Combining Collocations, Lexical and Encyclopedic Knowledge for Metonymy Resolution

Vivi Nastase and **Michael Strube**
EML Research gGmbH
Heidelberg, Germany
http://www.eml-research.de/nlp

Abstract

This paper presents a supervised method for resolving metonymies. We enhance a commonly used feature set with features extracted based on collocation information from corpora, generalized using lexical and encyclopedic knowledge to determine the preferred sense of the potentially metonymic word using methods from unsupervised word sense disambiguation. The methodology developed addresses one issue related to metonymy resolution – the influence of local context. The method developed is applied to the metonymy resolution task from SemEval 2007. The results obtained, higher for the countries subtask, on a par for the companies subtask – compared to participating systems – confirm that lexical, encyclopedic and collocation information can be successfully combined for metonymy resolution.

1 Introduction

Metonymies are a pervasive phenomenon in language. They occur because in communicating, we use words as pointers to a larger body of knowledge, that encompasses various facets of the concept evoked by a given word.

A listener need not understand the cello to be moved by its playing, just as it is unnecessary for a rider to understand technical jargon; all that matters is sensation, and here the **Kawasaki** *excels. The* **cockpit** *is sensibly designed, with a narrow* **front seat** *portion ...*

Kawasaki is a company, it has an organization, facilities, employees, it makes specific products. In the context above, the company name stands in for its products – motorcycles. Motorcycles have parts, the *cockpit* and *front seat* are some of them, and this provides the discourse links between the two sentences. Constraints on the interpretation of a word w in context comes both from the local and global context, and are applied to the information/knowledge evoked by w. The local constraints come from the words with which w is (grammatically) related to. The global constraints come from the domain/topic of the text, discourse relations that span across sentences.

Metonymic words have a rather small number of possible interpretations (also called readings) which occur frequently (Markert and Nissim, 2002). Idiosyncratic interpretations are also possible, but very rare. One can view the possible interpretations of a potentially metonymic word (PMW) as corresponding to the word's possible senses (Nissim and Markert, 2003), bringing the task close to word sense disambiguation.

The approach to metonymy resolution presented here is supervised, with unsupervised feature enrichment. We apply techniques inspired by unsupervised word sense disambiguation, which allow us to go beyond the annotated data provided in training, and quantify the restrictions imposed on the interpretation of a PMW by its grammatically related neighbours through collocation information extracted from corpora. The only annotation required for the corpora are automatically induced part-of-speech tags from which we obtain grammatical relations through regular expression matching over sequences of parts-of-speech. Collocation information is combined with lexical resources – WordNet – and encyclopedic knowledge extracted from Wikipedia to help us generalize the collocations found to determine higher level constraints on a word's grammatical collocates. In the example above, *Kawasaki* is grammatically related to the verb *excel* – it is its subject. To determine the most likely interpretation of *Kawasaki* given that it is in the subject relation with *excel* we look at all the nouns in the corpora

Proceedings of the 2009 Conference on Empirical Methods in Natural Language Processing, pages 910–918,
Singapore, 6-7 August 2009. ©2009 ACL and AFNLP

that appear as this verb's subjects, and estimate from this the preferences *excel* has for its subjects. Let us say the corpus contains the following collocations in subject position (with frequency information in parentheses): *player (4), musician (50), car (30), computer (12), camera (40), driver (55), bike (20) ....* The knowledge resources – WordNet, *isa* relations extracted from Wikipedia – will help generalize these collocations: *player, musician, driver* to `person` and *car, computer, camera, bike* to `artifact`. This together with frequency of occurrence are used to estimate the probability that the verb *excel* takes a `person` or `artifact`-type subject. These are *excel*'s selectional preferences towards certain collocates, and will help determine which possible interpretation for the PMW *Kawasaki* is appropriate in this context – `organization-for-people` or `organization-for-product`.

The paper continues with related work in Section 2 and the description of the data in Section 3. The representation used is introduced in Section 4. The results and the discussion are presented in Section 5. The paper wraps up with conclusions and future work.

