**Syntactic Pausing? Re-examining the associations**

**Overview** In this talk, we will discuss associations between syntactic units and silent pauses in a sample of diverse languages. The relation between pausing and syntactic boundaries has been rarely tested outside of major world languages, with the majority of studies performed in lab conditions utilising reading tasks and/or controlled stimuli. This work represents a first step in the wider testing of the purported correlation of pause length and syntactic unithood using naturalistic data.

**Background** Previous research suggests that pauses are more likely to occur at higher syntactic boundaries, where cognitive planning is required to encode the following unit of speech (e.g. Goldman-Eisler 1968, Cooper and Paccia-Cooper 1980). Later works suggest that prosodic structure and semantic heaviness is more important for the location and length of pauses (Gee and Grosjean 1983, Ferreira 1993). Longer pauses are hypothesised to reflect a heavier planning load, i.e. a higher level syntactic unit, such as a clause or sentence. Spontaneous data from French exhibited a trimodal distribution of pause length (Campione and Véronis 2002), but this was not further linked with syntactic or prosodic phrasing. This possible linkage goes largely uninvestigated.

**Data** To test whether these proposed correlations hold cross-linguistically, we use naturalistic monologues from 11 languages in the MultiCAST corpora (Arta, English, Mandarin, Nafsan, Kurdish, Sanzhi Dargwa, Tabasaran, Teop, Tondano, Tulil, Vera'a) (Haig and Schnell 2021). The files contain annotations for clause boundaries; in addition, we automatically extracted pauses from the provided audio using the silence recogniser in Praat, followed by manual correction. Both levels of annotation were then used to examine the association between clause boundaries and pauses across languages using R.

**Results** Our initial results support previous work advocating against a direct relation between syntax and prosody. Pauses are not strongly associated with clause boundaries in our dataset: between 73% and 94% of pauses occur within clauses rather than at their boundaries across languages in our dataset. This strongly suggests that clause boundaries are but one of the contexts in which we find silent pauses in spontaneous speech. However, we *do* find an association of clause boundaries with pauses, with the majority of main clause boundaries (74%) co-occurring with silent pauses. On the other hand, boundaries of dependent clauses are less likely to co-occur with a pause, independent of possible language and speaker effects. When we model the duration of silent pauses, we indeed find a small effect of main clause boundaries having slightly longer pauses than dependent clause boundaries (estimated mean difference of 80ms).

**Conclusion** Our study represents a first attempt to understand the phenomenon of pausing cross-linguistically by utilising corpus data from understudied languages. We end by outlining future research directions, including the possibility of investigating the associations between silent pauses and other types of units, including syntactic phrases and semantic units.

**References**

Campione, Estelle & Jean Véronis. 2002. A large-scale multilingual study of silent pause duration. *Proceedings of Eurospeech 2002*, 199-202.

Cooper, William E. & Jeanne Paccia-Cooper. 1980. *Syntax and speech*. Cambridge, MA: Harvard University Press.

Ferreira, Fernanda. 1993. Creation of prosody during sentence production. *Psychological Review* 100(2), 233-253.

Gee, James Paul & François Grosjean. 1983. Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology* 15, 411-458.

Goldman-Eisler, Frieda. 1968. *Psycholinguistics: Experiments in spontaneous speech*. London: The Academic Press.

Haig, Geoffrey & Schnell, Stefan (eds.). 2021. *Multi-CAST: Multilingual corpus of annotated spoken texts*. (multicast.aspra.uni-bamberg.de/) (accessed on 16/10/2021)