

Models for the Semantic Classification of Noun Phrases

**Dan Moldovan, Adriana Badulescu,
Marta Tatu, and Daniel Antohe**
Computer Science Department
University of Texas at Dallas
Dallas, Texas
moldovan@utdallas.edu

Roxana Girju
Department of Computer Science
Baylor University
Waco, Texas
girju@cs.baylor.edu

Abstract

This paper presents an approach for detecting semantic relations in noun phrases. A learning algorithm, called semantic scattering, is used to automatically label complex nominals, genitives and adjectival noun phrases with the corresponding semantic relation.

1 Problem description

This paper is about the automatic labeling of semantic relations in noun phrases (NPs).

The semantic relations are the underlying relations between two concepts expressed by words or phrases. We distinguish here between *semantic relations* and *semantic roles*. Semantic roles are always between verbs (or nouns derived from verbs) and other constituents (*run quickly*, *went to the store*, *computer maker*), whereas semantic relations can occur between any constituents, for example in complex nominals (*malaria mosquito* (CAUSE)), genitives (*girl's mouth* (PART-WHOLE)), prepositional phrases attached to nouns (*man at the store* (LOCATIVE)), or discourse level (*The bus was late. As a result, I missed my appointment* (CAUSE)). Thus, in a sense, semantic relations are more general than semantic roles and many semantic role types will appear on our list of semantic relations.

The following NP level constructions are considered here (cf. the classifications provided by (Quirk et al.1985) and (Semmelmeyer and Bolander 1992)): (1) *Compound Nominals* consisting of two consecutive nouns (eg *night club* - a TEMPORAL relation - indicating that club functions at night), (2) *Adjective Noun* constructions where the adjectival modifier is derived from a noun (eg *musical clock* - a MAKE/PRODUCE relation), (3) *Genitives* (eg *the door of the car* - a PART-WHOLE relation), and (4) *Adjective phrases* (cf. (Semmelmeyer and Bolander 1992)) in which the modifier noun is expressed

by a prepositional phrase which functions as an adjective (eg *toy in the box* - a LOCATION relation).

Example: “*Saturday’s snowfall* topped a *one-day record in Hartford*, Connecticut, with *the total of 12.5 inches*, the *weather service* said. The storm claimed its fatality Thursday, when *a car which was driven by a college student* skidded on *an interstate overpass in the mountains of Virginia* and hit *a concrete barrier*, police said”. (www.cnn.com - “Record-setting Northeast snowstorm winding down”, Sunday, December 7, 2003).

There are several semantic relations at the noun phrase level: (1) *Saturday’s snowfall* is a genitive encoding a TEMPORAL relation, (2) *one-day record* is a TOPIC noun compound indicating that record is about one-day snowing - an ellipsis here, (3) *record in Hartford* is an adjective phrase in a LOCATION relation, (4) *total of 12.5 inches* is an of-genitive that expresses MEASURE, (5) *weather service* is a noun compound in a TOPIC relation, (6) *car which was driven by a college student* encodes a THEME semantic role in an adjectival clause, (7) *college student* is a compound nominal in a PART-WHOLE/MEMBER-OF relation, (8) *interstate overpass* is a LOCATION noun compound, (9) *mountains of Virginia* is an of-genitive showing a PART-WHOLE/PLACE-AREA and LOCATION relation, (10) *concrete barrier* is a noun compound encoding PART-WHOLE/STUFF-OF.

1.1 List of Semantic Relations

After many iterations over a period of time we identified a set of semantic relations that cover a large majority of text semantics. Table 1 lists these relations, their definitions, examples, and some references. Most of the time, the semantic relations are encoded by lexico-syntactic patterns that are highly ambiguous. One pattern can express a number of semantic relations, its disambiguation being provided by the context or world knowledge. Often semantic relations are not disjoint or mutually exclusive, two or more appearing in the same lexical construct. This

is called *semantic blend* (Quirk et al. 1985). For example, the expression “*Texas city*” contains both a LOCATION as well as a PART-WHOLE relation.

Other researchers have identified other sets of semantic relations (Levi 1979), (Vanderwende 1994), (Sowa 1994), (Baker, Fillmore, and Lowe 1998), (Rosario and Hearst 2001), (Kingsbury, et al. 2002), (Blaheta and Charniak 2000), (Gildea and Jurafsky 2002), (Gildea and Palmer 2002). Our list contains the most frequently used semantic relations we have observed on a large corpus.

