

Prosodic segmentation and grammatical analysis in cross-linguistic corpora

Amina Mettouchi^{1,2}, Martine Vanhove²

¹ Ecole Pratique des Hautes Etudes, ² CNRS LLACAN, Paris

aminamettouchi@icloud.com, martine.vanhove@cnrs.fr

Mittwoch,
04.03.2020
16:30–17:00
VMP5 2101/2105

The development of documentary linguistics and its toolkit of transcription and annotation software has paved the way for the use, in typological research, of spoken corpora from a variety of lesser-described indigenous languages.

The main prerequisite when one compiles such sound-aligned annotated corpora, is to decide how to segment the data, so as to do justice to its spoken nature. Prosodic segmentation (Himmelman 2006), in prosodic units of variable types (intonation unit, paratone, breath groups etc.) reflects that nature far better than segmentation into clauses for instance, which, on top of raising the non-trivial question of the delimitation of clauses (how to identify their initial and final boundaries), creates a residue (speech chunks of various types, cf Chafe 1994).

Prosodic segmentation is not only relevant because of the spoken nature of the data, but also because prosodic groups are essential processing units for speech (Kibrik & Podlesskaya 2006); this is true not only for discourse or information structure, but also for core grammatical functions. The proposed presentation will provide evidence for that claim, using corpora whose prosodic segmentation is detailed in (Izre'el & Mettouchi 2015): CorpAfroAs (<http://corpafroas.huma-num.fr>, 2007–2012) and CorTypo (<http://cortypo.huma-num.fr>, 2013–2018).

The presentation will focus on grammatical relations, showing how their encoding involves not only morphological marking and word order, but also, and crucially, prosodic segmentation. Production phenomena in particular (e.g. prosodic lengthening and disfluencies around boundaries) will be used as evidence for the coding role of prosodic boundaries. Ultimately, the presentation's aim is to foster fruitful discussions on the nature and level of interaction between prosody and core grammar, and on the consequences of such interaction on corpus transcription, segmentation and annotation choices.

References: Chafe, W. (1994). *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing*. Chicago: University of Chicago Press. Himmelman, N. (2006). Prosody in linguistic documentation. In Gippert, J., N. Himmelman & U. Mosel (eds.), *Essentials of Language Documentation*. Berlin, New York: Mouton de Gruyter, 163–185. Izre'el, S. & A. Mettouchi (2015). Representation of speech in CorpAfroAs: Transcriptional strategies and prosodic units. In Mettouchi, A., M. Vanhove & D. Caubet (eds.), *Corpus-based Studies of Lesser-described Languages: The CorpAfroAs corpus of spoken AfroAsiatic languages*. *Studies in Corpus Linguistics* 68. John Benjamins: Amsterdam, Philadelphia, 13–41. Kibrik, A.A. & Podlesskaya, V.I. (2006). Problema segmentacii ustnogo diskursa i kognitivnaja sistema govorjashchego (Segmentation of spoken discourse and the speaker's cognitive system). In *Kognitivnye issledovanija*, Vol. 1, Solovjev, V. D. (ed.). Moscow: Institut psixologii RAN, 138–158.

AG 10