The Knowledge Required to Interpret Noun Compounds

James Fan, Ken Barker and Bruce Porter
Department of Computer Sciences
University of Texas at Austin
Austin, TX 78712 USA
{jfan, kbarker, porter}@cs.utexas.edu

Abstract

Noun compound interpretation is the task of determining the semantic relations among the constituents of a noun compound. For example, "concrete floor" means a floor made of concrete, while "gymnasium floor" is the floor region of a gymnasium. We would like to enable knowledge acquisition systems to interpret noun compounds, as part of their overall task of translating imprecise and incomplete information into formal representations that support automated reasoning. However, if interpreting noun compounds requires detailed knowledge of the constituent nouns, then it may not be worth doing: the cost of acquiring this knowledge may outweigh the potential benefit.

This paper describes an empirical investigation of the knowledge required to interpret noun compounds. It concludes that the axioms and ontological distinctions important for this task are derived from the top levels of a hierarchical knowledge base (KB); detailed knowledge of specific nouns is less important. This is good news, not only for our work on knowledge acquisition systems, but also for research on text understanding, where noun compound interpretation has a long history.

1 Introduction

Knowledge acquisition involves building knowledge bases (KBs) from the information provided by standard sources of expertise, such as people and texts. In addition to extracting relevant information, knowledge acquisition involves re-expressing the information in a formal, machine-sensible language. In general, this is difficult because the information is initially expressed in natural languages, and these expressions are notoriously imprecise and incomplete. However, the goal of our project is to improve knowledge acquisition methods by automating the translation of some kinds of expressions from natural languages to formal ones. We call this task *loose-speak interpretation*.

Several kinds of expressions are good candidates for loose-speak interpretation by knowledge acquisition systems. For example, noun compounds omit information that can often be inferred, *e.g. concrete floor* is "a floor *made of* concrete", while *gymnasium floor* is "the floor *region* of a gymnasium". Another candidate is metonymic expressions. These expressions contain incompatible terms and must be expanded to make meaningful phrases. For example, the statement *Joe read Shakespeare*, means "Joe read *a text written by* Shakespeare".

Relation Name	Example
material	marble statue
object-of	troop movement
isa	pine tree
location	sea mammal
product	protein production
part-of	human lung

Table 1: Some of the common semantic categories used in noun compound studies in computational linguistics.

This paper focuses on the first kind of loose-speak: interpreting noun compounds in the context of knowledge acquisition. A noun compound is a sequence of nouns composed of a head noun and one (or more) modifiers. The head noun determines the type of the whole compound (with few exceptions), and the modifiers specialize the type from the head noun. Although we limit our study to only pairs of nouns, our results can be applied to longer noun compounds by bracketing them into pairs of nouns (with few exceptions), and then interpreting each pair [Liberman and Sproat, 1992; Pustejovsky and Bergler, 1993; Barker, 1998].

The computational linguistics community has studied noun compound interpretation extensively [Leonard, 1984; Downing, 1977; Levi, 1979; Finin, 1986; Stal, 1996; Fabre, 1996; Lauer and Dras, 1994; Barker, 1998; Vanderwende, 1994]. In these studies, the task is to select a single semantic category for each pair of nouns. The selection is made from a list of about 20 semantic categories, such as *part-of, material* and *object-of*. See Table 1 for examples.

Our task is more general. Rather than selecting a single semantic category, our task is to find a sequence of semantic relations that links two nouns in a compound. Semantic relations are a list of about 50 thematic roles such as *agent, object, has-part, location, ...,*. For example, given *animal virus*, a traditional interpretation may classify this as a location category (*animal virus* is a virus in an animal). A loose-speak interpretation may be composed of a combination of semantic relations, such as: “an *animal virus* is a virus that is the agent of an invade, such that the object of the invade is the cell part of an animal”.

