

Geheugengebaseerd Leren als datagedreven techniek om lexeemspecifieke variatie te onderzoeken

Variatie in argumentstructuur wordt vaak op een hoog niveau van abstractie bestudeerd, waarbij taalkundigen dezelfde formele alternantie bij verschillende werkwoorden en verschillende objecten tegelijkertijd onderzoeken. Pijpops en Speelman (2017) gaan bijvoorbeeld op die manier te werk bij hun studie van de alternantie tussen de Nederlandse transitieve constructie en de reflexieve constructie (zie ook Röthlisberger et al. 2017; Geleyn 2017). Die aanpak houdt echter het gevaar in dat betekenisverschillen of verschillen tussen regio's die specifiek zijn aan bepaalde lexemen onder de mat geveegd worden. Er gaan dan ook stemmen op om meer oog te hebben voor de lagere niveaus van abstractie, bijvoorbeeld door variatie apart te onderzoeken voor verschillende werkwoorden, objecten, etc. (Boas 2010; Lehmann and Schneider 2012; Perek 2014).

Zo'n benadering was lange tijd onmogelijk vanwege dataschaarste, maar dat probleem is niet meer onoverkomelijk nu steeds grotere corpora beschikbaar worden. Er stellen zich echter twee nieuwe moeilijkheden. Ten eerste zouden we erg veel verschillende alternanties moeten onderzoeken. Als we bijvoorbeeld de Nederlandse datiefalternantie voor elk werkwoord apart zouden willen bekijken, levert dat alleen al 252 aparte alternanties op (Colleman 2009: 597). Ten tweede worden we geconfronteerd met een gebrek aan hypotheses. Als een alternantie inderdaad aangestuurd wordt door een onderscheid dat verschilt van lexeem tot lexeem, wordt het erg moeilijk om op voorhand te bedenken wat dat onderscheid kan zijn. We hebben met andere woorden nood aan een techniek die (i) schaalbaar is, zodat verschillende alternanties vrij snel na elkaar onderzocht kunnen worden, en (ii) datagedreven is, d.w.z. gericht op hypothese generatie.

In deze presentatie willen we Geheugengebaseerd Leren (GGL) voorstellen als zo'n techniek (Daelemans en van den Bosch 2005). Als gevalstudie kijken we naar de alternantie tussen de Nederlandse transitieve constructie en de prepositioneel-intransitieve constructie met *naar*, zoals in (1)-(3).

- (1) *Technici zoeken nu (naar) de oorzaak van de rook.*
- (2) *Pas na enige tijd was hij genoeg gekomen van de emoties om (naar) de politie te telefoneren.*
- (3) *Ik kwam tot bezinning, hapte (naar) lucht, propte de brief in mijn zak, (...)*

De techniek bestaat uit drie stappen. In de eerste stap kiezen we de kenmerken die de GGL-analyses zullen gebruiken om de alternantie te voorspellen. Aangezien GGL geen model bouwt, mogen deze kenmerken erg uiteenlopend zijn, erg veel verschillende waardes hebben en zelfs collineair zijn. In de tweede stap voeren we een hele reeks GGL-analyses uit voor verschillende werkwoorden en zelfs voor unieke combinaties van een werkwoord en een object. Hiervan houden we enkel de analyses over die een hoge C-index scoren, wat aangeeft dat ze mogelijk een onderscheid op het spoor zijn dat nuttig is om de alternantie correct te voorspellen. Ten derde bekijken we de *gain ratios* van die analyses. Deze *gain ratios* geven voor elk kenmerk aan hoe nuttig het is om de alternantie te voorspellen. Op basis daarvan stellen we een hypothese op over een mogelijk onderscheid dat de alternantie stuurt. Tot slot worden die hypotheses getest op data die uit de GGL-analyses gehouden zijn.

Referenties

- Boas, Hans. 2010. The syntax-lexicon continuum in Construction Grammar. A case study of English communication verbs. *Belgian Journal of Linguistics* 24(1). 54–82.
- Colleman, Timothy. 2009. Verb disposition in argument structure alternations: a corpus study of the dative alternation in Dutch. *Language Sciences* 31(5). 593–611.
- Daelemans, Walter and Antal van den Bosch. 2005. *Memory-based language processing*. Cambridge: Cambridge University Press.
- Geleyn, Tim. 2017. Syntactic variation and diachrony The case of the Dutch dative alternation. *Corpus Linguistics and Linguistic Theory* 13(1). 65–96.
- Lehmann, Hans Martin and Gerold Schneider. 2012. Syntactic variation and lexical preference in the dative-shift alternation. In Joybrato Mukherjee & Magnus Huber (eds.), *Corpus Linguistics and Variation in English*, 65–75. Amsterdam: Rodopi.
- Perek, Florent. 2014. Rethinking constructional polysemy: The case of the English conative construction. In Dylan Glynn & Jus Robinson (eds.), *Polysemy and synonymy: Corpus methods and applications in cognitive linguistics*, 61–85. Amsterdam/Philadelphia: John Benjamins.
- Pijpops, Dirk and Dirk Speelman. 2017. Alternating argument constructions of Dutch psychological verbs. A theory-driven corpus investigation. *Folia Linguistica* 51(1). 207–251.
- Röthlisberger, Melanie, Jason Grafmiller and Benedikt Szendrői. 2017. Cognitive indigenization effects in the English dative alternation. *Cognitive Linguistics* 28(4). 673–710.