

# Quantitative models of referential choice: Lexical anaphora in English

Nils Norman Schiborr (University of Bamberg, Germany)

A long line of research in the tradition of Chafe (1976) has examined speakers' selection of referring expressions in natural discourse, generally framing this choice in terms of information packaging (Prince 1981, Givón 1983, Gundel et al. 1993, Gordon & Hendrick 1997), and espousing some notion of “activation states” (Chafe 1976, 1994) or “accessibility” (Ariel 1990). According to these theories, the choice of particular expressions is driven largely by the properties of the preceding discourse, and is sensitive to certain morphosyntactic, semantic, and prosodic features. Two of the limitations of many of these approaches, however, is that they base their claims on small data sets of chiefly written language, and often offer insufficient insights into their methodologies.

This paper presents ongoing research into the factors underpinning the choice between full, lexical noun phrases (e.g. *the woman, Jane*) and non-lexical expressions (e.g. *she*, zero anaphora) for references to given discourse entities in natural spoken discourse. It explores these dimensions using a nuanced, bottom-up approach based on spoken, spontaneous corpus data from the Multi-CAST collection (Schiborr 2015). These data have been annotated for the form and grammatical role of referring expressions, their information status, as well as their referential identity (see Haig & Schnell 2014; Schiborr et al. 2018).

Preliminary findings suggest that discourse-based factors such as recency and prominence serve as strong indicators for referential choice, but to different extents for different grammatical roles, human and non-human referents, and specific semantic classes of predicates. In addition to contemporary English, this paper offers a brief outlook on the larger typological perspective, which suggests that the aforementioned tendencies are likely to be crosslinguistically stable.

## References

- Ariel, Mira. 1990. *Accessing noun-phrase antecedents*. London: Routledge.
- Chafe, Wallace. 1976. Givenness, contrastiveness, definiteness, subjects, topics, and point of view. In Charles N. Li (ed.), *Subject and topic*, 25–55. New York: Academic Press.
- Chafe, Wallace. 1994. *Discourse, consciousness, and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: The University of Chicago Press.
- Givón, Talmy (ed.). 1983. *Topic continuity in discourse* (Typological Studies in Language 3). Amsterdam: John Benjamins.
- Gordon, Peter C. & Randall Hendrick. 1997. Intuitive knowledge of linguistic co-reference. *Cognition* 62(2). 325–370.
- Gundel, Jeanette K., Nancy Hedberg & Ron Zacharski. 1993. Cognitive status and the form of referring expressions in discourse. *Language* 69(2). 274–307.
- Haig, Geoffrey & Stefan Schnell 2014. *Annotations using GRAID (Grammatical relations and animacy in discourse): Introduction and guidelines for annotators*, version 7.0. [multicast.aspra.uni-bamberg.de#annotations](http://multicast.aspra.uni-bamberg.de#annotations). (2 April, 2019.)
- Prince, Ellen F. 1981. Toward a taxonomy of given-new information. In Peter Cole (ed.), *Radical pragmatics*, 223–255. New York: Academic Press.
- Schiborr, Nils N. 2015. Multi-CAST English. In Geoffrey Haig & Stefan Schnell (eds.), *Multi-CAST: Multilingual corpus of annotated spoken texts*. [multicast.aspra.uni-bamberg.de#english](http://multicast.aspra.uni-bamberg.de#english). (2 April, 2019.)
- Schiborr, Nils N., Stefan Schnell & Hanna Thiele. 2018. *RefIND — Referent indexing in natural-language discourse: Annotation guidelines*, version 1.1. [multicast.aspra.uni-bamberg.de#annotations](http://multicast.aspra.uni-bamberg.de#annotations). (2 April, 2019.)