Experiments with an Annotation Scheme for a Knowledge-rich Noun Phrase Interpretation System

Roxana Girju
University of Illinois at Urbana-Champaign
girju@uiuc.edu

Abstract

This paper presents observations on our experience with an annotation scheme that was used in the training of a state-of-the-art noun phrase semantic interpretation system. The system relies on cross-linguistic evidence from a set of five Romance languages: Spanish, Italian, French, Portuguese, and Romanian. Given a training set of English noun phrases in context along with their translations in the five Romance languages, our algorithm automatically learns a classification function that is later on applied to unseen test instances for semantic interpretation. As training and test data we used two text collections of different genre: Europarl and CLUVI. The training data was annotated with contextual features based on two state-of-the-art classification tag sets.

1 Introduction

Linguistically annotated corpora are valuable resources for both theoretical and computational linguistics. They have played an important role in any aspect of natural language processing research, from supervised learning to evaluation, and have been used in many applications such as Syntactic and Semantic Parsing, Information Extraction, and Question Answering.

A long-term research topic in linguistics, computational linguistics[1], and artificial intelligence has been the semantic interpretation of noun phrases (NPs). The basic problem is simple to define: given a noun phrase constructed out of a pair of concepts expressed by words or phrases, $c_1 - c_2$, one representing the head and the other the modifier, determine the semantic relationship between the two concepts. For example, a compound *family estate* should be interpreted as the estate OWNED BY the family; an NP such as *dress of silk* should be interpreted as denoting a dress MADE FROM silk. The problem, while simple to state is hard to solve. The reason is that the meaning of these constructions is most of the time ambiguous or implicit.

Currently, the best-performing English NP interpretation methods in computational linguistics focus mostly on two consecutive noun instances (noun compounds) and are either (weakly) supervised, knowledge-intensive (Rosario and Hearst, 2001), (Rosario et al., 2002), (Moldovan et al., 2004), (Pantel and Pennacchiotti, 2006), (Pennacchiotti and Pantel, 2006), (Kim and Baldwin, 2006), (Snow et al., 2006), (Girju et al., 2005; Girju et al., 2006), or use statistical models on large collections of unlabeled data (Berland and Charniak, 1999), (Lapata and Keller, 2004), (Nakov and Hearst, 2005), (Turney, 2006). Unlike unsupervised models, supervised knowledge-rich approaches rely heavily on large sets of annotated training data. For example, we previously showed (Girju et al., 2006) that, for

[1]In the past few years at many workshops, tutorials, and competitions this research topic has received considerable interest from the computational linguistics community: Workshop on Multiword Expressions at COLING/ACL 2006, 2004, 2003; Computational Lexical Semantics Workshop at ACL 2004; Tutorial on Knowledge Discovery from Text at ACL 2003; Shared task on Semantic Role Labeling at CONLL 2005, 2004 and at SENSEVAL 2005.

Proceedings of the Linguistic Annotation Workshop, pages 168–175,
Prague, June 2007.

the task of automatic detection of part-whole relations, our system's learning curve reached a plateau at 74% F-measure when trained on approximatively 10,000 positive and negative examples.

Interpreting NPs correctly requires various types of information from world knowledge to complex context features. Since the training data needs to be as accurate as possible, many of such features are manually identified and annotated. Thus, the annotation process is an important task that requires not only considerable amount of time, but also experience with various annotation schemas and tools, and a good understanding of the research topic. Moreover, the extension of the noun phrase interpretation task to other natural languages brings forward new annotation issues.

This paper presents observations on our experience with an annotation scheme that was used in the training of a state-of-the-art noun phrase semantic interpretation system (Girju, 2007). The system relies on cross-linguistic evidence from a set of five Romance languages: Spanish, Italian, French, Portuguese, and Romanian. Given a training set of English noun phrases in context along with their translations in the five Romance languages, our algorithm automatically learns a classification function that is later on applied to unseen test instances for semantic interpretation. As training and test data we used two text collections of different genre: Europarl[2] and CLUVI[3]. The training data was annotated with contextual features based on two state-of-the-art classification tag sets: Lauer's set of 8 prepositions (Lauer, 1995) and our list of 22 semantic relations. The system achieved an accuracy of 77.9% (Europarl) and 74.31% (CLUVI).