Besides the work on semantic roles, considerable interest has been shown in the automatic interpretation of complex nominals, and especially of compound nominals. The focus here is to determine the semantic relations that hold between different concepts within the same phrase, and to analyze the meaning of these compounds. Several approaches have been proposed for empirical noun-compound interpretation, such as syntactic analysis based on statistical techniques (Lauer and Dras 1994), (Pustejovsky et al. 1993). Another popular approach focuses on the interpretation of the underlying semantics. Many researchers that followed this approach relied mostly on hand-coded rules (Finin 1980), (Vanderwende 1994). More recently, (Rosario and Hearst 2001), (Rosario, Hearst, and Fillmore 2002), (Lapata 2002) have proposed automatic methods that analyze and detect noun compounds relations from text. (Rosario and Hearst 2001) focused on the medical domain making use of a lexical ontology and standard machine learning techniques.

2 Approach

2.1 Basic Approach

We approach the problem top-down, namely identify and study first the characteristics or feature vectors of each noun phrase linguistic pattern, then develop models for their semantic classification. This is in contrast to our prior approach ((Girju, Badulescu, and Moldovan 2003a)) when we studied one relation at a time, and learned constraints to identify only that relation. We study the distribution of the semantic relations across different NP patterns and analyze the similarities and differences among resulting *semantic spaces*. We define a semantic space as the set of semantic relations an NP construction can encode. We aim at uncovering the general aspects that govern the NP semantics, and thus delineate the semantic space within clusters of semantic relations. This process has the advantage of reducing the annotation effort, a time consuming activity. Instead of manually annotating a corpus for each semantic relation, we do it only for each syntactic pattern and get a clear view of its semantic space. This syntactico-semantic approach allows

us to explore various NP semantic classification models in a unified way.

This approach stemmed from our desire to answer questions such as:

1. What influences the semantic interpretation of various linguistic constructions?
2. Is there only one interpretation system/model that works best for all types of expressions at all syntactic levels? and
3. What parameters govern the models capable of semantic interpretation of various syntactic constructions?

2.2 Semantic Relations at NP level

It is well understood and agreed in linguistics that concepts can be represented in many ways using various constructions at different syntactic levels. This is in part why we decided to take the syntactico-semantic approach that analyzes semantic relations at different syntactic levels of representation. In this paper we focus only on the behavior of semantic relations at NP level. A thorough understanding of the syntactic and semantic characteristics of NPs provides valuable insights into defining the most representative **feature vectors** that ultimately drive the discriminating learning models.

Complex Nominals

Levi (Levi 1979) defines complex nominals (CNs) as expressions that have a head noun preceded by one or more modifying nouns, or by adjectives derived from nouns (usually called denominal adjectives). Most importantly for us, each sequence of nouns, or possibly adjectives and nouns, has a particular meaning as a whole carrying an implicit semantic relation; for example, “*spoon handle*” (PART-WHOLE) or “*musical clock*” (MAKE/PRODUCE).

CNs have been studied intensively in linguistics, psycho-linguistics, philosophy, and computational linguistics for a long time. The *semantic interpretation* of CNs proves to be very difficult for a number of reasons. (1) Sometimes the meaning changes with the head (eg “*musical clock*” MAKE/PRODUCE, “*musical creation*” THEME), other times with the modifier (eg “*GM car*” MAKE/PRODUCE, “*family car*” POSSESSION). (2) CNs’ interpretation is knowledge intensive and can be idiosyncratic. For example, in order to interpret correctly “*GM car*” 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, “*USA city*” can be regarded as a LOCATION as well as a PART-WHOLE relation. (4) Interpretation of CNs 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” (cf. (Downing 1977)).

