

Variability of German Question Tags

Yulia Clausen Tatjana Scheffler Manfred Stede

{firstname.lastname}@uni-potsdam.de

SFB 1287 “Limits of Variability in Language: Cognitive, Grammatical, and Social Aspects”
University of Potsdam

In this work we analyze German question tags across different media channels. The large inventory of German question tags leads to great dialectal and pragmatic variability (ex. (1)). While tag questions (TQs) are characteristic of conversational speech, they also appear in scripted (The OpenSubtitles¹ corpus, (Lison and Tiedemann, 2016)), and written conversations (Twitter, (Scheffler, 2014)). We address two interrelated questions: (i) Which aspect of conceptual orality (Koch and Oesterreicher, 1985) in media channels facilitates the use of TQs? (ii) What do technological and social settings of the channels say about the pragmatics of the individual question tags?

- (1) Du musst nicht in die Schule, {ne, oder, wa, ja, ...}?
You must not in the school, right?
'You don't have to go to school, do you?'

We examine the occurrence of TQs in German Twitter, telephone (Karins et al., 1997) and scripted conversations through quantitative methods. We find that TQs are most frequent in telephone speech, although they also feature prominently in the other corpora (Figure 1). This indicates that TQs are an important method for establishing and maintaining common ground in conversations, whether spoken, scripted, or written. We further analyze the pragmatic context of a sample of question tags in the three corpora, including e.g. their co-occurrence with modal particles and the speaker's certainty (annotated based on examples in context). The usage pattern reveals significant differences regarding specific question tags across corpora. Overall, TQs frequently occur in all studied corpora, which points to the fact that they are licensed by interactive conversations rather than the spoken mode.

References

- Karins, Krisjanis, et al. 1997. CALLHOME German Transcripts LDC97T15. Web Download.
Philadelphia: Linguistic Data Consortium.
- Koch, Peter, and Wulf Oesterreicher. 1985. Sprache der Nähe-Sprache der Distanz.
Romanistisches Jahrbuch 36:15-43.
- Lison, Pierre, and Jörg Tiedemann. 2016. Opensubtitles2016: Extracting large parallel corpora from movie and TV subtitles. In *Proceedings of the 10th International Conference on Language Resources and Evaluation (LREC)*.
- Scheffler, Tatjana. 2014. A German Twitter snapshot. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC)*, 2284-2289. Reykjavik, Iceland.

¹ <http://www.opensubtitles.org/>

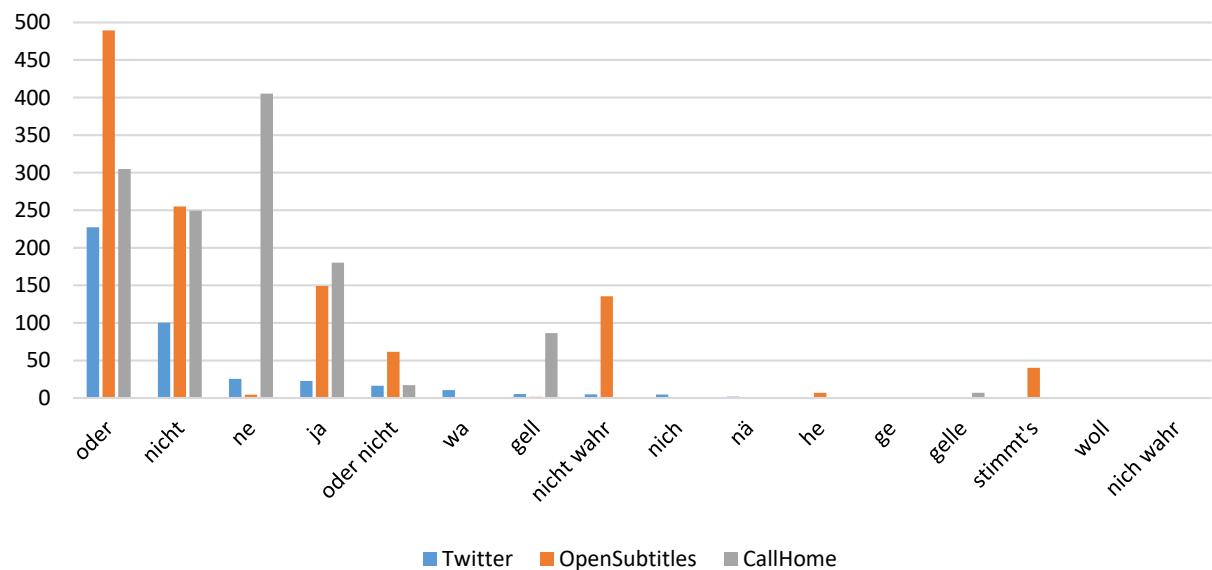


Figure 1: Number of occurrences of German tags in the different types of media per 1 Mio. tokens.