The paper is organized as follows. Section 2 presents a summary of linguistic considerations of noun phrases. In Section 3 we describe the list of semantic interpretation categories used along with observations regarding their distribution on the two different cross-lingual corpora. Section 4 presents the data used along with observations on corpus annotation and inter-annotator agreement. Finally, Section 5 offers some discussion and conclusions.

2 Linguistic considerations of noun phrases

The automatic discovery of semantic relations must start with a thorough understanding of the linguistic aspects of the underlying relations. These considerations are not only employed as features in the supervised noun phrase interpretation model, but they are also used in the annotation process.

Noun phrases can be compositional when their meaning is derived from the meaning of the constituent nouns (e.g., *door knob* – PART-WHOLE, *kiss in the morning* – TEMPORAL), or idiosyncratic, when the meaning is a matter of convention (e.g., *soap opera*, *sea lion*). NPs can also express metaphorical names (eg, *ladyfinger*), proper names (e.g., *John Doe*), and binomial (dvandva) compounds in which neither noun is the head (e.g., *player-coach*).

NPs can also be classified into *synthetic* (verbal) and *root* (non-verbal) constructions. It is widely held (Levi, 1978), (Selkirk, 1982) that the modifier noun of a synthetic noun compound, for example, may be associated with a theta-role of the verbal head. For instance, in *truck driver*, the noun *truck* satisfies the THEME relation associated with the direct object in the corresponding argument structure of the verb *to drive*.

Studied cross-linguistically, noun phrases can express variations from one language to another. For example, English compounds of the form N_1 N_2 (e.g., *wood stove*) usually translate in Romance languages as N_2 P N_1 (e.g., *four á bois* (French) – *stove at/to wood*). Romance languages have very few N N compounds and they are of limited semantic categories, such as TYPE (e.g., *legge quadro* (Italian) – *framework law*). Moreover, while English N N compounds are right-headed (e.g., *framework*/modifier *law*/head), Romance compounds are left-headed (e.g., *legge*/head *quadro*/modifier).

For this research we focus only on English–Romance compositional noun phrases of the type N N and N P N and disregard metaphorical and

[2]http://www.isi.edu/koehn/europarl/
This corpus contains over 20 million words in eleven official languages of the European Union covering the proceedings of the European Parliament from 1996 to 2001.

[3]CLUVI - Linguistic Corpus of the University of Vigo Parallel Corpus 2.1 - http://sli.uvigo.es/CLUVI/. CLUVI is an open text repository of parallel corpora of contemporary oral and written texts in some of the Romance languages, such as Galician, French, Spanish, Portuguese, Basque parallel text collections.

proper names. In the following section we present two different state-of-the-art classification sets used in NP interpretation.

3 Lists of semantic classification relations

Although researchers (Downing, 1977), (Jespersen, 1954) argued that noun compounds, and NPs in general, encode an infinite set of semantic relations, many agree (Finin, 1980), (Levi, 1978) there is a limited number of relations that occur with high frequency in these constructions. However, the number and the level of abstraction of these frequently used semantic categories are not agreed upon. They can vary from a few prepositions (Lauer, 1995) to hundreds and even thousands more specific semantic relations (Finin, 1980). The more abstract the categories, the more noun phrases are covered, but also the more room for variation as to which category a phrase should be assigned. Lauer (Lauer, 1995), for example, considers a set of eight prepositions as semantic classification categories that can link the head and the modifier nouns in a noun compound: *of, for, with, in, on, at, about*, and *from*. However, according to this classification, the noun compound *love story*, for instance, can be classified both as *story* **of** *love* and *story* **about** *love*. The main problem with these abstract categories is that much of the meaning of individual compounds is lost, and sometimes there is no way to decide whether a form is derived from one category or another. On the other hand, lists of very specific semantic relations are difficult to build as they usually contain a very large number of predicates, such as the list of all possible verbs that can link the noun constituents. Finin (Finin, 1980), for example, uses semantic categories such as "**dissolved in**" to build interpretations of compounds such as "*salt water*" and "*sugar water*".

