Acquisition of Semantic Lexicons: Using Word Sense Disambiguation to Improve Precision

Bonnie J. Dorr and Doug Jones

Department of Computer Science and
Institute for Advanced Computer Studies
University of Maryland
A. V. Williams Building
College Park, MD 20742
{bonnie,jones}@umiacs.umd.edu

Abstract

This paper addresses the problem of large-scale acquisition of computational-semantic lexicons from machine-readable resources. We describe semantic filters designed to reduce the number of incorrect assignments (i.e., improve precision) made by a purely syntactic technique. We demonstrate that it is possible to use these filters to build broad-coverage lexicons with minimal effort, at a depth of knowledge that lies at the syntax-semantics interface. We report on our results of disambiguating the verbs in the semantic filters by adding WordNet[1] sense annotations. We then show the results of our classification on unknown words and we evaluate these results.

1 Introduction

This paper addresses the problem of large-scale acquisition of computational-semantic lexicons from machine-readable resources. We describe semantic filters designed to reduce the number of incorrect assignments (i.e., improve precision) made by a purely syntactic technique. We demonstrate that it is possible to use these filters to build broad-coverage lexicons with minimal effort, at a depth of knowledge that lies at the syntax-semantics interface. We report on our results of disambiguating the verbs in the semantic filters by adding WordNet sense annotations. We then show the results of our classification on unknown words and we evaluate these results.

As machine-readable resources (i.e., online dictionaries, thesauri, and other knowledge sources) become readily available to NLP researchers, automated acquisition has become increasingly more attractive. Several researchers have noted that the average time needed to construct a lexical entry by hand can be as much as 30 minutes (see, e.g., (Neff and McCord, 1990; Copestake et al., 1995; Walker and Amsler, 1986)). Given that most large-scale NLP applications require lexicons of 20-60,000 words, automation of the acquisition process has become a necessity.

Previous research in automatic acquisition focuses primarily on the use of statistical techniques, such as bilingual alignment (Church and Hanks, 1990; Klavans and Tzoukermann, 1996; Wu and Xia, 1995), or extraction of syntactic constructions from online dictionaries and corpora (Brent, 1993; Dorr, Garman, and Weinberg, 1995). Others who have taken a more knowledge-based (interlingual) approach (Lonsdale, Mitamura, and Nyberg, 1996) do not provide a means for systematically deriving the relation between surface syntactic structures and their underlying semantic representations. Those who have taken more argument structures into account, e.g., (Copestake et al., 1995), do not take full advantage of the systematic relation between syntax and semantics during lexical acquisition.

[1]We used Version 1.5 of WordNet, available at http://www.cogsci.princeton.edu/~wn.

We adopt the central thesis of Levin (1993), i.e., that the semantic class of a verb and its syntactic behavior are predictably related. We base our work on a correlation between semantic classes and patterns of grammar codes in the Longman's Dictionary of Contemporary English (LDOCE) (Procter, 1978). We extend this work by coupling the syntax-semantics relation with a pre-defined association between WordNet (Miller, 1985) word senses and Levin's verbs in order to group the full set of LDOCE verbs into semantic classes.

While the LDOCE has been used previously in automatic extraction tasks (Alshawi, 1989; Farwell, Guthrie, and Wilks, 1993; Boguraev and Briscoe, 1989; Wilks et al., 1989; Wilks et al., 1990) these tasks are primarily concerned with the extraction of other types of information including syntactic phrase structure and broad argument restrictions or with the derivation of semantic structures from definition analyses. The work of Sanfilippo and Poznanski (1992) is more closely related to our approach in that it attempts to recover a syntactic-semantic relation from machine-readable dictionaries. However, they claim that the semantic classification of verbs based on standard machine-readable dictionaries (e.g., the LDOCE) is "a hopeless pursuit [since] standard dictionaries are simply not equipped to offer this kind of information with consistency and exhaustiveness."

