

HET
PROEFSCHRIFT
VAN

• • •
Gosse Bouma

Computertaalkunde

Redactie: Wim Daniëls

Er is een groeiende behoefte aan software-produkten die gebruik maken van taalkundige kennis. Bekende voorbeelden van dergelijke produkten zijn tekstverwerkers die automatisch woorden afbreken en spelling corrigeren. Daarnaast zijn er ook programma's die in het onderwijs gebruikt worden als onderdeel van het lees- en schrijfonderwijs, die het produceren van woordenboeken ondersteunen, vertalers behulpzaam zijn bij hun werk, en spraak omzetten in geschreven tekst en vice versa.

Al deze programma's maken op enig niveau gebruik van taalkundige regels. Voor al deze programma's geldt dat ze verbeterd zouden kunnen worden als ze maar over meer flexibel inzetbare taalkundige kennis zouden beschikken. Gosse Bouma beschrijft in zijn proefschrift een mogelijkheid om dat te verwezenlijken.

• NIET-EXACTE TAAL

Taalkundigen stellen regels op die de structuur beschrijven van een taal, zoals het Nederlands, of van een bepaald aspect van taal, zoals het verband tussen intonatie en betekenis. Computertaalkundigen proberen deze regels zó precies en eenvoudig te maken dat een computer ze kan hanteren. Bij het opstellen van zulke regels doen zich twee problemen voor. In de eerste plaats onderscheidt de taalkunde zich van de exacte vakken doordat taalkundige kennis zich niet laat reduceren tot een klein aantal abstracte wetten. Dit betekent dat zelfs een grammatica die maar een klein deel van het Nederlands be-

schrijft, al zal bestaan uit een flink aantal regels.

In de tweede plaats leidt het opstellen van algemene regels voor een zo grillig object als natuurlijke taal onvermijdelijk tot uitzonderingen. Neem bijvoorbeeld de zinsbouw van elliptische constructies als 'De jongens sliepen bij hun vader en de meisjes bij hun moeder.' Voor dergelijke constructies wordt vaak als regel gegeven dat het ontbrekende werkwoord in 'de meisjes bij hun moeder' identiek moet zijn met het werkwoord in het eerste deel van de zin. Dat deze regel lang niet altijd opgaat, blijkt als we het voorbeeld veranderen in 'De jongens sliepen bij hun vader en Maria bij haar moeder.'

• OVERKOEPELENDE REGELS

Om een zinvol software-taalprogramma te krijgen, moeten ook uitzonderingen op de regel hanteerbaar gemaakt worden voor de computer. Volgens Gosse Bouma kan dat — zonder de taal daarmee in een knellend keurslijf te persen — via de methode van het zogenoemde 'default-redeneren'. Stel dat er regels opgesteld moeten worden voor de vervoeging van Nederlandse werkwoorden.

•
**Een goed computer-
programma heeft niet
alleen weet van regels
maar ook van
uitzonderingen**
•

Een van de regels is dat het voltooid deelwoord van regelmatige werkwoorden eindigt op -t of -d. Nu is er een groep half-onregelmatige werkwoorden (zoals *lachen*) die volgens precies dezelfde regels vervoegd worden als de regelmatige werkwoorden, behalve dat het voltooid deelwoord (*gelachen*) op -en eindigt. We kunnen deze groep natuurlijk als een aparte klasse opvatten, maar een intuïtief veel aansprekender oplossing is te zeggen dat ze voldoen aan alle regels voor regelmatige werkwoorden met uitzondering van de regel voor voltooide deelwoorden.

Een computerprogramma dat niet alleen weet heeft van regels maar ook van uitzonderingen, maakt het mogelijk om precies deze formulering te kiezen. Een werkwoord als *werken* bijvoorbeeld wordt dan gedefinieerd als een regelmatig werkwoord, wat betekent dat alle regels voor regelmatige werkwoorden daarop van toepassing zijn. *Lachen* wordt gedefinieerd als een regelmatig werkwoord waarvan het voltooid deelwoord *gelachen* is. Het programma weet nu dat voor de vervoeging van het werkwoord *lachen* de regels voor regelmatige werkwoorden van toepassing zijn met uitzondering van de regel voor voltooide deelwoorden, omdat in de definitie van *lachen* al staat vermeld wat de vorm van het voltooid deelwoord is. Het programma maakt hier gebruik van een algemene 'default-regel': het toepassen van een algemene regel wordt geblokkeerd als er expliciete informatie beschikbaar is die met de regel in strijd is. Een bijkomend voordeel van deze aanpak is dat het aantal afzonderlijke regels dat nodig is om de vervoeging van werkwoorden te beschrijven, sterk wordt verminderd. <

Gosse Bouma - *Nonmonotonicity and
Categorial Unification Grammar*, Rijksuni-
versiteit Groningen, 1993