

Tellen met Taal

Het meten van variatie in zinsbouw in Nederlandse dialecten

Marco René Spruit

Taalkundige afstand

Iedereen weet dat de afstand tussen Amsterdam en Utrecht kleiner is dan de afstand tussen Amsterdam en Groningen. Zo voelt ook iedereen aan dat de taalkundige afstand tussen het Brabants en het Fries groter is dan de taalkundige afstand tussen het Brabants en het Limburgs. Voor geografische afstanden hebben we een meetkundige eenheid waarmee we precies kunnen aangeven hoe ver plaatsen uiteen liggen: de kilometer. Voor taalkundige afstanden wordt binnen het project "Determinanten van Dialect Variatie" (DDV) een meetkundige eenheid ontwikkeld. De mogelijkheid om de taalkundige afstand tussen de verschillende varianten van een taal te kunnen berekenen, is onder meer van belang voor onderzoek naar de relaties tussen dialecten en talen.

We richten ons in dit project op factoren die bepalend zijn voor de taalkundige afstand tussen dialecten. Ook allerlei sociale en politieke factoren kunnen de taalkundige afstand tussen dialecten beïnvloeden, maar dit project concentreert zich op de taalkundige factoren.

De taalkundige afstand wordt mede bepaald op basis van verschillen in uitspraak, woordkeus en zinsbouw. Tot voor kort werd veelal verondersteld dat er relatief weinig verschillen op het niveau van zinsbouw te vinden zijn binnen de Nederlandse dialecten. Het project "Syntactische Atlas van de Nederlandse Dialecten" (SAND) heeft echter aangetoond dat de honderden bestudeerde Nederlandse dialecten wel degelijk rijkelijk variëren op zinsbouwniveau. Sjef Barbiers heeft in *Respons 5* een beschrijving van dit project gegeven. Het huidige DDV-project maakt dankbaar gebruik van deze unieke verzameling gegevens, zowel om de onderlinge samenhang tussen de Nederlandstalige dialecten te onderzoeken, als om de rol van zinsbouw te bepalen als factor in de taalkundige afstand tussen taalvariëteiten. Op Afbeelding 1 kunt u zien welke dialecten onderzocht zijn.

De aard van de variatie in zinsbouw: twee voorbeelden

In de "Syntactische Atlas van de Nederlandse Dialecten" is op Kaart 68a de variatie in het gebruik van het wederkerend voornaamwoord "zich" in zinnen zoals "Jan herinnert zich dat verhaal wel" vastgelegd. In sommige dialecten blijkt men "Jan herinnert hem dat verhaal wel" te zeggen. In weer andere dialecten komt de vorm "Jan herinnert z'n eigen dat verhaal wel" voor. Kaart 68a toont de geografische distributie van de zinsvarianten voor deze taalsituatie.

Nog een voorbeeld: Kaart 14b van de atlas toont de variatie in het gebruik van het voegwoord in zinnen als "Het lijkt wel of er iemand in de tuin staat". In sommige dialecten zegt men "Het lijkt wel dat er iemand in de tuin staat", terwijl men in andere dialecten weer de vormen "Het lijkt wel of dat er iemand in de tuin staat" of "Het lijkt wel as er staat iemand in de tuin" gebruikt. De laatste voorbeeldzin toont bovendien aan dat de onderzochte variatie zich niet enkel op het niveau van

mogelijke functiewoorden afspeelt, maar tevens op het niveau van mogelijke woordvolgordes. Het werkwoord "staan" in deze bijzin bevindt zich immers op een andere positie in de zin dan bij de eerder genoemde varianten.

Het samenvoegen van gerelateerde taalverschijnselen

De "Syntactische Atlas van de Nederlandse Dialecten" bevat naast de twee beschreven taalverschijnselen nog ruim honderd andere taalsituaties waarvan de variatie op zinsbouwniveau is vastgelegd. Het vernieuwende van het DDV-onderzoek is nu dat voor het eerst gekeken wordt naar wetmatigheden in zinsbouwvariatie door meerdere taalverschijnselen als samengestelde eenheid te onderzoeken. Het idee hierachter is, dat zich door het samenvoegen van gerelateerde taalverschijnselen een algemeen variatiepatroon ontvouwt. Deze statistische benadering kijkt voorbij de details om de globale contouren te kunnen ontwaren, als het ware op zoek naar het bos voorbij de bomen.

Ook verkennen we de relaties tussen de verschillende taalsituaties. Een mogelijk verband tussen de eerder genoemde voorbeelden kan zijn dat mensen die "Jan herinnert z'n eigen dat verhaal wel" zeggen, ook meestal "Het lijkt wel as er staat iemand in de tuin" zeggen. Het blootleggen van dergelijke verbanden zou kunnen leiden tot nieuwe inzichten in de algemene werking van het taalsysteem.

Kwantitatieve taalkunde

Taalkundig onderzoek zoals dat binnen het DDV-project wordt gedaan, noemt men ook wel kwantitatieve taalkunde. Dit type onderzoek richt zich op grote hoeveelheden taalkundige gegevens om inzicht te krijgen in de achterliggende principes. Het is een mooie aanvulling op het traditionele kwalitatieve taalkundig onderzoek, waarin slechts één of enkele taalkundige verschijnselen tot in detail onderzocht worden. Waar kwalitatief onderzoek in de diepte gaat, concentreert kwantitatief onderzoek zich op de breedte - een ander perspectief dus.

