
A functional stylistics for register and genre

Thomas Haider
 Universität Heidelberg
haider@cl.uni-heidelberg.de

Alexis Palmer
 University of North Texas
palmer@cl.uni-heidelberg.de

Donnerstag
 09.03.2017
 10:00 – 10:30
 B4 1, 0.04

Work on register for German is rather scarce, compared to English. We narrow this gap by (i) developing a theoretically grounded comparative typology for genre and register analysis, (ii) compiling a corpus of German register and genre out of DeReKo (Kupietz et al., 2010), (iii) developing high accuracy machine learning algorithms for supervised text genre classification, (iv) providing an extensive description of prototypical register classes through the agglomeration of style feature loadings. We thus contribute both to linguistic theory building and to research on genre adaptation of NLP methods within the LiMo project.

Register can be understood as the function between *situational context* and *linguistic analysis* (Biber & Conrad, 2009) or as *functional stylistics* i.e. the distribution of style features in communicative areas (Jakobson, 1960). In designing the corpus typology (taxonomy) we make use of prototype theory (Rosch, 1973) that affords the embedding of a basic-level category (communicative purpose) within a super-level (social text use) and a sub-genre (topic modelling). We model the prototypical text classes using dimensions such as *document size*, a ‘natural’ distribution of *topics*, or the *dominant register* of a genre.

We then develop a machine learning system for register and genre classification. We focus on a high-quality feature extraction pipeline which implements style features through numerous morphosyntactic annotations, lexicons of linguistic nature (i.a. verb classes), psycholinguistic word norms (i.a. concreteness), and topic models. We identify the best models for supervised text genre classification, achieving results comparable to the state-of-the-art for English. Finally, we show that our work is also applicable to English. Here, we are able to identify pervasive functional dimensions using PCA (to wit, interpersonal and narrative dimensions in literary genre).

References: • Biber, D. and Conrad, S. (2009). *Register, genre, and style*. Cambridge University Press.
 • Jakobson, R. (1960). *Closing statement: Linguistics and poetics*. *Style in language*, 350:377.
 • Kupietz, M., Belica, C., Keibel, H., and Witt, A. (2010). The german reference corpus DeReKo: A primordial sample for linguistic research. In LREC.
 • Rosch, E. H. (1973). Natural categories. *Cognitive psychology*, 4(3):328–350.

AG13