Epilogue to "Resolving lexical ambiguity computationally with spreading activation and Polaroid Words"

Graeme Hirst
Department of Computer Science
University of Toronto
Toronto, Ontario M5S 3G4, Canada
gh@cs.toronto.edu
www.cs.toronto.edu/∼gh

14 April 2003

Now that the ideas of software agents and of object-oriented programming are ubiquitous, it's much easier to explain Polaroid Words than it was in 1988 when the paper was published, or in 1982, when the work commenced (Hirst and Charniak 1982):[1] PWs are communicating asynchronous processes that work together, swapping information about words and thematic roles, adjuncts, and arguments so that each can complete its own little disambiguation task. External to the processes, and used by them, is an extensive (but hypothetical) knowledge source in which word senses are represented and from which the semantic distance between senses can be determined. The processes run even as the sentence is being parsed and interpreted, so that they can help the parser with attachment decisions.

Because PWs required a resource whose eventual existence was merely hypothesized, and because there were no test data available, it was obviously not possible to empirically test or evaluate the system or its underlying ideas in any serious way. Rather, PWs and a knowledge source for marker passing were built for a few interesting words, and the system was run on a few interesting constructed examples to ensure that it did "the right thing". This approach to evaluation was quite normal in the milieu in which the research was carried out and didn't seem to worry anyone at the time: CL had not yet achieved its empirical orientation. Since the advent of the Senseval competitions (see Adam Kilgarriff's commentary in this volume), the field has been developing methods and data for the formal empirical evaluation of word-sense disambiguation systems. These competitions have demonstrated just how difficult the task is and how much (or how little) progress we have made.[2]

[1] In fact, the idea in the paper that has fared the worst since it was published is the metaphor in the name "Polaroid Words" itself. Advances in digital imaging have changed the idea of what an "instant picture" is; many readers of the paper in this volume will never have seen the gradual development, over a minute or two, of a Polaroid SX-70 "instant" picture. A more-contemporary metaphor would be progressive GIFs, themselves already obsolescent as bandwidth rapidly increases.

[2] Although the 1988 paper contained a number of promises for continued development of Polaroid Words, they were never taken

The abstract notion of the semantic or conceptual distance between concepts is perhaps the most important theme of the paper. Understanding and using the idea of semantic or conceptual distance has become a much more active research area in computational linguistics as suitable resources have become available. One such resource is online thesauri. For example, in 1991, Morris and Hirst developed a set of rules for semantic distance that relied on *Roget's Thesaurus*. They used this to segment text at points where the topic changed. Subsequently, when WordNet became available (Miller 1990), many other researchers began to use the connections in it as the basis for measures of semantic distance between word senses, both for word-sense disambiguation and for other applications; see Budanitsky (1999) for an extensive survey. Hirst and St-Onge (1998) and Budanitsky and Hirst (2001) used a semantic-distance measure to detect and correct real-word spelling errors in text: a word that is semantically distant in all its senses from all others in its context, but which is a spelling variation of one that would be semantically close, is possibly an error.

But as many have discovered, the conceptual distance represented by one link in WordNet varies widely, and this inconsistency inherently limits the consistency of measures of semantic distance in WordNet. Moreover, resources that make *all* the kinds of connections that PWs assumed are still only on the horizon: although WordNet includes meronymy and holonymy relations and cause and entailment relations between verbs, many other relationships are omitted, such as purpose (*broom–sweep*), typical-agent (*bark–dog*), and typical-action (*television–watch*). A resource that is more like that which PWs assumed is FrameNet (Johnson and Fillmore 2000), still under construction, the goal of which is to describe the complete conceptual structures that underlie the meanings of words. (In the absence of ideal resources for determining the conceptual distance between two words, many other approaches have been developed that rely, in one way or another, on the similarity of behavior of the words—for example, occurrence in similar contexts or similar documents. For a survey, see Dagan (2000) and Lebart and Rajman (2000).)

Even as the paper was being written, the complexity of the mechanisms and resources that PWs invoked was challenged by the work of Michael Lesk, who in 1986 proposed a method for word-sense disambiguation that used no more than string-matching in a resource that already existed: an online dictionary. Lesk's method (reprinted in this volume) picked the sense whose dictionary definition had the greatest number of words occurring in the context of the target word in the text. This was a reductio, almost a parody, of PWs: a marker-passing connection between two words is approximated by the occurrence of one in the definition of the other; syntactic structure is simply ignored. Lesk's method was intended for use in information-retrieval systems, resolving ambiguity in queries or answers; he didn't worry about case roles, about helping a parser with attachment decisions, or about working with a semantic interpretation process aimed at a deep level of "understanding".

And, generally speaking, many of the approaches to word-sense disambiguation that were explored in the early 1990s (exemplified by several papers in this section of the present volume) and the subsequent decade (Edmonds, Mihalcea, and Saint-Dizier 2002; Agirre and Edmonds 2004) share these characteristics. That is,

beyond the version that it described.

they are less ambitious than Polaroid Words: they rely on an existing resource and try to see how much can be done with it, rather than aiming for a complete solution and hypothesizing a resource that this necessitates.

