A Unified Knowledge Based Approach for Sense Disambiguation and Semantic Role Labeling

Peter Z. Yeh, Bruce Porter, and Ken Barker
Department of Computer Sciences
The University of Texas at Austin
Austin, TX 78712
{pzyeh,porter,kbarker}@cs.utexas.edu

Abstract

In this paper, we present a unified knowledge based approach for sense disambiguation and semantic role labeling. Our approach performs both tasks through a single algorithm that matches candidate semantic interpretations to background knowledge to select the best matching candidate. We evaluate our approach on a corpus of sentences collected from various domains and show how our approach performs well on both sense disambiguation and semantic role labeling.

Introduction

One of the goals of AI is to build Natural Language Understanding (NLU) systems that can produce rich semantic representations for tasks such as question answering, machine translation, and information extraction.

A key requirement of NLU systems is mapping the syntactic output from a parser to the corresponding semantic representation. To produce, for example, the semantic representation (see Figure 1 right) for a sentence like:

"The man's hand hit the top of the table."

most NLU systems first produce a syntactic representation that captures the syntactic relationships (e.g. subject, direct object, etc.) between the atomic constituents (i.e. nouns, verbs, adjectives, and adverbs) in the sentence (see Figure 1 left). These systems then select from an ontology the most appropriate concept and semantic relation for each constituent and syntactic relationship respectively. For example, *Region* may be the most appropriate concept for "top" (given the context), and *instrument* may be the most appropriate semantic relation for "subject".

In this paper, we report on our approach for selecting the appropriate concepts and semantic relations. These two tasks are forms of sense disambiguation and semantic role labeling respectively and much research has been conducted on each task individually. This has produced many different solutions such as (Lesk 1986; Banerjee & Pedersen 2002; Mihalcea & Moldovan 2000; Mihalcea, Tarau, & Figa 2004; Patwardhan, Banerjee, & Pedersen 2003; Vasilescu, Langlais, & Lapalme 2004; Gildea & Jurafsky 2002; Pradhan *et al.* 2003; 2005; Hacioglu 2004; Swier & Stevenson 2005).

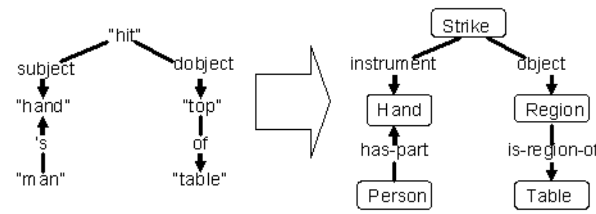

Figure 1: **Left:** The syntactic representation of our example sentence which captures the syntactic relationships between the constituents. **Right:** The semantic representation of the sentence.

Unfortunately, research in sense disambiguation usually focuses on determining just the meaning of a word within a context, and research in semantic role labeling usually focuses on determining just the semantic relation between a verb and its arguments. We have found it useful to combine these two tasks in order to produce rich semantic representations.

Our approach performs these two tasks through a single process – one of matching candidate semantic interpretations to background knowledge to select the interpretation with the best match.

We evaluate our approach by embedding it in a controlled (simplified) language system (Clark *et al.* 2002; 2005) to perform sense disambiguation and semantic role labeling. We chose a corpus made up of sentences from four domains – chemistry, pollution prevention, employee safety, and nuclear deterrence – and show how our approach performs well on both tasks.

Knowledge Requirements

Our approach requires a rich ontology to perform sense disambiguation and semantic role labeling, but existing resources provide limited semantics. The semantics of concepts in WordNet (Fellbaum 1998), for example, are limited mostly to hypernyms, meronyms, and synonyms. FrameNet (Baker, Fillmore, & Lowe 1998), on the other hand, provides richer semantics, but the focus is primarily on the semantic roles played by the syntactic arguments of verbs.

We want an ontology which provides representations with rich semantics that are linguistically motivated. Hence, we chose the *C*omponent *Lib*rary (CLib) built by Barker *et al.* (Barker, Porter, & Clark 2001). At the core of the CLib is a domain independent upper ontology with about 80 semantic relations and about 500 generic concepts (i.e. events and entities). These 500 concepts can be composed (and extended) to build domain specific ones. The CLib has over 2500 domain specific concepts in the military, chemistry, biology, and office domains.

