

Local and Global Context for Supervised and Unsupervised Metonymy Resolution

Vivi Nastase
HITS gGmbH
Heidelberg, Germany
vivi.nastase@h-its.org

Alex Judea
University of Stuttgart
Stuttgart, Germany
alexander.judea@ims.uni-stuttgart.de

Katja Markert
University of Leeds
Leeds, UK
K.Markert@leeds.ac.uk

Michael Strube
HITS gGmbH
Heidelberg, Germany
michael.strube@h-its.org

Abstract

Computational approaches to metonymy resolution have focused almost exclusively on the local context, especially the constraints placed on a potentially metonymic word by its grammatical collocates. We expand such approaches by taking into account the larger context. Our algorithm is tested on the data from the metonymy resolution task (Task 8) at SemEval 2007. The results show that incorporation of the global context can improve over the use of the local context alone, depending on the types of metonymies addressed. As a second contribution, we move towards unsupervised resolution of metonymies, made feasible by considering ontological relations as possible readings. We show that such an unsupervised approach delivers promising results: it beats the supervised most frequent sense baseline and performs close to a supervised approach using only standard lexico-syntactic features.

1 Introduction

With the exception of explicit tasks in metonymy and metaphor analysis, computational treatment of language relies on the assumption that the texts to be processed have a literal interpretation. This contrasts with the fact that figurative expressions are common in language, as exemplified by the metonymy in the excerpt from a Wikipedia article in Example 1 and another in Example 2 from the SemEval 2007 metonymy resolution task (Markert and Nisim, 2009).

- (1) In the gold medal game, *Canada* defeated the American team 2-0 to win their third consecutive gold.
- (2) This keyword is only required when your relational database is *Oracle*.

The *defeating* in Example 1 will not be done by the country as such, but by a team representing the country in a sporting event. Hence, in a metonymy a potentially metonymic expression or word (here *Canada*) stands for a conceptually related entity (here, people of Canada). In the second Example, a company name (*Oracle*) stands for a product (database) developed by the company.

Metonymy resolution can be important for a variety of tasks. Textual entailment may need metonymy resolution (Bentivogli et al., 2007): for example, we would like to be able to induce from Example 1 the hypothesis

The Canadian team won

Leveling and Hartrumpf (2008) show that metonymy recognition on location proper names helps geographical information retrieval by excluding metonymically used place names from consideration (such as Example 1 or the use of *Vietnam* for the *Vietnam war*). Metonymies also frequently interact with anaphora resolution (Nunberg, 1995; Markert and Hahn, 2002), as in Example 1 where the metonymic use of *Canada* is referred to by a plural pronoun afterward (*their*).

Metonymies can be quite regular: company names can be used for their management or their products, country names can be used for associated sports teams. Following from this, the currently

prevalent set-up for metonymy resolution — as in the SemEval 2007 task — provides a manually compiled list of frequent readings or metonymic patterns such as *organization-for-product* for pre-specified semantic classes (such as organizations) as well as annotated examples for these patterns so that systems can then treat metonymy resolution as a (supervised) word sense disambiguation task. However, this approach needs novel, manual provision of readings as well as annotated examples for each new semantic class.

In contrast, we will see readings as relations between the potentially metonymic word (PMW) and other concepts in a large concept network, a priori allowing all possible relations as readings. We base this approach on the observation that metonymic words stand in for concepts that they are related with — e.g. the part for the whole, the company for the product. These readings are obtained on the fly and are therefore independent of manually provided, preclassified interpretations or semantic classes, leading eventually to the possibility of unsupervised metonymy resolution. We achieve this by first linking a PMW to an article in Wikipedia. Then we extract from a large concept network derived from Wikipedia the relations surrounding the PMW.

As there will be (many) more than one such relation, these need to be ranked or scored. We achieve this in a probabilistic framework where we condition the probability of a relation on the context of the PMW. This ranking showcases our second major innovation in that the flexibility of our framework allows us to incorporate a wider context than in most prior approaches. Let us consider the indications for metonymic readings and its interpretation in Example 1, on the one hand, and Example 2, on the other hand. In Example 1, the grammatical relation to the verb *defeat* and the verb’s selectional preferences indicate the metonymy. We will call all such grammatically related words and the grammatical relations *the local context* of the PMW. Such types of local context have been used by most prior approaches (Pustejovsky, 1991; Hobbs et al., 1993; Fass, 1991; Nastase and Strube, 2009, among others). However, Example 2 shows that the local context can be ambiguous or often weak, such as the verb *to be*. In these examples, the wider context (*database, key-*

word) is a better indication for a metonymy but has not been satisfactorily integrated in prior approaches (see Section 2). We here call all words surrounding the PMW but not grammatically related to it *the global context*.

