

Encoding vs. decoding. Why do language users make sentence structure explicit?

Language is shaped by processing pressures from production, or encoding, and reception, or decoding (Hawkins 2004). Evidence of psycholinguistic experiments indicate that when both pressures counteract one another, the latter generally takes precedence (see a.o. Ferreira and Dell (2000) and references cited therein). In this study, we aim to complement this work with corpus research. The employed case study concerns the Dutch verb *zoeken* ‘to search’, where language producers have the choice whether or not to explicitly mark the object using the preposition *naar* ‘to’, as in (1)-(2).

- (1) *We zoeken alternatieven.* (Sonar corpus, Oostdijk et al. 2013, WR-P-P-G-0000254655.p.11.s.5)
‘We are looking for alternatives.’
- (2) *Wij zoeken dan wel naar alternatieven.* (Sonar corpus, Oostdijk et al. 2013, WR-P-P-G-0000488037.p.6.s.3)
‘We, then, look for alternatives.’

Using data from the Sonar corpus, we find that the likelihood of *naar* increases as the object becomes more complex (Figure 1). There are at least three possible ways to explain this relation, however. The first is that the strictly unnecessary preposition helps the addressee decode the sentence, and expressing the preposition is therefore especially called for when the object is complex (cf. Rohdenburg's (1996) Complexity Principle). The second is that *naar* functions as a way to buy time for the producer to formulate a complex object. Finally, the third states that *naar* is preferred with more complex objects because it allows the producer to extrapose such objects to postfield position. This study will attempt to disentangle these three possible explanations.

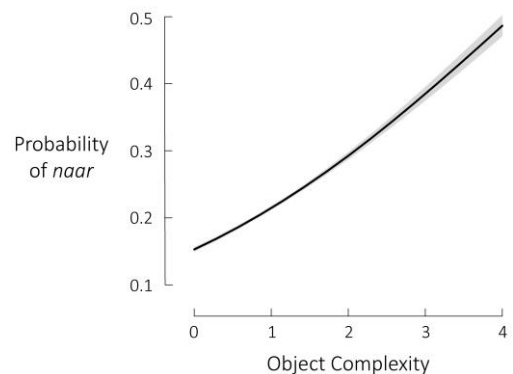


Figure 1: As the object becomes more complex, *naar* is more likely to be expressed.¹

References

- 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.
- Oostdijk, Nelleke, Martin Reynaert, Véronique Hoste and Ineke Schuurman. 2013. The Construction of a 500-Million-Word Reference Corpus of Contemporary Written Dutch. *Theory and Applications of Natural Language Processing*. 219–247.
- Rohdenburg, Günter. 1996. Cognitive Complexity and Increased Grammatical Explicitness in English. *Cognitive Linguistics* 7(2). 149–182.

¹ Effect plot of *Object Complexity* in a logistic regression model controlling for country and corpus component ($p < 0.0001$, coefficient = 0.41). *Object Complexity* was measured as the natural logarithm of the number of words of the object. The grey area represent confidence intervals.