Criteria for Manual Clustering of Verb Senses

Cecily Jill Duffield, Jena D. Hwang, Susan Windisch Brown,
Dmitriy Dligach, Sarah E. Vieweg, Jenny Davis, Martha Palmer
({cecily.duffield, hwangd, susan.brown, dmitriy.dligach, sarah.vieweg, jennifer.davis, martha.palmer}@colorado.edu)
Departments of Linguistics and Computer Science
University of Colorado
Boulder, C0 80039-0295, USA

Key words: word sense disambiguation, annotation, inter-annotator agreement, syntactic/semantic features

Introduction

Word sense ambiguity poses significant obstacles to accurate and efficient information extraction and automatic translation. Successful disambiguation of polysemous words in NLP applications depends on determining an appropriate level of granularity of sense distinctions, especially for verbs. WordNet, an important and widely used lexical resource, uses fine-grained distinctions that provide subtle information about the particular usages of various lexical items (Felbaum, 1998). When used as a resource for annotation of various genres of text, this fine level of granularity has not been conducive to high rates of inter-annotator agreement (ITA) or high automatic tagging performance. Annotation of verb senses as described by coarse-grained Proposition Bank framesets may result in higher ITA scores, but the blurring of distinctions between verb senses with similar argument structures may fail to alleviate the problems posed by ambiguity. Our goal in this project is to create verb sense distinctions at a middle level of granularity that allow us to capture as much information as possible from a lexical item while still attaining high ITA scores and high system performance in automatic sense disambiguation. We have demonstrated that clear sense distinctions improve annotator productivity and accuracy, which results in a corresponding improvement in system performance. Training on this new data, Chen, Schein, Ungar and Palmer, (2006) report 86.7% accuracy for verbs using a smoothed maximum entropy model and rich linguistic features (just over 70% for fine-grained senses). This paper focuses on the methodology used to create the sense groupings, with a particular emphasis on the types of features that are most accessible to human annotators who are not linguists.

Various criteria are considered when disambiguating senses and creating sense groupings for the verbs, including frequent lexical usages and collocations, syntactic features and alternations, and semantic features, similarly to the groupings for Senseval2 (Palmer, Dang & Felbaum, 2007). Our highest priority is to create clear distinctions among sense groupings that will be easily understood by the annotators and consequently result in high rates of inter-annotator agreement. We have found that the most successful approach is to cluster senses intuitively on a verb-by-verb basis, distinguishing sense groupings with features that are easily grasped by all annotators. Such features include specific domain usages, as in legal, financial, and social uses (distinguishing two senses of *integrate* in, "Over two-thirds of the teachers report they *integrated* the arts into their subjects," and "The movie is set in 1971, when TC Williams High *integrated* blacks and whites."); a specific syntactic construction, such as a required locative prepositional phrase (distinguishing the sense of *open* in, "The master bedroom *opens* to a large terrace," from that in "The door won't *open*."); and the features of nominal arguments, such as an agentive subject (separating senses of *indicate* in "These symptoms *indicate* a serious illness," and "He *indicated* the right road by nodding towards it.")

More theoretical features for distinguishing groupings have proven to be less successful. Annotators not familiar with linguistics were confused by concrete/abstract distinctions and such aspectual features as continuative or stative. Therefore, they are now rarely used to label sense groupings. Such concepts, when used, are more likely to be described in prose commentary. Certain compositional features of verbs such as *manner* and *path* have also proven to be confusing for annotators, and resulted in decreased annotator agreement. Verb sense groupings that do not receive high ITA scores in initial rounds of annotation are revised, often prioritizing the use of the more successful features illustrated above with examples.

Acknowledgements

We gratefully acknowledge the support of the National Science Foundation Grant NSF-0415923, Word Sense Disambiguation, and the Defense Advanced Research Projects Agency GALE program, Contract No. HR0011-06-C-0022, a subcontract from the BBN-AGILE Team.

References

Chen, J., Schein, A., Ungar, L., & Palmer, M. (2006). An empirical study of the behavior of word sense disambiguation. *Proceedings of NAACL-HLT 2006*. New York City, NY.

Fellbaum, C. (Ed.). (1998). *WordNet: An on-line lexical database and some of its application*s. MIT Press, Cambridge, MA.

Palmer, M., Dang, H.T., and Fellbaum, C. (to appear, 2007). Making fine-grained and coarse-grained sense distinctions, both manually and automatically. *Journal of Natural Language Engineering*