Semantic Relations

One type of knowledge in the CLib is semantic relations – the targets for semantic role labeling. These semantic relations fall into three general categories:

1. Relations between an event and an entity – e.g. *agent*, *instrument*, *object*, *destination*, etc. These relations are in the spirit of case roles proposed by Fillmore (Fillmore 1971) and others (Baker, Fillmore, & Lowe 1998).
2. Relations between entities – e.g. *has-part*, *possesses*, *material*, etc.
3. Relations between events – e.g. *caused-by*, *prevents*, *enables*, etc.

Each semantic relation also has information about its syntactic realization (Barker 1998) – i.e. how the relation surfaces in a sentence. For example, *agent* can surface as a prepositional phrase marked by the preposition "by" (e.g. "A ball was hit by a man"), and *instrument* can surface in the same way (e.g. "A ball was hit by a stick"). From this information, we can look up the possible semantic relations for each syntactic one – e.g. some possible semantic relations for "by" are *agent*, *instrument*, *caused-by*, and *is-beside*.

Events and Entities

A second type of knowledge in the CLib covers events and entities – the targets for word sense disambiguation. Each event is similar to the frames in FrameNet (Baker, Fillmore, & Lowe 1998) and encodes knowledge about the participants in the event, where and when the event occurred, and other events that are caused (or prevented). For example, the *Move* event (see Figure 2) encodes knowledge about what is moved (i.e. the object), the origin, the destination, etc. The *Transfer* event (see Figure 2 also) encodes knowledge about what is transferred (i.e. the object), the donor, the recipient, etc.

Each entity encodes knowledge about its parts, its spatial relationship to other entities, and the possible roles it can play. For example, the representation of *Cell* (see Figure 2) says a cell encloses a nucleus and is part of an organism.

Each concept in the CLib is annotated with appropriate senses from WordNet 2.0 (Fellbaum 1998) to provide information about the concept's lexical realization. For example, the *Move* event is annotated with the WordNet senses of travel#1, go#1, move#1, etc.

Our approach

Our approach performs sense disambiguation and semantic role labeling by generating candidate interpretations and matching them against background knowledge to select the best one.

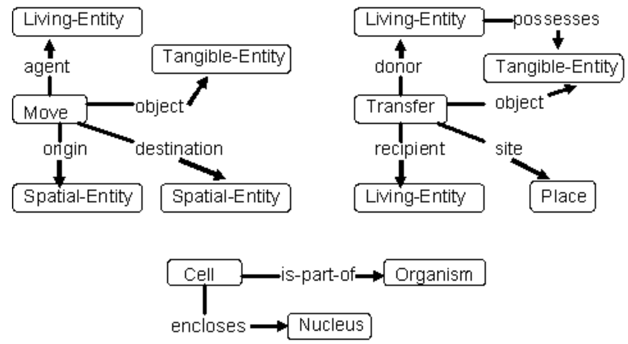

Figure 2: Encodings for the concepts of *Move*, *Transfer*, and *Cell*. We draw the encodings for these concepts as conceptual graphs.

Semantic Matcher

We start by describing briefly the semantic matcher used by our approach – for a complete discussion see (Yeh, Porter, & Barker 2003).

This semantic matcher takes two representations (encoded in a form similar to conceptual graphs (Sowa 1984)) and uses taxonomic knowledge (regarding both concepts and relations) to find the largest connected subgraph in one representation that is isomorphic to a subgraph of the other (it ignores degenerate – i.e. single node – subgraphs). This matcher then uses a library of about 200 transformation rules to shift the representations to improve the match. This improvement might enable other (non-degenerate) subgraphs to match isomorphically, which in turn might enable more transformation rules, and so on until the match improves no further.

