

Linguistic Units in the Writing Process. Systematic Study of Planning, Revising, and Editing

Dr. phil. Cerstin Mahlow

Writing is one of the main communicative acts; every day we produce large amounts of written text. And writing is a complex task: it involves the coordination of numerous cognitive, linguistic, and technical aspects. While linguistic aspects of the texts produced are widely researched in linguistics and computational linguistics, the *process* of producing linguistic units, such as multi-word expressions, co-reference chains, and discourse structures is still an underresearched topic.

In this project I will systematically explore the production of multi-word expressions with a focus on discourse elements, based on keystroke-logging data for ca. 90 writing sessions. Preliminary studies show that such units are produced in a way that indicates they are actually processed and produced as fixed units by the writer. The results of these studies will be used to design controlled experiments to investigate specific linguistic units.

Keystroke-logging of writing sessions allows researchers to reconstruct the writing process, we can link the product and the process and explore how specific phrases have been produced, which revisions and editing have taken place, whether authors actually produce the main argument as the very first or whether they re-arrange their argumentative essays various times. However, keystroke-logging produces large amounts of data, which can only be analyzed automatically. I will evaluate which grammar formalism (including its implementation) is the most suitable to handle the incremental nature of text production. The results of this evaluation will influence the development of detailed automatic processing of writing process data.