In our approach we integrate both the local and the global context in our probabilistic framework. For the local context, we compute the selectional preferences for the words related to the PMW from a corpus of English Wikipedia articles and generalize them in the Wikipedia concept network, thus (automatically) providing a set of *abstractions* — general concepts in the network that capture the semantic classes required by the local context. In the next step we compute probabilities of the global context surrounding the PMWs under each (locally required) abstraction, and combine this with the selectional preferences of the grammatically related words. That we can integrate local and global context in one probabilistic but also knowledge-based framework is possible because we combine two descriptions of meaning — ontological and distributional — by exploiting different sources of information in Wikipedia (category-article hierarchy and article texts).

We compute the probabilities of the relations (= readings) between the concept corresponding to the PMW and its directly related concepts. These can be used either (i) as additional features in a supervised approach or (ii) directly for unsupervised resolution. We do both in this paper and show that (i) the supervised approach using both local and global context can outperform one using just local context, dependent on the semantic class studied and (ii) that an unsupervised approach — although lower than the supervised one — outperforms the supervised most frequent reading baseline and performs close to a standard supervised model with the basic set of lexico-syntactic features (Nissim and Markert, 2005).

2 Related Work

The word sense disambiguation setting for metonymy resolution as developed by Nissim and Markert (2005) and used for the SemEval 2007 task (Markert and Nissim, 2009) uses a small, pre-specified number of frequently occurring readings.

The approaches building on this work (Farkas et al., 2007; Nicolae et al., 2007, among others) are supervised, mostly using shallow surface features as well as grammatical relations.¹ Most effective in the SemEval task as summarized in Markert and Nissim (2009) has been the local, grammatical context, with the two systems relying on the global context or the local/global context in a BOW model (Leveling, 2007; Poibeau, 2007) not outperforming the most frequent reading baseline. We believe that might be due to the lack of a link between the local and global context in these approaches — in our work, we condition the global context on the abstractions and selectional preferences yielded by the local context and achieve better results.

Lapata (2003), Shutova (2009) as well as Roberts and Harabagiu (2011) deal with the issue of logical metonymy, where the participant stands in for the full event: e.g. *Mary enjoyed the book.*, where *book* stands in for *reading the book*, and this missing event (*reading*) can be inferred from a corpus. Utiyama et al. (2000), Lapata (2003) propose a probabilistic model for finding the correct interpretation of such metonymies in an unsupervised manner. However, these event type metonymies differ from the problem dealt with in our paper and the SemEval 2007 task in that their recognition (i.e. their distinction from literal occurrences) is achieved simply by grammatical patterns (a noun instead of a gerund or to-infinitive following the verb) and the problem is limited to interpretation.

Our view of relations in a concept network being the interpretations of metonymies is strongly reminiscent of older work in metonymy resolution such as Hobbs et al. (1993), Fass (1991), Markert and Hahn (2002) or the use of a generative lexicon and its relations in Pustejovsky (1991), which also are unsupervised. However, these approaches lacked scalability due to the use of small hand-modeled knowledge bases which our use of a very large Wikipedia-derived ontology overcomes. In addition, most of these approaches (Fass, 1991; Hobbs et al., 1993; Pustejovsky, 1991; Harabagiu, 1998) rely on the view that metonymies violate selectional restrictions in their immediate, local context, usually those

¹Brun et al. (2007) is semi-supervised but again relies on the local grammatical context.

imposed by the verbs on their arguments. As can be seen in the Example 2, this misses metonymies which do not violate selectional restrictions. Nastase and Strube (2009) use more flexible probabilistic selectional preferences instead of strict constraint violations as well as WordNet as a larger taxonomy but are also restricted to the local context. Markert and Hahn (2002) do propose a treatment of metonymies that takes into account the larger discourse in the form of anaphoric relations between a metonymy and the prior context. However, they constrain discourse integration to potential PMWs that are definite NPs and the context to few previous noun phrases. In addition, their framework uses a strict rule-based ranking of competing readings that cannot be easily extended.

The work presented here also relies on a concept network, built automatically from Wikipedia. This resource provides us with links between entities in the text, and also a variety of ontological relations for the PMW, that will allow us to identify a wide variety of metonymic interpretations. Our approach combines information from the concept network with automatically acquired selectional preferences as well as a possibility to combine in a probabilistic framework the influence of the local and global context on the interpretation of a potentially metonymic word.

3 The Approach

The approach we present takes into account both the local, grammatical, context and the larger textual context of a potentially metonymic word. Figure 1 presents a graphical representation of our approach.

