Ontologizing Semantic Relations

Marco Pennacchiotti
ART Group - DISP
University of Rome "Tor Vergata"
Viale del Politecnico 1
Rome, Italy
pennacchiotti@info.uniroma2.it

Patrick Pantel
Information Sciences Institute
University of Southern California
4676 Admiralty Way
Marina del Rey, CA90292
pantel@isi.edu

Abstract

Many algorithms have been developed to harvest lexical semantic resources, however few have linked the mined knowledge into formal knowledge repositories. In this paper, we propose two algorithms for automatically ontologizing (attaching) semantic relations into WordNet. We present an empirical evaluation on the task of attaching *part-of* and *causation* relations, showing an improvement on *F*-score over a baseline model.

1 Introduction

NLP researchers have developed many algorithms for mining knowledge from text and the Web, including facts (Etzioni et al. 2005), semantic lexicons (Riloff and Shepherd 1997), concept lists (Lin and Pantel 2002), and word similarity lists (Hindle 1990). Many recent efforts have also focused on extracting binary semantic relations between entities, such as entailments (Szpektor et al. 2004), *is-a* (Ravichandran and Hovy 2002), *part-of* (Girju et al. 2003), and other relations.

The output of most of these systems is flat lists of lexical semantic knowledge such as "*Italy is-a country*" and "*orange similar-to blue*". However, using this knowledge beyond simple keyword matching, for example in inferences, requires it to be linked into formal semantic repositories such as ontologies or term banks like WordNet (Fellbaum 1998).

Pantel (2005) defined the task of *ontologizing* a lexical semantic resource as linking its terms to the concepts in a WordNet-like hierarchy. For example, "*orange similar-to blue*" ontologizes in WordNet to "*orange#2 similar-to blue#1*" and "*orange#2 similar-to blue#2*". In his framework, Pantel proposed a method of inducing ontological co-occurrence vectors [1] which are subsequently used to ontologize unknown terms into WordNet with 74% accuracy.

In this paper, we take the next step and explore two algorithms for ontologizing binary semantic relations into WordNet and we present empirical results on the task of attaching *part-of* and *causation* relations. Formally, given an instance (*x*, *r*, *y*) of a binary relation *r* between terms *x* and *y*, the ontologizing task is to identify the WordNet senses of *x* and *y* where *r* holds. For example, the instance (*proton, PART-OF, element*) ontologizes into WordNet as (*proton#1, PART-OF, element#2*).

The first algorithm that we explore, called the *anchoring approach*, was suggested as a promising avenue of future work in (Pantel 2005). This bottom up algorithm is based on the intuition that *x* can be disambiguated by retrieving the set of terms that occur in the same relation *r* with *y* and then finding the senses of *x* that are most similar to this set. The assumption is that terms occurring in the same relation will tend to have similar meaning. In this paper, we propose a measure of similarity to capture this intuition.

In contrast to anchoring, our second algorithm, called the *clustering approach*, takes a top-down view. Given a relation *r*, suppose that we are given every *conceptual instance* of *r*, i.e., instances of *r* in the upper ontology like (*particles#1, PART-OF, substances#1*). An instance (*x*, *r*, *y*) can then be ontologized easily by finding the senses of *x* and *y* that are subsumed by ancestors linked by a conceptual instance of *r*. For example, the instance (*proton, PART-OF, element*) ontologizes to (*proton#1, PART-OF, element#2*) since *proton#1* is subsumed by *particles* and *element#2* is subsumed by *substances*. The problem then is to automatically infer the set of con-

[1] The ontological co-occurrence vector of a concept consists of all lexical co-occurrences with the concept in a corpus.

ceptual instances. In this paper, we develop a clustering algorithm for generalizing a set of relation instances to conceptual instances by looking up the WordNet hypernymy hierarchy for common ancestors, as specific as possible, that subsume as many instances as possible. An instance is then attached to its senses that are subsumed by the highest scoring conceptual instances.