2 Related Work

Analysis of metonymies as a linguistic phenomenon dates back at least to the 1930s (Stern, 1931), and are increasingly recognized as an important phenomenon to tackle in the interest of higher level language processing tasks, such as anaphora resolution (Harabagiu, 1998; Markert and Hahn, 2002), question answering (Stallard, 1993) or machine translation (Kamei and Wakao, 1992).

Until the early 90s, the main view about metonymies was that they violate semantic constraints in their immediate context. To resolve metonymies then amounts to detecting violated constraints, usually from those imposed by the verbs on their arguments (Pustejovsky, 1991; Hobbs et al., 1993; Fass, 1991). Markert and Hahn (2002) showed that this approach misses metonymies which do not violate selectional restrictions. In this case referential cohesion relations may indicate that the literal reading is not appropriate and give clues about the intended metonymic interpretation.

Markert and Nissim (2003) have combined observations from the linguistic analysis of metonymies with results of corpus studies. Linguistic research has postulated that (i) conventional metonymic readings are very systematic; (ii) unconventional metonymies can be created on the fly and their interpretation is context dependent; (iii) metonymies are frequent. The fact that most metonymic interpretations are systematic and correspond to a small set of possible readings allow the metonymy resolution to be modelled as a classifier learning task. Markert and Nissim (2002) and Nissim and Markert (2003) have shown that conventional metonymies can be effectively resolved using a supervised machine learning approach. Moreover, grammatically related words are crucial in determining the interpretation of a PMW. The shortcoming is that manually annotated data is in short supply, and the approach suffers from data sparseness. To address this problem, Nissim and Markert (2003) proposed a word similarity-based method. They use Lin's thesaurus (Lin, 1998) to determine how close two lexical heads are, and use this instead of the more restrictive identity constraint when comparing two instances. This technique is complex, requiring smoothing, multiple iterations over the thesaurus and hybrid methods to allow a back-off to grammatical roles.

The supervised approach to resolving metonymies was encouraged by the metonymy resolution task at the semantic evaluation exercise SemEval 2007 (Markert and Nissim, 2007). The participating systems in this task were varied. Most of them (four out of five) have used supervised machine learning techniques. The systems that beat the baseline used either the grammatical annotations provided by the organizers (Farkas et al., 2007; Nicolae et al., 2007), or a robust and deep (not freely available) parser (Brun et al., 2007). These systems represented instances in a manner similar to (Nissim and Markert, 2005). They used additional manually built resources – WordNet, FrameNet, Levin's verb classes, manually built lists of "trigger" words – to generalize the existing features. Brun et al. (2007) also used the British National Corpus (BNC) for computing the distance between words based on their syntactic distribution.

While lexical resources and corpora are used to estimate word similarity, all these systems rely exclusively on the data provided by the organizers – instance representation captures only information that can be derived from or between the data points provided. The approach presented here goes beyond the given data, and induces from corpora measures that allow the system to determine what are the preferences of the words surrounding a PMW towards each of PMW's possible readings. The technique employed is adapted from unsupervised word sense disambiguation (WSD). In short, we use the local grammatical context as it is commonly used in WSD approaches, to guide the system in choosing the reading that fits best. The benefits of using grammatical information for automatic WSD were first explored by Yarowsky (1995) and Resnik (1996) in unsupervised approaches to disambiguating single words in context. The method described here uses automatically induced selectional preferences, computed from sense-untagged data, similar to Nastase (2008).

3 Data

We work with the data from the metonymy resolution task at SemEval 2007 (Markert and Nissim, 2007), generated based on a scheme developed by Markert and Nissim (2003).

The metonymy resolution task at SemEval 2007 consisted of two subtasks – one for resolving country names, the other for companies. For each subtask there is a training and a test portion. Figure 1 shows the text fragment for one sample, and Table 1 the data statistics. The *reading* column shows the possible interpretations of a PMW for countries and companies respectively. For example, `org-for-product` would be the interpretation of the PMW *Kawasaki* in the example shown in the introduction.

Occurrences of country and company names were annotated with a small number of possible readings, as shown in Table 1. This reflects previous analyses of the metonymy phenomenon, which showed that there is a rather small number of possible interpretations that appear more frequently (Markert and Nissim, 2002). Special interpretations are very rarely encountered.