This library of transformation rules is based on the CLib's upper ontology and has been used by Yeh *et al.* to improve matching in the domains of battle space planning (Yeh, Porter, & Barker 2003) and office equipment purchasing (Yeh, Porter, & Barker 2005). Each rule is an instance of the pattern "transfers through" (Lenat & Guha 1990) which has the following form:

$$C_1 \xrightarrow{r_1} C_2 \xrightarrow{r_2} C_3 \Rightarrow C_1 \xrightarrow{r_1} C_3$$

where C_i is a concept and r_j is a relation. Example rules include:

$$Event_1 \overset{object}{\longrightarrow} Entity_2 \overset{haspart}{\longrightarrow} Entity_3$$
$$\Longrightarrow Event_1 \overset{object}{\longrightarrow} Entity_3$$

$$Entity_1 \overset{haspart}{\longrightarrow} Entity_2 \overset{haspart}{\longrightarrow} Entity_3$$
$$\Longrightarrow Entity_1 \overset{haspart}{\longrightarrow} Entity_3$$

The first rule encodes *part descension* (i.e. acting on a whole also acts on its parts), and the second rule encodes the transitivity of the *has-part* relation.

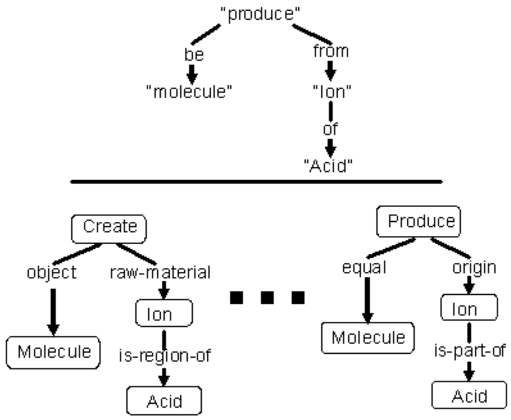

Figure 3: **Top:** The syntactic representation of our example sentence. **Bottom:** Some of the candidate interpretations.

A transformation rule is applicable to a representation if its antecedent subsumes a subgraph of the representation. A rule is applied by *joining* (Sowa 1984) the rule's consequent with the subgraph that matched the rule's antecedent.

Generate Candidates

Our approach generates a set of possible candidate interpretations from the syntactic representation of a sentence. To generate, for example, the candidate interpretations for:

"A molecule is produced from an ion of an acid."

our approach takes a shallow syntactic representation of this sentence (see Figure 3 top) and looks up the possible CLib concepts (i.e. senses) and relations for each constituent and syntactic relationship respectively (see previous section). The possible CLib concepts for "produce" are *Create*, *Produce*, etc.; the possible CLib relations for "from" are *origin*, *donor*, *raw-material*, *caused-by*, etc.; and so on.

Our approach generates all possible combinations from the resulting CLib concepts and relations. The concepts and relations in each combination replace their syntactic counterparts to produce a candidate semantic interpretation (which is in a form similar to conceptual graphs). Figure 3 (bottom) shows some of the candidate interpretations for our example sentence.

To reduce the number of candidate interpretations, our approach removes any that are invalid. A candidate interpretation is invalid if any of its semantic relations relate concepts outside the relations' scope. For example, *caused-by* relates *Produce* to *Ion* in some of the candidate interpretations for our example sentence, but *Ion* is outside the scope of *caused-by* because it is not an event.

Match Candidates

Each candidate interpretation G is matched to concepts in the ontology referenced in the interpretation. We will call this list of referenced concepts C. The concept in C that best matches G is removed from C, and the score for this match is added to the overall match score for G. This score is based on the semantic similarity between the candidate interpretation and the referenced concept and is computed using:

$$\left(\sum_{(n_i, n_j)} \frac{1}{d(n_i, n_j) + 1} \right) \div |G| \tag{1}$$

where (n_i, n_j) is a pair of matched concepts (or relations), d is the taxonomic distance (i.e. minimum number of steps) between two concepts (or relations) in an ontology, and $|G|$ is the total number of concepts (and relations) in G.

If G has unmatched information left, then our approach repeats the above step, but information (i.e. concepts and relations) in G that were matched previously are discarded. This process is repeated until all the information in G is matched or the match improves no further.

After all candidate interpretations have been processed, our approach selects the candidate with the best match score as the final interpretation of the sentence.

