

**The typology of the temporal fine structure of speech:  
Lengthening and pauses in spoken corpora from 51 languages**

Frank Seifart & Ludger Paschen

Leibniz-Zentrum Allgemeine Sprachwissenschaft(ZAS), Berlin

In this talk we will first present DoReCo (Language Documentation Reference Corpus, see <https://doreco.huma-num.fr/>), a brand-new, openly available CLLD resource consisting of time-aligned and morphologically annotated spoken language corpora on 51 small and often endangered languages from around the world. These corpora were extracted from language documentation collections and processed into a uniform format in collaboration with experts on these languages. We will then introduce an ongoing research program that uses this resource to study cross-linguistic variation regarding pause probabilities and the local acceleration and deceleration of speech such as phrase-final lengthening. DoReCo's broad cross-linguistic coverage allows us to reveal areal prosodic patterns, as well as dependencies between pauses and phonetic de-/acceleration, on the one hand, and language-specific grammatical structure, such as basic word order and phonological features, on the other hand.