VERBOCEAN: Mining the Web for Fine-Grained Semantic Verb Relations

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 verbs could benefit many NLP tasks. We present a semi-automatic method for extracting fine-grained semantic relations between verbs. We detect similarity, strength, antonymy, enablement, and temporal happens-before relations between pairs of strongly associated verbs using lexico-syntactic patterns over the Web. On a set of 29,165 strongly associated verb pairs, our extraction algorithm yielded 65.5% accuracy. Analysis of error types shows that on the relation *strength* we achieved 75% accuracy. We provide the resource, called VERBOCEAN, for download at *http://semantics.isi.edu/ocean/*.

1 Introduction

Many NLP 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 NLP 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 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 semi-automatically discovers fine-grained verb semantics by querying the Web using simple lexico-syntactic patterns. The verb relations we discover are similarity, strength, antonymy, enablement, and temporal relations. Identifying these relations over 29,165 verb pairs results in a broad-coverage resource we call VERBOCEAN. 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 Relevant 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

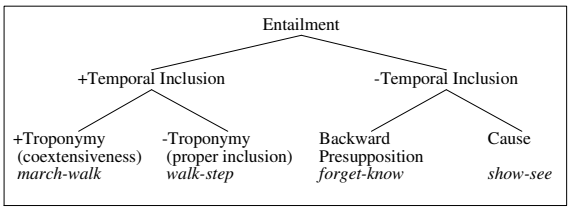

Figure 1. Fellbaum's (1998) entailment hierarchy.

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 / nonparticipation 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). Recently, Moldovan et al. (2004) present a learning algorithm to detect 35 fine-grained noun phrase relations.

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. Reliability of estimating bigram counts on the web via search engines has been investigated by Keller and Lapata (2003).

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 between, e.g., backward presupposition (*forget :: know*, where *know* must have happened before *forget*) from proper temporal inclusion (*walk :: step*). In formulating our set of relations, we have relied on the finer-grained analysis, explicitly breaking out the temporal precedence between entities.

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 4.4, 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 to strong, are: *taint :: poison, permit :: authorize, surprise :: startle, startle :: shock.* Some instances of strength sometimes map to WordNet's troponymy relation.

Strength, a 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*.

Table 1. Semantic relations identified in VERBOCEAN. *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	wound :: kill	synonyms or siblings	N
antonymy	open :: close	antonymy	Y
enablement	fight :: win	cause	N
happens-before	buy :: sell; marry :: divorce	cause entailment, no temporal inclusion	N

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*.

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 with sufficient frequency using previous work by Lin and Pantel (2001), as described in Section 4.4. Next, for each verb pair, we tested lexico-syntactic patterns, calculating a score for each possible semantic relation as described in Section 4.2. Finally, as described in Section 4.3, we compare the strengths of the individual semantic relations and, preferring the most specific and then strongest relations, output a consistent set as the final output. As a guide to consistency, we use a simple theory of semantics indicating which semantic relations are subtypes of other ones, and which are compatible and which are mutually exclusive.

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. We used 50 verb pairs and the overall process took about 25 hours.

We use a total of 35 patterns, which are listed in Table 2 along with the estimated frequency of hits.

Table 2. Semantic relations and the 35 surface patterns used to identify them. Total number of patterns for that relation is shown in parentheses. In patterns, "*" matches any single word. Punctuation does not count as words by the search engine used (Google).