In this research we experiment with two sets of semantic classification categories defined at different abstraction levels. The first is a core set of 22 semantic relations (22 SRs), set which was identified by us from the linguistics literature and from various experiments after many iterations over a period of time (Moldovan and Girju, 2003)[4]. We proved empirically that this set is encoded by noun – noun pairs in noun phrases and is a subset of our larger list of 35 semantic relations. This list, presented in Table 1 along with examples and semantic argument frames, is general enough to cover a large majority of text semantics while keeping the semantic relations to a manageable number. A semantic argument frame is defined for each semantic relation and indicates the position of each semantic argument in the underlying relation. For example, "Arg_1 is part of (whole) Arg_2" identifies the part (Arg_1) and the whole (Arg_2) entities of this relation. This representation is important since it allows to distinguish between different arrangements of the arguments for given relation instances. For example, most of the time, in N N compounds Arg_1 precedes Arg_2, while in N P N constructions the position is reversed (Arg_2 P Arg_1). However, this is not always the case as shown by N N instances such as "*ham*/Arg1 *sandwich*/Arg2" and "*door*/Arg2 *knob*/Arg1". These argument frames were introduced to provide consistent guide to the annotators to easily test the goodness-of-fit of the relations.

The second set is Lauer's list of 8 prepositions and can be applied only to noun–noun compounds. We selected these two state-of-the-art sets as they are of different size and contain semantic classification categories at different levels of abstraction. Lauer's list is more abstract and, thus capable of encoding a large number of noun compound instances found in a corpus, while our list contains finer grained semantic categories. Details about the coverage of these semantic lists on the two different corpora (Europarl and CLUVI), how well they solve the interpretation problem of noun phrases, and the mapping from one list to another are provided in a companion paper (Girju, 2007).

4 The data

For a better understanding of the semantic relations encoded by N N and N P N instances, we analyzed the semantic behavior of these constructions on a large cross-linguistic corpora of examples. Our intention is to answer questions such as:

(1) *What syntactic constructions are used to translate the English instances to the target Ro-*

[4]There are also other lists of semantic relations used by the research community (e.g., (Barker and Szpakowicz, 1998)), but they overlap considerably with our list of 22-SR.