Furthermore, computational linguists approach the noun compound interpretation task armed with a corpus of examples, but little or no knowledge about the constituent nouns. Typical solutions are based on statistical patterns discovered in the corpus of examples. In contrast, we approach the task in the context of deeper knowledge of the constituent nouns – their taxonomic classification, at least – but few examples of noun compounds, let alone a corpus.

2 Interpreting noun compounds during knowledge acquisition

During knowledge acquisition, the domain expert (or, more generally, the knowledge source) may provide a noun compound in any dialogue where the system expects a noun. We say that our knowledge acquisition system successfully interprets the noun compound if it finds a sensible sequence of semantic relations between the head noun and its modifier and builds a correct formal representation of the noun compound.

If noun compound interpretation requires *a priori*, detailed knowledge of the head noun and its modifier, then the cost of acquiring this knowledge may overshadow the benefit of interpreting the compound. If, on the other hand, noun compounds can be successfully interpreted without much knowledge about the specific constituent nouns, then the problem is avoided: a knowledge acquisition system might interpret one concept (the noun compound) with little more than skeletal knowledge of the related concepts (the constituent nouns). Knowledge bases tend to grow in this uneven way – following the lead of the knowledge sources providing expertise – and a knowledge acquisition system should support it.

The purpose of this study is to determine what sort of knowledge is required to interpret noun compounds, and how this knowledge might be obtained. Before delving into the details of the study, it is important to understand what we are *not* attempting to do.

We are not presenting a novel algorithm for general noun compound interpretation. Our algorithm is quite simple and is derived from previous research. Also, we are not introducing a new type of knowledge representation or a novel technique of automated reasoning. Finally, we are not using a new, comprehensive knowledge base. We built two of them rather quickly and we’re using another – not built for this task – “off the shelf.”

In summary, what we *are* doing is evaluating the knowledge requirements of a standard search algorithm applied to a variety of typical knowledge bases through a series of ablation studies.

3 Experimental design

The challenge in measuring an algorithm’s sensitivity to knowledge base content is that the results may vary across domains and across knowledge bases. We attempt to neu-

```
Given a noun compound of the form <C1, C2>:

  0. RESULT = nil
  1. breadth−first search starting from C1:
     if the current level is deeper than maximum depth, then
       step 2
     if the current level has C2 or concepts that are superclasses/subclasses of C2,
       then
       append RESULT with the paths from C1 to these concepts and go to step 2
     else
       breadth−first search the next level along all the semantic relations
  2. breadth−first search starting from C2:
     if the current level is deeper than maximum depth, then
       step 3
     if the current level has C1 or concepts that are superclasses/subclasses of C1,
       then
       append RESULT with the paths from C2 to these concepts and go to step 3
     else
       breadth−first search the next level along all the semantic relations
  3. sort RESULT in ascending order based on the lengths of the paths in RESULT
Return RESULT
```

Figure 1: Our noun compound interpretation algorithm. It consists of two breadth-first searches in steps 1 and 2. The first starts from C_1 and looks for C_2 or any superclasses or subclasses of C_2. The second search starts from C_2 and looks for C_1 or any superclasses or subclasses of C_1. The results are combined, sorted based on path length, and returned.

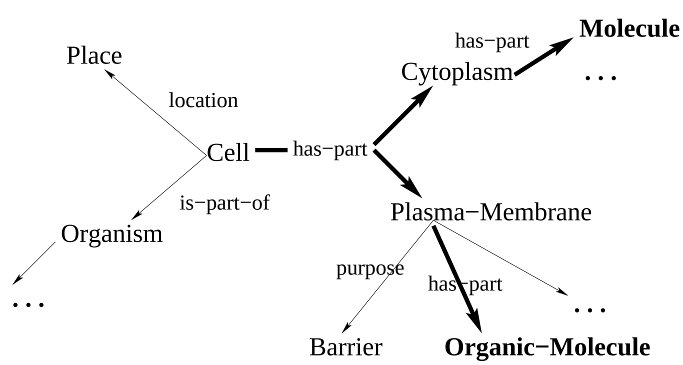


Figure 2: An example of how the algorithm works. Given the noun compound *cell lipid* (lipid is a fat molecule), the bold lines show the path found by breadth-first search starting at Cell and ending at Molecule and Organic-Molecule, both superclasses of lipid.

tralize these factors by replicating our study in three domains with quite different knowledge bases.

