Automatic Discovery of Part–Whole Relations

Roxana Girju*
University of Illinois at Urbana-Champaign

Adriana Badulescu†
Language Computer Corporation

Dan Moldovan‡
Language Computer Corporation

An important problem in knowledge discovery from text is the automatic extraction of semantic relations. This paper presents a supervised, semantically intensive, domain independent approach for the automatic detection of part–whole relations in text. First an algorithm is described that identifies lexico-syntactic patterns that encode part–whole relations. A difficulty is that these patterns also encode other semantic relations, and a learning method is necessary to discriminate whether or not a pattern contains a part–whole relation. A large set of training examples have been annotated and fed into a specialized learning system that learns classification rules. The rules are learned through an iterative semantic specialization *(ISS) method applied to noun phrase constituents. Classification rules have been generated this way for different patterns such as genitives, noun compounds, and noun phrases containing prepositional phrases to extract part–whole relations from them. The applicability of these rules has been tested on a test corpus obtaining an overall average precision of 80.95% and recall of 75.91%. The results demonstrate the importance of word sense disambiguation for this task. They also demonstrate that different lexico-syntactic patterns encode different semantic information and should be treated separately in the sense that different clarification rules apply to different patterns.*

1. Introduction

The identification of semantic relations in text is at the core of Natural Language Processing and many of its applications. Detecting semantic relations between various text segments, such as phrases, sentences, and discourse spans, is important for automatic text understanding (Rosario, Hearst, and Fillmore 2002; Lapata 2002; Morris and Hirst 2004). Furthermore, semantic relations represent the core elements in the organization of lexical semantic knowledge bases intended for inference purposes. Recently, there has been a renewed interest in text semantics as evidenced by the international

* Computer Science Department, University of Illinois at Urbana-Champaign, Urbana, IL 61801, E-mail: girju@uiuc.edu.
† Language Computer Corporation, 1701 N. Collins Blvd. Suite 2000, Richardson, TX 75080, E-mail: adriana@languagecomputer.com.
‡ Language Computer Corporation, 1701 N. Collins Blvd. Suite 2000, Richardson, TX 75080, E-mail: moldovan@languagecomputer.com.

Submission received: 21 October 2003; revised submission received: 7 March 2005; accepted for publication: 1 August 2005.

participation in the Senseval 3 Semantic Roles competition,[1] the associated workshops,[2] and numerous other workshops.

An important semantic relation for many applications is the *part–whole* relation, or *meronymy*. Let us notate the part–whole relation as PART(*X, Y*), where *X* is part of *Y*. For example, the compound nominal *door knob* contains the part–whole relation PART*(knob, door)*. Part–whole relations occur frequently in text and are expressed by a variety of lexical constructions as illustrated in the text below.

(1) The car's mail messenger is busy at work in the *mail car* as the *train* moves along. Through the open *side door* of the *car*, moving scenery can be seen. The worker is alarmed when he hears an unusual sound. He peeks through the *door*'s *keyhole* leading to the *tender* and *locomotive cab* and sees the two bandits trying to break through the *express car door*.[3]

There are several part–whole relations in this text: 1) the *mail car* is part of the *train*, 2) the *side door* is part of the *car*, 3) the *keyhole* is part of the *door*, 4) the *cab* is part of the *locomotive*, 5) the *tender* is part of the *train*, 6) the *locomotive* is part of the *train*, 7) the *door* is part of the *car*, and 8) the *car* is part of the *express* train (in the compound noun *express car door*).

This paper provides a supervised, knowledge-intensive method for the automatic detection of part–whole relations in English texts. Based on a set of positive (encoding meronymy) and negative (not encoding meronymy) training examples provided and annotated by us, the algorithm creates a decision tree and a set of rules that classify new data. The rules produce semantic conditions that the noun constituents matched by the patterns must satisfy in order to exhibit a part–whole relation. For the discovery of classification rules we used C4.5 decision tree learning (Quinlan 1993). The learned function is represented by a decision tree transformed into a set of if–then rules. The decision tree learning searches a complete hypothesis space from simple to complex hypotheses until it finds a hypothesis consistent with the data. Its bias is a preference for the shorter tree that places high information gain attributes closer to the root.

For training purposes we used WordNet, and the LA Times (TREC9)[4] and SemCor 1.7[5] text collections. From these we formed a large corpus of 27,963 negative examples and 29,134 positive examples of well distributed subtypes of part–whole relationships which provided a comprehensive set of classification rules. The rules were tested on

1 http://www.senseval.org/senseval3.
2 The Computational Lexical Semantics Workshop at the 2004 Human Language Technology (HLT/NAACL) conference; the first and second Workshops on Text Meaning and Interpretation at the HLT/NAACL-03 and the 2004 Association for Computational Linguistics conference (ACL), respectively; the first and second Workshops on Multiword Expressions at ACL 2003 and 2004; the ACL 2005 Workshop on Deep Lexical Acquisition.
3 This example is an excerpt from a review of the 1903 movie "The Great Train Robbery" (http://filmsite.org/grea.html).
4 TREC 9 is a text collection provided by NIST for the Question Answering competition (TREC-QA) at the TExt Retrieval Conference in 2000. It contains 3 GBytes of news articles from the Wall Street Journal, Financial Times, LA Times, Financial Report, AP Newswire, San Jose Mercury News, and Foreign Broadcast Information Center from 1989 to 1994.
5 The SemCor collection (Miller et al., 1993) is a subset of the Brown Corpus and consists of 352 news articles distributed into three sets in which the nouns, verbs, adverbs, and adjectives have been manually tagged with their corresponding WordNet senses and part-of-speech tags using Brill's tagger (1995).

two different text collections (LA Times and Wall Street Journal) obtaining an overall average precision of 80.95% and recall of 75.91%.

In this paper we do not distinguish between situations when whole objects consist of parts that are *always* present, or parts that are only *sometimes* present. For example, it might be relatively easy to pin down the parts of a car (e.g., *four wheels, one engine*, as ever present parts of a car irrespective of its type) as compared to enumerating all the components of a sandwich (e.g., *two layers of cheese* and/or *salami*, *two slices of bread*, that depend on the type of sandwich). In our experiments we focus only on part–whole instances that are mentioned in the corpus employed and on those provided by general-purpose lexical knowledge bases such as WordNet,[6] whether the parts are just *sometimes* constituents of the entity considered or are *always* present. We do not check for the validity of these instances (e.g., whether the instance "*wood is part of a sandwich*" is true or not). Based on a large training corpus of positive and negative part–whole examples, our system infers what type of objects are parts and wholes. Also, our system does not take into consideration modality information such as knowledge about the possibility, certainty, or probability of existence of part–whole relations.

The paper is organized as follows. Section 2 presents a summary of previous work on meronymy from several perspectives. Section 3 gives a detailed classification of the lexico-syntactic patterns used to express meronymy in English texts and a procedure for finding these patterns. Section 4 describes a method for learning semantic classification rules, while Section 5 shows the results obtained for discovering the part–whole relations by applying the classification rules on two distinct test corpora. Section 6 comments on the method's limitations and extensions, and Section 7 discusses the relevance of the task to NLP applications. Conclusions are offered in Section 8.

2. Previous Work on Meronymy

Historically, **part–whole** or **meronymy** relations have played an important role in linguistics, philosophy, and psychology mainly because a clear understanding of part–whole relations requires a deep interaction of logic, semantics, and pragmatics as they provide the tools needed for our understanding of the world. The part–whole relation has been considered a fundamental ontological relation since the atomists (Plato, Aristotle, and the Scholastics). They were the first to give a systematic characterization of parts and wholes, the relation between them, and the inheritance properties of this relation. However, most of the investigations of part–whole relations have been made since the beginning of the 20th century.

The **logical/philosophical** studies of meronymy were concerned with formal theories of parts (mereologies), wholes, and their relation in the context of formal ontology. This school of thought advocates a single, universal, and transitive part-of relation used for modeling various domains such as time and space.

Simons (1986) criticized this standard extensional view and proposed a more adequate account that offers an axiomatic representation of the *part-of* relation as a strict partial-ordering relation. The axioms considered were: *existence* (if A is a part of B then both A and B exist), *asymmetry* (if A is a part of B then B is not a part of A), *supplementarity* (if A is a part of B then B has a part C disjoint of A), and *transitivity* (if

6 For example, in WordNet 1.7 the only part listed for the concept *sandwich* is *bread*.

A is a part of B and B is a part of C then A is a part of C). In 1991, Simons (1991) added two more axioms: *extensionality* (objects with the same parts are identical) and *existence of mereological sum* (for any number of objects there exists a whole that consists exactly of those objects).

Linguistics and **cognitive psychology** researchers focused on different part–whole relations and their role as semantic primitives. Since there are different ways in which something can be expressed as part of something else, many researchers have claimed that meronymy is a complex relation that "should be treated as a collection of relations, not as a single relation" (Iris, Litowitz, and Evens 1988).

Based on psycholinguistic experiments and the way in which the parts contribute to the structure of the wholes, Winston, Chaffin, and Hermann (1987) determined six types of part–whole relations: (1) COMPONENT–INTEGRAL OBJECT, (2) MEMBER–COLLECTION, (3) PORTION–MASS, (4) STUFF–OBJECT, (5) FEATURE–ACTIVITY, and (6) PLACE–AREA.

They also proposed three *relation elements* (*functional, homeomerous*, and *separable*) to further classify the six types of meronymy relations. The *functional* relational element indicates that the part has a function with respect to its whole, whereas *homeomerous* means that the part is identical to the other parts making up the whole. The *separable* relational element shows that the part can be separated from the whole. For example, the relation *wheel–car* is a COMPONENT–INTEGRAL part–whole relation that is *functional, non-homeomerous* and *separable*. This means that the wheel has a specific function with respect to the car, does not resemble the other parts of the car, and can be separated from the car.

The COMPONENT–INTEGRAL relation is the relation between components and the objects to which they belong. Integral objects have a structure, their components are separable and have a functional relation with their wholes. For example, *kitchen–apartment* and *aria–opera* are COMPONENT–INTEGRAL relations.

The MEMBER–COLLECTION relation represents membership in a collection. Members are parts, but they cannot be separated from their collections and do not play any functional role with respect to their whole. For example, *soldier–army, professor–faculty*, and *tree–forest* are MEMBER–COLLECTION relations.

PORTION–MASS captures the relations between portions and masses, extensive objects, and physical dimensions. The parts are separable and similar to each other and to the wholes which they comprise, and do not play any functional role with respect to their whole. For example, *slice–pie* and *meter–kilometer* are PORTION–MASS relations.

The STUFF–OBJECT category encodes the relations between an object and the stuff of which it is partly or entirely made. The parts are not similar to the wholes that they comprise, cannot be separated from the whole, and have no functional role. For example, *steel–car* and *alcohol–wine* are STUFF–OBJECT relations.

The FEATURE–ACTIVITY relation captures the semantic links within features or phases of various activities or processes. The parts have a functional role, but they are not similar or separable from the whole. For example, *paying–shopping* and *chewing–eating* are FEATURE–ACTIVITY relations.

PLACE–AREA captures the relation between areas and special places and locations within them. The parts are similar to their wholes, but they are not separable from them. For example, *oasis–desert* and *Guadalupe Mountains National Park–Texas* are PLACE–AREA relations.

In this paper we use the Winston, Chaffin, and Hermann classification as a criterion for building the training corpus to provide a wide coverage of such subtypes of part–whole relations.

In **computational linguistics,** although a considerable amount of work has been done on semantic relation detection,[7] the work most similar to the task of identifying part–whole semantic relations is that of Hearst (1992) and Berland and Charniak (1999).

Hearst developed a method for the automatic acquisition of hypernymy relations by identifying a set of frequently used and mostly unambiguous lexico-syntactic patterns. For example, *countries, such as England* indicates a hypernymy relation between the words *countries* and *England*. In her paper, she mentions that she tried applying the same method to meronymy, but without much success, as the patterns detected also expressed other semantic relations. This is consistent with our study of part–whole lexico-syntactic patterns presented in this paper.

In 1999, Berland and Charniak applied statistical methods to a very large corpus[8] to find part–whole relations. Using Hearst's method, they focused on a small set of genitive patterns and a list of six seeds representing whole objects (*book, building, car, hospital, plant*, and *school*). Their system's output was an ordered list of possible parts according to some statistical metrics (e.g., the log-likelihood metric (Dunning 1993)). Although the training corpus used is very large, the coverage of the algorithm is small due to the limited number of patterns used and the small number of wholes allowed. Moreover, certain words, such as those ending in *-ing*, *-ness*, or *-ity*, were ruled out. Their accuracy is 55% for the first 50 ranked parts and 70% for the first 20 ranked parts. As a baseline, they considered as potential parts the head nouns immediately surrounding the target whole object and ranked them based on the same statistical metric. The baseline accuracy was 8%.

While Berland and Charniak's method focuses solely on identifying parts given a whole, our task targets the identification of both parts and wholes.

Hearst, and Berland and Charniak observed that for ambiguous whole words, such as *plant*, the method produces the weakest part list of the six seeds considered. Although they don't provide a one-to-one comparison, Berland and Charniak mention that their method outperforms Hearst's pattern matching algorithm mainly due to the very large corpus used. However, neither approach addresses the pattern ambiguity problem, i.e., patterns such as genitives that can express different semantic relations in different contexts (*the dress of silk* encodes a part–whole relation, but *the dress of my girl* does not). The ambiguity of these patterns explains our rationale for choosing an approach based on a machine learning method to discover discriminating rules automatically.

3. Lexico-Syntactic Patterns that Express Meronymy

The automatic discovery of any semantic relation must start with a thorough understanding of the lexical and syntactic forms used to express that relation. Since there are many ways in which something can be part of something else, there is a variety of lexico-syntactic structures that can express a meronymy semantic relation.

7 Besides the work on semantic roles (Charniak 2000; Gildea and Jurafsky 2002; Thompson, Levy, and Manning 2003), considerable interest has been shown in the automatic interpretation of various noun phrase-level constructions, such as noun compounds. The focus here is to determine the semantic relations that link the two noun constituents. The best-performing noun compound interpretation systems have employed either symbolic (Finin 1980; Vanderwende 1994) or statistical techniques (Pustejovsky, Bergler, and Anick 1993; Lauer and Dras 1994; Lapata 2002) relying on rather ad hoc, domain-specific, hand-coded semantic taxonomies, or on statistical patterns in a large corpus of examples, respectively.

8 The North American News Corpus (NANC) of 1 million words.

There are *unambiguous* lexical expressions that always convey a part–whole relation. For example:

(2) *The substance* consists of *three ingredients.*

(3) *The cloud was* made of *dust.*

(4) *Iceland is a* member of *NATO.*

In these cases the simple detection of the patterns leads to the discovery of part–whole relations.

On the other hand, there are many *ambiguous* expressions that are explicit but convey part–whole relations only in some contexts. The detection of meronymy in these cases is based on extracting semantic features of constituents and checking whether or not these features match the classification rules. For example, *The horn is part of the car* is meronymic whereas *He is part of the game* is not.

In the case of meronymy, since there are numerous unambiguous and ambiguous patterns, we devised a method to find these patterns and rank them in the order of their frequency of use. Our intention is to detect the most frequently occurring patterns that express meronymy and provide an algorithm for their automatic detection and disambiguation in text.

3.1 An Algorithm for Finding Lexico-Syntactic Patterns

In order to identify lexico-syntactic forms that express part–whole relations and determine their distribution over a very large corpus, we used the following algorithm inspired by Hearst's (1998) work:

Step 1. Pick pairs of concepts C_i, C_j among which there is a part–whole relation
For this task, we used the information provided by WordNet 1.7 (Fellbaum 1998). In WordNet, the nouns are organized into nine hierarchies, each hierarchy being identified by its corresponding root concept: {*abstraction*}, {*act*}, {*entity*}, {*event*}, {*group*}, {*phenomenon*}, {*possession*}, {*psychological feature*}, and {*state*}. The nouns are grouped in concepts or synsets; a concept consisting of a list of synonymous word senses. For example, {*mother*#1, *female parent*#1} is a WordNet concept. Besides concepts, WordNet contains 11 semantic relations: HYPONYMY (IS–A), HYPERNYMY (REVERSE IS–A), MERONYMY (PART–WHOLE), HOLONYMY (REVERSE PART–WHOLE), ENTAIL, CAUSE–TO, ATTRIBUTE, PERTAINYMY, ANTONYMY, SYNSET (SYNONYMY), and SIMILARITY. The part–whole relations in WordNet are further classified into three basic types: MEMBER-OF (e.g., *UK#1* IS-MEMBER-OF *NATO#1*), STUFF-OF (e.g., *carbon#1* IS-STUFF-OF *coal#1*), and PART-OF (e.g., *leg*#3 IS-PART-OF *table*#2) which includes all the other part–whole relations described in the Winston, Chaffin and Hermann (WCH) classification.