2 Relevant Work

Several researchers have worked on ontologizing semantic resources. Most recently, Pantel (2005) developed a method to propagate lexical co-occurrence vectors to WordNet synsets, forming ontological co-occurrence vectors. Adopting an extension of the distributional hypothesis (Harris 1985), the co-occurrence vectors are used to compute the similarity between synset/synset and between lexical term/synset. An unknown term is then attached to the WordNet synset whose co-occurrence vector is most similar to the term's co-occurrence vector. Though the author suggests a method for attaching more complex lexical structures like binary semantic relations, the paper focused only on attaching terms.

Basili (2000) proposed an unsupervised method to infer semantic classes (WordNet synsets) for terms in domain-specific verb relations. These relations, such as (*x*, *EXPAND*, *y*) are first automatically learnt from a corpus. The semantic classes of *x* and *y* are then inferred using *conceptual density* (Agirre and Rigau 1996), a WordNet-based measure applied to all instantiation of *x* and *y* in the corpus. Semantic classes represent possible common generalizations of the verb arguments. At the end of the process, a set of syntactic-semantic patterns are available for each verb, such as:

(*social_group#1, expand, act#2*)
(*instrumentality#2, expand, act#2*)

The method is successful on specific relations with few instances (such as domain verb relations) while its value on generic and frequent relations, such as *part-of*, was untested.

Girju et al. (2003) presented a highly supervised machine learning algorithm to infer semantic constraints on *part-of* relations, such as (*object#1, PART-OF, social_event#1*). These constraints are then used as selectional restrictions in harvesting *part-of* instances from ambiguous lexical patterns, like "*X of Y*". The approach shows high performance in terms of precision and recall, but, as the authors acknowledge, it requires large human effort during the training phase.

Others have also made significant additions to WordNet. For example, in eXtended WordNet (Harabagiu et al. 1999), the glosses in WordNet are enriched by disambiguating the nouns, verbs, adverbs, and adjectives with synsets. Another work has enriched WordNet synsets with topically related words extracted from the Web (Agirre et al. 2001). Finally, the general task of word sense disambiguation (Gale et al. 1991) is relevant since there the task is to ontologize each term in a passage into a WordNet-like sense inventory. If we had a large collection of sense-tagged text, then our mining algorithms could directly discover WordNet attachment points at harvest time. However, since there is little high precision sense-tagged corpora, methods are required to ontologize semantic resources without fully disambiguating text.

3 Ontologizing Semantic Relations

Given an instance (*x*, *r*, *y*) of a binary relation *r* between terms *x* and *y*, the ontologizing task is to identify the senses of *x* and *y* where *r* holds. In this paper, we focus on WordNet 2.0 senses, though any similar term bank would apply.

Let S_x and S_y be the sets of all WordNet senses of *x* and *y*. A *sense pair*, s_{xy}, is defined as any pair of senses of *x* and *y*: $s_{xy}=\{s_x, s_y\}$ where $s_x \in S_x$ and $s_y \in S_y$. The set of all sense pairs S_{xy} consists of all permutations between senses in S_x and S_y.

In order to attach a relation instance (*x*, *r*, *y*) into WordNet, one must:

- *Disambiguate* *x* and *y*, that is, find the subsets $S'_x \subseteq S_x$ and $S'_y \subseteq S_y$ for which the relation *r* holds; and
- *Instantiate* the relation in WordNet, using the synsets corresponding to all correct permutations between the senses in S'_x and S'_y. We denote this set of *attachment points* as S'_{xy}.

If S_x or S_y is empty, no attachments are produced.

For example, the instance (*study, PART-OF, report*) is ontologized into WordNet through the senses $S'_x=\{survey\#1, study\#2\}$ and $S'_y=\{report\#1\}$. The final attachment points S'_{xy} are:

(*survey#1, PART-OF, report#1*)
(*study#1, PART-OF, report#1*)

Unlike common algorithms for word sense disambiguation, here it is important to take into consideration the semantic dependency between the two terms *x* and *y*. For example, an entity that is *part-of* a study has to be some kind of informa-

tion. This knowledge about mutual selectional preference (the preferred semantic class that fills a certain relation role, as *x* or *y*) can be exploited to ontologize the instance.

In the following sections, we propose two algorithms for ontologizing binary semantic relations.

