

Book Reviews

WordNet: An Electronic Lexical Database

Christiane Fellbaum (editor)
(Rider University and Princeton University)

Cambridge, MA: The MIT Press
(Language, speech, and communication series), 1998, xxii+423 pp; hardbound, ISBN 0-262-06197-X, \$50.00

Reviewed by
Dekang Lin
University of Manitoba

WordNet is perhaps the most important and widely used lexical resource for natural language processing systems up to now. *WordNet: An Electronic Lexical Database*, edited by Christiane Fellbaum, discusses the design of WordNet from both theoretical and historical perspectives, provides an up-to-date description of the lexical database, and presents a set of applications of WordNet. The book contains a foreword by George Miller, an introduction by Christiane Fellbaum, seven chapters from the Cognitive Sciences Laboratory of Princeton University, where WordNet was produced, and nine chapters contributed by scientists from elsewhere.

Miller's foreword offers a fascinating account of the history of WordNet. He discusses the presuppositions of such a lexical database, how the top-level noun categories were determined, and the sources of the words in WordNet. He also writes about the evolution of WordNet from its original incarnation as a dictionary browser to a broad-coverage lexicon, and the involvement of different people during its various stages of development over a decade. It makes very interesting reading for casual and serious users of WordNet and anyone who is grateful for the existence of WordNet.

The book is organized in three parts. Part I is about WordNet itself and consists of four chapters: "Nouns in WordNet" by George Miller, "Modifiers in WordNet" by Katherine Miller, "A semantic network of English verbs" by Christiane Fellbaum, and "Design and implementation of the WordNet lexical database and search software" by Randee Teng. These chapters are essentially updated versions of four papers from Miller (1990). Compared with the earlier papers, the chapters in this book focus more on the underlying assumptions and rationales behind the design decisions. The description of the information contained in WordNet, however, is not as detailed as in Miller (1990).

The main new additions in these chapters include an explanation of sense grouping in George Miller's chapter, a section about adverbs in Katherine Miller's chapter, observations about autohyponymy (one sense of a word being a hyponym of another sense of the same word) and autoantonymy (one sense of a word being an antonym of another sense of the same word) in Fellbaum's chapter, and Teng's description of the Grinder, a program that converts the files the lexicographers work with to searchable lexical databases.

The three papers in Part II are characterized as "extensions, enhancements and

new perspectives on WordNet". Marti Hearst's Chapter 5, "Automated discovery of WordNet relations," investigates automatic detection of WordNet-style lexicosemantic relationships in large corpora, using rules such as this:

"such NP_0 as $NP_1, \dots, \{and/or\} NP_k$ " $\implies NP_0$ is a hypernym of $NP_1, \dots,$ and NP_k .

Similar techniques have been adopted in many previous approaches, as she notes. She also sketches a procedure for discovering new patterns, although it was not implemented. She manually inspected 200 instances that matched one of her patterns. About 20% of the hypothesized hypernym relations were already in WordNet. About 30% were not in WordNet but were classified as "good" or "pretty good." The rest were errors of various kinds.

Chapter 6, "Representing verb alternations in WordNet" by Kohl, Jones, Berwick, and Nomura, augments WordNet with Beth Levin's (1993) classification of English verbs. Since the underlying hypothesis of Levin's work is that semantic properties of words determine their syntactic properties, it would be extremely interesting to see the result of superimposing an independently constructed semantic structure, WordNet, onto Levin's verb classifications. It is a pity, therefore, that this enhancement did not make it into WordNet 1.6, as predicted in the book.

Chapter 7, "The formalization of WordNet by methods of relational concept analysis" by Uta Priss, attempts to formalize WordNet using set-theoretic concepts. According to Priss, "[the theoretical analysis] does not provide a complete system of axioms for semantic relations, but it can facilitate the investigation of the logical properties of those relations" (p. 179). She shows three fragments of WordNet where the relationships could be better structured. It is not clear, unfortunately, how the formalization could identify these fragments, leaving one to wonder whether it is simply a fancy way to state something obvious.

The chapters in Part III are about applications that use WordNet in a variety of ways: as a list of word senses (Chapters 8 and 9), as a taxonomy hierarchy (Chapter 10), and as a semantic network (Chapters 11–16).

Chapters 8 and 9, "Building semantic concordances" by Landes, Leacock, and Tengi and "Performance and confidence in a semantic annotation task" by Fellbaum, Grabowski, and Landes, are concerned with SemCor, a 250,000-word corpus in which all the open-class words are tagged with word senses from WordNet. This corpus can serve several purposes, such as giving feedback to lexicographers about the appropriateness and completeness of word senses and providing frequency information and example sentences for word senses. While SemCor is probably too small for statistical learning, it is certainly large enough to act as a test bed for word sense disambiguation systems.

Chapter 8 describes the construction process of SemCor, the software tool used (called ConText), the user interface of ConText, and analysis of errors and inconsistency. It also covers the training of taggers and some quality control issues.