Since the part and whole concepts provided by WordNet can belong to almost any WordNet noun hierarchy, we randomly selected 100 pairs of part–whole concepts that were well distributed over all nine WordNet noun hierarchies, the three WordNet meronymic relations, and the six types of part–whole relations of WCH. Two annotators with computational linguistic knowledge classified the WordNet meronymic relations into the WCH's six part–whole types. According to our annotations, the MEMBER-

OF WordNet relations correspond to Winston, Chaffin, and Hermann's MEMBER–COLLECTION relations, STUFF-OF relations correspond to WCH's STUFF–OBJECT relations, and the PART-OF correspond to the other four WCH relations. The annotators obtained a 100% agreement in mapping the MEMBER-OF to MEMBER–COLLECTION, STUFF-OF to STUFF–OBJECT. The PART-OF relations were mapped to the other four types of WCH relations with an average agreement of 98%. A third judge (one of the authors) checked the correctness of all the mappings and decided on the non-agreed instances. This mapping ensures that the 100 general-purpose WordNet pairs cover most of the possible types of part–whole relations in text.

Table 1 shows only 50 pairs from the set of 100 WordNet part–whole pairs and their distribution among the WordNet hierarchies and the part–whole types provided by WordNet and the WCH taxonomy. For example, the pair *Bucharest#1–Romania#1* is a PART-OF relation in WordNet, but based on the Winston, Chaffin, and Hermann classification it can be further classified as a more specific meronymy relation, PLACE–AREA.

For the purpose of this research, we lumped together all part–whole types in the classification of Winston et al.[9] However, the method presented in the paper is applicable to extracting subtypes of part–whole relations; separate annotations for each type would be necessary.

Step 2. Search a corpus and extract lexico-syntactic patterns that link a pair of part–whole concepts

For each pair of part–whole noun concepts determined above, search the Internet or any other large collection of documents and retain only the sentences containing that pair. Since our intention is to demonstrate that the automatic procedure proposed here is domain independent, we chose two distinct text collections: SemCor 1.7 and the LA Times from TREC-9. From each collection we randomly selected 10,000 sentences, which were searched for the pair of concepts selected. Since the LA Times collection is not word-sense disambiguated, we searched for sentences containing the pair of nouns without considering their senses. Out of these sentences, only some contained the part–whole pairs selected in Step 1. We manually inspected these sentences and picked only those in which the pairs involved meronymy. For example, the sentence *I can feel my fingers and close my hand* contains the meronymic pair *finger–hand*, but in this context the relationship is not expressed. From these sentences we manually extracted meronymic lexico-syntactic patterns. Table 2 shows for each collection the number of sentences used, the number of sentences that contain the studied concept pairs, the number of sentences that contain part–whole relations, and the number of unique patterns discovered from those sentences. Seven of the unique patterns occurred in both SemCor and the LA Times.

In order to extract the patterns from the SemCor collection we used its gold standard word sense annotations to our advantage and looked for the occurrences of concepts (word with the sense) in the corpus. This explains the large difference between the number of sentences discovered in the two corpora. The SemCor patterns thus extracted did not need manual validation, since the noun concept pairs were always in a part–whole relation.

9 The WCH categories were also used by the annotators to better distinguish between positive and negative examples.

Table 1
The list of fifty selected part–whole relation pairs used as input for the lexico-syntactic pattern identification procedure. *WN Type* is the part–whole type from WordNet and *WCH Type* is the part–whole type from the Winston, Chaffin and Hermann taxonomy.

Part concept	Whole concept	The Part Hierarchy	The Whole Hierarchy	WN Type	WCH Type
act	play	abstraction	abstraction	PART-OF	PORTION–MASS
air	wind	entity	phenomenon	STUFF-OF	STUFF–OBJECT
artillery	battery	entity	group	MEMBER-OF	MEMBER–COLLECTION
bodyguard	guard	entity	group	MEMBER-OF	MEMBER–COLLECTION
Bucharest	Romania	entity	entity	PART-OF	PLACE–AREA
cellulose	paper	entity	entity	STUFF-OF	STUFF–OBJECT
chew	eating	entity	phenomenon	PART-OF	FEATURE–ACTIVITY
chorus	song	group	entity	PART-OF	PORTION–MASS
computer	computer network	entity	entity	PART-OF	COMPONENT–INTEGRAL
door	car	entity	entity	PART-OF	COMPONENT–INTEGRAL
epileptic_seizure	epilepsy	event	state	PART-OF	FEATURE–ACTIVITY
executive	government	group	group	MEMBER-OF	MEMBER–COLLECTION
fight	war	entity	act	PART-OF	FEATURE–ACTIVITY
finger	hand	entity	entity	PART-OF	COMPONENT–INTEGRAL
foot	leg	entity	entity	PART-OF	COMPONENT–INTEGRAL
gentle_breeze	Beaufort_scale	phenomenon	abstraction	MEMBER-OF	MEMBER–COLLECTION
Gibraltar	Europe	group	entity	PART-OF	PLACE–AREA
hand	arm	entity	entity	PART-OF	COMPONENT–INTEGRAL
inch	foot	abstraction	abstraction	PART-OF	PORTION–MASS
iron	steel	entity	entity	STUFF-OF	STUFF–OBJECT
knee	leg	entity	entity	PART-OF	COMPONENT–INTEGRAL
leg	chair	entity	entity	PART-OF	COMPONENT–INTEGRAL
letter	alphabet	abstraction	abstraction	MEMBER-OF	MEMBER–COLLECTION
liquid_assets	capital	possession	possession	PART-OF	MEMBER–COLLECTION
lock	door	entity	entity	PART-OF	COMPONENT–INTEGRAL

Table 1
(cont.)

Part concept	Whole concept	The Part Hierarchy	The Whole Hierarchy	WN Type	WCH Type
member	organization	entity	group	MEMBER-OF	MEMBER–COLLECTION
memory	computer	entity	entity	PART-OF	COMPONENT–INTEGRAL
metacarpus	hand	entity	entity	PART-OF	COMPONENT–INTEGRAL
meter	kilometer	abstraction	abstraction	PART-OF	PORTION–MASS
middle_age	adulthood	abstraction	abstraction	PART-OF	FEATURE–ACTIVITY
myocardial_infarct	heart_attack	state	event	PART-OF	FEATURE–ACTIVITY
number	series	abstraction	entity	MEMBER-OF	MEMBER–COLLECTION
oxygen	air	entity	entity	STUFF-OF	STUFF–OBJECT
palm	hand	entity	entity	PART-OF	COMPONENT–INTEGRAL
pavement	road	entity	entity	STUFF-OF	STUFF–OBJECT
paw	cat	entity	entity	PART-OF	COMPONENT–INTEGRAL
people	world	group	group	MEMBER-OF	MEMBER–COLLECTION
promenade	ball	entity	group	PART-OF	FEATURE–ACTIVITY
Romania	Europe	entity	entity	PART-OF	PLACE–AREA
Romanian	Romania	entity	entity	MEMBER-OF	MEMBER–COLLECTION
Sahara	Africa	entity	entity	PART-OF	PLACE–AREA
shower	bath	entity	entity	PART-OF	COMPONENT–INTEGRAL
snow	snowball	entity	entity	STUFF-OF	STUFF–OBJECT
symptom	disease	psych_feature	state	PART-OF	FEATURE–ACTIVITY
tympanum	ear	entity	entity	PART-OF	COMPONENT–INTEGRAL
volume	set	entity	group	MEMBER-OF	MEMBER–COLLECTION
water	ice	entity	entity	STUFF-OF	STUFF–OBJECT
Waterloo	Belgium	entity	entity	PART-OF	PLACE–AREA
window	car	entity	entity	PART-OF	COMPONENT–INTEGRAL
wing	angel	entity	psych_feature	PART-OF	COMPONENT–INTEGRAL

3.2 Taxonomy of Part–Whole Patterns

From the 535 part–whole relations detected from the 20,000 SemCor and LA Times sentences, 493 (92.15%) were expressed by phrase-level patterns and 42 (7.85%) by sentence-level patterns. Overall, there were 42 unique meronymic lexico-syntactic patterns, of which 31 were phrase-level patterns and 11 sentence-level patterns.

Recall our notation for the part–whole relation PART(X, Y), where X is part of Y.

a. Phrase-level patterns
Here, the part and whole concepts are included in the same phrase. For example, for the pattern NP_X PP_Y the noun phrase that contains the part and the prepositional phrase that contains the whole are found in the same noun phrase. *The engine in the car* is an instance of this pattern where X is the part (*engine*) and Y is the whole (*car*).

b. Sentence-level patterns
In these constructions, the part–whole relation is intrasentential. The patterns contain specific verbs and the part and the whole can be found inside noun phrases or prepositional phrases that contain specific prepositions. A frequent such pattern is NP_Y *verb* NP_X, where NP_X is the noun phrase that contains the part, NP_Y is the noun phrase that contains the whole and the *verb* is restricted (see Table 2 of Appendix A). For instance, *the cars have doors* is an instance of this pattern.

An extension of this pattern is NP_X *verb* NP_Z PP_Y, with NP_Z containing the words *part* or *member*. An example is: *The engine is a part of the car*; NP_X — the *engine*, PP_Y — *of the car*, and the *verb* — *to be*.

In some instances the meronymic constructions contained combinations, conjunctions and/or disjunctions, of parts and wholes. For example, NP_{X1X2} PP_Y (e.g., *wheels and engine of a car*) is a form of the pattern NP_X PP_Y. This observation enabled us to generalize the list of patterns. A summary of phrase-level and sentence-level meronymic patterns along with their extensions and generalizations is provided in Appendix A.

Based on our observations of the corpus used for the pattern identification procedure and based on the results obtained by others (Evens et al. 1980), we have concluded that the lexico-syntactic patterns encoding meronymy can be classified according to their semantic similarity and frequency of occurrence into the clusters presented in Table 3. The clusters contain lexico-syntactic patterns that have similar semantic behavior. We also noticed that more than a half of cluster 4's patterns are very rare; for example, *X branch of Y*; *In Y, X_1 verb X_2*; or *In Y packed to X*. Overall, this cluster covers less than 7% of the part–whole patterns discovered. Thus, for the purpose of this research we considered only the first three clusters of lexico-syntactic patterns expressing meronymy.

Table 2
Number of sentences and patterns containing the 100 part–whole pairs in each text collection considered.

Collection	Number of sentences	Number of sentences containing the pairs	Number of sentences containing part–whole relations	Number of patterns
SemCor	10,000	87	48	12
LA Times	10,000	1,988	487	30

Table 3
Clusters of lexico-syntactic patterns classified based on their semantic similarity and their frequency of occurrence in the 20,000 sentence corpus used in the part–whole pattern identification procedure.

Cluster	Patterns	Freq.	Coverage	Examples
C1. genitives and verb *to have*	NP_X *of* NP_Y NP_Y's NP_X NP_Y have NP_X	282	52.71%	eyes *of* the baby girl*'s* mouth The table *has* four legs.
C2. noun compounds	NP_{XY} NP_{YX}	86	16.07%	door knob turkey pie
C3. preposition	NP_Y PP_X NP_X PP_Y	133	24.86%	A bird *without* wings cannot fly. A room *in* the house.
C4. other	others	34	6.36%	The Supreme Court *is a branch of* the Government.

This pattern classification criterion is justified, in part, by our desire to verify whether or not the automatic approach proposed here is generally applicable not only for the genitive cluster patterns (cluster 1) (Girju, Badulescu, and Moldovan 2003), but also for more complex types, such as noun compounds (cluster 2) and prepositional constructions (cluster 3). Our intuition that the proposed patterns have different semantic behavior, and thus have to be treated separately in distinct clusters, is partially justified by a linguistic analysis summarized in Section 3.3 and supported by our empirical results from Section 5.3. In the remainder of the paper, we refer to these clusters as the genitives (cluster 1), noun compounds (cluster 2), preposition (cluster 3), and other (cluster 4) clusters.

We also noticed that some patterns, such as the genitive and preposition clusters, prefer the part and the whole in a certain position. For example, in *of*–genitives the part is mostly in the first position (modifier), and the whole in the second (head) (e.g., *door of the car*), while in *s*–genitives the positions are reversed (e.g., *car's door*). The verb *to have* requires parts in the second position, while noun compounds have a preference for them in the second position (e.g., *car has door* and *car door*, respectively). In the preposition cluster patterns, for the preposition *in* the part is usually in the first position (e.g., *door in the car*) and for the preposition *with* the positions are reversed (e.g., *car with four doors*).

However, there are also exceptions. For instance, in some *of*–genitives the part can occupy the second position (e.g., *flock of birds*) and in some noun compounds it can be present in the first position (e.g., *ham sandwich*). In the corpus used for pattern identification these exceptions are rare. Therefore, we will not consider the patterns NP_Y *of* NP_X and NP_{XY} in our experiments. If such examples are encountered, the part and the whole concepts are wrongly identified, representing one source of errors.

Berland and Charniak (1999) also used Hearst's algorithm to find part–whole patterns. However, they focused only on the first five patterns that occur frequently in their corpus. These patterns are subsumed by our clusters as shown in Table 4. They noticed that the last three patterns are ambiguous and decided to use only the first two in their experiments.

3.3 The Ambiguity of Part–Whole Lexico-Syntactic Patterns

From the list of lexico-syntactic patterns thus extracted, we noticed that some of these part–whole constructions always refer to meronymy, but most of them are ambiguous,

Table 4
The patterns used by Berland and Charniak and the corresponding cluster patterns used by us.

Berland and Charniak patterns	Our cluster patterns	Example
NN_{whole} *'s* NN_{part}	NP_Y *'s* NP_X	girl*'s* mouth
NN_{part} *of* (the|a) (JJ|NN) NN_{whole}	NP_X *of* NP_Y	eyes *of* the baby
NN_{part} *in* (the|a) (JJ|NN) NN_{whole}	NP_X PP_Y	ball *in* red box
NN_{parts} *of* NN_{wholes}	NP_X *of* NP_Y	doors *of* cars
NN_{parts} *in* NN_{wholes}	NP_X PP_Y	quotations *in* articles

in the sense that they express a part–whole relation only in some particular contexts and only between specific pairs of nouns. For example, NP_1 *is member of* NP_2 always refers to meronymy, but this is not true for NP_1 *has* NP_2. In most cases, the verb *to have* has the sense of *to possess*, and only in some particular contexts refers to meronymy.

Table 5 presents a summary of some of the most frequent part–whole lexico-syntactic patterns we observed, classified based on their ambiguity.

Below we discuss further the ambiguities encountered in the patterns of the first three clusters.

The Semantic Ambiguity of Genitive Constructions
In English there are two kinds of genitives: the *s*-genitive and the of-*genitive*. A characteristic of the genitives is that they are very ambiguous, as the constructions can be given various interpretations (Moldovan and Badulescu 2005). For instance, genitives can encode relations such as PART–WHOLE (*Mary's hand*), POSSESSION (*Mary's car*), KINSHIP (*Mary's sister*), PROPERTY/ATTRIBUTE HOLDER (*Mary's beauty*), DEPICTION–DEPICTED (*Mary's painting* — if it depicts her), SOURCE-FROM (*Mary's birth city*), or

Table 5
Examples of meronymic expressions based on their ambiguity.

