

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

Dirk Pijpops, Dirk Spielman, Stefan Grondelaers, Freek Van de Velde

Fonds Wetenschappelijk Onderzoek Vlaanderen (FWO)

QLVL, University of Leuven

Radboud University of Nijmegen

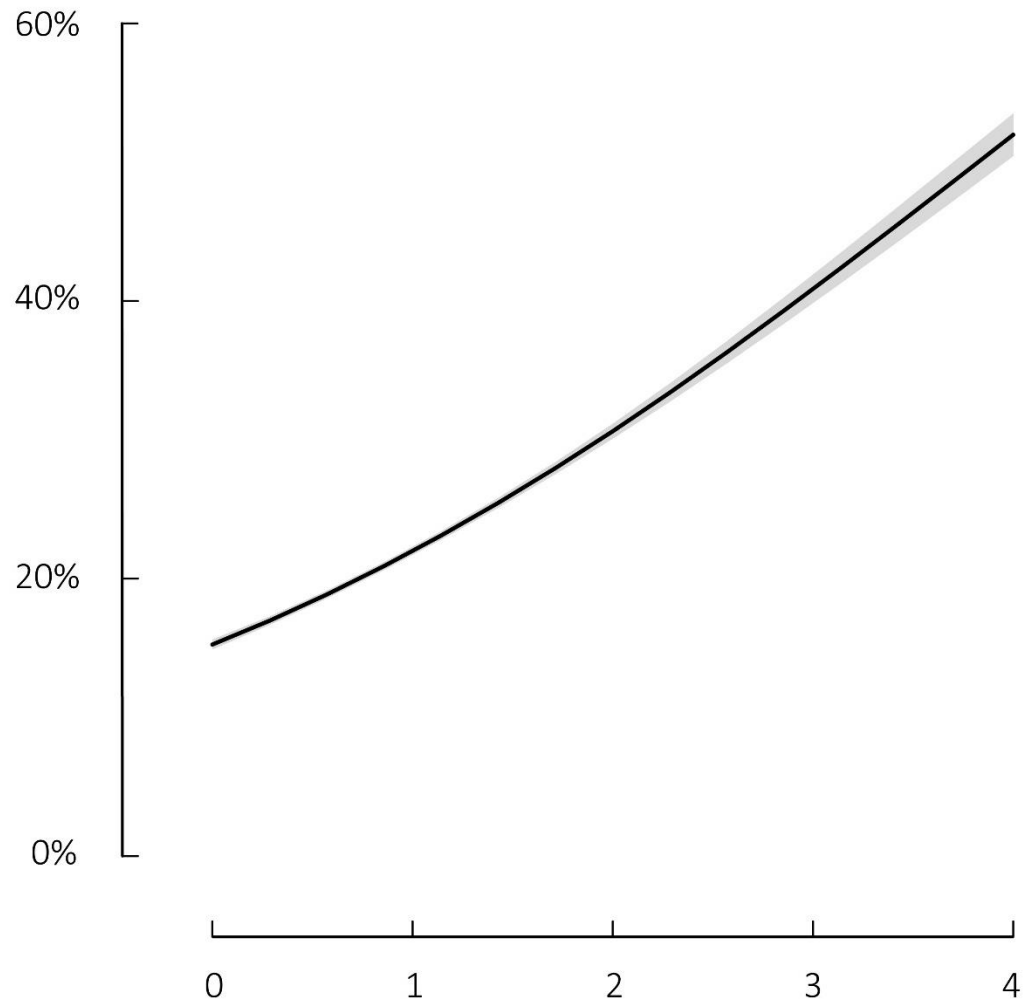
*De provincie zoekt **(naar)** een oplossing.*

*Ze zoekt **(naar)** een juridische oplossing zonder het onderdeel waarover het Vlaams parlement een belangenconflict heeft ingeroepen.*

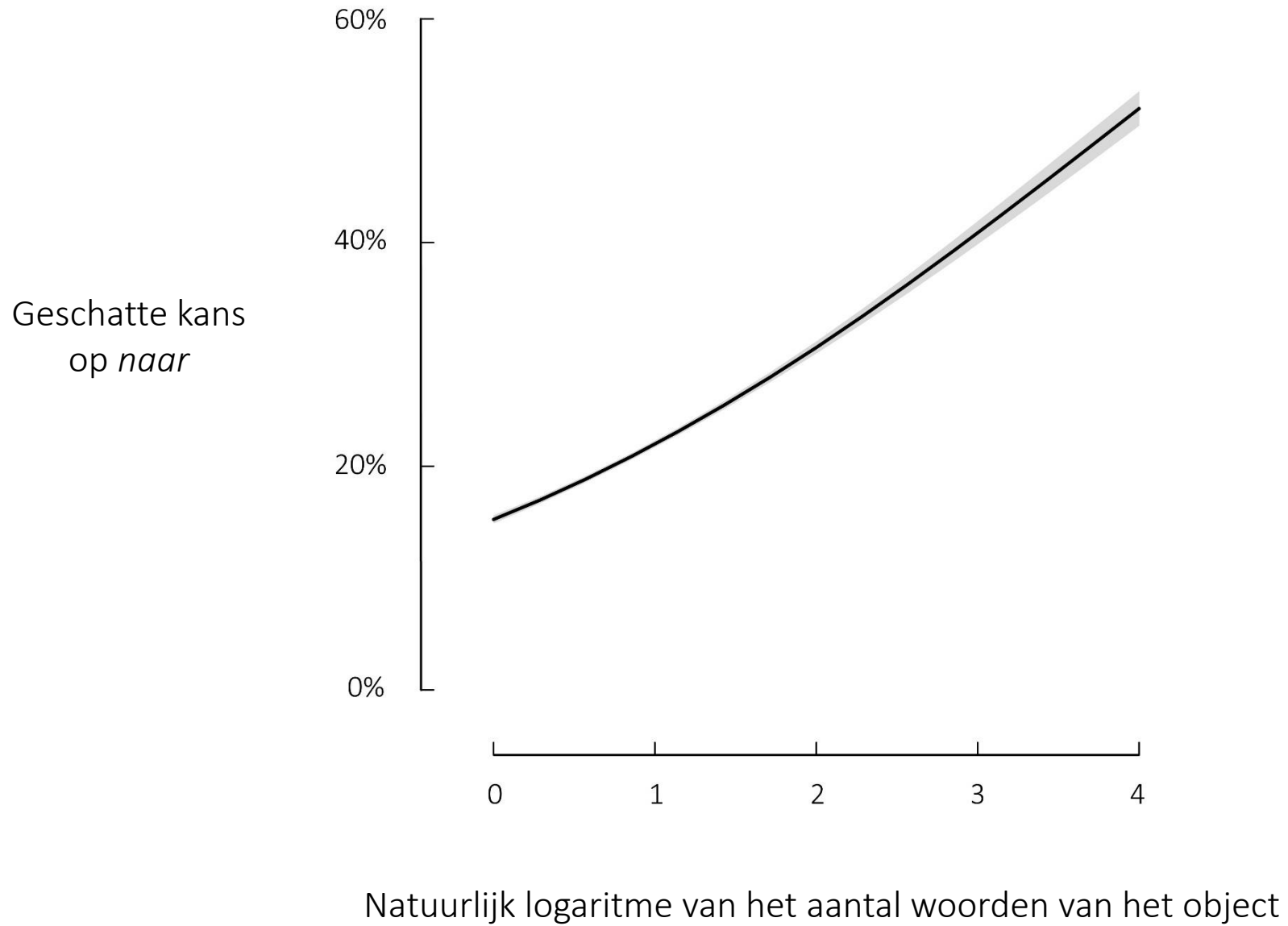
*De provincie zoekt een oplossing.*

*Ze zoekt **naar** een juridische oplossing zonder het onderdeel waarover het Vlaams parlement een belangenconflict heeft ingeroepen.*

