Large-Scale Extraction of Fine-Grained Semantic Relations between Verbs

Timothy Chklovski and Patrick Pantel

Information Sciences Institute
University of Southern California
4676 Admiralty Way
Marina del Rey, CA 90292
{timc,pantel}@isi.edu

Abstract. Broad-coverage repositories of semantic relations between actions could benefit many NLP tasks, as well as tasks related to reasoning and inference. We present an automatic method for extracting fine-grained semantic relations, addressing relations between verbs. We detect similarity, strength, antonymy, enablement, and temporal relations between pairs of verbs with high mutual information using lexico-syntactic patterns over the Web. On a set of 26,118 strongly associated verb pairs, our extraction algorithm yielded 56.5% accuracy[1]. On the relations *strength* and *similarity*, we achieved 79.6% and 66.7% accuracy respectively.

1 Introduction

Many tasks, such as question answering, summarization, and machine translation could benefit from broad-coverage semantic resources such as WordNet (Miller 1990) and EVCA (English Verb Classes and Alternations) (Levin 1993). These extremely useful resources have very high precision entries but have important limitations when used in real-world tasks due to their limited coverage and prescriptive nature (i.e. they do not include semantic relations that are plausible but not guaranteed). For example, it may be valuable to know that if someone has bought an item, they may sell it at a later time. WordNet does not include the relation "*X buys Y*" *happens-before* "*X sells Y*" since it is possible to sell something without having bought it (e.g. having manufactured or stolen it).

Verbs are the primary vehicle for describing events and expressing relations between entities. Hence, verb semantics could help in many natural language processing (NLP) tasks that deal with events or relations between entities. For NLP as well as reasoning and inference tasks which require canonicalization of natural language statements or derivation of plausible inferences from such statements, a particularly valuable resource is one which (i) relates verbs to one another and (ii) provides broad coverage of the verbs in the target language.

In this paper, we present an algorithm that automatically discovers fine-grained verb semantics by querying the Web using simple lexico-syntactic patterns. The verb

[1] The relations are available for download at *http://semantics.isi.edu/ocean/*.

relations we discover are similarity, strength, antonymy, enablement, and temporal relations. Our approach extends previously formulated ones that use surface patterns as indicators of semantic relations between nouns (Hearst 1992; Etzioni 2003; Ravichandran and Hovy 2002). We extend these approaches in two ways: (i) our patterns indicate verb conjugation to increase their expressiveness and specificity and (ii) we use a measure similar to mutual information to account for both the frequency of the verbs whose semantic relations are being discovered as well as for the frequency of the pattern.

2 Related Work

In this section, we describe application domains that can benefit from a resource of verb semantics. We then introduce some existing resources and describe previous attempts at mining semantics from text.

2.1 Applications

Question answering is often approached by canonicalizing the question text and the answer text into logical forms. This approach is taken, *inter alia*, by a top-performing system (Moldovan et al. 2002). In discussing future work on the system's logical form matching component, Rus (2002 p. 143) points to incorporating entailment and causation verb relations to improve the matcher's performance. In other work, Webber et al. (2002) have argued that successful question answering depends on lexical reasoning, and that lexical reasoning in turn requires fine-grained verb semantics in addition to troponymy (*is-a* relations between verbs) and antonymy.

In multi-document summarization, knowing verb similarities is useful for sentence compression and for determining sentences that have the same meaning (Lin 1997). Knowing that a particular action happens before another or is enabled by another is also useful to determine the order of the events (Barzilay et al. 2002). For example, to order summary sentences properly, it may be useful to know that selling something can be preceded by either buying, manufacturing, or stealing it. Furthermore, knowing that a particular verb has a meaning stronger than another (e.g. *rape* vs. *abuse* and *renovate* vs. *upgrade*) can help a system pick the most general sentence.

In lexical selection of verbs in machine translation and in work on document classification, practitioners have argued for approaches that depend on wide-coverage resources indicating verb similarity and membership of a verb in a certain class. In work on translating verbs with many counterparts in the target language, Palmer and Wu (1995) discuss inherent limitations of approaches which do not examine a verb's class membership, and put forth an approach based on verb similarity. In document classification, Klavans and Kan (1998) demonstrate that document type is correlated with the presence of many verbs of a certain EVCA class (Levin 1993). In discussing future work, Klavans and Kan point to extending coverage of the manually constructed EVCA resource as a way of improving the performance of the system. A wide-coverage repository of verb relations including verbs linked by the similarity relation will provide a way to automatically extend the existing verb classes to cover more of the English lexicon.