Types of Part–Whole Expressions	Positive Examples (part–whole)	Negative Examples (not part–whole)
Unambiguous	The *parts of* an airplane *include* the engine, ..	
	The substance *consists of* three ingredients.	
	One of the air's *constituents* is oxygen.	
	The cloud was *made of* dust.	
	Iceland is a *member of* NATO.	
Ambiguous	The horn is *part of* the car.	He is *part of* the game (PARTICIPANT–EVENT)
	The table *has* four legs.	Kate *has* four cats. (POSSESSION)
	The girl*'s* mouth is sensual.	Mary*'s* brother is cute. (KINSHIP)
	The eyes *of* the baby are blue.	The dress *of* my niece is blue. (POSSESSION)
	Each *door knob* was made of silver.	Dallas is a modern *Texas city*. (LOCATION)
	It was the girl *with* blue eyes.	The woman *with* triplets received a lot of attention. (KINSHIP)

MAKE-PRODUCE (*Mary's novel* — if Mary wrote it). Thus, any attempt to interpret genitive constructions has to deal with the semantic analysis of the two noun constituents. Sometimes world knowledge or more contextual information is necessary to identify the correct semantic relation (e.g., *Mary's novel* might mean the novel written by Mary, read by Mary, or dreamed about by Mary).

The Semantic Ambiguity of the Verb *To Have*

According to WordNet 1.7, the verb *to have* in transitive constructions has 21 different senses, such as to *possess, feature, need, get, undergo, be confronted with, accept, suffer from*, and many others. Although the senses enumerated in WordNet represent a rather disparate set with no well defined semantic connection among them, the verb *to have* can participate in many different semantic structures and has been studied extensively in the linguistics community (Freeze 1992; Schafer 1995; Jensen and Vikner 1996).

The semantic relations encoded by the verb *to have* are quite similar to those realized by genitive constructions. Some researchers (Jensen and Vikner 1996) offered a detailed analysis for the purpose of capturing the most important semantic features of the verb *to have*. Their hypothesis is based on the idea that, semantically, the verb *to have* has a sense of its own derived from the semantic interpretation of the close context or the sentence in which it occurs. Let's consider the following sentences: (a) *Kate has a sister* (KINSHIP), (b) *Kate has a cat* (POSSESSION), and (c) *Kate has green eyes* (PART–WHOLE). The meaning of the verb *to have* in these situations is derived from the semantic information encoded in both the subject and the object.

The Semantic Ambiguity of Noun Compounds

Noun compounds (NCs) are noun sequences of the type N_1 N_2 .. N_n that have a particular meaning as a whole. NCs have been studied intensively in linguistics (Levi 1978), psycholinguistics (Downing 1977), and computational linguistics (Spärck Jones 1983; Lauer and Dras 1994; Rosario and Hearst 2001) for a long time. The interpretation of NCs focuses on the detection and classification of a comprehensive set of semantic relations between the noun constituents. This task has proved to be very difficult due to the complex semantic aspect of noun compounds:

1. NCs have implicit semantic relations: for example, *spoon handle* (PART–WHOLE).
2. NCs' interpretation is knowledge intensive and can be idiosyncratic: For example, *GM car* (in order to correctly interpret this compound we have to know that GM is a car-producing company).
3. There can be many possible semantic relations between a given pair of word constituents. For example, *linen bag* can mean *bag made of linen* (PART–WHOLE), as well as *bag for linen* (PURPOSE).
4. Interpretation of NCs can be highly context-dependent. For example, *apple juice seat* can be defined as "*seat with apple juice on the table in front of it*" (Downing 1977).

The Semantic Ambiguity of Prepositional Constructions

In English and various other natural languages, prepositions play a very important role both syntactically and semantically in the phrases, clauses, and sentences in which they occur. Semantically speaking, prepositional constructions can encode various semantic relations, their interpretations being provided most of the time by the underlying context. For instance, in the following examples the preposition *with* encodes different semantic relations: (a) *It was the girl* with *blue eyes* (MERONYMY), (b) *The baby* with *the red ribbon is cute* (POSSESSION), and (c) *The woman* with *triplets received a lot of attention* (KINSHIP).

The variety and ambiguity of these constructions show the complexity and importance of our task. We have seen that the interpretation of these constructions depends heavily on the meaning of the two noun constituents. To get the meaning of the nouns we rely on a word sense disambiguation system that takes into consideration surrounding contexts of the words.

4. A Machine Learning Algorithm for Automatic Discovery of Classification Rules

4.1 Approach

In this section we propose a method for the automatic discovery of rules that discriminate whether or not a selected pattern instance is meronymic. First a corpus is prepared and patterns from clusters C1–C3 are identified. The approach relies on the assumption that the semantic relation between two noun constituents representing the part and the whole can be detected based on nouns' semantic features.

This procedure applies to ambiguous constructions. The unambiguous constructions don't have to be processed since they lead unmistakably to part–whole relations.

The system learns automatically classification rules that check semantic features of noun constituents. The classification rules are learned through an *iterative semantic specialization* (ISS) procedure applied on the noun constituents' semantic features provided by the WordNet lexical knowledge base (Fellbaum 1998). ISS starts by mapping the training noun pairs to the corresponding top WordNet noun concepts using hypernymy chains. Then, it builds a learning tree by recursively splitting the training corpus into unambiguous and ambiguous examples based on the semantic information provided by the WordNet noun hierarchies. The learning tree is built top-down, one level at a time, each level corresponding to a specialization iteration. The internal nodes represent sets of ambiguous examples at various levels of specialization, while the leaves contain unambiguous examples. The ambiguous examples are further specialized with next-level WordNet concepts. The process is repeated recursively until there are no more ambiguous examples. For each set of unambiguous positive and negative examples at each level in the downward descent, we apply Quinlan's C4.5 algorithm and learn classification rules of the form *if X is/is not of a WordNet semantic class A and Y is/is not of WordNet semantic class B, then the instance is/is not a part–whole relation*.

4.2 Preprocessing Part–Whole Lexico-Syntactic Patterns

Since our discovery procedure is based on the semantic information provided by WordNet, we need to preprocess the noun phrases (NPs) extracted by the three clusters considered and identify the potential part and the whole concepts. For each NP we keep only the largest sequence of words (from left to right) defined in WordNet. For example, from the noun phrase *brown carving knife* the procedure retains only *carving knife*, since this concept is defined in WordNet. For each such sequence of words, we manually annotate it with its WordNet sense in context. For the example above we annotated the noun phrase with sense #1 (*carving knife#1*), since in that context it had sense #1 in

WordNet (for this concept WordNet lists only one sense, defined as "a large knife used to carve cooked meat"). Table 6 shows a few examples of patterns from different clusters and the results of this preprocessing step.

4.3 Building the Training Corpus

In order to learn the classification rules, we used the SemCor 1.7 and TREC 9 text collections, and the part–whole information provided by WordNet. From the SemCor collection we selected 19,000 sentences. Another 100,000 sentences were randomly extracted from the LA Times articles of TREC 9. As SemCor 1.7 is already annotated with part-of-speech tags and WordNet senses, we part-of-speech tagged only the LA Times collection using Brill's tagger (1995). A corpus "A" was thus created from the selected sentences of each text collection. Each sentence in this corpus was then parsed using the syntactic parser developed by Charniak (2000). Focusing only on sentences containing the lexico-syntactic patterns in each cluster C1–C3, we manually annotated nouns in the patterns with their corresponding WordNet senses (with the exception of those from SemCor), as shown in Section 4.2, and marked all candidate instances that encoded a part–whole relation as positives, and negatives otherwise. In the corpus, 66% of the annotated instances were PART-OF relations, 14% STUFF-OF, and 20% MEMBER-OF.

Moreover, WordNet 1.7 contains 27,636 part–whole relations linking various noun concepts. As this information is very valuable for training purposes, we tried to see which of the selected patterns match these pairs. For each WordNet part–whole pair, we formed inflected queries (to capture singular and plural instances) and searched the Web, the largest on-line general purpose text collection, using Altavista. From the first 100 retrieved documents, we selected and syntactically parsed only those sentences containing pairs within cluster patterns. We manually validated those instances and registered which cluster(s) of patterns could extract the pair. All these sentences formed a second corpus, corpus "B". For instance, for the pair *door#4–car#1* we searched Altavista for documents containing both words *car* and *door*. Then, we retrieved all the sentences that contained the two words in at least one of the target patterns. As a result, we obtained sentences containing the pair of words linked by patterns such as *door of car*, *car's door*, *car has door*, *car with four doors*, *car door*, etc.

Overall, the 27,636 WordNet pairs were linked by the genitive cluster patterns, while the noun compound and preposition clusters extracted only some subsets of these pairs. Some part–whole pairs were linked by patterns that belong to more than one cluster. For instance, *door knob* is a pair that usually belongs to the noun compound cluster, but it can also be selected by the genitive cluster (e.g., *knob of the door*) and the preposition cluster (e.g., *the door with the iron knob*).

Table 6
Examples of identifying the potential Part and Whole concepts for different clusters.

Cluster	Example	Potential Part(X) and Whole(Y) concepts	Positive or negative example
C1. genitives	the door of the car	the $[door\#4]_X$ of the $[car\#1]_Y$	positive
	my friend's car	my $[friend\#1]_Y$'s $[car\#1]_X$	negative
C2. noun compounds	car door company	$[car\_door\#1]_X$ $[company\#1]_Y$	negative
		$[car\#1]_X$ $[door\#4]_Y$	positive
C3. prepositions	window from the car	$[window\#2]_X$ from the $[car\#1]_Y$	positive

Corpus “B” was used only to convince us that the part–whole pairs selected from WordNet were representative, ie., present in the patterns considered. Indeed corpus “B” pairs were found in at least one of the cluster patterns. While corpus ”A” consists of positive and negative examples from LA Times and SemCor collections, corpus ”B” contains only positive instances as they are WordNet part–whole pair concepts. Moreover, although corpus ”B” has a different distribution than corpus ”A”, the noun pairs from WordNet are general-purpose and always encode a part–whole relation.

Table 7 shows the statistics for the positive and negative training examples for each cluster. In the genitive cluster, for example, there were 18,936 such pattern instances, of which 325 encoded part–whole relations, while 18,611 did not. Thus, for the genitive cluster we used a training corpus of 27,961 positive examples (325 pairs of concepts in a part–whole relation extracted from corpus “A” and 27,636 extracted from WordNet as selected pairs) and 18,611 negative examples (the non-part–whole relations extracted from corpus “A”).

4.4 Inter-Annotator Agreement

The part–whole relation discovery procedure proposed in this paper was trained and tested on a large corpus of human annotated examples (a part of the LA Times collection for both training and testing, and a part of the Wall Street Journal (WSJ) collection for testing). The annotators, two researchers in computational semantics, decided whether an example pair encoded a part–whole relation or not. The examples were disambiguated in context: the annotators were given the pairs and the sentence in which they occurred. The two annotators' task was to determine the correct senses of the two noun constituents and then decide if the relation is meronymic or not. A third researcher decided on the non-agreed word senses and relations. The annotators were also provided with the list of subtypes of meronymy relations proposed by (Winston, Chaffin, and Hermann 1987) as a guideline for detecting part–whole relations. If an example contained one of the six meronymy subtypes, the annotators tagged that example as positive (part–whole); otherwise they tagged it as a negative example.

The annotators' agreement was measured using the kappa statistic (Siegel and Castellan 1988), one of the most frequently used measures of inter-annotator agreement for classification tasks:

$$K = \frac{Pr(A) - Pr(E)}{1 - Pr(E)}, \tag{1}$$

where $Pr(A)$ is the proportion of times the raters agree and $Pr(E)$ is the probability of agreement by chance.

Table 7
Training corpora statistics for each of the three clusters considered.

Cluster	Positive examples: from WordNet as evidenced by corpus “B”	Positive examples: from Corpus “A”	Negative examples: from Corpus “A”
C1. genitives	27,636	325	18,611
C2. noun compounds	142	625	6,601
C3. prepositions	111	295	2,751

The K coefficient is 1 if there is total agreement among the annotators, and 0 if there is no agreement other than that expected to occur by chance. This coefficient measures how well annotators agree at identifying both positive and negative instances of meronymic relations.

Table 8 shows the inter-annotator agreement on the part–whole classification task for each of the three clusters considered in both training and test phases of the part–whole relation discovery procedure.

On average, the K coefficient is close to 0.85, showing a good level of agreement, for all clusters in the training and test data. This can be explained by the instructions the annotators received prior to annotation and by their expertise in lexical semantics. The results also show that even for more productive genitive and noun compound examples, the sentence-level context was enough to disambiguate the examples most of the time.

4.5 Iterative Semantic Specialization (ISS) Learning

Iterative Semantic Specialization Learning is an iterative process that learns a decision tree and classification rules by mapping the semantic features of the noun pairs to the WordNet noun hierarchies. The procedure starts with a generalized version of the training examples as pairs of top WordNet noun concepts using hypernymy chains. The examples are then split into unambiguous and ambiguous. The ambiguous examples are further specialized with next-level WordNet concepts. The process is repeated recursively until there are no more ambiguous examples. For each set of unambiguous positive and negative examples at each level in the downward descent, we apply Quinlan's C4.5 algorithm and learn classification rules. As will be shown in Section 5.3, the algorithm is applied separately to each of the three clusters considered for optimal results.

The Iterative Semantic Specialization (ISS) Learning Algorithm

Input: Positive and negative meronymic examples of pairs of concepts. The concepts are WordNet words semantically disambiguated in context (tagged with their corresponding WordNet senses).

Output: Classification rules in the form of semantic selectional restrictions on the *modifier* and *head* concepts using WordNet IS–A hierarchy information.

Table 8
The inter-annotator agreement on the part–whole classification task for each of the three clusters considered in both training and test phases of the part–whole relation discovery procedure.

Corpus	Cluster	Kappa agreement
LA Times	genitives	0.878
(training and testing)	noun compounds	0.826
	prepositions	0.811
	genitives	0.880
WSJ (testing)	noun compounds	0.862
	prepositions	0.836

Step 1. Generalizing the training examples
Initially, the training corpus consists of examples that have the format ⟨part#sense; whole#sense; target⟩, where target can be either *Yes* or *No*, depending whether the relation between the part and whole is meronymy or not: for example, ⟨aria#1, opera#1, Yes⟩. From this initial set of examples an intermediate corpus was created by expanding each example using the following format:

⟨part#sense, class_part#sense, whole#sense, class_whole#sense; target⟩,

where class_part and class_whole correspond to the WordNet top semantic classes of the part and whole concepts, respectively. For instance, the previous example becomes ⟨aria#1, entity#1, opera#1, abstraction#6, Yes⟩.

From this intermediate corpus a generalized set of training examples is built, retaining only the semantic classes and the target value. At this point, the generalized training corpus contains three types of examples:

1. Positive examples ⟨X_hierarchy#sense, Y_hierarchy#sense, Yes⟩
2. Negative examples ⟨X_hierarchy#sense, Y_hierarchy#sense, No⟩
3. Ambiguous examples ⟨X_hierarchy#sense, Y_hierarchy#sense, Yes/No⟩

The third situation occurs when the training corpus contains both positive and negative examples for the same hierarchy types. For example, both the relationships ⟨apartment#1, woman#1, No⟩ and ⟨hand#1, woman#1, Yes⟩ are mapped into the more general type ⟨entity#1, entity#1, Yes/No⟩. However, the first example is negative (a POSSESSION relation), while the second one is a positive example.

Step 2. Learning classification rules for unambiguous examples
For the unambiguous examples in the generalized training corpus (those that are either positive or negative), rules are determined using C4.5. In this context, the features are the components of the relation (the part and, respectively the whole) and the values of the features are the corresponding WordNet semantic classes (the furthest ancestor in WordNet of the corresponding concept).

With the first two types of examples, the unambiguous ones, a new training corpus was created on which we applied C4.5 using a 10-fold cross validation. The corpus is split in ten permutations, 9/10 training and 1/10 testing, and the output is represented by 10 sets of rules and default values generated from these unambiguous examples.

The rules obtained are if–then rules with the part and whole noun semantic senses as preconditions. The default value is the most probable value for the target value and is used to classify unseen instances of that type when no other rule applies. It can be either Yes or No, corresponding to the possible values of the target attribute (part–whole relation or not).

The rules were ranked according to their frequency of occurrence and average accuracy obtained for each particular set. In order to use the best rules, we decided to keep only those that had a frequency above a threshold (occurring in at least 7 of the 10 sets of rules) and an average accuracy greater than or equal to 50%.

In order to minimize the redundancies that may occur during the learning process, rules with the same classification value as the default value are ignored. The idea is that the default rule incorporates all the rules with the same target value.

