

Workshop proposal:

Incorporating the spoken signal into grammatical typology

Laura BECKER (Department of Linguistics, University of Freiburg, Germany),
Naomi PECK (Department of Linguistics, University of Freiburg, Germany)

Description of the Workshop

The written bias in typology

Most languages are primarily spoken, with only a minority of languages or speaker communities developing a writing system. We can estimate that less than 10% of all languages spoken today have developed writing to the extent that they have a literary tradition, with the other 90% being exclusively spoken or written to a much lesser extent (Ong 1982: 7). Despite the primacy of the spoken mode for language, most work on grammar has relied upon the study of written representations, with typology being no exception. The information on grammatical structures needed for a typological study is usually extracted from transcribed examples in reference grammars or taken from a typological database such as WALS or Grambank, which are based on written resources of languages themselves. Even typological or cross-linguistic studies that use corpus data often have to rely on written records, given that most cross-linguistic corpus collections (such as the Universal Dependency treebanks) are based on compilations of written data.

Evidence for the spoken signal affecting grammar

Despite the general reliance on written data, typologists have begun to seriously consider the impact of the spoken signal on grammar. Two notable long-term projects, MultiCAST (Haig & Schnell 2021) and DoReCo (Seifart et al. 2024), have compiled annotated and time-aligned cross-linguistic spontaneous speech corpora, providing invaluable resources for typological corpus studies that consider phonetic and prosodic information for grammatical analysis.

Several typological studies have investigated phonetic properties in relation to grammar with spontaneous speech data from typologically distinct languages. A number of studies explored how phone duration helps to segment the continuous speech signal. Seifart et al. (2021) show that words are systematically lengthened in utterance-final positions across languages. Similarly, Blum et al. (2024) find that consonant lengthening marks the beginning of words. Furthermore, we have direct evidence for grammatical systems being sensitive to durational effects.

Seifart et al. (2018) show that nouns slow down speech compared to verbs, and Becker (Submitted) confirms that high-frequency grammatical markers are phonetically shortened compared to phonologically comparable but less frequent markers in the world's languages.

Similarly, we have evidence that prosody interacts with grammatical structure, especially when it comes to prosodic boundaries and intonation units. For instance, Mettouchi (2018) argues that prosodic integration is key to understanding grammatical relations in Kabyle (Afro-Asiatic). More broadly, Himmelmann (2014, 2022) argues that prosodic boundaries constrain how separate linguistic elements can coalesce phonologically and become grammatical units. This is supported by Peck & Becker (2024), who revealed complex interactions between syntactic boundaries and silent pauses. Similarly, Reinöhl & Casaretto (2018) use evidence from prosodic unithood in historical poems to explain the absence of potential grammaticalization processes in Modern Indo-Aryan languages.

Aim of the workshop

In this workshop, our aim is to bring together typologists who explore how the unique properties associated with the spoken signal are related to grammatical structures across languages. Our objective is to gain a better understanding of how phonetic and prosodic properties interact with other levels of grammatical structures, how they can affect language change and grammaticalization, and what methods we have and need to study the effect of the spoken signal on grammar from a typological perspective.

Topics of the workshop

Potential topics include, but are not limited to:

- empirical investigations of spoken language phenomena in relation to grammar (cross-linguistic studies, single language studies, both experimental or corpus-based)
- explorations of the interaction of suprasegmental properties (e.g. tone, stress, pitch) and spoken language processes (e.g. pausing, durational modulation) with grammar
- explanations for grammatical phenomena which rely on properties of the spoken signal
- methodological reflections on how we can incorporate properties of the spoken signal in synchronic and/or diachronic studies
- methodological investigations on how our practices of writing spoken data influence typological analyses

References

Becker, Laura. Submitted. Frequency effects in verbal argument indexing: A spoken typology approach. *Language*.

- Blum, Frederic, Ludger Paschen, Robert Forkel, Susanne Fuchs & Frank Seifart. 2024. Consonant lengthening marks the beginning of words across a diverse sample of languages. *Nature Human Behaviour*. 1–12.
- Haig, Geoffrey & Stefan Schnell. 2021. *Multi-CAST: Multilingual corpus of annotated spoken texts*.
- Himmelman, Nikolaus P. 2014. Asymmetries in the prosodic phrasing of function words: Another look at the suffixing preference. *Language* 90(4). 927–960.
- Himmelman, Nikolaus P. 2022. Prosodic phrasing and the emergence of phrase structure. *Linguistics* 60(3). 715–743.
- Mettouchi, Amina. 2018. The interaction of state, prosody and linear order in Kabyle (Berber): Grammatical relations and information structure. In Mauro Tosco (ed.), *Afro-Asiatic: Data and perspectives*, 261–285. Amsterdam: John Benjamins.
- Ong, Walter J. 1982. *Orality and Literacy: The Technologizing of the Word*. New York: Methuen.
- Peck, Naomi & Laura Becker. 2024. Syntactic Pausing? Re-examining the associations in spontaneous speech data. *Linguistics Vanguard* 10(1). 223–237.
- Reinöhl, Uta & Antje Casaretto. 2018. When grammaticalization does NOT occur: Prosody-syntax mismatches in Indo-Aryan. *Diachronica* 35(2). 238–276.
- Seifart, Frank, Ludger Paschen & Matthew Stave (eds.). 2024. *Language Documentation Reference Corpus (DoReCo) 2.0*. Lyon.
- Seifart, Frank, Jan Strunk, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Nikolaus P. Himmelman & Balthasar Bickel. 2021. The extent and degree of utterance-final word lengthening in spontaneous speech from 10 languages. *Linguistics Vanguard* 7(1). 20190063.
- Seifart, Frank, Jan Strunk, Swintha Danielsen, Iren Hartmann, Brigitte Pakendorf, Søren Wichmann, Alena Witzlack-Makarevich, Nivja H. de Jong & Balthasar Bickel. 2018. Nouns slow down speech across structurally and culturally diverse languages. *Proceedings of the National Academy of Sciences* 115(22). 5720–5725.