Within the framework of the SemEval task, metonymy resolution is evaluated on the given test data, on three levels of granularity: coarse – distinguish between `literal` and `non-literal` readings; medium – distinguish between `literal`, `mixed` and `non-literal` readings; fine – identify the specific reading of the target word/words (potentially metonymic word - PMW).

4 Representation

The method presented in this paper is a supervised learning method, along the same general lines as the systems which participated in the SemEval competition. As such, it represents each PMW in the data through features that describe its context and some semantic characteristics. The minimum set of necessary features is taken to be that presented by Nissim and Markert (2005), and proved to be effective in solving metonymies. These are the *M&N features* (Markert and Nissim features). We expand on these features and estimate preferences from words in a PMW's context towards specific PMW interpretations. These constitute the *selectional preference features*. Finally, Wikipedia is a source of facts which can be used to derive information that can bias the decision towards certain interpretations for a PMW. Each of these features are described in more detail in the following subsections.

reading	train	test
locations	925	908
literal	737	721
mixed	15	20
othermet	9	11
obj-for-name	0	4
obj-for-representation	0	0
place-for-people	161	141
place-for-event	3	10
place-for-product	0	1
organizations	1090	842
literal	690	520
mixed	59	60
othermet	14	8
obj-for-name	8	6
obj-for-representation	1	0
org-for-members	220	161
org-for-event	2	1
org-for-product	74	67
org-for-facility	15	16
org-for-index	7	3

Table 1: Reading distributions

4.1 M&N features

The features used by Nissim and Markert (2005) are:

- grammatical role of PMW (subj, obj, ...);
- lemmatized head/modifier of PMW (announce, say, ...);
- determiner of PMW (def, indef, bare, demonst, other, ...);

XML tagged text

```
<sample id=”samp114”>
<bnc:title> Computergram international </bnc:title>
<par>
LITTLE FEAR OF MICHELANGELO
The computer industry equivalent of “Small earthquake in Chile” ...
The Michelangelo computer virus that received worldwide attention last year is expected to cause even fewer problems this Saturday than it did when it struck last year, a team of <annot><org reading=”literal”> IBM </org></annot> researchers said.
</par>
</sample>
```

Grammatical annotations

```
SampleID|Lemma|PMW|GrRole|Reading
samp114|researcher|IBM|premod|literal
samp4|be|Williams Holdings|subj|literal
samp5|parent|Fujitsu Ltd|app|mixed
samp5|have|Fujitsu Ltd|subj|mixed
samp5|keep|Fujitsu Ltd|subj|mixed
samp8|against|IBM|pp|literal
```

POS tags

```
<bnc:s id=”samp114-bncCNJ-s341”> ...
<bnc:w id=”samp114-bncCNJ-s343-w29”
bnc:type=”NN0”> team </bnc:w> <bnc:w
id=”samp114-bncCNJ-s343-w30”
bnc:type=”PRF”> of </bnc:w> <annot> <org
reading=”literal”> <bnc:w possmeto=”yes”
id=”samp114-bncCNJ-s343-w31”
bnc:type=”NP0”> IBM </bnc:w> </org>
</annot> <bnc:w
id=”samp114-bncCNJ-s343-w32”
bnc:type=”NN2”> researchers </bnc:w> ...
```

Figure 1: Sample annotation

- grammatical number of PMW (sg, pl);
- number of grammatical roles in which the PMW appears in its current context;
- number of words in PMW;

All these features can be extracted from the grammatically annotated and POS tagged data provided by the organizers.

4.2 Selectional preference features

The grammatical relations and the connected words are important to describe the local context of the target PMW. Because of the limited amount of annotated data (a few thousand instances), lemmas of PMW’s grammatically related words will make for very sparse data that a machine learning system would not be able to generalize over. Nissim and Markert (2003) and the teams participating in the metonymy resolution task have then supplemented their systems with Lin’s thesaurus, WordNet, Beth Levin’s verb groups, FrameNet information, or manually designed lists of words to generalize the grammatically related words and thus find shared characteristics across instances of metonymies in text.