2.2 Existing resources

Some existing broad-coverage resources on verbs have focused on organizing verbs into classes or annotating their frames or thematic roles. EVCA (English Verb Classes and Alternations) (Levin 1993) organizes verbs by similarity and participation / non-participation in alternation patterns. It contains 3200 verbs classified into 191 classes. Additional manually constructed resources include PropBank (Kingsbury et al. 2002), FrameNet (Baker et al. 1998), VerbNet (Kipper et al. 2000), and the resource on verb selectional restrictions developed by Gomez (2001).

Our approach differs from the above in its focus. We relate verbs to each other rather than organize them into classes or identify their frames or thematic roles. WordNet does provide relations between verbs, but at a coarser level. We provide finer-grained relations such as strength, enablement and temporal information. Also, in contrast with WordNet, we cover more than the prescriptive cases.

2.3 Mining semantics from text

Previous web mining work has rarely addressed extracting many different semantic relations from Web-sized corpus. Most work on extracting semantic information from large corpora has largely focused on the extraction of *is-a* relations between nouns. Hearst (1992) was the first followed by recent larger-scale and more fully automated efforts (Pantel and Ravichandran 2004; Etzioni et al. 2004; Ravichandran and Hovy 2002).

Turney (2001) studied word relatedness and synonym extraction, while Lin et al. (2003) present an algorithm that queries the Web using lexical patterns for distinguishing noun synonymy and antonymy. Our approach addresses verbs and provides for a richer and finer-grained set of semantics.

Semantic networks have also been extracted from dictionaries and other machine-readable resources. MindNet (Richardson et al. 1998) extracts a collection of triples of the type “*ducks have wings*” and “*duck capable-of flying*”. This resource, however, does not relate verbs to each other or provide verb semantics.

3 Semantic relations among verbs

In this section, we introduce and motivate the specific relations that we extract. Whilst the natural language literature is rich in theories of semantics (Barwise and Perry 1985; Schank and Abelson 1977), large-coverage manually created semantic resources typically only organize verbs into a flat or shallow hierarchy of classes (such as those described in Section 2.2). WordNet identifies synonymy, antonymy, troponymy, and cause. As summarized in Figure 1, Fellbaum (1998) discusses a finer-grained analysis of entailment, while the WordNet database does not distinguish be-

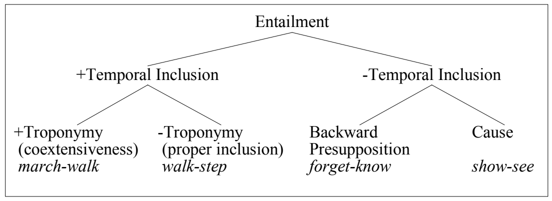

Figure 1. Fellbaum's (1998) entailment hierarchy.

Table 1. Semantic relations we identify. *Siblings* in the WordNet column refers to terms with the same troponymic parent, e.g. *swim* and *fly*.

Semantic Relation	*Example*	*Alignment with WordNet*	*Symmetric*
similarity	transform :: integrate	synonyms or siblings	Y
strength	push :: nudge	synonyms or siblings	N
antonymy	open :: close	antonymy	Y
enablement	wash :: clean	cause	N
happens-before	buy :: have; marry :: divorce	cause; entailment, no temporal inclusion	N
happens-while	chew :: eat snore :: sleep	entailment proper temporal inclusion, no troponymy	N

tween, e.g., proper temporal inclusion (*walk :: step*) from backward presupposition (*forget :: know*). In formulating our set of relations, we have relied on the finer-grained analysis.

In selecting the relations to identify, we aimed at both covering the relations described in WordNet and covering the relations present in our collection of strongly associated verb pairs. We relied on the strongly associated verb pairs, described in Section 4.3, for computational efficiency. The relations we identify were experimentally found to cover 99 out of 100 randomly selected verb pairs.

Our algorithm identifies six semantic relations between verbs. These are summarized in Table 1 along with their closest corresponding WordNet category and the symmetry of the relation (whether V_1 *rel* V_2 is equivalent to V_2 *rel* V_1).

