# Genre knowledge as an asset for NLP?: A classification experiment with German roommate search posts

**Sophie Decher**
***Fraunhofer Institut für Kommunikation, Informationsverarbeitung und Ergonomie (FKIE)***
sophie.decher@fkie.fraunhofer.de

Natural language processing (NLP) models are becoming ever larger and more powerful, yet pragmatic elements of language remain a challenge. The output of such statistical language models is highly dependent on the amount, type, and quality of training data used. The current study aims to investigate whether an NLP model can learn to recognize genre elements in written texts.

Genres have been defined as "class[es] of communicative events, the members of which share some set of communicative purposes" (Swales 1990) and as "institutionalized template[s] for social interaction" (Yates & Orlikowski 2002). Prototypical instances of a particular genre share microstructural and macrostructural elements (cf. Van Dijk 1980). Users familiar with the genre can recognize and reproduce these elements; expert users may exploit genre conventions and manipulate elements to further their communicative goals. Knowledge of genre conventions is integral for pragmatic awareness and pragmatic competence in both L1 and nonnative speakers (Ifantidou 2011).

An example of one such genre is arguably the German roommate search post. These texts are written with the communicative goal of finding a room in a shared apartment with one or more roommates (referred to as a *Wohngemeinschaft* or *WG*). Successful exemplars of this genre share certain formulaic elements—authors often include their current occupation, place of origin, and reason for moving, for example. An annotation scheme for the micro- and macrostructural elements in this genre was developed based on a self-compiled, manually annotated corpus of 72 online roommate search posts from WG-Gesucht.de. To validate this proposed annotation scheme, a model was trained with the annotated texts and unseen data was used to test whether the structural elements can be identified automatically. The results have implications for the roommate search post as a possible new genre and for training methods in NLP.

**References:** • Ifantidou, E. (2011). Genres and pragmatic competence. *Journal of Pragmatics, 43*, 327–346. • Swales, J. M. (1990). *Genre Analysis: English in academic and research settings.* Cambridge/New York/Melbourne: Cambridge University Press. • van Dijk, T. A. (1980). *Macrostructures: An interdisciplinary study of global structures in discourse, interaction, and cognition.* Hillsdale, New Jersey: Lawrence Erlbaum Associates. • Yates, J., & Orlikowski, W. (2002). Genre systems: Structuring interaction through communicative norms. *The Journal of Business Communication, 39*(1), 13–35.