HAL open science

Leveraging different meronym discovery methods for bridging resolution in French

Emmanuel Lassalle, Pascal Denis

To cite this version:

Emmanuel Lassalle, Pascal Denis. Leveraging different meronym discovery methods for bridging resolution in French. DAARC 2011 - 8th Discourse Anaphora and Anaphor Resolution Colloquium, Oct 2011, Faro, Portugal. pp.35-46, 10.1007/978-3-642-25917-3 . hal-00655916

HAL Id: hal-00655916

https://inria.hal.science/hal-00655916v1

Submitted on 3 Jan 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Leveraging different meronym discovery methods for bridging resolution in French

Emmanuel Lassalle and Pascal Denis

Alpage, INRIA & Univ. Paris Diderot
Sorbonne Paris Cité, F-75205 Paris, France
emmanuel.lassalle@ens-lyon.org,pascal.denis@inria.fr

Abstract. This paper presents a statistical system for resolving bridging descriptions in French, a language for which current lexical resources have a very low coverage. The system is similar to that developed for English by [20], but it was enriched to integrate meronymic information extracted automatically from both web queries and raw text using syntactic patterns. Through various experiments on the DEDE corpus [8], we show that although still mediocre the performance of our system compare favorably to those obtained by [20] for English. In addition, our evaluation indicates that the different meronym extraction methods have a cumulative effect, but that the text pattern-based extraction method is more robust and leads to higher accuracy than the Web-based approach.

1 Introduction

Bridging descriptions are a special kind of anaphora whose interpretation requires not only identifying an antecedent, but also inferring a specific relation linking it to the anaphor [6]. In this paper, we focus on *mereological* bridging anaphora (that is, cases wherein the inferred relation is a *part-whole* relation). An illustrative English example is given in (1):

(1) <u>The car</u> will not move. <u>The engine</u> is broken.

The resolution of bridging anaphora currently represents one of the most challenging tasks in discourse processing. It is considerably much harder than standard coreferential anaphora resolution for which shallow predictors (like distance, string matching, or morphosyntactic agreement) have been shown to be rather effective [15,25,18]. Part of the challenge is due to an important information bottleneck. Lexical resources like *WordNet* are still too poor and uneven in coverage to provide a realistic solution [22]. In turn, more recent approaches to bridging resolution have turned to web-based extraction methods to fill this "knowledge gap" [4,20]. To date, the most complete and best-performing approach combines focus and lexical distance predictors using machine learning techniques [20].

While there has been a lot of research on resolving bridging anaphora in English [22,28,4,20], much fewer work exist for other languages (see [27] for an attempt on German). In this paper, we develop a system that performs bridging

resolution in French; to our knowledge, this is first such system for this language. Note that languages other than English, the knowledge gap is even wider, since lexical resources are typically scarcer. Our system directly builds upon the system developed for English by [20], and enriches it in several ways. First, we refine the search engine queries to include a wider range of meronym-holonym extraction patterns (some of them specific to French). Second, and more substantially, we augment the system with an original iterative pattern-based relation extraction method from raw text.

Our system is evaluated on *DEDE* [8], a corpus of French definite descriptions annotated for bridging. Initial experiments indicate that our system achieves performance that compares favorably to those of [20] for English under the same simple evaluation protocols. We found that the two meronym discovery approaches have a cumulative effect for resolution accuracy, and further comparison between the two methods suggests extraction from raw text is less noisy. But we also show that, when tested in a more realistic setting, the system still has rather low accuracy (at least for French). Our analysis reveals that the current features are still not discriminative enough, which in turns suggests the need for acquiring additional relational data.

The rest of this paper is organized as follows. Section 2 briefly presents the corpus we used. In section 3, we first describe a free lexical database for French, as well as our two methods for extracting meronyms. Section 4 details various experiments wherein we use these different resources for bridging resolution. A detailed comparison of the two extraction methods is given in section 5. Finally, section 6 discusses related work and section 7 offers some conclusions and directions for future work.

2 DEDE: a French corpus for bridging descriptions

For this study, we use *DEDE*[1], a French corpus annotated with definite descriptions [8], built from *Le Monde* articles in 1987. The corpus is not only the first of its kind in French with a high agreement rate between annotators, but it also introduces a fine-grained classification of definite descriptions.