For instance, after running C4.5 on the unambiguous set for the *abstraction#6–abstraction#6* example, we obtained a list of five rules and a default value No, as shown in Table 9. Rules 1 and 5 were discarded as they were incorporated into the default class. Rules 3 and 4 were also discarded as their frequency did not pass the threshold of 7. Thus, rule 2 remains the only applicable rule.

After filtering the rules that have the default value or do not pass the frequency and accuracy thresholds, there might be cases in which the set of remaining rules is empty.

Step 3. Specializing ambiguous examples
Since C4.5 cannot be applied to ambiguous examples, we recursively specialize them to eliminate the ambiguity. The specialization procedure is based on the IS–A information provided by WordNet. Initially, each semantic class represented the root of one of the noun hierarchies in WordNet. By specialization, the semantic class is replaced with the corresponding hyponym for that particular sense, i.e., the concept immediately below in the hierarchy.

For this task, we again considered the intermediate training corpus of examples.

For instance, the examples ⟨leg#2, entity#1, bee#1, entity#1, Yes⟩ and ⟨beehive#1, entity#1, bee#1, entity#1, No⟩ that caused the ambiguity ⟨entity#1, entity#1, Yes/No⟩, were replaced with ⟨leg#2, thing#12, bee#1, organism#1, Yes⟩ and ⟨beehive#1, object#1, bee#1, organism#1, No⟩, respectively. This intermediate example is thus generalized in the less ambiguous examples ⟨thing#12, organism#1, Yes⟩ and ⟨object#1, organism#1, No⟩. This way, we specialize the ambiguous examples with more specific values for the attributes. The specialization process for this particular example is shown in Figure 1.

Although this specialization procedure eliminates a proportion of the ambiguous examples, there is no guarantee it will work for all the ambiguous examples of this type. This is because the specialization splits the initial hierarchy into smaller distinct subhierarchies, with the examples distributed over this new set of subhierarchies. For the examples described above, the procedure eliminates the ambiguity through specialization of the semantic classes into new ones: *thing#12–organism#1* and *object#1–organism#1*.

However, not all the examples can be disambiguated after only one specialization. For the examples ⟨leg#2, bee#1, Yes⟩ and ⟨world#7, bee#1, No⟩, the procedure generalizes *abstraction#6–abstraction#6* into the ambiguous example ⟨entity#1, entity#1, Yes/No⟩ and then specializes it in the ambiguous example ⟨part#7, organism#1, Yes/No⟩. After one specialization the ambiguity still remains.

Steps 2 and 3 are repeated until there are no more ambiguous examples. The general architecture of this procedure is shown in Figure 2.

Table 9
The list of rules for the iteration generated by the unambiguous subset of the ambiguous example ⟨abstraction#6, abstraction#6, yes/no⟩. 'Yes' means part–whole relation, while 'No' means non-part–whole relation. The global default target value of this unambiguous node is *No*. Note that rules 3 and 4 are discarded as their frequency is below 7, and rules 1 and 5 were also discarded as incorporated in the default class No.

Rule no.	Part Class	Whole Class	Target	Accuracy value	Frequency
1	measure#3	abstraction#6	*No*	92.51	9
2	time#5	abstraction#6	Yes	79.21	9
3	abstraction#6	time#5	Yes	85.70	1
4	abstraction#6	measure#3	Yes	63.00	1
5	abstraction#6	attribute#2	*No*	93.00	1
Default			No		

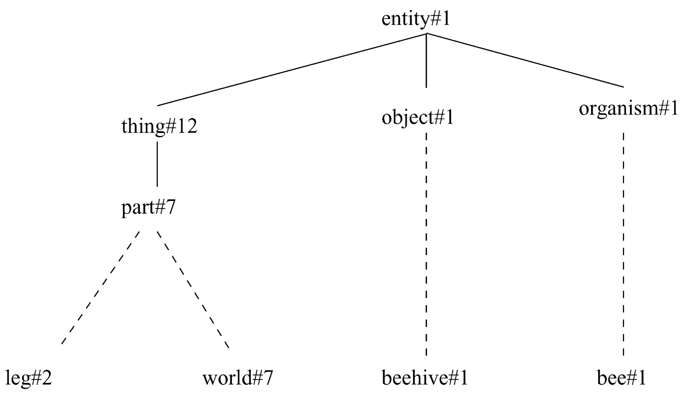

Figure 1
The specialization of examples ⟨leg#2, entity#1, bee#1, entity#1, Yes⟩, ⟨beehive#1, entity#1, bee#1, entity#1, No⟩, and ⟨world#7, entity#1, bee#1, entity#1, No⟩ with the corresponding WordNet semantic classes.

We observed that after the first generalization, 99.72% of the examples were ambiguous. After each specialization, the percentage decreases. For instance, after one level of specialization, 97.36% of the examples for *entity#1–entity#1*, 96.05% for *abstraction#6–abstraction#6*, and 97.56% for *entity#1–group#1* were ambiguous.

Table 10 presents a sample of the iterations produced by the program to specialize the genitive cluster ambiguous example *abstraction#6–abstraction#6*. Each indentation corresponds to a specialization iteration.

The training corpus considered for this research required on average 2.5 and at most five levels of specialization.

The next section describes the construction of classification rules, the experiments, and the results obtained.

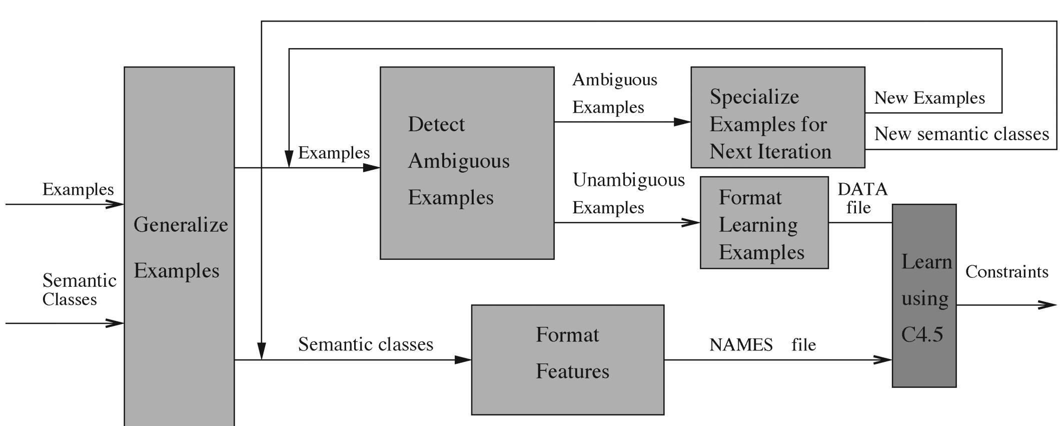

Figure 2
Diagram of the ISS system.

Table 10
A sample iteration produced by the ISS procedure for the genitive cluster *abstraction#6–abstraction#6* ambiguous example. The italicized examples are unambiguous.

abstraction#6–abstraction#6
    attribute#2–attribute#2
    attribute#2–measure#3
    attribute#2–relation#1
    relation#1–attribute#2
    measure#3–measure#3
    measure#3–relation#1
    time#5–time#5
    relation#1–time#5
    relation#1–relation#1
        magnitude_relation#1–magnitude_relation#1
        part#1–part#1
        social_relation#1–part#1
            communication#2–language_unit#1
        social_relation#1–social_relation#1
            communication#2–communication#2
                signal#1–message#2
                signal#1–written_communication#1
                written_communication#1–written_communication#1
                    writing#2–writing#2
                message#2–written_communication#1

5. Formulating Classification Rules and Applying them to Discover Part–Whole Relations

5.1 Building the Learning Tree

The ISS learning procedure presented in the previous section builds a learning tree by recursively splitting the training corpus into unambiguous and ambiguous examples, based on the semantic information provided by the WordNet noun hierarchies. The learning tree is built top-down, one level at a time, each level corresponding to a specialization iteration. The internal nodes represent ambiguous examples at various levels of specialization, while the leaves contain sets of unambiguous examples. For instance, Figure 3 shows the learning tree corresponding to the specialization from Table 10.

Initially, the learning tree contains only a dummy root node that provides no information. After the generalization done in step 1 of the ISS learning procedure, all the initial examples are mapped into corresponding pairs of top noun semantic classes in WordNet and split into unambiguous and ambiguous sets based on their target function. All these new sets of examples form the first level of the learning tree.

The learning tree has two types of nodes: unambiguous nodes, corresponding to the sets of unambiguous examples from each iteration (e.g., nodes 1.1, 1.3.1, and 1.4.1 from Figure 3) and ambiguous nodes, corresponding to each ambiguous example from each iteration (e.g., nodes 1.2, 1.3, 1.4, and 1.4.2 from Figure 3). Each node has associated with it a pair {R, D} representing a set of rules and a default value. The set of rules represents the rules to be used for classifying the new instances and the default value represents the target value (Yes if an instance is a part–whole relation and No if the instance is

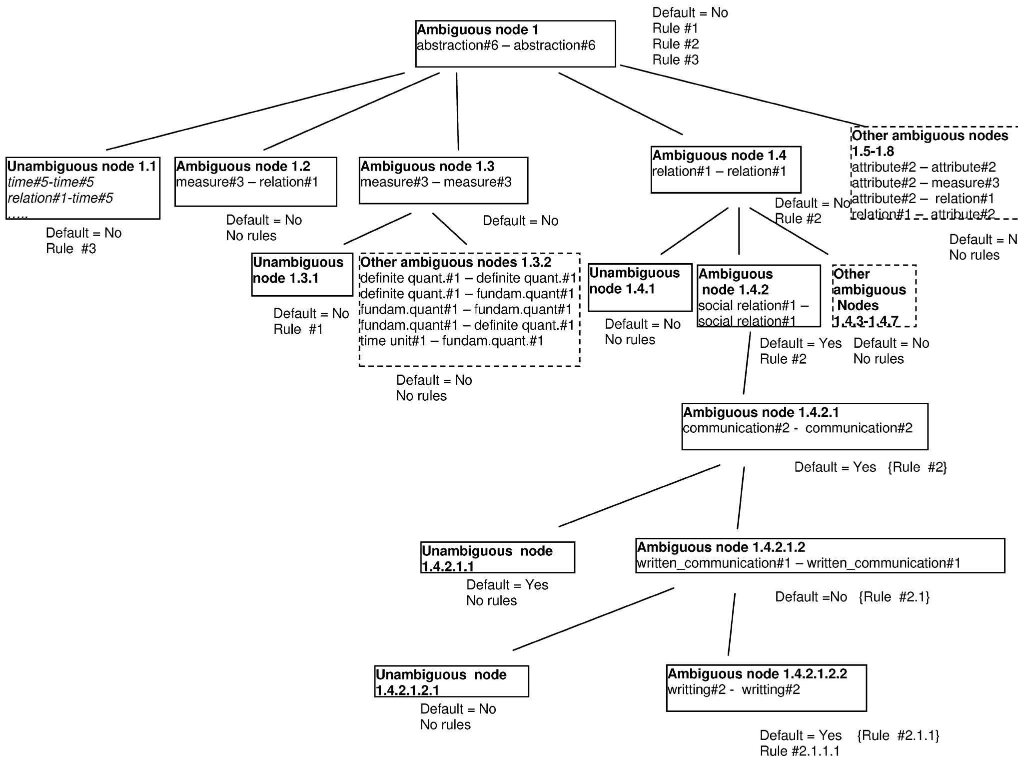

Figure 3
A snapshot of the learning subtree *abstraction#6–abstraction#6* on which the combination and propagation algorithm is exemplified. Each node has an associated set of rules and a default value. The rule number references are for the "No." column from Table 11.

not a part–whole relation) that should be returned if none of the rules classify the new instances.

After learning the classification rules in Step 2 of the ISS procedure, all the unambiguous nodes have default values and some have rules.

5.2 Formulating the Classification Rules

In order to generate an overall set of classification rules, we traverse the learning tree in a bottom-up fashion, applying the rules generated at each level in this order. The rationale of this approach is that the rules closer to the bottom are more specific, and thus more accurate. At each level, the idea is to combine the rules associated with each sibling node and propagate the result to the parent. The combination and propagation steps are applied recursively until the root is reached. The combination phase guarantees that the rules to be combined are applied in a particular order at each level.

Figure 4 shows a typical tree corresponding to one iteration of the ISS procedure on which we will explain the combination and propagation algorithm. Node L represents an internal node containing an ambiguous example. Through specialization, the learning procedure generated a set of unambiguous examples represented by the leaf L_U, and a sequence of n ambiguous examples represented by the internal nodes L_{A_1}, L_{A_2}, .. L_{A_n}.

Table 11
The rules and default value learned for the genitive cluster for the *abstraction#6–abstraction#6* ambiguous example. "Val." is the target value, "Acc." is the rules' accuracy, and "Fr." is their occurrence frequency. The numbering style used in the "No." column is intended to indicate rules at different specialization levels.

No.	Part Class	Whole Class	Val.	Acc.	Fr.	Example
	abstraction#6	**abstraction#6**	No			*glory#2–past#1*
1	linear_measure#3	measure#3	Yes	63	9	*centimeter#1–decimeter#1*
2	communication#2	communication#2	Yes			*act#3–play#1*
2.1	written_comm.#1	written_comm.#1	No			*text#1–act#3*
2.1.1	writing#2	writing#2	Yes			*New_Testament#1–Bible#1*
2.1.1.1	matter#6		No	79.98	9	*text#1–act#3*
2.2	indication#1	message#2	No	73.25	10	*copy#1–recommendation#1*
2.3	message#2	communication#2	No	79.72	8	*irony#1–play#1*
3	time#5	abstraction#6	Yes	79.21	9	*carboniferous#1–paleozoic#1*
Default			No			*glory#2–past#1*

The values associated with the ambiguous nodes (rules and default values) are generated through propagation from lower levels.

Rule combination and propagation algorithm:

Input: Pairs of rules and associated default values for each unambiguous and ambiguous node: $\{R_U, D_U\}, \{R_{A_1}, D_{A_1}\}, \{R_{A_2}, D_{A_2}\}, ..\{R_{A_n}, D_{A_n}\}$;

Output: A pair of rules and default value for parent node: $\{R_L, D_L\}$.

Step 1. Propagating the default value to the parent node: $D_L \leftarrow D_U$

The default value of the unambiguous examples (D_U) will be directly propagated to the parent as the global default value of the subtree L (D_L). For example, the default value for the unambiguous node 1.1 from Figure 3 is No and it will propagate to the parent node *abstraction#6–abstraction#6* (node 1 in Figure 3).

If there is no unambiguous node L_U (and therefore default value D_U), the default value for the first ambiguous example is propagated to L. For instance, for the ambiguous node 1.4.2 (*social_relation#1–social_relation#1*), there were no unambiguous ex-

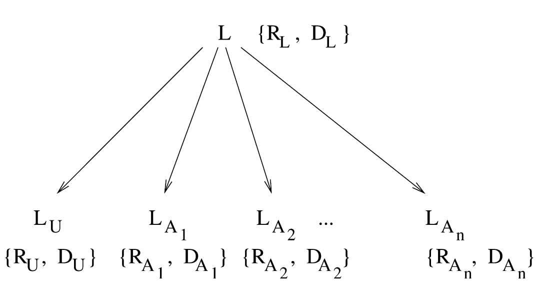

Figure 4
A part of the learning tree generated by the ISS learning procedure. The pairs of rules and default value associated with the parent node are generated through propagation of the combined pairs of rules and default values of the children.

amples; and therefore the default value from the node 1.4.2.1 (*written_ communication#2–written_ communication#2*) will be used.

Step 2. Propagating the rules from an ambiguous node with the same default value to the parent node: $R_L \leftarrow \{R_{A_i} | D_{A_i} = D_L, 1 \leq i \leq n\}$

The ambiguous nodes are the first to be tested. All the rules associated with the ambiguous nodes having the same default value as the global one are applied with the highest priority. For instance, all the ambiguous nodes for *abstraction#6–abstraction#6* (nodes 1.2–1.8 in Figure 3) received a default value of No through propagation from their descendents. Since the default value for this node is No, it will receive all their rules (Rules 1 and 2 from Table 12).