No.	Semantic Relations	Default argument frame	Examples
1	POSSESSION	Arg_1 POSSESSES Arg_2	*family#2/Arg_1 estate#2/Arg_2*
2	KINSHIP	Arg_1 IS IN KINSHIP REL. WITH Arg_2	*the boy#1/Arg_1's sister#1/Arg_2*
3	PROPERTY	Arg_2 IS PROPERTY OF Arg_1	*lubricant#1/Arg_1 viscosity#1/Arg_2*
4	AGENT	Arg_1 IS AGENT OF Arg_2	*investigation#2/Arg_2 of the crew#2/Arg_1*
5	TEMPORAL	Arg_2 IS TEMPORAL LOCATION OF Arg_1	*morning#1/Arg_2 news#3/Arg_1*
6	DEPICTION-DEPICTED	Arg_1 DEPICTS Arg_2	*a picture#1 Arg_1 of the nice#1/Arg_2*
7	PART-WHOLE	Arg_2 IS PART OF (whole) Arg_1	*faces#1/Arg_2 of children#1/Arg_1*
8	HYPERNYMY (IS-A)	Arg_2 IS A Arg_1	*daisy#1/Arg_2 flower#1/Arg_1*
9	CAUSE	Arg_1 CAUSES Arg_2	*scream#1/Arg_2 of pain#1/Arg_1*
10	MAKE/PRODUCE	Arg_1 PRODUCES Arg_2	*chocolate#2/Arg_2 factory#1/Arg_1*
11	INSTRUMENT	Arg_2 IS INSTRUMENT OF Arg_1	*laser#1/Arg_2 treatment#1/Arg_1*
12	LOCATION	Arg_2 IS LOCATED IN Arg_1	*castle#1/Arg_2 in the desert#1/Arg_1*
13	PURPOSE	Arg_2 IS PURPOSE OF Arg_1	*cough#1/Arg_2 syrup#1/Arg_1*
14	SOURCE	Arg_2 IS SOURCE OF Arg_1	*grapefruit#2/Arg_2 oil#3/Arg_1*
15	TOPIC	Arg_2 IS TOPIC OF Arg_1	*weather#1/Arg_2 report#2/Arg_2*
16	MANNER	Arg_2 IS MANNER OF Arg_1	*performance#3/Arg_1 with passion#1/Arg_2*
17	MEANS	Arg_2 IS MEANS OF Arg_1	*bus#1/Arg_2 service#1/Arg_1*
18	EXPERIENCER	Arg_1 IS EXPERIENCER OF Arg_2	*the girl#1/Arg_1's fear#1/Arg_2*
19	MEASURE	Arg_2 IS MEASURE OF Arg_1	*cup#2/Arg_2 of sugar#1/Arg_1*
20	RESEMBLANCE/TYPE	Arg_2 RESEMBLES OR IS A TYPE OF Arg_1	*framework#1/Arg_1 law#2/Arg_2*
21	THEME	Arg_2 IS THEME OF Arg_1	*acquisition#1/Arg_1 of stock#1/Arg_2*
22	BENEFICIARY	Arg_1 IS BENEFICIARY OF Arg_2	*reward#1/Arg_2 for the finder#1/Arg_1*
		OTHERS	*altar#1 boys#1*

Table 1: The set of 22 semantic relations along with examples interpreted in context and the semantic argument frame.

mance languages and vice-versa? (cross-linguistic syntactic mapping),

(2) *What semantic relations do these constructions encode?* (cross-linguistic semantic mapping),

(3) *What is the corpus distribution of the semantic relations per each syntactic construction?*, and finally

(4) *What is the role of English and Romance prepositions in the NP interpretation?*

Thus, we collected the data from two text collections with different distributions and of different genre, Europarl and CLUVI.

The Europarl text collection

Europarl is a parallel corpora of over 20 million words in eleven official languages of the European Union covering the proceedings of the European Parliament from 1996 to 2001. The corpus was assembled by combining four of the bilingual sentence-aligned corpora made public as part of the freely available Europarl corpus. Specifically, the Spanish-English, Italian-English, French-English and Portuguese-English corpora were automatically aligned based on exact matches of English translations. Then, only those English sentences which appeared verbatim in all four language pairs were considered. The resulting English corpus contained 10,000 sentences which were syntactically parsed (Charniak, 2000). From these we extracted the first 3,000 NP instances (N N: 48.82% and N P N: 51.18%).

The CLUVI text collection

CLUVI (Linguistic Corpus of the University of Vigo) is an open text repository of parallel corpora of contemporary oral and written languages, resource that besides Galician also contains literary text collections in other Romance languages. We focused only on the English-Portuguese and English-Spanish literary parallel texts from the works of John Steinbeck, H. G. Wells, J. Salinger, among others. Using the CLUVI search interface we created a sentence-aligned parallel corpus of 2,800 English-Spanish and English-Portuguese sentences. The English versions were automatically parsed after which each N N and N P N instance thus identified was manually mapped to the corresponding translations. The resulting corpus contains 2,200 English instances with a distribution of 26.77% N N and 73.23% N P N.