The corpus contains 4,910 descriptions classified in 5 main categories (autonomous, coreferential, associative, situational and non-referential) themselves further divided into subcategories, so that the corpus provides an appropriate resource for evaluating a complete system for processing definite descriptions. As we are solely interested in bridging resolution for now, we only detail the "associative" class (that is bridging).

The "associative" class contains 530 anaphoras of 4 different types (see [8] for details): 322 *MERO* (meronymy, e.g. *a tree/the trunk*), 87 *CIRC* (modifier-modified relation, e.g. *Paris/the region*), 97 *REL* (predicate-argument, e.g. *two robbers/the attack*), 24 *MOD* (relation introduced by a modifier e.g. *in July/the next month*). The *MERO* subclass involves 296 distinct holonym-meronym pairs.

[1] *http://www.cnrtl.fr/corpus/dede/*

It is worth noticing that the *MERO* subclass of *DEDE* collapses together several kinds of part-whole relation: it contains classical component-object relations (e.g., *trunk/tree*), unnecessary parts (e.g., *wine/diner*), and functions (e.g., *mayor/city*).

Some of bridging relations involve a proper name as antecedent such as: *steering committee/La Société Générale*. That should be taken into account in resolution. For more details on part-whole relations, please refer to [30,11].

3 Methods for meronymy extraction

Having access to lexical and encyclopedic knowledge is an essential ingredient for resolving definite bridging, for it is precisely the existence of an inferable relation that permits bridging. Static databases like WordNet suffer from a low coverage impacting the performance in bridging resolution. In French, similar databases are even smaller than in English, which further motivates the need for automatic lexical relation extraction methods. Another point is that bridging relations do not always correspond to a lexical relation (e.g. *fragment/pot*, about archaeology) so that we cannot expect to find some of those relations in a static lexicon, but perhaps in texts if their usage is frequent enough.

3.1 A static lexical database

As there is no *free* equivalent to the *Princeton WordNet*[2] for French, we use the *WOLF* database[3] [23], a resource built from *WordNet* and several multilingual resources. It has a total of $32,351$ synsets, which is more than the *French EuroWordNet* ($22,121$ synsets), but much less than English WordNet ($177,659$ synsets). Crucially, *WOLF* only covers 2.7% of meronyms in *DEDE*, whereas *WordNet* covers 38.8% of the corpus used in [22]. We will see during the experiment that the contribution of this resource in bridging resolution is very small because of this low coverage. In particular, many lexical distance calculations will be impossible because of this low coverage.

3.2 Querying meronyms from the Web

Recent approaches to bridging resolution compensate the lack of lexical data by using web-based methods [4,20] (see also [12,13] for related tasks). The idea is to query a search engine for syntactic patterns taken to instantiate a particular lexical relation, and use the number of hits returned as an estimate of the plausibility of that relation. In English, the pattern mostly used to instantiate meronymy is *"the X of the Y"*.

We also adopt this method (using *Yahoo!*[4]), with some modifications. In French, the equivalent of *"the X of the Y"* has different realizations depending on

[2] *http://wordnet.princeton.edu/*
[3] *https://gforge.inria.fr/projects/wolf/*
[4] *http://developer.yahoo.com/search/boss/*

the grammatical genders and numbers of X and Y. Because we used lemmatized lexical heads, we only took singular Xs and Ys[5], which gives the three initial patterns: *"le X du/de/de la Y"*.

Because the first patterns are not in one-to-one correspondence with the meronymy relation, we defined another set of nine syntactical patterns, based on the possessive. Our purpose is to combine them to obtain less noisy results. Different settings can be applied to combine the patterns, whether we want a better coverage or a better precision: we set a threshold that corresponds to a minimum of nonzero matches among all the patterns required to accept a meronym candidate. In our experiments, meronym candidates are *(potential antecedent, anaphora)* pairs for the *MERO* class of *DEDE*. The Web-based method has a very high coverage: with a threshold set at value 1, 92.9% of all meronyms were positively identified (=275/296 different meronyms in *DEDE*, or 292/322 annotated bridges). We will give more detailed results in section 5.