Step 3. Propagating the rules from an ambiguous node with the opposite default value to the parent node: $R_L \leftarrow R_L \cup \{\textit{if } A_j \textit{ then } R_{A_j} \cup D_{A_j} | D_{A_j} \neq D_L, 1 \leq j \leq n\}$

The remaining ambiguous nodes have associated with them a different default value (a non-default value). Since the two nodes have opposite default values, the default value (D_{A_j}) needs to be used when the rules for the child node (A_j) do not hold. Therefore, a new rule, specific to the example A_j, needs to be created, for handling all the instances of A_j: *if* A_j *then* $R_{A_j} \cup D_{A_j}$.

For example, the ambiguous node 1.4.2.1.2 (*written_communication#1–written_ communication#1*) has the default value No. Its only ambiguous node (node 1.4.2.1.2.2 : *writing#2–writing#2*) has the default value Yes. Therefore, a specific rule (Rule 2.1.1 from Table 12) needs to be created for the example (*Part=writing#2 and Whole=writing#2*),

Table 12
The list of rules obtained for the ambiguous example *abstraction#6–abstraction#6* for the genitive cluster.

1	if Part is linear_measure#3 and Whole_Class is measure#3
	then It is a part–whole relation
2	if Part is communication#2 and Whole is communication#2
	then
2.1	if Part is written_communication#1 and Whole is written_communication#1
	then
2.1.1	if Part is writing#2 and Whole is writing#2
	then
2.1.1.1	if Part is matter#6
	then It is not a part–whole relation
	else It is a part–whole relation
	else It is not a part–whole relation
	else
2.2	if Part is indication#1 and Whole is message#2
	then It is not a part–whole relation
	else
2.3	if Part is message#2 and Whole is communication#2
	then It is not a part–whole relation
	else It is a part–whole relation
3	if Part is time#5 and Whole is abstraction#6
	then It is a part–whole relation

that applies its rule (Rule 2.1.1.1: *if Part=matter#6 then No*) and returns its default value (a non-default value) for the other cases:

if Part=writing#2 and Whole=writing#2	- the ambiguous example A_j
then	
if Part=matter#6 then No (Is not part–whole)	- the rules R_{A_j}
else Yes (Is part–whole)	- the non-default value D_{A_j}

Step 4. Propagating the rules learned from an unambiguous node to the parent node: $R_L \leftarrow R_L \cup R_U$

Last, the rules learned from the unambiguous examples propagate to the parent node. They are applied last, since they are more general than the other rules. For example, after running C4.5 on the unambiguous set for the *abstraction#6–abstraction#6* ambiguous example (Node 1 in Figure 3), and eliminating the non-satisfactory rules (see Table 9), we obtained only Rule 3: *if Part is time#5 and Whole is abstraction#6 then Yes* and the default value No (see Table 11). The rule is propagated to the parent node *abstraction#6– abstraction#6* and applied last.

In the end, the rules learned from the unambiguous examples are propagated to the parent node *L*. The procedure repeats until the top node of the tree is reached. After the combination and propagation procedure finishes, the root node contains the complete set of rules. The default value is added as a last rule, for classifying the instances that are not captured by the rules.

A sample of the rules obtained using the ISS procedure for the genitive cluster is shown in Table 11 in the order in which they were applied and propagated to the *abstraction#6–abstraction#6* node. Table 12 shows a translation of these rules into if–then–else rules.

The meaning of a rule *Part_Class Whole_Class Val* is *if Part is Part_Class and Whole is Whole_Class, then It is a part–whole relation (Val. = Yes) or not (Val. = No)*. For example, Rule 1 is *if Part is a linear_measure#3 and Whole is a measure#3, then It is a part–whole relation*.

5.3 Classification Rules for Each Cluster

In this section we present the classification rules learned for each cluster using the ISS learning procedure. We also performed various experiments to study the similarities and differences among clusters, especially to determine whether or not the classification rules learned for a particular cluster can be applied with high accuracy to other clusters.

A. Experiments with the genitive cluster

The most frequently used set of part–whole lexico-syntactic patterns is represented by the genitive cluster. Tables 13 shows some of the classification rules learned for this cluster by the ISS learning procedure in the order provided by the combination and propagation algorithm. The full list of classification rules is shown in Tables 1 and 2 from Appendix B. The unambiguous set at level 1 of the learning tree did not generate any rules. The rule labeled Default in Table 13 shows the learning tree global default value (No). The tables of classification rules show only the frequency and accuracy of the rules generated at the unambiguous nodes.

Table 13
A sample of the rules learned for the genitive cluster. The full list is provided in Table 1, Appendix B. "Val." means target value (No or Yes), "Acc." is the rules' accuracy, and "Fr." is the frequency with which they occurred. The numbering style used in the "No." column is intended to indicate rules learned at different specialization levels.

No.	Part Class	Whole Class	Val.	Acc.	Fr.	Example
	abstraction#6	**abstraction#6**	No			*glory#2–past#1*
1	linear_measure#3	measure#3	Yes	63	9	*centimeter#1–decimeter#1*
	abstraction#6	**entity#1**	No			*age#1–earth#1*
4	shape#2	artifact#1	Yes			*point#8–knife#2*
4.1	shape#2	structure#1	No	67.62	10	*diameter#2–plug#1*
4.2	shape#2	surface#1	No	67.62	10	*square#1–pegboard#1*
	abstraction#6	**group#1**	No			*history#3–regiment#1*
9	abstraction#6	biological_group#1	Yes	92.44	10	*year#3–montia#1*
10	relation#1	arrangement#2	Yes	79.40	9	*medium_frequency#1–electromagnetic_spectrum#1*
	abstraction#6	**phenomenon#1**	No			*cause#2–death#2*
11	shape#2	physical_phenomenon#1	Yes			*dewdrop#1–dew#1*
	abstraction#6	**psychological_feature#1**	No			*amount#1–work#4*
12	measure#3	structure#3	Yes	95.64	10	*August#1–Gregorian_calendar#1*
	entity#1	**phenomenon#1**	No			*keeper#2–flame#1*
13	point#2	physical_phenomenon#1	Yes			*storm_center#3–storm#1*
14	object#1	process#2	Yes			*ferric_oxide#1–rust#3*
	phenomenon#1	**entity#1**	No			
18	process#2	organism#1	Yes			*meiosis#1–anapsid#1*

Table 13
(cont.)

No.	Part Class	Whole Class	Val.	Acc.	Fr.	Example
18.1	process#2	person#1	No	76.70	8	*growth#2–child#2*
	phenomenon#1	**phenomenon#1**	No			*influence#4–action#6*
19	natural_phenomenon#1	natural_phenomenon#1	Yes			*meteor#1–meteor_shower#1*
	possession#2	**entity#1**	No			*cost#1–home#2*
20	territory#2	entity#1	Yes	69.84	9	*united_states_virgin_islands#1–virgin_islands#1*
	psychological_feature#1	**psychological_feature#1**	No			
22	knowledge_domain#1	knowledge_domain#1	Yes			*agrology#1–agronomy#1*
23	**entity#1**	**entity#1**	Yes			*door#4–car#1*
23.1	causal_agent#1	causal_agent#1	No			*lethal_dose#1–opium#1*
23.2	causal_agent#1	location#1	No			*taxi_driver#1–Los_Angeles#1*
23.3	causal_agent#1	object#1	No			*dose#1–malathion#1*
23.4	point#2	body_of_water#1	No	94.46	10	*headwaters#1–nile#1*
23.5	region#1	body_of_water#1	No	91.79	10	*east_side#1–river#1*
23.6	line#11	region#3	No			*direction#1–park#1*
23.6.1	admin_district#1	admin_district#1	Yes			*Alaska#1–United_States#1*
23.	Default		Yes			*door#4–car#1*
26	**group#1**	**group#1**	Yes			*genus_amoeba#1–amoebida#1*
26.1	social_group#1	people#1	No			*dictatorship#1–proletariat#1*
26.2	group#1	people#1	No	83.86	8	*demi-monde#1–high_society#1*
26.3	arrangement#2	collection#1	No	82.22	10	*classification#2–family#4*
26.4	social_group#1	collection#1	No	82.22	10	*circle#2–law#2*
26.	Default		Yes			*genus_amoeba#1–amoebida#1*
	Default		No			

Overall, for the genitive cluster the ISS procedure obtained 27 complex sets of classification rules.

B. Experiments with the noun compound cluster

Taking into consideration the results already obtained for the genitive cluster, there are three possible approaches for detecting part–whole relations using the *Y X* and *X Y* patterns:

a. [C1] Use the classification rules obtained for the genitive cluster.

b. [C1 + C2] Determine new classification rules collectively for the genitive and noun compound clusters (*Y's X*; *X of Y*; *Y have X*; and *Y X*).

c. [C2] Determine classification rules only for the noun compound cluster (*Y X; X Y*).

Table 14 shows the results obtained for the noun compound cluster using these three approaches. As one can observe, the best approach is to use only the classification rules generated by the noun compound cluster training examples. The recall increases significantly when new classification rules are learned for both the genitive and noun compound clusters, while the precision jumps considerably when the classification rules are learned only from the noun compound cluster examples. These statistics indicate that the genitive and noun compound clusters encode different semantic information, and consequently should be treated separately.

Table 15 shows the classification rules learned only for the noun compound cluster.

C. Experiments with the preposition cluster

Taking into consideration the results obtained for the previous two clusters, there are five possible approaches for detecting part–whole relations using *X prep Y* and *Y prep X* patterns:

a. [C1] Use the classification rules obtained for the genitive cluster.

b. [C2] Use the classification rules obtained for the noun compound cluster.

c. [C1 + C3] Determine new classification rules for all the patterns in the genitive and preposition clusters (*Y's X*; *X of Y*; and *Y have X*; *Y prep X*; and *X prep Y*).

Table 14
The results obtained for each of the three approaches for the *Y X*; *X Y* patterns applied on the LA Times test corpus.

Results	Genitives (C1)	Genitives + Noun compounds (C1 + C2)	Noun compounds (C2)
Precision	48.43%	52.98%	79.02%
Recall for cluster	58.08%	73.46%	75.33%
F-measure	52.82%	61.56%	77.13%

Table 15
The semantic classification rules learned for the noun compound cluster. "Val." means target value (No or Yes), "Acc." is the rules' accuracy, and "Fr." is the frequency with which they occurred.

No.	Part Class	Whole Class	Val.	Acc.	Fr.	Example
	abstraction#6	**abstraction#6**	No			*art#4–advertising#1*
1	time_period#1	time_period#1	Yes			*afternoon#1–Wednesday#1*
2	message#2	written_communication#1	Yes			*index#4–back_matter#1*
3	written_communication#1	message#2	Yes			*zip_code#1–address#6*
	abstraction#6	**entity#1**	No			*address#1–restaurant#1*
4	communication#2	musical_composition#1	Yes	50	7	*lyric#1–ballad#1*
5	written_communication#1	creation#2	Yes	50	7	*zip_code#1–address#6*
	abstraction#6	**psychological_feature#1**	No			*theorem#1–decomposition#1*
6	attribute#2	information#3	Yes	50	7	*head#10–abscess#1*
	act#2	**group#1**	No			*consolidation#2–school#1*
7	act#2	people#1	Yes			*president#6–class#1*
	entity#1	**abstraction#6**	No			*book#1–recipe#1*
8	surface#1	communication#2	Yes	67.62	10	*head#27–coin#1*
8.1	horizontal_surface#1		No			*dais#1–medal#1*
	entity#1	**entity#1**	No			*advocate#1–child#1*
9	object#1	body_of_water#1	Yes			*water#1–pond#1*
10	covering#2	instrumentality#3	Yes	96.50	10	*roof#–car#1*
11	way#6	structure#1	Yes	90.66	10	*stairway#1–building#1*
12	opening#10	artifact#1	Yes	84.30	10	*window#2–bus#1*
13	covering#2	structure#1	Yes	82.22	10	*roof#1–building#1*
14	artifact#1	covering#2	Yes	67.09	10	*top#11–roof#1*
15	instrumentality#3	covering#2	Yes			*lock#1–lid#2*
16	instrumentality#3	instrumentality#3	Yes			*accelerator#1–car#1*
16.1	conveyance#3	instrumentality#3	No	93.53	10	*alarm#2–seismograph#1*
16.2	furnishings#1	instrumentality#3	No	83.56	10	*stand#4–magazine#1*
16.3	means#2	instrumentality#3	No	76.79	10	*magazine#1–telescope#1*

Table 15
(cont.)

No.	Part Class	Whole Class	Val.	Acc.	Fr.	Example
16.4	equipment#1	instrumentality#3	No	72.73	10	*recorder#1–pen#1*
17	region#1	location#1	Yes	67.62	10	*boundary#1–city#1*
18	region#3	district#1	Yes	50	8	*city#–California#*
19	region#1	organism#1	Yes	50	8	*crown#8–tree#1*
	entity#1	**group#1**	No			*mine#1–navy#1*
20	causal_agent#1	people#1	Yes	50	9	*administrator#1–school#1*
21	organism#1	people#1	Yes	50	9	*secretary#2–press#1*
22	organism#1	social_group#1	Yes			*chancellor#1–university#1*
23	plant#2	social_group#1	No	67.62	10	*rice#1–U.S.#1*
	group#1	**group#1**	No			*government#1–military#1*
24	social_group#1	set#5	Yes	50	8	*leader#1–party#1*
Default			No			*officer#1–narcotic#1*

Table 16
The results obtained for each of the five approaches for the *Y prep X* and *X prep Y* patterns in the preposition cluster applied on the LA Times test corpus. C1 refers to the genitive cluster, C2 to the noun compound cluster, and C3 to the preposition cluster.

Results	C1	C2	C1 + C3	C2 + C3	C3	C1 + C2 + C3
Precision	46.98%	61.54%	4.81%	36.36%	82.56%	40.74%
Recall for cluster	61.37%	54.55%	8.26%	36.36%	62.83%	15.06%
F-measure	53.25%	57.84%	6.18%	36.36%	71.36%	22.78%

d. [C2 + C3] Determine new classification rules for all the patterns in the noun compound and preposition clusters (*Y X*, *Y prep X* and *X prep Y*).

e. [C3] Determine classification rules only for the preposition cluster patterns (*Y prep X* and *X prep Y* patterns).

f. [C1 + C2 + C3] Determine new classification rules for all the patterns in all three clusters (*Y's X*; *X of Y*; *X Y*; *Y X*, and *Y have X*; *Y prep X* and *X prep Y*).

Table 16 shows the results obtained for the preposition cluster patterns in each of the five approaches used. One can observe that the preposition cluster alone provides the best results over all other combinations. These statistics are also consistent with the results obtained for the noun compound cluster experiments. The best approach is to use only the classification rules generated by the preposition cluster training examples.

Table 17 shows the classification rules learned only for the preposition cluster patterns.

5.4 Results for Discovering Part–Whole Relations

In order to test the classification rules for the extraction of part–whole relations, we selected two different text collections: the LA Times news articles from TREC 9 and the Wall Street Journal (WSJ) articles from Treebank2.[10] From each collection we randomly selected 10,000 sentences that formed two distinct test corpora. This corpus was parsed and disambiguated using a state-of-the-art domain independent Word Sense Disambiguation system that has an accuracy of 71% when disambiguating nouns in texts (Novischi et al. 2004). In cases in which the noun constituents were not in WordNet, we used an in-house Named Entity Recognizer (NERD) that has a 96% F-measure on MUC6 data.

The part–whole relations extracted by the ISS system were validated by comparing them with the gold standard for the test set obtained through inter-annotator agreement.

We define the *precision*, *recall*, and *F-measure* performance metrics in this context:

$$Precision = \frac{\text{Number of correctly retrieved relations}}{\text{Number of relations retrieved}} \tag{2}$$

10 Treebank2 is a text collection developed at UPenn consisting of a million words of 1989 Wall Street Journal material.

Table 17
The semantic classification rules learned for the preposition cluster. "Val." means target value (No or Yes), "Acc." is the rules' accuracy, and "Fr." is the frequency with which they occurred.