3.1 The algorithm

Our noun compound interpretation task can be viewed as follows: given a knowledge base encoded as a conceptual graph, and a pair of nouns corresponding to two nodes in the graph, find a path of semantic relations between them. The algorithm we used is a breadth-first search algorithm on a knowledge base. The algorithm (see figure 1) is given a noun compound of the form $< C_1, C_2 >$ where C_1 and C_2 are the KB concepts that are mapped from the constituents of the given noun compound. Each of the first two steps conducts a breadth-first search of the knowledge base along all semantic relation arcs. The first search starts from C_1 and looks for C_2 or any superclass or subclass of it. The second search starts from C_2 and looks for C_1 or any superclass or subclass of it. Step 3 combines the results, sorted by path length.

Figure 2 is an example of how the algorithm works. Given the noun compound *cell lipid* (lipid is a fat molecule), the search begins at Cell, then traverses all the semantic relations from Cell, such as *has-part* and *location*. The search stops at two nodes, described by these paths through the conceptual graph: “the organic-molecule part of the plasma membrane of the cell” and “the molecule part of the cytoplasm part of the cell”. The bold lines indicate paths found and returned.

There are two reasons why the algorithm stops when the search finds either a superclass or a subclass of the goal. When a subclass of the goal is found, by inductive reasoning we infer there is a path between the starting concepts and the goal concept: any instance of the subclass is an instance of the class itself. When a superclass of the goal is found, by abductive reasoning we infer there may be a path between the starting concepts and the goal concept: an instance of the superclass may be an instance of the class itself. Empirical results show that this combination of inductive and abductive reasoning is sufficient for our purpose.

3.2 The performance task

The interpretation of a noun compound is considered correct if the first path returned by the algorithm provides a *sensible* interpretation, as judged by a human oracle. If multiple paths of the same lengths are returned, then one path is randomly selected as the first path. Here is an example of one such *sensible* interpretation. Given “computer expert”, the interpretation “a computer acting as an expert” is correct, even though another interpretation (“an expert on computers”) might be the first to come to mind. This criterion is used for three reasons. First, our main focus is not the absolute performance of the algorithm applied to a particular knowledge base, but rather the relative contribution of various parts of the ontology, so any consistent standard used to evaluate whether an interpretation is correct will suffice. Second, this criterion is used in previous approaches such as [Barker, 1998] [Vanderwende, 1994]. Using the same criterion makes it easier to compare the results with other reported studies. Third, when data sets are extracted from text as pairs of nouns, the associated context is lost, and it is hard to know the “right” interpretation without any context.

The metrics we used to evaluate performance are precision and recall. Precision and recall are defined as follows [Jurafsky and Martin, 2000]:

$$Precision = \frac{\#\ of\ correct\ answers\ given\ by\ system}{\#\ of\ answers\ given\ by\ system}$$

$$Recall = \frac{\#\ of\ correct\ answers\ given\ by\ system}{total\ \#\ of\ possible\ correct\ answers}$$

For our experiment, the correct answers are the correctly interpreted noun compounds. Answers given by the system are the noun compounds for which interpretations have been found. Finally, all possible correct answers are all the noun compounds in the test set. Precision estimates the likelihood of a correct interpretation when an interpretation is found; recall is a measurement of the coverage, which suggests the likelihood of a correct interpretation when a noun compound is given.

3.3 Data sets

To avoid getting results that are skewed to a particular domain or knowledge representation, we used a variety of quite different data sets. The first consists of 224 noun compounds from a college-level cell biology text [Alberts *et al.*, 1998]. The second consists of 294 noun compounds from a small engine repair manual. The third data set consists of 224 compounds from a Sun Sparcstation manual. The nouns used in these data sets are mapped to the corresponding concepts in knowledge bases on these topics.