Similarity. As Fellbaum (1998) and the tradition of organizing verbs into similarity classes indicate, verbs do not neatly fit into a unified *is-a* (troponymy) hierarchy. Rather, verbs are often similar or related. Similarity between action verbs, for example, can arise when they differ in connotations about manner or degree of action. Examples extracted by our system include *maximize :: enhance*, *produce :: create*, *reduce :: restrict*.

Strength. When two verbs are similar, one may denote a more intense, thorough, comprehensive or absolute action. In the case of change-of-state verbs, one may denote a more complete change. We identify this as the *strength* relation. Sample verb pairs extracted by our system, in the order *weak :: strong*, are: *taint :: poison, permit :: authorize, surprise :: startle, startle :: shock.*

This subclass of similarity has not been identified in broad-coverage networks of verbs, but may be of particular use in natural language generation and summarization applications.

Antonymy. Also known as semantic opposition, antonymy between verbs has several distinct subtypes. As discussed by Fellbaum (1998), it can arise from switching thematic roles associated with the verb (as in *buy :: sell, lend :: borrow*). There is also antonymy between stative verbs (*live :: die, differ :: equal*) and antonymy between sibling verbs which share a parent (*walk :: run*) or an entailed verb *(fail* :: *succeed* both entail *try*).

Antonymy also systematically interacts with the *happens-before* relation in the case of restitutive opposition (Cruse 1986). This subtype is exemplified by *damage :: repair*, *wrap :: unwrap*. In terms of the relations we recognize, it can be stated that *restitutive-opposition*(V_1, V_2) = *happens-before*(V_1, V_2), and *antonym*(V_1, V_2). Examples of antonymy extracted by our system include: *assemble :: dismantle; ban :: allow; regard :: condemn, roast :: fry*.

Enablement. This relation holds between two verbs V_1 and V_2 when the pair can be glossed as V_1 *is accomplished by* V_2. Enablement is classified as a type of causal relation by Barker and Szpakowicz (1995). Examples of enablement extracted by our system include: *assess :: review* and *accomplish :: complete*.

Happens-before. This relation indicates that the two verbs refer to two temporally disjoint intervals or instances. WordNet's *cause* relation, between a causative and a resultative verb (as in *buy :: own*), would be tagged as instances of *happens-before* by our system. Examples of the *happens-before* relation identified by our system include *marry :: divorce*, *detain :: prosecute, enroll :: graduate*, *schedule :: reschedule*, *tie :: untie*.

Happens-while. This relation indicates proper temporal inclusion, either of a repeating activity (*chew :: eat*) or an event (*find :: study*). In some cases also classified as *happens-while*, it may be difficult to say if the temporal inclusion is necessarily strict, as in *say :: announce*.

4 Approach

We discover the semantic relations described above by querying the Web with Google for lexico-syntactic patterns indicative of each relation. Our approach has two stages. First, we identify pairs of highly associated verbs co-occurring on the Web in sufficient volume. These pairs are extracted using previous work by Lin and Pantel (2001), as described in Section 4.3. Next, for each verb pair, we tested lexico-syntactic patterns, outputting the first detected relation[2].

[2] In effect, we are making the simplifying assumption that at most one relation needs to be detected. This assumption may be relaxed in future work.

Table 2. Semantic relations and samples of the 33 surface patterns used to identify them. In patterns, "*" matches any single word. Punctuation does not count as words by the search engine used (Google). Relations are shown in the order of testing.

SEMANTIC RELATION	*Surface Patterns*	$Hits_{est}$ *for patterns*
happens-while	to X while Ying; Xed while Ying	6,752,541
strength	X and even Y; Yed or at least Xed	2,172,811
happens-before	Xed * and then Yed; to X and eventually Y	4,074,935
enablement	Xed * by Ying the; to X * by Ying or	2,348,392
antonymy	to X * but Y; Xed * * but Yed	18,040,916
nonequivalence-but-similarity*	both Xed and Yed; X rather than Y	1,777,755
broad similarity*	Xed and Yed; Xs and Ys; to X and Y	174,797,897

*nonequivalence-but-similarity and broad-similarity were later combined into a single category, *similarity*, and are treated as a single category in the rest of our discussion.

4.1 Lexico-syntactic patterns

The lexico-syntactic patterns were manually selected by examining pairs of verbs in known semantic relations. They were refined to decrease capturing wrong parts of speech or incorrect semantic relations.