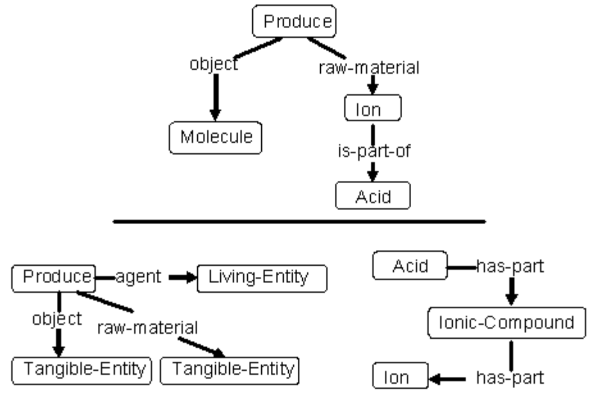

Figure 4: **Top:** The first candidate interpretation G for our example sentence. **Bottom:** Encodings for two of the concepts (i.e. Produce and Acid) referenced in this candidate interpretation.

Continuing with our previous example, our approach matches the first candidate interpretation G – shown at the top of Figure 4 – to the concepts of *Produce*, *Ion*, *Molecule*, and *Acid* (see Figure 4 bottom). G best matches *Produce* (see Figure 5 top) with a score of 0.49. In the CLib, $d(Molecule, Tangible\text{-}Entity) = 4$, $d(Ion, Tangible\text{-}Entity) = 4$, and the remaining matches for G (i.e. *Produce*, *raw-material*, and *object*) have a taxonomic distance of 0 because they are matched to the same concepts (and relations).

G still has unmatched information – i.e. the *Ion* is part of the *Acid*. So, our approach will match G with the remaining concepts in C – i.e. *Ion*, *Molecule*, and *Acid*. *Acid* would be a good match, but *Acid* has part an *Ionic-Compound* which has part an *Ion* – see Figure 5 bottom. This mismatch between what is said and what is represented in the ontology is very common, but our semantic matcher can resolve these mismatches using its library of transforms. For example, the transitivity of parts rule (given previously) will transform the

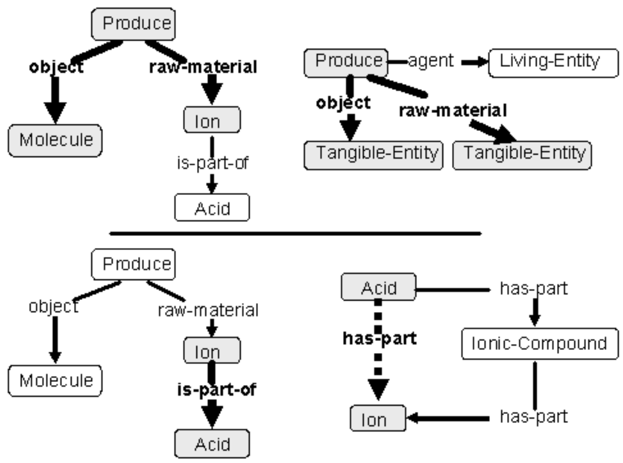

Figure 5: **Top:** The matches between G and *Produce* are shown in **bold**. **Bottom:** The matches between G and *Acid* are shown in **bold**. The dashed edge resulted from applying a transformation. Note, *is-part-of* is the inverse of *has-part*, so this edge can be reversed.

Acid concept by adding a *has-part* relation connecting the *Acid* to the *Ion*, and this transformation allows the remaining information in G to be matched – see Figure 5 bottom.

This is the best match with a score of 0.29 – $d(Acid, Acid) = 0$ and $d(has\text{-}part, has\text{-}part) = 0$. The match for *Ion* is discarded because it was matched previously, so the overall score for G is 0.78.

Evaluation

We evaluate the performance of our approach on the tasks of sense disambiguation and semantic role labeling. We embed our approach in a controlled language system (Clark *et al.* 2002; 2005) – the natural language interface for a question answering system that requires rich representations for reasoning. This system is similar to the ones described in (Friedland & others 2004).

Corpus

Our corpus contains 196 sentences collected from human subjects from four domains – chemistry, pollution prevention, employee safety, and nuclear deterrence. Example sentences from each domain include:

1. "An acid reacts with a base to transfer a proton to the base."
2. "The narrator on a beach says people cause water pollution."
3. "The tool injured the employee's right thumb."
4. "Musharraf said Pakistan expected India will test missiles."