On the one hand, the word/term to be interpreted (the potentially metonymic word/term – PMW) is mapped onto a concept in the concept network (Section 3.3), which gives us access to the conceptual relations (\mathcal{R}_i) between the PMW and other concepts ($c_x \in \mathcal{C}_{\mathcal{R}_i}$). On the other hand, any word w grammatically related to the PMW via a grammatical relation r provides us with semantic restrictions on the interpretation of the PMW, namely preferred semantic classes A_j (we call them *abstractions*) and a selectional preference score.² These are automatically

²We restrict the grammatical context that provides selectional preferences to verbs or adjectives grammatically related

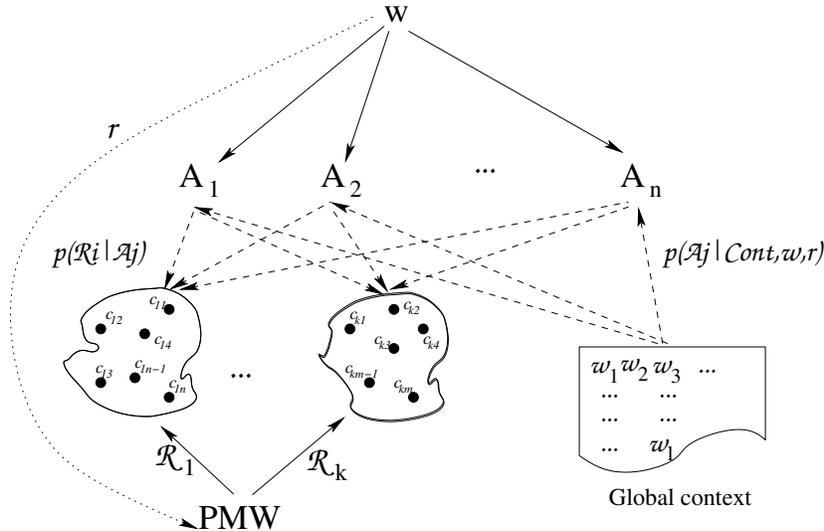


Figure 1: Metonymy resolution using selectional preferences A_j derived from local context w and r , semantic relations \mathcal{R}_i to the PMW from a concept network, and the global context surrounding a term to be interpreted

acquired by using a corpus of Wikipedia articles and a repository of encyclopedic knowledge (presented in Section 3.1), as described in detail in 3.2. Because the abstractions A_j and the PMW’s related concepts (c_x) come from the same structured resource, we can compute the probabilities for each \mathcal{R}_i given the grammatically related word w and the grammatical relation r . The global context can also easily be added to the computation, as the probability of each word in the context relative to an abstraction A_j can be computed through the resource’s *is_a* hierarchy and its link to Wikipedia articles. This is detailed in Section 3.4.

3.1 A concept network obtained from Wikipedia

We use a Wikipedia article dump (January 2011) which provided over 3.5 million English articles, interconnected through a hierarchy of categories and hyperlinks. This partly structured repository is transformed into a large-scale multilingual concept network, whose nodes are concepts corresponding to articles or categories in Wikipedia (Nastase et al., 2010). Concepts in this network are connected through a variety of semantic relations (e.g. *is a*, *member of*, *nationality*) derived from category names and infoboxes. The version of WikiNet used

had 3,707,718 nodes and 49,931,266 relation instances of 494 types, and is freely available³.

WikiNet is used here as a concept inventory, and its links and structure to generalize more specific concepts identified in texts to general concepts. The fact that nodes in WikiNet correspond to articles/categories in Wikipedia is used to link article texts in Wikipedia to general concepts, for the purpose of computing various probability scores (detailed in Section 3.4).

3.2 Selectional preferences and abstractions

To compute selectional preferences we use the set of English Wikipedia articles, which describe specific concepts. Wikipedia contributors are encouraged to insert hyperlinks, which link important terms in an article to the corresponding articles. A hyperlink consists of two parts, the actual link (i.e. a URL) and a phrase to appear in the text. Hyperlinks then constitute a bridge from the textual level to the conceptual level without the need for word sense disambiguation. We exploit these links to gather concept arguments for verbs and adjectives, and generalize these using the concept network built from Wikipedia.

The corpus of Wikipedia articles was first enriched with hyperlinks, making the “one sense per

to the PMW.

³<http://www.h-its.org/english/research/nlp/download/wikinet.php>

Algorithm 1 computeSelPrefs(G, WkN)

Input: G – grammatical relation triples
WkN – WikiNet
 M – maximum number of generalization steps

Output: Γ

```
 $\Gamma = \{\}$ 
for all  $(w, r)$  such that  $(c, r, w) \in G$  do
   $S = \{(c, f) | f \text{ is the frequency of } (c, r, w) \text{ in } G\}$ 
   $\Gamma_{w,r} = S$ 
   $mdl = MDL(\Gamma_{w,r}, S)$ 
  for all  $i = 1, M$  do
     $\Gamma' = abstract(S, WkN)$ 
     $mdl_{\Gamma'} = MDL(\Gamma', S)$ 
    if  $mdl_{\Gamma'} < mdl$  then
       $\Gamma_{w,r} = \Gamma'$ 
   $\Gamma = \{\Gamma_{w,r}\} \cup \Gamma$ 
return  $\Gamma$ 
```

