

Abstract von Sabine Schulte im Walde

Semantic verb classifications, i.e., groupings of verbs according to semantic properties, are of great interest to both theoretical and computational linguistics. In theoretical linguistics, verb classes are a useful means to organise verbs with respect to common properties, such as meaning components (Koenig and Davis, 2001), or shared argument structure (Levin, 1993). In computational linguistics, semantic verb classifications represent a valuable source of underspecification, by generalising over the verbs according to their shared properties.

Specifically regarding the pervasive problem of data sparseness in the processing of natural language, such classifications have been used in computational applications such as word sense disambiguation (Dorr and Jones, 1996; Kohomban and Lee, 2005), machine translation (Prescher et al., 2000; Koehn and Hoang, 2007), document classification (Klavans and Kan, 1998), and statistical lexical acquisition in general (Merlo and Stevenson, 2001; Schulte im Walde, 2006).

This talk presents a novel approach to a semantic classification of verbs, that incorporates selectional preferences as common verb properties. Similarly to previous related work (Pereira et al., 1993; Rooth et al., 1999), we rely on the Expectation-Maximisation (EM) Algorithm as a soft-clustering technique, and model verb classification by probabilistic class membership of verbs and their semantic properties. In contrast to earlier work, we choose a more complex set of semantic properties: rather than directly using bilexical head dependencies between verbs and (direct object) nouns as clustering dimensions, we abstract over the noun dimension by selectional preferences. Consequently, a semantic class generalizes over verb senses (as one dimension), and selectional preferences (as a second dimension).

The talk will introduce the classification approach in some detail, by providing (a) an overview of the parameters of the clustering technique (the input, the probabilistic model, and the implementation), and (b) potential interpretations and applications, relying on examples for English and German verbs.