Others have also argued that the task of simplifying lexical entries on the basis of broad semantic class membership is complex and, perhaps, infeasible (see, e.g., Boguraev and Briscoe (1989)). However, a number of researchers (Fillmore, 1968; Grimshaw, 1990; Gruber, 1965; Guthrie et al., 1991; Hearst, 1991; Jackendoff, 1983; Jackendoff, 1990; Levin, 1993; Pinker, 1989; Yarowsky, 1992) have demonstrated conclusively that there is a clear relationship between syntactic context and word senses; it is our aim to exploit this relationship for the acquisition of semantic lexicons.

We first describe the LDOCE verb classification resulting from a purely syntactic approach to deriving semantic classes. We then describe a semantic filter designed to reduce the number of incorrect assignments made by the syntactic technique; we show how this filter can be enhanced with a method that accounts for multiple word senses. Finally we show the results of our classification of unknown verbs, and we evaluate these results. Our results clearly indicate that the resolution of polysemy is a key component to developing an effective semantic filter.

2 Verb Classification Based on Syntactic Behavior

We build on the syntactic filter approach of (Dorr, Garman, and Weinberg, 1995), in which verbs were automatically classified into semantic classes using syntactic encodings in LDOCE. This earlier approach produced a ranked assignment of verbs to the semantic classes from (Levin, 1993) based on syntactic tests (e.g., whether a verb occurs in a dative construction such as *Mary gave John the book*).[2] The syntactic approach alone was demonstrated to classify Levin verbs with 47% accuracy (i.e., 1812 correct verb classifications out of 3851 possible assignments).

The measure of success used in the purely syntactic approach is flawed in that the "accuracy" factor was based on the number of correct assignments in the five top-ranked assignments produced by their algorithm. A better measure of the efficacy of the algorithm would be to examine the ratio of correct assignments to the total number of assignments. The algorithm in (Dorr, Garman, and Weinberg, 1995) is correct only 13% of the time (1812 correct assignments out of 13761 total assignments) if given up to 5 assignments per verb. If given up to 15 assignments, the situation

[2] Levin's semantic classes are labeled with numbers ranging from 9 to 57; the actual number of semantic classes is 191 (not 46) due to many class subdivisions under each major class. These 191 classes cover 2813 verbs that occur in the LDOCE. Since verbs may occur in multiple classes, the number of possible assignments of LDOCE verbs into classes is 3851.

would deteriorate further: even though 2607 out of 3851 possible assignments would be correct, these correct assignments constitute only 6.5% of the total number of assignments made by the algorithm.

We borrow terminology from Information Filtering (see, e.g., (Lewis, 1992)) to characterize these results. In particular, *Recall* is the number of correct categorizations the algorithm gives divided by the number of correct categorizations already given in the database. *Precision*, on the other hand, is the number of correct categorizations that the algorithm gives divided by the total number of categorizations that it gave. In these terms, the algorithm in (Dorr, Garman, and Weinberg, 1995) achieves a recall of 67.7%, but a precision of 6.5% if given up to 15 semantic class assignments per verb.

In addition to low precision, the purely syntactic filter described above was tested only on verbs that are in (Levin, 1993) and it did not take into account the problem of multiple word senses. The remainder of this paper describes the formulation and refinement of semantic filters that increases the precision of this earlier experiment, while extending the coverage to novel verbs (i.e., ones not occurring in (Levin, 1993)) and addressing the polysemy problem.

3 Semantic Filter: Increasing Precision

We take as our starting point 7767 LDOCE verbs, approximately 5000 of which do not occur in Levin's classes. Each of these verbs was assigned up to 15 possible semantic classes, ranked by the degree of likelihood that the verb belongs to that class, giving a total of 113,106 ranked assignments. As described above, the syntactic filter discovers 2607 of the 3851 assignments of LDOCE verbs found in Levin's semantic classes. These assignments are particularly interesting because we know they are correct, and we can see how high the program ranks the correct assignments.