4.1 Corpus annotation

For each corpus, each NP instance was presented separately to two experienced annotators[5] in a web interface in context along with the English sentence and its translations. Since the corpora do not cover some of the languages (Romanian in Europarl and CLUVI, and Italian and French in CLUVI), three other native speakers of these languages and fluent in English provided the translations which were added to the list.

WordNet senses

The two computational semantics annotators had to tag each English constituent noun with its corresponding WordNet sense[6]. If the word was not found in WordNet the instance was not considered.

Tagging each noun constituent with the corresponding WordNet sense in context is important not only as a feature employed in the training models, but also as guidance for the annotators to select the right semantic relation. For instance, in the following sentences, *daisy flower* expresses a PART-WHOLE relation in (1) and a IS-A relation in (2) depending on the sense of the noun *flower* (cf. WordNet 2.1: *flower#2* is a "reproductive organ of angiosperm plants especially one having showy or colorful parts", while *flower#1* is "a plant cultivated for its blooms or blossoms").

(1) "Usually, more than one *daisy#1 flower#2* grows on top of a single stem."

(2) "Try them with orange or yellow flowers of red-hot poker, solidago or other late *daisy#1 flowers#1*, such as rudbeckias and heliopsis."

In cases where noun senses were not enough for relation selection, the annotators had to rely on a larger context provided by the sentence and its translations as shown below.

Semantic argument frame

The annotators were also asked to identify the translation phrases, tag each instance with the corresponding semantic relation, and identify the semantic arguments Arg_1 and Arg_2 in the semantic argument frame of the corresponding relation.

[5]The annotators have extensive expertise in computational semantics and are fluent in at least two of the Romance languages considered for this task.

[6]For the purpose of this research we used WordNet 2.1.

Thus, since the order of the semantic arguments in an NP is not fixed (Girju et al., 2005), the annotators were presented with the semantic argument frame for each of the 22 semantic relations and were asked to tag the NP instances accordingly. For example, in PART-WHOLE instances such as *chair*/Arg2 *arm*/Arg1 the part *arm* follows the whole *chair*, while in *button*/Arg1 *shirt*/Arg2 the order is reversed.

Translation instances

In the annotation process the annotators were asked to identify and use, if necessary, the five corresponding translations as additional information in selecting the semantic relation. Since only N N and N P N noun phrase constructions were considered, the annotators had to discard those instances encoded by different syntactic constructions in the Romance languages.

For instance, the context provided by the Europarl English sentence in (3) below does not give enough information for the disambiguation of the English noun phrase "*judgment of the presidency*" which can mean either AGENT or THEME. The annotators had to rely on the Romance translations in order to identify the correct meaning in context (in this case THEME): *valoración sobre la Presidencia* (Es.), *avis sur la présidence* (Fr.), *giudizio sulla Presidenza* (It.), *veredicto sobre a Presidência* (Port.), *evaluarea Presendiţiei* (Ro.)[7].

(3)

En.: "If you do , *our final judgment of the Spanish presidency* will be even more positive than it has been so far."

Es.: "Si se hace, nuestra valoración sobre la Presidencia española del Consejo será aún mucho más positiva de lo que es hasta ahora."

Fr.: "Si cela arrive, notre avis sur la présidence espagnole du Conseil sera encore beaucoup plus positif que ce n'est déjà le cas."

It.: "Se ci riuscirà il nostro giudizio sulla Presidenza spagnola sarà ancora più positivo di quanto non sia stato finora."

[7]En. means English, Es. – Spanish, Fr. – French, It. – Italian, Port. – Portuguese, and Ro. – Romanian.

Port.: "Se isso acontecer, o nosso veredicto sobre a Presidência espanhola será ainda muito mais positivo do que o actual."
Ro.: "Dacă are loc, evaluarea Preşedinţiei spaniole va fi încă mai pozitivă decât până acum."