3.1 Method 1: *Anchor Approach*

Given an instance (*x*, *r*, *y*), this approach fixes the term *y*, called the *anchor*, and then disambiguates *x* by looking at all other terms that occur in the relation *r* with *y*. Based on the principle of distributional similarity (Harris 1985), the algorithm assumes that the words that occur in the same relation *r* with *y* will be more similar to the correct sense(s) of *x* than the incorrect ones. After disambiguating *x*, the process is then inverted with *x* as the anchor to disambiguate *y*.

In the first step, *y* is fixed and the algorithm retrieves the set of all other terms *X*' that occur in an instance (*x*', *r*, *y*), *x*' ∈ *X*'[2]. For example, given the instance (*reflections, PART-OF, book*), and a resource containing the following relations:

(*false allegations, PART-OF, book*)
(*stories, PART-OF, book*)
(*expert analysis, PART-OF, book*)
(*conclusions, PART-OF, book*)

the resulting set *X*' would be: {*allegations*, *stories*, *analysis*, *conclusions*}.

All possible permutations, $S_{xx'}$, between the senses of *x* and the senses of each term in *X*', called $S_{x'}$, are computed. For each sense pair $\{s_x, s_{x'}\} \in S_{xx'}$, a similarity score $r(s_x, s_{x'})$ is calculated using WordNet:

$$r(s_x, s_{x'}) = \frac{1}{d(s_x, s_{x'}) + 1} \times f(s_{x'})$$

where the distance $d(s_x, s_{x'})$ is the length of the shortest path connecting the two synsets in the hypernymy hierarchy of WordNet, and $f(s_{x'})$ is the number of times sense $s_{x'}$ occurs in any of the instances of *X*'. Note that if no connection between two synsets exists, then $r(s_x, s_{x'}) = 0$.

The overall sense score for each sense s_x of *x* is calculated as:

$$r(s_x) = \sum_{s_{x'} \in S_{x'}} r(s_x, s_{x'})$$

Finally, the algorithm inverts the process by setting *x* as the anchor and computes $r(s_y)$ for each sense of *y*. All possible permutations of senses are computed and scored by averaging $r(s_x)$ and $r(s_y)$. Permutations scoring higher than a threshold τ_1 are selected as the attachment points in WordNet. We experimentally set $\tau_1 = 0.02$.

3.2 Method 2: *Clustering Approach*

The main idea of the clustering approach is to leverage the lexical behaviors of the two terms in an instance as a whole. The assumption is that the general meaning of the relation is derived from the combination of the two terms.

The algorithm is divided in two main phases. In the first phase, *semantic clusters* are built using the WordNet senses of all instances. A semantic cluster is defined by the set of instances that have a common semantic generalization. We denote the *conceptual instance* of the semantic cluster as the pair of WordNet synsets that represents this generalization. For example the following two *part-of* instances:

(*second section, PART-OF, Los Angeles-area news*)
(*Sandag study, PART-OF, report*)

are in a common cluster represented by the following conceptual instance:

[*writing#2, PART-OF, message#2*]

since *writing#2* is a hypernym of both *section* and *study*, and *message#2* is a hypernym of *news* and *report*[3].

In the second phase, the algorithm attaches an instance into WordNet by using WordNet distance metrics and frequency scores to select the best cluster for each instance. A good cluster is one that:

- achieves a good trade-off between generality and specificity; and
- disambiguates among the senses of *x* and *y* using the other instances' senses as support.

For example, given the instance (*second section, PART-OF, Los Angeles-area news*) and the following conceptual instances:

[*writing#2, PART-OF, message#2*]
[*object#1, PART-OF, message#2*]
[*writing#2, PART-OF, communication#2*]
[*social_group#1, PART-OF, broadcast#2*]
[*organization#, PART-OF, message#2*]

the first conceptual instance should be scored highest since it is both not too generic nor too specific and is supported by the instance (*Sandag study, PART-OF, report*), i.e., the conceptual instance subsumes both instances. The second and

[2] For semantic relations between complex terms, like (*expert analysis, PART-OF, book*), only the head noun of terms are recorded, like "*analysis*". As a future work, we plan to use the whole term if it is present in WordNet.

