Identifying Semantic Relations in Context: Near-misses and Overlaps

Alla Rozovskaya and Roxana Girju
University of Illinois at Urbana-Champaign
{rozovska,girju}@illinois.edu

Abstract

This paper addresses the problem of semantic relation identification for a set of relations difficult to differentiate: *near-misses* and *overlaps*. Based on empirical observations on a fairly large dataset of such examples we provide an analysis and a taxonomy of such cases. Using this taxonomy we create various contingency sets of relations. These semantic categories are automatically identified by training and testing three state-of-the-art semantic classifiers employing various feature sets. The results show that in order to identify such near-misses and overlaps accurately, a semantic relation identification system needs to go beyond the ontological information of the two nouns and rely heavily on contextual and pragmatic knowledge.

Keywords

lexical semantics; semantic relations; machine learning

1 Introduction

Although semantic relations have been studied for a long time both in linguistics and natural language processing, they received special attention recently due to research done in various knowledge-rich tasks such as question answering [3, 17], information retrieval [11], and textual entailment [24].

The identification of semantic relations between nominals is the task of recognizing the relationship between two nouns in context. For example, the noun pair (cycling, happiness) encodes a Cause-Effect relation in the sentence *He derives great joy and happiness from cycling*. This task requires several local and global decisions needed for relation identification. This involves the meaning of the two noun entities along with the meaning of other words in context.

The problem, while simple to state is hard to solve. The reason is that the meaning encoded by the two nouns is not always explicitly stated in context. Despite the encouraging results obtained by the participating systems at the SemEval-2007 - Task 4: *Classification of Semantic Relations between Nominals* [9], the problem needs further analysis. For example, the set of semantic relations considered for this problem needs to be better understood. Thus, a more thorough analysis of semantic relations needs to be done before building systems capable of recognizing them automatically in context. Particular attention should be given to those semantic relations that are difficult to differentiate (near-misses) and those relations that coexist in some particular contexts (overlaps). Consider for example the following sentences:

(1)   a.   I got home and big $\langle e1 \rangle$branches$\langle /e1 \rangle$ had fallen off the $\langle e2 \rangle$tree$\langle /e2 \rangle$ into the driveway.

    b.   He fell off the $\langle e1 \rangle$tree$\langle /e1 \rangle$ and hit every $\langle e1 \rangle$branch$\langle /e1 \rangle$ on the way down.

(2)   Whisk together the mustard, vinegar and two $\langle e1 \rangle$teaspoons$\langle /e1 \rangle$ of the remaining $\langle e2 \rangle$lemon juice$\langle /e2 \rangle$.

(1)a and (1)b are near-misses since the same noun-noun concept pair *branch - tree* encodes Origin-Entity in (1)a and Part-Whole in (1)b (the branches are still part of the tree). In example (2) the noun-noun concept pair *teaspoon - lemon_juice* encodes Part-Whole (in particular Portion-Mass, a subtype of Part-Whole relations), but also Measure and Content-Container, so these three relations coexist in the context of the same sentence.

These semantic relations are difficult to differentiate, and thus pose a challenging problem to the learning models. Some of these relations can coexist only in some contexts, and this overlap is not genuine but is influenced by contextual and pragmatic factors. Consider, for example, Part-Whole, Content-Container, and Measure. These relations coexist for some special classes of nouns (e.g., *glass, cup* can mean either container or quantity) which have a dual semantic nature and thus, performing what is called a metonymic shift. For example, a simple analysis of the hits returned by Google for the noun phrase *glass of wine* showed that its interpretation is highly contextual: ".. I enjoyed/broke a glass of wine", etc. Here, the verb selects either the *wine* or the *glass* as point of focus. In case the focus is *wine*, the meaning is Measure, and since liquid substances come in containers then it also implies Content-Container. However, when the focus is on *glass*, the Content-Container interpretation does not necessarily imply Measure (the glass might not be full). This is just an example of many other clusters of such relations which need to be further analyzed.

Although there have been recent attempts in this direction (the consideration of near-misses as negative examples for each semantic relation at SemEval 2007 - Task 4), to our knowledge there is no systematic study of clusters of closely related and overlapping semantic relations.

In this paper we provide an analysis of a set of five most frequently occurring semantic relations which are near-misses (Part-Whole, Origin-Entity, Purpose) and overlaps (Part-Whole, Measure, Content- Container). Moreover, we compare the performance of three state-of-the-art relation identification systems which employ different feature sets: (1) an improved implementation of a supervised model, SemScat2 [2], (2) the SNoW machine learning architecture [23], and (3) a competitive 1007 SemEval-Task 4 system [1]. The systems were trained and tested on a corpus of 1,000 examples.