These sentences are written in controlled (simplified) English which differs from unrestricted English in the following ways: there are no pronouns, there is a default preference for how prepositional phrases are attached, and only simple clauses and sentence structures are allowed.

These simplifications, however, do not make the task of selecting the appropriate senses and semantic relations for a sentence any easier. The average length of each sentence is 8.95 words. The total number of sense choices is 798, and each choice has on average 4.86 sense options. The total number of semantic relation choices is 608, and each choice has on average 6.98 semantic relation options.

To form the answer key, we asked two human subjects to select the appropriate senses and semantic relations. Each subject was given the syntactic representation for each sentence along with the possible CLib concepts (i.e. senses) and relations for each constituent and syntactic relationship respectively. The subjects were then asked to pick the appropriate semantic concepts and relations for each syntactic one. The kappa (Carletta 1996) agreement for sense and relation selection are 85.59% (1/5 chance agreement) and 89.44% (1/7 chance agreement) respectively.

Evaluation: Sense Disambiguation

To establish a baseline for comparison, we constructed a system which always picks the most frequent sense (given by WordNet) for each word. Further, we compare our approach to a well established knowledge based algorithm for word sense disambiguation – the Lesk algorithm (Lesk 1986). Recent adaptations of this algorithm (Banerjee & Pedersen 2002) disambiguate a word by comparing its WordNet glosses with those of its neighbors and selecting the sense with the most overlap. For our implementation, we have Lesk select the most frequent sense when there is no overlap to improve recall, and we map the WordNet senses selected by Lesk to the corresponding CLib concepts using the WordNet annotations in the CLib.

	Our Approach	**Lesk**	**Baseline**
Grader1			
Precision	86.59	75.69	74.06
Recall	84.96	75.69	74.06
Grader2			
Precision	93.87	78.70	77.07
Recall	92.11	78.70	77.07

Table 1: Each system's performance (given as percentages) on sense disambiguation for graders 1 and 2. N = 798. Precision and recall were the same for Lesk because it selected a sense for all 798 words. The Baseline had the same precision and recall for the same reason.

The systems were tested on our corpus. Their sense choices were compared against our answer key and graded using precision (i.e. the number of correct answers over the total number of answers given by a system) and recall (i.e. the number of correct answers over the total number of answers given by the gold standard) – see Table 1.

Our approach performed significantly better than Lesk and the baseline for both precision and recall across both graders ($p < 0.01$ for the $\mathcal{X}^2$ test in each case). We believe our approach performed better because it uses transformations to resolve mismatches, performs both sense disambiguation and semantic role labeling concurrently, and uses a rich ontology.

To verify this hypothesis, we constructed three additional systems. The first system (called NoXForm) is just like our approach except it does not use transformations. The second system (called NoReln) is just like our approach except we ablate information about semantic relations by replacing them with *relation* (the most general relation in the CLib) for all candidate interpretations. The final system (called Lesk+CLib) is like the Lesk algorithm except it looks for overlaps between CLib concepts instead of WordNet glosses. The performance of these three systems is shown in Table 2.

	NoXForm	**NoReln**	**Lesk+CLib**
Grader1			
Precision	86.21	78.21	75.81
Recall	75.19	75.56	75.81
Grader2			
Precision	93.39	84.18	79.70
Recall	81.45	81.33	79.70

Table 2: The three additional systems' performance on sense disambiguation for graders 1 and 2.

As expected, removing transformations causes recall to drop significantly because many mismatches between candidate interpretations and concepts in the ontology could not be resolved. Ablating information about the semantic relations not only reduces recall, but precision also. This shows that performing sense disambiguation and semantic role labeling concurrently better guides interpretation. Surprisingly, adapting the Lesk algorithm to use the CLib (instead of WordNet) has little affect on either precision or recall. Hence, these results show that the difference in performance is not in the background knowledge used, but in the algorithms themselves.