No.	Part Class	Whole Class	Val.	Acc.	Fr.	Example
	abstraction#6	**abstraction#6**	No			*force#7–past#1*
1	statement#1	speech#2	Yes			*announcement#1–news_conference#1*
2	signal#1	message#2	Yes			*letter#2–alphabet#1*
3	writing#2	writing#2	Yes			*addendum#1–back_matter#1*
4	linear_measure#1	measure#3	Yes	50	8	*inch#1–foot#1*
	entity#1	**entity#1**	No			*child#2–husband#1*
5	artifact#1	artifact#1	Yes			*door#1–room#1*
5.1	creation#2	artifact#1	No	93.17	10	*book#1–audiocassette#1*
5.2	artifact#1	creation#2	No	87.74	10	*charcoal#2–watercolor#1*
5.3	fabric#1	artifact#1	No	82.22	10	*knit#1–tie#1*
5.4	artifact#1	float#4	No	82.20	10	*outboard_motor#1–raft#1*
5.5	artifact#1	excavation#3	No	76.79	10	*adit#1–mine#1*
5.6	artifact#1	surface#1	No	68.13	10	*cargo_container#1–main_deck#1*
5.7	line#18	artifact#1	No	67.62	10	*rope#1–walkway#1*
5.8	way#6	way#6	No	67.62	10	*path#2–door#2*
5.9	container#1	furnishings#1	No	81.42	10	*bottle#1–wardrobe#1*
6	natural_object#1	natural_object#1	Yes			*pistil#1–flower#1*
7	region#1	object#1	Yes	67.09	10	*foot#3–shoe#1*
8	region#3	artifact#1	Yes	50	7	*seat#1–hall#3*
9	part#4	object#1	Yes	50	7	*auto_accessory#1–car#1*
	entity#1	**group#1**	No			*weapon#1–troop#2*
10	causal_agent#1	social_group#1	Yes			*member#1–association#1*
	group#1	**group#1**	No			*delegation#1–Washington#3*
11	people#1	social_group#1	Yes			*youth#2–high_school#1*
Default			No			*automobile#1–garage#1*

$$Recall = \frac{\text{Number of correctly retrieved relations}}{\text{Number of correct relations}} \tag{3}$$

$$F - measure = \frac{2}{\frac{1}{Precision} + \frac{1}{Recall}} \tag{4}$$

Tables 18 and 19 show the overall results obtained by the ISS system on the Wall Street Journal (WSJ) and on the LA Times collections of news articles, respectively. The results obtained for each cluster are summarized in Tables 1 and 2 in Appendix C.

Overall, on the WSJ test set the system obtained 82.87% precision and 79.09% recall on these three clusters. Besides the 373 relations corresponding to the three clusters, 33 other meronymy relations (406 − 373) were found in the corpus corresponding to part–whole lexico-syntactic patterns that were not studied in this paper, giving us a global part–whole relation coverage (recall) of 72.66%.

The ISS system's results were compared to four baseline measures. Baseline1 shows the results obtained by the system with no word sense disambiguation (WSD), using only sense#1 (the most frequent sense in WordNet) for the pair of concepts. In Baseline2, the system considered WSD and applied the specialization algorithm, but ran C4.5 only once on all the unambiguous sets of specialized training examples representing all the leaves of the learning tree. Baseline3 shows the results obtained without generalizing the concepts; and Baseline4 shows the results obtained with automatic word sense disambiguation (WSD) on the training corpus as opposed to the manual word sense disambiguation used for ISS training.

From the baselines' results for both the WSJ and LA Times text collections, one can see the importance of the WSD and IS–A generalization/specialization features to the extraction of the part–whole relations.

Table 18
The number of part–whole relations obtained and the accuracy in the WSJ collection.

Results	Baseline1 (No WSD)	Baseline2 (One learning)	Baseline3 (No generalization)	Baseline4 (Using WSD for training)	ISS System
Precision	7.72%	7.73%	15.71%	53.57%	82.87%
Pattern recall	24%	43%	2.95%	27.87%	79.09%
Relation recall	10.81%	19.37%	2.71%	25.86%	72.66%
F-measure	3.56%	6.02%	4.97%	36.67%	82.05%

Table 19
The number of part–whole relations obtained and the accuracy in the LA Times collection.

Results	Baseline1 (No WSD)	Baseline2 (One learning)	Baseline3 (No generalization)	Baseline4 (Using WSD for training)	ISS System
Precision	2.10%	3.24%	24.34%	48.22%	79.03%
Pattern recall	11.61%	42.86%	6.25%	20.61%	85.30%
Relation recall	3.02%	11.16%	5.80%	30.05%	79.15%
F-measure	11.68%	13.1%	9.98%	28.88%	80.94%

Figure 5 shows the learning curve where the classifier is trained on an incrementally increasing number of training data instances. The learning curve was determined by applying the training rules obtained through specialization on the LA Times test corpus annotated with automatic WSD. If for 1,000 positive and 1,000 negative examples the F-measure is only 35%, for 5,000 it increases to 70%, for 10,000 to 74%, for 15,000 to 77%, and it stabilizes at 87% for 20,000 examples. The learning curve shows that the ISS system obtains an F-measure of about 75% with only 16.8% of the training data.

5.5 Comparison with Previous Work

In this section we compare our work with two other approaches most similar to our task of part–whole semantic relation detection.

Berland and Charniak (1999) limit their approach to single words denoting some entities that have recognizable parts, such as *car* and *building*. As they also observe, this approach causes errors, such as the detection of *conditioner is part of car* instead of *air conditioner is part of car*. Our system is considerably more knowledge intensive, but more general in the sense that it relies on WordNet and NERD to detect both single word and multiple word concepts in context. Moreover, their system was tested only on a working list of predefined highly probable wholes for their corpus based on the genitive syntactic patterns. In contrast, the ISS system can disambiguate any pair of concepts, provided they are in WordNet or can be classified by NERD.

In order to eliminate a part of the data ambiguities, Berland and Charniak apply an ad hoc filtering procedure to eliminate those instances that represent properties or qualities of objects, such as those ending in *-ing*, *-ity*, and *-ness*. Our procedure is general enough to treat both positive and negative example instances.

Using the genitive patterns they find parts of a predefined list of wholes from a large text collection. Our method, however, determines if two noun concepts are in a part–whole relation or not. By generalizing the method to all the parts and wholes from our testing corpus, the accuracy of the system will fall. On the other hand, to be able to test the system on their six whole concepts we would need thousands of positive and negative examples for each such word. For instance, for the word *book*, Berland

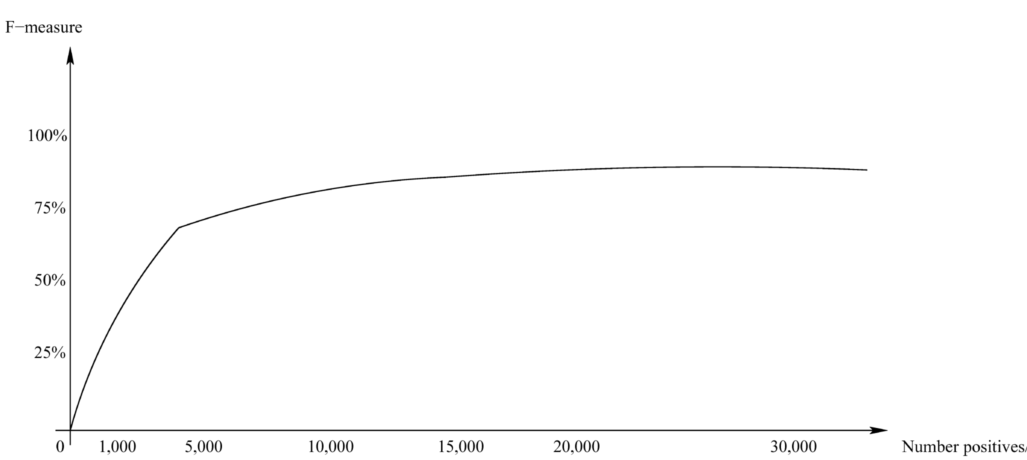

Figure 5
The learning curve for the number of learning examples.

and Charniak had almost 2,000 examples for the top 50 ranked parts. Unfortunately, in our LA Times testing corpus we couldn't find more than ten parts for each of their proposed whole objects. Therefore, we are unable to replicate their work using our text collection.

ISS algorithm is based on an iterative semantic specialization method that allows us to go deeper into the semantic complexity problem of the patterns considered. To the best of our knowledge, ISS is the only noun-phrase interpretation system that uses word sense disambiguation. One other noun compound interpretation system, SENS (Vanderwende 1995), used IS–A generalizations, and considered only the first sense of the noun constituents. The current state-of-the-art approaches in automatic detection of semantic roles (Gildea and Jurafsky 2002) have tried to use lexico-semantic hierarchies, such as WordNet, to generalize from lexical noun features. However, they also rely on the first sense listed for each noun occurring in the training data. Our experiments indicate the importance of WSD in extracting part–whole semantic relations.

6. Limitations and Extensions

The difficulty of detecting part–whole relations is due to a variety of factors ranging from syntactic analysis, to semantic and pragmatic information. In this section we analyze the sources of errors occurring in our experiments and present some possible improvements.

To arrive at an interpretation of the pair of words selected by the cluster patterns, it is first necessary to identify that both words are nouns, and not other parts of speech. For example, if Brill's tagger mis-tags an adjective or verb as a noun, then the ISS system will also be affected.

Our classification rule learning approach is based on the WordNet semantic classes of the two concepts that represent the part and the whole, respectively. Thus, if the WSD system fails to annotate the concepts with the correct senses, the ISS system can generate wrong semantic classes, which leads to wrong conclusions. For example, the WordNet concept *end* has 14 senses corresponding to 6 semantic classes (*entity, abstraction, event, psychological feature, state*, and *act*) (see Table 20). However, not all the senses refer to a part–whole relation (e.g., senses 4, 6, 8, 9, 11, and 14 do not). Some senses corresponding to both positive and negative examples are mapped into the same semantic class (e.g., senses 7 and 8). In this case, the classification error will not affect the final result as it is eliminated in the specialization phase. However, when a part–whole sense of *end* is mapped erroneously into a semantic class that is representative of negative examples, then the error might propagate to the final classification rule.

For some words, WordNet does not have all their senses. For example, the concepts *import* and *export* are not listed in WordNet as denoting the act of importing/exporting commodities from a foreign country. Thus, relations such as *import of sweater* and *export of milk* are mis-classified. Similar examples are *participant* and *beneficiary* for which WordNet lists only the senses corresponding to people and not to other entities, such as countries (e.g., a country can be one of the participants at a NATO meeting).

When a noun is too specific to be found in WordNet, we rely on a named entity recognizer (NERD). NERD identifies people, organizations, and other information extraction categories and annotates them accordingly. However, NERD doesn't always provide the correct annotation. For example, in the phrase *attorney of York*, it identifies

Table 20
The semantic classes and part–whole status for all the senses of the concept *end* in WordNet.

Sense No.	Semantic Class	Part–Whole Relation?	Examples
1	entity	Yes	*end of line*
2	abstraction	Yes	*end of year*
3	event	Yes	*end of movie*
4	psychological feature	No	*the end justifies the means*
5	psychological feature	Yes	*end of section*
6	state	No	*glorious end of the experiment* (sense of *destruction, death*)
7	entity	Yes	*end of box*
8	entity	No	*end hold the pass*
9	entity	No	*both ends wrote encouraging thoughts*
10	entity	Yes	*end of town*
11	act	No	-
12	abstraction	Yes	*In the end of the presentation*
13	entity	Yes	*the end of the cloth*
14	act	No	*the end from the line of scrimmage*

York as a name of a person and tags it with sense#1. However, *York#1* is defined in WordNet as *the House of York, the English royal house that reigned from 1461 to 1485*. Consequently, the ISS system will consider *York#1* a *group* instead of an *entity*, yielding an erroneous result.

The WSD tool identifies noun compounds and annotates them with the corresponding WordNet sense. For instance in the sentence "*... by/IN simply/RB/1 redesigning/VBG/1 how/WRB/1 a/DT car_door/NN/1 is/VBZ assembled/VBN/1*" the system annotated the concept *car door* with its WordNet sense (sense 1). This way, the ISS system considers the two words as a concept and not as a noun compound encoding a part–whole relation. The majority of noun compounds from the test corpus are names of people (e.g., *Andrea West, Mr. Moore*), dates (e.g., *Oct 12, Monday afternoon*), names of institutions (e.g., *Bank of America, Planters Corp., Research Inc., Johnson* & *Johnson*), or numbers (e.g., *six days, five years*). After analyzing the ambiguous pairs of nouns in noun compound instances, we noticed that only a few of them were positive examples. This error can be easily fixed by disabling the labeling of noun compounds with word senses.

Another class of errors involves the position of part and whole concepts. For example, the part–whole instance *band#1 of people#1* is detected by the pattern NP_X of NP_Y and the system classifies erroneously *band* as part, and *people* as whole. One way to overcome this is to further classify the patterns based on selectional restrictions on their constituent nouns (e.g., group nouns in *of*–genitives have different positions for the part and whole concepts).

We present in Table 21 the types of errors and their frequency of occurrence for each cluster and overall.

Although our approach takes context into account through the use of word sense disambiguation, it does so in a limited way, without access to the general discourse and pragmatic context within which a pair of nouns is embedded. Various researchers (Spärck Jones 1983; Lascarides and Copestake 1998; Lapata 2002) showed that the interpretation of noun compounds, for example, may be influenced by discourse and pragmatic knowledge. For instance, the discourse context provided by the following

Table 21
Error types statistics measured on the Wall Street Journal corpus for the ISS system.

	Clusters				
Error Type	Genitives	Noun compounds	Preposition	All	Error (%)
WSD System	57	36	26	119	53.85%
NERD module	11	5	9	25	11.31%
Brill's tagger	7	4	8	19	8.6%
Missing WordNet sense	10	1	7	18	8.14%
Part and Whole identification	5	10	0	15	6.79%
Classification rules	7	2	5	14	6.33%
Unseen examples					
Noun compound annotation	0	9	2	11	4.98%
Total	97	67	57	221	100%

sentences prefers the PURPOSE interpretation (*bag for cotton clothes*) of the noun compound *cotton bag* over the PART–WHOLE meaning (*bag made of cotton*) (cf. (Lapata 2002)):

(5) *Mary sorted her clothes into various bags made from plastic.*

(6) *She put her skirt into the cotton bag.*[11]

Encoding discourse knowledge is thus necessary. However, this is an open research problem and involves considerable manual annotation effort.

Furthermore, our experiments focused on the detection of part–whole relations in compositional constructions. A more general approach would consider lexicalized instances as well. Pragmatic knowledge is particularly important for the interpretation of lexicalized constructions, such as *soap opera*. The meaning of lexicalized instances is usually captured by semantic lexicons and dictionaries.

Finally, the approach presented here can be extended to other semantic relations encoded by the cluster patterns considered. The only part–whole elements used in this algorithm were the patterns and the examples. Thus the learning and the validation procedures are generally applicable and we intend to generalize the method for the detection of other semantic relations, such as KINSHIP and PURPOSE. So far, we have obtained encouraging results for a list of 35 general-purpose semantic relations encoded by genitives (Moldovan and Badulescu 2005), by noun compounds (Girju et al. 2005), and different noun phrase-level patterns including genitives, noun compounds, and the preposition patterns (Moldovan et al. 2004).

The drawback of the method presented here, as for other very precise learning methods, is that the number of training examples needs to be very large. If a certain class of negative or positive examples is not seen in the training data (and therefore it is not captured by the classification rules), the system cannot classify its instances. Thus, the larger and more diverse the training data, the better the classification rules.

11 These sentences were introduced in (Lapata 2002).

Table 22
The components of the AH–64A Apache Helicopter found on Web documents.

AH–64A Apache Helicopter		
	Hellfire air-to-surface missile	
		millimeter wave seeker
	70mm Folding Fin Aerial rocket	
	30mm Cannon camera	
	armaments	
	General Electric 1700-GE engine	
	4-rail launchers	
	four-bladed main rotor	
	anti-tank laser guided missile	
	Longbow millimetre wave fire control radar	
		integrated radar frequency interferometer
	rotating turret	
	tandem cockpit	
		Kevlar seats

7. Importance to NLP Applications

Since part–whole semantic relations occur frequently in text and have been recognized as fundamental ontological relations since ancient times, their discovery is paramount for applications such as Question Answering, automatic ontology construction, textual inferencing, and others. For questions like *What parts does General Electric manufacture?*, *What are the components of X*, *What is Y made of?*, and many more, the discovery of part–whole relations is necessary to assemble the right answer.