3.3 Extracting meronyms from raw text

To further alleviate the "knowledge gap", we set up a system that iteratively extracts lexical relations and corresponding syntactic patterns from raw text. This type of approach has proven quite efficient on problems similar to bridging [13,2,9,24]. Although it was developed independently, our system is most similar to [1] in the sense that it alternatively extracts syntactic patterns and pairs. However, we enhance the system by considering two kinds of pattern: (i) linear patterns (i.e., a n-gram between 2 words) as well as (ii) dependency patterns (i.e., a path between 2 words in the dependency tree). In the latter case, we use two kinds of pattern (common ancestor or domination, see figure 1).

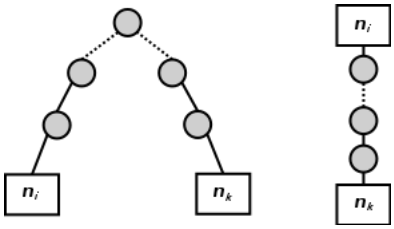

Fig. 1. Two types of dependency pattern.

We used a preprocessed corpus of 142M words from newspaper *L'Est Républicain*[6]. The system starts from a database of meronyms and syntactic patterns (linear or in dependencies) and iteratively extracts new pairs and patterns to enrich the initial database. To avoid to noisy results, we only retain pairs of nouns

[5] Another reason is the time it takes to query a search engine with meronym candidates: for all the patterns we used, it took more than a week.

[6] *http://www.cnrtl.fr/corpus/estrepublicain/*

appearing within two patterns or more (about 50,000 pairs). We also discard all pairs of patterns occurring less than a threshold of 5 (an empiric value).

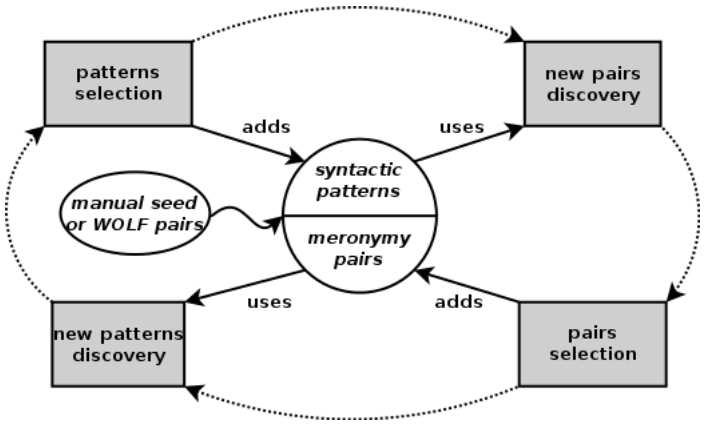

Fig. 2. Our system for extracting meronyms.

The system was initialized with two different types of seed data. When starting with a small set of manually crafted pairs and patterns, it retrieved 0.5% (=54) of meronyms in *WOLF* and 11.5% (=34) of meronyms in *DEDE*. When launched with the known meronyms in *WOLF* (with no patterns), the system was able to retrieve up to 16.2% (=48/296 different meronyms in *DEDE*, or 57/322 annotated bridges) of *distinct* meronyms in *DEDE* (by extracting 135 linear patterns and 42 dependency patterns), a much better coverage than *WOLF* but well below the Web-based method. This will be the final setting for our resolution system. A more restrictive selection of extracted pairs (e.g. with a ranking method like in [1]) would have reduced that coverage (some of *DEDE* meronyms were not found frequently by the system). But our purpose was mainly to reach an acceptable coverage while still filtering noisy pairs. Section 5 provides further comparison between the two extraction methods.

4 Experiments on bridging resolution

4.1 Task and system setup

Our system performs the following task: given a definite description known to be a meronymic bridging anaphora, find its correct antecedent among previously occurring NPs. We avoid the complete inference task on *DEDE* (jointly resolving bridging and coreference, finding the lexical relation, and the antecedent) to fully evaluate the performances of the system (and extraction methods) on meronymy.