Expliciete  
markering



Complexiteit van de constituent



*De Indiërs aarzelen te investeren.*

*De Indiërs aarzelen **om** nu plots een substantieel deel van hun vermogen in ons bedrijf te investeren.*

(Bouma 2017: 65)

Korte  
infinitieszinnen



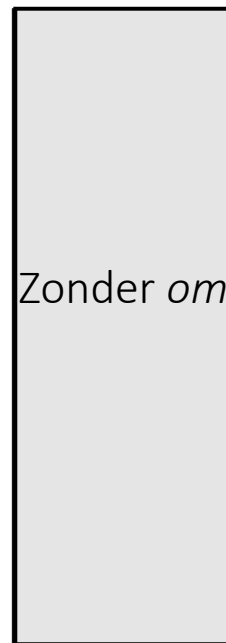
Met *om*

Middellange  
infinitieszinnen



Met *om*

Lange  
infinitieszinnen



Met *om*

(Data uit het TwNC, dank aan Robbert de Troij)

*De koning gaf de man een cadeau.*

*De koning gaf een cadeau **aan** de grote, met stuurse blik voor zich starende man die geflankeerd werd door zijn drie schattige, maar luidruchtige kinderen en zijn robuuste, van snoepjes, snacks en andere lekkernijen voorziene vrouw.*

(Lammertyn 2009: 20, Coleman 2006: 578)



*I think he is a powerful man*

*I think **that** my direct boss is a much more powerful man within our company than his job description makes him out to be.*

(Shank, Bogaert & Plevoets 2016)

Complexe constituenten worden vaak  
door een functiewoord ingeleid

Maar waarom?

Omdat sprekers  
dom zijn

Omdat er wel eens  
koffie gemorst wordt

Omdat hoorders  
dom zijn

Productie



Comprehensie



Kanaal

*Ze zoekt ... een juridische oplossing zonder het onderdeel waarover het Vlaams parlement een belangenconflict heeft ingeroepen.*

*Ze zoekt naar een juridische oplossing zonder het onderdeel waarover het Vlaams parlement een belangenconflict heeft ingeroepen.*

Concomitante signalen

Productie → Distributie → Cognitie → Comprehensie

Productie

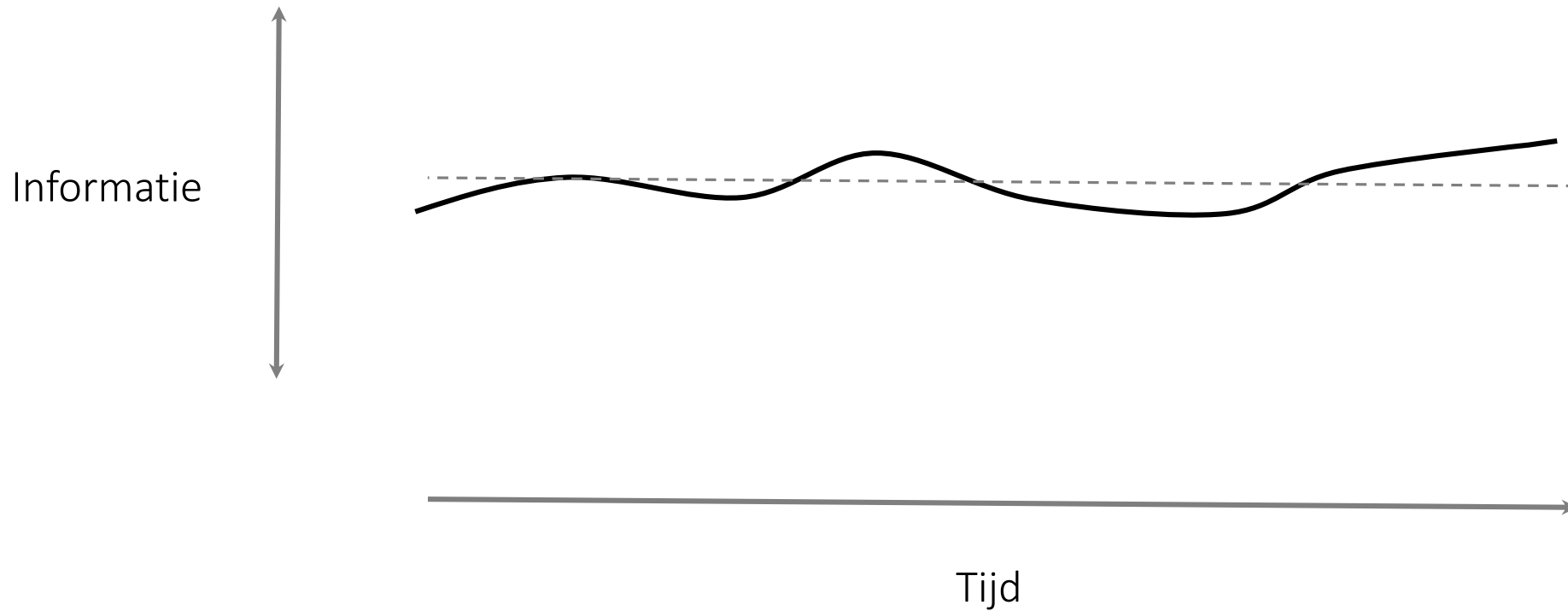
Lorem ipsum dolor sit amet,  
consectetur adipiscing elit. Nam  
vitae mi urna.

Comprehensie





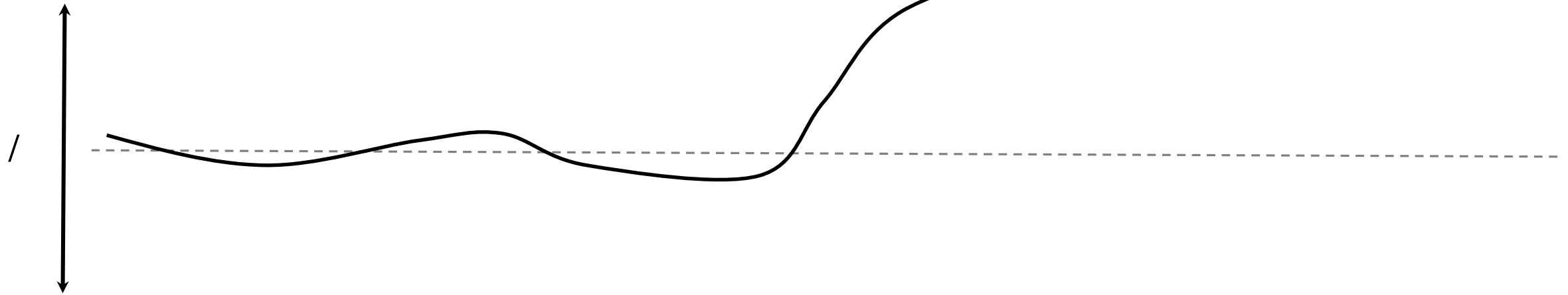
1. De kanalen van menselijke taal zijn informatiekanaalen
2. Elk informatiekanaal dat blootstaat aan ruis, heeft een optimaal niveau van informatiedensiteit



