

Het bos van de Nederlandse zinsbouw? Ik dacht van ik weet de weg wel.

Liesbeth Augustinus, KU Leuven

De constructie van syntactisch geannoteerde corpora of *treebanks* biedt boeiende mogelijkheden voor de studie van de Nederlandse zinsbouw. Dergelijke corpora stellen onderzoekers in staat om empirisch onderzoek te voeren naar fenomenen die moeilijk te extraheren zijn uit platte corpora. De relevante onderdelen van een syntactische constructie staan immers niet noodzakelijk naast elkaar. Zo kunnen een determinator en de kern van een nominale constituent van elkaar gescheiden worden door een willekeurig aantal modificeerders (bv. *die grote rode kerstbal*). Het is zeer moeilijk om dergelijke constructies op een systematische manier uit een corpus te halen op basis van lexicale elementen of woordsoortannotatie alleen. Treebanks maken het mogelijk om abstractie te maken van het lexicale niveau om de gewenste constructies te vinden, in dit geval door te zoeken naar het label 'NP'.

Naast de voordelen die treebanks kunnen bieden, zijn er ook een aantal uitdagingen die het gebruik van dergelijke corpora met zich meebrengt. Zo vergt de syntactische annotatielaag kennis van specifieke zoekstrategieën om de corpora efficiënt te doorzoeken. Voor het Nederlands kan dat met behulp van de zoekmachine GrETEL (<http://gretel.ccl.kuleuven.be/>).

Aan de hand van constructies met het quotatieve *van*, zoals in (1-2), zullen we illustreren hoe treebanks kunnen bijdragen tot het onderzoek van de Nederlandse zinsbouw.

- (1) Ik had gisteren nog zoiets van dat lukt me wel.
- (2) Ik dacht van ik zal dat morgen wel doen.

In dergelijke constructies wordt *van* gecombineerd met een zin in de directe rede. Voor een beschrijving van het fenomeen, zie Hoekstra (2010), Coppen (2010) en Broekhuis & Corver (2015).

We vergelijken de idiomatische *had-zoiets-van*-constructie in (1) met constructies die meer variatie vertonen in de keuze van het selecterende werkwoord (*denken, vragen, weten, zeggen...*). Daarbij zoeken we een antwoord op de vraag waarom die werkwoorden gevolgd kunnen worden door het quotatieve *van* en andere semantisch verwante werkwoorden niet (bv. *menen, vermoeden, geloven...*). Tenslotte brengen we aan de hand van corpusdata in kaart hoe dergelijke zinnen zich verhouden tot constructies waarin *van* een beknopte bijzin selecteert, zoals in (3), en een gewone directe rede, zoals in (4).

- (3) Ik vroeg nog van even te wachten.
- (4) Ik dacht: "Ik zal dat morgen wel doen".

Bij iedere stap in dit onderzoek komen zowel de voordelen als de uitdagingen van het gebruik van treebanks aan bod.

Referenties

- Broekhuis, H. & N. Corver (2015). *Syntax of Dutch. Verbs and Verb Phrases*, vol. 2. Amsterdam University Press.
- Coppen, P.-A. (2010). Bericht van de innerlijke stem. Synchronie en diachronie van de *heb-zoiets-van*-constructie. *Nederlandse Taalkunde*, 15(1):33–53.
- Hoekstra, E. (2010). *Van* als markeerder van zinnen in de directe en indirecte rede in het Fries en het Nederlands. *Leuvense Bijdragen*, 96:169–188.