Our system relies heavily on [20]. This system uses standard classification techniques to combine focus features (such utterance distance and first-mentionhood)

and lexical features (scores calculated from *WordNet* and *Google* results). Our feature set (see Table 1) enriches that of [20] with new lexical features, and an extra salience feature (salience is enriched by the grammatical function of the antecedent). *WOLF* is used to calculate lexical similarity and directly as a meronymy database. Extracted relations are used as a database (with an associated *Yahoo!* score for the Web-based method).

group	feature	description	type
salience	*utterance distance*	# of sentences separating p and a	*num*
	NPs distance	# of NPs separating p and a	*num*
	"is first mention"	p is the first mention of a sentence	*bool*
	grammatical function	the function of p (sub, obj, dep, ...)	*class*
known relations	*WOLF meronyms*	(a,p) is a meronymy in WOLF	*bool*
	extracted meronyms	(a,p) is an extracted meronymy	*bool*
lexical distance	*WOLF distance*	$\frac{L(ShortestPath(a,p))}{L(MaxLengthPath(a,p))}$ or 0 if no path	*num*
search engine	*Yahoo! score*	number of pages, on 12 patterns	*num/bool*

Table 1. Feature set (a is the anaphora and p a potential antecedent).

Like [20], we have experimented with a naive Bayes and multi-layer perceptron classifier, but we found no significant differences in performance. Results will therefore be reported only for the naive Bayes classifier (for which distances are modelled by exponential laws, and *Yahoo!* scores by a Gaussian).

4.2 Corpus preprocessing

DEDE corpus was preprocessed in several ways. First, we recovered accentuations: these were absent from the released texts. We performed NPs chunking (with TreeTagger[7]), as well as dependency parsing (using MSTParser[8]). We also separated texts in much finer and balanced documents than the originals (original documents contain several articles with non-related topics). All NP chunks that precede an anaphor in the text are considered as potential antecedents.

4.3 Evaluation and results

We use two different types of evaluation procedures. For comparison, we first follow the evaluation protocol of [20] and first report accuracy figures for positive and negative instances classification (all results were obtained by doing 5-fold cross-validations). Under this evaluation we obtain from 88.12% (salience only)

[7] *http://www.ims.uni-stuttgart.de/projekte/corplex/TreeTagger/*
[8] We used the retrained version of the MSTParser [14] described in [5].

to 90.0% (all features)[9]. These high scores come in part from the rareness of the bridging relation, and in turn the overwhelming number of negative instances. Recall on positives was around 76% with salience features, and lexical features did not bring any significant improvement.

baseline	accuracy	accuracy in [20]
random (previous) mention	1.7%	-
random mention, in first position of a sentence	1.7%	-
random mention in the previous 5 sentences	1.8%	4%
mention with best yahoo! score	1.9%	13.8%
mention in first position with best yahoo! score	2.0%	12%
random mention in the previous sentence	2.0%	19%
random in the previous 5 sentences, first position	2.0%	21.3%
best yahoo! score in the previous 5 sentences	3.9%	-

Table 2. Baselines comparison with [20]

A second, more realistic evaluation consists in evaluating the accuracy in finding the correct antecedent for a given bridging anaphora. For comparison, we used various baselines similar to those in [20]. As shown in table 2, the accuracy rates are consistently much lower than for equivalent baselines on Poesio *et al.*'s corpus, suggesting that the number of potential antecedents is probably much higher in *DEDE* corpus than the corpus used in [20]. Note that the low score of the baseline based on *Yahoo!* score suggests that the number of *hits* is not necessarily a good indicator.

features	accuracy
salience	7.9%
+ WOLF (distance/meronyms)	12.6%
+ Yahoo! score	22.4%
+ extracted meronyms	23.6%

Table 3. Overall resolution accuracy results.

We tested our system by successively adding features to the classifier (see table 3). First, note that the salience features make the system better than the baselines. Lexical features increase performance, and the different lexical resources appear to have a cumulative effect on performance (suggesing that they are somehow complementary). Under this evaluation, our best system obtains an

[9] The best score obtained in [20] among several classifier is 79.3%

overall accuracy of 23.6%, which is significantly lower than the scores reported by [20] on their corpus[10] but comparatively much better than the baselines[11]. The results were almost the same when ignoring *Yahoo!* scores and using extracted meronyms as a database (with a threshold on *Yahoo!* score to filter noisy pairs).

4.4 Error analysis

Preliminary error analysis reveals that (i) the correct antecedent was in the top 10 candidates (out of 206 on average) in $66,5\%$ of cases; (ii) many errors are due to selecting an antecedent with the same lexical head as the true one; (iii) the system fails in most cases where the antecedent is a proper name[12]; (iv) the system sometimes selects a meronym of the true antecedent (e.g. *room/first floor/hotel*). These results suggest that current features are still not discriminative enough, and point to a need for acquiring additional relational data using a more accurate model for selecting new pairs and patterns.