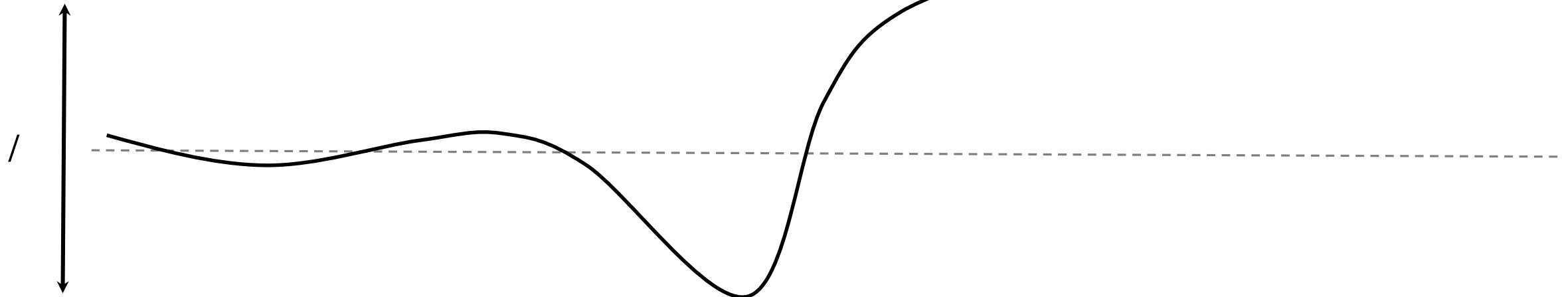
1. De kanalen van menselijke taal zijn informatiekanaalen
2. Elk informatiekanaal dat blootstaat aan ruis, heeft een optimaal niveau van informatiedensiteit
3. Informatie =  $-\log_2(\text{Voorspelbaarheid})$
4. Complexe constituenten zijn vaak specifiek, specifieke constituenten zijn vaak onvoorspelbaar

*In de praktijk zoekt de directie dan ... een alternatief*

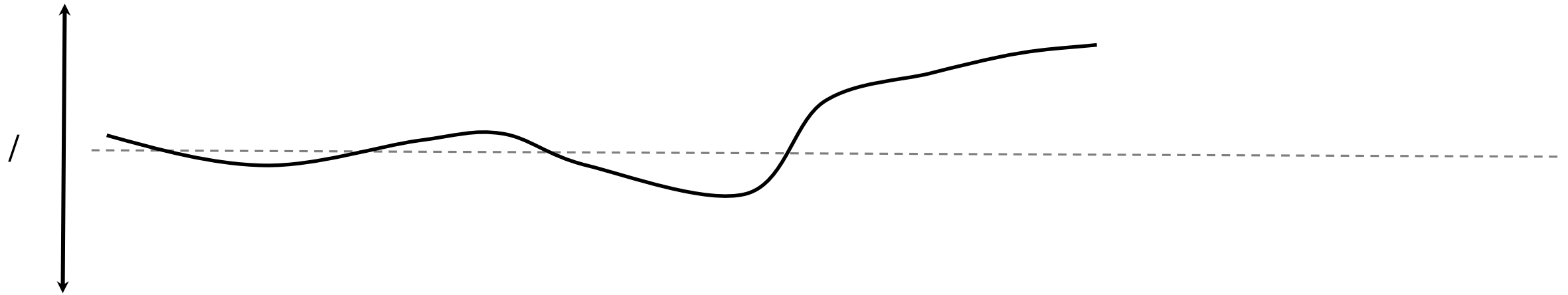
*De afgelopen anderhalf jaar zocht Ten Brink ... een 'groen alternatief' voor het, niet bepaald schoon verlopende, proces waarbij geurstoffen zoals mint, kaneel en citroen worden geproduceerd uit alcohol*



*De afgelopen anderhalf jaar zocht Ten Brink ... een 'groen alternatief' voor het , niet bepaald schoon verlopende , proces waarbij geurstoffen zoals mint , kaneel en citroen worden geproduceerd uit alcohol*



*De afgelopen anderhalf jaar zocht Ten Brink naar een 'groen alternatief' voor het , niet bepaald schoon verlopende , proces waarbij geurstoffen zoals mint , kaneel en citroen worden geproduceerd uit alcohol*



*De afgelopen anderhalf jaar zocht Ten Brink naar een 'groen alternatief' voor het , niet bepaald schoon verlopende , proces waarbij geurstoffen zoals mint , kaneel en citroen worden geproduceerd uit alcohol*

1. Menselijke taal is een informatiekanaal
  2. Elk informatiekanaal dat blootstaat aan ruis, heeft een optimaal niveau van informatiedensiteit
  3. Informatie =  $-\log_2(\text{Voorspelbaarheid})$
  4. Complexe constituenten zijn vaak specifiek, specifieke constituenten zijn vaak onvoorspelbaar
  5. Optionele functiewoorden bevatten nauwelijks/geen inherente informatie
  6. Door ze in te leiden voor een functiewoord, maak je complexe constituenten voorspelbaarder
- ⇒ Complexe constituenten worden vaak ingeleid door een functiewoord

Productie



Kanaal

Comprehensie



*Ze zoekt ... een juridische oplossing zonder het onderdeel waarover het Vlaams parlement een belangenconflict heeft ingeroepen.*



*Ze zoekt naar een juridische oplossing zonder het onderdeel waarover het Vlaams parlement een belangenconflict heeft ingeroepen.*

Altruïsme van de producent

Distributie → Comprehensie → Cognitie → Productie

Data

- Alle 79.410 voorkomens van *zoeken* met een object uit het Sonar-corpus
- Vaste uitdrukkingen uitgesloten: *zijn heil zoeken bij* (959), *ergens niets/niks te zoeken hebben* (728), *zijn toevlucht zoeken* (959)
- Object in het achterveld uitgesloten (6454)
- Onbekend land uitgesloten (4726)
- 55.232 zonder *naar*, 10.354 met *naar*
- 1000 voorkomens manueel gecontroleerd → 18
- Onderscheid: werkwoord voor object vs. object voor werkwoord

We **zoeken** (naar) de oorzaak, maar hebben nog geen idee.

(Naar) de oorzaken **zoeken** we al maanden

We zijn dus wel gedwongen nu al (naar) een goede vervanger te **zoeken**.

... als je (naar) een oplossing **zoekt** die perfect aansluit bij je bancaire behoeften.

Productie



*De Wereldraad van Kerken heeft dat niet gedaan, omdat hij van begin af aan*

*... een derde weg tussen het communistische oostblok en het vrije, kapitalistische westen*

*... zocht.*

- Des te langer het voorafgaande object
- Des te minder de producent zich al wil vastpinnen op een werkwoord
- Des te minder geneigd hij/zij zal zijn *naar* te gebruiken.

