doen bijvoorbeeld met oogcontact en gebaren. Ielka van der Sluis legt uit.

10

En verder...

Overzees shoppen 14
Taaltechnologie: een informatievak 16
NOTaS Nieuws 18

DIXIT is dood, leve DIXIT

2005 is zonder meer een jaar van verandering geweest. Team voor Taal transformeerde in het veel ambitieuzere Taal in Bedrijf, en met de nodige ophef is de spelling gewijzigd. Waar iedereen enthousiast blijkt over de actieve manier waarop onder andere NOTaS de praktische toepassingen van TST heeft belicht, daar is niet iedereen even enthousiast over de nieuwe spelling.

Vlak na de officiële introductie van de nieuwe spellingregels, de spellingshervorming, barstte er een forse discussie in allerlei media los. Een discussie die overigens even snel luwde als dat hij opkwam. De teneur was echter duidelijk: onzinnig, inconsequent, en vooral overbodig. De details zijn mij inmiddels ontschoten, maar het curieuze woord 'ideeëloos' blijft wel bij. Linde van den Bosch, algemeen secretaris van de Nederlandse Taalunie, mag dan in Binnenlands Bestuur (4 november dit jaar) zeggen dat 'dat maar zo'n klein onderdeel is van de taal', het blijft een monstrum. Ik durf dan ook te voorspellen dat bij de volgende spellingshervorming 'ideeënloos' weer de gangbare spelling is.

De belangrijkste verandering, vanuit het perspectief van DIXIT wel te verstaan, gaat pas in 2006 zijn beslag krijgen. DIXIT blijft DIXIT, maar dan anders. Volgend jaar zal er naast een elektronische versie met enige regelmatig een 'papieren' thema-DIXIT verschijnen. Maar de

belangrijkste verandering is dat DIXIT een grote broer krijgt, in de vorm van een jaarboek. Eind volgend jaar, rond september, zal de redactie de belangrijkste ontwikkelingen van het afgelopen jaar en de belangrijkste artikelen op het TST-vakgebied bundelen.

Een jaarboek is eigenlijk een logische aanvulling of extensie, omdat een jaarboek, letterlijk, de ruimte biedt om zowel de diepte als de breedte in te gaan. Om bijvoorbeeld technische ontwikkelingen te duiden en te plaatsen binnen mogelijke toepassingsvelden. Op deze manier denkt de redactie om de huidige ondertitel, 'tijdschrift over toegepaste taal- en spraaktechnologie', nog meer recht te kunnen doen.

Maar goed, dat is pas volgend jaar. Als aperitief kunt u in dit nummer onder andere een uitgebreid verslag lezen van het congres Taal in Bedrijf. Het artikel 'Innovatiesubsidies werken averechts' plaatst serieuze vraagtekens bij de huidige wijze waarop met innovatiegelden wordt omgegaan. En het artikel 'Dat witte paard daar' ten slotte laat zich het best omschrijven als voor de gevorderde TST-liefhebber.

Rest mij iedereen een rustige en hopelijk witte kerst toe te wensen, en natuurlijk een 'ideeërijk' 2006.

Pierre Pieterse
Hoofdredacteur DIXIT

Taal in Bedrijf

Op dinsdag 22 november was het dan zover: Taal in Bedrijf! Deze opvolger van het twee jaar eerder door NOTaS georganiseerde 'Team voor Taal?' werd dankzij een grote subsidie (€50K) van de Nederlandse Taalunie groots aangepakt. In het Evoluon in Eindhoven werden drie sprekers voor het plenaire ochtendprogramma gevraagd en 27 voor de sessies in de middag. Een live-verslag van Arjan van Hessen.

Het doel van deze (niet-NOTaS) bijeenkomst was om (potentiële) afnemers van Taal en Spraaktechnologie met leveranciers en ontwikkelaars in contact te brengen. Er bestaat nog steeds een soort koudwatervrees op het gebied van TST-toepassingen en men wilde proberen die weg te nemen door vooral echte (lees: commerciële) gebruikers van TST-applicaties aan het woord te laten. Uiteraard is er niet sprake van één type TST-applicatie en dus werd een heel scala aan onderwerpen met sprekers geregeld zodat alle onderwerpen in meer of mindere mate aan bod zouden komen, en alle bezoekers op z'n minst één onderwerp van zijn/haar gading zou vinden.

| ARJAN VAN HESSEN |

Wat cijfers

- 370 mensen hadden zich vooraf ingeschreven, 30 deden het de dag zelf zodat we in potentie 400 geïnteresseerden hadden. 30 mensen hadden vooraf laten weten toch niet te kunnen komen en 80 mensen kwamen gewoon niet opdagen, zodat er uiteindelijk 290 deelnemers waren. Gegeven de korte voorbereidingstijd (9 weken) is dit erg netjes!
- Een aantal onderwerpen (overheid & publieke diensten en TST voor mensen met communicatieve beperkingen) werden als dusdanig nuttig bestempeld, dat nu al besloten is er later nog aparte bijeenkomsten over te houden.

OCHTENDPROGRAMMA

De dag begon met een korte film van 3 studenten van de Hogeschool voor Kunsten uit Utrecht. Veel herrie en flitsend gefilmd, gaf het een korte impressie van een aantal bekende toepassingen van taal en spraaktechnologie. Natuurlijk zouden sommigen andere keuzes gemaakt hebben, maar op deze manier kreeg je wel een indruk van hetgeen jonge buitenstaanders er van vinden.

Na een inleidend praatje door Linde van den Bos, voorzitter van de Nederlandse Taalunie, hield Co Berendsen een vertelling. Co, in het dagelijkse leven partner bij Boer & Croon NeXstrategy, vertelde een soort sprookje over het toepassen van Taal en Spraaktechnologie in het verbeteren van klantcontacten. Hij benadrukte dat het essentieel is om op een andere manier na te denken over de manier waarop bedrijven met hun klanten omgaan. In zijn dagelijkse leven komt hij het dikwijls tegen dat men klanten vooral als lastig ziet: ze stellen vragen, kosten tijd, en leiden de aandacht maar af. Deze simpele, maar al te ware boodschap, was waarschijnlijk wel bekend bij veel bezoekers uit het bedrijfsleven, maar bood zeker voor de aanwezige onderzoekers een frisse kijk op de manier waarop het (helaas) vaak toegaat.

Na de pauze, hield Luc Soete een voor mij inspirerende lezing over de veranderende samenleving. Luc Soete, hoogleraar op de Universiteit van Maastricht, had het over de implicaties die de verschuiving van informatiemaatschappij naar communicatie maatschappij voor ons allen heeft. Hij liet zien dat de bijdrage van ICT niet erg zichtbaar is in de welvaartsoverzichten van onze wereld, gegeven de huidige meetmethoden die hiervoor gebruikt worden. Zo introduceerde hij het begrip informatie-inflatie. In alle overzichten wordt slechts gesteld dat informatie beschikbaar is, maar niet wat de kwaliteit ervan is. Die kwaliteit is de laatste jaren enorm verbeterd, maar

het is volstrekt niet duidelijk op welke wijze dat tot een 'betere' samenleving heeft bijgedragen.

Als afsluiter kwam Hans Appel, CTO van Sun Microsystems, met een knallende presentatie over het nut om anders met creativiteit om te gaan. In zijn optiek wordt er bij de inzet van ICT (waaronder taal en spraaktechnologie) nog steeds te veel geredeneerd vanuit het gezichtspunt van een boekhouder. Bedrijven beschermen angstvallig hun producten, en leveranciers van content houden alles het liefst achter slot en grendel. Hans liet zien dat de komst van *open source software* en de enorme verspreidingsmogelijkheden die Internet biedt, dit traditionele model aan het wankelen brengen. Uiteraard moeten bedrijven nog steeds geld verdienen, maar de wijze waarop moet anders.

