

# Incorporating Pite Saami: Combining Lexicography, Morphophonology, Syntax in Automatic Corpus Creation

Joshua Wilbur

Albert-Ludwigs-Universität Freiburg

The main goal of the project “Syntactic Structures in Pite Saami: A corpus-based exploration of 130 years of variation and change” is to ascertain syntactic patterns in Pite Saami (Uralic, Sweden) based mainly on empirical evidence found in a heavily annotated corpus. These annotations ideally track information on lemma, part of speech, morphological categories and English glosses. The first stages of the project focus on creating such a corpus out of both my own recordings and heritage materials. Completing such annotations by hand is exceptionally time consuming, so I have focused instead on developing a way to annotate automatically using computational methods. In doing so, lexicographic, morphophonological and syntactic data are combined together to make this possible. In this talk, I will describe how the interplay between these different areas of linguistics inform one another, and come together to make this possible, as well as outline the challenges encountered.