

Janne Bondi Johannessen: The Nordic Dialect Corpus.

In this talk I will present and demonstrate the Nordic Dialect Corpus, a joint effort following collaboration between dialectologists, linguists and language technologists in the networks ScanDiaSyn and NORMS. The corpus has been developed as a research tool for the Nordic dialects of six countries: Denmark, Faroes, Finland, Iceland, Norway, and Sweden. Both w.r.t. contents and usability this corpus offers a wide range of options. The dialects contained in the corpus are each represented by at least one type of transcription, and these are linked directly to audio and video. They are all in the process of being grammatically tagged. The dialect in most cases represent the speech of both younger and older people, and they represent most areas where the language is spoken. The usability of the corpus is taken care of by the general user interface Glossa, developed at the Text Laboratory, which ensures that all kinds of contents and metavariables regarding the corpus are searchable in a unified and simple user interface, and that the search results can be further processed.