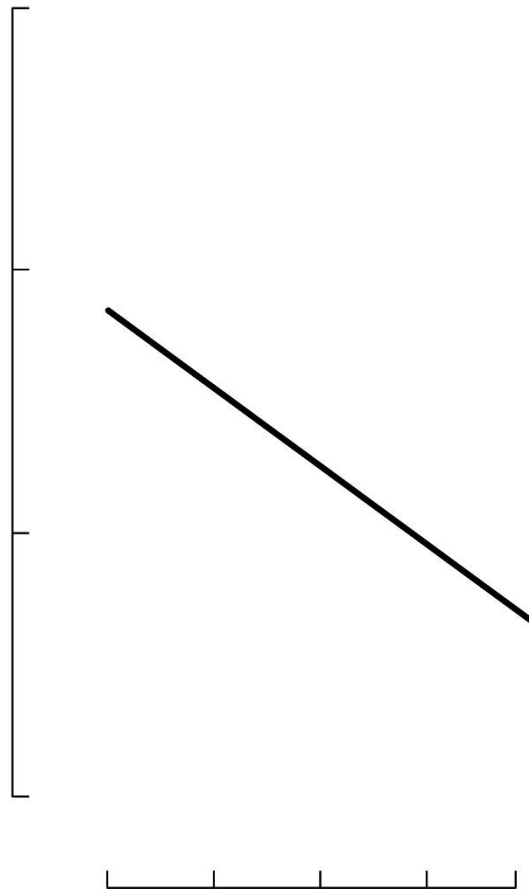


Object voor werkwoord

Werkwoord voor object

Kans op *naar*

Kans op *naar*



Complexiteit van het object

Complexiteit van het object



Kanaal



Als *naar* voor het werkwoord staat,

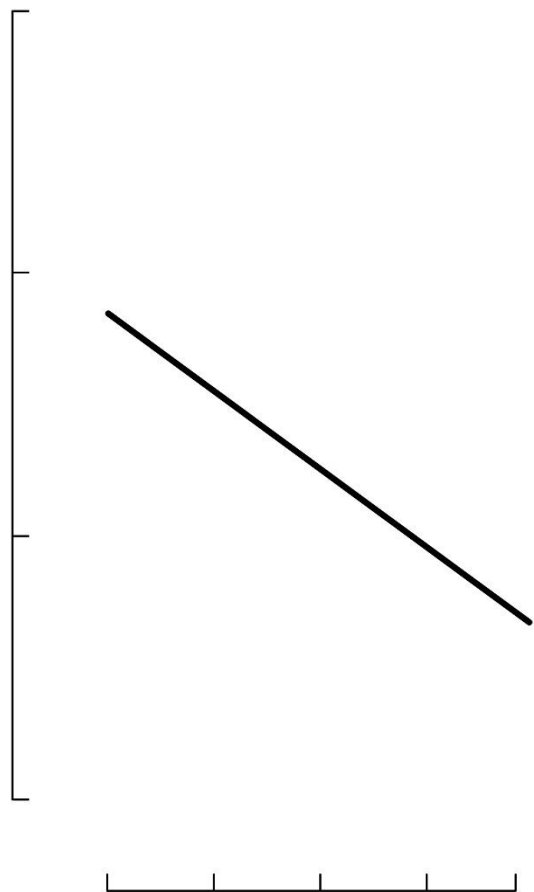
- geeft het al veel informatie over het te volgen werkwoord
- Is het al informationeel zwaar
- Wil je het niet combineren met informationeel zware constituent

Object voor werkwoord

Werkwoord voor object

Kans op *naar*

Kans op *naar*



Complexiteit van het object

Complexiteit van het object

$$\begin{aligned}
\Delta I_{\text{naar}} &= I(\text{naar} \mid \textit{preverbaal object}) - I(\text{naar} \mid \textit{postverbaal object}) \\
&= I(\text{zoeken} \mid \textit{preverbaal direct object}) - I(\text{zoeken} \mid \textit{preverbaal prepositioneel object}) \\
&\approx I(\text{zoeken} \mid \textit{preverbaal object lemma}) - I(\text{zoeken} \mid \textit{preverbaal object lemma \& naar}) \\
&= -\log_2 p(\text{zoeken} \mid \textit{preverbaal object lemma}) + \log_2 p(\text{zoeken} \mid \textit{preverbaal object lemma \& naar})
\end{aligned}$$