Genitives

The semantic interpretation of genitive constructions

No.	Semantic Relation	Definition / Example
1	POSSESSION	an animate entity possesses (owns) another entity; (<i>family estate; the girl has a new car.</i>), (Vanderwende 1994)
2	KINSHIP	an animated entity related by blood, marriage, adoption or strong affinity to another animated entity; (<i>Mary's daughter; my sister</i>); (Levi 1979)
3	PROPERTY/ ATTRIBUTE-HOLDER	characteristic or quality of an entity/event/state; (<i>red rose; The thunderstorm was awful.</i>); (Levi 1979)
4	AGENT	the <i>doer</i> or instigator of the action denoted by the predicate; (<i>employee protest; parental approval; The king banished the general.</i>); (Baker, Fillmore, and Lowe 1998)
5	TEMPORAL	time associated with an event; (<i>5-o'clock tea; winter training; the store opens at 9 am</i>), includes DURATION (Navigli and Velardi 2003).
6	DEPICTION-DEPICTED	an event/action/entity depicting another event/action/entity; (<i>A picture of my niece.</i>),
7	PART-WHOLE (MERONYMY)	an entity/event/state is part of another entity/event/state (<i>door knob; door of the car</i>), (Levi 1979), (Dolan et al. 1993),
8	HYPERNYMY (IS-A)	an entity/event/state is a subclass of another; (<i>daisy flower; Virginia state; large company, such as Microsoft</i>) (Levi 1979), (Dolan et al. 1993)
9	ENTAIL	an event/state is a logical consequence of another; (<i>snoring entails sleeping</i>)
10	CAUSE	an event/state makes another event/state to take place; (<i>malaria mosquitoes; to die of hunger; The earthquake generated a Tsunami</i>), (Levi 1979)
11	MAKE/PRODUCE	an animated entity creates or manufactures another entity; (<i>honey bees; nuclear power plant; GM makes cars</i>) (Levi 1979)
12	INSTRUMENT	an entity used in an event/action as instrument; (<i>pump drainage; the hammer broke the box</i>) (Levi 1979)
13	LOCATION/SPACE	spatial relation between two entities or between an event and an entity; includes DIRECTION; (<i>field mouse; street show; I left the keys in the car</i>), (Levi 1979), (Dolan et al. 1993)
14	PURPOSE	a state/action intended to result from a another state/event; (<i>migraine drug; wine glass; rescue mission; He was quiet in order not to disturb her.</i>) (Navigli and Velardi 2003)
15	SOURCE/FROM	place where an entity comes from; (<i>olive oil; I got it from China</i>) (Levi 1979)
16	TOPIC	an object is a topic of another object; (<i>weather report; construction plan; article about terrorism</i>); (Rosario and Hearst 2001)
17	MANNER	a way in which an event is performed or takes place; (<i>hard-working immigrants; enjoy immensely; he died of cancer</i>); (Blaheta and Charniak 2000)
18	MEANS	the means by which an event is performed or takes place; (<i>bus service; I go to school by bus.</i>) (Quirk et al.1985)
19	ACCOMPANIMENT	one/more entities accompanying another entity involved in an event; (<i>meeting with friends; She came with us</i>) (Quirk et al.1985)
20	EXPERIENCER	an animated entity experiencing a state/feeling; (<i>Mary was in a state of panic.</i>); (Sowa 1994)
21	RECIPIENT	an animated entity for which an event is performed; (<i>The eggs are for you</i>) ; includes BENEFICIARY; (Sowa 1994)
22	FREQUENCY	number of occurrences of an event; (<i>bi-annual meeting; I take the bus every day</i>); (Sowa 1994)
23	INFLUENCE	an entity/event that affects other entity/event; (<i>drug-affected families; The war has an impact on the economy.</i>);
24	ASSOCIATED WITH	an entity/event/state that is in an (undefined) relation with another entity/event/state; (<i>Jazz-associated company.</i>)
25	MEASURE	an entity expressing quantity of another entity/event; (<i>cup of sugar; 70-km distance; centennial rite; The jacket cost \$60.</i>)
26	SYNONYMY (NAME)	a word/concept that means the same or nearly the same as another word/concept; (<i>Marry is called Minnie</i>); (Sowa 1994)
27	ANTONYMY	a word/concept that is the opposite of another word/concept; (<i>empty is the opposite of full</i>); (Sowa 1994)
28	PROBABILITY OF EXISTENCE	the quality/state of being probable; likelihood (<i>There is little chance of rain tonight</i>); (Sowa 1994)
29	POSSIBILITY	the state/condition of being possible; (<i>I might go to Opera tonight</i>); (Sowa 1994)
30	CERTAINTY	the state/condition of being certain or without doubt; (<i>He definitely left the house this morning</i>);
31	THEME	an entity that is changed/involved by the action/event denoted by the predicate; (<i>music lover; John opened the door.</i>); (Sowa 1994)
32	RESULT	the inanimate result of the action/event denoted by the predicate; includes EFFECT and PRODUCT. (<i>combustion gases; I finished the task completely.</i>); (Sowa 1994)
33	STIMULUS	stimulus of the action or event denoted by the predicate (<i>We saw [the painting]. I sensed [the eagerness] in him. I can see [that you are feeling great].</i>) (Baker, Fillmore, and Lowe 1998)
34	EXTENT	the change of status on a scale (by a percentage or by a value) of some entity; (<i>The price of oil increased [ten percent]. Oil's price increased by [ten percent].</i>); (Blaheta and Charniak 2000)
35	PREDICATE	expresses the property associated with the subject or the object through the verb; (<i>He feels [sleepy]. They elected him [treasurer].</i>) (Blaheta and Charniak 2000)

Table 1: A list of semantic relations at various syntactic levels (including NP level), their definitions, some examples, and references.

is considered problematic by linguists because they involve an implicit relation that seems to allow for a large variety of relational interpretations; for example: “*John’s car*”-POSSESSOR-POSSESSEE, “*Mary’s brother*”-KINSHIP, “*last year’s exhibition*”-TEMPORAL, “*a picture of my niece*”-DEPICTION-DEPICTED, and “*the desert’s oasis*”-PART-WHOLE/PLACE-AREA. A characteristic of these constructions is that they are very productive, as the construction can be given various interpretations depending on the context. One such example is “*Kate’s book*” that can mean the book Kate owns, the

book Kate wrote, or the book Kate is very fond of.