Algorithm 2 MDL(Γ, S)

Input: $\Gamma = \{(c, f)\}$ – a scored list of concepts
 S – the set of observations (concept collocates)

Output: MDL(Γ, S)

```
 $\hat{\theta} = \langle f_1, \dots, f_n \rangle; (c_i, f_i) \in \Gamma$ 
remove  $\{(c, f) \in \Gamma | f = 1\}$  // parameter description
length :
 $L(\hat{\theta} | \Gamma) = \frac{|\Gamma|-1}{2} * \log(|S|)$  // data description length :
for all  $(c, f) \in \Gamma$  do
   $L(S | \Gamma, \hat{\theta}) = L(S | \Gamma, \hat{\theta}) + f * \log(\frac{f}{hyponyms(c) * |\Gamma|})$ 
return  $L(\hat{\theta} | \Gamma) - L(S | \Gamma, \hat{\theta})$ 
```

Algorithm 3 abstract(S, WkN)

Input: $S = \{(c, f) | (w, R, c) \in G\}$
WkN – WikiNet

Output: S'

```
 $S' = \{\}$ 
for  $c | (c, -) \in S$  do
  while  $c$  has only one is_a link do
     $c = c', (c, is\_a, c') \in WkN$ 
     $C = \{(c', c) | (c, is\_a, c') \in WkN\}$ 
    for  $(c', c) \in C$  do
      if  $(c', f') \in S'$  then
        replace  $(c', f')$  with  $(c', f' + \frac{f}{|C|})$ ,  $(c, f) \in S$ 
        in  $S'$ 
      else
         $S' \cup = \{(c', f)\}, (c, f) \in S$ 
// Remove hyponyms.
for all  $\{(c, c') \in S' | (c', is\_a, c) \in WkN\}$  do
  // update frequency  $f$  of  $c$ 
   $f_c = f_c + f_{c'}, f \in S$ 
  delete  $c'$ 
return  $S'$ 
```

discourse” assumption – a phrase that appears associated with a hyperlink once in the article body will be associated with the same hyperlink throughout the article (this applies to the article title as well, which is not hyperlinked in the article itself). This new version of the corpus was then split into sentences, and those without hyperlinks were removed. The remaining 18 million sentences were parsed with a parallelized version of Ensemble⁴ (Surdeanu and Manning, 2010), and we extracted G , the set of all grammatical relations of the type (*verb, dependency, hyperlink*) and (*adjective, dependency, hyperlink*), with the hyperlinks resolved to their corresponding node (concept) in the network ($|G| = 1,578,413$ triples). For each verb and adjective in the extracted collocations, and for each of their dependency relations, their collocates were generalized in the network defined by the hypernym/hyponym relations in WikiNet following a method similar to the Minimum Description Length principle (Li and Abe, 1998).

Essentially, we aimed to determine a small set of (more general) concepts that describe the set of collocates for a word w and grammatical relation r . Starting from the concept collocates gathered, we go upwards following WikiNet’s *is_a* links, and for each node found that covers at least N concept collocates (N is a parameter, $N=2$ in the experiments presented here), the MDL score of the node is computed (Algorithm 2). We place a limit M on the number of upward steps in the hierarchy ($M=3$ in our experiments). The disjoint set of nodes that has the lowest overall MDL score is chosen (Γ), and for each node in this cut (which we call *abstraction*), we compute the selectional preference score, based on the number of concepts it dominates.

As an example, for the verb *defeat*, the corpus leads to collocations such as⁵:

defeat

nsubj

Earle Page (10357) – 8, Manuela Maleeva (1092361) – 7, New York Yankees (10128601) – 5, Tommy Haas (1118005) – 5, ...

obj

⁴<http://www.surdeanu.name/mihai/ensemble/>

⁵The format is:
Article name (Article Id) – frequency.

New York Yankees (10128601) – 9, Oakland Athletics (11641124) – 6, Phoenix Suns (11309373) – 4, Jason Suttie (10080653) – 3, Ravana (100234) – 3, ...

Determining abstractions and selectional preferences leads to the following information⁶:

defeat

nsubj
 Martial artists (118977183) – 0.5, Person (219599) – 0.3518, Interest (146738) – 0.037, ...

obj
 Video games (9570081) – 0.25, British games (24489088) – 0.25, Person (219599) – 0.1445, Interest (146738) – 0.1341, ...