*Gemiddelde  $\Delta_{\text{naar}} \approx 3,7$  bits*

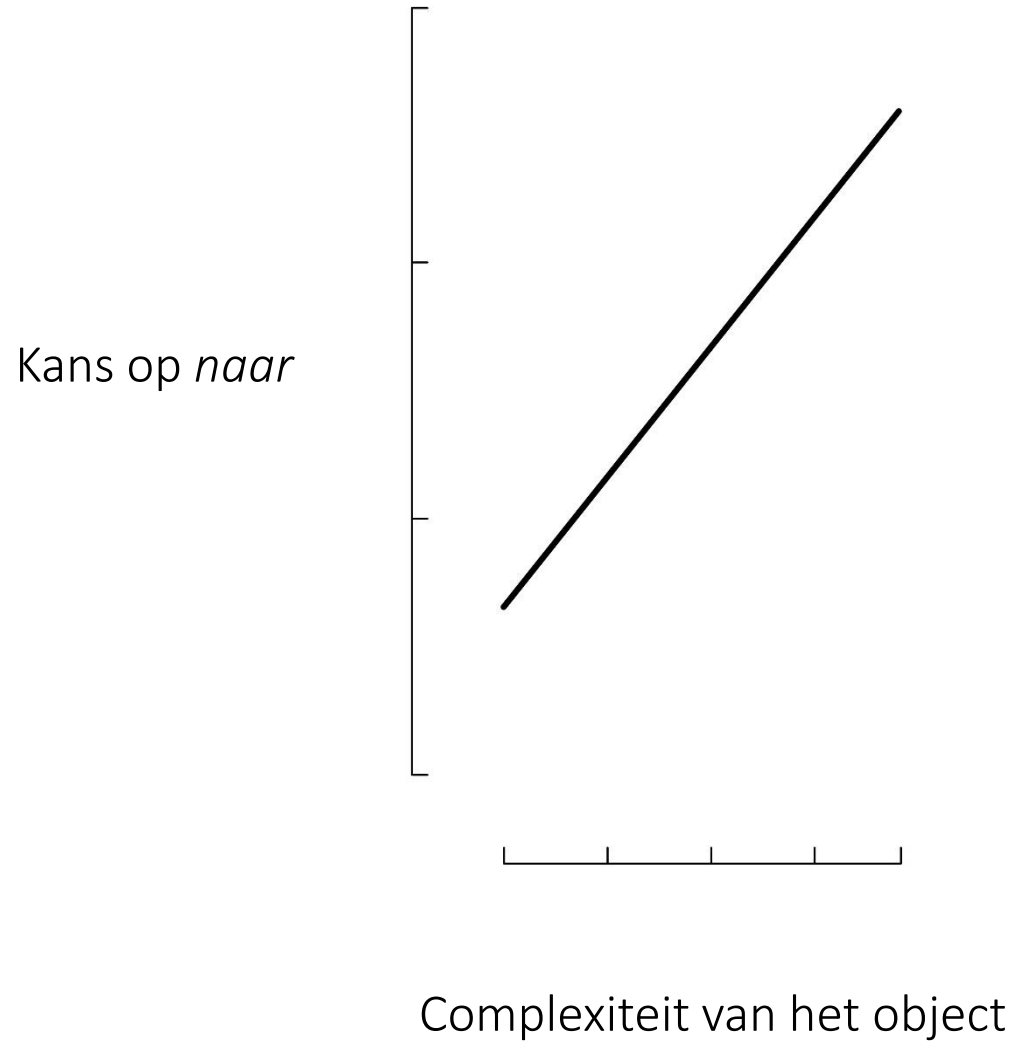
Comprehensie



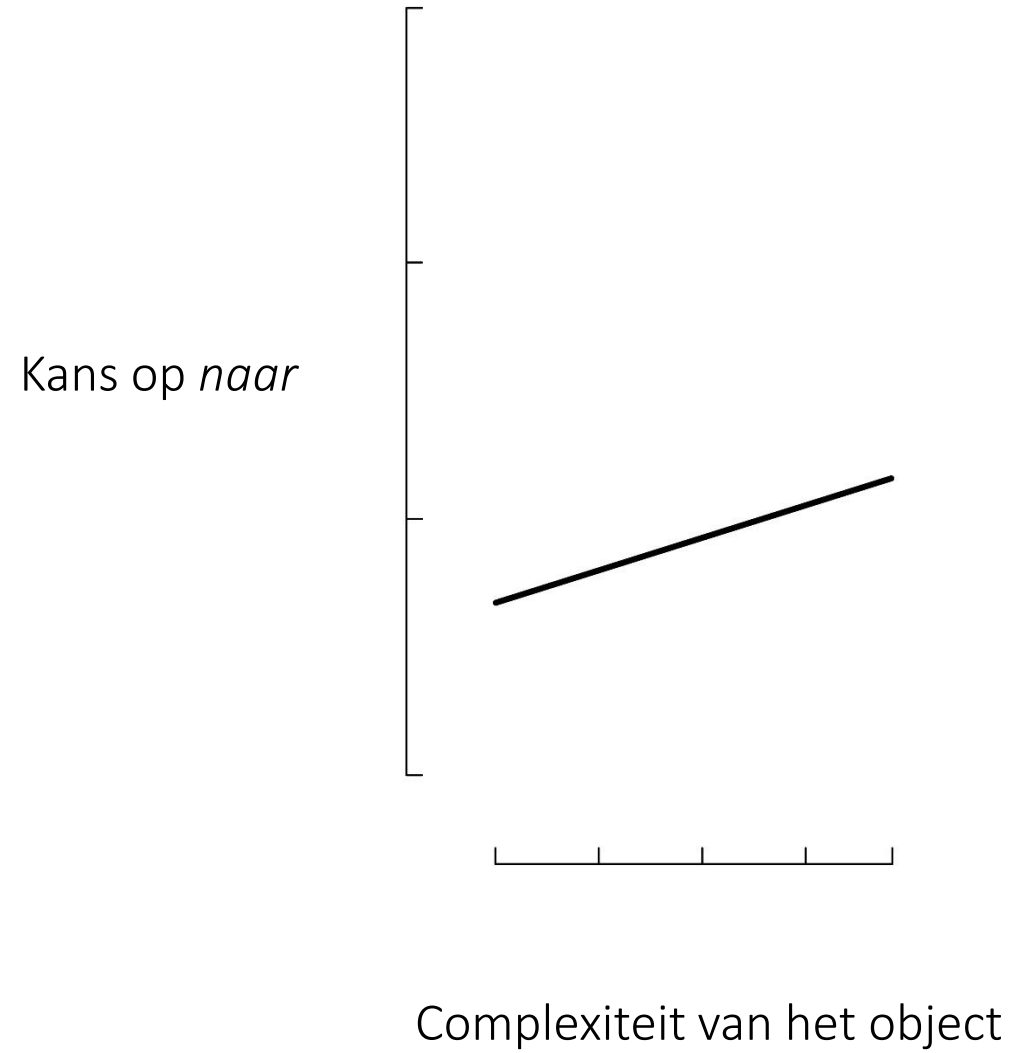
*De Wereldraad van Kerken heeft dat niet gedaan, omdat hij van begin af aan **naar** een derde weg tussen het communistische oostblok en het vrije, kapitalistische westen zocht.*

- Des te complexer het voorafgaande object
- Des te nuttiger het is alvast informatie over het werkwoord te krijgen
- Des te belangrijker het is om *naar* te horen/lezen

Object voor werkwoord



Werkwoord voor object





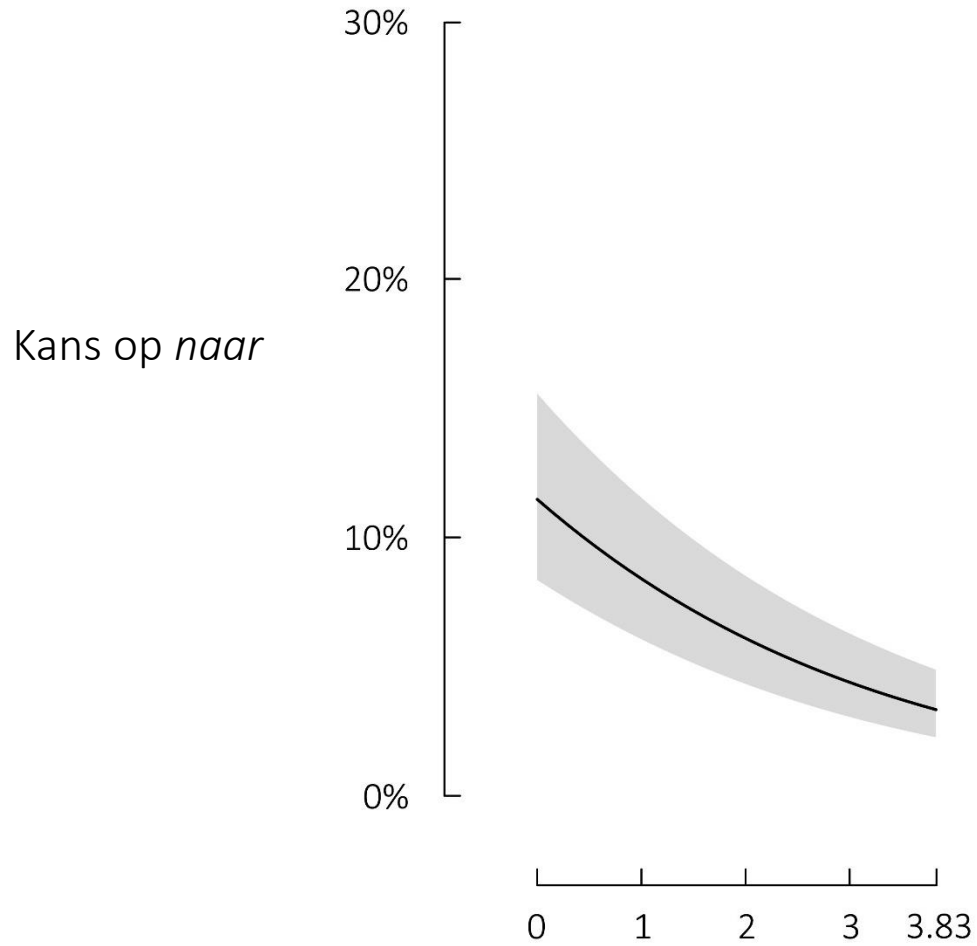
# Regressiemodel: respons met/zonder *naar*

- Fixed effects
  - Complexiteit van het object
  - Werkwoord-object volgorde
  - Interactie tussen Complexiteit van het object en Werkwoord-object volgorde
  - Zinstype: hoofdzin, bijzin
  - Werkwoordsvorm: finiet, infinitief, deelwoord
  - Land: België, Nederland
- Random effects
  - Corpus component: ondertitels, krantenmateriaal, auto cues,...
  - Semantische cluster