Although many patterns may indicate the relations, we use a total of 33 patterns. Some representatives are shown in Table 2. Note that our patterns specify the tense of the verbs they accept. When instantiating these patterns, we conjugate as needed. For example, "*both Xed and Yed*" instantiates on *sing* and *dance* as "*both sung and danced*".

4.2 Testing for a semantic relation

In this section, we describe how the presence of a semantic relation is detected. We test the relations in the order specified in Table 2. We adopt an approach inspired by mutual information to measure the strength of association, denoted $S_p(V_1, V_2)$, between three entities: a verb pair V_1 and V_2 and a lexico-syntactic pattern p:

$$S_p(V_1, V_2) = \frac{P(V_1, p, V_2)}{P(p) \times P(V_1) \times P(V_2)} \quad \textbf{(1)}$$

The probabilities in the denominator are difficult to calculate directly from search engine results. For a given lexico-syntactic pattern, we need to estimate the frequency of the pattern instantiated with appropriately conjugated verbs. For verbs, we need to estimate the frequency of the verbs, but avoid counting other parts-of-speech (e.g. *chair* as a noun or *painted* as an adjective). Another issue is that some relations are symmetric (we treat *similarity* and *antonymy* as symmetric), while others are not (*strength*, *enablement*, *happens-while*, *happens-before*). For symmetric relations only, the verbs can fill the lexico-syntactic pattern in either order. To address these issues, we estimate $S_p(V_1, V_2)$ using:

$$S_P(V_1,V_2) \approx \frac{\dfrac{hits(V_1, p, V_2)}{N}}{\dfrac{hits_{est}(p)}{N} \times \dfrac{hits(\text{"}to\ V_1\text{"}) \times C_v}{N} \times \dfrac{hits(\text{"}to\ V_2\text{"}) \times C_v}{N}} \qquad \textbf{(2)}$$

for asymmetric relations and

$$S_P(V_1,V_2) \approx \frac{\dfrac{hits(V_1, p, V_2)}{N} + \dfrac{hits(V_2, p, V_1)}{N}}{\dfrac{2 * hits_{est}(p)}{N} \times \dfrac{hits(\text{"}to\ V_1\text{"}) \times C_v}{N} \times \dfrac{hits(\text{"}to\ V_2\text{"}) \times C_v}{N}} \qquad \textbf{(3)}$$

for symmetric relations.

Here, *hits(S)* denotes the number of documents containing the string *S*, as returned by Google. *N* is the number of words indexed by the search engine ($N \approx 7.2 \times 10^{11}$), C_v is a correction factor to obtain the frequency of the verb V in all tenses from the frequency of the pattern "to V". Based on several verbs, we have estimated $C_v = 8.5$. Because pattern counts, when instantiated with verbs, could not be estimated directly, we have computed the frequencies of the patterns in a part-of-speech tagged 500M word corpus and used it to estimate the expected number of hits $hits_{est}(p)$ for each pattern. We estimated the *N* with a similar method.

We say that the semantic relation indicated by lexico-syntactic patterns *p* is present between V_1 and V_2 if

$$\sum_p S_p(V_1, V_2) > C_1 \qquad \textbf{(4)}$$

As a result of tuning the system, $C_1 = 8.5$.

Additional test for asymmetric relations. For the asymmetric relations, we require not only that $\sum_p S_P(V_1, V_2)$ exceed a certain threshold, but that there be strong asymmetry of the relation:

$$\frac{\sum_p S_p(V_1, V_2)}{\sum_p S_p(V_1, V_2)} = \frac{\sum_p hits(V_1, p, V_2)}{\sum_p hits(V_2, p, V_1)} > C_2 \qquad \textbf{(5)}$$

Tuning on 50 verb pairs has yielded $C_2 = 7$.

4.3 Extracting highly associated verb pairs

To exhaustively test the more than 64 million unordered verb pairs for WordNet's more than 11,000 verbs would be computationally intractable. Instead, we use a set of highly associated verb pairs output by a paraphrasing algorithm called DIRT. Since we are able to test up to 4000 verb pairs per day on a single machine (we issue at most

40 queries per test and each query takes approximately 0.5 seconds), we are able to test several dozen associated verbs for each verb in WordNet in a matter of weeks.