3.3 Linking the PMW to the concept network

In our environment, linking the PMW to the concept network is equivalent to finding its corresponding concept in our ontology, WikiNet. We see this corresponding concept as the literal reading of the PMW. Doing so is a non-trivial task (see the Cross-Lingual Link Discovery task at NTCIR-9 (Tang et al., 2011) and the Cross-Lingual Entity Linking task – part of the Knowledge Base Population track – at TAC 2011⁷). In our particular setting, where we use the metonymy data from SemEval 2007, the domain of the PMW is well defined: locations and companies, respectively. Using these constraints, finding the corresponding Wikipedia articles is much simplified, by using the category hierarchy and constraining the concepts to fall under the *Geography* and *Companies* categories respectively. When multiple options are present, we find instead a matching disambiguation page. In this case we pick the article that is listed first on this disambiguation page. On a manually checked random sample, the accuracy of the approach was 100% (on a sample of 100 PMWs).

3.4 Scoring conceptual relations with local and global context

We work under the assumption that the concept corresponding to the PMW is related to the possible interpretations through a semantic relation, in particular one that is captured in the concept network. After

⁶The format is:

Concept name (Concept Id) – selectional preference score.

⁷<http://nlp.cs.qc.cuny.edu/kbp/2011/>

countries	: <i>Administrator_of,</i>	<i>Architect_of,</i>
	<i>Based_in,</i>	<i>Built_in,</i>
	<i>Continent,</i>	<i>...</i>
companies	: <i>Association,</i>	<i>Brand,</i>
	<i>Company,</i>	<i>Distributed_by,</i>
	<i>Executive_of,</i>	<i>...</i>

Table 1: Example conceptual relations

establishing the connection to the resource by linking the PMW to the concept c_{PMW} corresponding to its literal interpretation (see Section 3.3), we extract the relations in which it is involved ($\mathcal{R}_i, i = 1, k$), and the concepts it is connected to through these relations ($\mathcal{C}_{\mathcal{R}_i} = \{c_x | (c_{PMW} \mathcal{R}_i c_x)\}$). Table 1 shows examples of conceptual relations extracted for companies and countries.

We are interested in computing the likelihood of a conceptual relation being the correct interpretation of a PMW, given its local and global context $p(\mathcal{R}_i | Cont, w, r)$.

3.4.1 The local context

The local context considered in this work are all grammatically related verbs and adjectives w and their associated grammatical relation r . The grammatical analysis (see Section 3.2) provides the set of abstractions corresponding to the grammatically related word w and grammatical relation r : $A_j, j = 1, n$. Remember that these are local context constraints on the interpretation of the PMW.

Through the knowledge resource used we can establish and quantify connections between each c_x and A_j , and thus between each \mathcal{R}_i and A_j :

$$(3) \quad p(\mathcal{R}_i | A_j) = \sum_{x \in \mathcal{C}_{\mathcal{R}_i}} p(c_x | A_j)$$

where $p(c_x | A_j)$ is the probability of concept c_x under abstraction A_j , which is computed based on the semantic relations in WikiNet:

$$p(c_x | A_j) = \sum_H \prod_{h_i \in H} p(h_i | h_{i+1})$$

where H is in turn each path from c_x to A_j following *is_a* links in WikiNet, starting with c_x (i.e. $h_0 = c_x$) and ending in A_j . $p(h_i | h_{i+1})$ is the probability of the child node h_i given its ancestor h_{i+1} . Within this work we assume a uniform probability distribution in each node:

$$p(h_i|h_{i+1}) = \frac{1}{|\text{descendants}(h_{i+1})|}$$

Through this, it is straightforward that $\sum_{c_x} p(c_x|A_j) = 1$ when c_x ranges over all concepts subsumed by A_j , and is thus a valid probability distribution.

3.4.2 The global context

The abstractions obtained before are concepts. We extract all nodes in the network subsumed by these concepts, and their corresponding articles in Wikipedia (if they have one). This produces “abstraction-specific” article sets, based on which we compute the probability of the global context of a PMW for each abstraction. We are interested in the probability of an abstraction, given the context and the word w and grammatical relation r , which we compute as:

$$p(A_j|Cont, w, r) = \frac{p(Cont|A_j, w, r) * p(A_j, w, r)}{p(Cont, w, r)}$$

which, considering that $p(Cont, w, r)$ is the same for a given context, we approximate as

$$p(A_j|Cont) \approx p(Cont|A_j) * p(A_j, w, r)$$

$p(A_j, w, r) = p(A_j|w, r) * p(w, r)$, and we approximate it through the computed selectional preference $p(A_j|w, r)$, since $p(w, r)$ is constant for a given example to analyze.

$$\begin{aligned} p(Cont|A_j, w, r) &= \sum_{j=1}^n p(Cont|A_j) p(A_j|w, r) \\ &= \sum_{j=1}^n \left(\prod_{l=1}^m p(w_l|A_j) \right) p(A_j|w, r) \end{aligned}$$

where $Cont$ is the global context consisting of m words $w_l, l = 1, m$.⁸

⁸The global context therefore could be all words in a text or all words in a sentence or any other token-based definition in our framework. As the SemEval 2007 data gives metonymic examples in a three-sentence context we use all the words in the 3 sentences as our global context.

$$p(w_l|A_j) = \frac{\text{count}(w_l, A_j)}{|A_j|}$$

where A_j is the set of articles subsumed by abstraction A_j , and $\text{count}(w_l, A_j)$ is the number of times word w_l appears in the article collection A_j .