Thus, the features that contribute to the semantic interpretation of genitives are: the nouns’ semantic classes, the type of genitives, discourse and pragmatic information.

Adjective Phrases are prepositional phrases attached to nouns acting as adjectives (cf. (Semmelmeier and Bolander 1992)). Prepositions play an important role both syntactically and semantically. Semantically speaking, prepositional constructions can encode various se-

semantic relations, their interpretations being provided most of the time by the underlying context. For instance, the preposition “with” can encode different semantic relations: (1) It was the girl *with* blue eyes (MERONYMY), (2) The baby *with* the red ribbon is cute (POSSESSION), (3) The woman *with* triplets received a lot of attention (KINSHIP).

The conclusion for us is that in addition to the nouns semantic classes, the preposition and the context play important roles here.

In order to focus our research, we will concentrate for now only on noun - noun or adjective - noun *compositional constructions* at NP level, ie those whose meaning can be derived from the meaning of the constituent nouns (“door knob”, “cup of wine”). We don’t consider metaphorical names (eg, “ladyfinger”), metonymies (eg, “Vietnam veteran”), proper names (eg, “John Doe”), and NPs with coordinate structures in which neither noun is the head (eg, “player-coach”). However, we check if the constructions are *non-compositional (lexicalized)* (the meaning is a matter of convention; e.g., “soap opera”, “sea lion”), but only for statistical purposes. Fortunately, some of these can be identified with the help of lexicons.

2.3 Corpus Analysis at NP level

In order to provide a unified approach for the detection of semantic relations at different NP levels, we analyzed the syntactic and semantic behavior of these constructions on a large open-domain corpora of examples. Our intention is to answer questions like: (1) *What are the semantic relations encoded by the NP-level constructions?*, (2) *What is their distribution on a large corpus?*, (3) *Is there a common subset of semantic relations that can be fully paraphrased by all types of NP constructions?*, (4) *How many NPs are lexicalized?*

The data

We have assembled a corpus from two sources: Wall Street Journal articles from TREC-9, and eXtended WordNet glosses (XWN) (<http://xwn.hlt.utdallas.edu>). We used XWN 2.0 since all its glosses are syntactically parsed and their words semantically disambiguated which saved us considerable amount of time. Table 2 shows for each syntactic category the number of randomly selected sentences from each corpus, the number of instances found in these sentences, and finally the number of instances that our group managed to annotate by hand. The annotation of each example consisted of specifying its feature vector and the most appropriate semantic relation from those listed in Table 1.

Inter-annotator Agreement

The annotators, four PhD students in Computational Semantics worked in groups of two, each group focusing on one half of the corpora to annotate. Noun - noun

(adjective - noun, respectively) sequences of words were extracted using the Lauer heuristic (Lauer 1995) which looks for consecutive pairs of nouns that are neither preceded nor succeeded by a noun after each sentence was syntactically parsed with Charniak parser (Charniak 2001) (for XWN we used the gold parse trees). Moreover, they were provided with the sentence in which the pairs occurred along with their corresponding WordNet senses. Whenever the annotators found an example encoding a semantic relation other than those provided or they didn’t know what interpretation to give, they had to tag it as “OTHERS”. Besides the type of relation, the annotators were asked to provide information about the order of the modifier and the head nouns in the syntactic constructions if applicable. For instance, in “owner of car”-POSSESSION the possessor *owner* is followed by the possessee *car*, while in “car of John”-POSSESSION/R the order is reversed. On average, 30% of the training examples had the nouns in reverse order.

Most of the time, one instance was tagged with one semantic relation, but there were also situations in which an example could belong to more than one relation in the same context. For example, the genitive “city of USA” was tagged as a PART-WHOLE/PLACE-AREA relation and as a LOCATION relation. Overall, there were 608 such cases in the training corpora. Moreover, the annotators were asked to indicate if the instance was lexicalized or not. Also, the judges tagged the NP nouns in the training corpus with their corresponding WordNet senses.