Lin and Pantel (2001) describe an algorithm called DIRT (Discovery of Inference Rules from Text) that automatically learns paraphrase expressions from text. It is a generalization of previous algorithms that use the distributional hypothesis (Harris 1985) for finding similar words. Instead of applying the hypothesis to words, Lin and Pantel applied it to paths in dependency trees. Essentially, if two paths tend to link the same sets of words, they hypothesized that the meanings of the corresponding paths are similar. It is from paths of the form *subject-verb-object* that we extract our set of associated verb pairs. Hence, this paper is concerned only with relations between transitive verbs.

A path, extracted from a parse tree, is an expression that represents a binary relation between two nouns. A set of paraphrases was generated for each pair of associated paths. For example, using a 1.5GB newspaper corpus, here are the 20 most associated paths to "*X solves Y*" generated by DIRT:

```
Y is solved by X, X resolves Y, X finds a solution to Y, X
tries to solve Y, X deals with Y, Y is resolved by X, X ad-
dresses Y, X seeks a solution to Y, X does something about
Y, X solution to Y, Y is resolved in X, Y is solved through
X, X rectifies Y, X copes with Y, X overcomes Y, X eases Y,
X tackles Y, X alleviates Y, X corrects Y, X is a solution
to Y, X makes Y worse, X irons out Y
```

DIRT only outputs pairs of paths that it has syntactic evidence of being in some semantic relation. We used these as our set to extract finer-grained relations.

5 Experimental results

In this section, we empirically evaluate the accuracy of our system.

5.1 Experimental setup

We studied 26,118 pairs of verbs. Applying DIRT to a 1.5GB newspaper corpus[3], we extracted 4000 paths that consisted of single verbs in the relation *subject-verb-object* (i.e. paths of the form "*X verb Y*") whose verbs occurred in at least 150 documents on the Web. For example, from the 20 most associated paths to "*X solves Y*" shown in Section 4.3, the following verb pairs were extracted:

```
solves :: resolves
solves :: addresses
solves :: rectifies
solves :: overcomes
solves :: eases
solves :: tackles
solves :: corrects
```

[3] The 1.5GB corpus consists of San Jose Mercury, Wall Street Journal and AP Newswire articles from the TREC-9 collection.

Table 3. First five randomly selected pairs along with the system tag (in bold) and the judges' responses.

PAIRS WITH SYSTEM TAG (IN BOLD)	CORRECT JUDGE 1	CORRECT JUDGE 2	PREFERRED SEMANTIC RELATION JUDGE 1	PREFERRED SEMANTIC RELATION JUDGE 2
X rape Y **is stronger than** X abuse Y	Yes	Yes	is stronger than	is stronger than
X accomplish Y **is enabled by** X complete Y	Yes	Yes	is accomplished by	is accomplished by
X achieve Y **is enabled by** X boost Y	Yes	Yes	is accomplished by	is accomplished by
X annotate Y **is similar to** X translate Y	No	Yes	has no relation with	is an alternative to
X further Y **is stronger than** X attain Y	No	No	happens before	happens before

Table 4. Accuracy of system-discovered relations.

	ACCURACY *Tags Correct*	ACCURACY *Preferred Tags Correct*	ACCURACY *Baseline Correct*
Judge 1	56%	52%	26%
Judge 2	57%	44%	33%
Average	56.5%	48%	29.5%

5.2 Accuracy

To evaluate the accuracy of the system, we ran it on 100 randomly selected pairs and classified each according to the semantic relations described in Section 3. We presented the classifications to two human judges. The adjudicators were asked to judge whether or not the system classification was acceptable. Since the semantic relations are not disjoint (e.g. *mop* is both stronger than and similar to *sweep*), multiple relations may be appropriately acceptable for a given verb pair. The judges were also asked to identify their preferred semantic relations (i.e. that relation which seems most plausible). Table 3 shows the first five randomly selected pairs along with the judges' responses.

Table 4 shows the accuracy of the system. The baseline system consists of labeling each pair with the most common semantic relation, *similarity*, which occurs 29 times. The Kappa statistic (Siegel and Castellan 1988) for the task of judging system tags as correct and incorrect is $\kappa = 0.74$ whereas the task of identifying the preferred semantic relation has $\kappa = 0.697$. For the latter task, the two judges agreed on 72 of the 100 semantic relations. 72% gives an idea of an upper bound for humans on this task. Of these 72 relations, the system achieved a higher accuracy of 61.1%.

