

Variation in complementizer choice between *if* and *whether*: corpus evidence for real-world decisions

This study uses corpus data to explain choices in everyday language use to examine the variation between *if* and *whether* as complementizers in interrogative subordinate clauses. In closed interrogative clauses, speakers may choose between *if* and *whether*, e.g., *I'm not sure if/whether that happened*. In contrast to the abundance of research on the variation between retention and omission of the complementizer *that* (see, e.g., the overview in Szmrecsanyi & Kolbe-Hanna, to appear), only few studies deal with the choice between *if* and *whether* in this related syntactic context. Zieglschmid (1929) and Steinbach (1929) describe prescriptive and futile efforts to promote the exclusive use of *if* as conditional subordinator and of *whether* for interrogatives. Biber et al. (1999, 691-693) report that *if* is more frequent than *whether* in conversation, but controlled by a smaller range of matrix verbs. Kolbe (2008, 129–136) finds that in British dialects *if* is also slightly more frequent overall. Both studies detect a preference for *whether* after the matrix verb *know*. In the British dialect data, *whether* is more likely to be used in the North, by older speakers and by women.

To deepen the understanding of the variation between *if* and *whether*, this study relies on data from ICE-GB, ICE-Ireland and ICE-New Zealand to analyse the distribution of both complementizers in different text types and in different varieties of English. It employs different statistical analyses (logistic regression, conditional inferences trees, random forest) to pinpoint the most important predictors of *if* and *whether*, whilst taking account of idiosyncratic features such as speakers' individual preferences (see Tagliamonte and Baayen 2012 for an overview and comparison of these and other statistical tools). Cognitive variables that have been proven to influence the choice between the retention of the complementizer *that* and its omission (e.g. morphosyntactic features in matrix and embedded clauses, length of the embedded clauses) are included as potential predictors. This study therefore explores the impact of cognitive factors when the choice is not between retention and omission of a complementizer, but between a more explicit (*whether*) and a less explicit (*if*) option (cf. Rohdenburg 1996 for the notion of explicitness).

The analysis reveals that the strongest overall predictor of the variation between *if* and *whether* is the matrix verb controlling the interrogative clause. There is a strong preference for *if* after *ask*, *see* and *wonder*, which indicates that speakers store and retrieve the corresponding sequences in their mental lexicon. When the embedded clause is longer than seven words and thus more complex in production as well as in processing, *whether* is more frequent. In ICE-Ireland, *if* is more frequent than elsewhere.

By combining register, regional variety and cognitive factors as predictors of the choice between *if* and *whether*, this study sheds light on the weighting of external and internal factors in making linguistic choices and on the importance of cognitive factors in complementizer choice in general. Thus, it adds to our insight into the production of linguistic output that draws on real-life experience.

Works cited

- Biber, D., Conrad, S. & Leech, G. (1999). *Longman grammar of spoken and written English*. Harlow, UK: Longman.
- Kolbe, D. (2008). Complement clauses in British Englishes (Unpublished doctoral dissertation).

- Rohdenburg, G. (1996). Cognitive complexity and increased grammatical explicitness in English. *Cognitive Linguistics* 7, 149-182.
- Steinbach, R. (1929). On usage in English. *American Speech* 4, 161-177.
- Szmrecsanyi, B. & Kolbe-Hanna, D. (to appear). New ways of analyzing dialect grammars: Complementizer omission in traditional British English dialects. In: S. Grondelaers & R. van Hout (Eds.), *New ways of analyzing syntactic variation*. Berlin, Boston: de Gruyter Mouton.
- Tagliamonte, S. & H. Baayen (2012). Models, forests, and trees of York English: *Was/were* variation as a case study for statistical practice. *Language Variation and Change*, 24 (2), 135-178.
- Zieglschmid, A. J. F. (1929). "If" for "whether". *American Speech* 5 (1), 50-51.