The notion of selectional restrictions used in metonymy resolution – meaning the restrictions imposed on the interpretation of a PMW by its context – is similar to the notion of selectional preferences from word sense disambiguation – meaning the preferences of a word for the senses of the words in its context. We import this notion, and compute selectional preferences for the words in a PMW’s (grammatical) neighbourhood, and allow them to influence the chosen reading for the PMW. Applying methods from unsupervised WSD allow us to estimate such preferences from (sense/metonymy) untagged corpora.

A potentially metonymic word (or phrase) has a small number of possible readings. These can be viewed as possible senses, and the task is to choose the one that fits best in the given context. The preference for each possible sense can be determined based on the PMW’s grammatically related words. To estimate these sense preferences we use grammatical collocations extracted from the British National Corpus (BNC), detected using regular expression matching over sequences of POS using the Word Sketch Engine (Kilgarriff et al., 2004). The scores are computed following a technique similar to Nastase (2008), which is illustrated using the following example:

The Kawasaki drives well, steers brilliantly both under power and in tight corners ...

The PMW *Kawasaki* is involved in the following grammatical relations in the previous sentence:

(drive,subject,Kawasaki)
(steer,subject,Kawasaki)

SampleID	Lemma	PMW	GrRole	Reading	act	animal	artifact	...	person	...
samp190	say	Sun	subj	org-for-members	0.00056	0.01171	0.01958	...	0.61422	...
samp190	claim	Sun	subj	org-for-members	0.00198	0.00099	0.00893	...	0.50211	...

Table 2: Grammatical annotation file enhanced with selectional preference estimates

The BNC provides the collocations *(drive,subject,X)* and *(steer,subject,Y)*, to determine what kind of subject *drive* and *steer* prefer, in “word-POS:frequency” format:

drive	subject	chauffeur-n:12, engine-n:30, car-n:62, taxi-n:13, motorist-n:10, disk-n:15, truck-n:11, man-n:75, ...
steer	subject	power-n:6, car-n:3, sport-n:2, firm-n:2, boy-n:2, government-n:2, man-n:2, people-n:2 ...

The target whose interpretation must be determined is *Kawasaki*. If for a potentially metonymic word representing a company name, there are the following possible interpretations: `company`, `member/person`, `product/artifact`, `facility`, `name`, we compute the preference for each of these interpretations based on the extracted collocations. For the verb *drive* for example, the collocations *engine, car, taxi, truck* are all *artifacts* (according to WordNet), and thus vote for the `product/artifact` reading, while *chauffeur, motorist, man* are all *person*, and vote for the `member/person` reading. Preferences from different grammatical relation for the same PMW are summed.

Formally, we choose the PMWs’ “senses” – a set of words which are close to the possible readings of metonymic words in the data. In this work, these senses are the WordNet 3.0 supersenses:

`S =` $\{$ `act, animal, artifact, attribute, body, cognition, communication, event, feeling, food, group, location, motive, object, person, phenomenon, plant, possession, process, quantity, relation, shape, state, substance, time` $\}$.

Because none of these can be seen as a sense for “company”, the list is supplemented with `company` and `organization`. Granted, there is no 1:1 mapping from these supersenses to PMW readings, but find such a strict correspondence is not necessary because the context preferences for each of these senses are used as features, and the mapping to PMW readings is found through a supervised learned model.

To compute the preference of a word w in the grammatical context of a PMW t (the target) towards each of t’s possible senses, we consider each relation (w, R, t), where R is the grammatical relation. The set C of word collocations are extracted from the BNC

$$C = \{(w, R, w_j : f_j)|(w, R, w_j) \in BNC, \\ f_j \text{is the frequency of occurrence}\}$$

and used to compute a preference score P_{s_i} for each sense $s_i \in \mathsf{S}$:

$$P_{s_i} = \frac{\sum_{(w,R,w_{i,j}:f_{i,j})\in C_{s_i}} f_{i,j}}{\sum_{(w,R,w_j:f_j)\in C} f_j}$$

where

$$C_{s_i} = \{(w, R, w_j : f_j)|(w, R, w_j : f_j) \in C; \\ \text{supersense}(w_j, s_i) \parallel \text{isa}(w_j, s_i)\}.$$

supersense(w_j, s_i) is true if s_i is a supersense of one of w_j’s senses;

isa(w_j, s_i) is true if s_i is a hypernym of one of w_j’s senses in WordNet, or is a fact extracted from Wikipedia.