The annotators’ agreement was measured using the Kappa statistics, one of the most frequently used measure of inter-annotator agreement for classification tasks: $K = \frac{Pr(A) - Pr(E)}{1 - Pr(E)}$, where $Pr(A)$ is the proportion of times the raters agree and $Pr(E)$ is the probability of agreement by chance. The K coefficient is 1 if there is a total agreement among the annotators, and 0 if there is no agreement other than that expected to occur by chance.

Table 3 shows the semantic relations inter-annotator agreement on both training and test corpora for each NP construction. For each construction, the corpus was split into 80/20 training/testing ratio after agreement.

We computed the K coefficient only for those instances tagged with one of the 35 semantic relations. For each pattern, we also computed the number of pairs that were tagged with OTHERS by both annotators, over the number of examples classified in this category by at least one of the judges, averaged by the number of patterns considered.

The K coefficient shows a fair to good level of agreement for the training and testing data on the set of 35 relations, taking into consideration the task difficulty. This can be explained by the instructions the annotators received prior to annotation and by their expertise in lexical semantics. There were many heated discussions as well.

	Wall Street Journal					eXtended WordNet 2.0
	CNs		Genitives		Adjective Phrases	Complex Nominals
	NN	AdjN	's	of		
No. of sentences	7067	5381	50291	27067	14582	51058
No. of instances	5557	500	2990	4185	3502	12412
No. of annotated instances	2315	383	1816	3404	1341	1651

Table 2: Corpus statistics.

Kappa Agreement (1 - 35)					
Complex Nominals		Genitives		Adjective Phrases	OTHERS
NN	Adj N	's	of		
0.55	0.68	0.66	0.65	0.67	0.69

Table 3: The inter-annotator agreement on the semantic annotation of the NP constructions considered on both training and test corpora. For the semantic blend examples, the agreement was done on one of the relations only.

2.4 Distribution of Semantic Relations

Even noun phrase constructions are very productive allowing for a large number of possible interpretations, Table 4 shows that a relatively small set of 35 semantic relations covers a significant part of the semantic distribution of these constructions on a large open-domain corpus. Moreover, the distribution of these relations is dependent on the type of NP construction, each type encoding a particular subset. For example, in the case of of-genitives, there were 21 relations found from the total of 35 relations considered. The most frequently occurring relations were PART-WHOLE, ATTRIBUTE-HOLDER, POSSESSION, LOCATION, SOURCE, TOPIC, and THEME. By comparing the subsets of semantic relations in each column we can notice that these semantic spaces are not identical, proving our initial intuition that the NP constructions cannot be alternative ways of packing the same information. Table 4 also shows that there is a subset of semantic relations that can be fully encoded by all types of NP constructions. The statistics about the *lexicalized* examples are as follows: N-N (30.01%), Adj-N (0%), s-genitive (0%), of-genitive (0%), adjective phrase (1%). From the 30.01% lexicalized noun compounds, 18% were proper names.

This simple analysis leads to the important conclusion that the NP constructions must be treated separately as their semantic content is different. This observation is also partially consistent with other recent work in linguistics and computational linguistics on the grammatical variation of the English genitives, noun compounds, and adjective phrases.

We can draw from here the following conclusions:

1. Not all semantic relations can be encoded by all NP syntactic constructions.
2. There are semantic relations that have preferences over particular syntactic constructions.

2.5 Models

2.5.1 Mathematical formulation

Given each NP syntactic construction considered, the goal is to develop a procedure for the automatic labeling of the semantic relations they encode. The semantic relation derives from the lexical, syntactic, semantic and contextual features of each NP construction.

Semantic classification of syntactic patterns in general can be formulated as a learning problem, and thus benefit from the theoretical foundation and experience gained with various learning paradigms. This is a multi-class classification problem since the output can be one of the semantic relations in the set. We cast this as a supervised learning problem where input/ output pairs are available as training data.

An important first step is to map the characteristics of each NP construction (usually not numerical) into feature vectors. Let's define with \mathbf{x}_i the feature vector of an instance i and let X be the space of all instances; ie $\mathbf{x}_i \in X$.

The multi-class classification is performed by a function that maps the feature space X into a semantic space S ,

$f : X \rightarrow S$, where S is the set of semantic relations from Table 1, ie $r_j \in S$.

Let T be the training set of examples or instances $T = (\mathbf{x}_1 r_1, \mathbf{x}_2 r_2, \dots, \mathbf{x}_l r_l) \subseteq (X \times S)^l$ where l is the number of examples \mathbf{x} each accompanied by its semantic relation label r . The problem is to decide which semantic relation r to assign to a new, unseen example \mathbf{x}_{l+1} . In order to classify a given set of examples (members of X), one needs some kind of measure of the similarity (or the difference) between any two given members of X . Most of the times it is difficult to explicitly define this function, since X can contain features with numerical as well as non-numerical values.