3.4 Knowledge bases

The knowledge bases(KBs) used for our experiments are made of a set of concepts and a set of axioms associated with each concept. The concepts are related to one another through subsumption relations to form a hierarchy. The axioms on one concept are assertions of semantic relations between that concept and other concepts. For example *Action* is a concept in our KB that describes things that *happen*. The axioms on Action include things such as “every Action has an object, which is an Entity”, where *Entity* is another concept in the KB to describe things that *are*. Axioms are either local or inherited from the axioms of a superclass.

Despite these commonalities, the KBs differ significantly. First they differ in terms of how they were built. The knowledge base for the biology text was built using the generic Component Library [Barker *et al.*, 2001] to answer end-of-the-chapter style questions, as one of the challenge problems for DARPA’s Rapid Knowledge Formation project [Clark *et al.*, 2001]. The knowledge bases for the other two data sets (the small engine repair manual and the Sparcstation manual) were built “on top of” the knowledge in WordNet [Fellbaum, 1998]. We augmented WordNet with the upper ontology of the generic Component Library plus about ten concepts that are important to each of the two domains (whose partonomies are not complete in WordNet). Through this process, we encoded 416 concepts in about 50 man-hours. The advantage of using WordNet as the foundation for these knowledge bases is two-fold: it includes most of the terms used in the data sets,

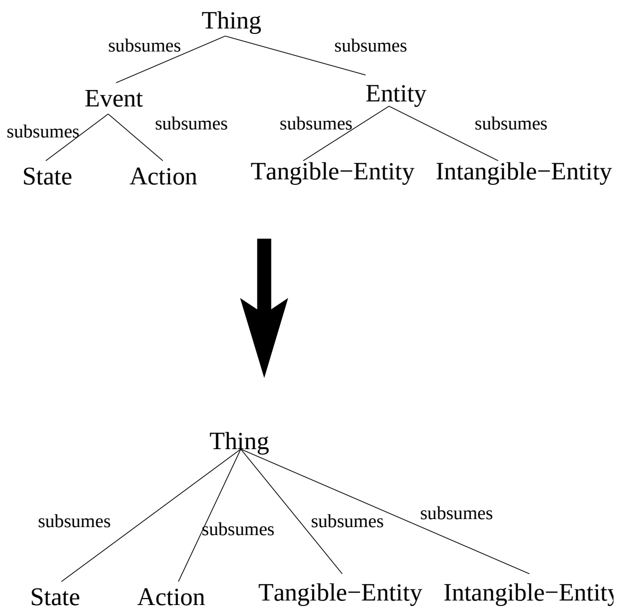


Figure 3: The ablation of level 1 on a sample taxonomy. Notice that the concepts "Entity" and "Event" are removed from the taxonomy. Classes on level 2 are promoted to level 1 and they are the direct subclasses of "Thing".

linked with both taxonomic and partonomic relations, and it is widely available and well used. The KBs also differ in content. Other than the shared upper ontology of the generic Component Library, they have few concepts in common.

3.5 Ablation steps

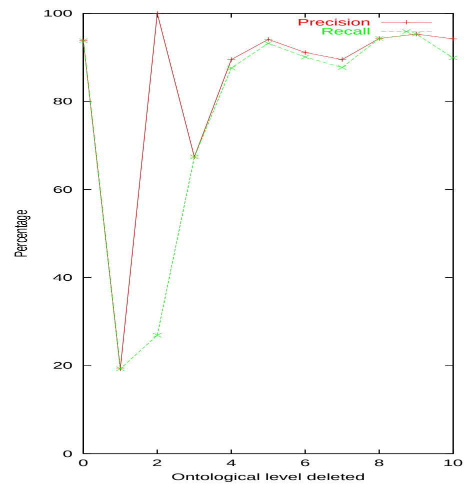
The sensitivity of the algorithm to each level of the ontology is measured through a series of ablations. When a level is ablated, the concepts on that level and all their axioms are deleted from the knowledge base. The superclasses of the subclasses of these concepts are changed to the superclasses of the concepts being deleted. As a special case, when the 0th level (the root level concept) is deleted, it is replaced by a generic concept of "Thing". Because the root level concept is vacuous, deleting it has no affect. As an example, figure 3 illustrates the ablation of level 1.