[3] Again, here, we use the syntactic head of each term for generalization since we assume that it drives the meaning of the term itself.

the third conceptual instances should be scored lower since they are too generic, while the last two should be scored lower since the sense for *section* and *news* are not supported by other instances. The system then outputs, for each instance, the set of sense pairs that are subsumed by the highest scoring conceptual instance. In the previous example:

(*section#1, PART-OF, news#1*)
(*section#1, PART-OF, news#2*)
(*section#1, PART-OF, news#3*)

are selected, as they are subsumed by [*writing#2, PART-OF, message#2*]. These sense pairs are then retained as *attachment points* into WordNet.

Below, we describe each phase in more detail.

Phase 1: Cluster Building

Given an instance (x, r, y), all sense pair permutations $s_{xy}=\{s_x, s_y\}$ are retrieved from WordNet. A set of *candidate conceptual instances*, C_{xy}, is formed for each instance from the permutation of each WordNet ancestor of s_x and s_y, following the hypernymy link, up to degree τ_2.

Each candidate conceptual instance, $c=\{c_x, c_y\}$, is scored by its degree of generalization as follows:

$$r(c) = \frac{1}{(n_x+1)\times(n_y+1)}$$

where n_i is the number of hypernymy links needed to go from s_i to c_i, for $i \in \{x, y\}$. $r(c)$ ranges from [0, 1] and is highest when little generalization is needed.

For example, the instance (*Sandag study, PART-OF, report*) produces 70 sense pairs since *study* has 10 senses and *report* has 7 senses. Assuming $\tau_2=1$, the instance sense (*survey#1, PART-OF, report#1*) has the following set of candidate conceptual instances:

C_{xy}	n_x	n_y	r(c)
(survey#1, PART-OF,report#1)	0	0	1
(survey#1, PART-OF,document#1)	0	1	0.5
(examination#1, PART-OF,report#1)	1	0	0.5
(examination#1, PART-OF,document#1)	1	1	0.25

Finally, each candidate conceptual instance c forms a cluster of all instances (x, r, y) that have some sense pair s_x and s_y as hyponyms of c. Note also that candidate conceptual instances may be subsumed by other candidate conceptual instances. Let G_c refer to the set of all candidate conceptual instances subsumed by candidate conceptual instance c.

Intuitively, better candidate conceptual instances are those that subsume both many instances and other candidate conceptual instances, but at the same time that have the least distance from subsumed instances. We capture this intuition with the following score of c:

$$score(c) = \frac{\sum_{g\in G_c} r(g)}{|G_c|} \times \log|I_c| \times \log|G_c|$$

where I_c is the set of instances subsumed by c. We experimented with different variations of this score and found that it is important to put more weight on the distance between subsumed conceptual instances than the actual number of subsumed instances. Without the *log* terms, the highest scoring conceptual instances are too generic (i.e., they are too high up in the ontology).

Phase 2: Attachment Points Selection

In this phase, we utilize the conceptual instances of the previous phase to attach each instance (x, r, y) into WordNet.

At the end of Phase 1, an instance can be clustered in different conceptual instances. In order to select an attachment, the algorithm selects the sense pair of x and y that is subsumed by the highest scoring candidate conceptual instance. It and all other sense pairs that are subsumed by this conceptual instance are then retained as the final attachment points.

As a side effect, a final set of conceptual instances is obtained by deleting from each candidate those instances that are subsumed by a higher scoring conceptual instance. Remaining conceptual instances are then re-scored using $score(c)$. The final set of conceptual instances thus contains unambiguous sense pairs.

4 Experimental Results

In this section we provide an empirical evaluation of our two algorithms.

4.1 Experimental Setup

Researchers have developed many algorithms for harvesting semantic relations from corpora and the Web. For the purposes of this paper, we may choose any one of them and manually validate its mined relations. We choose *Espresso*[4], a general-purpose, broad, and accurate corpus harvesting algorithm requiring minimal supervision. Adopt-

[4] Reference suppressed – the paper introducing *Espresso* has also been submitted to COLING/ACL 2006.

SYSTEM	PRECISION	RECALL	F-SCORE
BL	54.0%	31.3%	39.6%
AN	40.7%	47.3%	43.8%
CL	**57.4%**	**49.6%**	**53.2%**

Table 1. System precision, recall and *F*-score on the *part-of* relation.