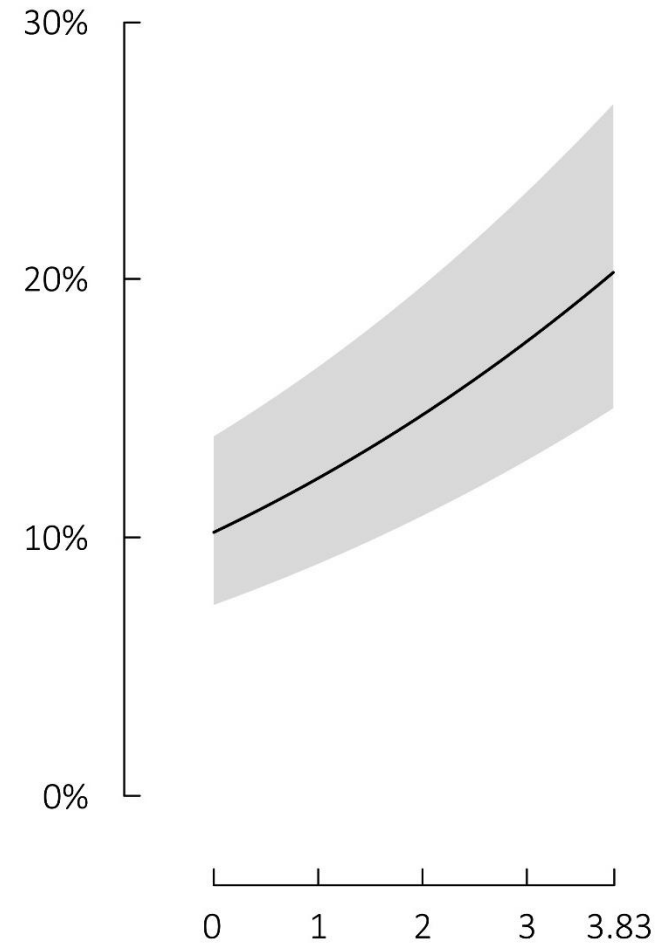
# Semantische cluster

- Groepering van objectlemma's op basis van hun betekenis
  - *Aanlegplaats, opvanglocatie, slaapplek,...* : één groep
  - *Formule, methode, techniek,...* : één groep
  - ...
- Levels
  - 50 groepen substantivische objectlemma's
  - Voornaamwoorden: elk apart
  - Unieke naam
  - Telwoord

### Object voor werkwoord



### Werkwoord voor object



AIC = 51.146  
C-index = 0,734  
Cond. nr.  $\kappa$ : 6,28

Natuurlijk logaritme van het aantal woorden van het object

Natuurlijk logaritme van het aantal woorden van het object

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

Productie

Kanaal

Comprehensie

Omdat sprekers  
dom zijn

Omdat er wel eens koffie  
gemorst wordt

Omdat hoorders  
dom zijn

Dank u!

Pijpops, Dirk, Dirk Speelman, Stefan Grondelaers & Freek Van de Velde. 2018.

**Comparing explanations for the Complexity Principle. Evidence from argument realization.**

*Language and Cognition* 10(3). 514–543.

dirk.pijpops@kuleuven.be

# Referenties

- Arnold, J., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: the role of constituent ordering. *Journal of Memory and Language*, 51(1), 55–70.
- Baayen, R. H. (2008). *Analyzing linguistic data: a practical introduction to statistics using R*. Cambridge University Press.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013). *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.4. Retrieved from http://cran.r-project.org/package=lme4
- Bloem, J. (2016). Evaluating automatically annotated treebanks for linguistic research. In P. Baski, M. Kupietz, H. Lingen, A. Witt, A. Barbaresi, H. Biber, ... S. Clemenatis (Eds.), *Proceedings of the 4th Workshop on Challenges in the Management of Large Corpora (CMLC-4)* (pp. 8–14). Mannheim: Institut für Deutsche Sprache.
- Bloem, J., Versloot, A., & Weerman, F. (2014). Applying automatically parsed corpora to the study of language variation. In *Proceedings of COLING 2014: the 25th International Conference on Computational Linguistics: technical papers: August 23–29, 2014, Dublin, Ireland* (pp. 1974–1984).
- Bloem, J., Versloot, A., & Weerman, F. (2017). Verbal cluster order and processing complexity. *Language Sciences*, 60, 94–119.
- Bock, K., Irwin, D., & Davidson, D. (2004). Putting First Things First. In *The Interface of Language, Vision, and Action: Eye Movements and the Visual World* (pp. 224–250). New York: Psychology Press.
- Bolinger, D. (1972). *That's that*. The Hague: Mouton.
- Bolinger, D. (1980). Wanna and the Gradience of Auxiliaries. In G. Brettschneider & C. Lehmann (Eds.), *Wege zur Universalienforschung: sprachwissenschaftliche Beiträge zum 60. Geburtstag von Hansjakob Seiler* (pp. 292–299). Tübingen: Gunter Narr.
- Bouma, G. (2017). Om-omission. In M. Wieling, M. Kroon, G. van Noord, & G. Bouma (Eds.), *From Semantics to Diacotometry* (pp. 65–73). Groningen: College Publications.
- Bouma, G., & Kloosterman, G. (2002). Querying Dependency Treebanks in XML. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC)* (pp. 1686–1691).
- Bouma, G., & Kloosterman, G. (2007). Mining Syntactically Annotated Corpora with XQuery. In *Proceedings of the Linguistics Annotation Workshop (ACL 07)* (pp. 17–24).
- Brennan, S., & Williams, M. (1995). The feeling of another's knowing: prosody and filled pauses as cues to listeners about the metacognitive states of speakers. *Journal of Memory and Language*, 34(3), 383–398.
- Bresnan, J., Cueni, A., Nikitina, T., & Baayen, R. H. (2007). Predicting the dative alternation. In G. Bouma, I. Krämer, & J. Zwarts (Eds.), *Cognitive Foundations of Interpretation* (pp. 69–94). Amsterdam: Royal Netherlands Academy of Science.
- Bybee, J. (2006). From Usage to Grammar: The Mind's Response to Repetition. *Language*, 82(4), 711–733.
- Clark, H. (1996). *Using language*. Cambridge: Cambridge University Press.
- Clark, H. (2004). Pragmatics of Language Performance. In L. R. Horn & G. Ward (Eds.), *The Handbook of Pragmatics* (pp. 344–364). Walden: Blackwell.
- Clark, H., & Fox Tree, J. (2002). Using uh and um in spontaneous speaking. *Cognition*, 84(1), 73–111.
- Clark, H., & Murphy, G. (1982). Audience Design in Meaning and Reference. *Advances in Psychology*, 9, 287–299.
- Coleman, Timothy. 2006. De Nederlandse dativealternatie. Een constructieel en corpusgebaseerd onderzoek. Dissertation Ghent University.
- Collard, P., Corley, M., MacGregor, L., & David, D. (2008). Attention Orienting Effects of Hesitations in Speech: Evidence from ERPs. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(3), 696–702.
- Corley, M., & Hartsuiker, R. (2003). Hesitation in speech can um help a listener understand. In R. Alterman & D. Kirsh (Eds.), *Proceedings of the 25th Annual Conference of the Cognitive Science Society* (pp. 276–281). Boston: Cognitive Science Society.
- Cover, T., & Thomas, J. (1991). *Elements of information theory*. Hoboken: Wiley-Interscience.
- Elness, J. (1984). That or zero? A look at the choice of object clause connective in a corpus of American English. *English Studies*, 65, 519–368.
- Fenk-Oczlon, & (2001). Familiarity, information flow, and linguistic form. In J. Bybee & P. Hopper (Eds.), *Frequency and the Emergence of Linguistic Structure (Typological Studies in Language 45)* (pp. 431–448). Amsterdam: John Benjamins.
- Fenk, A., & Fenk-Oczlon, G. (1993). Menzerath's law and the constant flow of linguistic information. In R. Köhler & B. Rieger (Eds.), *Contributions to quantitative linguistics* (pp. 11–31). Dordrecht: Kluwer Academic Publishers.
- Ferreira, V., & Dell, G. (2000). Effect of Ambiguity and Lexical Availability on Syntactic and Lexical Production. *Cognitive Psychology*, 40(4), 296–340.
- Ferreira, V., & Hudson, M. (2011). Saying “that” in Dialogue: The Influence of Accessibility and Social Factors on Syntactic Production. *Language and Cognitive Processes*, 26(10), 1736–1762.
- Ferreira, V., & Schotter, E. (2013). Do verb bias effects on sentence production reflect sensitivity to comprehension or production factors? *The Quarterly Journal of Experimental Psychology*, 66(8), 1548–1571.
- Fox, J. (2003). Effect displays in R for generalised linear models. *Journal of Statistical Software*, 8, 1–27.
- Fox Tree, J., & Clark, H. (1997). Pronouncing “the” as “thee” to signal problems in speaking. *Cognition*, 62(2), 151–167.
- Garnsey, S., Pearlmutter, N., Myers, E., & Lotocky, M. (1997). The Contributions of Verb Bias and Plausibility to the Comprehension of Temporarily Ambiguous Sentences. *Journal of Memory and Language*, 37(1), 58–93.
- Geenan, S., & Macdonald, M. (2009). Linking production and comprehension processes: The case of relative clauses. *Cognition*, 111(1), 1–23.
- Gleitman, L., January, D., Nappa, R., & Trueswell, J. (2007). On the Give and Take between Event Apprehension and Utterance Formulation. *Journal of Memory and Language*, 57(4), 544–569.
- Goldberg, A. E. (1995). *Constructions: a construction grammar approach to argument structure*. Chicago: University of Chicago press.
- Goldberg, A. E. (1999). The emergence of the semantics of argument structure constructions. In B. MacWhinney (Ed.), *Emergence of Language* (pp. 197–212). Hillsdale: Lawrence Erlbaum Associates.
- Gries, S. T. (2002). The influence of processing on grammatical variation: particle placement in English. In N. Dehé, R. Jackendoff, A. McIntyre, & S. Urban (Eds.),