(Het multimedia bedrijf Noterik heeft het ochtendprogramma gefilmd zodat iedereen dit nog eens rustig kan bekijken. De link naar deze opnamen is: http://www.bestuuronline.nl/taalinbedrijf/tib_22nov.aspx)

LUNCH

De exact 1honderdminuten durende uitstekende lunch was een ideale mogelijkheid om: te netwerken, bijgepraat te worden over net gestarte Stevin-projecten, en om direct in contact te komen met bedrijven die daadwerkelijk taal- en spraaktechnologie ontwikkelen dan wel implementeren. In de grote hal stonden ongeveer twintig informatiestands met TST-gerelateerde bedrijven (Telecats, Knowledge Concepts, Dutcheer, Linguistic Systems, CMGLogica, Human Inference, Linguistica, TextKernel, Q-go en natuurlijk Polderland), TST-organisaties (NOTaS, TST-centrale en de Nederlandse Taalunie) en lopende, TST-gerelateerde onderzoeksprojecten. Samen besloegen deze stands waarschijnlijk 90 procent van het taal- en spraaktechnologie potentieel in Vlaanderen en Nederland!

Het aardige van deze uiteenlopende mix van groepen was dat men eenvoudig met elkaar in contact kwam. De academische wereld kon kennis nemen van de bedrijven (wat doen ze nu met onze onderzoeksresultaten?), en het bedrijfsleven kon laagdrempelig kennismaken van de nieuwe TST-mogelijkheden (wat kunnen we



Luc Soete



Hans Appel





Luc Soete, Linde v.d Bos en
Lisanne Teunissen



Co Berendsen

Was er ook kritiek? Jazeker!

Een veel gehoorde klacht was dat, hoewel de ochtendsprekers zelf wel in orde waren, hun onderwerpen wel wat (erg) ver van taal en spraaktechnologie afstonden: veel deelnemers hadden liever iets meer TST-inhoudelijke praatjes gehoord. Een ander duidelijk teleurstellend feit was het geringe aantal Vlaamse deelnemers (slechts 9%). Na het echec van Lernout & Hauspie zijn er slechts zéér weinig Vlaamse TST-bedrijven overgebleven. Ook is er blijkbaar minder belangstelling van Vlaamse bedrijven voor het toepassen van TST hoewel het feit dat het evenement in Nederland werd georganiseerd en er vooral Nederlandse gebruikers een praatje kwamen houden, hier ongetwijfeld ook een rol bij gespeeld zal hebben. De deelnemende bedrijven ten slotte, waren tevreden en spraken van een hoge kwaliteit van de bezoekers: geen wereldverbeteraars, geen mensen met de ultieme oplossing en geen consultants die zichzelf kwamen aanprijzen, maar geïnteresseerde bezoekers die veel van de hoed en de rand wilde weten.

samen doen? wat kunnen we over een paar jaar verwachten?). Dat dit netwerken aansloeg, bleek wel uit het grote aantal deelnemers dat om stipt 14:00 door de organisatoren naar de parallelle sessies gestuurd moest worden.

MIDDAGPROGRAMMA

De middag was vooral bedoeld om potentiële gebruikers te informeren over de wijze waarop TST-applicaties in bedrijven geïmplementeerd kunnen worden. Het waren ditmaal nadrukkelijk *niet* de makers of de implementators van de applicaties die aan het woord kwamen. Nee, échte gebruikers vertelde over het waarom, het hoe en het uiteindelijk resultaat van het toepassen van TST. Er waren in het totaal 27 workshops verdeeld over 9 verschillende interesse velden. Te weten:

- Onderwijs
- Toerisme, cultuur & vrije tijd
- Transport & logistiek
- Overheid & publieke diensten
- Media
- Gezondheidszorg
- Telecom
- Financiële sector
- TST voor mensen met communicatieve beperkingen.

Hoewel niet alle sessies even druk bezocht werden (met name de laatste sessies van 16:00 tot 17:00 waren wat leeg), was op het einde nog meer dan 50 procent van de genodigden aanwezig.

EVALUATIE

De officiële evaluatie van de dag toonde aan wat de organisatoren al voorvoelden: grosso modo was het een groot succes! Meer dan driekwart van de deelnemers gaf aan het als geslaagd te beschouwen en een volgende keer zeker terug te komen.

Hoewel het, gezien de hoge kosten van dit soort bijeenkomsten, niet zeker is of er binnen de looptijd van het Stevin-project nogmaals zo'n beurs georganiseerd zal worden, zullen er wel 'dedicated follow-ups' komen die voor een kleiner publiek wat dieper op de materie in zullen gaan. Als bovendien de grote aantallen professionele contacten in de nabije toekomst leiden tot een aantal aansprekende TST-projecten, dan zullen alle gestelde doelen gehaald zijn!

Innovatiesubsidies werken averechts

Innovatie en subsidie gaan niet samen. Weliswaar is extra geld altijd welkom, maar in de praktijk is de weg er naar toe meer dan een barrière. Sterker: het hindert de bedrijfsvoering. Norbert Mergen, ex-ondernemer, gaat voor de gelegenheid weer eens op de stoel van de ondernemer zitten.

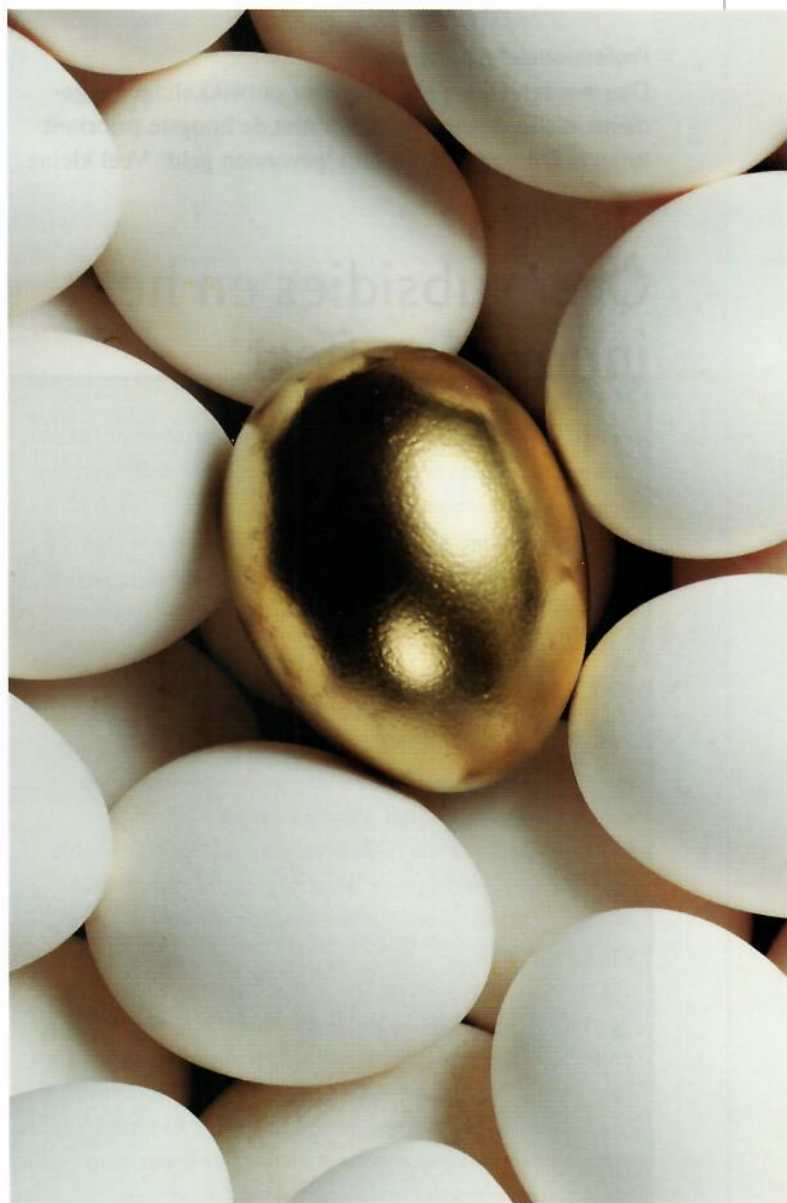
| NORBERT MERGEN |

Ik moet mijn hart luchten. Het wordt me te veel. Recent stond er een bericht in de krant dat de Nederlandse universiteiten snel buitenlands bloed moeten krijgen, omdat een kennisachterstand ontstaat. Kenniseconomie zou een van de beleidsspeerpunten van het kabinet Balkenende II (en een kroonjuweel van D'66) moeten zijn. De minister-president in persoon nam dit onderwerp voor zijn rekening. Maar de Normen en Waarden-discussie had voorrang. Pas op 4 februari 2005 – dat is dus bijna twee jaar na het regeerakkoord – vraagt 'het kabinet het Innovatieplatform het idee voor een innovatieakkoord tussen alle betrokken partijen bij de kenniseconomie, nader uit te werken. Op basis van dat onderzoek beslist het kabinet over deelname.'

Dit doet mijn tenen krommen. We zijn tot een stroperige maatschappij geworden, verstrikt in vormelijkheden, regels en procedures. En iedere ambtenaar die dreigt een beslissing te moeten nemen, kruipt schielijk weg achter stapels met richtlijnen of roept ten einde raad een onderzoeksc commissie in het leven. Zo schiet het niet op.

Gehinderd door subsidie

Ik ben ondernemer (gewees). Dat wil zeggen: mijn zakenpartner en ik, waren tot eind 2003 eigenaar van een middelgroot tamelijk innoverende en best wel succesvolle middelgrote onderneming op het gebied van de taaltechnologie. We deden fundamenteel onderzoek en creëerden



werkelijk nieuwe producten. Hoewel er vanwege de overheid en de EU allerlei subsidie- en stimuleringsregelingen in het leven werden geroepen, hebben we ons daar weinig door laten hinderen. Ja, u leest het goed: hinderen. Want een beetje ondernemer met een goed idee gaat niet meedoen aan een loterij. Voor veel subsidieregelingen is namelijk een gelimiteerd bedrag beschikbaar. Dan geldt 'op is op'. En soms er is een sluitingsdatum. Een bedrijf moet een voorstel indienen. Daarin moet je de woorden gebruiken waarop de beoordelaars geilen – genetische

algoritmen, complexe parser, desambiguering, backtrack-mechanisme, morfologische systemen, multi-level pattern analysing system, enzovoort – want als je dat niet doet, kan je het meteen al vergeten. Als je pech hebt, wordt het plan na maanden beoordeeld en, als je nog meer pech hebt, voor subsidie afgewezen. Hierop bestaat overigens een redelijke kans. Of je krijgt een schier eindeloze reeks vragen van mensen die eigenlijk niet snappen waar het over gaat. Op zulke onzekerheden wil je als ondernemer niet bouwen.

Professionele 'geldlopers'

Dus worden alleen plannen voor ontwikkelingen ingediend, die in de directiekamer niet de hoogste prioriteit krijgen. Dit onder het motto 'gevonden geld'. Veel kleine

Over subsidies en het innovatie-effect

Ik heb dit betoog op 12 september 2005 geschreven. In de tussentijd hebben diverse berichten in de pers de kern van het verhaal geschraagd.

Op 22 september publiceert de Evert Vermeerstichting de resultaten van het onderzoek naar de bestemming van de Europese landbouwsubsidies: de kleine boeren profiteren er nauwelijks van. Concerns als Nestlé, Avebe, Campina haalden elk honderden miljoenen euro's binnen. Maar ook Mars (40 miljoen), KLM, Shell Chemie, Dow Chemicals, DSM Gist, Heineken, Bavaria en Oranjeboom zijn samen goed voor enkele honderden miljoenen, zo meldt NRC Handelsblad. Van de in totaal 1,26 miljard landbouwsubsidie die Nederland jaarlijks krijgt, blijft bar weinig over voor de kleine boeren, waar het uiteindelijk om te doen was. De werkelijke landbouw is er niet beter van geworden.

Op 26 oktober publiceert Trouw een interview met minister Brinkhorst. Ruim baan voor de kenniswerker, luidt de kop. En de subkop: Bureaucratie schaadt Nederlandse economie. De minister pleit ervoor de toelating van buitenlandse kenniswerkers te versoepelen. Op aandrang van het Innovatieplatform, waar de minister zelf deel van uit maakt, was de regeling een jaar geleden al versoepeld, maar dat bleek ontoereikend. De gemeenten stellen zich wat hem betreft te formeel op als het gaat om toelating van 'personen die wat toevoegen'. Daags erna publiceert dezelfde krant een grafiek waaruit blijkt dat Nederland vergeleken met landen als Zweden, de VS, Canada, Duitsland, Finland, Japan, Frankrijk en Groot-Brittannië bijzonder weinig investeert in R&D, software en hoger onderwijs. Duidelijk is dat er structureel iets moet veranderen. En de oplossing is dan niet het binnenhalen van vliegtuigen vol met hoogopgeleide Indiërs. En ook niet het op dezelfde voet doorgaan met het toepassen van subsidieregelingen.

en middelgrote ondernemingen hebben trouwens de tijd en de middelen niet om aan de rompslomp van dat circus mee te doen. Terwijl alom bekend is dat innovatie door met name kleine ondernemingen wordt gedragen. Grotere ondernemingen daarentegen, hebben een speciale man of zelfs een team aangesteld om op deze manier geld binnen te harken. En er zijn adviesbureaus die weinig anders doen dan voor een percentage van het ontvangen bedrag een deel van het werk uit handen nemen. Ik ken bedrijven die overwegend van subsidie bestaan – of bestonden. Vaak universitaire *spin offs*. Zij dienen plan na plan in en creëren zo hun primaire geldstroom. Maar een product gebruiksklaar maken en op de markt brengen kunnen ze niet. Nee, een ondernemer die gelooft in de technische en markthaalbaarheid van zijn idee bewandelt andere wegen dan deze.

En dan heb ik het nog niet gehad over randvoorwaarden die aan diverse regelingen worden gesteld, zoals het samenwerken met buitenlandse ondernemingen of een universiteit – wel leuk voor de *spin offs* op het nabijgelegen *brain park*. Als samenwerking voor de ontwikkeling van een innovatief product nodig is, gebeurt het vanzelf wel. In de praktijk wordt deze echter te vaak gezocht met slechts als enig doel de geldruif leeg te halen.

TST en de NOTaS-oppepper

Voor NOTaS hebben wij, dat wil zeggen vooral wijlen Geert Kobus, ons hard gemaakt voor een flinke oppepper van de taal- en spraaktechnologie in de Lage Landen. Het Nederlands is weliswaar geen wereldtaal, maar ook geen kleine taal. En het verschilt wezenlijk van het Amerikaans/Engels. Behoudens een paar middelgrote bedrijven, wordt de taal- en spraaktechnologie hier voornamelijk gedragen door kleinere ondernemingen. Voor een aantal producten is de Nederlandstalige markt gewoonweg te klein om op economische aantrekkelijke basis onderzoek en ontwikkeling te plegen. De keerzijde hiervan is dat, als er in Nederland en Vlaanderen *geén* eigen ontwikkelingen worden gedaan, de Nederlandse taal in Engels/Amerikaans technologische modellen wordt gepropt. Met de gevolgen van dien. Dus als hier geen rol is weggelegd voor de Nederlandse (en Vlaamse) overheid, zou ik het niet meer weten.

Academische vragen en geleerd navelstaren

We hebben diverse gesprekken gevoerd. Wat mij frappeerde, is de enorme kloof die gaapt tussen enerzijds de universitaire (wetenschappelijke) en ambtelijke wereld en anderzijds het bedrijfsleven. Er heerst achterdocht omdat men elkaar niet begrijpt – beter: niet wil begrijpen. Zwart-wit gesteld heeft de wetenschappelijke wereld er geen flauw benul van hoe bedrijven in de maatschappij opereren. Bij begrippen als 'commercieel haalbaar' trekt men een vies en onbegrijpend gezicht. Er sprake van navelstaren op een eiland. Men weet nauwelijks wat er in bedrijven aan onderzoek en ontwikkeling wordt gedaan, want dat is platvloers – de eigen wetenschappelijke bezigheden

zijn daar ver boven verheven. Het tegendeel is soms waar: in een onderzoek als de BaTaVo (Basis Taal- en spraaktechnologie Voorzieningen) wordt gesteld dat bepaalde essentiële bouwstenen ontbreken en moeten worden ontwikkeld, terwijl deze pakweg zestien jaar geleden al zijn gemaakt in het bedrijfsleven. En op basis van deze BaTaVo is er een subsidieplan gekomen. Ik heb universiteiten bezocht en gezien wat men aan onderzoek doet; soms is men met interessante dingen bezig en soms wordt het wiel opnieuw uitgevonden. Want, hoezeer ook nagestreefd, tot op heden zijn de meeste pogingen om universiteiten met bedrijfsleven samen te laten werken vruchteloos gebleken. Met als gevolg dat veel van datgene dat binnen een universiteit is bereikt na afloop op de plank blijft liggen en in de vergetelheid raakt. Je zou toch zeggen dat opleiding en onderzoek in dienst van de samenleving – in dit geval: bedrijfsleven – moeten staan. Maar het omgekeerde lijkt het geval. Er is door de jaren heen een groot verschil in beleving van de werkelijkheid ontstaan en een vervreemding van de maatschappij. Hierin vinden de ambtelijke en de wetenschappelijke wereld elkaar. Een van de pluspunten van NOTaS is dat het de universitaire wereld en het bedrijfsleven wil binden. Een brug wil slaan en wederzijds begrip wil stimuleren.

Wadlopen in diep slijk

Zo had NOTaS eens een stoutmoedig plan om de ontwikkeling van taal- en spraaktechnologie te bevorderen. Een plan, dat de muur tussen de wetenschappelijke wereld en het bedrijfsleven zou slechten en de taal- en spraaktechnologie de nodige stimulans zou geven. Een plan dat zich bovendien zelf terugverdienende. Van de overheid werd verwacht dat ze het gevraagde kader zou scheppen en een startsubsidie zou geven. Maar de gesprekken met vertegenwoordigers van De Taalunie, EZ, OC&W, NWO, enzovoort verliepen als wadlopen in diep slijk. De Taalunie is een intergouvernementele organisatie die, in principe, moet waken over het Nederlands (in de TST). NWO financiert wetenschappelijk onderzoek aan Nederlandse universiteiten en instituten 'via bijna 170 verschillende subsidies'. EZ wil het bedrijfsleven graag stimuleren, maar voor OC&W gelden andere waarden. Het plan dat NOTaS indiende paste niet binnen bestaande subsidieregelingen en structuren, zei men. Dat was een probleem. Als vertegenwoordiger uit het bedrijfsleven kwam ik er gaandeweg achter dat de geldstromen worden beheerst door de mensen die er het allergrootste belang aan hechten zichzelf, hun afdeling, hun instituut te laten voortbestaan. Je kunt ertegen vechten. Of je kunt meedoen, in de hoop dat er iets wordt bereikt. Dan kost het nog erg veel tijd en is het uiteindelijke rendement gering. Ongeacht welke subsidieregeling dan ook, is het opmerkelijk dat in de commissies die daar toezicht op moeten houden en/of ingediende plannen moeten beoordelen, het authentieke bedrijfsleven niet of nauwelijks is vertegenwoordigd. Of het moeten de wetenschappers zijn, die zich in deeltijd aan een bedrijf hebben verbonden, dan wel de

vertegenwoordigers van een van de universitaire *spin offs* – ons kent ons.

Warme deken der bureaucratie

Het is vervreemdend te ervaren dat wat we in de naoorlogse jaren hebben opgebouwd om de samenleving te faciliteren, zich uiteindelijk tegen het oorspronkelijke doel lijkt te keren. Eind van het liedje is dat een nog steeds goed plan, gedragen door het bedrijfsleven *én* de universiteiten, wordt uitgekleeft en dat wat rest binnen bestaande kaders wordt gewrongen. Want, als puntje bij paaltje komt, kiezen de universitaire instellingen – ook zij die lid van NOTaS zijn – eieren voor hun geld en voor de gebruikelijke weg van de gegarandeerde subsidiestroom. Natuurlijk zijn er enkele ideeën uit het plan overgenomen. Maar niet de kern. En dat wat is gekopieerd, wordt gesmoord door de deken van de ambtenarij, waardoor het doel ervan uiteindelijk niet of slechts ten dele wordt bereikt.

Na BaTaVo STEVIN

Dus nu is er STEVIN (Spraa- en Taaltechnologische Essentiële Voorzieningen In het Nederlands). Een subsidieprogramma voor universiteiten om de in de BaTaVo genoemde ontbrekende essentiële componenten te ontwikkelen (...). Het leeuwendeel van het voor taal- en spraaktechnologie beschikbaar gestelde geld gaat zo weer naar ambtelijke en universitaire instellingen. *If you do what you did, you get what you got.*

De regering heeft het Innovatieplatform in het leven geroepen om de kenniseconomie te verwezenlijken. Haar missie luidt: 'De innovatiekracht van Nederland versterken, zodat ons land in 2010 weer een koploper is in de Europese kenniseconomie. Dat betekent dat Nederland een land moet worden waar volop ruimte is voor excellentie, ambitie en ondernemerschap van mensen en organisaties.' Het is nu bijna 2006.

Ik stel voor dat het Innovatieplatform met spoed de in zestig jaar zo zorgvuldig opgebouwde heilige huisjes omver haalt, waardoor realisme en pragmatisme weer een kans hebben. Want alleen dan zullen subsidies en stimuleringsregelingen het rendement hebben dat wordt beoogd.

Pionieren en innoveren

Vooralsnog verandert er weinig, vrees ik. Gelukkig zijn er, als vanouds, de witte raven; bedrijven die zich opwerpen om als eerste nieuwe technologie te implementeren. Want er is met taal- en spraaktechnologie erg veel mogelijk. Lang niet al datgene dat mogelijk is wordt al toegepast en ontwikkelingen, al dan niet in Nederland geïnitieerd, schreiden voort. Deze bedrijven, die een pioniersrol willen vervullen in de toegepaste technologie, creëren zo een voorsprong op hun concurrenten en genieten in de publiciteit eveneens een innoverend imago. En daar is helemaal niets mis mee. Dixit.

‘Dat witte paard daar’

In de ontwikkeling naar meer robuuste en efficiënte mens-computer systemen speelt het gebruik van meer ‘modaliteiten’ dan alleen taal, een grote rol. De verwachting is dat de communicatie tussen mens en computer veel sneller en beter verloopt wanneer meer communicatieve middelen dan alleen taal gebruikt worden, omdat mensen dat onderling ook doen bijvoorbeeld met oogcontact en gebaren.

De systemen waarin op dit moment de interactie met de gebruiker via meerdere modaliteiten verloopt, zijn helaas niet in evenwicht. Dit komt omdat de ontwikkeling zich vooral heeft gericht op de **interpretatie** van multimodale input (bijvoorbeeld combinaties van taal en penmuis), terwijl de **generatie** van multimodale output weinig aandacht heeft gekregen. In dit artikel wordt geïnventariseerd welke aspecten van belang zijn bij het genereren van multimodale output. Hierbij wordt vooral gekeken naar een kerntaak in de output generatie, namelijk het genereren van referentiële uitingen. Referentiële uitingen, zijn verwijzende uitdrukkingen die mensen gebruiken om objecten te identificeren. Deze uitdrukkingen kunnen uitsluitend bestaan uit taal, zoals ‘dat witte paard onder die boom’, maar het betreffende paard kan ook geïdentificeerd worden door het aan te wijzen, of aan te raken of door een combinatie van taal en wijzen: ‘dit paard hier’ of ‘dat witte paard daar’ waarbij de spreker zijn wijsvinger uitsteekt in de richting van het paard.

TEMPLATES & CONTEXT

In mens-computer systemen steunt de ontwikkeling van multimodale output vooral op de vooruitgang in Natuurlijke Taal Generatie; de taak in de natuurlijke taalverwerking die machinetaal, zoals bijvoor-

beeld de data uit een knowledgebase, omzet naar menselijke taal die uit-gesproken of -geschreven kan worden. De talige output wordt in deze systemen gecombineerd met andere visuele modaliteiten, zoals grafische representatie en gebaren. In de meeste systemen, worden hiervoor elementaire middelen zoals templates gebruikt, die toegepast kunnen worden in simpele slot-filling dialogen. De output heeft zodoende, ongeacht de dialoogcontext, steeds dezelfde **vorm** met een variabele **inhoud**. In bijvoorbeeld een schaak-hulpprogramma kan de zin “Je kunt X naar Y verplaatsen” gegenereerd worden, waarbij in plaats van X een schaakstuk (toren, koningin) ingevuld kan worden en Y ruimte geeft voor de nieuwe positie van het stuk (e2-f4). Een systeem kan zo’n eenvoudig template vaker gebruiken door alleen de referentiële uitingen op de posities X en Y aan te passen. Meer ontwikkelde systemen vereisen echter contextgevoelige generatie, waarbij de output afhankelijk wordt gesteld van de specifieke situatie in de interactie. In het schaakvoorbeeld is ‘het paard’ als variabele X moeilijk te interpreteren als de gebruiker twee paarden op het bord heeft staan, terwijl ‘het paard aan de linkerkant’ of ‘het paard op positie b4’ in zo’n geval misschien wel een unieke beschrijving geeft.

SYNCHRONISATIE

In multimodale systemen moet de output van meerdere modaliteiten zodanig gecombineerd worden dat de gebruiker deze gemakkelijk kan interpreteren. Zo zou in een grafische representatie van het schaakbord het systeem kunnen zeggen: “Dit paard kun je naar positie Y verplaatsen”, terwijl het paard en positie Y even gaan knipperen (blinking). Op deze manier kunnen de talige en de grafische output gesynchroniseerd worden.

Twee belangrijke vragen die in de ontwikkeling van geavanceerdere multimodale presentaties een rol spelen zijn daarom:

| IELKA VAN DER SLUIS |

(1) Op welke manier wordt de generatie van multimodale uitingen bepaald door de context? en (2) Welke factoren bepalen het gebruik van een modaliteit of een combinatie van modaliteiten? Naar aanleiding van deze vragen wordt in dit artikel ingegaan op de manier waarop menscomputer communicatie verbeterd kan worden.

OBJECT IDENTIFICATIE

Een taak in veel multimodale systemen, zoals hierboven gedemonstreerd met het schaakhulpprogramma is het identificeren van een bepaald object in een visuele context; een context die zowel voor de gebruiker als voor het systeem toegankelijk is. Er zijn verschillende manieren om een object te identificeren. Object identificatie kan bijvoorbeeld gebeuren door een paar eigenschappen van het object te noemen, of door het object op het scherm te laten oplichten (highlighting) of knipperen. Een andere mogelijkheid is het gebruik van een Embodied Conversational Agent (ECA); een computeranimatie met 'menselijke eigenschappen', die met de gebruiker op een natuurlijke manier kan communiceren middels spraak en gebaar. Zo'n ECA kan objecten **aanwijzen** in combinatie met een talige referentiële uiting. In situaties waarin een volledig talige beschrijving erg complex zou zijn (bijvoorbeeld in een domein met veel identieke objecten), kan highlighting of een wijsactie de meest efficiënte manier zijn om een object te identificeren. Vergeleken met de uiting 'de derde pion van links in de tweede rij' is het aanwijzen van de pion een stuk overzichtelijker. Bij het ontwerpen van meer geavanceerde systemen is het van belang om te weten hoe de taal en/of gebaren aangeboden moeten worden zodat de gebruiker begrijpt wat het systeem bedoelt. Om dit goed te kunnen doen, is het noodzakelijk om te weten hoe zulke multimodale referentiële uitingen door mensen worden geproduceerd.

MENSELIJKE MULTIMODALE UITINGEN

Referentiële, context-afhankelijke uitingen die een wijsactie bevatten komen in menselijke communicatie veel voor. De **context** van deze uitingen kan simpelweg opgesplitst worden in een 'discourse context' (wat wordt er gezegd) en een 'perceptive context' (wat wordt er waargeno-



men). In het algemeen is er meer talige informatie nodig om een object te identificeren dat zich niet in de discourse context bevindt, dan om een object te identificeren waarover recent gesproken is.

Bijvoorbeeld in de output “Wanneer je het paard op b4 verplaatst, wordt het paard op b4 niet geslagen”, zou de tweede referentie naar ‘het paard op b4’ vervangen kunnen worden met ‘het’ of ‘het paard’. Een andere belangrijke factor is de hoeveelheid **inspanning** die zowel voor de spreker als voor de hoorder nodig is om een object te identificeren. In coöperatieve dialogen probeert een spreker om tegelijkertijd zijn eigen inspanning en de inspanning van de hoorder zo klein mogelijk te maken. Het doel van de spreker is om de identificatie voor de hoorder gemakkelijk te maken en tegelijkertijd zijn eigen inspanning bij het produceren van de referentiële uiting te minimaliseren door precies voldoende (niet te veel en niet te weinig) informatie te geven. Identificatie van het paard met behulp van de referentiële uiting ‘het witte paard aan de linkerkant van het bord op positie b4 dat bedreigd wordt door de zwarte loper’ kost zowel om te produceren als om te interpreteren nodeloos veel moeite wanneer het stuk ook uniek te beschrijven is met ‘het paard op b4’.

Behalve op het in balans brengen van de hoeveelheid informatie heeft de benodigde inspanning ook betrekking op het soort informatie dat gebruikt wordt. In sommige gevallen is een wijsactie de optimale manier om aan een object te refereren, terwijl in andere gevallen een talige beschrijving of een combinatie van de twee veel geschikter is. Bovendien volgt uit het minimaliseren van inspanning niet dat sprekers objecten zo minimaal mogelijk beschrijven. Het blijkt juist, dat mensen geneigd zijn om objecten te ‘over-specificeren’; om identificatie voor de hoorder makkelijker te maken, wordt meer informatie toegevoegd dan strikt noodzakelijk is, bijvoorbeeld door te zeggen ‘het witte paard aan de linkerkant op b4’, terwijl ‘het paard op b4’ voldoende zou zijn.

AUTOMATISCHE MULTIMODALE UITINGEN

Huidige algoritmes die multimodale referentiële uitingen genereren, produceren meestal tamelijk simpele beschrijvingen,

waarbij relatief rechtlijnige, context-onafhankelijke criteria bepalen of een wijsactie toegevoegd moet worden of niet. In de meeste gevallen worden nauwkeurige wijsacties gegenereerd die een object eenduidig identificeren, waarbij de talige informatie enkel een lidwoord en een zelfstandig naamwoord bevat. In het geval van het schaakhulpprogramma zou dat ‘het paard’ zijn, waarbij het betreffende paard gehighlight wordt of waarbij een ECA het paard met zijn wijsvinger aanraakt. Meer geavanceerde algoritmes zouden gebruik moeten maken van contextgestuurde modellen die een nauw verband beschrijven tussen de talige informatie en de wijsacties. Hierbij zouden, net als in menselijke communicatie, verschillende **typen wijsacties** gegenereerd kunnen worden: nauwkeurige en onnauwkeurige. Het bereik van een nauwkeurige wijsactie bevat precies 1 object, terwijl het bereik van een onnauwkeurige wijsactie meer objecten kan bevatten. Stel je bijvoorbeeld een schaakhulp-ECA voor die aan de rechterkant van het schaakbord staat. Wanneer deze ECA aan ‘het paard op b4’ wil refereren, zal hij om het schaakbord heen moeten lopen om het paard nauwkeurig aan te wijzen. Het zou daarom veel handiger zijn als de ECA zich niet in alle gevallen zou moeten verplaatsen, maar ook vanaf een afstand het paard kan aanwijzen (‘het witte paard aan de andere kant’). Een directe consequentie van zo’n model is dat het aantal talige eigenschappen en het type wijsactie dat nodig is om een onderscheidende multimodale referentiële uiting te genereren, wederzijds afhankelijk zijn.

PROMINENTIE

Bij het genereren van een referentiële uiting, kan afhankelijk van de context bepaald worden welk type wijsactie in de huidige situatie het meest geschikt is en hoeveel talige informatie er gebruikt moet worden. In de context kan bijvoorbeeld gedefinieerd worden welke objecten prominent zijn en welke niet. De prominentie van objecten hangt daarbij af van tenminste drie factoren:

- Talige context: in de context van dit artikel is het paard tamelijk prominent en hoeft op dit moment in de tekst daarom niet meer beschreven te worden als ‘het witte paard op b4’.

- **Perceptuele focus:** wanneer er net aan het paard gerefereerd is, zijn de stukken die in de buurt van het paard staan wellicht iets prominenter dan de stukken die zich aan de andere kant van het bord bevinden.
- **Inherente prominentie:** vanwege opvallende eigenschappen springen bepaalde objecten in een domein meer in het oog dan andere; op het schaakbord is dat bijvoorbeeld de koning, omdat elke speler daar maar één van heeft.

In menselijke communicatie hebben prominente objecten veelal **minder informatie** nodig voor identificatie dan niet prominente objecten. De koning van de schaker kan daarom simpelweg aangeduid worden met 'de koning', terwijl één van de acht pionnen een uitgebreidere beschrijving vereist dan 'de pion'. Door de hoeveelheid inspanning die nodig is voor object-identificatie te combineren met de prominentie van het object in het domein, kan bepaald worden welke referentiële uiting een algoritme het beste kan genereren in een gegeven context.

EVALUATIE

Evaluatie van algoritmes die referentiële uitingen genereren is moeilijk; linguïstische corpora zijn slecht bruikbaar omdat hierin de objecten, hun eigenschappen en hun context niet bekend zijn. Evaluatie van multimodale referentiële uitingen is nog moeilijker vanwege de schaarste van multimodale corpora en vanwege het feit dat de redenen waarom sprekers besluiten welke modaliteit te gebruiken, verborgen zijn. Deze problemen zijn op te lossen door het gebruik van **productie-experimenten** waarin de deelnemers objecten identificeren met behulp van spraak en gebaar. Op deze manier kan, op basis van gecontroleerde input, spontane multimodale data verzameld worden waarmee de output van een algoritme vergeleken kan worden.

CONCLUSIE

Op weg naar geavanceerdere multimodale mens-computer systemen kun je zeggen dat er alleen al voor wat betreft referentiële uitingen nog een hoop te doen valt. In de huidige systemen, die veelal gebruik maken van templates en slot-filling, kan de communicatie sterk verbeterd worden

door de implementatie van algoritmen die gebruik maken van informatie uit de context. Toevoeging van wijsacties aan de referentiële uitingen zorgt hierbij voor een natuurlijke interactie, zoals gebruikers in mens-mens communicatie ook gewend zijn. Door het gebruik van wijsacties kunnen complexe talige beschrijvingen, die zowel het systeem als de gebruiker veel moeite kosten, worden vermeden. Aan de andere kant roept het toevoegen van wijsacties in de output ook nieuwe vragen op met betrekking tot het gebruik van wijsacties en de combinatie van wijsacties en taal. In dit artikel zijn een aantal aspecten aangestipt die van belang zijn in de ontwikkeling van algoritmes die multimodale referentiële uitingen genereren. Zo is vastgesteld dat het belangrijk is om zowel de discourse context als de perceptive context te modelleren. De prominentie van de objecten in het domein speelt hierbij een grote rol. Daarnaast is de inspanning die gevraagd kan worden van de gebruiker een geschikte maat om te bepalen hoeveel en wat voor soort informatie gegenereerd moet worden om een object te identificeren. Om er achter te komen of de output van een systeem voor de gebruiker werkelijk sneller en beter te interpreteren is, is het echter van belang om te weten wat er precies gebeurt in de menselijke communicatie. Dit vereist het nodige onderzoek naar de manier waarop mensen aan objecten refereren: wat ze zeggen, wanneer ze wijzen en op wat voor manier ze hun wijsacties en wat ze zeggen met elkaar combineren.



Dit artikel is een vereenvoudigde samenvatting van Ielka's proefschrift 'Multimodal Reference, studies in the automatic generation of multimodal referring expressions' dat op 19 December 2005 aan de Universiteit van Tilburg is verdedigd.

Overzees shoppen

| KHALIL SIMA'AN |

Ongeveer een jaar geleden kreeg ik een email van Owen Rambow (Columbia University, New York) met een uitnodiging om deel te nemen aan de 'Summer Language Engineering Workshop' aan Johns Hopkins University (JHU). Deze workshop duurt zes volle weken en staat bekend als zeer intensief. Het kostte mij ook meer dan zes weken de tijd om tot een besluit te komen omtrent mijn deelname. In eerste instantie wilde ik graag voor maximaal drie weken deel nemen aan de workshop. Dat was geen optie, zo bleek al gauw, en ik moest toch aanwezig zijn voor de volle duur van de workshop, of helemaal niet. Het was Baltimore of Amsterdam gedurende de hele zomer. Het werd Baltimore.

De 2005-editie (de 13de) van deze workshop behoort tot een serie workshops die wordt georganiseerd door het 'Centre for Language and Speech Processing' (CLSP),

JHU in Baltimore. De initiatiefnemer, professor Fred Jelinek, alom bekend van de informatietheorie, spraakherkenning en probabilistische taalmodellen, heeft tot doel gesteld het verbeteren van de 'state-of-the-art' in de taaltechnologie. Hoewel het een workshop heet, is dit geen workshop in de bekende zin van het woord (zoals een soort miniconferentie).

Bij deze serie workshops gaat het jaarlijks om drie gerelateerde projecten, ieder in een bepaald subgebied van taalverwerking en taaltechnologie. Deze drie projecten zijn geselecteerd in een peer-to-peer selectie systeem uit ongeveer dertig projectaanvragen. Naast de hoofdaanvragers worden er wereldwijd drie tot vijf onderzoekers per project uitgenodigd om een het projectteam te versterken. Vervolgens worden er een aantal Ph.D. en 'ungraduate' (bachelor en master) studenten gekozen aan de hand van een landelijke competitie.

De workshop wordt gesponsord door het National Science Foundation (NSF), en wel zeer royaal voor Nederlandse begrippen. Voor zes weken moet de CLSP-groep woon- en werkruimtes bieden aan 32 mensen, er moet computerarchitectuur aangepast worden (clusters nota bene), reisschema's bijgehouden worden, gastsprekers gezocht en uitgenodigd worden, pre-workshop bijeenkomsten worden gehouden ter voorbereiding, en budgetten worden beheerd voor reis en dagelijkse behoeftes. Er zijn maar zeer weinig onderzoeksgroepen ter wereld die een workshop van deze omvang kunnen organiseren, en ook nog zo geslaagd!

Dit jaar vormde 'statistical parsing' (parsing = ontleden) het hoofdthema voor drie projecten:

- machinaalvertalen met behulp van parsing;
- detectie van 'structurele events' in spraak met behulp van parsing (het



- detecteren van zinsgrenzen en het markeren van onvolkomenheden van gesproken taal zoals herhalingen, correcties, incomplete phrases enzovoort);
- en het verkrijgen van een statistische parser voor een taal (Arabische dialect - gesproken) waarvoor weinig resources bestaan uit de resources van een gerelateerde taal waarvoor meer resources bestaan (Modern Standaard Arabisch - geschreven).

Voorafgaand aan de zes weken van de workshop wordt er een 'summerschool' georganiseerd om vooral de studenten voor te bereiden op de inhoud van de workshops. Daarna barsten er zes hectische weken los! De teams hebben elkaar al twee keer ontmoet ter voorbereiding, maar zodra de workshop aanvangt, begint het echte werk.

In een gigantische zaal zijn er drie hoeken ingericht met bureaus, stoelen en computers voor de drie teams. Een aantal andere kamers wordt ingericht als lounge-ruimtes waar eten en drank beschikbaar zijn voor de vele momenten van frustratie. Door de open werkruimte ben je gauw bekend met iedere andere onderzoeker, zodat je na zes weken interactie wel iedereen erg goed leert kennen. En omdat Baltimore niet echt New York is, zorgt de CLSP groep voor avondprogramma's en levert het informatie over leuke restaurants en gezellige bars.

En de zes weken gaan voorbij zonder dat ik het eigenlijk heb gemerkt: het werktempo is zo hoog, de interactie zo intens, de discussies zo vruchtbaar, de lezingen zo interessant en de teams zo hecht dat er weinig tijd is om over andere zaken dan onderzoek te denken. En dat blijkt nog leuk ook!

Terug in Amsterdam, waar het gezellig is en waar bier beter smaakt, bekruipt mij als

onderzoeker een gevoel van onmacht. Zes weken heb ik gezien hoe taaltechnologie wordt bedreven als een volwaardige topsport waarin zowel de overheid als bedrijven (zoals IBM) veel investeren. Dat blijkt geen onderzoek te zijn dat gericht is op 'het Amerikaans Engels', maar onderzoek naar taaltechnologie als zodanig. En de onderzoeker is de koning, althans voor even. Dat staat in scherp contrast tot de gewone dagen van de onderzoeker op zijn thuisbasis, waar hij dubbel zoveel tijd moet besteden aan onderwijs dan collega's in de VS met een vergelijkbare functie.

Oké, ons bier is beter, maar wat taaltechnologie betreft lijkt ons land een 'minikenisland' te willen blijven. Waar het budget te klein is, er minachting heerst voor technologie en engineering, de taalkunde geen relatie vertoont tot de taaltechnologie, de computerlinguïstiek weinig met computermodellen te maken heeft, de taalverwerking tussen (de Exacte) wal en (de Geesteswetenschappen) schip is gevallen, zal het moeilijk zijn om deze lange zin automatisch te ontleden. Maar let vooral niet op mijn woorden, want ik ben nog maar een paar maanden terug uit Baltimore!

Khalil Sima' an is universitair docent natuurlijke taalverwerking en statistische leeralgoritmen en is verbonden aan de Universiteit van Amsterdam



Taaltechnologie: een informaticavak

Midden jaren negentig deed ik mijn promotieonderzoek als AIO van de vakgroep Informatica van de Universiteit Maastricht, toen nog Rijksuniversiteit Limburg. Ik voelde me er thuis. De vakgroep was een levendige speeltuin van hackers en nerds, die zich hevig konden opwinden over caching-methoden voor schaakcomputers, tweedimensionale representaties van hypertextstructuren, rotatie-invariante visuele objectherkenning, en zelflerende robots die balletjes voor zich uit trapt en elkaar vreemde talen aanleerden.

Als ik vertelde over mijn werk rondom de automatische inductie van beslissingsbomen, dan kreeg ik vragen om meer uitleg, verwijzingen naar gerelateerd werk, en tips om mijn code sneller te maken. Als ik vertelde over mijn toepassingsgebied taal, dan luisterde men geduldig. Taal was mijn ding; mijn collega's werkten met andere domeinen als rechtspraak, spelletjes, en beeldherkenning.

| ANTAL VAN DEN BOSCH |

Vakgroepvoorzitter professor Jaap van den Herik liet zijn vakgroep breed werken op het snijvlak van de kracht van menselijke cognitie en brute computerkracht. Taalverwerking is een voor de hand liggend onderdeel van dit brede onderzoeks-

gebied, dat ooit heel stoer 'kunstmatige intelligentie' werd genoemd.

Zoals ik eerder heb beargumenteerd in deze column, is computerlinguïstiek (en taaltechnologie, maar ik ga niet soebatten over een mogelijk verschil tussen deze termen) een informaticavak. Het is daarom vaak ondergebracht in informaticafaculteiten. De Nederlandse situatie, en die van sommige van de ons direct omringende landen in West-Europa, wijkt hiervan af: in ons land zijn we de letterenfaculteitmedewerkers die 'iets met computers' doen, en die voor het gemak met de taalkundigen en de logici worden gecombineerd, met de typische historisch achterwaartse gerichte denkwijze die ons (West-)Europeanen zo scherp onderscheidt van de Noord-Amerikanen.

Proceedings

De Proceedings van de ACL van 2005 zijn na te lezen op <http://acl.ldc.upenn.edu/P/P05/>. Op de website van de ACL anthology (<http://acl.ldc.upenn.edu/>), een voortreffelijk initiatief om alle proceedings die ooit in internationale bijeenkomsten over computerlinguïstiek zijn verschenen, elektronisch beschikbaar te stellen. Zonder een specifieke paper aan te duiden (behalve wellicht het 'best paper' van David Chiang, A Hierarchical Phrase-Based Model for Statistical Machine Translation) beveel ik de lezer van harte aan eens door de lijst van 77 geaccepteerde papers te bladeren.

Goed, we zijn 66 taalkundigen, en de informatica-aspecten in ons werk staan grotendeels ten dienste van de ontdekking en ontwikkeling van taalkundige kennis. Laat ik de navelstreng niet doorknippen. In feite zijn we bezig de taalkunde in groeiende mate te voorzien van een informaticageblaseerd instrumentarium. Ook de niet-computationele taalkundigen om me heen erkennen dat meer en meer, en hoewel deze sociolinguïsten, psycholinguïsten en taaltypologen zich niet snel



informatici zullen noemen, zijn ze zich hard aan het inwerken in programmeertalen, methoden voor corpusverzameling en -analyse, en methoden voor de creatie en evaluatie van databaseerde voorspelende modellen.

Mijn argument verslapt naarmate ik meer toegeef dat we ons in een multidisciplinair, veranderend vakgebied bevinden dat zichzelf enigszins aan het heruitvinden is op basis van nieuwe golven van empirische onderzoeksmethoden en de opkomst van 'cognitive science' als een wetenschapsgebied. Nee, ik meende wel degelijk wat ik eerder schreef: ons vakgebied is een informaticavak.

Mijn stelligheid is gevoed door mijn bezoek aan ACL-2005, de grootste jaarlijkse conferentie van computerlinguïsten, dit jaar gehouden in Ann Arbor, een studentenstadje nabij Detroit, eind juni van dit jaar. Het is een conferentie waar ik me thuis voel om precies dezelfde reden dat ik me ooit in Maastricht thuis voelde: hier lopen informatici rond met state-of-the-art

kennis op het gebied van leeralgoritmes en statistische modelleertechnieken.

Toevallig, zou je bijna zeggen, passen ze het allemaal toe op taal, waarvan men alles weet wat nodig is. Maar de nieuwigheid in al het werk van vandaag zit 'm in de algoritmes en de modellen, en de hoge accuraatheden die ermee gehaald worden in de realistische evaluaties van tegenwoordig.

De ACL van dit jaar markeerde ook het punt waarop (en de uitzondering bevestigt de regel) voor het eerst alle gepresenteerde papers gebruik maakten van empirische datagedreven methoden. Het gebied verandert, en boekt vooruitgang, door voortschrijdende technologie die een algemenere achtergrond heeft in de informatica. Belangrijk is ook om vast te stellen dat het gebied niet alleen volgt, maar ook bijdraagt aan die ontwikkelingen. Evenals de bio-informatica is de computerlinguïstiek bijvoorbeeld een bekende en gewaardeerde bron van ontwikkelingen op het gebied van het analyseren van sequenties (of dat nu letterreeksen of DNA-sequenties zijn).

We feliciteren...

Successvolle STEVIN-indienaars

Veel NOTaS-deelnemers hebben een erg druk jaar achter de rug. We feliciteren de diverse deelnemers met hun succesvolle aanvragen binnen STEVIN. Zowel bij de tweede ronde voor de grote onderzoeksprojecten als bij de eerste ronde voor demonstratieprojecten zijn NOTaS-deelnemers betrokken, of zelfs projectleiders bij gehonoreerde voorstellen. Het STEVIN-programma blijft voor allen die in Nederland en Vlaanderen betrokken zijn bij taal- en spraaktechnologie een heel belangrijke bron van financiering, en een katalysator bij het stimuleren van nieuwe samenwerkingsverbanden tussen kennisinstellingen en bedrijven. De verschillende netwerkbijeenkomsten geven telkens weer een mooie kans om met wetenschappers, onderzoekers en ontwikkelaars te brainstormen over toekomstige ontwikkelingen en trends op het gebied van taal- en spraaktechnologie.

We danken...

Het bestuur

Achter de schermen is het bestuur van NOTaS ook in 2005 druk bezig geweest. Aan het eind van dit jaar nemen we afscheid van Rik Schuts van Van Dale Lexicografie. Hij wordt opgevolgd door Hans Jongebloed van Dutchear. De rest van het bestuur blijft onveranderd en bestaat uit Dr Jeannine Beeken (TST-Centrale), Professor Harry Bunt (Universiteit van Tilburg), Dr Arjan van Hessen (Universiteit van Twente en Telecats), Debbie Kenyon-Jackson (Cumlingua Language for Life), Dr Nelleke Oostdijk (CLST, Radboud Universiteit Nijmegen), en Jan Schalkwijk (Knowledge Concepts). NOTaS dankt alle bestuursleden voor hun tijd en inzet in 2005.

De organisatie van Taal in Bedrijf

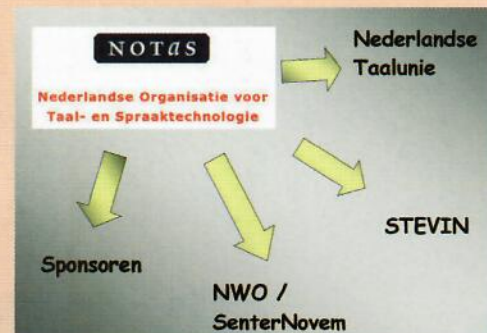
We zijn ook Martin Spitholt (Telecats), Holger Wandt (Human Inference), en Arjan van Hessen dankbaar voor hun inspanningen namens de deelnemers om het NOTaS-initiatief Taal in Bedrijf een groot succes te maken. Hierover leest u meer in deze DIXIT. Dankzij financieel en organisatorisch steun van STEVIN, NTU en SenterNovem hebben meer dan 300 deelnemers een boeiende leerzame dag gehad. De rol van NTU als make-laar en schakelaar is weer met succes ingezet.

We verheugen ons op...

STEVIN-resultaten

De resultaten van een aantal projecten uit de eerste STEVIN-ronde komen in 2006 vrij. Ook de eerste demonstratieprojecten hopen we in het najaar te kunnen presenteren. Er komt een tweede, grote ronde voor demonstratie-

projecten waarin we hopen nog meer hoogwaardige voorstellen van verschillende NOTaS-deelnemers te zien.



De e-DIXIT en het TST-jaarboek

Het wordt een grote stap en we kijken uit naar de elektronische versie van DIXIT die volgend jaar voor het eerst zal verschijnen. In het bijzonder zijn we erg enthousiast over het eerste TST-Jaarboek dat in het najaar van 2006 wordt gepubliceerd. Zorg ervoor dat u ook hierin komt te staan en dat u op de verzendlijst staat. Arjan van Hessen blijft namens het bestuur deze belangrijk taak trekken.

De 'new-look' deelnemersbijeenkomsten

De volgende algemene NOTaS-deelnemersvergadering wordt gehouden op vrijdag 10 februari om 10 uur bij Dutchear in Delft (www.dutchear.nl). Aansluitend is er ook een netwerklunch om kennis te maken met nieuwe deelnemers en ideeën voor 'new-look' bijeenkomsten in 2006 uit te wisselen. Het bestuur komt op 27 januari 2006 bij elkaar.

In de tussentijd blijft de website van www.taaluniversum.org/tst interessant om in de gaten houden. U kunt daar onder andere overzichten van TST-events vinden. Regelmatige updates uit de wereld van taal- en spraaktechnologie zijn ook online beschikbaar op <http://taaluniversum.org/taal/technologie/nieuwsbrief/>

En last but not least...

...wenst het bestuur u, uw gezinnen en collega's een heel fijne Kerst en een succesvol, gezond en gelukkig 2006 toe.

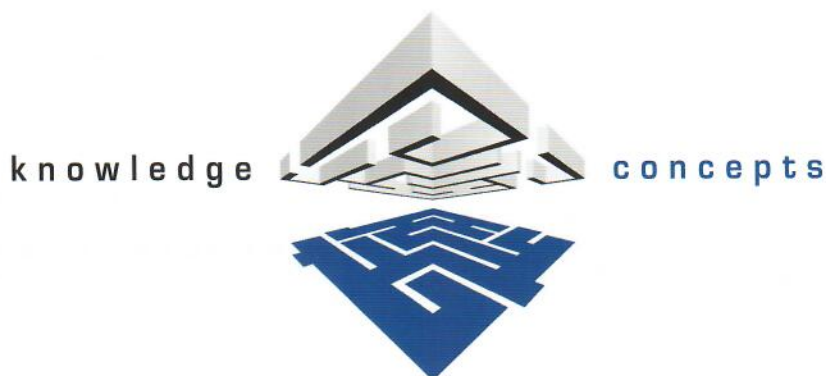
Debbie Kenyon-Jackson, Voorzitter NOTaS

NOTaS

Deelnemers van de Stichting NOTaS zijn:

Human Inference Enterprise B.V. • Universiteit van Tilburg
Faculteit Letteren • TNO Management Consultants
• Knowledge Concepts • Polderland Language & Speech
Technology B.V. • TeleCats • TNO-TPD • Universiteit Twente,
HMI • Van Dale Lexicografie • Dutchear • Q-go B.V.
• Asknow Solutions B.V. • Comsys B.V. • Radboud Universiteit,
CLST • FNB • Stichting Instituut voor Nederlandse Lexicologie,
TST-Centrale • Em@ilco
Sponsor: Nederlandse Taalunie
(Meer informatie: www.notas.nl)

Wij werken aan de derde generatie Intranet



De Handboog 9, 5283 WR Boxtel, Nederland
Tel: +31(0)411 610802 | Fax: +31(0)411 611027
E-mail: info@knowledge-concepts.com | Web: www.knowledge-concepts.com



Een prikkelende vraag aan alle verzekeraars:

'U doet er alles aan om het uw klant zo gemakkelijk mogelijk te maken... maar waarom lukt dat niet via uw website?'

Uw website is een uitstekend middel voor het verkopen van verzekeringsproducten maar dan moeten uw klanten het verhaal wél begrijpen. Veel verzekeraars maken het hun klanten niet makkelijk een keuze te maken en direct online een verzekering af te sluiten. Een gemiste kans.

Met de oplossingen van Q-go kunt u online de dialoog met uw klant aangaan. Het resultaat is dat u online meer verkoopt en betere service verleent tegen lagere kosten. In tegenstelling tot zoekmachines begrijpt Q-go de betekenis van vragen die klanten stellen en geeft meteen gericht antwoord. U genereert gedetailleerde klantinformatie waarmee u de interactie met uw klanten al binnen enkele weken volledig optimaliseert.

Q-go is dé toonaangevende Europese aanbieder van oplossingen voor Online Marketing & Selfservice. Met onze oplossingen helpen wij bedrijven meer rendement te halen uit hun online dienstverlening.

Op dit moment verwerken wij bijna vier miljoen vragen per maand voor o.a.:



Wilt u weten wat Q-go voor u kan doen? Kijk op www.q-go.nl of bel 020 531 3800.



Online Marketing & Self Service