3.4.3 Putting it all together

This enables us now to compute $p(\mathcal{R}_i|Cont, w, r)$ based on the formulas 3, 4:

$$p(\mathcal{R}_i|Cont, w, r) = \sum_{j=1}^n (p(\mathcal{R}_i|A_j) * p(A_j|Cont, w, r))$$

4 Experiments

The computed probabilities for each conceptual relation (= potential readings) of the PMW in the concept network can be used as features in a supervised framework or directly as an unsupervised prediction, returning the most likely conceptual relation given the context as the required reading.

Although the latter is our ultimate goal, to allow comparison with related work from the metonymy resolution task (Task 8) at SemEval 2007, we first investigate the supervised set-up. We then simulate the unsupervised setting in Section 4.3.

4.1 Data

We use the data from the metonymy resolution task (Task 8) at SemEval 2007. It consists of training and test data for country and company names which are potentially metonymic. Table 2 shows the statistics of the data, and the possible interpretations for the PMWs. The training-test division was achieved randomly so that the test data can have metonymic readings for which no training data exists, showing again the limitations of a supervised approach of prespecified readings.

Grammatical features The features used by Nissim and Markert (2005), and commonly used for the supervised classification of metonymy readings (Markert and Nissim, 2009):

- grammatical role of PMW (subj, obj, ...);
- lemmatized head/modifier of PMW (announce, say, ...);

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 2: Statistics for the Task 8 data

- determiner of PMW (def, indef, bare, demonstr, other, ...);
- 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.

The annotations provided are dependency relations, many of which contain a preposition as an argument (e.g. (to, *pp*, UK) from the example ... *the visit to the UK of ...*). Such relations are not informative, but together with the head that dominates the prepositional complement (e.g. *visit to*) they may be. Because of this, we process the provided annotations and add wherever possible to the simple prepositions the head of their subsuming constituent. This would change the above mentioned dependency to (visit, *prep-to*, UK).

Semantic relations as features To evaluate the proposed approach we use the PMW’s conceptual relations as features. The feature values are the $p(\mathcal{R}_i|Cont, w, r)$ scores.

For the “countries” portion of the data this adds 109 semantic relation features, and for companies 29 features. Table 1 showed examples of these new features.

4.2 Supervised learning

We use the SMO classifier in the WEKA machine learning toolkit (Witten and Frank, 2000) with its standard settings, training on the SemEval 2007 (Task 8) training set.

Table 3 shows the results of various configurations on the test data, in comparison with a most frequent reading baseline (assigning literal to all PMWs) as well as a system M&N that shows the results computed using only the features proposed by Nissim and Markert (2005). In addition, we compare to the best results⁹ at SemEval 2007 (SE_{max}) and Nastase and Strube (2009) (N09). Nastase and Strube (2009) added WordNet supersenses as features, and their values are selectional preferences computed with reference to WordNet. These are similar to our abstractions, which in our approach serve to link the local and the global context to the ontological relations, but do not appear as features.

Our system SP shows the results obtained using the M&N features plus the conceptual relation features conditioned on both local and global context whereas SP_{local} and SP_{global} use conceptual relations conditioned on local ($p(A_j|Cont, w, r) \approx p(A_j|w, r)$) or global context ($p(A_j|Cont, w, r) \approx p(A_j|Cont) = \sum_{j=1}^n (\prod_{l=1}^m p(w_l|A_j))$) only.

While the differences in overall accuracies are small, there are significant differences in classifying individual classes, as shown in Tables 4 – 5¹⁰, where the *distrib.* column shows the class distribution in the test data. It is interesting to note that, in our setting, the global context is more useful than the local

⁹We show the best result for each category, not necessarily from the overall best performing system. This holds for Tables 4 and 5 as well.