- Verb-particle explorations* (pp. 169–288). Berlin/New York: Mouton de Gruyter.
- Gries, S. T. (2003). *Multifactorial analysis in corpus linguistics : a study of particle placement*. New York: Continuum.
- Gries, S. T. (2015). The most under-used statistical method in corpus linguistics: multi-level (and mixed-effects) models. *Corpora*, 10(1), 95–125.
- Grondelaers, S. (2000). *De distributie van niet-anaforsich er buiten de eerste zinsplaats: sociolinguïstische, functionele en psycholinguïstische aspecten van er's status als presentatief signaal*. Dissertation University of Leuven.
- Grondelaers, S., van den Bosch, A., Speelman, D., & van Hout, R. (2015). Comparing memory-based learning and regression approaches in the explanation of syntactic variation and change in Belgian and Netherlandic Dutch. In *New Ways of Analyzing Variation (NVAW 43)*. October 26, Chicago.
- Grondelaers, S., & Speelman, D. (2007). A variationist account of constituent ordering in presentative sentences in Belgian Dutch. *Corpus Linguistics and Linguistic Theory*, 3(2), 161–193.
- Grondelaers, S., Speelman, D., Drieghe, D., Brysbaert, M., & Geeraerts, D. (2009). Introducing a new entity into discourse: Comprehension and production evidence for the status of Dutch er “there” as a higher-level expectancy monitor. *Acta Psychologica*, 130(2), 153–160.
- Grondelaers, S., Speelman, D., & Geeraerts, D. (2003). De distributie van er in het gesproken Nederlands. In *Paper presented at the workshop spraakmakende spraak*. Nijmegen.
- Grondelaers, S., Speelman, D., & Geeraerts, D. (2008). National variation in the use of er “there”. Regional and diachronic constraints on cognitive explanations. In G. Kristiansen & R. Dirven (Eds.), *Cognitive Sociolinguistics. Language Variation, Cultural Models, Social Systems* (pp. 153–203). Berlin/New York: Mouton de Gruyter.
- Haeseryn, W., Romijn, K., Geerts, G., de Rooij, J., & van den Toorn, M. (1997). *Algemene Nederlandse Spraakkunst [General Dutch Grammar]*. Groningen: Nijhoff.
- Harrell, F. J., with contributions from Charles Dupont, & many others. (2017). Hmisc: Harrell Miscellaneous. R package version 4.0-3. Retrieved from https://cran.r-project.org/package=Hmisc
- Haspelmath, M. (2008). Frequency vs. Iconicity in Explaining Grammatical Asymmetries. *Cognitive Linguistics*, 19(1), 1–33.
- Hawkins, J. (2002). Symmetries and asymmetries. Their grammar, typology and parsing. *Theoretical Linguistics*, 28(2), 95–149.
- Hawkins, J. (2004). *Efficiency and complexity in grammars*. Oxford: Oxford University Press.
- Heller, B. (2018). *Stability and fluidity in syntactic variation world-wide. The genitive alternation across varieties of English*. Dissertation University of Leuven.
- Hopper, P., & Thompson, S. A. (1980). Transitivity in Grammar and Discourse. *Language*, 56(2), 251–299.
- Hoosier, D., & Lemeshow, S. (2000). *Applied logistic regression* (2nd edn.). New York: Wiley.
- Jaeger, F. T. (2005). Optional that indicates production difficulty: Evidence from disfluencies. In *Proceedings of DIS'05, Disfluency in Spontaneous Speech Workshop* (pp. 103–109). Aix-en-Provence.
- Jaeger, F. T. (2006). *Redundancy and Syntactic Reduction in Spontaneous Speech*. Dissertation Stanford University.
- Jaeger, F. T. (2010). Redundancy and Reduction: Speakers Manage Syntactic Information Density. *Cognitive Psychology*, 61(1), 23–62.
- Jaeger, F. T. (2011). Corpus-based research on language production: Information density and reducible subject relatives. In E. Bender & J. Arnold (Eds.), *Language from a cognitive perspective: Grammar, usage, and processing. Studies in honor of Tom Wasow* (pp. 161–197). Stanford: CSLI Publications.
- Jaeger, F. T. (2013). Production preferences cannot be understood without reference to communication. *Frontiers in Psychology*, 4.
- Jaeger, F. T., Levy, R., Wasow, T., & Orr, D. (2005). The Absence of that is Predictable if a Relative Clause is Predictable. In *Proceedings of Architectures and Mechanisms of Language Processing*. Ghent.
- Jaeger, F. T., & Wasow, T. (2005). Production-complexity driven variation: Relativizer omission in non-subjectextracted relative clauses. In *Paper presented at the 18th annual CUNY conference on sentence processing*.
- Kirby, S. (1999). *Function, selection, and innateness : the emergence of language universals*. Oxford: Oxford University Press.
- König, E., & Gast, V. (2009). *Understanding English-German contrasts* (2nd edn.). Berlin: Erich Schmidt.
- Konopka, A. E. (2012). Planning Ahead: How World Experience with Structures and Words Changes the Scope of Linguistic Planning. *Journal of Memory and Language*, 66(1), 143–162.
- Krajčić, T., & Brennan, S. (2005). Prosodic Disambiguation of Syntactic Structures: For the Speaker or for the Addressee? *Cognitive Psychology*, 50(2), 194–231.
- Lammertyn, Ine. 2009. Historisch corpusonderzoek. Het gebruik van de dubbelobject-constructie en de aan-constructie van de 17e eeuw tot nu. Master's Thesis Ghent University.
- Langacker, R. W. (1991). *Foundations of cognitive grammar: descriptive application*. Stanford: Stanford University Press.
- Levinson, S. (2000). *Presumptive meanings: the theory of generalized conversational implicature*. Cambridge: Cambridge : MIT press..
- Levshina, N., & Heylen, K. (2014). A radically data-driven Construction Grammar: Experiments with Dutch causative constructions. In R. Boogaart, T. Collemare, & G. Rutten (Eds.), *Extending the Scope of Construction Grammar* (pp. 17–46). Berlin: Mouton de Gruyter.
- Levy, R., & Jaeger, F. T. (2007). Speakers optimize information density through syntactic reduction. In B. Schölkopf, J. Platt, & T. Hoffman (Eds.), *Advances in Neural Information Processing Systems 19* (pp. 849–856). MIT Press.
- MacDonald, M. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 226.
- MacDonald, M., & Thornton, R. (2009). When language comprehension reflects production constraints: Resolving ambiguities with the help of past experience. *Memory & Cognition*, 37(8), 1177–1186.
- Menin, L., & Duffield, C. J. (2014). Looking for a ‘Gold Standard’ to measure language complexity: what psycholinguistics and neurolinguistics can (and cannot) offer to formal linguistics. In F. Newmeyer & L. Preston (Eds.), *Measuring Grammatical Complexity* (pp. 281–302). Oxford: Oxford University Press.
- Müller, S. (2006). Phrasal or lexical Constructions? *Language*, 82(4), 850–883.
- Müller, S., & Wechsler, S. (2014). Lexical approaches to argument structure. *Theoretical Linguistics*, 40(1–2), 1–76.
- Oostdijk, N., Reynaert, M., Hoste, V., & Schuurman, I. (2013a). *SoNaR User Documentation*.
- Oostdijk, N., Reynaert, M., Hoste, V., & Schuurman, I. (2013b). The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch. In P.