4 Data analysis

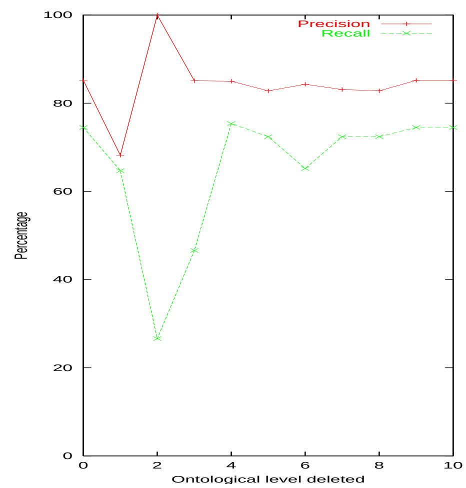
In this section, we analyze the experimental data to determine the contribution of each level of the knowledge bases' ontology to the task of interpreting noun compounds, and to find plausible explanations for the results.

4.1 Which levels of the taxonomy are most important?

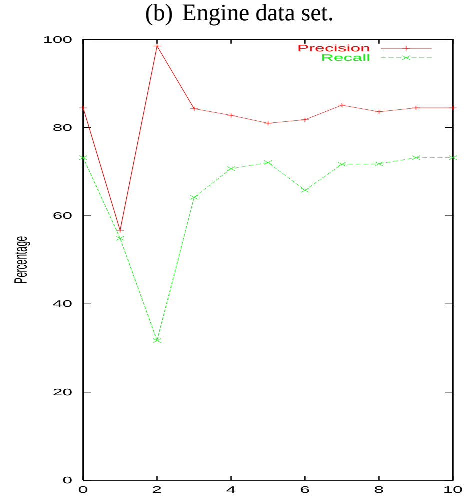
Figures 4(a–c) show the effect of ablating different levels of the ontology for the three knowledge bases in terms of precision and recall. Without any ablation, both precision and recall are around 80% across all three knowledge bases. Ablating level one causes a big drop in both precision and recall. Ablating level two introduces a big gap between precision



(a) Biology data set.



(b) Engine data set.



(c) Sparc data set.

Figure 4: The impact on precision and recall of ablating different levels of the ontology on the 3 data sets.

and recall. The gap indicates that the algorithm does not find interpretations for many noun compounds. As lower levels in the ontology are ablated one at a time, the impact diminishes and performance improves to the level of a knowledge base with no ablations.

The contribution of the first two levels of the ontology is observed across all three data sets and knowledge bases. This pattern strongly suggests that top levels of the ontology are most important for the noun compound interpretation task.

4.2 Why are the top levels of the taxonomy the most important?

Perhaps the top levels of the ontology are more important than lower levels because they contain many axioms important to the noun compound interpretation task. Ablating these levels has the affect of deleting the associated axioms, not only from the concepts at these levels, but also from all concepts below them in the taxonomy (because the axioms are passed by inheritance from superclasses to subclasses).