5 Web-based method vs. extraction from text

In order to gain a better understanding of the influence of the different meronym extraction methods on resolution, we perform some additional evaluation. When setting up the two methods, we were mainly concerned with maximizing coverage (to make up for the lack of coverage of WOLF). On that metric alone, the web-based extraction technique clearly outperforms both WOLF and the raw text extraction. Given that the two approaches produce noisy results, we also have to compare them on the basis of precision.

The evaluation protocol we propose is simple: all pairs (potential antecedent, anaphora) of *DEDE* are tested as meronym candidates (there are $45,605$ different pairs). A pair is marked as positive if it has been extracted by the method, and negative otherwise. Only meronyms marked in *DEDE* are considered true positives, and all the other pairs are negatives. We aware that this evaluation clearly over-estimates the number of false positives, but the point is meronyms that are not marked in *DEDE* are a small portion of all possible pairs and the aim of this test is to compare the noise of our two methods.

Looking first at the web-based approach, we compared extracted pairs to *DEDE* meronyms. For this, we considered a pair to be extracted if its *Yahoo!* score was above an arbitrary threshold. First, we tested all patterns separately and grouped them by gender variation (results in table 4).

[10] They report an impressive 92.5%, but their evaluation protocol consists in finding the correct antecedent for 6 anaphora randomly chosen in the evaluation set, which is absolutely not significant if not repeated a large number of times (but in that case, cross-validations would be wiser).

[11] Baselines are the only point of comparison we can have with [20] on this task.

[12] Proper nouns antecedents require a two steps inference (e.g. identify *Technip* as a company, identify the meronymy *employees/company*).

	TP	FP	prec.	recall	F1-score
X **d'** *Y*	74	6904	1.06 %	25.00 %	2.03 %
X **de** *Y*	245	29081	0.84 %	82.77 %	1.65 %
comb.	248	29781	0.83 %	83.78 %	1.64 %
X **du** *Y*	165	16159	1.01 %	55.74 %	1.99 %
X **de l'** *Y*	69	6924	0.99 %	23.31 %	1.89 %
X **de la** *Y*	135	13383	1.00 %	45.61 %	1.95%
comb.	252	28760	0.87 %	**85.14 %**	1.72 %
Y **dont le** *X*	72	5749	1.24 %	24.32 %	2.35 %
Y **dont l'** *X*	38	2949	1.27 %	12.84 %	2.31 %
Y **dont la** *X*	47	3731	1.24 %	15.88 %	2.31 %
comb.	144	12008	1.18 %	48.65 %	2.31 %
Y **et son** *X*	100	7483	1.32 %	33.78 %	2.54 %
Y **et sa** *X*	41	3569	1.14 %	13.85 %	2.10 %
comb.	136	10858	**1.24** %	45.95 %	**2.41 %**
X **de son** *Y*	98	7602	1.27 %	33.11 %	2.45 %
X **de sa** *Y*	33	4175	0.78 %	11.15 %	1.47 %
comb.	129	11189	1.14 %	43.58 %	2.22 %

Table 4. testing patterns independently on *DEDE* meronyms (best results were bold-faced)

French translations of *"the X of the Y"* (*"X du/de l'/de la Y"*) have the best coverage but one of the lowest precision. Patterns based on the possessive have a low coverage but they also less noisy. Finally, patterns obtaining the best F1-score are *"X et son/et sa Y"*.

Next we defined a more restrictive criterion to classify meronyms: a pair is classified as a meronym if it has been observed by the search engine with strictly more than x patterns, x being a parameter that varies from 0 to 11[13]. From the precision, recall and F1 values given in table 5 we see that combining the patterns leads to a better F1-score than using patterns separately: at $x = 7$, the F1-score has doubled (5.10%) compared to the best F1-score for isolated patterns. This motivates the use of several patterns with the web-based method.

Now turning to meronyms extracted from raw text (using the settings mentioned in section 3, we found 48 true positives and $1,096$ false positives, which gives a precision of 4.20%, a recall of 16.22%, and reaches a F1 of 6.66%, which is better than the F1-score for all settings of the web-based method (the comparison is acceptable because, as mentioned in section 4, *Yahoo!* scores do not have a strong influence on the performances of our resolution system compared to using a threshold on it and considering extracted pairs as a lexical database). Extracting meronyms from text seems not only less noisy than web-based methods, but

[13] Value 0 corresponds to combining all 12 patterns by disjunction.

