

Startpunt: het vakgebied taalkunde » taalvariatie » zinsbouwvariatie

Het vakgebied zinsbouwvariatie bestudeert onder meer de verschillende volgordes van de woorden in een zin. Ook de regelmatigheden in de opbouw van woorden worden geanalyseerd indien deze samenhangen met de omgeving van de zin. Belangrijk is dat de zinsvarianten dezelfde betekenis uitdrukken. De volgende vier zinnen uit de eerste Syntactische Atlas van de Nederlandse Dialecten (SAND) illustreren de rijke variatie op zinsbouwniveau binnen de Nederlandse dialecten:

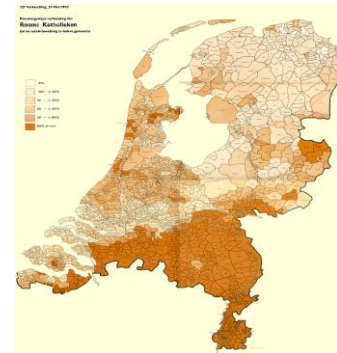
- (1) “Het lijkt wel of er iemand in de tuin staat.” (weergegeven als de donkerbruine locaties op Figuur 1)
- (2) “Het lijkt wel dat er iemand in de tuin staat.” (weergegeven als de groene locaties op Figuur 1)
- (3) “Het lijkt wel of dat er iemand in de tuin staat.” (weergegeven als de lichtbruine locaties op Figuur 1)
- (4) “Het lijkt wel of er staat iemand in de tuin.” (weergegeven als de donkerblauwe locaties op Figuur 1)



Figuur 1: Kaart 14b uit SAND1 toont zeven zinsbouwvariabelen waaronder “Het lijkt wel dat er iemand in de tuin staat.” (SPRUIT 2008:27).



Figuur 2: Een geaggregeerd overzicht van zinsbouwvariatie in de Nederlandse dialecten gebaseerd op computationele afstanden (SPRUIT ET AL., 2010).



Figuur 3: 12e Volkstelling, 31 Mei 1947 - Percentsgewijze verhouding der Rooms Katholieken tot de totale bevolking in iedere gemeente (HEEK, 1954:213)

Van probleem tot resultaat: het uniform identificeren van de determinanten van taalvariatie

Dit onderzoeksvoorstel stelt zich tot doel om een beter inzicht te verkrijgen in de determinanten van taalvariatie in het algemeen, en in zinsbouwvariatie binnen Nederlandse taalvariëteiten in het bijzonder.

Deze brede probleemstelling sluit op een vernieuwende wijze aan op een reeds rijke onderzoekstraditie. De onderzoeksaanpak omvat onder meer een combinatie van tegelijkertijd onderzochte niveaus van taalabstractie, een bewezen procescyclische aanpak, op het taaldomein toegespitste algoritmiek, en de inmiddels ruim voorradige hoeveelheid beschikbare onderzoeksgegevens.

Er worden twee types determinanten onderscheiden: *intralinguïstische* en *extralinguïstische*. Het LIKE onderzoeksplatform maakt het mogelijk om op *uniforme* wijze te bepalen welke taalkundige wetmatigheden ten grondslag liggen aan geografische distributiepatronen van zinsbouwvariatie, alsmede variatiepatronen tussen taalkundige niveaus, alsook distributiepatronen tussen taalkundige en niet-taalkundige aspecten zoals geloofsovertuiging, politieke overtuiging, en opleidingsniveau.

Om een voorbeeld te geven: SPRUIT (2008:130) bemerkt een mogelijke samenhang tussen de geografische scheidslijn op zinsbouwniveau, die grofweg ten noorden van Noord-Brabant loopt in Figuur 2, en de traditionele scheidslijn tussen de Rooms-katholieke en de Protestantse geloofsgemeenschappen in Figuur 3.

Het langetermijndoel van dit onderzoek is om een mogelijke rol te kunnen spelen bij het vormgeven van de toekomstige onderzoeksagenda binnen de Geesteswetenschappen. In dit verband is het nog nuttig op te merken dat via onder meer het DANS instituut recentelijk verscheidene gegevensbronnen in digitale vorm beschikbaar zijn gekomen, zoals die in Figuur 3, waardoor de LIKE onderzoeksopzet operationaliseerbaar is geworden.

**Aanpak: via duale onderzoeksbenadering naar geïntegreerde antwoorden**

---

Dit voorstel sluit aan op de innovatieve benadering zoals beschreven in (SPRUIT, 2009). Het integreert en verdiept van twee gerelateerde onderzoekslijnen. De eerste onderzoekslijn heeft de exploratieve *data mining* technieken in (SPRUIT, 2007) als startpunt, waarin onder meer de blootgelegde relatie tussen de volgende twee taalsituaties wordt beschreven:

- (A) “We geloven dat gieder niet zo slim zijn als wij.” (*voornaamwoord als onderwerp in tweede persoon enkelvoud*)  
(B) “Ze gelooft dat gij eerder thuis bent dan ik.” (*voornaamwoord als onderwerp in tweede persoon meervoud*)

Bovenstaande wetmatigheid is te lezen als: *INDIEN* men in een dialect (A) zegt, *DAN* zegt men ook (B). Een mogelijke taalkundige interpretatie hiervoor is: het meervoudige voornaamwoord “gieder” behoort tot hetzelfde paradigma als het enkelvoudige voornaamwoord “gij”.

De tweede onderzoekslijn bouwt voort op (SPRUIT *ET AL.* 2010) dat kwantitatieve methoden introduceert waarmee de relaties tussen de taalkundige niveaus mede bepaald kunnen worden, alsook de rol van geografie als een onderliggende factor die taalvariatieverschijnselen beïnvloedt. De LIKE procesbenadering omvat de benodigde methodologie om de onderzoeksresultaten op uniforme wijze te integreren.

**Samenwerking: partners in taal & technologie**

---

Dit onderzoek bevindt zich op het snijpunt van zinsbouwvariatie, data mining algoritmiek en IT methode-ontwikkeling, en wordt daarom uitgevoerd in nauwe samenwerking met de volgende drie onderzoeksgroepen: Variatielinguïstiek/Meertens Instituut (prof. Sjef Barbiers), Algoritmische Data-analyse/Universiteit Utrecht (prof. Arno Siebes) en Informatiekunde/Universiteit Utrecht (prof. Sjaak Brinkkemper).