

Boekbesprekingen

M.N.M. Boot, Homographie, ein Beitrag zur automatischen Wortklassenzuweisung in der Computerlinguistik, Academisch proefschrift, Utrecht 1979

Boot ontwikkelde een procedure ten behoeve van de automatische woordklassetoewijzing van structuurwoorden in het Duits. Dit kan volgens de auteur ook uitgebreid worden op inhoudswoorden en voor talen die met het Duits vergelijkbaar zijn, wat van groot belang kan zijn voor sociale wetenschappers die met geschreven teksten te maken hebben.

Met de term "homographie" wordt het probleem van de automatische woordklastoewijzing aangeduid dat daaruit bestaat dat gelijk geschreven woorden verschillende betekenissen kunnen hebben. Door middel van de door Boot ontwikkelde procedure lijkt dit probleem van de meerduidigheid van structuurwoorden tot een redelijke oplossing gekomen te zijn. (Bij herhaling van de procedure op een willekeurig gekozen tekst bedroeg het percentage correcte oplossingen 99,4%.)

In het eerste hoofdstuk wordt de geschiedenis van de automatische tekstanalyse behandeld. Voorts worden de gangbare linguïstische modellen voor automatische tekstanalyse besproken (hoofdstuk 2). In het derde hoofdstuk wordt het gehanteerde model geïntroduceerd. Boots ICS-model bestaat uit twee componenten t.w. een contextvrije input component (IC) en een gedeelte dat gebruik maakt van de zgn. string properties (S). De IC component bepaalt voor iedere woordsoort afzonderlijk (zoals bijv. lidwoord, prepositie etc.) de syntactische en grafematische frasestructuur. De paragrafen 3.1.3.4.5 en 3.3.5.1 geven een illustratie van de IC component.

De string component heeft betrekking op de virtuele plaats van het woord in de phrase en wordt door middel van de informatie uit de IC component bepaald.

Op grond van de beschrijving van de andere modellen (hoofdstuk 2) valt op te maken dat het gebruik van extra linguïstische (=grafematische) informatie een interessante vernieuwing is.

Aan de hand van het bovenvermelde regelmechanisme worden beslissingen genomen die tot interpretatie leiden. Daarna wordt de beslissingsstructuur omgezet in een computerprogramma dat als output de codering van de woordklasse oplevert.

Hoofdstuk 4 illustreert de procedure aan de hand van konkrete toepassingen. De programma's zijn in PL/1 geschreven.

Samenvattend valt op te merken dat Boots procedure niet alleen aandacht van computerlinguïsten verdient maar ook van sociale wetenschappers die in automatisering van inhoudsanalyse geïnteresseerd zijn.

I.N. Gallhofer