The results show that in order to identify such near-misses and overlaps accurately, a semantic relation iden-

tification system needs to go beyond the ontological information of the two nouns and rely heavily on contextual and pragmatic knowledge.

The paper is organized as follows. In the next section we present previous work, followed by an analysis of semantic relations. In particular, we provide a classification of clusters of near-miss and overlapping relations based on empirical observations. In Section 4 we present the models employed. Finally, we present various experiments and discuss the results.

2 Previous Work

Most of the attempts in the area of noun - noun semantic interpretation have tackled the problem either out of context (mostly in linguistics: [26], [15]) or in different limited syntactic contexts (linguistics and computational linguistics), such as noun–noun compounds and other noun phrases (e.g., "N preposition N", "N, such as N"), and "N verb N". More recently, in the datasets provided as part of the SemEval-2007, Task 4, the nouns could occur anywhere in the sentence. Moreover, in what concerns the set of semantic relations used, state of the art systems follow roughly two main directions: interpretation based on semantic similarity with previous seen examples [22], [16], [20], and semantic disambiguation relative to an underlying predicate or semantically-unambiguous paraphrase [13], [12].

Most methods employ rich ontologies, such as WordNet or look for local paraphrases (such as "N prep. N" or "N verb N") and disregard the sentence context in which the nouns occur, partly due to the lack of annotated contextual data on which they are trained and tested, and partly due to the claim that axioms and ontological distinctions are more important than the information derived from the context in which the nouns occur. We also support this claim, based on the results of our recent experiments [2] which show that some relations such as Part-Whole, Origin-Entity, and Content-Container are better fitted for an ontological approach than others. However, even these relations are difficult to identify from a pool of examples containing near misses. For example, at SemEval 2007, Origin-Entity was identified as one of the most difficult relation. 99% and 73% of the 11 B-type systems (WordNet-based) identified the relation as one of the three, and respectively two most difficult relations to classify.

In this paper we show that for near miss and overlapping contextual semantic and pragmatic data, semantic interpretation systems need to explore both the linguistic context of the sentence and the context of use (pragmatics).

3 Semantic Relations: Analysis of Near-misses and Overlaps

There are to date several sets of semantic relations that have been widely used in the computational linguistics community. In 1995 Lauer [14] proposed a set of 8 prepositions as semantic classification categories: {*of, for, with, in, on, at, about, from*}. Others [20] have used more specific relations organized into a two-level hierarchy, splitting 5 relations in the top level (Causal, Participant, Spatial, Temporal, Quality) into 30 more specific. Moldovan & Girju

[18] presented a list of 35 semantic relations which overlaps considerably with that of Nastase & Szpakowicz [20].

In 2007, the SemEval-Task 4 organizers introduced a collection of 7 semantic relations which were chosen from the most frequently used ones in the literature: Cause-Effect, Instrument-Agency, Product-Producer, Origin-Entity, Theme-Tool, Part-Whole, and Content-Container.

Thus, these semantic relations need to be studied in more detail in order to build accurate relation classifiers.

A closer look at the inter-annotator agreements reported in the computational linguistics literature on various relations and the annotation comments from the freely available SemEval-Task 4 dataset [9] shows that some relations cluster together either as near-misses or overlaps. For example, Girju et al. [5] report a Kappa inter-annotator agreement of about 0.83 on Part-Whole[1], while Panachiotti & Pantel [21] list an agreement of about 0.73 on two non overlapping relations, Part-Whole and Cause-Effect. For larger sets of semantic relations the inter-annotator agreement is much lower. For example, Girju et al. [7] report a Kappa agreement of about 0.6 on a set of 35 relations and SemEval organizers report an average agreement of about 70.3% on the 7 SemEval relations (a much higher agreement was obtained after discussions). Moreover, the SemEval annotators' comments and suggestions made and listed as part of the released datasets, along with our own observations on various data collections show that annotation disagreements are mainly attributed to the fact that various semantic relations can occur in very similar contexts or can even coexist/overlap in the context of the same sentence. These relations form what we call a `contingency set`.

