

Alternative Annotations of Word Usage

Diana McCarthy,
Department of Informatics,
University of Sussex
Falmer BN1 9QJ, UK
dianam@sussex.ac.uk

Abstract

Right from Senseval's inception there have been questions over the choice of sense inventory for word sense disambiguation (Kilgarriff, 1998). While researchers usually acknowledge the issues with predefined listings produced by lexicographers, such lexical resources have been a major catalyst to work on annotating words with meaning. As well as the heavy reliance on manually produced sense inventories, the work on word sense disambiguation has focused on the task of selecting the single best sense from the predefined inventory for each given token instance. There is little evidence that the state-of-the-art level of success is sufficient to benefit applications. We also have no evidence that the systems we build are interpreting words in context in the way that humans do. One direction that has been explored for practical reasons is that of finding a level of granularity where annotators and systems can do the task with a high level of agreement (Navigli et al., 2007; Hovy et al., 2006). In this talk I will discuss some alternative annotations using synonyms (McCarthy and Navigli, 2007), translations (Sinha et al., 2009) and WordNet senses with graded judgments (Erk et al., to appear) which are not proposed as a panacea to the issue of semantic representation but will allow us to look at word usages in a more graded fashion and which are arguably better placed to reflect the phenomena we wish to capture than the 'winner takes all' strategy.

References

- Katrin Erk, Diana McCarthy, and Nick Gaylor. Investigations on word senses and word usages. In Proceedings of ACL-IJCNLP 2009, to appear.
- Eduard Hovy, Mitch Marcus, Martha Palmer, Lance Ramshaw, and Ralph Weischedel. Ontonotes: The 90% solution. In *Proceedings of the HLT-NAACL 2006 workshop on Learning word meaning from non-linguistic data*, New York City, USA, 2006. Association for Computational Linguistics.
- Adam Kilgarriff. Gold standard datasets for evaluating word sense disambiguation programs. *Computer Speech and Language*, 12(3):453–472, 1998.
- Diana McCarthy and Roberto Navigli. SemEval-2007 task 10: English lexical substitution task. In *Proceedings of (SemEval-2007)*, pages 48–53, Prague, Czech Republic, 2007.
- Roberto Navigli, Kenneth C. Litkowski, and Orin Hargraves. SemEval-2007 task 7: Coarse-grained English all-words task. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-2007)*, pages 30–35, Prague, Czech Republic, 2007.
- Ravi Sinha, Diana McCarthy, and Rada Mihalcea. Semeval-2010 task 2: Cross-lingual lexical substitution. In *Proceedings of the NAACL-HLT Workshop SEW-2009*, Boulder, Colorado, USA, 2009.