

# Annotating Questions under Discussions in Authentic Data

Kordula De Kuthy  
University of Tübingen  
kdk@sfs.uni-tuebingen.de

The information structure of sentences is receiving increased interest in linguistics as the attention has shifted from the analysis of isolated sentences to the question how information is packaged in sentences analyzed in context. In order to connect the information structure of sentences to the overall structure of the discourse, an analysis in terms of *Questions under Discussion* is proving to be a useful tool.

According to Roberts' (2012) account, natural discourse in general serves to answer hierarchically ordered Questions under Discussion (QUDs). These implicit QUDs can be used to account for the information structure of utterances in context: the part of a sentence contained in the formulation of the current question is called the *background*, while the part which provides the actual answer is the *focus*.

The notion of implicit QUDs has also been referenced in corpus-based research attempting to analyze the information structure of naturally occurring, authentic data (e.g., Ritz et al., 2008; Calhoun et al., 2010). Yet these approaches were only rewarded with limited success in terms of achieving agreement, arguably because the task of identifying QUDs was not made explicit.

In this talk reporting joint work with Arndt Riester, we introduce our methodology for a combined analysis of data in terms of both discourse and information structure, integrating an explicitly spelled out notion of QUDs. We identify the necessary steps of an analysis procedure based on QUDs and demonstrate the method on authentic data taken from a German spoken-language corpus. We formulate pragmatic principles that allow us to analyze the discourse structure, formulate adequate QUDs, and analyze the information structure of individual utterances in the discourse. Based on the authentic data analysis, we illustrate that the formulation of QUDs can be successfully guided by the formulated principle and that QUDs play a crucial role in accounting for discourse structural configurations. At the same time, they also provide an objective means to determine the information structure, including both the focus-background divide as well as not-at-issue content.

## References

- Calhoun, S., J. Carletta, J. Brenier, N. Mayo, D. Jurafsky, M. Steedman, and D. Beaver (2010). The NXT-format switchboard corpus: A rich resource for investigating the syntax, semantics, pragmatics and prosody of dialogue. *Language Resources and Evaluation* 44, 387–419.
- Ritz, J., S. Dipper, and M. Götze (2008). Annotation of information structure: An evaluation across different types of texts. In *Proceedings of the 6th International Conference on Language Resources and Evaluation*, Marrakech, Morocco, pp. 2137–2142.
- Roberts, C. (2012). Information structure in discourse: Towards an integrated formal theory of pragmatics. *Semantics and Pragmatics* 5(6), 1–69.