To determine the *supersense* and *isa* relation we use WordNet 3.0, and a set of 7,578,112 *isa* relations extracted by processing the page and category network of Wikipedia[1] (Nastase and Strube, 2008). The collocations extracted from BNC contain numerous named entities, most of which are not part of WordNet. If an *isa* relation between a collocate from the corpus w_j and a possible sense of a PMW s_i cannot be established using supersense information (for the supersenses) or through transitive closure in the hypernym-hyponym hierarchy in WordNet (for `company`

[1] `http://www/eml-research.de/nlp/download/wikirelations.php`

and `organization`) for any sense of w_j, it is tried against the Wikipedia-based links.

This process transforms the grammatical annotation file and enhances it with the collocation estimates, as shown in Table 2 (compare this with a sample of the original file presented in Figure 1).

4.3 Product and event features

Farkas et al. (2007) observed that using the PMWs themselves as features leads to improvement on determining the reading for organization names, and postulate that this is because some company names are more likely to be used in a metonymic way. This is often the case with companies that make products which are commonly used (cars, for example).

Brun et al. (2007) note that certain locations, such as *Vietnam*, are more likely to be used with an *event* reading than others locations. Generally, locations strongly associated with events tend to be used to refer to the event, and more often have a `place-for-event` interpretation rather than a `literal` one.

These two observations have lead us to mine for these pieces of information in the Wikipedia relations, and to add two more features for a target PMW:

has-product will take a value of 1 if any of the PMW's hypernyms (according to the *isa* relations extracted from Wikipedia) contains the string *manufacturer,* will have the value 0 otherwise;

has-event will have the value 1 if any of the PMW's hypernyms refers to an event (movements/operations/riots), and value 0 otherwise.

4.4 Data representation

As mentioned before, the representation built can be seen as consisting of roughly three subsets of features:

- **the M&N features** proposed by Nissim and Markert (2005). To combine the grammatical information from all relations, we transform the grammatical relations into features (as opposed to values). For a relation *subject* for example, we generate a binary `subject` feature that indicates whether for a given target this grammatical relation is filled or not, and a `subject_lemma` feature , whose value is the lemma of the grammatically related word.
- **the selectional preference scores**. Each of these features corresponds to one of the elements of S, presented above. These features combine the selectional preferences of all the grammatical relations for one target PMW.
- **product and event information from Wikipedia** – has-product and has-event.

The grammatical annotation file consists of one entry for each grammatical relation in which a PMW appears. For the final representation, information about all relations of a given PMW is compressed into one instance. Because the basic features were binarized, and instead of having one *grammatical role* feature now each possible grammatical relation has its own feature, combining several entries for one PMW is easy, as it only implies setting the correct value for the grammatical relations that are valid in the PMWs context.

The final representation consists of 63 features + class feature for the subset for company PMWs, 59 features + class feature for the subset containing countries PMWs. The sample ID and the PMW itself were not part of this representation.

5 Results

The models for determining a PMW's correct interpretation are learned on the training data provided, and evaluated on the test portion, using the answer keys and evaluation script provided with the data. For learning the models we use Weka (Witten and Frank, 2005), and select the final learning algorithms based on 10-fold cross-validation on the training data. We have settled on support vector machines (SMO in Weka), and we use the learner's default settings.

