

Ingenieur van de taal

Gents ingenieur maakt eerste corpus-gebaseerde woordenboek voor Afrikaanse talen

ZUID-AFRIKA. Tot 11 april zit hij in Zuid-Afrika: de Gentse dr. ing. Gilles-Maurice de Schryver MSc. Bruggen noch wegen hebben iets te maken met zijn uitgebreide bezoek aan Kaapstad en Pretoria. Aanleiding is een wel heel speciaal woordenboek dat de ingenieur van de Universiteit Gent heeft samengesteld: een vertalend schoolwoordenboek Noord-Sotho - Engels, gepubliceerd door de gerenommeerde Oxford University Press.

De Vlaamse ingenieur lanceert daarmee het eerste corpusgebaseerde woordenboek voor Afrikaanse talen ter wereld. Voorwaar een wonderbaarlijk huzarenstuk.

"Toch niet", reageert Gilles-Maurice de Schryver. "Dat is uiteindelijk niet zo vreemd. Het is zoals vragen: 'Hoe belandt een ingenieur bij Google?' of een tijdje terug: 'Waarom vraagt Microsoft aan ingenieurs om spellingcheckers samen te stellen?'. Tegenwoordig heeft men alsmear meer ingenieurs nodig om met taal om te gaan. Ik heb in dat opzicht 'geluk', omdat ik naast een diploma als micro-elektronicus ook een diploma in de linguïstiek van de Afrikaanse talen behaalde. Een ideaal uitgangspunt dus om als computerlinguïst aan het werk te gaan. Startpunt is steeds de bouw van grote elektronische taalcorpora, die dan met eigen technieken gemanipuleerd worden. Om er zo woordenlij-

sten allerhande, al dan niet gelemmatiseerd en/of gemarkeerd voor woordklassen, uit te puren. Die zaken zijn goud waard in onderhandelingen met Google, Microsoft en dus ook met woordenboekuitgevers. Zonder mijn achtergrond als ingenieur zou ik nooit het soort woordenboeken kunnen maken dat ik nu maak."



Aan het werk met een moedertaalspreker in Pretoria.

Warm gevoel

Gilles-Maurice de Schryver is verbonden aan de vakgroep Afrikaanse talen en culturen van de UGent. Hij maakt, analyseert en gebruikt elektronische corpora om een nieuw soort woordenboeken samen te stellen voor de Bantoe-talen. In zijn doctoraat ontwierp hij daarvoor een stappenplan en hij illustreerde hoe dat moet worden gebruikt voor de officiële talen van Zuid-Afrika. Nu, drie jaar later, ligt zijn eerste woordenboek dat volgens die methode is samengesteld, in de winkel: een vertalend Noord-Sotho - Engels schoolwoordenboek, gepubliceerd door Oxford University Press, zonder enige twijfel de belangrijkste en grootste uitgever voor referentiewerken.

Gilles-Maurice de Schryver: "In 1987 kwam het eerste corpusgebaseerde woordenboek ter wereld uit, Cobuild, een verklarend woordenboek voor het Engels. Dat project vond heel wat navolging, overal ter wereld, maar nog niet op het Afrikaans continent. Nu, twintig jaar later, is dat



Overhandiging van het woordenboek aan de Belgische Ambassadeur in Zuid-Afrika, Jan F. Mutton.



Gilles-Maurice de Schryver met eigen werk in de Protea-boekhandel in Pretoria

Bio Gilles-Maurice de Schryver

Ing. Gilles-Maurice de Schryver werd geboren op 15 september 1971 in Sint-Amandsberg. Hij behaalde achtereenvolgens een master of science in de micro-elektronica (1995), een master of arts in corpuslexicografie (1999) en een doctoraat in de computationele linguïstiek (2005).

Hij is auteur of coauteur van 12 boeken en 69 peer-reviewed artikels, vooral over corpus- en woordenboekonderwerpen rond Afrikaanse talen en taalonafhankelijke lexicografiesoftware.

De Schryver heeft 86 artikels op conferenties voorgesteld en heeft zelf tien conferenties en seminaries georganiseerd.

Hij doceert Afrikaanse talen aan de Universiteit Gent en is buitengewoon professor aan het Xhosa-departement van de University of the Western Cape, Zuid-Afrika.

Tot slot is hij ook stichtend lid van de African Language Technology en bestuurslid van Afrilex, Euralex en Asialex, initiator van PangaeaLex en medestichter van TshwaneDJe HLT.

PD



Compare-merge in Zuid-Afrika.

Noord-Sotho-woordenboek het eerste corpusgebaseerde woordenboek voor een Afrikaanse taal. Dat geeft alvast mij een warm gevoel."

**“Tegenwoordig heeft men
alsmaar meer ingenieurs nodig
om met taal om te gaan.
Ik heb in dat opzicht 'geluk',
omdat ik naast een diploma
als micro-elektronicus ook
een diploma in de linguïstiek
van de Afrikaanse talen
behaalde.”**



Marketingmeeting in Kaapstad.

Bantoe

Concreet betekent het dat de selectie van alle trefwoorden gebaseerd is op gebruiksfrequenties.

De Schryver: "Op microstructureel niveau werden de betekenissen afgeleid uit het gebruik zoals te zien in de corpora, en die betekenissen werden geordend volgens hun frequentie."

Tot slot komt elke voorbeeldzin ook uit het corpus, zodat 'echt' taalgebruik geïllustreerd wordt.

"Voor de Bantoe-talen en het Noord-

Sotho in het bijzonder heb ik ook heel wat nieuwe conventies ontworpen om op een gebruiksvriendelijke manier de stap van een Bantoe-grammatica naar die van het Engels - en vice versa - te kunnen zetten", zegt de linguïst-ingenieur.

"Naast het feit dat dit woordenboek volledig gebaseerd is op corpusmateriaal werd het ook binnen het kader van 'Simultaneous feedback'/ 'Simultane terugkoppeling' samengesteld met behulp van de woordenboeksoftware TshwaneLex. Dat werd ontwikkeld in het kader van mijn doctoraatsonderzoek. Ook werden zelfontworpen 'multidimensionale lexicografische raders' gebruikt om de woordenboektekst over het alfabet correct uit te balanceren."

De Schryver - what's in a name - werkt op dit ogenblik aan een tweede woordenboek en een derde staat ook al op de sporen.

Meer informatie: www.oxford.co.za

**Tekst: Peter DUPONT
Foto's: Ing. Gilles-Maurice
DE SCHRYVER MSc**

WAAR

start employer branding?

GO wijst carrièrestarters de kortste weg naar hun eerste werkgever. Dus werkgevers die hun rekruteringscommunicatie en branding willen afstemmen op de aanstormende talenten van morgen kunnen vertrouwen op GO.

Meer informatie: Ron Maes, ron.maes@roularta.be

GO^{BE}
EMPLOYER BRANDING