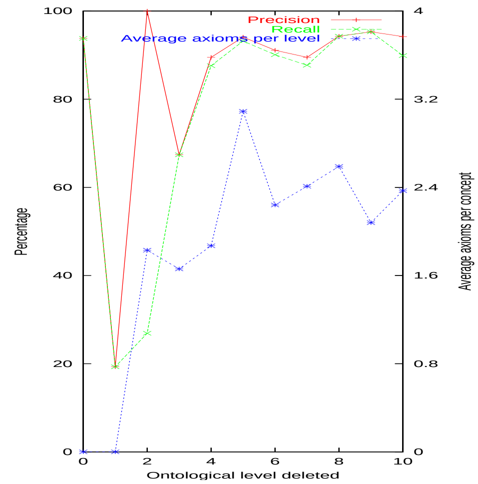
To verify this hypothesis, we counted the number of axioms associated with concepts at each level of the ontology. There are various ways to count axioms – we chose a simple and conservative one. For this study, we considered an axiom about concept C_1 to be an assertion of a semantic relation between instances of C_1 and instances of another concept C_2. For example, one axiom associated with the concept *Car* is: “every car has 4 wheels”. We counted only local (non-inherited) axioms, and we counted in minimum fashion, so this axiom is counted once, not multiple times.

Figures 5(a–c) compare the impact of ablating different ontological levels (in terms of precision and recall) with the average number of axioms per concept on each level. Ablating shows that upper levels of the ontology are most impactful, yet they contain relatively few axioms; lower levels of ontology are least impactful, yet they contain relatively more axioms.

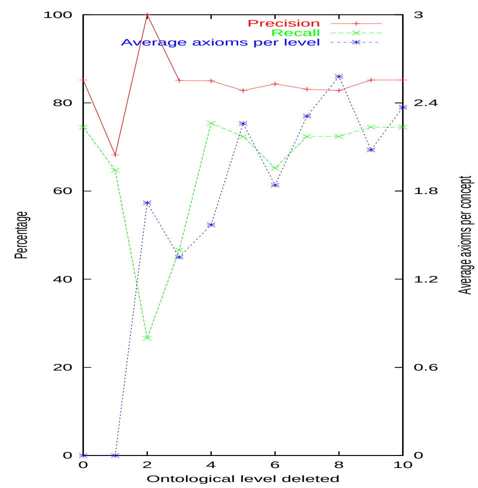
We conclude, therefore, the number of axioms in top levels of the ontology is not what makes these levels important. Rather, we suggest that they are important for noun compound interpretation for two reasons:

1. Top levels of the ontology include concepts that make important ontological distinctions. For example, ablating the level that introduces *Entity* and *Event* blurs the distinction between widely different classes of concepts, which causes the search to stop with erroneous results. Consequently, many more interpretations are returned, and because we only use the first interpretation, the probability of it being correct is reduced.

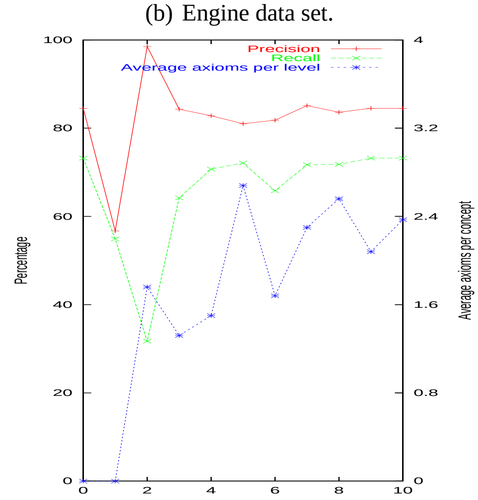
2. Although they contain relatively few axioms, axioms in top-level concepts are important for the task. The top-level ontology contains the most frequently used axioms, such as the fact that “every Action involves an object that is acted upon”. These axioms are used in the search as a step along the way. Deleting these axioms makes it difficult to find an interpretation for many of the noun compounds, thereby causing recall to lag behind precision.



(a) Biology data set.



(b) Engine data set.



(c) Sparc data set.

Figure 5: Comparison of the impact of ablating different levels of the ontology with the average number of axioms on each level.

5 Discussion

This paper reports an encouraging result: interpreting noun compounds does not require detailed knowledge of the constituent nouns. Rather, it requires only that the nouns be correctly placed in a taxonomy, and that the taxonomy include the ontological distinctions and axioms commonly found in domain-independent upper levels. These requirements are easily met in the context of knowledge acquisition, which is our focus.