For this research, we focused on the SemEval-Task 4 datasets and the publicly available cluvi-europarl text collection[2] [4]. The cluvi-europarl collection is presented in the SemEval Task 4 format and is based on a set of 22 semantic relations overlapping with that of one used at SemEval 2007. The collection contains 2,031 (1,023 europarl and 1,008 cluvi) instances. The Kappa values were obtained on europarl (N N: 0.61; N P N: 0.67) and cluvi (N N: 0.56; N P N: 0.68).

In the next subsections we present the data used in this research, an evaluation of the frequently occurring set of such semantic relations, and propose a classification of contingency relations.

3.1 SemEval Task 4: Classification of Semantic Relations between Nominals

The SemEval 2007 task on semantic relations between nominals is to identify the underlying semantic relation between two nouns in the context of a clause. Since there is no concensus on the number and abstraction level of semantic relations, the SemEval effort focused on seven frequently occurring semantic relations listed by many researchers in their lists of relations [20, 6, 8]: Cause–Effect, Instrument–Agency, Product–Producer, Origin–Entity, Theme–Tool, Part–Whole, and Content–Container. The dataset provided consists of a definition file and 140

---

[1] Girju et al. [5] trained the annotators providing explicit annotation schemas based on a well defined classification of 6 subtypes of Part-Whole relations [25].

[2] This collection is freely available at:
$http://apfel.ai.uiuc.edu/resources.html$.

training and about 70 test sentences for each of the seven relations considered. The definition file for each relation includes a detailed definition, restrictions and conventions, and prototypical positive and near–miss negative examples. For example, the Part–Whole relation is defined as follows [9]:

Definition Part–Whole(X, Y) is true for a sentence S that mentions entities X and Y iff:

(a) X and Y appear close in the syntactic structure of S (we do not assign the relation to entities from separate clauses in a composite sentence);

(b) according to common sense, the situation described in S entails that X is the part of Y.

(c) X and Y follow the constraints proposed by Winston et al. 1987 [25] in a classification into six specialized types of the Part-Whole relation, of which we consider five [..]

Winston et al. 1987 [25] performed psycholinguistic experiments to identify Part–Whole instances based on the way in which the parts contribute to the structure of the wholes. Here detailed restrictions are listed for X and Y for five subtypes of Part–Whole relations: Component–Integral (e.g., *wheel–car*), Member–Collection (e.g., *soldier–army*), Portion–Mass (e.g., *slice–pie*), Stuff–Object (e.g., *silk–dress*), and Place–Area (e.g., *oasis–desert*).

For each relation, the instances were selected by applying wild–card search patterns on Google. The patterns were built manually, using the approach of Hearst 1992 [10] and Nakov & Hearst 2006 [19]. Examples of queries which potentially select Part–Whole instances are "* is part of *", "* has *", "* contains *".

Each SemEval-Task4 organizer was responsible for a particular semantic relation for which they collected a corpus of instances. Each instance in this corpus was then annotated by two other organizers. In each training and test example sentence, the nouns were identified and manually labeled with their corresponding WordNet 3.0 senses (given as sense keys). The average inter-annotator agreement on relations (true/false) after the independent annotation step was 70.3%, and the average agreement on WordNet sense labels was 71.9%. In the process of arriving at a consensus between annotators, the definition of each relation was revised to cover explicitly cases where there had been disagreement.

Table 1 shows all seven relations considered along with the positive/negative instance distribution and examples.

Moreover, each example was accompanied by the heuristic pattern (query) the relation organizer used to extract the sentence from the web and the position of the arguments in the relation. Positive and negative instances of the Part–Whole relation are listed in the examples (3) and (4) below. Part–Whole relations are semantically similar to other relations such as Origin–Entity, and Content–Container, and thus difficult to differentiate automatically. Instances encoding these relations are called near–miss examples, as shown in (4).

(3) 026 "He caught her $<e_1>$arm$</e_1>$ just above the $<e_2>$wrist$</e_2>$."
WordNet(e_1) = "arm%1:08:00::", WordNet(e_2) = "wrist%1:08:00::", Part-Whole(e_2, e_1) = "true", Query = "* just above the *"
Comment: Component–Integral object

(4) "Not sure what brand of model it came from but the $<e_1>$wings$</e_1>$ are from a $<e_2>$trashed plane$</e_2>$ my buddy had."
Comment: Origin–Entity

The example in (4) is interpreted by inferring that the wings have been taken from a plane of which they used to be part. This goes way beyond sentential context into very complex inferences about our knowledge about the world.