Note that the features, thus space X , vary from an NP pattern to another and the classification function will be pattern dependent. The novelty of this learning problem is the feature space X and the nature of the discriminating

No.	Semantic Relations	Frequency(%)				Adjective Phrases	Examples
		CNs		Genitives			
		NN	AdjN	's	of		
1	POSSESSION	2.91	9.44	14.55	3.89	1.47	"family estate"
2	KINSHIP	0	0	7.94	3.23	0.20	"woman with triplets"
3	ATTRIBUTE-HOLDER	12.38	7.34	8.96	10.77	4.09	"John's coldness"
4	AGENT	4.21	10.49	9.75	0.98	3.46	"the investigation of the crew"
5	TEMPORAL	0.82	0	11.96	0.53	7.97	"last year's exhibition"
6	DEPICTION-DEPICTED	0	0	1.49	2.86	0.20	"a picture of my niece"
7	PART-WHOLE	19.68	10.84	27.38	31.70	5.56	"the girl's mouth"
8	IS-A (HYPERNYMY)	3.95	1.05	0	0.04	1.15	"city of Dallas"
9	ENTAIL	0	0	0	0	0	
10	CAUSE	0.08	0	0.23	0.57	1.04	"malaria mosquitoes"
11	MAKE/PRODUCE	1.56	2.09	0.86	0.98	0.63	"shoe factory"
12	INSTRUMENT	0.65	0	0.07	0.16	0.94	"pump drainage"
13	LOCATION/SPACE	7.51	20.28	1.57	3.89	18.04	"university in Texas"
14	PURPOSE	6.69	3.84	0.07	0.20	5.45	"migraine drug"
15	SOURCE	1.69	11.53	2.51	5.85	2.94	"president of Bolivia"
16	TOPIC	4.04	4.54	0.15	4.95	9.13	"museum of art"
17	MANNER	0.40	0	0	0	0.20	"performance with style"
18	MEANS	0.26	0	0	0	0.10	"bus service"
19	ACCOMPANIMENT	0	0	0	0	0.42	"meeting with friends"
20	EXPERIENCER	0.08	2.09	0.07	0.16	2.30	"victim of lung disease"
21	RECIPIENT	1.04	0	3.54	2.66	2.51	"Josephine's reward"
22	FREQUENCY	0.04	7.00	0	0	0.52	"bi-annual meeting"
23	INFLUENCE	0.13	0.35	0	0	0.73	"drug-affected families"
24	ASSOCIATED WITH	1.00	0.35	0.15	0.16	2.51	"John's lawyer"
25	MEASURE	1.47	0.35	0.07	13.88	1.36	"hundred of dollars"
26	SYNONYMY	0	0	0	0	0	
27	ANTONYMY	0	0	0	0	0	
28	PROBABILITY	0	0	0	0	0	
29	POSSIBILITY	0	0	0	0	0	
30	CERTAINTY	0	0	0	0	0	
31	THEME	6.51	1.75	3.30	6.26	9.75	"car salesman"
32	RESULT	0.48	0	0.07	0.41	1.36	"combustion gas"
33	STIMULUS	0	0	0	0	0	
34	EXTENT	0	0	0	0	0	
35	PREDICATE	0.04	0	0	0	0.10	
	OTHERS	23.19	6.64	5.19	5.77	15.73	"airmail stamp"
	Total no. of examples	100% (2302)	100% (286)	100% (1271)	100% (2441)	100% (953)	

Table 4: The distribution of the semantic relations on the annotated corpus after agreement. The number in parentheses represent number of examples for a particular pattern.

function f derived for each syntactic pattern.

2.5.2 Feature space

An essential aspect of our approach below is the word sense disambiguation (WSD) of the content words (nouns, verbs, adjectives and adverbs). Using a state-of-the-art open-text WSD system, each word is mapped into its corresponding **WordNet 2.0** sense. When disambiguating each word, the WSD algorithm takes into account the surrounding words, and this is one important way through which *context* gets to play a role in the semantic classification of NPs.

So far, we have identified and experimented with the following NP features:

1. Semantic class of head noun specifies the WordNet sense (synset) of the head noun and implicitly points to all its hypernyms. It is extracted automatically via a word sense disambiguation module. The NP semantics is influenced heavily by the meaning of the noun constituents. Example: "car manufacturer" is a kind of manufacturer that MAKES/PRODUCES cars.
2. Semantic class of modifier noun

specifies the WordNet synset of the modifier noun. In case the modifier is a denominal adjective, we take the synset of the noun from which the adjective is derived. Example: "musical clock" - MAKE/PRODUCE, and "electric clock"- INSTRUMENT.