Semantic relations

Whenever the annotators found an example encoding a semantic relation or a preposition paraphrase other than those provided or they didn't know what interpretation to give, they had to tag it as OTHER-SR and OTHER-PP, respectively . For example, in the CLUVI sentences (4) and (5) below, the noun phrases *melody of the pearl* and *cry of death* (the cry announcing death) were tagged as OTHER-SR since here the context of the sentences does not indicate the association between the two nouns. Moreover, noun compound instances such as *the corner box* and *knowledge searches* were tagged as OTHER-PP (*box* **in** *the corner*, *searches* **after** *knowledge*).

(3) LPE-284: "And because the need was great and the desire was great, the little secret *melody of the pearl* that might be was stronger this morning." (En.)

(4) LPE-1582: "And then Kino's brain cleared from its red concentration and he knew the sound - the keening, moaning, rising hysterical cry from the little cave in the side of the stone mountain, *the cry of death*." (En.)

Moreover, most of the time one instance was tagged with one semantic relation, and respectively preposition paraphrase, but there were also situations in which an example could belong to more than one classification category in the same context. For example, *Texas city* is tagged as PART-WHOLE/PLACE-AREA, but also as a LOCATION relation using the 22-SR classification category, and respectively as *of, from, in* based on the 8-PP category (e.g., *city of Texas*, *city from Texas*, and *city in Texas*). Other instances, however, can encode a total of three semantic relations in a particular context. One such instance is *cup#2 of hot_chocolate#1* in example (6) below, which was tagged in CLUVI as MEASURE/OTHER(CONTENT-CONTAINER)/LOC. Sense #2 of *cup* in WordNet refers to "the quantity the cup will hold" (cf. WordNet 2.1), thus mostly indicating a MEASURE relation.

(5) 557-AGU: "Wouldn't you like a cup of hot chocolate before you go?" (En.)

However, since most hot beverages (such as tea, coffee, and chocolate) are served in cups, it stands to reason that the instance can be easily paraphrased as a cup holding hold chocolate. Although our current NP interpretation system (Girju, 2007) does not differentiate between LOCATION and CONTENT-CONTAINER (as other researchers (Tyler and Evans, 2003)[8], we consider CONTENT-CONTAINER as a special type of LOCATION), we capture them in our annotation scheme.

Other examples of multiple annotations are MEASURE/PART-WHOLE (e.g., *an abundance of buildings, a bunch of guys*), Overall, 0.5% Europarl and 6.9% CLUVI instances were tagged with more than one semantic relation, and almost all noun compound instances were tagged with more than one preposition.

Thus, the annotated instances used in the corpus analysis and system training phases have the following format: <NP$_{En}$;NP$_{Es}$; NP$_{It}$; NP$_{Fr}$; NP$_{Port}$; NP$_{Ro}$; target>. The word *target* is one of the 23 (22 + OTHER) semantic relations or one of the eight prepositions considered. For example, <*judgment#2/Arg_1 of presidency#2/Arg_2; valoración sobre la Presidencia; avis sur la présidence; giudizio sulla Presidenza; veredicto sobre a Presidência; evaluarea Preşedinţiei*; THEME>.

4.2 Inter-annotator agreement

The annotators' agreement was measured using 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 annotators 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.

[8](Tyler and Evans, 2003) cite child language acquisition studies which show there is a strong cognitive relationship between LOCATION and CONTENT-CONTAINER.

The Kappa values obtained on each corpus are shown in Table 2. We also computed the number of pairs that were tagged with OTHER by both annotators for each semantic relation and preposition paraphrase, over the number of examples classified in that category by at least one of the judges. For the noun compound instances that encoded more than one classification category, the agreement was done on one of the relations only.

The agreement obtained for the Europarl corpus is higher than the one for CLUVI on both classification sets. This is partially explained by the distribution of semantic relations in both corpora. Overall, the K coefficient shows a fair to good level of agreement for the corpus data on the set of 22-SRs, taking into consideration the task difficulty. The level of agreement for the prepositional paraphrases was much higher. All these can be explained by the instructions the annotators received prior to the annotation and by their expertise in lexical semantics.

