

## Subordination in spoken language: A crosslinguistic corpus study

**Introduction** Studies have argued that spoken or informal varieties tends to show simpler syntax than written or formal varieties of the same language (Biber 1995, Chafe & Tannen 1987, Miller & Weinert 1998, Redeker 1984), e.g. in terms of subordination. Other studies suggested that this may result in systematic crosslinguistic differences between languages with a literacy tradition and languages without (Dąbrowska forthcoming, Martowicz 2011, Mithun 1984). Because most empirical work has focused on single (often Indo-European) languages regarding the long-term influence of register and modality on syntactic complexity, it is still not clear to what extent subordination differs due to language-internal grammatical properties across languages. Therefore, this study examines the properties of complex sentences in comparable, spoken corpus data of 12 typologically diverse languages.

**Corpus & methodology** The dataset is taken from the multicast corpus (Haig & Schnell 2019) which contains natural monologic data from Arta, Cypriot Greek, English, Mandarin, Nafsan, Northern Kurdish, Sanzhi Dargwa, Teop, Tondano, Tulil, and Vera'a, with 1000-5700 clauses per language. The annotation distinguishes between relative, complement, and adverbial clauses for subordination. In order to assess the syntactic complexity of each language, the study compares (amongst others) the relative frequencies of each subordinate clause type, the proportion of clauses per utterance, the average length of clauses, the average number of lexical, pronominal, and zero referents per clause type.

**Selected results** While the proportions of relative clauses are relatively stable across languages (0.03-0.17 per utterance), the proportions of adverbial and complement clauses differ to a larger extent (0.01-0.32 and 0.01-0.23). Fitting Bayesian logistic regression models to predict the clause proportions for each language confirms greater real differences for the latter two clause types. In addition, the frequencies of adverbial and complement clauses are not independent of each other; they have a correlation of 0.83 across languages. Many of the crosslinguistic differences observed can be accounted for by morphosyntactic properties of individual languages: For instance, a generally low proportion of subordination in Mandarin is due to the fact that clausal dependency in the sense of subordination can often not be distinguished morphosyntactically from coordination. The language with the highest proportion of subordination is Sanzhi Dargwa; it uses a high number of non-finite converbs which are dependent clauses. Other than that, the complexity of main and dependent clauses seem relatively similar across languages. The average length of main clauses is 5-8 words; the average length of dependent clauses is 6-10 words; another Bayesian regression model confirms these crosslinguistic similarities.

**Implications** This study offers a first quantitative approach to comparing the syntactic complexity of spoken data in a larger number of typologically distinct languages. The findings suggest that while some clausal properties are stable across languages, the distribution of certain types of subordinate clauses varies to a great extent conditioned by other grammatical properties of the languages. Thus, such crosslinguistic differences have to be taken into account when examining general diachronic effects of written language on subordination.

- Biber, Douglas. 1995. *Dimensions of register variation*. Cambridge: Cambridge University Press.
- Chafe, Wallace & Deborah Tannen. 1987. The Relation between Written and Spoken Language. *Annual Review of Anthropology* 16. 383–407.
- Dąbrowska, Ewa. Forthcoming. How writing changes language. In Anna Mauranen & Svetlana Vetchinikova (eds.), *Language Change: The Impact of English as a Lingua Franca*. Cambridge: Cambridge University Press.
- Haig, Geoffrey & Stefan Schnell. 2019. *Multi-CAST: Multilingual Corpus of Annotated Spoken Texts*. <https://multicast.aspra.uni-bamberg.de>.
- Martowicz, Anna. 2011. *The origin and functioning of circumstantial clause linkers: A cross-linguistic study*. Edinburgh: University of Edinburgh dissertation.
- Miller, Jim & Regina Weinert. 1998. *Spontaneous spoken language: syntax and discourse*. Oxford: Clarendon.
- Mithun, Marianne. 1984. How to Avoid Subordination. *Annual Meeting of the Berkeley Linguistics Society* 10(0). 493–509.
- Redeker, Gisela. 1984. On differences between spoken and written language. *Discourse Processes* 7. 43–55.