Chapter 9 is a description and analysis of an experiment to measure the accuracy and confidence in the semantic tagging task. The subjects in the experiment are 17 taggers (undergraduate and graduate students). They each tagged 254 polysemous words in a 660-word passage. For each sense tag, they also indicated the degree of confidence for the tag by assigning a number from 1 (highly certain) to 5 (highly uncertain). The taggers were divided into two groups. In the lexicon given to the first group (8 taggers), the senses of words were listed in descending order of their frequencies. In the lexicon given to the second group (9 taggers), they were in random

order. The sense tags assigned by the taggers were compared with answer keys created by expert lexicographers. The percentage of agreement between the taggers and the experts, as well as between the taggers themselves, were measured. It was found that when word senses were ordered by their frequencies, the tagger-expert agreement was 75.2% and intertagger agreement was 79.7%; when word senses were randomly ordered, the tagger-expert agreement was 72.8% and intertagger agreement was 79.9%. The explanation for the higher intertagger agreement was that "naïve speakers" (taggers) have a mental lexicon different from that of the lexicographers. The even bigger difference for the group using randomly ordered senses was explained by the hypothesis that, under such a condition, "the taggers must have examined all senses rather carefully before making a selection" (p. 220). It is surprising to me that the authors would resort to these two additional hypotheses when the observations could be explained by a hypothesis they have already made, i.e., taggers examine word senses in the order they are listed and ignore the rest of the list when they find one sense that is satisfactory to them. This common search strategy means that taggers tend to make the same mistakes, and consequently explains the higher intertagger agreement. The group using frequency-ordered senses got higher tagger-expert agreement than the other group because this search strategy works better under that condition.

Philip Resnik's Chapter 10, "WordNet and class-based probabilities," estimates the probabilities of the concepts (synsets) in WordNet with an untagged text corpus. The probability distribution can then be used to determine the "selectional preference strength" of verbs. For example, the verb *drink* has a much stronger selectional preference on its object than the verb *have*. The most interesting aspect of this work is that it combines symbolic knowledge about linguistic relationships with statistical knowledge about language use. With the addition of statistical knowledge, the relationships in WordNet can be quantitatively differentiated. With the symbolic taxonomy in WordNet, probabilities can be distributed over classes, as well as words.

The next four chapters deal with the word sense disambiguation problem in one way or another. The disambiguation algorithms in the four chapters are based on the same assumption: in the local context of the target word (the word to be disambiguated), one can expect to find other words that are closely related to the intended meaning of the target word. Given this assumption, the intended meaning of a word can be identified by scoring its potential senses with the potential senses of words in the local context or by finding connections among the senses of the word to senses of other words in the context and eliminating those senses that are not involved in any connection.

In "Combining local context and WordNet similarity for word sense identification" by Leacock and Chodorow, senses of the target word are scored by their similarity to the senses of other words in the local context (e.g., ± 2 words). The authors also combined this method with a naïve Bayes type of algorithm and showed that the combination resulted in significant improvements (about 5%).

Ellen Voorhees's Chapter 12, "Using WordNet for text retrieval," is based on her earlier work (Voorhees 1993). Potential senses of the target word are scored by totaling the frequencies of the words in their respective "hoods." The hood of a word's sense is the maximal portion of WordNet that contains the sense but not any other sense of the word. The main finding in her experiment with word sense disambiguation in query expansion is the following: when sense disambiguation is perfectly correct, query expansion with WordNet can improve the performance of short queries, but it does not make any significant difference with long queries; when the disambiguation algorithm is less than perfect, query expansion can even hurt the retrieval performance.

Hirst and St-Onge's Chapter 12 is titled "Lexical chains as representations of context for the detection and correction of malapropisms." "A malapropism is the confounding of an intended word with another word of similar sound or spelling that has a quite different and malapropos meaning, for example *an ingenuous* [for *ingenious*] *machine for peeling oranges*." A lexical chain is a path that connects senses of a group of words in a document. Each link in the path is a lexicosemantic relationship in WordNet and is associated with a weight that indicates the strength of the relationship. Hirst and St-Onge imposed a set of constraints on the paths to "ensure the path corresponds to a reasonable relation between the source and the target word." For example, a hyponym link cannot be followed later on by a hypernym link. In other words, after the context is narrowed down, it must not be enlarged again. An alarm is raised if a word is not connected through a lexical chain to any other word in the context, but a similarly spelled word would be. Hirst and St-Onge tested their system on a 322,645-word corpus with 1,409 malapropisms. Their results showed that alarms were raised for 28.2% of the malapropisms and the false-alarm rate is 87.5%. The basic idea in this chapter is the same as that of Morris and Hirst (1991), which used *Roget's Thesaurus* instead of WordNet. However, the algorithm of Morris and Hirst was not implemented due to the lack of lexical resources.