¹⁰The detailed results for previous approaches are reproduced from (Nastase and Strube, 2009). We include only the classes that have a non-zero F-score for at least one of the presented approaches.

task ↓ method →	baseline	SE_{max}	N09	M&N	SP	SP_{local}	SP_{global}	SP_{unsup}
LOCATION-COARSE	79.4	85.2	86.1	83.4	85.8	83.0	85.0	82.5
LOCATION-MEDIUM	79.4	84.8	85.9	82.3	85.7	82.7	84.6	82.3
LOCATION-FINE	79.4	84.4	85.0	81.3	84.7	82.1	83.8	81.8
ORGANIZATION-COARSE	61.8	76.7	74.9	74.0	77.0	76.4	76.8	73.4
ORGANIZATION-MEDIUM	61.8	73.3	72.4	69.4	74.6	74.0	74.4	70.3
ORGANIZATION-FINE	61.8	72.8	71.0	68.5	72.8	71.9	72.7	68.3

Table 3: Accuracy scores

task ↓ method →	distrib.	SE_{max}	N09	SP	task ↓ method →	distrib.	SE_{max}	N09	SP
LOCATION-COARSE					ORGANIZATION-COARSE				
literal	79.4	91.2	91.6	91.4	literal	61.8	82.5	81.4	82.7
non-literal	20.6	57.6	59.1	58.5	non-literal	38.2	65.2	61.6	65.5
LOCATION-MEDIUM					ORGANIZATION-MEDIUM				
literal	79.4	91.2	91.6	91.4	literal	61.8	82.5	81.4	82.7
metonymic	18.4	58.0	61.5	61.6	metonymic	31.0	60.4	58.7	63.1
mixed	2.2	8.3	16	9.1	mixed	7.2	30.8	26.8	27.4
LOCATION-FINE					ORGANIZATION-FINE				
literal	79.4	91.2	91.6	91.4	literal	61.8	82.6	81.4	82.7
place-for-people	15.5	58.9	61.7	61.1	org-for-members	19.1	63.0	59.7	66.5
place-for-event	1.1	16.7	0	0	org-for-product	8.0	50.0	44.4	35.0
obj-for-name	0.4	66.7	0	0	org-for-facility	2.0	22.2	36.3	45.5
mixed	2.2	8.3	16	9.1	org-for-name	0.7	80.0	58.8	44.4
					mixed	7.2	34.3	27.1	27.4

Table 4: Fine-grained results for each classification task for countries (F-scores)

Table 5: Fine-grained results for each classification task for companies (F-scores)

one for resolving metonymies. Combining local and global evidence improves over both, indicating that the information they provide is not redundant.

For companies the difference is small in terms of accuracy, but in classification of individual classes the difference in performance is higher, but because of the small data size not statistically significant.

Countries in WikiNet have a high number of surrounding relations, because they are used as categorization criteria for professionals, for example, which generates fine-grained relations such as *Administrator_of*, *Ambassador_of*, *Chemist_of* Such a fine grained distinction between different professions for people in a country is not necessary, or indeed, desirable, for the metonymy resolution task. The results show that despite this shortcoming, the results are on par with the state-of-the-art, but in future work we plan to explore the task of relation generalization and its impact on the current task.

4.3 Simulating unsupervised metonymy resolution

In an unsupervised metonymy resolution approach, we would assign as interpretation the conceptual relation whose probability given the PMW, global and local contexts is highest. To simulate then the unsupervised metonymy resolution task, we make the relation features (used in the supervised approach) binary, where for each instance the relation that has highest probability has the value 1, the others 0. In case of ties there will be multiple features with the value 1 and if all features have the value 0, we set the value of relation *IS_A* (corresponding to a literal interpretation) to 1.

Using only the relation features simulates an unsupervised approach – this set-up learns a mapping between the relations used as features and the metonymy classes in the data used. Column

SP_{Unsup} in Table 3 shows the results obtained in this configuration. As expected the results are lower, but still close the supervised method when using only grammatical features (M&N). The results also significantly beat the baseline (apart from the Location-Fine setting). This indicates that relations are a viable substitute for manually provided classes in an unsupervised framework.

5 Conclusion

We have explored the usage of local and global context for the task of metonymy resolution in a probabilistic framework. The global context has been rarely used for the task of determining the intended reading of a potentially metonymic word (PMW) in context. We rely on automatically computed selectional preferences, extracted from a corpus of Wikipedia articles, and generalized based on a concept network also extracted from Wikipedia. Despite relying on automatically derived resources, the presented approach produces results on-a-par with current state-of-the-art systems. The method described here is also a step towards the unsupervised resolution of metonymic words in context, by taking into account knowledge about the concept corresponding to the literal interpretation of the PMW, and its relations to other concepts. This framework would also allow for exploring the metonymy resolution phenomena in various languages (since Wikipedia and WikiNet are multilingual), and investigate whether the same relations apply or different languages have different metonymic patterns.

Acknowledgments

Katja Markert is the recipient of an Alexander-von-Humboldt Fellowship for Experienced Researchers. This work was financially supported by the EC-funded project CoSyne (FP7-ICT-4-24853) and the Klaus Tschirra Foundation. We thank the reviewers for the helpful comments, and Helga Krämer-Houska for additional support for conference participation.