	TP	FP	prec.	recall	F1-score
# *patterns* > 0	275	34319	0.79 %	**92.91 %**	1.58 %
# *patterns* > 1	246	27664	0.88 %	83.11 %	1.74 %
# *patterns* > 2	193	19661	0.97 %	65.20 %	1.92 %
# *patterns* > 3	150	12718	1.17 %	50.68 %	2.28 %
# *patterns* > 4	116	7569	1.51 %	39.19 %	2.91 %
# *patterns* > 5	72	3926	1.80 %	24.32 %	3.35 %
# *patterns* > 6	36	1452	2.42 %	12.16 %	4.04 %
# *patterns* > 7	20	469	4.09 %	6.76 %	**5.10 %**
# *patterns* > 8	7	100	**6.54 %**	2.36 %	3.47 %
# *patterns* > 9	0	14	0 %	0 %	-
# *patterns* > 10	0	4	0 %	0 %	-
# *patterns* > 11	0	0	-	0 %	-

Table 5. testing patterns combinations on *DEDE* meronyms (best results were bold-faced)

it also has the advantage of being faster than web queries, and the possibility to exploit more information (e.g. dependency parses or contextual information).

6 Related work

Bridging anaphora resolution was investigated by [28,20,4,27], among others. Earlier systems were rule-based, while more recent ones use machine-learning techniques. We can find similar work (web-based methods) on similar problems such as nominal anaphora [12] and metonymy [16]. [19] points out the problem of lack of knowledge with static lexical resources on bridging anaphora resolution and examines the contribution of lexical extraction techniques on this task.

Pattern-based lexical relation extraction from raw text dates back to [9], but fully automatic methods have now become available [1]. Recently, many advances have been made in open information extraction [31,32,10] and can be also applied to discovering lexical relations.

Extraction of relations from the Web using patterns were used in bootstrapping methods for discovering named entities or relations [3,7] as well as directly in the resolution of problems requiring lexical knowledge [4,20,12,13].

7 Conclusion and perspectives

We have set up and compared two approaches for extracting meronyms. We have improved the web-based method for French by using several patterns and combining them, and we have developed an original method pattern-based extraction method from raw text. We have then used the acquired lexical data as features in a system for resolving bridging anaphora in the case of a part-whole

relation. In the system, salience is a very relevant feature to localize antecedent in a region (in restricting the antecedent space), while lexical features must provide a better precision in this region to find the actual antecedent. However, our results show that current methods for discovering meronyms are still not sufficient to get reach good results in the resolution process. Finally, we have shown that the web-based method, even when it combines several syntactic patterns, is more noisy that extraction from raw text although it provides a wider coverage.

Improving the resolution system will first require more accurate data, both lexical and encyclopedic. In fact, some of the required data, such as proper nouns classes, can be retrieved from Wikipedia, and some other data, like lexical relations between common nouns, might be inferred by relational learning. Then, transforming Wikipedia into an ontology (like in [26,17]) and relational learning become central issues for a proper treatment of bridging. While the improvement of web-based methods appears limited, there are more possible extensions in relational learning that can be applied to text pattern-based methods. We now plan to extend these techniques and apply them to definite bridging processing.

Acknowledgments

We thank the *DAARC 2011* anonymous reviewers for their helpful comments.

References

1. Eugene Agichtein and Luis Gravano. Snowball: extracting relations from large plain-text collections. In *Proc. of DL*, pages 85–94, 2000.
2. Matthew Berland and Eugene Charniak. Finding parts in very large corpora. In *Proc. of ACL*, pages 57–64, 1999.
3. Sergey Brin. Extracting patterns and relations from the world wide web. In *Selected papers from the International Workshop on The World Wide Web and Databases*, pages 172–183, 1999.
4. Razvan Bunescu. Associative anaphora resolution: A web-based approach. In *Proc. of EACL Workshop on the Computational Treatment of Anaphora*, 2003.
5. Marie Candito, Joakim Nivre, Pascal Denis, and Enrique Henestroza Anguiano. Benchmarking of statistical dependency parsers for french. In *Proc. of COLING*, pages 108–116, 2010.
6. