

# **\*TALŌ's LANGUAGE TECHNOLOGY**

## **Die taal en die passende tegnologie van \*TALŌ**

*Dr.J.C.Woestenburg,*

*\*TALŌ b.v.,  
Lijsterlaan 379,  
1403 AZ Bussum,  
The Netherlands.*

*tel: +31 35 69 32 801  
fax: +31 35 69 75 993  
e-mail: [info@talo.nl](mailto:info@talo.nl)  
<http://www.talo.nl/>*

\*TALŌ, wat in 1986 in Nederland gestig en ook daar gesetel is, spesialiseer in taal. Hierdie algehele betrokkenheid by taal lê ingebed in die naam van ons organisasie: \*TALŌ is die rekonstruksie van 'n Germaanse oervorm, wat self nie oorgelewer is *nie* — vandaar die asterisk vooraan. Van die vorm TALŌ, wat beteken 'inkeping op 'n kerfstof', stam ons woorde TAAL, TEL en VERTEL.

Ons het die \*TALŌ woordskeidings- en speltoetsingsprogramme soos beskou vanuit 'n psigologiese, kognitiewe en taalperspektief, bedink, ontwerp en uitgebou. Die stelsels is gebaseer op die basiseenhede van 'n taal, naamlik die onderliggende lettergrepe. Die stelsel is toepasbaar op elke alfabetiese taal binne die Indo-Europese taalfamilie. By die ontwerp van die stelsels het ons natuurlik voorsiening gemaak vir die probleme wat spruit uit die groot onderlinge verskille tussen die lettergreepverdeling van die onderskeie Europese tale. Ons het dit die taalmodel genoem, 'n beskrywing van die wesentlike elemente wat die taal bepaal en die trefsekerheid bevorder.

### **Woordskeiding wat rekening hou met betekenis**

Ons het in 1986 die eerste woordskeidingsprogram vir WordPerfect ontwikkel en daarna die beginsels op die ander Europese tale toegepas. Ons het sedertdien 42 taalmodules ontwikkel, en ondersteun ook minderheids- en klein tale.

Die woordskeidings, gebaseer op 'n taalmodel, verskil in belangrike mate van die Amerikaanse insig om elke taal sonder diepere kennis van die taal self, te behandel. So 'n insiglose benadering lewer moeilikhede op vir die tipies Afri-

kaanse gebruik om onder sekere omstandighede nie die dubbele s in samestellings te gebruik nie. Ons skryf nie "korings-skuur" nie, maar "koring-skuur"; tog is dit wel "Afrikaans-spreker". Ons sal die samestelling nie skei as "koringskuur" nie, selfs al bestaan die woord "kuur". Die lettergrepe word dus deur betekenis bepaal.

## **Speltoetsing wat rekening hou met betekenis**

Die speltoetser is eweneens gebaseer op die eenheid van die lettergreep en nie op afsonderlike willekeurige kombinasies van letters nie. Ons het ook nie die fout gemaak van die aanmaak van permutasies van woorde nie. So mag dus nie gesê word dat twee woorde vier nuwe kombinasies oplewer nie. Sommige van die kombinasies is leksikale onsin en dra nie by tot sinvolle aanwysings nie. Die kombinasies veroorsaak selfs foute. Ons kom hierop terug.

## **Die kompleksiteit**

Vanweë die gemeenskaplike herkoms van al die Europese tale kan die beginsel in gelyke mate toegepas word op sowel Afrikaanse as Nederlands. Die rede waarom Afrikaans nie aan die hand van die Engelse metode ontleed kan word nie, is tog duidelik. Anders as Engels word ons samestellings aaneengeskrif. Die grens tussen die samestellings word bepaal deur betekenis en deur eie gewoonte sedert die ontwikkeling van die taal van Jan van Riebeeck.

So sal die toetsing van "registrasietoestel" negatief wees en "registreertoestel" korrek, maar die Amerikaanse sagteware ken nie die woord des onderskeids nie en wil beide vorme aanvaar. Die leksikale betekenisverskille is wesentlik. 'n Voorbeeld uit Fins is: Jumala (God) in die woord "jumalakeskeinen" wat gedeel moet word as "jumala-keskeinen" (teosentries). "Jumalanilma" beteken egter dreigende weer en is Gods besit oor die weer, te skei as "jumalan-ilma" en nie as "jumala-nilma" nie.

Dit beteken dat kombinasies van woorde die verbode vrug is vir snelle oplossings van toetsing en skeiding, met ander woorde van foute.

\*TALŌ het sig daarom ten doel gestel om woordeboeke saam te stel met woorde wat eg gebruik word in die praktyk en baie volledig is. Dit beteken dat woordeboeke alle verbuigings en vorme van woorde besit, hoewel die aantal en kompleksiteit tussen tale verskil.

Die grootste woordeboek is Sweeds met 1.4 miljoen woorde, gevolg deur Duits met 850.000 woorde. Nederlands (420.000 woorde) en Afrikaans (220.000 woorde) is kleiner, omdat die gebruik van die naamvalle ontbreek en omdat die werkwoorde in Afrikaans minder vorme ken. Die woordeboeke word reëlmatig uitgebrei met nuwe woorde. In die Europese gemeenskap is BSE (malbeessiekte) aktueel en as gevolg daarvan ontstaan nuutskeppings, iets wat ongetwyfeld

ook in Suid-Afrika sal voorkom. Ons brei die woordeboeke dan ook reëlmatig uit.