SYSTEM	PRECISION	RECALL	F-SCORE
BL	**45.0%**	25.0%	32.1%
AN	41.7%	32.4%	**36.5%**
CL	40.0%	**32.6%**	35.9%

Table 2. System precision, recall and *F*-score on the *causation* relation.

ing a bootstrapping approach, *Espresso* takes as input a few seed instances of a particular relation and iteratively learns surface patterns to extract more instances.

Test Sets

We experiment with two relations: *part-of* and *causation*. The *causation* relation occurs when an entity produces an effect or is responsible for events or results, for example (*virus, CAUSE, influenza*) and (*burning fuel, CAUSE, pollution*). We manually built five seed relation instances for both relations and apply *Espresso* to a dataset consisting of a sample of articles from the Aquaint (TREC-9) newswire text collection. The sample consists of 55.7 million words extracted from the Los Angeles Times data files. *Espresso* extracted 1,468 *part-of* instances and 1,129 *causation* instances. We manually validated the output and randomly selected 200 correct relation instances of each relation for ontologizing into WordNet 2.0.

Gold Standard

We manually built a gold standard of all correct attachments of the test sets in WordNet. For each relation instance (*x*, *r*, *y*), two human annotators selected from all sense permutations of *x* and *y* the correct attachment points in WordNet. For example, for (*synthetic material, PART-OF, filter*), the judges selected the following attachment points: (*synthetic material#1, PART-OF, filter#1*) and (*synthetic material#1, PART-OF, filter#2*). The kappa statistic (Siegel and Castellan Jr. 1988) on the two relations together was K = 0.73.

Systems

The following three systems are evaluated:

- **BL**: the baseline system that attaches each relation instance to the first (most common) WordNet sense of both terms;
- **AN**: the *anchor approach* described in Section 3.1.
- **CL**: the *clustering approach* described in Section 3.2.

4.2 Precision, Recall and *F*-score

For both the *part-of* and *causation* relations, we apply the three systems described above and compare their attachment performance using precision, recall, and *F*-score. Using the manually built gold standard, the precision of a system on a given relation instance is measured as the percentage of correct attachments and recall is measured as the percentage of correct attachments retrieved by the system. Overall system precision and recall are then computed by averaging the precision and recall of each relation instance.

Table 1 and Table 2 report the results on the *part-of* and *causation* relations. We experimentally set the CL generalization parameter τ_2 to 5 and the τ_1 parameter for AN to 0.02.

4.3 Discussion

For both relations, CL and AN outperform the baseline in overall *F*-score. For *part-of*, Table 1 shows that CL outperforms BL by 13.6% in *F*-score and AN by 9.4%. For *causation*, Table 2 shows that AN outperforms BL by 4.4% on *F*-score and CL by 0.6%.

The good results of the CL method on the *part-of* relation suggest that instances of this relation are particularly amenable to be clustered. The generality of the *part-of* relation in fact allows the creation of fairly natural clusters, corresponding to different sub-types of *part-of*, as those proposed in (Winston 1983). The *causation* relation, however, being more difficult to define at a semantic level (Girju 2003), is less easy to cluster and thus to disambiguate.

Both CL and AN have better recall than BL, but precision results vary with CL beating BL only on the *part-of* relation. Overall, the system performances suggest that ontologizing semantic relations into WordNet is in general not easy.

The better results of CL and AN with respect to BL suggest that the use of comparative semantic analysis among corpus instances is a good way to carry out disambiguation. Yet, the BL

method shows surprisingly good results. This indicates that also a simple method based on word sense usage in language can be valuable. An interesting avenue of future work is to better combine these two different views in a single system.

The low recall results for CL are mostly attributed to the fact that in Phase 2 only the best scoring cluster is retained for each instance. This means that instances with multiple senses that do not have a common generalization are not captured. For example the *part-of* instance (*wings, PART-OF, chicken*) should cluster both in [*body_part#1, PART-OF, animal#1*] and [*body_part#1, PART-OF, food#2*], but only the best scoring one is retained.

5 Conceptual Instances: Other Uses

Our *clustering approach* from Section 3.2 is enabled by learning conceptual instances – relations between mid-level ontological concepts. Beyond the ontologizing task, conceptual instances may be useful for several other tasks. In this section, we discuss some of these opportunities and present small qualitative evaluations.