Chapter 14, "Temporal indexing through lexical chaining" by Al-Halimi and Kazman, discusses the use of trees, instead of paths, to connect related words in transcripts of audio tapes. The goal of their application is to retrieve segments of audio tapes that are relevant to a query. This is achieved by creating a lexical tree for the query and retrieving the tape segments with lexical trees that are most similar to it.

Chapter 15, "COLOR-X: Using knowledge from WordNet for conceptual modeling" by Burg and van de Riet, is only tangentially related to computational linguistics. Their basic idea is the following: Since the conceptual models of software systems involve many classes of entities and relationships that are represented in WordNet, why not retrieve them from WordNet so that the software designers do not have to come up with the relationships themselves?

The last chapter, "Knowledge processing on an extended WordNet" by Harabagiu and Moldovan, treats WordNet as a semantic network. A marker-passing algorithm similar to that of Charniak (1986) and Norvig (1989) was employed to make "text inferences." The markers are claimed to be "intelligent markers" that could enforce their own constraints. However, the "intelligence" of the markers is not explicitly described in the paper. The paper contains more elaborate examples than earlier marker-passing papers. Unfortunately, that seems to be all. The algorithm is not implemented nor tested with real data.

One problem with the last three chapters is the lack of proper evaluation of the proposed algorithms. Al-Halimi and Kazman evaluated their lexical-tree-building algorithm by comparing its output on a single 1,800-word article with keywords selected by an unspecified number of human subjects. Neither Burg and van de Riet nor Harabagiu and Moldovan performed any form of evaluation.

Since WordNet is a large-scale lexical resource, without quantitative evaluation, it may be impossible to predict how well an algorithm will work or even whether or not it will work at all. The following example is found in Voorhees's chapter (p. 294):

The nouns *nail*, *hammer*, and *carpenter* are all good hints that the intended sense of *board* is the 'lumber' sense. However, within WordNet a nail is a fastener, which in turn is a device, so *nail* would help select the 'control panel' sense of *board*. Similarly, a hammer is a tool, which is an implement, which is an article of commerce, so *hammer* would

help select the ‘dining table’ sense of *board*. Finally, a carpenter is a worker, which is a person, which is both an agent and a life form, which are both things. Thus, *carpenter* would not help select any sense of *board*.

Similarly, as pointed out by Hirst and St-Onge, in WordNet, *stew* and *steak* are not closely related, but *public* and *professional* are.

About half of the chapters are revised versions of their authors’ earlier publications in journals or reasonably accessible conference proceedings. Perhaps for this reason, readers who look for brand-new ideas in the book may feel somewhat disappointed. On the other hand, given the importance of WordNet, it is convenient to have them in a single collection. Furthermore, it offers a historical perspective of WordNet and a relatively complete coverage of its applications.

The book also highlights some common issues that arise from different applications. For example, all the application papers that are related to word sense disambiguation expressed the need for what Hirst and St-Onge called “situation relations” (p. 318), which connect entities involved in the same event or scenario, such as *Nasdaq—share* and *hospital—physician*. One of George Miller’s assumptions about WordNet is that lexical knowledge can be separated from other types of knowledge. Incorporation of such relations in WordNet would mean the abandonment of this assumption, as situation relations do not seem to be part of lexical knowledge. Another possibility is that situation relations should be acquired from corpus data, instead of being encoded in WordNet. However, none of the chapters explored this idea.

I found the discussions about lexicosemantic relationships in Part I most insightful and thought-provoking. As a description of the software, however, papers in Part I are not as systematic and organized as those of Miller (1990) (which are included in the WordNet software distribution at <http://www.cogsci.princeton.edu/~wn/>). There is a great deal of variation in the quality of papers in Parts II and III. Overall, I consider the book to be worthwhile.

References

- Charniak, Eugene. 1986. A neat theory of marker passing. In *Proceedings of the 5th National Conference on Artificial Intelligence (AAAI-86)*, volume 1, pages 584–588.
- Levin, Beth. 1993. *English Verb Classes and Alternations*. University of Chicago Press.
- Miller, George A., editor. 1990. WordNet: An on-line lexical database. Special issue of *International Journal of Lexicography*, 3(4).
- Morris, Jane and Graeme Hirst. 1991. Lexical cohesion computed by thesaural relations as an indicator of the structure of text. *Computational Linguistics*, 17: 21–48.
- Norvig, Peter. 1989. Marker passing as a weak method for text inferencing. *Cognitive Science*, 13: 569–620.
- Voorhees, Ellen M. 1993. Using WordNet to disambiguate word senses for text retrieval. In *Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 171–180. ACM Press.

Dekang Lin is Associate Professor of Computer Science at the University of Manitoba. His research interests include principle-based broad-coverage parsing, information extraction, word sense disambiguation, and learning from parsed corpora. Lin’s address is: Department of Computer Science, University of Manitoba, Winnipeg, Manitoba, Canada, R3T 2N2; e-mail: lindek@cs.umanitoba.ca