Dr. Sabine Bartsch, Prof. Dr. Elke Teich, Monica Holtz
Technische Universität Darmstadt
{bartsch, teich, holtz}@linglit.tu-darmstadt.de

Corpus-based register profiling: The case of mechanical engineering texts

Features of domain specific language are hard to trace systematically; text corpora can serve as a useful source of authentic data in order to get a balanced picture of the distribution and co-occurrence of the features characteristic of a particular register.

The project reported on in this paper aims to develop a profile of the register of its target domain, data processing in construction, a subdiscipline of mechanical engineering, by establishing a matrix of interacting linguistic features that jointly characterise the register of that domain.

We report on the design, compilation and annotation of a text corpus from the target domain which covers a variety of text types ranging from academic articles over course materials presented on the web to traditional text books, thus comprising expert-to-expert as well as teacher-student communication.

The study incorporates systemic functional linguistics with corpus and computational linguistic techniques as its theoretical and methodological basis. This combination allows the systematic investigation of features at the lexical level (collocation) as well as features at the grammatical level (colligation).

References:

Biber, Douglas, Susan Conrad, Randi Reppen. 1998. Corpus Linguistics. Investigating Language Structure and Use. Cambridge, New York, Melbourne: Cambridge University Press.
Halliday, MAK 2004. An introduction to functional grammar. London: Arnold. (Third edition revised by C. Matthiessen).
Matthiessen, C. 1999. The system of TRANSITIVITY: an exploratory study of text-based profiles. Functions of Language 6.1.