The concepts and part–whole relations acquired from a collection of documents can be useful in answering difficult questions that normally can not be handled based solely on keyword matching and proximity. As the level of difficulty increases, Question Answering systems need richer semantic resources, including ontologies and larger knowledge bases. Consider the question *What does the AH–64A Apache helicopter consist of?* For questions like this, the system must extract all the components the war helicopter has. Unless an ontology of such army attack helicopter parts exists in the knowledge base, which in an open domain situation is highly unlikely, the system must first acquire from the document collection all the pieces the helicopter is made of. These parts can be scattered all over the text collection, so the Question Answering system has to gather together these partial answers into a single and concise hierarchy of parts. This technique is called **answer fusion** (Girju 2001).

Using a state-of-the-art Question Answering system (Moldovan et al. 2002) adapted for **answer fusion** and including the ISS system as a module, the question presented above was answered by searching the Internet (the website for the Defence Industries—army at www.army-technology.com). The QA system started with the question focus *helicopter* and extracted and disambiguated all the meronymy relations using the ISS module. Table 22 shows the taxonomic ontology created for this question (presenting all the parts of a whole).

For example, the relation "*AH–64 Apache helicopter* has part *Hellfire air-to-surface missile*" was determined from the sentence *AH–64 Apache helicopter has a Longbow-millimetre wave fire control radar and a Hellfire air-to-surface missile*. Only the heads of the noun phrases were considered as they occur in WordNet (i.e., *helicopter* and *air-to-surface missile*, respectively).

Ontologies[12] are used more and more as means to boost the accuracy of natural language application systems (Moldovan and Girju 2001). Semantically richer ontologies can be built by incorporating more semantic relations in addition to the traditional IS–A relation. Part–whole is an excellent example of such relations. Recently, Tatu and Moldovan (2005) have shown that semantic relations such as part–whole can be combined with other relations using a semantic calculus for the purpose of improving the performance of a textual inference system.

8. Conclusions

In this paper we presented a supervised, knowledge-intensive approach to the automatic detection of part–whole relations encoded by the three most frequent clusters of syntactic constructions: (1) genitives and NP *have* NP clauses, (2) noun compounds, and (3) other NP PP phrases. The detection of the part–whole relations is difficult due to the highly ambiguous nature of the syntactic constructions, as they can encode other relations than meronymy.

Our method for detection of part–whole relations discovers semi-automatically the part–whole lexico-syntactic patterns and learns automatically the semantic classification rules needed for the disambiguation of these patterns. We defined the task as a binary classification problem and used an approach that relies on the assumption that the semantic relation between two constituent nouns representing the part and the whole can be detected based on the components' semantic classification rules. The classification rules are learned automatically through an *iterative semantic specialization* (ISS) procedure applied on the noun constituents' semantic classes provided by WordNet. We successfully combined the results of decision tree learning with the WordNet IS-A hierarchy specialization for more accurate learning. We proved the method is domain independent.

The classification rules learned by our method and listed in several tables can be easily implemented to extract part–whole relations from text. However, to apply these rules a word sense disambiguation system for nouns is necessary.

Our experiments revealed the importance of word sense disambiguation and WordNet IS–A specialization. We have directly compared and contrasted the results of our system with a variety of baselines and have shown impressive results. Combination of word sense disambiguation information with IS-A semantic information in WordNet yields better performance over either WSD or IS-A specialization alone.

Our experiments also showed that the three cluster patterns considered are not alternative ways of encoding part–whole information. This observation is very important for various text understanding applications.

Moreover, the approach presented can be extended to other semantic relations since the learning procedures are generally applicable and yield good results for sufficiently large training corpora.

12 Gartner Group identified Ontologies as one of the leading IT technologies, ranked 3^{rd} in its list of top 10 technologies forecast for 2005.

Appendix A: Experiments with Meronymic Patterns

Tables 1, 2, and 3 present a summary of phrase-level and sentence-level meronymic patterns and their possible extensions.

Table 1
The phrase-level patterns determined with the pattern identification procedure in Section 3. "Fr." means *frequency*.

No.	Pattern	Fr.	Example
1	NP_X PP_Y — PP_Y starts with *of*	173	door of his car the executive of the new government
2	NP_X PP_Y — PP_Y starts with *in*	61	people in the world
3	NP_X PP_Y — NP_X ends with *branch* — PP_Y begins with *of*	14	They organized the executive branch of government
4	NP_X PP_Y — PP_Y starts with *from*	9	oxygen from air people from all over the world
5	NP_X PP_Y — PP_Y starts with *throughout*	6	people throughout the world
6	NP_X PP_Y — PP_Y starts with *at*	5	window at the rear of the building
7	NP_X PP_Y — PP_Y starts with *on*	16	five fingers on one hand
8	NP_X PP_Y — PP_Y starts with *to*	3	the door to the house
9	NP_X PP_Y — PP_Y starts with *around*	2	people around the world
10	NP_X PP_Y — PP_Y starts with *all over*	1	people all over the world
11	*X* and other *Z* of *Y*	1	in Romania and the other countries of Eastern Europe
12	NP_X PP_Y — PP_Y starts with *to* — NP_X ends with *damage*	1	severe ligament damage to the left knee
13	NP_X PP_Y — PP_Y starts with *onto*	1	pavement onto streets
14	NP_X PP_Y — PP_Y starts with *outside*	1	windows outside the court building
15	NP_Y PP_X — PP_X starts with *with*	7	the organization with 120 members The spiders with 6 legs are dangerous.
16	NP_Y PP_X — PP_X starts with *above*	4	They amputate his leg above the knee.
17	NP_Y's NP_X	71	car's engine organization's membership
18	PP_X PP_Y — PP_X starts with *in* — PP_Y starts with *in*	2	in an abdomen, in a reclining torso
19	PP_Y PP_X — PP_X starts with *on* — PP_Y starts with *on*	1	on her car, on her window
20	NP_X, NP_Y	10	Bucharest, Romania in Atlanta ,Ga.
21	PP_Y WHP_X — WHP_X starts with *whose*	4	The club whose membership

Table 1
(cont.)

No.	Pattern	Fr.	Example
22	NP_{X1X2} PP_Y — NP_{X1X2} ends with *branches* — NP_{X1X2} contains *and* or *or* — PP_Y begins with *of*	8	between the executive and legislative branches of government
23	NP_{X1X2} PP_Y — NP_{X1X2} contains *and* or *or* — PP_Y begins with *of*	1	the memory and other features of IBM-compatible personal computers.
24	NP_Y (NP_{X1X2}) — NP_{X1X2} contains *and* or *or*	1	the three states of Southern New England (Massachusetts, Connecticut, and Rhode Island)
25	NP_X (NP_Y)	1	red-bellied snake (Storeria)
26	NP_Y NP_X	42	He sell car doors
27	NP_Z–NP_X NP_Y	26	a one–act ballet
28	NP_Y NP_X NP_Z	12	faulty garage door lock computer memory chip four–door compact car
29	NP_X NP_Y	3	membership organization power window buildings
30	NP_X–NP_Y NP_Z	3	a play–act universe

Table 2
The sentence-level patterns determined with the pattern identification procedure in Section 3. "Fr." means *frequency*.

No.	Pattern	Fr.	Example
1	NP_Y *verb* NP_X - verbs: *carry, combine, comprehend, comprise, consist, contain , enclose, feature, have, hold, hold in, house, include, incorporate, inherit, integrate, receive, retain, subsume*	18	A car *has* wheels. The cake *contains* fresh fruits. Any car *includes* a spare tire. The patient *received* a new heart.
2	NP_Z *verb* NP_X PP_Y - PP_Y starts with *in, into, as* or *from* - verbs: *assemble, build, build in, carry, combine, compose, compound, comprehend, comprise, connect, consist, construct, contain , coordinate, create, embrace, enclose, enter, fabricate, feature, file, form, have, hold, hold in, house, include, incorporate, infix, inherit, insert, integrate, introduce, join, link, make, manufacture, merge, observe, organize, overlap, receive, retain, subsume, unify, unite, use*	13	They *constructed* the car *from* engine, doors, wheels. The price *includes* a membership *in* a good club. The system administrator *connect* the computers *into* a computer network. The state *uses* these soldiers *as* the main army. The colonel *organized* the soldiers *into* an elite army. The user *inserts* the file *into* his directory. The programmer *includes* the main procedure *in* the C source file.
3	NP_Z *verb* PP_X PP_Y - PP_Y starts with *through* - verbs: *drag, exit, leave*	2	They *dragged* him *out of* the car *through* the window.
4	NP_X *verb* NP_Y - verbs: *accommodate, add, admit, affiliate, appertain, be, bear, belong, build in, colligate, compose, compound, confine, constitute, dwell, embrace, encompass, fall in, form, get together, infiltrate, inhere, involve, join, let in, lie, lie in, make, make up, pertain, rejoin, repose, represent, reside, rest, sign up, take*	1	The member *joined* the organization in 1976. The oxygen *composes* the air. The man *infiltrates* the organization in 1999.

Table 2
(cont.)

No.	Pattern	Fr.	Example
5	*NP_X verb* PP_Y - verbs: *attached to, inhere in* (PP_Y starts with *in* or *to*)	1	The cytoplasm *inhere in* a cell.
6	NP_X *verb NP_Z* PP_Y - PP_Y starts with *of* - NP_Z is part or *member*	2	The engine *is a part of* a car. A rose *is a member of* genus Rosa.
7	*NP_X NP_Y verb* - verbs: *carry, combine, comprehend, comprise, consist, contain , enclose, feature, have, hold, hold in, house, include, incorporate, inherit, integrate, receive, retain, subsume*	1	*The* headlights the car *had* were blue.
8	*NP_{Y1} verb NP_X* to *NP_{Y2}* - verbs: *donate, give*	1	The man *donated* one of his kidneys *to* his sister.
9	In *PP_Y, NP_{X1} verb NP_{X2}* - verbs: *cover*	1	*In* a car, the car body *covers* the engine.
	in *NP_Y* packed to *NP_X*	1	. . . *in* a car *packed to* the windows with personal belongings. . .
10	*NP_Z PP_X verb NP_T PP_Y* –*PP_X* starts with *in* –*NP_T* contains an adjective superlative - verb: *be*	1	Infant mortality *in* Romania *is* now the highest *in* Europe.

Table 3
Extensions for lexico-syntactic patterns discovered in the 20,000 sentence corpus used in the pattern identification procedure in Section 3.

No.	Pattern	Example
1	NP_Y PP_X - PP_Y starts with *without*	A bird without wings cannot fly.
2	NP_X PP_Y X inside Y - PP_Y starts with *inside*	The walls inside the building had better colors.

Appendix B: Semantic Classification Rules for the Genitive Cluster

Tables 1 and 2 show the full list of semantic classification rules learned for the genitive cluster from all the ambiguous nodes.

Table 1
The semantic classification rules learned for the genitive cluster all ambiguous nodes (default value No). "Val." is the target value, "Acc." is the rules' accuracy, and "Fr." is their occurrence frequency. The indentations in the "No." column refer to rules at different specialization levels.

No.	Part Class	Whole Class	Val.	Acc.	Fr.	Example
	abstraction#6	**abstraction#6**	No			*glory#2–past#1*
1	linear_measure#3	measure#3	Yes	63	9	*centimeter#1–decimeter#1*
2	communication#2	communication#2	Yes			*act#3–play#1*
2.1	written_communication#1	written_communication#1	No			*text#1–act#3*
2.1.1	writing#2	writing#2	Yes			*New_Testament#1–Bible#1*
2.1.1.1	matter#6		No	79.98	9	*text#1–act#3*
2.2	indication#1	message#2	No	73.25	10	*copy#1–recommendation#1*
2.3	message#2	communication#2	No	79.72	8	*irony#1–play#1*
2.4	time#5	abstraction#6	Yes	79.21	9	*carboniferous#1–paleozoic#1*
	abstraction#6	**entity#1**	No			*age#1–earth#1*
4	shape#2	artifact#1	Yes			*point#8–knife#2*
4.1	shape#2	structure#1	No	67.62	10	*diameter#2–plug#1*
4.2	shape#2	surface#1	No	67.62	10	*square#1–pegboard#1*
5	measure#3	object#1	Yes			*drumstick#1–bird#2*
5.1	definite_quantity#1	artifact#1	No			*dozen#1–videotape#1*
5.2	indefinite_quantity#1	artifact#1	No			*lot#1–throttle#1*
5.3	linear_measure#1	artifact#1	No			*mile#1–quarters#1*
5.4	system_of_measurement#1	object#1	No	82.26	8	*bandwidth#1–receiver#1*
5.5	relative_quantity#1	object#1	No	79.72	8	*nothing#1–refrigerator#1*
6	position#7	artifact#1	Yes			*circle#6–theater#1*
6.1	placement#1		No	67.62	10	*density#2–pattern#3*
7	written_communication#1	instrumentality#3	Yes	85.70	10	*by-line#1–writing_arm#1*
8	shape#2	location#1	Yes	66.56	10	*point#8–arrowhead#1*
	abstraction#6	**group#1**	No			*history#3–regiment#1*

Table 1
(cont.)

No.	Part Class	Whole Class	Val.	Acc.	Fr.	Example
9	abstraction#6	biological_group#1	Yes	92.44	10	*year#3–montia#1*
10	relation#1	arrangement#2	Yes	79.40	9	*medium_frequency#1–electromagnetic_spectrum#1*
	abstraction#6	**phenomenon#1**	No			*cause#2–death#2*
11	shape#2	physical_phenomenon#1	Yes			*dewdrop#1–dew#1*
	abstraction#6	**psychological_feature#1**	No			*amount#1–work#4*
12	measure#3	structure#3	Yes	95.64	10	*August#1–Gregorian_calendar#1*
	entity#1	**phenomenon#1**	No			*keeper#2–flame#1*
13	point#2	physical_phenomenon#1	Yes			*storm_center#3–storm#1*
14	object#1	process#2	Yes			*ferric_oxide#1–rust#3*
	event#1	**entity#1**	No			
15	periodic_event#1	object#1	Yes	66.56	10	*wave#3–waveguide#1*
	event#1	**event#1**	No			*rerun#1–television_show#1*
16	happening#1	periodic_event#1	Yes	58.12	8	*flood#6–flood_tide#2*
	phenomenon#1	**abstraction#6**	No			*omission#3–pronoun#1*
17	atmospheric_phenomenon#1	communication#2	Yes			*gentle_breeze#1–beaufort_scale#1*
	phenomenon#1	**entity#1**	No			
18	process#2	organism#1	Yes			*meiosis#1–anapsid#1*
18.1	process#2	person#1	No	76.70	8	*growth#2–child#2*
	phenomenon#1	**phenomenon#1**	No			*influence#4–action#6*
19	natural_phenomenon#1	natural_phenomenon#1	Yes			*meteor#1–meteor_shower#1*
	possession#2	**entity#1**	No			*cost#1–home#2*
20	territory#2	entity#1	Yes	69.84	9	*united_states_virgin_islands#1–virgin_islands#1*
	possession#2	**possession#2**	No			*liquid_assets#1–capital#1*
21	assets#1	transferred_property#1	Yes			*cut#6–loot#1*
	psychological_feature#1	**psychological_feature#1**	No			
22	knowledge_domain#1	knowledge_domain#1	Yes			*agrology#1–agronomy#1*

Table 2
The semantic classification rules learned for the genitive cluster from all the ambiguous nodes with default value Yes. "Val." means target value (No or Yes), "Acc." is the rules' accuracy, and "Fr." is the frequency with which they occurred. The numbering style used in the "No." column is intended to indicate rules learned at different specialization levels.