One issue has that remained constant is what kinds of information in the text may be drawn upon as cues for disambiguation, and how close to the target word those cues should be. In the 1988 paper, the restrictions on communication between Polaroid Words arose from two competing principles: that any particular word or structural cue for disambiguation has quite a limited sphere of influence; and yet almost anything in a text or discourse is potentially a cue for disambiguation (compare McRoy 1992). In contemporary, classification-oriented systems for word sense disambiguation, the analogous dilemma is in the choice of features and the window size.

Since the Polaroid Words paper was published, my work has concentrated on related issues of lexical and semantic knowledge. In particular, in addition to the research on semantic distance cited above and research on the limitations of semantic representations (Hirst 1991), I have looked at the problems that arise in computational language understanding and generation when the distance between word meanings is *small*—that is, when two words are *near-synonyms*, differentiated by small nuances of meaning (DiMarco and Hirst 1993, Hirst 1995, Edmonds and Hirst 2002).

References

Agirre, Eneko and Edmonds, Philip (editors) (2004). *Word Sense Disambiguation: Algorithms and Applications*. Kluwer Academic Publishers.

Budanitsky, Alexander (1999). *Lexical Semantic Relatedness and its Application in Natural Language Processing.* Technical report CSRG-390, Department of Computer Science, University of Toronto, August 1999.
http://www.cs.toronto.edu/compling/Publications/Abstracts/Theses/Budanitsky-thabs.html

Budanitsky, Alexander and Hirst, Graeme (2001). "Semantic distance in WordNet: An experimental, application-oriented evaluation of five measures." *Workshop on WordNet and Other Lexical Resources*, Second meeting of the North American Chapter of the Association for Computational Linguistics, Pittsburgh, June 2001.
http://www.cs.toronto.edu/compling/Publications/Abstracts/Papers/Budanitsky+Hirst-2001-abs.html

Dagan, Ido (2000). "Contextual word similarity." In: Robert Dale, Hermann Moisl, and Harold Somers (editors), *Handbook of Natural Language Processing*, New York: Marcel Dekker, 2000, 459–475.

DiMarco, Chrysanne and Hirst, Graeme. "A computational theory of goal-directed style in syntax." *Computational Linguistics*, 19(3), September 1993, 451–499.
http://www.cs.toronto.edu/compling/Publications/Abstracts/Papers/DiMarco+Hirst-CL-93-abs.html

Edmonds, Philip and Hirst, Graeme (2002). "Near-synonymy and lexical choice." *Computational Linguistics*, 28(2), 105–144.

http://www.cs.toronto.edu/compling/Publications/Abstracts/Papers/Edmonds+Hirst-2002-abs.html

Edmonds, Philip; Mihalcea, Rada; and Saint-Dizier, Patrick (editors) (2002). *Proceedings of the Workshop on Word Sense Disambiguation: Recent Successes and Future Directions.* Philadelphia, July 2002. Association for Computational Linguistics

Hirst, Graeme (1991). "Existence assumptions in knowledge representation." *Artificial Intelligence*, 49, May 1991, 199–242. Reprinted in: Brachman, Ronald J.; Levesque, Hector J.; and Reiter, Raymond (editors). *Knowledge Representation*. Cambridge, MA: The MIT Press, 1992.

Hirst, Graeme (1995). "Near-synonymy and the structure of lexical knowledge." *Working notes, AAAI Symposium on Representation and Acquisition of Lexical Knowledge: Polysemy, Ambiguity, and Generativity*, Stanford University, March 1995, 51–56.

http://www.cs.toronto.edu/compling/Publications/Abstracts/Papers/Hirst-NearSynonyms-95-abs.html

Hirst, Graeme and Charniak, Eugene (1982). "Word sense and case slot disambiguation". *Proceedings of the Second National Conference on Artificial Intelligence (AAAI-82)*, Pittsburgh, August 1982, 95–98.

Hirst, Graeme and St-Onge, David (1998). "Lexical chains as representations of context for the detection and correction of malapropisms". In: Christiane Fellbaum (editor), *WordNet: An electronic lexical database*, Cambridge, MA: The MIT Press, 1998, 305–332.

Johnson, Christopher R. and Fillmore, Charles J. (2000). "The FrameNet tagset for frame-semantic and syntactic coding of predicate-argument structure" . *Proceedings of the 1st Meeting of the North American Chapter of the Association for Computational Linguistics*, Seattle, 56–62.

Lebart, Ludovic and Rajman, Martin (2000). "Computing similarity." In: Robert Dale, Hermann Moisl, and Harold Somers (editors), *Handbook of Natural Language Processing*, New York: Marcel Dekker, 2000, 477–505.

Lesk, Michael (1986). "Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone." *Proceedings, Fifth International Conference on Systems Documentation (SIGDOC '86)*, Toronto, 8–11 June 1986, 24–26. Reprinted in this volume.

McRoy, Susan (1992). "Using multiple knowledge sources for word sense discrimination." *Computational Linguistics*, 19(1), 1–30.

Miller, George (1990). "WordNet: An on-line lexical database." *International Journal of Lexicography*, 3(4), 235–312.

Morris, Jane and Hirst, Graeme (1991). "Lexical cohesion, the thesaurus, and the structure of text." *Computational Linguistics*, 17(1), 21–48.

http://www.cs.toronto.edu/compling/Publications/Abstracts/Papers/Morris+Hirst-91-abs.html

Roget's Thesaurus. Many editions and publishers.