2.5.3 Learning Models

Several learning models can be used to provide the discriminating function f . So far we have experimented with three models: (1) semantic scattering, (2) decision trees, and (3) naive Bayes. The first is described below, the other two are fairly well known from the machine learning literature.

Semantic Scattering. This is a new model developed by us particularly useful for the classification of compound nominals $N_1 N_2$ without nominalization. The semantic relation in this case derives from the semantics of the two noun concepts participating in these constructions as well as the surrounding context.

Model Formulation. Let us define with $SC^m = \{f_i^m\}$ and $SC^h = \{f_j^h\}$ the sets of semantic class features (ie,

WordNet synsets) of the NP modifiers and, respectively NP heads (ie features 2 and 1). The compound nominal semantics is distinctly specified by the feature pair $f_i^m f_j^h$, written shortly as f_{ij} . Given feature pair f_{ij} , the probability of a semantic relation r is $P(r|f_{ij}) = \frac{n(r, f_{ij})}{n(f_{ij})}$, defined as the ratio between the number of occurrences of a relation r in the presence of feature pair f_{ij} over the number of occurrences of feature pair f_{ij} in the corpus. The most probable relation \hat{r} is $\hat{r} = \operatorname{argmax}_{r \in R} P(r|f_{ij}) = \operatorname{argmax}_{r \in R} P(f_{ij}|r)P(r)$

Since the number of possible noun synsets combinations is large, it is difficult to measure the quantities $n(r, f_{ij})$ and $n(f_{ij})$ on a training corpus to calculate $P(r|f_{ij})$. One way of approximating the feature vector f_{ij} is to perform a semantic generalization, by replacing the synsets with their most general hypernyms, followed by a series of specializations for the purpose of eliminating ambiguities in the training data. There are 9 noun hierarchies, thus only 81 possible combinations at the most general level. Table 5 shows a row of the probability matrix $P(r|f_{ij})$ for $f_{ij} = \text{entity} - \text{entity}$. Each entry, for which there is more than one relation, is *scattered* into other subclasses through an iterative process till there is only one semantic relation per line. This can be achieved by specializing the feature pair's semantic classes with their immediate WordNet hyponyms. The iterative process stops when new training data does not bring any improvements (see Table 6).

2.5.4 Overview of the Preliminary Results

The f-measure results obtained so far are summarized in Table 7. Overall, these results are very encouraging given the complexity of the problem.

2.5.5 Error Analysis

An important way of improving the performance of a system is to do a detailed error analysis of the results. We have analyzed the sources of errors in each case and found out that most of them are due to (in decreasing order of importance): (1) errors in automatic sense disambiguation, (2) missing combinations of features that occur in testing but not in the training data, (3) levels of specialization are too high, (4) errors caused by metonymy, (6) errors in the modifier-head order, and others. These errors could be substantially decreased with more research effort.

A further analysis of the data led us to consider a different criterion of classification that splits the examples into nominalizations and non-nominalizations. The reason is that nominalization noun phrases seem to call for a different set of learning features than the non-nominalization noun phrases, taking advantage of the underlying verb-argument structure. Details about this approach are provided in (Girju et al. 2004).

3 Applications

Semantic relations occur with high frequency in open text, and thus, their discovery is paramount for many applications. One important application is *Question Answering*. A powerful method of answering more difficult questions is to associate to each question the semantic relation that reflects the meaning of that question and then search for that semantic relation over the candidates of semantically tagged paragraphs. Here is an example.

Q. Where have nuclear incidents occurred? From the question stem word *where*, we know the question asks for a LOCATION which is found in the complex nominal “*Three Mile Island*”-LOCATION of the sentence “The Three Mile Island nuclear incident caused a DOE policy crisis”, leading to the correct answer “*Three Mile Island*”.

Q. What did the factory in Howell Michigan make? The verb *make* tells us to look for a MAKE/PRODUCE relation which is found in the complex nominal “*car factory*”-MAKE/PRODUCE of the text: “The car factory in Howell Michigan closed on Dec 22, 1991” which leads to answer *car*.

Another important application is building semantically rich ontologies. Last but not least, the discovery of text semantic relations can improve syntactic parsing and even WSD which in turn affects directly the accuracy of other NLP modules and applications. We consider these applications for future work.