The task is defined as a binary classification problem. Thus, given a pair of nouns and their sentential context, a semantic interpretation system decides whether the nouns are linked by the target semantic relation. Based on the information employed, systems can be classified in four types of classes:

(A) systems that use neither the given WordNet synsets nor the queries,
(B) systems that use only WordNet senses,
(C) systems that use only the queries, and
(D) systems that use both WordNet senses and queries.
Detailed information about the SemEval-Task4 data and procedure can be found in [9].

3.2   The Data

We have identified some initial contingency sets of relations from the annotations, comments, definitions, and constraints provided as part of the SemEval datasets. Then we looked in cluvi and europarl datasets for examples involving these contingency sets. The most frequently occurring contingency sets we identified are {Part-Whole, Origin-Entity}, {Part-Whole, Purpose}, {Origin-Entity, Purpose}, {Part-Whole, Measure}, {Part-Whole, Content-Container}. It is interesting to note that most of these contingency sets involve Part-Whole.

As a next step, we relabeled the Part-Whole relations in the mentioned datasets with their five subtypes according to the context of the sentence. This was a relatively easy task since the five Part-Whole subtypes are well defined and many of the Part-Whole relations in SemEval and cluvi-europarl collections were already identified with these subtypes in the "Comment" sections.

For the other relations in the identified contingency sets, we selected only those examples in which the noun-noun pair was one of the five subtypes of Part-Whole. For example, Origin-Entity relations can hold between Component-Integral nouns (e.g., *apple - seed*: "The seeds were removed from the apple") as well as between Entity-Location nouns (e.g., *China - cup*: "I got the cup from China"). The rational was to focus only on semantic relations that are near misses.

Thus we built an initial corpus of 1,109 examples. The distribution is shown in Table 2.

In order to provide a fairly balanced corpus of examples for the set of semantic relations considered, we followed the procedure used by the SemEval annotators and searched the web using various relevant queries. Since we found only a few examples for Place-Area, Phase-Activity, and Indeterminate (the relation was not clear from the context) we did not include them in the final corpus.

Two annotators familiar with the task provided the semantic relations and the noun sense keys in context following the format used at SemEval 2007. This was somewhat a trivial task since all of the examples from SemEval

Relation	Training data		Test data		Example
	positive	total size	positive	total size	
Cause-Effect	52.14%	140	51.25%	80	laugh (cause) wrinkles (effect)
Instrument-Agency	50.71%	140	48.71%	78	laser (instrument) printer (agency)
Product-Producer	60.71%	140	66.67%	93	honey (product) bee (producer)
Origin-Entity	38.57%	140	44.44%	81	message (entity) from outer-space (origin)
Theme-Tool	41.43%	140	40.84%	71	news (theme) conference (tool)
Part-Whole	46.43%	140	36.11%	72	the door (part) of the car (whole)
Content-Container	46.43%	140	51.35%	74	apples (content) in the basket (container)

Table 1: *Data set statistics on each of the seven SemEval relations considered along with he positive/negative instance distribution and examples.*

Relations	Number of examples			Total
	SemEval	cluvi, europarl	web	
Component-Integral	25	140	3	168
Portion-Mass	3	2	167	172
Member-Collection	32	112	26	170
Stuff-Object	6	8	86	100
Place-Area	0	6	4	10
Phase-Activity	0	3	5	8
Origin-Entity	44	6	31	81
Measure	6	22	126	154
Purpose	0	26	95	121
Content-Container	70	18	32	120
Indeterminate	0	0	5	5
Total		1,109		

Table 2: *Semantic relation counts in all the text collections considered.*

and cluvi-europarl collections had the nouns already annotated with corresponding sense keys. The annotators, however, paid special attention to the semantic relation annotation. They provided new labels if they did not agree with the initial annotation (in case of SemEval and cluvi and europarl datasets) or if they thought multiple relations are possible. After this, a third judge analyzed the conflicting cases (total and partial disagreements) and identified 290 partial disagreements (overlaps among the labels proposed by the annotators per example) by collapsing the sets proposed by the annotators. The resulting sets were {Portion-Mass, Member-Collection, Measure}, {Portion-Mass, Measure, Content-Container}, {Portion-Mass, Member-Collection, Measure, Content-Container}, and {Measure, Content-Container}. The data was thus re-labeled to reflect these overlaps. The content of the resulting corpus (1,000 examples) is presented in Table 3 along with examples. Since Portion-Mass and Member-Collection both involve homeomerous parts, we collapsed them in one class called P-Whp – P-W with homeomerous parts (e.g., the whole comprises other parts similar with the part in question). These subtypes of Part-Whole are involved in similar overlaps.