References

Luisa Bentivogli, Elena Cabrio, Ido Dagan, Danilo Giampiccolo, Medea Lo Leggio, and Bernardo Magnini.

2007. Building textual entailment specialized data sets: A methodology for isolating linguistic phenomena relevant to inference. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, La Valetta, Malta, 17–23 May 2010.
- 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.
- Richárd Farkas, Eszter Simon, György Szarvas, and Dániel 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 metonymy and metaphor by computer. *Computational Linguistics*, 17(1):49–90.
- Sanda M. Harabagiu. 1998. Deriving metonymic coercions from WordNet. In *Proceedings of the Workshop on the Usage of WordNet in Natural Language Systems*, Montral, Quebec, Canada, 16 August, 1998, pages 142–148.
- Jerry Hobbs, Mark Stickel, Douglas Appelt, and Paul Martin. 1993. Interpretation as abduction. *Artificial Intelligence*, 63(1-2):69–142.
- Maria Lapata. 2003. Probabilistic text structuring: Experiments with sentence ordering. In *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, Sapporo, Japan, 7–12 July 2003, pages 545–552.
- Johannes Leveling and Sven Hartrumpf. 2008. On metonymy recognition for geographic information retrieval. *International Journal of Geographical Information Science*, 22(3):289–299.
- Johannes Leveling. 2007. FUH (FernUniversität in Hagen): Metonymy recognition using different kinds of context for a memory-based learner. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-1)*, Prague, Czech Republic, 23–24 June 2007, pages 153–156.
- Hang Li and Naoki Abe. 1998. Generalizing case frames using a thesaurus and the MDL principle. *Computational Linguistics*, 24(2):217–244.
- Katja Markert and Udo Hahn. 2002. Metonymies in discourse. *Artificial Intelligence*, 135(1/2):145–198.
- Katja Markert and Malvina Nissim. 2009. Data and models for metonymy resolution. *Language Resources and Evaluation*, 43(2):123–138.
- Vivi Nastase and Michael Strube. 2009. Combining collocations, lexical and encyclopedic knowledge for metonymy resolution. In *Proceedings of the 2009*

- Conference on Empirical Methods in Natural Language Processing*, Singapore, 6-7 August 2009, pages 910–918.
- Vivi Nastase, Michael Strube, Benjamin Börschinger, Căcilia Zirn, and Anas Elghafari. 2010. WikiNet: A very large scale multi-lingual concept network. In *Proceedings of the 7th International Conference on Language Resources and Evaluation*, La Valetta, Malta, 17–23 May 2010.
- 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. 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.
- Geoffrey Nunberg. 1995. Transfers of meaning. *Journal of Semantics*, 12(1):109–132.
- Thierry Poibeau. 2007. Up13: Knowledge-poor methods (sometimes) perform poorly. In *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval-1)*, Prague, Czech Republic, 23–24 June 2007, pages 418–421.
- James Pustejovsky. 1991. The generative lexicon. *Computational Linguistics*, 17(4):209–241.
- Kirk Roberts and Sanda M. Harabagiu. 2011. Unsupervised learning of selectional restrictions and detection of argument coercions. In *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing*, Edinburgh, UK, 27-29 July 2011, pages 980–990.
- Ekaterina Shutova. 2009. Sense-based interpretation of logical metonymy using a statistical method. In *Proceedings of the Joint Conference of the 47th Annual Meeting of the Association for Computational Linguistics and the 4th International Joint Conference on Natural Language Processing*, Singapore, 2–7 August 2009, pages 1–9.
- Mihai Surdeanu and Christopher D. Manning. 2010. Ensemble Models for Dependency Parsing: Cheap and Good? In *Proceedings of Human Language Technologies 2010: The Conference of the North American Chapter of the Association for Computational Linguistics*, Los Angeles, Cal., 2–4 June 2010, pages 649–652.
- Ling-Xiang Tang, Shlomo Geva, Andrew Trotman, Yue Xu, and Kelly Y. Itakura. 2011. Overview of the NTCIR-9 crosslink task: Cross-lingual link discovery. In *Proceedings of the 9th NII Test Collection for IR Systems Workshop meeting – NTCIR-9 Tokyo, Japan*, 6–9 December 2011.
- Masao Utiyama, Masaki Murata, and Hitoshi Isahara. 2000. A statistical approach to the processing of metonymy. In *Proceedings of the 18th International Conference on Computational Linguistics*, Saarbrücken, Germany, 31 July – 4 August 2000, pages 885–891.
- Ian H. Witten and Eibe Frank. 2000. *Data Mining: Practical Machine Learning Tools and Techniques with Java Implementations*. Morgan Kaufmann, San Diego, CA.