No.	Part Class	Whole Class	Val.	Acc.	Fr.	Example
23	**entity#1**	**entity#1**	Yes			*door#4–car#1*
23.1	causal_agent#1	causal_agent#1	No			*lethal_dose#1–opium#1*
23.2	causal_agent#1	location#1	No			*taxi_driver#1–Los_Angeles#1*
23.3	causal_agent#1	object#1	No			*dose#1–malathion#1*
23.4	point#2	body_of_water#1	No	94.46	10	*headwaters#1–nile#1*
23.5	region#1	body_of_water#1	No	91.79	10	*east_side#1–river#1*
23.6	line#11	region#3	No			*direction#1–park#1*
23.7	geographic_point#1	region#3	No	89.43	9	*corner#4–washington#1*
23.8	point#2	geographical_area#1	No	79.98	9	*stock_exchange#1–istanbul#1*
23.9	district#1	district#1	No			*commonwealth#1–puerto_rico#1*
23.9.1	admin_district#1	admin_district#1	Yes			*Alaska#1–United_States#1*
23.10	location#1	object#1	Yes			*Romania#1–Europe#1*
23.10.1	location#1	natural_object#1	No	89.55	10	*neighborhood#1–earth#1*
23.10.2	area#1	natural_object#1	No	50	8	*corner#1–earth#1*
23.10.3	location#1	person#1	No	87.90	10	*birthplace#1–Nixon#1*
23.11	object#1	causal_agent#1	No			*bottle#1–prescription_drug#1*
23.12	part#4	location#1	No	50	7	*map#1–Vietnam#1*
23.13	creation#2	extremity#4	No			*set#4–edge#1*
23.14	facility#1	region#3	No	93.74	10	*railway_station#1–Beijing#1*
23.15	creation#2	region#3	No	87.90	10	*miniature#2–warsaw#1*
23.16	equipment#1	admin_district#1	No			*reactor#2–iraq#1*
23.17	natural_object#1	point#2	No			*headland#1–top#3*
23.18	block#1	building_material#1	No	95.32	10	*slab#1–concrete#1*
23.19	artifact#1	paving_material#1	No	91.76	10	*slab#1–concrete#1*
23.20	creation#2	structure#1	No	95.32	10	*art#1–music-hall#1*
23.21	commodity#1	instrumentality#3	No			*shipment#1–capacitor#1*
23.22	flap#1	clothing#1	No			*hem#1–dress#1*
23.23	representation#2	creation#2	No	67.62	10	*spectacle#2–scenery#1*

Table 2
(cont.)

No.	Part Class	Whole Class	Val.	Acc.	Fr.	Example
23.24	design#4	covering#2	No	67.62	10	*colors#1–paint#1*
23.25	land#3	island#1	No	91.79	10	*continent#1–Atlantis#1*
23.26	appendage#3	instrumentality#3	No			*stock#7–artillery#1*
23.27	object#1	part#7	No			*slab#1–fat#2*
23.28	object#1	unit#6	No			*addition#1–sodium_nitrate#1*
23.29	organism#1	causal_agent#1	No			*supplier#1–cocaine#1*
23.30	organism#1	location#1	No			*ambassador#1–iraq#1*
23.31	organism#1	object#1	No			*author#1–book#1*
23.32	organism#1	organism#1	No			*assassin#1–Kennedy#1*
23.33	thing#12	entity#1	No	94.64	10	*something#1–America#1*
23.34	body_of_water#1	entity#1	No	68.18	8	*sea#1–interaction#1*
23.	Default		Yes			*door#4–car#1*
24	**entity#1**	**group#1**	Yes			*academician#1–academy#2*
24.1	entity#1	system#1	No	87.90	10	*river#1–ecosystem#1*
24.2	artifact#1	gathering#1	No			*rostrum#1–congress#2*
24.	Default		Yes			*academician#1–academy#2*
25	**entity#1**	**possession#2**	Yes			*Tuamotu_Archipelago#1–French_Polynesia#1*
25.1	organism#1	possession#2	No	85.50	8	*manager#1–investment_funds#1*
25.2	causal_agent#1	possession#2	No	85.30	8	*buyer#1–life_insurance#1*
25.	Default		Yes			*Tuamotu_Archipelago#1–French_Polynesia#1*
26	**group#1**	**group#1**	Yes			*genus_amoeba#1–amoebida#1*
26.1	social_group#1	people#1	No			*dictatorship#1–proletariat#1*
26.2	group#1	people#1	No	83.86	8	*demi-monde#1–high_society#1*
26.3	arrangement#2	collection#1	No	82.22	10	*classification#2–family#4*
26.4	social_group#1	collection#1	No	82.22	10	*circle#2–law#2*
26.	Default		Yes			*genus_amoeba#1–amoebida#1*
	Default		No			

Appendix C: Performance Results per Cluster

Tables 1 and 2 show the performance results obtained for each cluster considered on LA Times and WSJ test corpora.

Table 1
The number of part–whole relations obtained and the accuracy for each cluster and for all the clusters in the WSJ collection.

Results	Genitives cluster	Noun compounds cluster	Preposition cluster	All clusters
ISS system				
Number of patterns	1514	1217	1081	3812
Number of correctly retrieved relations	161	85	49	295
Number of relations retrieved	202	90	64	356
Number of correct relations for the pattern(s)	167	141	65	373
Number of correct relations				406
Precision	79.70%	94.44%	76.56%	82.87%
Recall for cluster(s)	96.41%	60.28%	75.38%	79.09%
Coverage				72.66%
F-measure	86.87%	77.13%	71.36%	82.05%
Baseline1 — No WSD				
Precision	6.04%	50%	45.45%	7.72%
Recall for the pattern(s)	36%	3.57%	22.73%	24%
Coverage				10.81%
F-measure	2.12%	7.41%	23.26%	3.56%
Baseline2 — One learning				
Precision	7.18%	37.5%	20%	7.73%
Recall for the pattern(s)	78%	10.71%	4.54%	43%
Coverage				19.37%
F-measure	6%	10%	3.57%	6.02%
Baseline3 — No Generalization				
Precision	28.21%	0%	0%	15.71%
Recall for the pattern(s)	6.59%	0%	0%	2.95%
Coverage				2.71%
F-measure	10.68%	1.12%	0%	4.97%
Baseline4 — WSD using system for training				
Precision	62.42%	50.74%	59.05%	53.57%
Recall for the pattern(s)	92.40%	68.67%	54.87%	27.87%
Coverage				25.86%
F-measure	74.51%	58.36%	56.88%	36.67%

Table 2
The number of part–whole relations obtained and the accuracy for each cluster and for all the clusters in the LA Times collection.

Results	Genitives cluster	Noun compounds cluster	Preposition cluster	All clusters
ISS system				
Number of patterns	4106	3442	2577	10125
Number of correctly retrieved relations	321	113	71	505
Number of relations retrieved	410	143	86	639
Number of correct relations for the pattern(s)	329	150	113	592
Number of correct relations				638
Precision	78.29%	79.02%	82.56%	79.03%
Recall for cluster(s)	97.57%	75.33%	62.83%	85.30%
Coverage				79.15%
F-measure	87.26%	73.59%	75.97%	80.94%
Baseline1 — No WSD				
Precision	1.16%	33.33%	38.46%	2.10%
Recall for the pattern(s)	12.07%	4.17%	16.67%	11.61%
Coverage				3.02%
F-measure	10.34%	6.66%	30.3%	11.68%
Baseline2 — One learning				
Precision	3.12%	12.5%	3.84%	3.24%
Recall for the pattern(s)	77.59%	8.33%	3.33%	42.86%
Coverage				11.16%
F-measure	13.15%	16.66%	7.4%	13.1%
Baseline3 — No Generalization				
Precision	34.29%	3.33%	0%	24.34%
Recall for the pattern(s)	10.94%	0.67%	0%	6.25%
Coverage				5.80%
F-measure	16.59%	1.12%	0%	9.98%
Baseline4 — WSD using system for training				
Precision	52.8%	54%	39.81%	48.22%
Recall for the pattern(s)	79.04%	57.45%	63.08%	20.61%
Coverage				30.05%
F-measure	63.31%	55.67%	48.81%	28.88%

Acknowledgments
We would like to thank Matthew Jones for his help in providing the gold-standard annotations for the training and test corpora used in this research. We are grateful for the constructive comments made by Robert Dale and anonymous reviewers that helped considerably clarify and improve the presentation. This work was partially supported by the Advanced Research and Development Activity/ Disruptive Technology Office.

References

Berland, Matthew and Eugene Charniak. 1999. Finding parts in very large corpora. In *Proceedings of the 37th Annual Meeting of the Association for Computational Linguistics (ACL 1999)*, pages 57–64, University of Maryland.

Brill, Eric. 1995. Transformation-based error-driven learning and natural language processing: A case study in part-of-speech tagging. *Computational Linguistics*, 21(4):543–566.

Charniak, Eugene. 2000. A maximum-entropy-inspired parser. In *Proceedings of the 1st Conference of the North American Chapter of the Association for Computational Linguistics (NAACL 2000)*, pages 132–139, Seattle, WA.

Downing, Pamela. 1977. On the creation and use of English compound nouns. *Language*, 53(4):810–842.

Dunning, Ted. 1993. Accurate methods for the statistics of surprise and coincidence. *Computational Linguistics*, 19:61–74.

Evens, Martha W., Bonnie C. Litowitz, Judith A. Markowitz, Raoul N. Smith, and Oswald Werner. 1980. Lexical-semantic relations: A comparative survey. *Linguistic Research*, pages 187–219.

Fellbaum, Christiane. 1998. *WordNet—An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Finin, Timothy W. 1980. *The Semantic Interpretation of Compound Nominals*. Ph.D. thesis, University of Illinois at Urbana-Champaign.

Freeze, Ray. 1992. Existentials and other locatives. *Language*, 68:553–595.

Gildea, Daniel and Daniel Jurafsky. 2002. Automatic labeling of semantic roles. *Computational Linguistics*, 28(3): 245–288.

Girju, Roxana. 2001. Answer fusion with on-line ontology development. In *Proceedings of the 2nd Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL 2001) - Student Research Workshop*, pages 23–28, Pittsburgh, PA.

Girju, Roxana, Adriana Badulescu, and Dan Moldovan. 2003. Learning semantic constraints for the automatic discovery of part-whole relations. In *Proceedings of the 3rd Human Language Technology Conference/ 4th Meeting of the North American Chapter of the Association for Computational Linguistics Conference (HLT-NAACL 2003)*, pages 80–87, Edmonton, Canada.

Girju, Roxana, Dan Moldovan, Marta Tatu, and Daniel Antohe. 2005. On the semantics of noun compounds. *Computer Speech and Language—Special Issue on Multiword Expressions (in press)*.

Hearst, Marti. 1992. Acquisition of hyponyms from large text corpora. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, pages 539–545, Nantes, France.

Hearst, Marti. 1998. Automated discovery of WordNet relations. In Christiane Fellbaum, editor, *An Electronic Lexical Database and Some of Its Applications*. MIT Press, Cambridge, MA, pages 131–151.

Iris, Madelyn, Bonnie Litowitz, and Martha Evens. 1988. Problems with part-whole relation. In M. W. Evens, editor, *Relational Models of the Lexicon: Representing Knowledge in Semantic Networks*. Cambridge University Press, Cambridge, pages 261–288.

Jensen, Per Anker and Carl Vikner. 1996. The double nature of the verb have. *LAMBDA*, 21:25–37.

Kingsbury, Paul, Martha Palmer, and Mitch Marcus. 2002. Adding semantic annotation to the Penn Treebank. In *Proceedings of the 2nd Human Language Technology Conference (HLT 2002)*, pages 252–256, San Diego, CA.

Lapata, Mirella. 2002. The disambiguation of nominalisations. *Computational Linguistics*, 28(3):357–388.

Lascarides, Alex and Ann Copestake. 1998. Pragmatics and word meaning. *Journal of Linguistics*, 34(2):387–414.

Lauer, Mark and Mark Dras. 1994. A probabilistic model of compound nouns. In *Proceedings of the 7th Australian Joint Conference on Artificial Intelligence*, pages 474–481, Armidale, Australia.

Levi, Judith. 1978. *The Syntax and Semantics of Complex Nominals*. Academic Press, New York.

Marcus, Mitchell P., Beatrice Santorini, and Mary Ann Marcinkiewicz. 1993. Building a large annotated corpus of English: The Penn Treebank. *Computational Linguistics*, 19(2):313–330.

Moldovan, Dan and Adriana Badulescu. 2005. A semantic scattering model for the automatic interpretation of genitives. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Proceesing (HLT/EMNLP 2005)*, pages 891–898, Vancouver, BC, Canada.

Moldovan, Dan, Adriana Badulescu, Marta Tatu, Daniel Antohe, and Roxana Girju. 2004. Models for the semantic classification of noun phrases. In *Proceedings of the Human Language Technology Conference (HLT-NAACL) 2004, Computational Lexical Semantics Workshop*, Boston, MA.

Moldovan, Dan and Roxana Girju. 2001. An interactive tool for the rapid development of knowledge bases. *International Journal on Artificial Intelligence Tools*, 10(1–2):65–86.

Moldovan, Dan, Sanda Harabagiu, Roxana Girju, Paul Morarescu, Finley Lacatusu, Adrian Novischi, Adriana Badulescu, and Orest Bolohan. 2002. LCC tools for question answering. In *Proceedings of the 11th Meeting of the Text Retrieval Conference (TREC 2002)*, pages 388–397, Gaithersburg, MD.

Morris, Jane and Graeme Hirst. 2004. Non-classical lexical semantic relations. In *Proceedings of the 4th Human Language Technology Conference / of the 5th Meeting of the North American Chapter of the Association for Computational Linguistics (HLT-NAACL 2004) - Workshop on Computational Lexical Semantics*, pages 46–51, Boston, MA.

Novischi, Adrian, Dan Moldovan, Paul Parker, Adriana Badulescu, and Bob Hauser. 2004. LCC's WSD systems for Senseval 3. In *Proceedings of Senseval 3 (ACL 2004)*, Barcelona, Spain.

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

Quinlan, Ross. J. 1993. *C4.5: Programs for Machine Learning*. Morgan Kaufmann, San Francisco, CA.

Resnik, Philip. 1996. Selectional constraints: An information-theoretic model and its computational realization. *Cognition*, 61:127–159.

Resnik, Philip and Marti Hearst. 1993. Structural ambiguity and conceptual relations. In *Proceedings of the 31st Meeting of the Association for Computational Linguistics (ACL 1993)-1st Workshop on Very Large Corpora: Academic and Industrial Perspectives*, pages 58–64, Ohio State University, Columbus, OH.

Rosario, Barbara and Marti Hearst. 2001. Classifying the semantic relations in noun compounds via a domain-specific lexical hierarchy. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP 2001)*, pages 82–90, Pittsburgh, PA.

Rosario, Barbara, Marti Hearst, and Charles Fillmore. 2002. The descent of hierarchy, and selection in relational semantics. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 247–254, University of Pennsylvania.

Schafer, Robin. 1995. The SLP/ILP distinction in have-predication. In M. Simons and T. Galloway, editors, *Proceedings from Semantics and Linguistic Theory V*. Cornell University Department of Linguistics, pages 292–309, Ithaca.

Siegel, Sidney and John Castellan. 1988. *Nonparametric Statistics for the Behavioral Science*. McGraw-Hill, New York.

Simons, Peter. 1987. *Parts. A Study in Ontology*. Clarendon Press, Oxford.

Simons, Peter. 1991. Part/whole II: Mereology since 1900. In H. Burkhardt and B. Smith, editors, *Handbook of Metaphysics and Ontology*. Philosophia, Munich, pages 672–675.

Spärck Jones, K. 1983. Compound noun interpretation problems. In F. Fallside and W. A. Woods, editors, *Computer Speech Processing*. Prentice-Hall, Englewood Cliffs, NJ, pages 363–380.

Tatu, Marta and Dan Moldovan. 2005. A semantic approach to recognizing textual entailmant. In *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP 2005)*, pages 371–378, Vancouver, BC, Canada.

Thompson, Cynthia A., Roger Levy, and Christopher Manning. 2003. A generative model for Framenet semantic role labeling. In *Proceedings of the 14th*

European Conference on Machine Learning (ECML 2003), pages 397–408, Cavtat-Dubrovnik, Croatia.

Vanderwende, Lucy. 1994. Algorithm for automatic interpretation of noun sequences. In *Proceedings of the 15th International Conference on Computational Linguistics (COLING 1994)*, pages 782–788, Kyoto, Japan.

Vanderwende, Lucy. 1995. *The Analysis of Noun Sequences using Semantic Information Extracted from On-Line Dictionaries*. Ph.D. thesis, Georgetown University.

Winston, Morton, Roger Chaffin, and Douglas Hermann. 1987. A taxonomy of part-whole relations. *Cognitive Science*, 11(4):417–444.