SEMANTIC RELATION	*Surface Patterns*	*$Hits_{est}$ for patterns*
narrow similarity (2)*	X ie Y Xed ie Yed	219,480
broad similarity (2)*	Xed and Yed to X and Y	154,518,326
strength (8)	X even Y Xed even Yed X and even Y Xed and even Yed Y or at least X Yed or at least Xed not only Xed but Yed not just Xed but Yed	1,016,905
enablement (4)	Xed * by Ying the Xed * by Ying or to X * by Ying the to X * by Ying or	2,348,392
antonymy (7)	either X or Y either Xs or Ys either Xed or Yed either Xing or Ying whether to X or Y Xed * but Yed to X * but Y	18,040,916
happens-before (12)	to X and then Y to X * and then Y Xed and then Yed Xed * and then Yed to X and later Y Xed and later Yed to X and subsequently Y Xed and subsequently Yed to X and eventually Y Xed and eventually Yed	8,288,871

*narrow- and broad- similarity overlap in their coverage and are treated as a single category, *similarity*, when postprocessed. Narrow similarity tests for rare patterns and $hits_{est}$ for it had to be approximated rather than estimated from the smaller corpus.

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 with patterns exemplified 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:

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

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 (*similarity* and *antonymy*), while others are not (*strength*, *enablement*, *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:

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

for asymmetric relations and

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

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

$$\sum_p S_p(V_1,V_2) > C_1$$

As a result of tuning the system on a tuning set of 50 verb pairs, $C_1 = 8.5$.

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

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

From the tuning set, $C_2 = 5$.

4.3 Pruning identified semantic relations

Given a pair of semantic relations from the set we identify, one of three cases can arise: (i) one relation is more specific (*strength* is more specific than *similarity*, *enablement* is more specific than *happens-before*), (ii) the relations are compatible (*antonymy* and *happens-before*), where presence of one does not imply or rule out presence of the other, and (iii) the relations are incompatible (*similarity* and *antonymy*).

It is not uncommon for our algorithm to identify presence of several relations, with different strengths. To produce the most likely output, we use semantics of compatibility of the relations to output the most likely one(s). The rules are as follows:

If the frequency was too low (less than 10 on the pattern "X * Y" OR "Y * X" OR "X * * Y" OR "Y * * X"), output that the statements are unrelated and stop.

If *happens-before* is detected, output presence of *happens-before* (additional relation may still be output, if detected).

If *happens-before* is not detected, ignore detection of *enablement* (because *enablement* is more specific than *happens-before*, but is sometimes falsely detected in the absence of *happens-before*).

If *strength* is detected, score of *similarity* is ignored (because *strength* is more specific than *similarity*).

Of the relations *strength, similarity, opposition and enablement* which were detected (and not ignored), output the one with highest S_p.

If nothing has been output to this point, output unrelated.