To create a semantic filter, we take a semantic class from Levin and extend it with related verbs from WordNet. We call this extended list a *semantic field*. Verbs that do not occur in the semantic field of a particular class fail to pass through the semantic filter for that class, by definition. We first examined different semantic relations provided by WordNet (synonymy, hyponymy, both synonyms and hyponyms, and synonyms of synonyms) in order to determine which one would be most appropriate for constructing semantic fields for each of Levin's 191 verb classes. We evaluated the performance of these different relations by examining the degree of class coverage of the relation using a prototypical verb from each class.[3]

For example, the Change of State verbs of the *break* subclass (Class 45.1) contains the verbs *break, chip, crack, crash, crush, fracture, rip, shatter, smash, snap, splinter, split, tear*. The full semantic field contains the union of the related verbs for every verb in the original Levin class. Thus, if we build our semantic field on the basis of the synonymy relation, all synonyms of verbs in a particular class would be legal candidates for membership in that class. For Class 45.1, using the synonymy relation would result in a field size of 185 (i.e., there are 185 WordNet synonyms for the 13 verbs in the class); by contrast, the hyponymy relation would yield a field size of 245.

To choose a relation to use for the semantic field, we looked at verbs semantically related to the prototypical verb in each class, and checked how many of the verbs in each class would be included in the filter. We examined several relations based on combinations of synonymy and hyponymy. We considered the best candidate to be the one that matched the greatest proportion of the verbs in Levin's semantic classes when given the prototype verb. The best relation, synonyms of the

[3]A verb is considered to be *prototypical* with respect to a class if it conforms to all of Levin's membership tests for that class. These tests are based on grammaticality of usage in certain well-defined contexts (e.g., the dative construction).

	All	Filtered
Total Assignments	40,248	4168
Right Assignments	2,607	2607
Wrong Assignments	37,641	1561
Precision (Right/Total)	6.5%	62.5%

Table 1: Increasing Precision with the Semantic Filter

prototype verb, matched an average of 20% of the Levin verbs, while having an average size of 11 verbs. The average size of Levin's semantic classes is 22 verbs.

Let us now look at the behavior of the synonymy-based semantic filter. Of the 113,106 assignments of LDOCE verbs to Levin classes given by the syntactic filter, 6029 (19%) pass through the semantic filter. Clearly, the semantic filter constrains the possible assignments, but the question to ask is whether the constraint improves the accuracy of the assignments. To answer this, we first examined the 2813 verbs in LDOCE that also appear in Levin to see if they matched Levin's categorization.

Without the semantic filter, the syntactic filter provides up to 15 semantic-class assignments for each of the 2813 verbs, giving 40,248 assignments, as shown in Table 1. 2,607 of these assignments (6.5%) are correct. When we add the semantic filter, the number of assignments drops to 4168, 10% of the unfiltered assignments. 2607 of these (62.5%) are correct, a twelve-fold improvement over the unfiltered assignments.

By *Right Assignments*, we mean: cases in which the system assigns a verb to a given Levin class, when that verb appears in that class in Levin's book. By *Wrong Assignments*, we mean: cases in which the system assigns a verb to a given Levin class, when that verb does not appear in that class in Levin's book.

It is important to point out that even though the semantic filter is based on words in Levin, it still sometimes categorized the Levin verb incorrectly. Since the filter is based on synonyms of Levin verbs, in some cases, a synonym of a verb from some other class will appear in the set that does not belong there. In this case, there are 1561 assignments known to be wrong, out of a total of 4168 assignments. For example, the verb *scatter* is a synonym of *break* in WordNet. Because the verb *break* occurs in each of these classes, the semantic filter based on synonyms assigns *scatter* to classes 10.6 (*Cheat* Verbs), 23.2 (*Split* Verbs), 40.8.3 (*Hurt* Verbs), 45.1 (*Break* Verbs), 48.1.1 (*Appear* Verbs). But the correct class for *scatter* is 9.7 (*Spray/Load* Verbs). This illustrates the difficulty of using an approach that does not account for multiple word senses. We will address this point further in section 3.

Setting aside the polysemy problem, we see that this semantic filter is very useful for reducing the number of incorrect assignments.

