Representing Underspecification by Semantic Verb Classes Incorporating Selectional Preferences

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 generalises over verb senses (as one dimension), and selectional preferences (as a second dimension), as illustrated by the fictitious example in Table 1: the left column *Verbs* shows a list of verbs that are ordered by the probability of being members of this class; the *Selectional Preferences* column presents a list of selectional preferences, also ordered by the class membership probability; each selectional preference is a triple $< frame, argument, concept >$, where the concept (e.g., *event*) has been determined as selectional preference description for the specific argument (e.g., prepositional phrase headed by *about*) within the specific frame type (e.g., subj+pp). The implicit assumption behind our clustering model is that verbs are assigned to a common class if they agree in their subcategorisation properties, as referred to by the selectional preferences.

Verb Class			
	Selectional Preferences		
Verbs	Frame	Argument in Frame	Preference
talk	subj:pp-about	pp-about	event
report	subj:obj	obj	event
negotiate	subj:pp-about	pp-about	phenomenon
discuss	subj:pp-about	subj	group
complain	subj:obj	subj	person
...	subj	subj	person
	...		

Table 1: Example class with verbs and selectional preferences.

The classification approach is introduced in some detail, by providing an overview of the parameters of the clustering technique,

1. the *input*: tuples with joint frequencies for verbs, frames and argument nouns are induced from parsed corpus data, e.g., *talk, subj:pp-about, president, education* → *freq=43*.

2. the *probabilistic model*: The probability of a verb-argument tuple $p(v, f, a_1, ..., a_n)$ is defined as the product of the prior probabilities for classes $p(c)$, verb and frame probabilities given the class $p(v|c)$ and $p(f|c)$, and selectional preference parameters.

3. the *implementation*: The clusters are implemented as a graph structure; estimation of the cluster parameters is performed by the Inside-Outside algorithm on data tuples, and maximisation is performed on the complete graph over all data tuples.

4. the *induction of selectional preferences*: The lexical taxonomy WordNet (Fellbaum, 1998) is exploited for selectional preference induction, applying a variant of a state-of-the-art approach using *Minimum Description Length* (Li and Abe, 1998).

5. the *interpretation* of the clustering results: The cluster analyses are interpreted, based on cluster membership probabilities and relating semantic class properties (i.e., selectional preferences) to verb senses and verbal polysemy.

References

Bonnie J. Dorr and Doug Jones. Role of Word Sense Disambiguation in Lexical Acquisition: Predicting Semantics from Syntactic Cues. In *Proceedings of the 16th International Conference on Computational Linguistics*, pages 322–327, Copenhagen, Denmark, 1996.

Christiane Fellbaum, editor. *WordNet – An Electronic Lexical Database*. Language, Speech, and Communication. MIT Press, Cambridge, MA, 1998.

Judith L. Klavans and Min-Yen Kan. The Role of Verbs in Document Analysis. In *Proceedings of the 17th International Conference on Computational Linguistics and the 36th Annual Meeting of the Association for Computational Linguistics*, pages 680–686, Montreal, Canada, 1998.

Philipp Koehn and Hieu Hoang. Factored Translation Models. In *Proceedings of the Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*, pages 868–876, Prague, Czech Republic, 2007.

Jean-Pierre Koenig and Anthony R. Davis. Sublexical Modality and the Structure of Lexical Semantics. *Linguistics and Philosophy*, 24:71–124, 2001.

Upali S. Kohomban and Wee Sun Lee. Learning Semantic Classes for Word Sense Disambiguation. In *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 34–41, Ann Arbor, MI, 2005.

Beth Levin. *English Verb Classes and Alternations*. The University of Chicago Press, 1993.

Hang Li and Naoki Abe. Generalizing Case Frames Using a Thesaurus and the MDL Principle. *Computational Linguistics*, 24(2):217–244, 1998.

Paola Merlo and Suzanne Stevenson. Automatic Verb Classification Based on Statistical Distributions of Argument Structure. *Computational Linguistics*, 27(3):373–408, 2001.

Fernando Pereira, Naftali Tishby, and Lillian Lee. Distributional Clustering of English Words. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics*, pages 183–190, Columbus, OH, 1993.

Detlef Prescher, Stefan Riezler, and Mats Rooth. Using a Probabilistic Class-Based Lexicon for Lexical Ambiguity Resolution. In *Proceedings of the 18th International Conference on Computational Linguistics*, 2000.

Mats Rooth, Stefan Riezler, Detlef Prescher, Glenn Carroll, and Franz Beil. Inducing a Semantically Annotated Lexicon via EM-Based Clustering. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics*, Maryland, MD, 1999.

Sabine Schulte im Walde. Experiments on the Automatic Induction of German Semantic Verb Classes. *Computational Linguistics*, 32(2):159–194, 2006.