3.3 Data Analysis

Based on the literature and our own observations with the corpus created and presented in the previous subsection and with other text collections, we identified two classes of contingency sets: near-misses and overlaps. We present next a detailed account of each type.

Overlaps
So far we have identified the following types of overlaps:

(A) Genuine, when two or more relations coexist in the same context,

(B) Indeterminate, when two or more relations are possible due to insufficient context information, and

(C) Ill-defined or too general, when two or more relations coexist since some of them are either ill-defined or too general. These include those which overlap with other relations in just one or few of their subtypes. Thus, they need to be revised and further refined.

Overlaps type (A) and (B) are valid overlaps, while overlaps of type (C) are not. These are exemplified in sentences (5) - (7) below. The genuine overlaps we have identified so far are directional. For instance, in (5) Place-Area entails Location (and not the other way around since there are other types of Location which are not Place-Area) and in (6) Measure entails Content-Container. However, the entailment in (6) is of pragmatic nature. This example suggests the idea of amount/measure and Content-Container coexists with Measure, but it is pragmatically inferred from it – since liquids, and especially coffee are usually served in cups. Thus, while the overlap in (5) always holds, the overlaps in (6) and (7) are most of the time resolved by the linguistic context (syntax, semantics) and the context of use (pragmatics).

(5) "Darfur is a ⟨e1⟩region⟨/e1⟩ in western ⟨e2⟩Sudan⟨/e2⟩, Africa.", Relation(e1,e2) = {Place-Area; Location}

(6) "Making a delicious ⟨e1⟩cup⟨/e1⟩ of ⟨e2⟩coffee⟨/e2⟩ is not a magical experience or a hit-and-miss stroke of luck." Relation(e2,e1) = {Measure; Content-Container}

(7) "I set on fire the ⟨e1⟩branches⟨/e1⟩ of the ⟨e2⟩tree⟨/e2⟩, Relation(e2,e1) = {Origin-Entity, Component-Integral}.

The interpretation of the noun pair in example (7) is indeterminate since there is not enough context, so both relations are possible. Sure, we can extend the context to include the entire paragraph or the document it came from. However, even in such situations the interpretation may remain indeterminate. Consider for example the instance *the girl's shoes* which can mean the shoes the girl made, dreams of, buys, wears, etc.

An example of relation which creates a type (C) overlap is Part-Whole. As mentioned in the previous section, this relation belongs to the following contingency set: {Origin-Entity, Purpose, Measure, Content-Container}. However, it does not interact in the same way with each of the relations in the set. For instance, it forms a near-miss set with Origin-Entity and Purpose and overlaps with Measure and Content-Container. The specialization of this relation into its 5 subtypes gives the following contingency relations: Near misses – {Component-Integral, Origin-

No.	Set of semantic categories	Number of instances	Examples
1	Component-Integral	168	"When the first transplant took place at St. Paul's in 1986, the vast majority of *patients* received a new *kidney* from a deceased donor."
2	$Part - Whole_{hp}$	36	"The *catalogue* contained *books* published in 1998 warning of the upcoming millennium bug and other similarly germane works, but neither of Brock's bestsellers."
3	Stuff-Object	100	"Typically, an unglazed *clay pot* is submerged for 15 to 30 minutes to absorb water."
4	Origin-Entity	81	"I got home and big *branches* had fallen off the *tree* into the drive way."
5	Measure	104	"Ensure that you don't lose a *drop* of *juice* by nestling your shellfish in salt."
6	Purpose	121	"023 "All he had on underneath was a phoney *shirt collar*, but no shirt or anything."
7	Content-Container	120	"Among the contents of the *vessel* were a set of carpenter's *tools*, several large storage jars, ceramic utensils..
8	$Part - Whole_{hp}$/ Measure	156	"The contents of the boxes included phone cards, disposable cameras and razors, travel-size toiletries, snack food, and *lots* of *candy*."
9	$Part - Whole_{hp}$/ Measure/ Content-Container	75	"I would have ripe olives and about a *cup* of that leftover *tea*."
10	Measure/ Content-Container	69	"Is enjoying a *glass* of red *wine* with dinner each evening beneficial to your health?"

Table 3: *The set of 10 semantic relation categories considered along with examples.*

Entity, Purpose}, and overlaps – {Portion-Mass, Member-Collection, Measure}, {Portion-Mass, Measure, Content-Container}, {Portion-Mass, Member-Collection, Measure, Content-Container}, and {Measure, Content-Container}. Similarly, other semantic relations may be decomposed into finer grain types, so more specific relation taxonomies may be built.