Conceptual instances represent common semantic generalizations of a particular relation. For example, below are two possible conceptual instances for the *part-of* relation:

[*person#1, PART-OF, organization#1*]
[*act#1, PART-OF, plan#1*]

The first conceptual instance in the example subsumes all the *part-of* instances in which one or more persons are part of an organization, such as:

(*president Brown, PART-OF, executive council*)
(*representatives, PART-OF, organization*)
(*students, PART-OF, orchestra*)
(*players, PART-OF, Metro League*)

Below, we present three possible ways of exploiting these conceptual instances.

Support to Relation Extraction Tools

Conceptual instances may be used to support relation extraction algorithms such as *Espresso*.

Most minimally supervised harvesting algorithm do not exploit *generic patterns*, i.e. those patterns with high recall but low precision, since they cannot separate correct and incorrect relation instances. For example, the pattern "*X of Y*" extracts many correct relation instances like "*wheel of the car*" but also many incorrect ones like "*house of representatives*".

Girju et al. (2003) described a highly supervised algorithm for learning semantic constraints on *generic patterns*, leading to a very significant increase in system recall without deteriorating precision. Conceptual instances can be used to automatically learn such semantic constraints by acting as a filter for generic patterns, retaining only those instances that are subsumed by high scoring conceptual instances. Effectively, conceptual instances are used as *selectional restrictions* for the relation. For example, our system discards the following incorrect instances:

(*week, CAUSE, coalition*)
(*demeanor, CAUSE, vacuum*)

as they are both part of the very low scoring conceptual instance [*abstraction#6, CAUSE, state#1*].

Ontology Learning from Text

Each conceptual instance can be viewed as a formal specification of the relation at hand. For example, Winston (1983) manually identified six sub-types of the *part-of* relation: *member-collection, component-integral object, portion-mass, stuff-object, feature-activity* and *place-area*. Such classifications are useful in applications and tasks where a semantically rich organization of knowledge is required. Conceptual instances can be viewed as an automatic derivation of such a classification based on corpus usage. Moreover, conceptual instances can be used to improve the ontology learning process itself. For example, our *clustering approach* can be seen as an *inductive* step producing conceptual instances that are then used in a *deductive* step to learn new instances. An algorithm could iterate between the induction/deduction cycle until no new relation instances and conceptual instances can be inferred.

Word Sense Disambiguation

Word Sense Disambiguation (WSD) systems can exploit the selectional restrictions identified by conceptual instances to disambiguate ambiguous terms occurring in particular contexts. For example, given the sentence:

"the board is composed by members of different countries"

and a harvesting algorithm that extracts the *part-of* relation (*members, PART-OF, board*), the system could infer the correct senses for *board* and *members* by looking at their closest conceptual instance. In our system, we would infer the attachment (*member#1, PART-OF, board#1*) since it is part of the highest scoring conceptual instance [*person#1, PART-OF, organization#1*].

CONCEPTUAL INSTANCE	SCORE	# INSTANCES	INSTANCES
[*multitude#3, PART-OF, group#1*]	2.04	10	(*ordinary people, PART-OF, Democratic Revolutionary Party*) (*unlicensed people, PART-OF, underground economy*) (*young people, PART-OF, commission*) (*air mass, PART-OF, cold front*)
[*person#1, PART-OF, organization#1*]	1.71	43	(*foreign ministers, PART-OF, council*) (*students, PART-OF, orchestra*) (*socialists, PART-OF, Iraqi National Joint Action Committee*) (*players, PART-OF, Metro League*)
[*act#2, PART-OF, plan#1*]	1.60	16	(*major concessions, PART-OF, new plan*) (*attacks, PART-OF, coordinated terrorist plan*) (*visit, PART-OF, exchange program*) (*survey, PART-OF, project*)
[*communication#2, PART-OF, book#1*]	1.14	10	(*hints, PART-OF, booklet*) (*soup recipes, PART-OF, book*) (*information, PART-OF, instruction manual*) (*extensive expert analysis, PART-OF, book*)
[*compound#2, PART-OF, waste#1*]	0.57	3	(*salts, PART-OF, powdery white waste*) (*lime, PART-OF, powdery white waste*) (*resin, PART-OF, waste*)

Table 3. Sample of the highest scoring conceptual instances learned for the *part-of* relation. For each conceptual instance, we report the score(*c*), the number of instances, and some example instances.