De kern van kwantitatief onderzoek zit in de overgang van het meten van *afzonderlijke* taalkundige verschillen naar *samengevoegde* verschillen tussen taalvariëteiten. Hiervoor is het nodig om numerieke waarden toe te kennen aan taalkundige verschijnselen. Na de vertaalslag van tekst-verschillen naar getal-verschillen wordt het mogelijk om te tellen met taal.

Overigens is het ook zonder deze vertaalslag mogelijk om meerdere taalverschijnselen als samengestelde eenheid te bestuderen. Men geeft bijvoorbeeld op een dialectkaart de begrenzingslijnen aan van de gebieden waarbinnen de onderzochte taalverschijnselen voorkomen. Een probleem van deze methode is echter dat de begrenzingen voor ieder taalverschijnsel anders zijn. Om tot een algemener beeld te komen, kan de onderzoeker dan de overlappende begrenzingslijnen van de betreffende taalverschijnselen samenvoegen op basis van zijn kennis, ervaring en intuïtie. Het is echter goed mogelijk dat een andere onderzoeker de overlappende lijnen anders zal interpreteren.

Een andere tekortkoming is dat de betrouwbaarheid van de samengevoegde begrenzingslijnen afneemt naar gelang het aantal taalverschijnselen dat opgenomen is op een dialectkaart. Deze problemen doen zich echter niet voor bij een kwantitatieve benadering, aangezien afzonderlijke begrenzingslijnen dan - als

het ware - opgeteld kunnen worden tot objectief samengevoegde begrenzingslijnen.

De zinsbouwafstand tussen dialecten

Ter illustratie van de vertaalslag van tekst-verschillen naar getal-verschillen volgt nu een fragment uit de metingprocedure om de afstand op zinsbouwniveau te bepalen tussen de dialecten van Lunteren en Veldhoven. In Lunteren kan men zowel "Jan herinnert zich dat verhaal wel" zeggen als "Jan herinnert z'n eigen dat verhaal wel". In Veldhoven zegt men altijd "Jan herinnert zich dat verhaal wel". In geen van beide dialecten is het mogelijk om "Jan herinnert hem dat verhaal wel" te zeggen. Op basis van deze mogelijke zinnen wordt de afstand tussen deze twee dialecten in drie stappen als volgt berekend:

1. De zinsvariant met "zich" komt voor in beide dialecten, zodat dit geen invloed heeft op de dialectafstand.
2. De zin met "hem" komt in beide dialecten niet voor, zodat ook dit geen invloed heeft op de dialectafstand.
3. De variant met "z'n eigen" komt wel voor in Lunteren maar niet in Veldhoven, zodat de dialectafstand wordt verhoogd met één punt.

De zinsbouwafstand tussen de dialecten van Lunteren en Veldhoven is op basis van dit fragment dus één punt, gedeeld door het aantal mogelijke zinnen: $1/3$. De uiteindelijke zinsbouwafstand tussen deze twee dialecten ontstaat na vergelijking van 485 taalsituaties. Op dezelfde wijze wordt de zinsbouwafstand tussen alle andere dialectparen uitgerekend. Dit resulteert in een enorme tabel met tienduizenden afstanden. Aangezien het ondoenlijk is om een dergelijke hoeveelheid getallen te interpreteren, zijn statistische analyses onontbeerlijk om de resultaten te kunnen analyseren en in kaart te brengen.

Afbeelding 2 laat de resultaten van een zogenaamde clusteranalyse zien. Deze kaart van het Nederlandse taalgebied toont de zes belangrijkste regio's op basis van de zinsbouwvariatie. U kunt onder meer Vlaanderen duidelijk als één geheel waarnemen. Ook Friesland en Limburg zijn als aaneensluitend gebied goed herkenbaar.

Tot slot: de voordelen

Wellicht vraagt u zich nu af wat de meerwaarde is van dergelijk werk indien gelijksoortige overzichtskarten reeds eerder zijn opgesteld door ervaren dialectologen. Daarom eindigen we met enkele voordelen van deze manier van onderzoek. Vanuit wetenschappelijk oogpunt is het een belangrijk voordeel dat deze karten op objectieve wijze - en volledig automatisch - zijn opgesteld op basis van expliciet geformuleerde criteria. Het resultaat is daarmee controleerbaar en verifieerbaar. Ook is de procedure zonder extra inspanning toepasbaar op taalgebieden waarvan nog geen door experts opgestelde karten beschikbaar zijn. Ten slotte kan met dezelfde precisie iedere combinatie van taalkundige verschijnselen als een samengesteld geheel worden bestudeerd. Taalkundig onderzoek vanuit een kwantitatief perspectief verrijkt kortom het Nederlandse dialectonderzoek.

Afbeelding 1: De onderzochte dialecten in de syntactische atlas.



Afbeelding 2: De zes belangrijkste Nederlandstalige dialectgebieden.