4 Performance on Novel Words

We now examine how well it performs on unknown words by constructing a semantic filter based on three different proportions of the original 2813 Levin verbs: (a) 50%, (b) 70%, and (c) 90%, chosen randomly.[4] We then checked whether the "unknown" verbs (those not used to construct

[4] We chose randomly selected subsets: First we selected a random 90% of the Levin verbs, then we chose 77.7% of those to give 70% of the Levin verbs. In turn, 71.4% of those give the verbs for the 50% study.

Semantic-Filter Assignments to Levin Classes

% Levin	*Original Assignments*		*Number of Guesses*			*Ratios*	
			Total	Wrong	Right	Precision	Recall
50%	known	1282	1752	470	1282	73.2%	100.0%
	novel	1325	841	429	412	49.0%	31.1%
70%	known	1798	2628	830	1798	68.4%	100.0%
	novel	809	663	360	303	45.7%	37.5%
90%	known	2341	3632	1291	2341	64.5%	100.0%
	novel	266	271	158	113	41.7%	42.5%
100%	all known	2607	4168	1561	2607	62.5%	100.0%

Original Syntactic-Filter Assignments to Levin Classes

% Levin	*Original Assignments*		*Number of Assignments*			*Ratios*	
			Total	Wrong	Right	Precision	Recovery
100%	Known	2607	40248	37641	2607	6.5%	100%

Table 2: Undisambiguated Synonyms

the semantic filter) were assigned to their correct classes.

Table 2 summarizes the recall and precision results for semantic filtering on these three different proportions of Levin verbs. Consider the rows that show the behavior of the experiment which uses 50% of Levin's verbs, and tries to guess the remaining verbs using synonymy. Recall that there are 2607 verbs all together. In this case, 1282 verbs were chosen at random to use in constructing the filter. We call these the "known" verbs. This leaves 1325 for use in evaluating the semantic filter—we call these the "novel" verbs. For the 1282 known verbs, the filter made 1752 assignments to semantic classes. There were 470 wrong assignments and 1282 right ones, giving a precision rate of 73.2% and recall rate of 100.0% .

5 The Effect of Disambiguation

As mentioned previously, the problem with the semantic filter we have defined is that it is not sensitive to multiple word senses of the particular verbs in the semantic classes. For example, there are 23 senses of the verb *break* in WordNet. This includes senses which correspond to the Change of State verbs, such as Sense 9, "break, bust, cause to break", the synonyms of which are *destroy, ruin, bust up, wreck, wrack*. But it also includes irrelevant senses, such as Sense 7, "break dance", the synonyms of which are *dance, do a dance, perform a dance*. Clearly, the semantic filter would behave better if we used word senses in creating the fields. As an attempt to address the polysemy problem, we conducted an exploratory study in which the verbs in Levin's semantic classes were disambiguated by hand: each verb received as many WordNet senses as were applicable.

The performance of the various filters is shown in Table 3. To see the effect of disambiguation, compare the difference between undisambiguated and disambiguated synonyms. Precision has increased from 62.5% to 85.3% . For novel verbs, in the experiment which uses 50% of the verbs and

Undisambiguated Synonyms				
%	*Known*		*Novel*	
Levin	Recall	Precision	Recall	Precision
100%	100.0%	62.5%	0.0%	0.0%
90%	100.0%	64.5%	42.5%	41.7%
70%	100.0%	68.4%	37.5%	45.7%
50%	100.0%	73.2%	31.1%	49.0%
Disambiguated Synonyms				
%	*Known*		*Novel*	
Levin	Recall	Precision	Recall	Precision
100%	100.0%	85.3%	0.0%	0.0%
90%	100.0%	86.2%	29.3%	63.9%
70%	100.0%	88.3%	26.1%	68.5%
50%	100.0%	91.7%	21.6%	70.8%
Disambiguated Hyponyms of Hypernyms				
%	*Known*		*Novel*	
Levin	Recall	Precision	Recall	Precision
100%	100.0%	37.7%	0.0%	0.0%
90%	100.0%	39.0%	68.8%	29.5%
70%	100.0%	41.5%	63.0%	31.1%
50%	100.0%	45.8%	58.6%	34.6%
Union of Disambiguated Synonyms with Hyponyms of Hypernyms				
%	*Known*		*Novel*	
Levin	Recall	Precision	Recall	Precision
100%	100.0%	37.6%	0.0%	0.0%
90%	100.0%	38.9%	69.5%	29.7%
70%	100.0%	41.4%	64.4%	31.5%
50%	100.0%	45.8%	59.6%	34.9%