Near-misses

Near misses are sets of mutually exclusive relations in the context of the same sentence. As shown above, in the empirical investigations of this research we identified the following set of near misses: {Component-Integral, Origin-Entity, Purpose}. For instance, although the pair *branch - tree* (as shown in the examples below) can encode Component-Integral, Origin-Entity and Purpose, only one relation is possible in a given context:

(8) "He grabbed the $\langle e1 \rangle$branches$\langle /e1 \rangle$ of the $\langle e2 \rangle$tree$\langle /e2 \rangle$ to get closer to the nest up high." Relation(e1,e2) = {Component-Integral}

(9) "He took the $\langle e1 \rangle$branches$\langle /e1 \rangle$ he cut from the old $\langle e2 \rangle$tree$\langle /e2 \rangle$ and burned them." Relation(e2,e1) = {Origin-Entity}

(10) "'These plastic $\langle e1 \rangle$branches$\langle /e1 \rangle$ are for the green $\langle e2 \rangle$tree$\langle /e2 \rangle$', said he while showing me how to assemble them." Relation(e2,e1) = {Purpose}.

Table 4 shows the contingency relations identified in this research, types of encountered overlaps, plus constraints observed on the data. Due to an insufficient number of examples, we have not performed any experiments with the {Content-Container, Stuff-Object} contingency set.

4 Models

In order to test the validity of the new set of contingency classes we trained and tested three state-of-the-art classifiers on the 1,000 sentence corpus: (1) our implementation of a supervised semantic interpretation model, Semantic Scattering2 [2], (2) the SNoW machine learning architecture [23], and (3) a competitive SemEval type-B system [1].

Semantic Scattering is a supervised model which uses only semantic information about the two nouns. It consists of a set of iterative procedures of specializations of the training examples on the WordNet IS-A hierarchy. Thus, after a set of necessary specialization iterations the method produces specialized examples from which the model learns a discrimination function.

We implemented the model and improved it. In our implementation we obtain similar performance, but with a much smaller number of training examples[2].

SNoW is a learning architecture that learns a sparse network of linear functions and can deal very well with a large number of features. SNoW has been used successfully in a number of NLP tasks. The features that we used include word-level and part-of-speech information of context words, as well as grammatical categories (subject, object) of the two nouns. All features were implemented as boolean features.

Our SemEval system is a type-B system which participated competitively in the evaluations of SemEval - Task 4 [1]. It makes use of the WordNet IS-A hierarchy to get semantic information about the two nouns, but it also employs various shallow contextual features.

The classification task is defined as a multi-class classification problem on different classification sets.

Contingency relations	Type of overlap	Constraints
{Part-Whole_hp, Content-Container, Measure}	P-Whp $\models$ M	1) Whole must exist before the parts; the part has to be homeomerous with other parts of the whole; it entails the idea of separation of the part from the whole.
	P-Whp $\models$ M $\models_p$ C-C	2) the semantic head noun is a container (this condition is in addition to those at 1) above)
	M $\models_p$ C-C	3) the head noun is a container (this condition is in addition to all of the above), but it is not P-W (the existence of the whole is not a condition for the existence of the parts)
	M	4) the head noun is not a container and the existence of the whole is not presupposed
	P-Wnp	5) when the parts are not homeomerous
{Component-Integral, Purpose, Origin-Entity}	$(C - I \cap PRP \cap O - E = \emptyset)$	6) mutually exclusive; C-I and O-E: encoded by specific instances; parts have a function in regard to the whole and can be (potentially) separated; PRP: encoded by generic instances
{Content-Container, Stuff-Object}	$(C - C \cap S - O = \emptyset)$	7) mutually exclusive; the whole and can be separated for C-C and it cannot for S-O.

Table 4: *Sets of contingency relations along with types of overlap and constraints.*

5 Experimental results

Using the three classifiers described in the previous section, we performed two sets of experiments on the annotated corpus. In the experiment set I each classifier was trained and tested with a 10-fold cross validation one-vs-all approach for each relation (as positive examples those annotated with the relation and as negative the remaining examples in the corpus). Table 5 shows that the best overall results are obtained by the SemEval system, while the worst results are obtained by SemScat2. These results can be partially explained by the fact that the SemEval and SNoW systems relied on contextual information, while SemScat relied only the WordNet information of the two nouns.