Corpus	Classification tag sets	Kappa Agreement		OTHER
		N N	N P N	
Europarl	**8-PP**	0.80	N/A	91%
	22-SR	0.61	0.67	78%
CLUVI	**8-PP**	0.77	N/A	86%
	22-SR	0.56	0.58	69%

Table 2: The inter-annotator agreement on the NP annotation on the two corpora. For the noun compound instances that encoded more than one semantic classification category, the agreement was done on one of the relations only. "N/A" means not applicable.

13.05% of Europarl[9] and 1.9% of CLUVI instances that could not be tagged with Lauer's prepositions were included in OTHER-PP category. About 99% of the Europarl N N instances encode TYPE relations (e.g., *framework law*), while in CLUVI most of them were TYPE (e.g., *nightmare sensation*), followed by OTHER-SR (e.g., *altar boys*), and IS-A (e.g., *Winchester carbine*).

From the initial corpus we considered those English instances that had all the translations encoded by N N and N P N. Out of these, we selected only 1,023 Europarl and 1,008 CLUVI instances encoded by N N and N P N in all languages considered and resulted after agreement[10]. We split the corpora using a 8:2 training - test ratio and used it to train and test our system. Details about the experiments and the results obtained are presented in (Girju, 2007).

5 Discussion and conclusions

In this paper we presented some observations on our experience with an annotation scheme that was used in the training of a state-of-the-art noun phrase semantic interpretation system. These observations are defined in the framework of a larger project. This project is to investigate various linguistic issues and develop specific language models for the interpretation of noun phrase constructions in Germanic, Romance, and other classes of languages.

Our approach to NP interpretation, and thus annotation procedure, is novel in several ways. We define the problem in a cross-linguistic framework and provide empirical observations on various annotation issues based on a set of two different corpora using two state-of-the-art classification tag sets: Lauer's prepositions and our list of 22 relations.

The linguistic implications are also important to mention here. The annotation investigations done in this research provide new insights into the research topic at hand, the semantic interpretation of noun phrases, in particular and the identification of semantic relations between nominals (irrespective of the syntactic constructions that link the two nouns), in general. One such linguistic aspect is the importance of context for this task. Sometimes, the local context of the noun phrase is not enough to disambiguate the underlying instances. For this, the annotators need to relay on world and domain specific knowledge and the entire context of the sentence, or consider a larger context window (from a simple paragraph including the sentence, to the discourse of the text) as shown below in (6), (7), and (8). In (6) and (7), for example, neither the context of the sentence, nor the context of their paragraph provide the meaning of the NPs. Many of the CLUVI instances tagged as OTHER-SR (such as *the music of the pearl* in (6)), are naming phrases – they were defined only once in the text collection and later on mentioned to refer to the initial concept.

In (8), on the other hand, the meaning of the NP *the destruction of the Palestinian Authority* is THEME and not AGENT as might be considered by default.

[9]Only 5.70% of the TYPE instances in the Europarl corpus were unique.

[10]The annotated corpora resulted in this research are available at http://apfel.ai.uiuc.edu.

(6) LPE-390: “And *the music of the pearl* rose like a chorus of trumpets in his ears.” (CLUVI)

(7) “Mr President, *the violent destruction of the State of Israel*.” (Europarl)

(8) “The spread of the settlements, the seizing of land, the curfews, the Palestinians imprisoned in their own villages, the summary executions, the ambulances prevented from reaching their destinations, the women giving birth at check points, *the destruction of the Palestinian Authority*: these are not mistakes or accidents.” (Europarl)

6 Acknowledgments

We would like to thank all the people who helped with the corpus creation and annotation, and those with whom we had nice discussions about various semantic relations. Without them this research wouldn't have been possible: Archna Bhatia, Gustavo Cavallin, Brian Drexler, Matt Garley, Tania Ionin, Matt Niemi, Dustin Parr, and Chris Struven. And last, but not least we like to thank the reviewers for their useful comments.

