As readers we are sensitive to various levels of structure within texts. The line of research described in this talk addresses two levels: the contribution of individual clauses and that of an intermediate level of text structure -- passages of several sentences or more. We follow Smith (2003) in discussing the Discourse Modes of individual text passages. Smith introduces the five modes of Narrative, Description, Report, Information, and Argument/Commentary and further identifies specific linguistic correlates of these different modes. One group of correlates points to the contributions made to the discourse by individual clauses of text. As a first step toward Discourse Mode classification, we address the problem of identifying and classifying references to situations expressed in written English texts. The situation entity (SE) classification task as construed here is the classification of clauses according to the type of entity they introduce to the discourse. The classification scheme we use includes, among others, events, states, abstract entities, and generic sentences.

In this talk I will discuss both previous and ongoing work on SE classification. In earlier work (Palmer et al. 2007), we annotate a portion of the Brown corpus with SE labels and then use this for training and testing models for SE classification. The two key strategies employed are (1) use of linguistically-motivated features and (2) implementation of SE classification as a sequence labeling task. We find that deep linguistic information improves classifier performance. Our results also provide empirical support for the notion of modes of discourse, as we see clear genre effects in SE classification. Our ongoing work refines the SE classification scheme from earlier work and aims to more carefully annotate a much larger corpus of texts from a number of different genres, thus laying the foundation for more extensive computational treatment of SE type and related linguistic features. Finally, I will present recent work on automatic classification for one such feature: inherent lexical aspect (Friedrich and Palmer 2014).
Kurzlebenslauf:
Alexis Palmer is a postdoctoral researcher at the Department of Computational Linguistics and Phonetics, Saarland University (Saarbrücken, Germany). Her current research focuses primarily on computational approaches to discourse and semantics, including work on discourse modes, situation entities, short-answer scoring, and working with learner language. In addition, Alexis has a second research focus on computational linguistics for low-resource languages; her 2009 PhD thesis "Semi-Automatic Annotation and Active Learning for Language Documentation" was completed at the University of Texas at Austin under the joint supervision of Jason Baldridge and Katrin Erk.
http://www.coli.uni-saarland.de/~apalmer/

Ansprechpartner:
Prof. Dr. Iryna Gurevych (gurevych@ukp.informatik.tu-darmstadt.de, Tel.: 06151/16-5411)