Table 3: Comparison of Filters

tries to guess the rest, the precision increases from 49.0% to 70.8% . But notice also that the recall decreases: with disambiguation (in the 50% study), recall drops from 31.1% for undisambiguated verbs to 21.6% for disambiguated verbs. The reason for this is that the undisambiguated filters contain numerous assignments which are correct but are included only accidentally.

Table 3 also shows the performance of two other semantic filters based on hyponyms. We found that using hyponyms of hypernyms (going up one level in abstraction, and then one level back down) gave much better recall than plain synonymy, although the precision is lower. We also built a filter based on the union of synonyms with hyponyms of hypernyms. The effect of the synonyms on this filter was negligible, presumably since synonyms are often hyponyms of hypernyms. The results for both of these filters are shown in Table 3.

6 Conclusion and Future Work

Our main result is that the semantic field substantially reduces the number of incorrect assignments given by the syntactic filter. One of our goals is to assign new verbs, i.e., all of the verbs in LDOCE, to the semantic classes of Levin. Since there are 7767 verbs in LDOCE, and there are 191 semantic classes in Levin, there are 1,483,497 potential assignments of verbs to these semantic classes. The syntactic filter reduces the number of assignments under consideration to 113,106 (7.6% of the number of potential assignments) while preserving 67% of the assignments we know to be correct. The various semantic filters in turn reduce the number of assignments further. For example, the broad semantic filter reduced the 113,106 verbs that passed through the syntactic filter down to 6029 assignments, 19% of the number of assignments based on syntax and 0.4% of the potential assignments.

Our goal throughout the acquisition task is to eliminate as many incorrect assignments as possible while preserving the correct assignments, and in this respect we are encouraged by the the behavior of the semantic filter on "unknown" verbs. Recall that to assess this behavior, we excluded randomly selected Levin verbs from the semantic filter, and saw how the filter behaved on these verbs.

Acknowledgements

The research reported herein was supported, in part, by Army Research Office contract DAAL03-91-C-0034 through Battelle Corporation, NSF NYI IRI-9357731, Alfred P. Sloan Research Fellow Award BR3336, and a General Research Board Semester Award. We would like to thank Julie Dahmer, Charles Lin, and David Woodard for their help in annotating the verbs. We would also like to thank Karen Kohl for permission to use her WordNet annotations for Part One of Levin's book as hints for WordNet senses for Part Two.

References

Alshawi, H. 1989. Analysing the Dictionary Definitions. In B. Boguraev and T. Briscoe, editor, *Computational Lexicography for Natural Language Processing*. Longman, London, pages 153–169.

Boguraev, B. and T. Briscoe. 1989. Utilising the LDOCE Grammar Codes. In B. Boguraev and T. Briscoe, editor, *Computational Lexicography for Natural Language Processing*. Longman, London, pages 85–116.

Brent, M. 1993. Unsupervised Learning of Lexical Syntax. *Computational Linguistics*, 19:243–262.

Church, K. and P. Hanks. 1990. Word Association Norms, Mutual Information and Lexicography. *Computational Linguistics*, 16:22–29.

Copestake, A., T. Briscoe, P. Vossen, A. Ageno, I. Castellon, F. Ribas, G. Rigau, H. Rodríguez, and A. Samiotou. 1995. Acquisition of Lexical Translation Relations from MRDS. *Machine Translation*, 9.

Dorr, B., J. Garman, and A. Weinberg. 1995. From Syntactic Encodings to Thematic Roles: Building Lexical Entries for Interlingual MT. *Machine Translation*, 9.