4.4 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 (Lin and Pantel 2001). 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
```

This list of associated paths looks tantalizingly close to the kind of axioms that would prove useful in an inference system. However, DIRT only outputs pairs of paths that have 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 VERBOCEAN[1].

5.1 Experimental setup

We studied 29,165 pairs of verbs. Applying DIRT to a 1.5GB newspaper corpus[2], 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 4.4, the following verb pairs were extracted:

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

[1] VERBOCEAN is available for download at http://semantics.isi.edu/ocean/.

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

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

Pairs with System Tag (in bold)	*Correct*		*Preferred Semantic Relation*	
	Judge 1	*Judge 2*	*Judge 1*	*Judge 2*
X absolve Y **is similar to** X vindicate Y	Yes	Yes	is similar to	is similar to
X bottom Y **has no relation with** X abate Y	Yes	Yes	has no relation with	has no relation with
X outrage Y **happens-after / is stronger than** X shock Y	Yes	Yes	happens-before / is stronger than	happens-before/ is stronger than
X pool Y **has no relation with** X increase Y	Yes	No	has no relation with	can result in
X insure Y **is similar to** X expedite Y	No	No	has no relation with	has no relation with

Table 4. Accuracy of system-discovered relations.

	Accuracy		
	Tags Correct	*Preferred Tags Correct*	*Baseline Correct*
Judge 1	66%	54%	24%
Judge 2	65%	52%	20%
Average	65.5%	53%	22%

Table 5. Accuracy of each semantic relation.

Semantic Relation	*System Tags*	*Tags Correct*	*Preferred Tags Correct*
similarity	41	63.4%	40.2%
strength	14	75.0%	75.0%
antonymy	8	50.0%	43.8%
enablement	2	100%	100%
no relation	35	72.9%	72.9%
happens before	17	67.6%	55.9%

5.2 Accuracy

We classified each verb pair according to the semantic relations described in Section 2. If the system does not identify any semantic relation for a verb pair, then the system tags the pair as having *no relation*. To evaluate the accuracy of the system, we randomly sampled 100 of these verb pairs, and presented the classifications to two human judges. The adjudicators were asked to judge whether or not the system classification was acceptable (i.e. whether or not the relations output by the system were correct). 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. those relations which seem most plausible). Table 3 shows five randomly selected pairs along with the judges' responses. The Appendix shows sample relationships discovered by the system.

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 33 times. The *Tags Correct* column represents the percentage of verb pairs whose system output relations were deemed correct. The *Preferred Tags Correct* column gives the percentage of verb pairs whose system output relations matched exactly the human's preferred relations. The Kappa statistic (Siegel and Castellan 1988) for the task of judging system tags as correct and incorrect is $\kappa = 0.78$ whereas the task of identifying the preferred semantic relation has $\kappa = 0.72$. For the latter task, the two judges agreed on 73 of the 100 semantic relations. 73% gives an idea of an upper bound for humans on this task. On these 73 relations, the system achieved a higher accuracy of 70.0%. The system is allowed to output the *happens-before* relation in combination with other relations. On the 17 *happens-before* relations output by the system, 67.6% were judged correct. Ignoring the *happens-before* relations, we achieved a *Tags Correct* precision of 68%.

Table 5 shows the accuracy of the system on each of the relations. The *stronger-than* relation is a subset of the *similarity* relation. Considering a coarser extraction where *stronger-than* relations are merged with *similarity*, the task of judging system tags and the task of identifying the preferred semantic relation both jump to 68.2% accuracy. Also, the overall accuracy of the system climbs to 68.5%.

As described in Section 2, WordNet contains verb semantic relations. A significant percentage of our discovered relations are not covered by WordNet's coarser classifications. Of the 40 verb pairs whose system relation was tagged as correct by both judges in our accuracy experiments and whose tag was not '*no relation*', only 22.5% of them existed in a WordNet relation.

5.3 Discussion

The experience of extracting these semantic 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 *enablement* and *strength* relations.

Corpus-based methods, including ours, hold the promise of wide coverage but are weak on discriminating senses. While we hope that applications will benefit from this resource as is, an interesting next step would be to augment it with sense information.

6 Future work

There are several ways to improve the accuracy of the current algorithm and to detect relations between low frequency 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 29,165 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*.

