Identifying the Causative Alternation in English

Esther Seyffarth

Heinrich-Heine-Universität Düsseldorf

esther.seyffarth@hhu.de

The causative alternation in English is an example of a *diathesis alternation* where semantic roles are encoded in different syntactic positions, as demonstrated in sentences (1) and (2):

(1) The number of students has decreased.
(2) We have decreased the number of students.

Identifying verbs that participate in this alternation is not trivial: The syntactic patterns of these verbs overlap with patterns of verbs with optional objects.

This poster presents and evaluates 8 different strategies for the automatic identification of English verbs in the causative alternation, based on statistics about syntactic patterns, distributional information about arguments of the verbs, perplexity scores calculated by a Recurrent Neural Network, and Verb-Net classes. Each approach is evaluated in three test conditions for two sources of gold data (Levin, 1993; VerbNet class annotations). The wide range of accuracy scores in our different test conditions shows how important reproducible evaluation conditions are for this type of task.

Earlier work in this area, such as McCarthy (2000), Schulte im Walde (2000), McCarthy (2001), Merlo and Stevenson (2001), tends to evaluate on small or hand-picked test sets. We compare our approaches to the work of Merlo and Stevenson (2001). They achieve an accuracy of 69% on the three-way distinction of unaccusative, unergative, and unexpressed-object verbs. One of our unsupervised approaches outperforms that of Merlo and Stevenson by 6% on a balanced test set. Our setup is simpler than that of Merlo and Stevenson, requiring only a dependency-parsed corpus from which to derive statistics, whereas their system makes use of features like animacy or causativity, which pose problems when a verb or its arguments are attested infrequently.

The methods we present can be applied to any role-switching alternation (McCarthy, 2001) in any language for which the necessary preprocessing tools are available. In future work, we plan to compare the performance of our approaches to the performance they achieve on other languages, which will be informative from a typological perspective.

References: • Karin Kipper et al. (2000): Class-Based Construction of a Verb Lexicon. *Proc. of AAAI 2000*, pp. 691-696. • Levin, B. (1993): *English Verb Classes and Alternations: A Preliminary Investigation.* University of Chicago press. • McCarthy, D. (2000): Using Semantic Preferences to Identify Verbal Participation in Role Switching Alternations. *Proc. of NAACL 2000*, pp. 256-263. • McCarthy, D. (2001): *Lexical Acquisition at the Syntax-Semantics Interface: Diathesis Alternations, Subcategorization Frames and Selectional Preferences.* Ph.D. thesis, University of Sussex. • Merlo, P., and Stevenson, S. (2001): Automatic Verb Classification Based on Statistical Distributions of Argument Structure. *Computational Linguistics* 27(3), pp. 373-408.