Farwell, D., L. Guthrie, and Y. Wilks. 1993. Automatically Creating Lexical Entries for ULTRA, a Multilingual MT System. *Machine Translation*, 8(3).

Fillmore, C.J. 1968. The Case for Case. In E. Bach and R.T. Harms, editor, *Universals in Linguistic Theory*. Holt, Rinehart, and Winston, pages 1–88.

Grimshaw, J. 1990. *Argument Structure*. MIT Press, Cambridge, MA.

Gruber, J.S. 1965. *Studies in Lexical Relations*. Ph.D. thesis, MIT, Cambridge, MA.

Guthrie, J., L. Guthrie, Y. Wilks, and H. Aidinejad. 1991. Subject-Dependent Co-occurrence and Word Sense Disambiguation. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, pages 146–152, University of California, Berkeley, CA.

Hearst, M. 1991. Noun Homograph Disambiguation Using Local Context in Large Text Corpora. In *Using Corpora*, University of Waterloo, Waterloo, Ontario.

Jackendoff, R. 1983. *Semantics and Cognition*. MIT Press, Cambridge, MA.

Jackendoff, R. 1990. *Semantic Structures*. MIT Press, Cambridge, MA.

Klavans, J.L. and E. Tzoukermann. 1996. Dictionaries and Corpora: Combining Corpus and Machine-readable Dictionary Data for Building Bilingual Lexicons. *Machine Translation*, 10.

Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago, IL.

Lewis, David Dolan. 1992. *Representation and Learning in Information Retrieval*. Ph.D. thesis, University of Massachusetts, Amherst.

Lonsdale, D., T. Mitamura, and E. Nyberg. 1996. Acquisition of Large Lexicons for Practical Knowledge-Based MT. *Machine Translation*, 9.

Miller, G. 1985. WORDNET: A Dictionary Browser. In *Proceedings of the First International Conference on Information in Data*, University of Waterloo Centre for the New OED, Waterloo, Ontario.

Neff, M. and M. McCord. 1990. Acquiring Lexical Data from Machine-Readable Dictionary Resources for Machine Translation. In *Third International Conference on Theoretical and Methodological Issues in Machine Translation of Natural Languages (TMI-90)*, Austin, Texas.

Pinker, S. 1989. *Learnability and Cognition: The Acquisition of Argument Structure*. MIT Press, Cambridge, MA.

Procter, P. 1978. *Longman Dictionary of Contemporary English.* Longman, London.

Sanfilippo, A. and V. Poznanski. 1992. The Acquisition of Lexical Knowledge from Combined Machine-Readable Dictionary Resources. In *Proceedings of the Applied Natural Language Processing Conference*, pages 80–87, Trento, Italy.

Walker, D. and R. Amsler. 1986. The Use of Machine-readable Dictionaries in Sublanguage Analysis. In R. Grishman and R. Kittredge, editors, *Analyzing Language in Restricted Domains.* Lawrence Erlbaum Associates, Hillsdale, New Jersey, pages 69–83.

Wilks, Y., D. Fass, C.M. Guo, J.E. McDonald, and T. Plate. 1990. Providing Machine Tractable Dictionary Tools. *Machine Translation*, 5(2):99–154.

Wilks, Y., D. Fass, C.M. Guo, J.E. McDonald, T. Plate, and B.M. Slator. 1989. A Tractable Machine Dictionary as a Resource for Computational Semantics. In B. Boguraev and T. Briscoe, editor, *Computational Lexicography for Natural Language Processing.* Longman, London, pages 85–116.

Wu, D. and X. Xia. 1995. Large-Scale Automatic Extraction of an English-Chinese Translation Lexicon. *Machine Translation*, 9.

Yarowsky, D. 1992. Word-Sense Disambiguation: Using Statistical Models of Roget's Categories Trained on Large Corpora. In *Proceedings of the Fourteenth International Conference on Computational Linguistics*, pages 454–460, Nantes, France.