Evaluation: Semantic Role Labeling

We could not make a direct comparison between our approach and the state of the art (Gildea & Jurafsky 2002; Pradhan *et al.* 2003; 2005; Hacioglu 2004; Swier & Stevenson 2005) on the task of semantic role labeling. Existing approaches use different background knowledge (e.g. FrameNet, VerbNet, etc.) – each with its own unique set of semantic relations. Furthermore, existing approaches focus more on verb-argument roles and less on semantic relations among other constituents (such as intra-phrasal roles and inter-clausal roles). Our approach focuses on these relations as well. Finally, existing approaches require a large corpus to train on, but we have insufficient data to train such systems. This is because our approach is a knowledge based one which requires background knowledge not training examples.

Instead, we compared our approach to a variant with transformations ablated (called NoXForm) and a baseline that always picks the most common semantic relation associated with a syntactic one. Each approach was tested on our corpus. Their semantic relation choices were compared against the answer key and graded using precision and recall – see Table 3.

Our approach performed significantly better than the baseline for both precision and recall across both graders ($p < 0.01$ for the $\mathcal{X}^2$ test). Our approach performed significantly better than NoXForm on recall across both graders ($p < 0.01$ for the $\mathcal{X}^2$ test). We attribute this improvement to the only difference between our approach and NoXForm – the use of transformations to resolve mismatches.

	Our Approach	**NoXForm**	**Baseline**
Grader1			
Precision	85.18	80.89	41.78
Recall	82.24	66.12	41.78
Grader2			
Precision	92.16	87.73	42.43
Recall	88.98	71.71	42.43

Table 3: Each system's performance (given as percentages) on semantic role labeling for graders 1 and 2. N = 608.

Future Work

The results we observed are encouraging, but there are several issues and open questions that need to be addressed.

We evaluated our approach on a corpus of controlled English sentences, but our approach should work for unrestricted English also. The syntactic representations for unrestricted English sentences are more complex, but the process of interpreting them should be the same – i.e. matching candidate interpretations to background knowledge. We plan to investigate this problem in the near future using off the shelf, state of the art parsing techniques for unrestricted English.

We used the CLib as background knowledge, but our approach should be able to use other ontologies – provided the ontology has information about how its concepts (and relations) are realized syntactically. The CLib provides background knowledge for candidate interpretations to be matched against, but there is no reason why other ontologies cannot serve this same purpose. We plan to conduct additional studies to investigate the effect of using different ontologies such as FrameNet (Baker, Fillmore, & Lowe 1998) or Cyc (Lenat & Guha 1990).

We found that generating possible candidate interpretations (even just valid ones) can still pose a problem for efficiency. Some sentences in our corpus had thousands of valid candidates and took over 20 seconds to interpret. We plan to investigate ways to alleviate this problem. For example, we might use a greedy strategy to terminate the search when a complete match is found or a memory model to eliminate weak candidates based on prior interpretations.

We also observed that our approach prefers interpretations based on complete theories in the ontology as opposed to incomplete ones. For example, the possible CLib concepts for "strike" are *Collide* and *Attack-by-Fire*, but the encoding for *Attack-by-Fire* is less elaborate in the CLib. Hence, our approach prefers to interpret sentences like:

"The minister ordered the missile strike Pakistan."

as the "missile" colliding with Pakistan because *Collide* results in a better match. This interpretation is still sensible, but it is suboptimal. We plan to study how much our system can "accommodate" an interpretation (in the presence of incomplete knowledge) before inappropriate results are given.

Finally, we plan to compare our approach to the state of the art in semantic parsing (Ge & Mooney 2005; Kate, Wong, & Mooney 2005). Like our approach, semantic parsing performs both sense disambiguation and semantic role labeling, but it is a data driven approach. Our approach is a knowledge based alternative that requires rich background knowledge but does not need to be trained. Although background knowledge is expensive to build also, it supports other tasks such as reasoning, explanation generation, etc. We want to compare these two approaches to evaluate each one's strengths and weaknesses.

Conclusion

In this paper, we presented a unified knowledge based approach for sense disambiguation and semantic role labeling. Our approach performs both tasks through a single process – one of matching candidate interpretations to background knowledge to select the best match.