References

- C. Baker, C. Fillmore, and J. Lowe. 1998. The Berkeley FrameNet Project. In *Proceedings of COLLING/ACL*, Canada.
- D. Blaheta and E. Charniak. 2000. Assigning function tags to parsed text. In *Proceedings of the 1st Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL)*, Seattle, WA.
- E. Charniak. 2001. Immediate-head parsing for language models. In *Proceedings of ACL*, Toulouse, France.
- W. Dolan, L. Vanderwende, and S. Richardson. 1993. Automatically deriving structured KBs from on-line dictionaries. In *Proceedings of the Pacific Association for Computational Linguistics Conference*.
- P. Downing. 1977. On the creation and use of English compound nouns. *Language*, 53(4), 810-842.
- T. Finin. 1980. The Semantic Interpretation of Compound Nominals. *Ph.D dissertation*, University of Illinois, Urbana, Illinois.
- D. Gildea and D. Jurafsky. 2002. Automatic Labeling of Semantic Roles. In *Computational Linguistics*, 28(3).

Relation no.	1	2	3	6	7	11	13	15	16	21	25	the rest
$P(r entity - entity)$	0.06103	0.11268	0.00939	0.04225	0.39437	0.01878	0.03286	0.25822	0.04694	0.01878	0.00469	0

Table 5: Sample row from the conditional probability table where the feature pair is entity-entity. The numbers in the top row identify the semantic relations (as in Table 4).

Level	Level 1	Level 2	Level 3	Level 4
Number of modifier features	9	52	70	122
Number head features	9	46	47	47
No. of feature pairs	57 out of 81	189 out of 2392	204 out of 3290	250 out of 5734
Number of features with only one relation	1	152	181	225
Average number of non-zero relations per line	2.7692	1.291	1.1765	1.144

Table 6: Statistics for the semantic class features by level of specialization.

Syntactic Pattern	Semantic Scattering	Decision Tree	Naive Bayes	
Complex Nominals	NN	43.22%	36.42%	36.18%
Genitives	AdjN	53.37%	33.33%	40.00%
	'S	50.64%	29.42%	25.42%
	Of	75.16%	39.94%	34.72%
Adjective Phrases		33.24%	36.76%	32.20%

Table 7: F-measure results for the semantic classification of NP patterns obtained with four learning models on a corpus with an 80/20 training/testing ratio. "NA" means not available.

- D. Gildea and M. Palmer. 2002. The Necessity of Parsing for Predicate Argument Recognition. In *Proceedings of the 40th Meeting of the Association for Computational Linguistics (ACL 2002)*.
- R. Girju, A. Badulescu, and D. Moldovan. 2003. Learning Semantic Constraints for the Automatic Discovery of Part-Whole Relations. In *Proceedings of the Human Language Technology Conference (HLT-03)*, Canada.
- R. Girju, A.M. Giuglea, M. Olteanu, O. Fortu, and D. Moldovan. 2004. Support Vector Machines Applied to the Classification of Semantic Relations in Nominalized Noun Phrases. In *Proceedings of HLT/NAACL 2004 - Computational Lexical Semantics workshop*, Boston, MA.
- P. Kingsbury, M. Palmer, and M. Marcus. 2002. Adding Semantic Annotation to the Penn TreeBank. In *Proceedings of the Human Language Technology Conference (HLT 2002)*, California.
- M. Lapata. 2002. The Disambiguation of Nominalisations. In *Computational Linguistics* 28:3, 357-388.
- M. Lauer and M. Dras. 1994. A probabilistic model of compound nouns. In *Proceedings of the 7th Australian Joint Conference on AI*.
- M. Lauer. 1995. Designing Statistical Language Learners: Experiments on Compound Nouns. In *PhD thesis*, Macquarie University, Sidney.
- Judith Levi. 1979. *The Syntax and Semantics of Complex Nominals*. New York: Academic Press.
- R. Navigli and P. Velardi. 2003. Ontology Learning and Its Application to Automated Terminology Translation. In *IEEE Intelligent Systems*.
- J. Pustejovsky, S. Bergler, and P. Anick. 1993. Lexical semantic techniques for corpus analysis. In *Computational Linguistics*, 19(2).
- R. Quirk, S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A comprehensive grammar of english language*, Longman, Harlow.
- B. Rosario and M. Hearst. 2001. Classifying the Semantic Relations in Noun Compounds via a Domain-Specific Lexical Hierarchy. In *the Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing (EMNLP 2001)*, Pittsburgh, PA.
- B. Rosario, M. Hearst, and C. Fillmore. 2002. The Descent of Hierarchy, and Selection in Relational Semantics. In *the Proceedings of the Association for Computational Linguistics (ACL-02)*, University of Pennsylvania.
- M. Semmelmeier and D. Bolander. 1992. *The New Webster's Grammar Guide*. Lexicon Publications, Inc.
- J. F. Sowa. 1994. *Conceptual Structures: Information Processing in Mind and Machine*. Addison Wesley.
- L. Vanderwende. 1994. Algorithm for automatic interpretation of noun sequences. In *Proceedings of COLING-94*, pg. 782-788.