Tables 3 and Table 4 show the results obtained, and the baseline and the best results from the SemEval task for comparison (Markert and Nissim, 2007). The baseline in Table 3 corresponds to classifying everything as the most frequent class – literal interpretation. The *M&N feat.* and *M&N feat.bin.* correspond to datasets that contain only the M&N features and the binarized versions of these features, respectively. *SemEval best* gives the best results obtained on each task in the SemEval 2007 task (Markert and Nissim, 2007). SMO_{wiki} are the results obtained with the complete feature set described in Section 4, and SMO_{SP} are the results obtained when only the new features are used – only selectional preference, has-product and has-event features (none of

task ↓ method →	baseline	SemEval best	SMO$_{wiki}$	SMO$_{SP}$	M&N$_{feat.}$	M&N$_{feat.bin.}$
LOCATION-COARSE	79.4	85.2	86.1	82.8	79.4	83.4
LOCATION-MEDIUM	79.4	84.8	85.9	82.6	79.4	82.3
LOCATION-FINE	79.4	84.1	85.0	82.0	79.4	81.3
ORGANIZATION-COARSE	61.8	76.7	74.9	66.6	73.8	74.0
ORGANIZATION-MEDIUM	61.8	73.3	72.4	65.0	69.8	69.4
ORGANIZATION-FINE	61.8	72.8	71.0	64.7	68.4	68.5

Table 3: Accuracy scores

task ↓ method →	base	max	SMO$_{wiki}$	SMO
LOCATION-COARSE				
literal	79.4	91.2	91.6	91.6
non-literal	20.6	57.6	59.1	58.8
LOCATION-MEDIUM				
literal	79.4	91.2	91.6	91.6
metonymic	18.4	58.0	61.5	61.5
mixed	2.2	8.3	16	8.7
LOCATION-FINE				
literal	79.4	91.2	91.6	91.6
place-for-people	15.5	58.9	61.7	61.7
place-for-event	1.1	16.7	0	0
place-for-product	1.1	0	0	0
obj-for-name	0.4	66.7	0	0
obj-for-rep	0	0	0	0
othermet	1.2	0	0	0
mixed	2.2	8.3	16	8.7
ORGANIZATION-COARSE				
literal	61.8	82.5	81.4	81.2
non-literal	38.2	65.2	61.6	60.7
ORGANIZATION-MEDIUM				
literal	61.8	82.5	81.4	81.2
metonymic	31.0	60.4	58.7	58.1
mixed	7.2	30.8	26.8	28.9
ORGANIZATION-FINE				
literal	61.8	82.6	81.4	81.2
org-for-members	19.1	63.0	59.7	59.2
org-for-event	0.1	0	0	0
org-for-product	8.0	50.0	44.4	44
org-for-facility	2.0	22.2	36.3	38.1
org-for-index	0.3	0	0	0
org-for-name	0.7	80.0	58.8	58.8
org-for-rep	0	0	0	0
othermet	1.0	0	0	0
mixed	7.2	34.3	27.1	29.3

Table 4: Detailed F-scores

the M&N features). The baseline for detailed reading results in Table 4 reflects the distribution of the classes in the test file. The *max* column shows the best performance for each task in the SemEval 2007 competition (Markert and Nissim, 2007). The SMO column shows the results of learning when Wikipedia information is not used to compute the values of the collocation, has-product and has-event features.

Nissim and Markert (2003) have shown that grammatical roles are very strong features. Experiments on the data represented exclusively through grammatical role features confirm this observation, as the results obtained using only the syntactic features (no lexical head information) give the same results as the *M&N feat.bin.* which does include lexical information.

On the location metonymies, the current approach performs better on all evaluation types (coarse, medium, fine) by 0.9, 1.1 and 0.9% points respectively. The improvement comes from recognizing better the metonymic readings, as it is apparent from the detailed F-score results in Table 4. For the coarse readings, the F-score for the `non-literal` reading is 1.5% points higher than the best performance at SemEval, and 2.5% and 7.7% points respectively for the `metonymic` and `mixed` readings for the medium and fine coarseness. It is interesting that the learning is quite successful even when only selectional preference and Wikipedia-based has-product and has-event features are used – the SMO_{SP} column in Table 3. The grammatical role and the related lemma were used to derive these collocation features, but they do not appear as such in the representation used for this batch of experiments.

For company metonymies the current approach does not perform better than the state-of-the-art. For these metonymies the syntactic information is not as useful. This is evidenced by the lower performance of the classifier that uses only syntactic information (column *M&N feat.bin.* in Table 3), despite the fact that the training dataset for companies is larger than the one for countries. This observation is further supported by the low results when using only selectional preference features. It indicates that for company metonymies the local context does not provide as strong clues as it does for locations. For such PMWs we should explore the larger context. We have made a start with the Wikipedia-based features built following the observation about companies and their products made by Farkas et al. (2007) and Brun et al. (2007). In future work we plan to analyse this matter further, and find a method to derive more such features, and without manually provided clues (such as *manufacturer* or *riots*).