Table 5 shows the accuracy of the system on each of the relations. The system did particularly well on the *strength* and *similarity* relations. However, the *happens-while* relation was hardly exercised. Only one of the five instances that the system tagged as a *happens-while* relation was judged correct and by only one of the two judges. Also, 35% of the errors the system made on the *no relation* tag were *antonymy* and 23%

Table 5. Accuracy of each semantic relation.

SEMANTIC RELATION	SYSTEM TAGS	*Tags Correct*	*Preferred Tags Correct*
similarity	36	66.7%	58.3%
strength	22	79.6%	65.9
antonymy	4	25.0%	25.0%
enablement	7	57.2%	50.0
happens before	7	28.6%	28.6
happens while	5	10.0%	0%
no relation	19	39.5%	31.6%

were *similarity*. This suggests that other patterns are needed to discover these two relations.

As described in Section 3, WordNet contains verb semantic relations. A significant percentage of our discovered relations are not covered by WordNet's coarser classifications. Of the 50 verb pairs whose system relation was tagged as correct by both judges in our accuracy experiments, only 34% of them existed in a WordNet relation.

5.3 Discussion

The experience of extracting these syntactic relations has clarified certain important challenges.

While relying on a search engine allows us to query a corpus of nearly a trillion words, some issues arise: (i) the number of instances has to be approximated by the number of hits (documents); (ii) the number of hits for the same query may fluctuate over time; and (iii) some needed counts are not directly available. We addressed the latter issue by approximating these counts using a smaller corpus.

We do not detect entailment with lexico-syntactic patterns. In fact, we propose that whether the entailment relation holds between V_1 and V_2 depends on the absence of another verb V_1' in the same relationship with V_2. For example, given the relation *marry happens-before divorce*, we can conclude that *divorce* entails *marry*. But, given the relation *buy happens-before sell*, we cannot conclude entailment since *manufacture* can also happen before *sell*. This also applies to the relations *happens-while* and *enablement*.

Corpus-based methods, including ours, hold the promise of wide coverage but are weak on discriminating senses.

6 Future work

There are several ways to improve the accuracy of the current algorithm and to detect relations between low volume verb pairs. One avenue would be to automatically learn or manually craft more patterns and to extend the pattern vocabulary (when developing the system, we have noticed that different registers and verb types require different patterns). Another possibility would be to use more relaxed patterns when the part of speech confusion is not likely (e.g. "eat" is a common verb which does not have a noun sense, and patterns need not protect against noun senses when testing such verbs).

Our approach can potentially be extended to multiword paths. DIRT actually provides two orders of magnitude more relations than the 26,118 single verb relations (subject-verb-object) we extracted. On the same 1GB corpus described in Section 5.1, DIRT extracted over 200K paths and 6M unique paraphrases. These provide an opportunity to create a much larger corpus of semantic relations, or to construct smaller, in-depth resources for selected subdomains. For example, we could extract that *take a trip to* is similar to *travel to*, and that *board a plane* happens before *deplane*.

Finally, as discussed in Section 5.3, entailment relations can be derived by processing the complete graph of the identified semantic relation.

7 Conclusion

We have demonstrated that certain fine-grained semantic relations between verbs are present on the web, and are extractable with a simple pattern-based approach[4]. In addition to discovering relations identified in WordNet, such as opposition and troponymy, we obtain strong results on enablement and strength relations (for which no wide-coverage resource is available). On a set of 26,118 associated verb pairs, experimental results show an accuracy of 56.5% in assigning similarity, strength, antonymy, enablement, and temporal relations.

Further work may refine extraction methods and further process the mined semantics to derive other relations such as entailment.

We hope to open the way to inferring implied, but not stated assertions and to benefit applications such as question answering, information retrieval, and summarization.

References

1. Baker, C.; Fillmore, C.; and Lowe, J. 1998. The Berkeley FrameNet project. In *Proceedings of COLING-ACL*. Montreal, Canada.
2. Barker, K.; and Szpakowicz, S. 1995. Interactive Semantic Analysis of Clause-Level Relationships. In *Proceedings of PACLING `95*. Brisbane.
3. Barwise, J. and Perry, J. 1985. Semantic innocence and uncompromising situations. In: Martinich, A. P. (ed.) *The Philosophy of Language*. New York: Oxford University Press. pp. 401–413.
4. Barzilay, R.; Elhadad, N.; and McKeown, K. 2002. Inferring strategies for sentence ordering in multidocument summarization. *JAIR*, 17:35–55.