## **Kombinasies, wat lei tot lukrake nuutskeppings**

Amerikaanse speltoetsingsprogramme maak ontledings van samestellings en nuutskeppings. Die programme probeer toets of die basiselemente van dié woorde bestaan. As die basiswoorde bestaan, word die onbekende woord as korrek aanvaar. Of anders word 'n lys voorstelle gemaak. Die lys voorstelle wat afgelei is van kombinasies, is veelal onsinnig. Soms word vergeet om spasies te laat, byvoorbeeld "decalulators" en kry dan die Nederlandse alternatiewe "dencalulators" (van denneboom), "dekcalulators" (dek van 'n skip) "odecalulators" ('n liefdesode), "theecalulators" (tee).

Die volgende soort fout is ook tiperend: 'n Nederlandse joernalis het gepraat van "tulpenmanie", wat toe deur die beroemde sagteware aanvaar is. Nou word dié voorbeeld in Nederlands as "tulpomanie" geskryf. Vir Afrikaans is die benadering toevallig juis "tulp(e)manie", maar ons sal skryf "Turkomanne" en nie "Turkemanne" nie. Die belang word verder onderstreep deur 'n vergissing as "tulp-mania" in die Landbouweekblad van Januarie vanjaar.

Die probleme van die kombinasies word oor die algemeen groter namate die woordlengte toeneem (bosluisontwikkelingsinhibeerder, enkelrigtingberoepsopleidingskolleges, hoëronderrysbestuursinligtingstelsel, katoenpluismeulenaarsvereniging, melkbeesprestasietoetsadvieskomitee). Die oplossings om lang woorde altyd te aanvaar en te swyg, kan ons nie begryp nie.

Ons het egter daarna gestreef om omvattende woordeboeke saam te stel, wat ten minste die vraag beantwoord hoe gespel sal word met gelyksoortige woorde as voorbeeld. Die probleme met samestellings het die gewoonte laat ontstaan om koppeltekens te gebruik as onderbrekingsteken aan die einde van 'n geskrewe reël (katoenpluis-meulenaarsvereniging of katoen-pluismeulenaarsvereniging), na gelang van waar die Amerikaanse sagteware 'n skeidingsfout gemaak het. Dié gebruik onderstreep die waarde wat ons heg aan volledige woordeboeke en die hoë noukeurigheid van woordskeidingstelsels. Noukeurigheid maak die tussenvoegings van koppeltekens oorbodig. Die probleem het ons ook in ander tale gekonstateer, en stylboeke van koerante probeer oorbodige tekens nou vermy (Stylboek NRC Handelsblad, 2000).

## **In druk en geskrif**

Ons speltoetsings- en woordskeidings-XTensions vir die grafiese program QuarkXPress het een van die belangrikste toepassings geword. Ons het die toepassings ontwikkel tesame met Techno Design, 'n firma wat spesialiseer in die druk- en publisasiewêreld. 'n Belangrik aspek was die sukses wat behaal is in die hervorming van die spelreëls vir Nederlands, Duits en Frans. Ten opsigte

van Noors is daar ook 'n hervorming aangekondig. Veral die XTension het veel bygedra tot oplossings vir die drukkers- en uitgewersbedryf.

Soos te verwagte, het die hervorming van die spelreëls nie sonder verwarring ontstaan nie. In Nederlands het die Taalunie (die regeringsinstansie) en Van Dale ('n outoriteit op leksikografiese gebied) elk sy eie reëls saamgestel. In Duitsland is die nuwe spelreëls eers na 'n uitspraak deur die Hoogste Hof (das Verfassungsgericht) aanvaar. Die koerante en nuusdienste het egter hul eie reëls saamgestel. In die lig daarvan het ons dus verskillende woordeboeke saamgestel.

Vir Suid-Afrika kom die taal-XTension met beide hooftale. Ons het nie vergeet van die belang/invloed van Engels nie, maar hoop om Afrikaans sodoende te bevorder, omdat ons gemerk het dat weinig sagtewareprogramme Afrikaans verstaan. Ons wil die drukkerswêreld aanmoedig om die taal XTensions eers 'n maand lank te beproef.

\*TALŌ het onlangs die speltoetsing uitgebrei na kombinasies van woorde, woordgroepe of rangskikkings. Die nuwe stelstel is bedoel vir koerante en ander instellings wat 'n eie stylboek elk na eie wens wil gebruik.

Ons het vir Afrikaans voorbeelde van die moontlikhede aangegee. So kan die hoofletter in onder andere plaasname byvoorbeeld gekontroleer word: Frans se Kraal, asook in Eerste Straat, Drie Werve; of as 'n joernalis "Grand Slam-titel" skryf en die koerant wil dit "Grand-Slam-titel" spel. Maar ook die foutiewe gebruik van ou naamvalle kan vasgestel word: "ten behoeve van", nie "te behoeve van". Natuurlik is die gebruik van leksikale rangskikkings van besondere belang, omdat die betekenisse bepaal hoe daar gespel word.

Vir Windows kan die Euro Speller 3.0 per e-pos by [info@talo.nl](mailto:info@talo.nl) aangevra word. Die voorbeeld woordskeidingsprogram "die EuroAsiaHyphenator" en EuroAsiaSpeller 5.0.2 kan afgelaai word van <http://www.talo.nl/> en die QuarkXPress XTensions is geplaas op die webtuiste van ons mede-ontwikkelaar by <http://www.techno-design.com/>. Dit is verder moontlik om spesiale uitgawes te ontwikkel in samewerking met Afrikaanse bedrywe.

Ons probeer op hierdie wyse in woord en daad te ondersteun. Ons sou ook graag persoonlik met u wou kennis maak.

Bussum, 15 Februarie 2001