Wikipedia derived information does not contribute very much, but as expected it is helpful to identify other classes than the `literal` one. It is helpful to detect the `mixed` class – 16% F-score when using Wikipedia information compared to 8.7% for the countries data when we estimate preferences using only WordNet. It also increases the performance on the `non-literal`, `metonymic` and `org-for-members` classes in coarse, medium and fine classification respectively for both countries and companies. There is a small improvement for recognizing the `org-for-product` reading for organizations when using Wikipedia-based features. It is an indication that the has-product feature is useful. We cannot draw conclusions about the has-event feature, as there are only 3 training instances for the `place-for-event` reading. The results are encouraging, as we have just scraped the surface of the information that Wikipedia can provide.

The corpus derived selectional preferences perform very well, especially for determining the reading of locations. Analysis of the data and the features gives some indication as to why this happens: in the grammatical annotations provided, when the PMW is a prepositional complement or has a prepositional complement, the grammatically related word is a preposition. We extract only grammatical collocations for open-class words, restricted by the grammatical relation of interest, so we do not extract collocations for prepositions. Location prepositions (*in, at, from*) are less ambiguous than others (e.g. *for*), which are more common for the organization data. We have attempted to bypass this problem by generating parses using the dependency output of the Stanford Parser (de Marneffe et al., 2006), and bypassing the preposition – incorporate it in the grammatical role (*pp_in*, for example), and using as lemma the head of the prepositional complement or the constituent which dominates the prepositional phrase, depending on the position of the PMW. Now we can use the grammatical relation and the associated open-class word to look for collocations. This approach did not lead to good results, because the quality of the automatic parses is far from the manually provided information.

6 Conclusions

We have explored the use of selectional preference scores derived from a sense untagged corpus as local constrains for determining the interpretation of potentially metonymic words. Such methods were previously successfully used for word sense disambiguation, and transfer nicely to the metonymy resolution task. Adding encyclopedic knowledge to the mix improved the results further, by filling in gaps for WordNet, and extracting information particular to PMW. We plan to expand on this, and find methods to extract more such features automatically, without manually provided clues.

For a more comprehensive treatment of metonymies one must take into consideration not only local context but also discourse relations. A possible avenue of research is to build upon coreference resolution systems, and use the mentions they detect and link to each other in a manner similar to using grammatical information and grammatically related words to determine constraints from a larger context. Determining the link between two mentions in a text can take advantage of encyclopedic knowledge, and the system's ability to infer the connection between the mentions.

There is much work on unsupervised word sense disambiguation. Working with untagged data gives a system access to a much larger information base. Since selectional preferences acquired from sense-untagged corpora have worked well for the metonymy resolution task, we plan to push further towards unsupervised metonymy resolution, putting to use the lessons learned from unsupervised WSD.

Acknowledgements

We thank the Klaus Tschira Foundation, Heidelberg, for financial support, and the organizers of the SemEval-2007 Task #8 *Metonymy Resolution* – Katja Markert and Malvina Nissim – for making the annotated data freely available.

References

Caroline Brun, Maud Ehrmann, and Guillaume Jacquet. 2007. XRCE-M: A hybrid system for named entity metonymy resolution. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-1),* Prague, Czech Republic, 23–24 June 2007, pages 488–491.

Marie-Catherine de Marneffe, Bill MacCartney, and Christopher D. Manning. 2006. Generating typed dependency parses from phrase structure parses. In *Proceedings of the 5th International Conference on Language Resources and Evaluation,* Genoa, Italy, 22–28 May 2006, pages 449–454.

Richard Farkas, Eszter Simon, Gyorgy Szarvas, and Daniel Varga. 2007. GYDER: Maxent metonymy resolution. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-1),* Prague, Czech Republic, 23–24 June 2007, pages 161–164.