[4] We plan to provide an online resource of verb semantics discovered from this work if accepted.

5. Cruse, D. 1992 Antonymy Revisited: Some Thoughts on the Relationship between Words and Concepts, in A. Lehrer and E.V. Kittay (eds.), Frames, Fields, and Contrasts, Hillsdale, NJ, Lawrence Erlbaum associates, pp. 289-306.
6. Etzioni, O.; Cafarella, M.; Downey, D.; Kok, S.; Popescu, A.; Shaked, T.; Soderland, S.; Weld, D.; and Yates, A. 2004. Web-scale information extraction in KnowItAll. To appear in *WWW-2004*.
7. Fellbaum, C. 1998. Semantic network of English verbs. In Fellbaum, (ed). *WordNet: An Electronic Lexical Database*, MIT Press.
8. Gomez, F. 2001. An Algorithm for Aspects of Semantic Interpretation Using an Enhanced WordNet. In *NAACL-2001*, CMU, Pittsburgh.
9. Harris, Z. 1985. Distributional Structure. In: Katz, J. J. (ed.), *The Philosophy of Linguistics*. New York: Oxford University Press. pp. 26–47.
10. Hearst, M. 1992. Automatic acquisition of hyponyms from large text corpora. In *COLING-92*. pp. 539–545. Nantes, France.
11. Kingsbury, P; Palmer, M.; and Marcus, M. 2002. Adding semantic annotation to the Penn TreeBank. In *Proceedings of HLT-2002*. San Diego, California.
12. Kipper, K.; Dang, H.; and Palmer, M. 2000. Class-based construction of a verb lexicon. In *Proceedings of AAAI-2000*. Austin, TX.
13. Klavans, J. and Kan, M. 1998. Document classification: Role of verb in document analysis. In *Proceedings COLING-ACL '98*. Montreal, Canada.
14. Levin, B. 1993. *English Verb Classes and Alternations: A Preliminary Investigation*. University of Chicago Press, Chicago, IL.
15. Lin, C-Y. 1997. Robust Automated Topic Identification. Ph.D. Thesis. University of Southern California.
16. Lin, D. and Pantel, P. 2001. Discovery of inference rules for question answering. Natural Language Engineering, 7(4):343–360.
17. Lin, D.; Zhao, S.; Qin, L.; and Zhou, M. 2003. Identifying synonyms among distributionally similar words. In *Proceedings of IJCAI-03*. pp.1492–1493. Acapulco, Mexico.
18. Miller, G. 1990. WordNet: An online lexical database. *International Journal of Lexicography*, 3(4).
19. Moldovan, D.; Harabagiu, S.; Girju, R.; Morarescu, P.; Lacatusu, F.; Novischi, A.; Badulescu, A.; and Bolohan, O. 2002. LCC tools for question answering. In *Notebook of the Eleventh Text REtrieval Conference (TREC-2002)*. pp. 144–154.
20. Palmer, M., Wu, Z. 1995. Verb Semantics for English-Chinese Translation Machine Translation, 9(4).
21. Pantel, P. and Ravichandran, D. 2004. Automatically labeling semantic classes. To appear in *Proceedings of HLT/NAACL-2004*. Boston, MA.
22. Ravichandran, D. and Hovy, E., 2002. Learning surface text patterns for a question answering system. In *Proceedings of ACL-02*. Philadelphia, PA.
23. Richardson, S.; Dolan, W.; and Vanderwende, L. 1998. MindNet: acquiring and structuring semantic information from text. In *Proceedings of COLING '98*.
24. Rus, V. 2002. Logic Forms for WordNet Glosses. Ph.D. Thesis. Southern Methodist University.
25. Schank, R. and Abelson, R. 1977. *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures.* Lawrence Erlbaum Associates.
26. Siegel, S. and Castellan Jr., N. 1988. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill.
27. Turney, P. 2001. Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. In *Proceedings ECML-2001*. Freiburg, Germany.
28. Webber, B.; Gardent, C.; and Bos, J. 2002. Position statement: Inference in question answering. In *Proceedings of LREC-2002*. Las Palmas, Spain.