System	P [%]	R [%]	F [%]
SemScat	60	52	55
SNoW	64	63	63
SemEval	72	60	65

Table 5: *The overall performance of the three systems using a 10-fold cross validation one-vs-all approach.*

In the second round of experiments we trained and tested the classifiers using a 10-fold cross validation, one-vs-contingency_set approach. Each classifier was trained per relation as in the previous experiments, but this time as negative examples we considered only those belonging to the corresponding contingency set. Thus, we split the 1,000 example corpus into three datasets corresponding to the following contingency sets: {Component-Integral, Origin-Entity, Purpose}, {Content-Container, Measure/Content-Container, P-Whp/Measure/Content-Container}, and {P-Whp/Measure, Measure, P-Whp}.

Tables 6, 7 and 8 show the results which vary per system and differ from those presented in Table 5. Particular attention should be given to SemScat which obtained the lowest results overall. It differentiates poorly between PW_{hp}/M/C-C and M/C-C, since most of the noun - noun pair examples had the same e1-e2 order (i.e., the

container followed by the content as in *a cup of soup* – Measure/Content-Container vs. *a cup of that soup* – Measure/Portion-Mass/Content-Container). SemScat performed much better on the last contingency set, in particular to identify PW_{hp}/M. This is explained by the fact that many of these examples are of the type *lots/bunch/couple of cats/flowers*. These are called *vague measure partitives* since they refer to both the amount (Measure) and the parts of the whole (Member-Collection).

This shows one more time that for relations which are very difficult to differentiate, the ontological information about the two nouns is not very helpful.

Better results are obtained by SNoW and the SNoW and Semeval systems due to their contextual features. In particular, these systems classified well the near miss examples in Table 6 due to various lexical and syntactic features such as verbs and the prepositions "from" and "for". The SemEval system however did not perform well for PW_{hp}/M/C-C and M/C-C since it disregards stop words, including determiners and definite articles which are very important here (in many PW_{hp}/M/C-C examples, the whole is preceeded by a definite article/determiner; e.g.: *a cup of that soup*).

We also performed a quick error analysis. In particular we looked at some of the examples which were misclassified by the Semeval system. Many of the examples required a combination of world knowledge about other words in context as well as pragmatic information. Instances (11) and (12) below indicate such cases. The systems labeled the instance as P-Whp/Measure/Content-Container due to lexical cues such as the verb *pour* and the determiners *that* and *this*. However, the correct interpretation is Measure/Content-Container since *tea* and *wine* here refer to a kind of tea, respectively wine (generic noun) and not to a particular one (specific noun). Example (12) is actually more problematic since it involves pragmatic knowledge.

(11) "I'd also pour you a $\langle e1\rangle$cup$\langle /e1\rangle$ of that apricot $\langle e2\rangle$tea$\langle /e2\rangle$ you like so you could sit and visit with me next week."

(12) The waiter stood politely near the table while Mary decided to order: "I'd like to try a $\langle e1 \rangle$glass$\langle /e1 \rangle$ of this $\langle e2 \rangle$wine$\langle /e2 \rangle$", said she pointing at the menu.

No. rel	Relation	SemScat	SNoW	SemEval system
1	C-I	61.2	87	86.3
4	O-E	57.3	77	78.0
6	PRP	59.1	87	88.7

Table 6: *The performance of the SemEval system on the semantic classification categories representing near-misses: Component-Integral (C-I), Origin-Entity (O-E), and Purpose (PRP).*

No. rel	Relation	SemScat	SNoW	SemEval system
7	C-C	63.7	84	85.3
9	PW_{hp}/M/C-C	54.4	83	75.6
10	M/C-C	56.8	72	71.1

Table 7: *The performance of the SemEval system the semantic classification categories representing overlaps: Content-Container (C-C), Part-Whole/Measure/Content-Container (PW_{hp}/M/C-C), and Measure/Content-Container (M/C-C).*

No. rel	Relation	SemScat	SNoW	SemEval system
2	PW_{hp}	64.0	50	74.3
5	M	66.2	84	78.7
8	PW_{hp}/M	70.0	80	85.1

Table 8: *The performance of the SemEval system the semantic classification categories representing overlaps: Part-Whole (PW_{hp}), Measure (M), Part-Whole/Measure (PW_{hp}/M).*

6 Discussion and Conclusions

This paper addresses the problem of semantic relation identification for a set of relations difficult to differentiate: *near-misses* and *overlaps*. Based on empirical observations on a fairly large dataset of such examples we provided an analysis and a taxonomy of such cases. Using this taxonomy we created various contingency sets of relations. These semantic categories were identified by training and testing three state-of-the-art semantic classifiers employing various feature sets. The results showed that relation identification systems need to rely on both the information provided by the linguistic context and the context of use (pragmatics).