Dan C. Fass. 1991. met*: A method for discriminating metonomy and metaphor by computer. *Computational Linguistics*, 17(1):49–90.

Sanda M. Harabagiu. 1998. Deriving metonymic coercions from WordNet. In *In Workshop on the Usage of WordNet in Natural Language Processing Systems,* Montreal, Canada, August 16, 1998, pages 142–148.

Jerry Hobbs, Mark Stickel, Douglas Appelt, and Paul Martin. 1993. Interpretation as abduction. *Artificial Intelligence*, 63(1-2):69–142.

Shin-ichiro Kamei and Takahiro Wakao. 1992. Metonymy: Reassessment, survey of acceptability, and its treatment in a machine translation system. In *Proceedings of the 30th Annual Meeting of the Association for Computational Linguistics*, Newark, Del., 28 June – 2 July 1992, pages 309–311.

Adam Kilgarriff, Pavel Rychly, Pavel Smrz, and David Tugwell. 2004. The Sketch Engine. In *Proceedings of the 11th International Congress of the European Association for Lexicography,* Lorient, France, 6–10 July 2004, pages 105–116.

Dekang Lin. 1998. An information-theoretic definition of similarity. In *Proceedings of the 15th International Conference on Machine Learning,* Madison, Wisc., 24–27 July 1998, pages 296–304.

Katja Markert and Udo Hahn. 2002. Metonymies in discourse. *Artificial Intelligence*, 135(1/2):145–198.

Katja Markert and Malvina Nissim. 2002. Metonymy resolution as classification task. In *Proceedings of the 2002 Conference on Empirical Methods in Natural Language Processing,* Philadelphia, Penn., 6–7 July 2002, pages 204–213.

Katja Markert and Malvina Nissim. 2003. Corpus-based metonymy analysis. *Metaphor and Symbol*, 18(3):175–188.

Katja Markert and Malvina Nissim. 2007. SemEval-2007 Task 08: Metonymy Resolution at SemEval-2007. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-1),* Prague, Czech Republic, 23–24 June 2007, pages 36–41.

Vivi Nastase and Michael Strube. 2008. Decoding Wikipedia category names for knowledge acquisition. In *Proceedings of the 23rd Conference on the Advancement of Artificial Intelligence,* Chicago, Ill., 13–17 July 2008, pages 1219–1224.

Vivi Nastase. 2008. Unsupervised all-words word sense disambiguation with grammatical dependencies. In *Proceedings of the 3rd International Joint Conference on Natural Language Processing,* Hyderabad, India, 7–12 January 2008, pages 757–762.

Cristina Nicolae, Gabriel Nicolae, and Sanda Harabagiu. 2007. UTD-HLT-CG: Semantic architecture for metonymy resolution and classification of nominal relations. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-1),* Prague, Czech Republic, 23–24 June 2007, pages 454–459.

Malvina Nissim and Katja Markert. 2003. Syntactic features and word similarity for supervised metonymy resolution. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics,* Sapporo, Japan, 7–12 July 2003, pages 56–63.

Malvina Nissim and Katja Markert. 2005. Learning to buy a Renault and talk to BMW: A supervised approach to conventional metonymy. In *Proceedings of the 6th International Workshop on Computational Semantics,* Tilburg, Netherlands, January 12-14, 2005.

James Pustejovsky. 1991. The generative lexicon. *Computational Linguistics*, 17(4):209–241.

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

David Stallard. 1993. Two kinds of metonymy. In *Proceedings of the 31st Annual Meeting of the Association for Computational Linguistics,* Columbus, Ohio, 22–26 June 1993, pages 87–94.

Gustaf Stern. 1931. *Meaning and Changes of Meaning*. Indiana University Press, Bloomington, Indiana. (1968; first published in Sweden 1931).

Ian H. Witten and Eibe Frank. 2005. *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufmann, San Francisco, Cal., 2nd edition.

David Yarowsky. 1995. Unsupervised word sense disambiguation rivalling supervised methods. In *Proceedings of the 33rd Annual Meeting of the Association for Computational Linguistics,* Cambridge, Mass., 26–30 June 1995, pages 189–196.