We reached this conclusion using a novel experimental method. We measured the contribution of each level of the ontology to the task of interpreting noun compounds. We ablated levels of the ontology one at a time, thereby conflating ontological distinctions and removing the axioms associated with concepts at each level. We found that the upper levels of the ontology for the KBs we used are the most important for noun compound interpretation. As successively lower levels in the ontology are ablated, the impact becomes insignificant

One of the challenges in measuring an algorithm's sensitivity to knowledge base content is that the results may vary across domains and across knowledge representations. We attempted to neutralize these factors by evaluating performance in three domains with quite different data sets and knowledge bases. The primary results were consistent across them.

Acknowledgments

Support for this research is provided by a contract from Stanford Research Institute as part of DARPA's Rapid Knowledge Formation project.

References

[Alberts *et al.*, 1998] Bruce Alberts, Dennis Bray, Alexander Johnson, Julian Lewis, Martin Raff, Keith Robert, Peter Walter, and Keith Roberts. *Essential Cell Biology: An Introduction to the Molecular Biology of the Cell*. Garland Publisher, 1998.

[Barker *et al.*, 2001] Ken Barker, B. Porter, and P. Clark. A library of generic concepts for composing knowledge bases. In *Proceedings of First International Conference on Knowledge Capture*, 2001.

[Barker, 1998] Ken Barker. *Semi-automatic Recognition of Semantic Relationships in English Technical Texts*. PhD thesis, University of Ottawa, Ottawa, Ontario, 1998.

[Clark *et al.*, 2001] Peter Clark, J. Thompson, K. Barker, B. Porter, V. Chaudhri, A. Rodriguez, J. Thomere, S. Mishra, Y. Gil, P. Hayes, and T. Reichherzer. Knowledge entry as the graphical assembly of components. In *Proceedings of First International Conference on Knowledge Capture*, 2001.

[Downing, 1977] Pamela A. Downing. On the creation and use of English compounds. *Language*, 53:810 – 842, 1977.

[Fabre, 1996] Cecile Fabre. Interpretation of nominal compounds: Combining domain independent and domain-specific information. In *Proceedings of Sixteenth International Conference on Computational Linguistics*, pages 364 – 369, 1996.

[Fellbaum, 1998] Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database*. The MIT Press, Boston, 1998.

[Finin, 1986] Timothy W. Finin. *Constraining the Interpretation of Nominal Compounds in a Limited Context*. Lawrence Erlbaum Assoc, New Jersey, 1986.

[Jurafsky and Martin, 2000] Daniel Jurafsky and James Martin. *Speech and Language Processing: An Introduction to Natural Language Processing Computational Linguistics, and Speech Recognition*. Prentice Hall, New Jersey, 2000.

[Lauer and Dras, 1994] Mark Lauer and Mark Dras. A probabilistic model of compound nouns. In *Proceedings of the 7th Australian Joint Conference on Artificial Intelligence*. World Scientific Press, 1994.

[Leonard, 1984] Rosemary Leonard. *The Interpretation of English Noun Sequences on the Computer*. Elsevier Science, Amsterdam, 1984.

[Levi, 1979] Judith Levi. *The Syntax and Semantics of Complex Nominals*. Academic Press, New York, 1979.

[Liberman and Sproat, 1992] Mark Liberman and Richard Sproat. *Stress and Structure of Modified Noun Phrases*. CSLI Publications, 1992.

[Pustejovsky and Bergler, 1993] James S. Pustejovsky and Anick P. Bergler. Lexical semantic techniques for corpus analysis. *Computational Linguistics*, 19(2):331 – 358, 1993.

[Stal, 1996] Wilco Ter Stal. *Automated Semantic Analysis of Compounds: Problem Identification and Literature Overview*. PhD thesis, University of Twente, The Netherlands, 1996.

[Vanderwende, 1994] Lucy Vanderwende. Algorithm for automatic interpretation of noun sequences. In *Proceedings of Fifteenth Intional Conference on Computational Linguistics*, pages 782 – 788, 1994.