- Spyns & J. Odijk (Eds.), *Essential Speech and Language Technology for Dutch, Theory and Applications of Natural Language Processing* (pp. 219–247). Heidelberg: Springer.
- Peters, F. (2015). *Argument structure in usage-based construction grammar: experimental and corpus-based perspectives*. Amsterdam/Philadelphia: John Benjamins.
- Pijpops, D., & Van de Velde, F. (2016). Constructional contamination: How does it work and how do we measure it? *Folia Linguistica*, 50(2), 543–581.
- Pijpops, D., & Van de Velde, F. (2018). Lectal contamination. How language-external variation becomes language-internal through language contact. In *Variationist Linguistics meets Contact Linguistics*. 21 May, Ascona.
- R Core Team. (2014). *R: A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna. Retrieved from http://www.r-project.org/
- Rohdenburg, G. (1996). Cognitive Complexity and Increased Grammatical Explicitness in English. *Cognitive Linguistics*, 7(2), 149–182.
- Rohdenburg, G. (2016). Testing two processing principles with respect to the extraction of elements out of complement clauses in English. *English Language and Linguistics*, 20(3), 463–486.
- Roland, D., Elman, J., & Ferreira, V. (2006). Why is “that”? Structural Prediction and Ambiguity Resolution in a Very Large Corpus of English Sentences. *Cognition*, 98(3), 245–272.
- Shank, C., Pleveots, K., & Bogaert, J. Van. (2016). A multifactorial analysis of that/zero alternation: the diachronic development of the zero complementizer with think, guess and understand. In J. Yoon & S. T. Gries (Eds.), *Corpus-based Approaches to Construction Grammar* (pp. 201–240). Amsterdam/Philadelphia: John Benjamins.
- Shannon, C. E. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27(3), 379–423.
- Smith, V., & Clark, H. (1993). On the Course of Answering Questions. *Journal of Memory and Language*, 32(1), 25–38.
- Speelman, D. (2014). Logistic regression: A confirmatory technique for comparisons in corpus linguistics. In D. Glynn & J. A. Robinson (Eds.), *Corpus Methods for Semantics: Quantitative studies in polysemy and synonymy* (pp. 487–533). Amsterdam: John Benjamins.
- Speelman, D., Heylen, K., & Geeraerts, D. (2018a). Introduction. In D. Speelman, K. Heylen, & D. Geeraerts (Eds.), *Mixed-Effects Regression Models in Linguistics* (pp. 1–10). Cham: Springer.
- Speelman, D., Heylen, K., & Geeraerts, D. (2018b). *Mixed-Effects Regression Models in Linguistics*. Cham: Springer.
- Speelman, D., Heylen, K., & Grondelaers, S. (2014). A bottom-up, data-driven operationalization of semantic classes and predictability in syntactic alternation research. In S. Grondelaers & R. van Hout (Eds.), *New ways of analyzing syntactic variation*. Berlin: Mouton de Gruyter.
- Tanner, D., & Bulkes, N. (2015). Cues, quantification, and agreement in language comprehension. *Psychonomic Bulletin & Review*, 22(6), 1753–1763.
- Tanner, D., Nicol, J., & Brahm, L. (2014). The time-course of feature interference in agreement comprehension: Multiple mechanisms and asymmetrical attraction. *Journal of Memory and Language*, 76, 195–215.
- Theijssen, D., Boves, L., Halteren, H., & Oostdijk, N. (2010). Evaluating automatic annotation: automatically detecting and enriching instances of the dative alternation. *Language Resources and Evaluation*, 46(4), 565–600.
- Thompson, S. A., & Mulac, A. (1991). The discourse conditions for the use of the complementizer that in conversational English. *Journal of Pragmatics*, 15(3), 237–251.
- Turney, P., & Pantel, P. (2010). From frequency to meaning: Vector space models of semantics. *Journal of Artificial Intelligence Research*, 37, 141–188.
- van den Bosch, A., Grondelaers, S., & Speelman, D. (n.d.). A Memory Based Account Of Constructional Differences Between Netherlandic And Belgium Dutch. In S. Grondelaers & R. van Hout (Eds.), *New Ways of Analyzing Syntactic Variation*. Berlin: Mouton de Gruyter.
- van Noord, G. & Speelman, D. (2010). At Last Parsing Is Now Operational. *TALN*, 20–42.
- van Noord, G., Bouma, G., Van Eynde, F., De Kok, D., Van der Linde, J., Schuurman, I., ... Vandeghinste, V. (2013). Large Scale Syntactic Annotation of Written Dutch: Lassy. In P. Spyns & J. Odijk (Eds.), *Essential Speech and Language Technology for Dutch* (pp. 219–247). Berlin: Springer.
- Willems, A., & De Sutter, G. (2015). Reassessing the effect of the complexity principle on PP Placement in Dutch. *Nederlandse Taalkunde*, 20(3), 339–366.
- Wolk, C., Bresnan, J., Rosenbach, A., & Szrmeccsanyi, B. (2013). Dative and genitive variability in Late Modern English: Exploring cross-constructual variation and change. *Diachronica*, 30(3), 382–419.
- Zifonun, G., Hoffmann, L., & Strecker, B. (1997). *Grammatik der deutschen Sprache [Grammar of the German language]*. Berlin: de Gruyter.
- Zipf, G. K. (1949). *Human behavior and the principle of least effort : an introduction to human ecology*. New York: Hafner.
- Zwart, J.-W. (2011). *The syntax of Dutch*. Cambridge: Cambridge University Press.