References

K. Barker and S. Szpakowicz. 1998. Semi-automatic recognition of noun modifier relationships. In *the Proceedings of the Association for Computational Linguistics / Conference on Computational Linguistics*.

M. Berland and E. Charniak. 1999. Finding Parts in Very Large Corpora. In *the Proceedings of the Association for Computational Linguistics (ACL)*, University of Maryland.

E. Charniak. 2000. A Maximum-entropy-inspired Parser. In *the Proceedings of the North American Chapter of the Association for Computational Linguistics (NAACL)*, Seattle, Washington.

P. Downing. 1977. On the Creation and Use of English Compound Nouns. *Language*, 53(4):810–842.

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

R. Girju, D. Moldovan, M. Tatu, and D. Antohe. 2005. On the semantics of noun compounds. *Computer Speech and Language*, 19(4):479–496.

R. Girju, A. Badulescu, and D. Moldovan. 2006. Automatic discovery of part-whole relations. *Computational Linguistics*, 32(1).

R. Girju. 2007. Improving the interpretation of noun phrases with cross-linguistic information. In *the Proceedings of the Association for Computational Linguistics (ACL)*, Prague.

O. Jespersen. 1954. *A Modern English Grammar on Historical Principles*. London.

S. N. Kim and T. Baldwin. 2006. In *the Proceedings of the Association for Computational Linguistics*, Sydney, Australia.

M. Lapata and F. Keller. 2004. The Web as a baseline: Evaluating the performance of unsupervised Web-based models for a range of NLP tasks. In *the Proceedings of the Human Language Technology Conference / North American Chapter of the Association of Computational Linguistics (HLT-NAACL)*.

M. Lauer. 1995. Corpus statistics meet the noun compound: Some empirical results. In *the Proceedings of Association for Computational Linguistics (ACL)*, Cambridge, Mass.

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

D. Moldovan and R. Girju. 2003. Knowledge discovery from text. In *the Tutorial Proceedings of the Association for Computational Linguistics (ACL)*, Sapporo, Japan.

D. Moldovan, A. Badulescu, M. Tatu, D. Antohe, and R. Girju. 2004. Models for the semantic classification of noun phrases. In *the Proceedings of the HLT/NAACL Workshop on Computational Lexical Semantics*, Boston, MA.

P. Nakov and M. Hearst. 2005. Search engine statistics beyond the n-gram: Application to noun compo und bracketing. In *the Proceedings of the Computational Natural Language Learning Conference*.

P. Pantel and M. Pennacchiotti. 2006. Espresso: Leveraging generic patterns for automatically harvesting semantic relations. In *the Proceedings of the International Conference for Computational Linguistics (COLING/ACL)*, Sydney, Australia.

M. Pennacchiotti and P. Pantel. 2006. Ontologizing semantic relations. In *the Proceedings of Conference on Computational Linguistics / Association for Computational Linguistics (COLING/ACL-06)*, Sydney, Australia. Association for Computational Linguistics.

B. Rosario and M. Hearst. 2001. Classifying the semantic relations in noun compounds. In *the Proceedings of the 2001 EMNLP Conference*.

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*.

E. Selkirk. 1982. Syntax of words. In *Linguistic Inquiry Monograph*. MIT Press.

R. Snow, D. Jurafsky, and A. Ng. 2006. Semantic taxonomy induction from heterogenous evidence. In *the Proceedings of the Conference on Computational Linguistics / Association for Computational Linguistics (COLING-ACL)*, Sydney, Australia.

P. Turney. 2006. Expressing implicit semantic relations without supervision. In *the Proceedings of the Conference on Computational Linguistics / Association for Computational Linguistics (COLING/ACL)*, Sydney, Australia.

A. Tyler and V. Evans. 2003. *Spatial Experience, Lexical Structure and Motivation: The Case of In*. In G. Radden and K. Panther. Studies in Linguistic Motivation. Berlin and New York: Mouton de Gruyter.