5.1 Qualitative Evaluation

Table 3 and Table 4 list samples of the highest ranking conceptual instances obtained by our system for the *part-of* and *causation* relations. Below we provide a small evaluation to verify:

- the *correctness* of the conceptual instances. Incorrect conceptual instances such as [*attribute#2, CAUSE, state#4*], discovered by our system, can impede WSD and extraction tools where precise selectional restrictions are needed; and
- the *accuracy* of the conceptual instances. Sometimes, an instance is incorrectly attached to a correct conceptual instance. For example, the instance (*air mass, PART-OF, cold front*) is incorrectly clustered in [*group#1, PART-OF, multitude#3*] since *mass* and *front* both have a sense that is descendant of *group#1* and *multitude#3*. However, these are not the correct senses of *mass* and *front* for which the *part-of* relation holds.

For evaluating *correctness,* we manually verify how many correct conceptual instances are produced by Phase 2 of the clustering approach described in Section 3.2. The claim is that a *correct* conceptual instance is one for which the relation holds for all possible subsumed senses. For example, the conceptual instance [*group#1, PART-OF, multitude#3*] is correct, as the relation holds for every semantic subsumption of the two senses. An example of an incorrect conceptual instance is [*state#4, CAUSE, abstraction#6*] since it subsumes the incorrect instance (*audience*, *CAUSE*, *new context*). A manual evaluation of the highest scoring 200 conceptual instances, generated on our test sets described in Section 4.1, showed 82% correctness for the *part-of* relation and 86% for *causation*.

For estimating the overall clustering *accuracy*, we evaluated the number of correctly clustered instances in each conceptual instance. For example, the instance (*business people, PART-OF, committee*) is correctly clustered in [*multitude#3, PART-OF, group#1*] and the instance (*law, PART-OF, constitutional pitfalls*) is incorrectly clustered in [*group#1, PART-OF, artifact#1*]. We estimated the overall accuracy by manually judging the instances attached to 10 randomly sampled conceptual instances. The accuracy for *part-of* is 84% and for *causation* it is 76.6%.

6 Conclusions

In this paper, we proposed two algorithms for automatically ontologizing binary semantic relations into WordNet: an *anchoring approach* and a *clustering approach*. Experiments on the *part-of* and *causation* relations showed promising results. Both algorithms outperformed the baseline on *F*-score. Our best results were on the *part-of* relation where the *clustering approach* achieved 13.6% higher *F*-score than the baseline.

The induction of conceptual instances has opened the way for many avenues of future work. We intend to pursue the ideas presented in Section 5 for using conceptual instances to: i) support knowledge acquisition tools by learning semantic constraints on extracting patterns; ii) support ontology learning from text; and iii) improve word sense disambiguation through selectional restrictions. Also, we will try different similarity score functions for both the clustering and the anchor approaches, as those surveyed in Corley and Mihalcea (2005).

CONCEPTUAL INSTANCE	*SCORE*	*# INSTANCES*	*INSTANCES*
[*change#3, CAUSE, state#4*]	1.49	17	(*separation, CAUSE, anxiety*) (*demotion, CAUSE, roster vacancy*) (*budget cuts, CAUSE, enrollment declines*) (*reduced flow, CAUSE, vacuum*)
[*act#2, CAUSE, state#3*]	0.81	20	(*oil drilling, CAUSE, air pollution*) (*workplace exposure, CAUSE, genetic injury*) (*industrial emissions, CAUSE, air pollution*) (*long recovery, CAUSE, great stress*)
[*person#1, CAUSE, act#2*]	0.64	12	(*homeowners, CAUSE, water waste*) (*needlelike puncture, CAUSE, physician*) (*group member, CAUSE, controversy*) (*children, CAUSE, property damage*)
[*organism#1, CAUSE, disease#1*]	0.03	4	(*parasites, CAUSE, pneumonia*) (*virus, CAUSE, influenza*) (*chemical agents, CAUSE, pneumonia*) (*genetic mutation, CAUSE, Dwarfism*)

Table 4. Sample of the highest scoring conceptual instances learned for the *causation* relation. For each conceptual instance, we report score(c), the number of instances, and some example instances.