If the entire database is viewed as a graph, we currently leverage and enforce only local consistency. It would be useful to enforce global consistency, e.g. V_1 *stronger-than* V_2, and V_2 *stronger-than* V_3 indicates that V_1 *stronger-than* V_3, which may be leveraged to identify additional relations or inconsistent relations (e.g. V_3 *stronger-than* V_1).

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

7 Conclusions

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. In addition to discovering relations identified in WordNet, such as opposition and enablement, we obtain strong results on strength relations (for which no wide-coverage resource is available). On a set of 29,165 associated verb pairs, experimental results show an accuracy of 65.5% in assigning similarity, strength, antonymy, enablement, and happens-before.

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.

Acknowledgments

The authors wish to thank the reviewers for their helpful comments and Google Inc. for supporting high volume querying of their index. This research was partly supported by NSF grant #EIA-0205111.

References

Baker, C.; Fillmore, C.; and Lowe, J. 1998. The Berkeley FrameNet project. In *Proceedings of COLING-ACL*. Montreal, Canada.

Barker, K.; and Szpakowicz, S. 1995. Interactive Semantic Analysis of Clause-Level Relationships. In *Proceedings of PACLING `95*. Brisbane.

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.

Barzilay, R.; Elhadad, N.; and McKeown, K. 2002. Inferring strategies for sentence ordering in multidocument summarization. *JAIR*, 17:35–55.

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.

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*.

Appendix. Sample relations extracted by our system.

Semantic Relation	*Examples*	*Semantic Relation*	*Examples*	*Semantic Relation*	*Examples*
similarity	maximize :: enhance produce :: create reduce :: restrict	enablement	assess :: review accomplish :: complete double-click :: click	happens before	detain :: prosecute enroll :: graduate schedule :: reschedule
strength	permit :: authorize surprise :: startle startle :: shock	antonymy	assemble :: dismantle regard :: condemn roast :: fry		

Fellbaum, C. 1998. Semantic network of English verbs. In Fellbaum, (ed). *WordNet: An Electronic Lexical Database*, MIT Press.

Gomez, F. 2001. An Algorithm for Aspects of Semantic Interpretation Using an Enhanced WordNet. In *NAACL-2001*, CMU, Pittsburgh.

Harris, Z. 1985. Distributional Structure. In: Katz, J. J. (ed.), *The Philosophy of Linguistics*. New York: Oxford University Press. pp. 26–47.

Hearst, M. 1992. Automatic acquisition of hyponyms from large text corpora. In *COLING-92*. pp. 539–545. Nantes, France.

Keller, F.and Lapata, M. 2003. Using the Web to Obtain Frequencies for Unseen Bigrams. *Computational Linguistics* 29:3.

Kingsbury, P; Palmer, M.; and Marcus, M. 2002. Adding semantic annotation to the Penn TreeBank. In *Proceedings of HLT-2002*. San Diego, California.

Kipper, K.; Dang, H.; and Palmer, M. 2000. Class-based construction of a verb lexicon. In *Proceedings of AAAI-2000*. Austin, TX.

Klavans, J. and Kan, M. 1998. Document classification: Role of verb in document analysis. In *Proceedings COLING-ACL '98*. Montreal, Canada.

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

Lin, C-Y. 1997. Robust Automated Topic Identification. Ph.D. Thesis. University of Southern California.

Lin, D. and Pantel, P. 2001. Discovery of inference rules for question answering. Natural Language Engineering, 7(4):343–360.

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.

Miller, G. 1990. WordNet: An online lexical database. *International Journal of Lexicography*, 3(4).

Moldovan, D.; Badulescu, A.; Tatu, M.; Antohe, D.; and Girju, R. 2004. Models for the semantic classification of noun phrases. To appear in *Proceedings of HLT/NAACL-2004 Workshop on Computational Lexical Semantics*.

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.

Palmer, M., Wu, Z. 1995. Verb Semantics for English-Chinese Translation Machine Translation, 9(4).

Pantel, P. and Ravichandran, D. 2004. Automatically labeling semantic classes. To appear in *Proceedings of HLT/NAACL-2004*. Boston, MA.

Ravichandran, D. and Hovy, E., 2002. Learning surface text patterns for a question answering system. In *Proceedings of ACL-02*. Philadelphia, PA.

Richardson, S.; Dolan, W.; and Vanderwende, L. 1998. MindNet: acquiring and structuring semantic information from text. In *Proceedings of COLING '98*.

Rus, V. 2002. Logic Forms for WordNet Glosses. Ph.D. Thesis. Southern Methodist University.

Schank, R. and Abelson, R. 1977. *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures.* Lawrence Erlbaum Associates.

Siegel, S. and Castellan Jr., N. 1988. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill.

Turney, P. 2001. Mining the Web for synonyms: PMI-IR versus LSA on TOEFL. In *Proceedings ECML-2001*. Freiburg, Germany.

Webber, B.; Gardent, C.; and Bos, J. 2002. Position statement: Inference in question answering. In *Proceedings of LREC-2002*. Las Palmas, Spain.