

## Dorothee Beermann & Atle Prange

### *Annotating and archiving natural language paradigms online*

TypeCraft (TC) is an online tool for the interlinearized glossing of natural language paradigms. The main purpose of annotation is to illuminate the meaning and the grammatical structure of natural language expressions, which in TC is stored in the form of sentence or phrasal level tokens, categorized according to the language and the construction(s) they represent. Our main goal in designing TC was to make linguistic glossing 'easy', 'interactive' and 'software independent', as well as to make relevant linguistic data available to a larger group of researchers. TC is a development of LinLab@ntnu, a research group for formal systems of natural languages at the Norwegian University of Science and Technology (NTNU), Trondheim. At present TC offers a free user interface for the retrieval of annotated linguistic paradigms. Search parameters are 'language' and 'construction'. For those that would like to use TC as an annotation tool, an account is created upon request. The use of TC does not require the installation of software. TC offers, next to a QuickStart page, glossaries that can be called upon under annotation. Under 'Help' the user finds explanation of all glosses known to TC and further explanations about basic linguistic parameters, which will aid the user in situating his own sentence paradigms within a network of linguistic construction types. TC is open to the use of new glosses, but signals the difference between glosses that are known and unknown to the system. TC offers an internal messaging system. Feedback to TC is possible at any point of the annotations process. In a comment field the annotator can provide additional linguistic information that will be stored together with the language tokens and can be retrieved in the TC public interface by other users of the system. TC is in its beta phase. The following improvements of the system are under development:

- \* Improved information sharing. We aim at making the search results available in several different formats. xml, latex, jpg and txt formats are currently underway.
- \* Improved communication: A forum that connects users and enables them to discuss and exchange information within the context of their work.
- \* Better support for Internet Explorer.
- \* at present TC data is not archived. With TC v1.0 we will start to archive TC resources.