

## **Antal van den Bosch, Robbert De Troij, Stefan Grondelaers & Dirk Speelman**

*Vissen naar alternanties: met ondertitels, automatische-vertaalsoftware, en de pivoting-aanpak op zoek naar nationale variatie in de Nederlandse grammatica*

Het is bekend dat Belgisch-Nederlands en Nederlands-Nederlands fonologische en lexicale verschillen vertonen (Geeraerts et al. 1999), maar op het terrein van de syntaxis wordt vaak beweerd dat nationale variatie minder belangrijk of zelfs quasi onbestaande is (Van Haver 1989: 41; Broekhuis et al. 2015: xiii). We zullen laten zien dat deze aanname geworteld is in ideologische vooringenomenheid (het geloof dat BD en ND verschillende oppervlakte-uitingen van dezelfde onderliggende motor zijn), en in het feit dat zowel leken als onderzoekers zich meestal niet bewust zijn van nationale verschillen in syntaxis die niet categorisch of sterk gemediatiseerd zijn. Gezien het feit dat interessante syntactische variabelen gemakkelijk onder de radar blijven (Nguyen en Eisenstein 2017) als we ons uitsluitend baseren op intuïties van onderzoekers, stellen we in dit praatje een computationele bottom-up benadering voor om ons visnet aanzienlijk breder en fijnmaziger te maken.

We baseren ons daarbij op de methode van Bannard & Callison-Burch (2005) en Van den Bosch (2014), en op een reusachtig parallel corpus van Belgisch- en Nederlands-Nederlandse vertalingen van Engelse ondertitels. We gebruiken automatische vertaalsoftware om plausibele “mappings” te vinden tussen een Engels n-gram en mogelijke Nederlandse vertalingen; op die manier oogsten we Nederlandse parafrases, d.w.z. stukken tekst van 2 tot 6 woorden lang die ongeveer dezelfde betekenis hebben.

In een vorige pilotstudie hebben we op basis van een niet-gestratificeerd ondertitelcorpus (lees: zonder subsets Belgisch- en Nederlands-Nederlands) onderzocht welke parafrases voor specifieke Engelse n-grammen syntactische variabelen vertegenwoordigden, in plaats van bijvoorbeeld idiomem. We vonden een aantal bekende alternanties (*er*-variatie, volgordevariatie, allerlei flectieverschijnselen), maar ook beloftevolle nieuwe patronen (zoals de systematische alternantie tussen verbale constructies (*als je ongesteld bent*) en nominale alternatieven (*tijdens de menstruatie*)). In dit praatje voegen we aan het onderzoek de cruciaalste variabele toe door op basis van afzonderlijke Belgisch- en Nederlands-Nederlandse corpora de nationaliteit van de parafrases in de analyse mee te nemen.

### **Referenties**

- Broekhuis, Hans, Norbert Corver en Riet Vos. 2015. *Syntaxis van het Nederlands: Werkwoorden en werkwoorden*, vol. 1. Amsterdam: Amsterdam University Press.
- Geeraerts, Dirk, Stefan Grondelaers en Dirk Speelman. 1999. *Convergentie en divergentie in de Nederlandse woordenschat: Een onderzoek naar kleding- en voetbaltermen*. Amsterdam: Meertens Instituut.
- Nguyen, Dong and Jacob Eisenstein. 2017. A kernel independence test for geographical language variation. *Computational Linguistics* 43(3). 567–592.
- Van Haver, Jozef. 1989. *Noorderman & Zuiderman: Het taalverdriet van Vlaanderen*. Tiel: Lannoo.