

Complexe constituenten worden vaak door een functiewoord ingeleid. Maar waarom?

Het complexiteitsprincipe van Rohdenburg (1996) stelt dat des te complexer een constituent is, des te vaker die vergezeld gaat van een expliciete functiemarkeerder, zoals een voorzetsel of een voegwoord. In het Nederlands zijn daarvan voorbeelden in overvloed. Zo verschijnt het voegwoord *om* vaker om complexe infinitiefzinnen in te leiden (Bouma 2017), duikt het voorzetsel *aan* vaker op voor complexe indirect objecten (Colleman 2006: 536–553), en bij het werkwoord *zoeken* zien we vaker het voorzetsel *naar* wanneer het object complex is. Zo is er bij (1) een hogere kans op *naar* dan bij (2).

- (1) *Ze zoekt naar een juridische oplossing zonder het onderdeel waarover het Vlaams parlement een belangenconflict heeft ingeroepen.* (Sonar-id: WS-U-E-A-0000331862.p.1.s.2, Oostdijk et al. 2013)
- (2) *De provincie Gelderland zoekt een oplossing.* (Sonar-id: WR-P-P-G-0000037003.p.1.s.4)

Tot zover zijn de meesten onder ons het eens. Waar echter wel sterke onenigheid over bestaat, is waarom die correlatie tussen complexiteit en explicietheid opgaat. Daarbij zijn de voorstellen in te delen zijn in drie groepen. Een eerste groep zoekt haar oorzaak in het cognitieve verwerkingsproces van de taalproducent (o.a. Ferreira and Dell 2000; MacDonald 2013), een tweede groep in het kanaal tussen taalproducent en -ontvanger (o.a. Fenk-Oczlon 2001; Jaeger 2010), en een derde groep in het cognitieve verwerkingsproces van de taalontvanger (o.a. Hawkins 2004; Kirby 1999: 31–62).

Volgens de eerste groep stelt het samenstellen van een complexe constituent, zoals “*een juridische oplossing zonder het onderdeel waarover het Vlaams parlement een belangenconflict heeft ingeroepen*”, hoge eisen aan het cognitieve taalverwerkingsvermogen van de producent. Door een functiemarkeerder in te voegen voor zo’n constituent, zou de producent tijd rekken om die samenstelling tot een goed einde te brengen, zonder daarvoor te moeten terugvallen op tussenwerpsels als *euh*.

De tweede groep wijt de correlatie aan ruis in het kanaal tussen producent en ontvanger, zoals achtergrondgeluid bij gesproken taal of koffievlekken bij geschreven taal. Door die ruis ontstaat het gevaar dat er informatie verloren gaat in dat kanaal. Des de compacter de informatie door het kanaal stroomt, des te meer informatie verloren dreigt te gaan bij de minste ruis. Dat betekent dat elk informatiekanaal dat blootstaat aan ruis, een eigen optimaal niveau van informatiedensiteit heeft. Boven dit niveau is het gevaar op substantieel informatieverlies te groot, terwijl onder het niveau de informatie-uitwisseling overdreven traag en dus inefficiënt gebeurt. Verder stelt men dat complexe constituenten vaak veel informatie bevatten. Om die piek in informatiedensiteit te reduceren, zou men er een functiemarkeerder voor plaatsen. Deze markeerder kondigt immers de constituent aan, zodat die voorspelbaarder wordt. Aangezien informatie gelijk is aan de negatieve logaritme van voorspelbaarheid, wordt de informatie van de constituent daarmee gereduceerd (Shannon 1948).

Ten slotte stelt de derde groep dat het interpreteren van een complexe constituent een zware taak is voor de taalontvanger. Om die taak te vereenvoudigen, verschijnt er vaker een functiemarkeerder voor zo’n constituent. Er blijft dan nog de vraag of de producent er telkens zelf voor kiest om zijn of haar ontvanger op die manier te helpen, dan wel of het complexiteitsprincipe het gevolg is van een meer indirecte druk op het taalsysteem.

We zullen de resultaten van ons onderzoek naar het werkwoord *zoeken* gebruiken om te arbitrereren tussen enerzijds de eerste en tweede groep verklaringen en anderzijds de derde. We maken met name een cruciaal onderscheid tussen de gevallen waarbij het object achter het werkwoord *zoeken* staat, en diegene waarbij het ervoor staat. Door de vorm van de correlatie in beide contexten na te gaan, terwijl we controleren voor een eventueel betekenisverschil en een aantal andere factoren, kunnen we de verklaringen uit elkaar trekken.

Referenties

- Bouma, Gosse. 2017. Om-omission. In Martijn Wieling, Martin Kroon, Gertjan van Noord & Gosse Bouma (eds.), *From Semantics to Dialectometry*, 65–73. Groningen: College Publications.
- Colleman, Timothy. 2006. De Nederlandse datiefalternantie. Een constructioneel en corpusgebaseerd onderzoek. Dissertation Ghent University.
- Fenk-Oczlon, Gertraud. 2001. Familiarity, information flow, and linguistic form. In Joan Bybee & Paul Hopper (eds.), *Frequency and the Emergence of Linguistic Structure [Typological Studies in Language 45]*, 431–448. Amsterdam: John Benjamins.
- Ferreira, Victor and Gary Dell. 2000. Effect of Ambiguity and Lexical Availability on Syntactic and Lexical Production. *Cognitive Psychology* 40(4). 296–340.
- Hawkins, John. 2004. *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Jaeger, Florian Tim. 2010. Redundancy and Reduction: Speakers Manage Syntactic Information Density. *Cognitive Psychology* 61(1). 23–62.
- Kirby, Simon. 1999. *Function, selection, and innateness: the emergence of language universals*. Oxford: Oxford University Press.
- MacDonald, Maryellen. 2013. How language production shapes language form and comprehension. *Frontiers in psychology* 4. 226.
- Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste and Ineke Schuurman. 2013. The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch. In Peter Spyns & Jan Odijk (eds.), *Essential Speech and Language Technology for Dutch, Theory and Applications of Natural Language Processing*, 219–247. Heidelberg: Springer.
- Rohdenburg, Günter. 1996. Cognitive Complexity and Increased Grammatical Explicitness in English. *Cognitive Linguistics* 7(2). 149–182.
- Shannon, C E. 1948. A Mathematical Theory of Communication. *Bell System Technical Journal* 27(3). Oxford, UK: Blackwell Publishing Ltd. 379–423.