The algorithms described in this paper may be applied to ontologize many lexical resources of semantic relations, no matter the harvesting algorithm used to mine them. In doing so, we have the potential to quickly enrich our ontologies, like WordNet, thus reducing the knowledge acquisition bottleneck. It is our hope that we will be able to leverage these enriched resources, albeit with some noisy additions, to improve performance on knowledge-rich problems such as question answering and textual entailment.

References

Agirre, E. and Rigau, G. 1996. Word sense disambiguation using conceptual density. In *Proceedings of COLING-96*. pp. 16-22. Copenhagen, Danmark.

Agirre, E.; Ansa, O.; Martinez, D.; and Hovy, E. 2001. Enriching WordNet concepts with topic signatures. In *Proceedings of NAACL Workshop on WordNet and Other Lexical Resources: Applications, Extensions and Customizations.* Pittsburgh, PA.

Basili, R.; Pazienza, M.T.; and Vindigni, M. 2000. Corpus-driven learning of event recognition rules. In *Proceedings of Workshop on Machine Learning and Information Extraction (ECAI-00)*.

Corley, C. and Mihalcea, R. 2005. Measuring the Semantic Similarity of Texts. In *Proceedings of the ACL Workshop on Empirical Modelling of Semantic Equivalence and Entailment*. Ann Arbor, MI.

Etzioni, O.; Cafarella, M.J.; Downey, D.; Popescu, A.-M.; Shaked, T.; Soderland, S.; Weld, D.S.; and Yates, A. 2005. Unsupervised named-entity extraction from the Web: An experimental study. *Artificial Intelligence*, 165(1): 91-134.

Fellbaum, C. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Gale, W.; Church, K.; and Yarowsky, D. 1992. A method for disambiguating word senses in a large corpus. *Computers and Humanities*, 26:415-439.

Girju, R.; Badulescu, A.; and Moldovan, D. 2003. Learning semantic constraints for the automatic discovery of part-whole relations. In *Proceedings of HLT/NAACL-03*. pp. 80-87. Edmonton, Canada.

Girju, R. 2003. Automatic Detection of Causal Relations for Question Answering. In *Proceedings of ACL Workshop on Multilingual Summarization and Question Answering*. Sapporo, Japan.

Harabagiu, S.; Miller, G.; and Moldovan, D. 1999. WordNet 2 - A Morphologically and Semantically Enhanced Resource. In *Proceedings of SIGLEX-99*. pp.1-8. University of Maryland.

Harris, Z. 1985. Distributional structure. In: Katz, J. J. (ed.) *The Philosophy of Linguistics. New York*: Oxford University Press. pp. 26–47.

Hindle, D. 1990. Noun classification from predicate-argument structures. In *Proceedings of ACL-90*. pp. 268–275. Pittsburgh, PA.

Lin, D. and Pantel, P. 2002. Concept discovery from text. In *Proceedings of COLING-02*. pp. 577-583. Taipei, Taiwan.

Pantel, P. 2005. Inducing Ontological Co-occurrence Vectors. In *Proceedings of ACL-05*. pp. 125-132. Ann Arbor, MI.

Ravichandran, D. and Hovy, E.H. 2002. Learning surface text patterns for a question answering system. In *Proceedings of ACL-2002*. pp. 41-47. Philadelphia, PA.

Riloff, E. and Shepherd, J. 1997. A corpus-based approach for building semantic lexicons. In *Proceedings of EMNLP-97*.

Siegel, S. and Castellan Jr., N. J. 1988. Nonparametric Statistics for the Behavioral Sciences. McGraw-Hill.

Szpektor, I.; Tanev, H.; Dagan, I.; and Coppola, B. 2004. Scaling web-based acquisition of entailment relations. In *Proceedings of EMNLP-04*. Barcelona, Spain.

Winston, M.; Chaffin, R.; and Hermann, D. 1987. A taxonomy of part-whole relations. *Cognitive Science*, 11:417–444.