We evaluated our approach on a corpus of sentences from various domains, and showed how our approach performed well on both sense disambiguation and semantic role labeling. We also proposed reasons why our approach performed well – 1) it uses transformations to resolve mismatches between what is said and what is encoded in background knowledge, and 2) it performs both sense disambiguation and semantic role labeling concurrently which better guides the interpretation process.

Acknowledgments

We would like to thank the reviewers for helpful comments and suggestions and Peter Clark, John Thompson, Saurabh Baji, and Geoff King for their help on this project.

References

Baker, C.; Fillmore, C.; and Lowe, J. 1998. The Berkeley FrameNet Project. In *COLING*.

Banerjee, S., and Pedersen, T. 2002. An Adapted Lesk Algorithm for Word Sense Disambiguation Using WordNet. In *CICLING*.

Barker, K.; Porter, B.; and Clark, P. 2001. A Library of Generic Concepts for Composing Knowledge Bases. In *KCAP*.

Barker, K. 1998. *Semi-Automatic Recognition of Semantic Relationships in English Technical Texts*. Ph.D. Dissertation, University of Ottawa.

Carletta, J. 1996. Assessing Agreement on Classification Tasks: The Kappa Statistic. In *COLING*.

Clark, P.; Duncan, L.; Holmback, H.; Jenkins, T.; and Thompson, J. 2002. A Knowledge-Rich Approach to Understanding Text about Aircraft Systems. In *HLT*.

Clark, P.; Harrison, P.; Jenkins, T.; Thompson, J.; and Wojcik, R. 2005. Acquiring and Using World Knowledge Using a Restricted Subset of English. In *FLAIRS*.

Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Fillmore, C. 1971. Some Problems for Case Grammar. In *22nd Annual Round Table Meeting on Linguistics and Language Studies*.

Friedland, N., et al. 2004. Project Halo: Towards a Digital Aristotle. *AI Magazine* 25(4).

Ge, R., and Mooney, R. 2005. A Statistical Semantic Parser that Integrates Syntax and Semantics. In *CoNLL*.

Gildea, D., and Jurafsky, D. 2002. Automatic Labeling of Semantic Roles. *Computational Linguistics* 28(3).

Hacioglu, K. 2004. Semantic Role Labeling Using Dependency Trees. In *COLING*.

Kate, R.; Wong, Y.; and Mooney, R. 2005. Learning to Transform Natural to Formal Languages. In *AAAI*.

Lenat, D., and Guha, R. 1990. *Building Large Knowledge-Based Systems*. Addison-Wesley.

Lesk, M. 1986. Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. In *5th International Conference on Systems Documentation*.

Mihalcea, R., and Moldovan, D. 2000. An Iterative Approach to Word Sense Disambiguation. In *FLAIRS*.

Mihalcea, R.; Tarau, P.; and Figa, E. 2004. PageRank on Semantic Networks, with Application to Word Sense Disambiguation. In *COLING*.

Patwardhan, S.; Banerjee, S.; and Pedersen, T. 2003. Using Measures of Semantic Relatedness for Word Sense Disambiguation. In *CICLING*.

Pradhan, S.; Hacioglu, K.; Ward, W.; Martin, J.; and Jurafsky, D. 2003. Semantic Role Parsing: Adding Semantic Structure to Unstructured Text. In *ICDM*.

Pradhan, S.; Ward, W.; Hacioglu, K.; Martin, J.; and Jurafsky, D. 2005. Semantic Role Labeling Using Different Syntactic Views. In *ACL*.

Sowa, J. F. 1984. *Conceptual Structures: Information Processing in Mind and Machine*. Addison-Wesley.

Swier, R., and Stevenson, S. 2005. Exploiting a Verb Lexicon in Automatic Semantic Role Labelling. In *HLT/EMNLP*.

Vasilescu, F.; Langlais, P.; and Lapalme, G. 2004. Evaluating Variants of the Lesk Approach for Disambiguating Words. In *LREC*.

Yeh, P.; Porter, B.; and Barker, K. 2003. Using Transformations to Improve Semantic Matching. In *KCAP*.

Yeh, P.; Porter, B.; and Barker, K. 2005. Matching Utterances to Rich Knowledge Structures to Acquire a Model of the Speaker's Goal. In *KCAP*.