The taxonomy of near-miss and overlapping relations presented here is by no means exhaustive and we intend to extend it in future research. Moreover, we would like to explore ways to learn the contingency sets automatically.

References

[1] B. Beamer, S. Bhat, B. Chee, A. Fister, A. Rozovskaya, and R. Girju. UIUC: A Knowledge-rich Approach to Identifying Semantic Relations between Nominals. In *SemEval-2007*, 2007.

[2] B. Beamer, A. Rozovskaya, and R. Girju. Automatic Semantic Relation Extraction with Multiple Boundary Generation. In *The National Conference on Artificial Intelligence (AAAI)*, 2008.

[3] R. Girju. Automatic Detection of Causal Relations for Question Answering. In *Association for Computational Linguistics (ACL 2003), Workshop on "Multilingual Summarization and Question Answering - Machine Learning and Beyond"*. 2003.

[4] R. Girju. Improving the interpretation of noun phrases with cross-linguistic information. In *Proceedings of ACL*, 2007.

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

[6] R. Girju, A. Giuglea, M. Olteanu, O. Fortu, O. Bolohan, and D. Moldovan. Support vector machines applied to the classification of semantic relations in nominalized noun phrases. In *In Proceedings of the HLT/NAACL Workshop on Computational Lexical Semantics.*, Boston, MA., 2004.

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

[8] R. Girju, D. Moldovan, M. Tatu, and D. Antohe. On the Semantics of Noun Compounds. *Computer Speech and Language - Special Issue on Multiword Expressions*, 2005.

[9] R. Girju, P. Nakov, V. Nastase, S. Szpakowicz, P. Turney, and D. Yuret. Semeval-2007 task 04: Classification of semantic relations between nominals. In *SemEval-2007*, 2007.

[10] M. Hearst. Acquisition of Hyponyms from Large Text Corpora. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING-92)*, pages 539–545, Nantes, France, 1992.

[11] C. S. G. Khoo, S. H. Myaeng, and R. N. Oddy. Using cause-effect relations in text to improve information retrieval precision. *Information Processing & Management*, 37(1):119–145, 2001.

[12] S. N. Kim and T. Baldwin. Interpreting semantic relations in noun compounds via verb semant ics. In *International Computational Linguistics Coference (COLING)*, 2006.

[13] M. Lapata. The disambiguation of nominalisations. *Computational Linguistics*, 28(3):357–388, 2002.

[14] M. Lauer. Corpus statistics meet the noun compound: Some empirical results. In *In the Proceedings of the Association for Computational Linguistics*, pages 47–54, Cambridge, Mass, 1995.

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

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

[17] D. Moldovan, C. Clark, S. Harabagiu, and D. Hodges. A semantically and contextually enriched logic prover for question answering. *Journal of Applied Logic*, 5:49–69, 2005.

[18] D. Moldovan and R. Girju. Knowledge discovery from text. In *The Tutorial Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, 2003.

[19] P. Nakov and M. Hearst. Using verbs to characterise noun-noun relations. In *Proceedings of the 12th International Conference on Artificial In telligence: Methodology, Systems and Applications*, 2006.

[20] V. Nastase and S. Szpakowicz. Augmenting wordnet's structure using ldoce. In *CICLing-2003*, 2003.

[21] M. Pennacchiotti and P. Pantel. Ontologizing semantic relations. In *COLING/ACL-06*, 2006.

[22] B. Rosario and M. Hearst. Classifying the semantic relations in noun compounds. In *Proceedings of the 2001 Conference on Empirical Methods in Natural Language Processing*, 2001.

[23] D. Roth. The SNoW learning architecture. Technical Report UIUCDCS-R-99-2101, UIUC Computer Science Department, May 1999.

[24] M. Tatu and D. I. Moldovan. A Logic-Based Semantic Approach to Recognizing Textual Entailment. In *Proceedings of the Association for Computational Linguistics - Poster session*, 2006.

[25] M. Winston, R. Chaffin, and D. Hermann. A taxonomy of part-whole relations. *Cognitive science*, 11(4):417–444, 1987.

[26] K. Zimmer. Some general observations about nominal compounds. *Stanford Working Papers on Linguistic Universals*, 5:C1-C21, 1971.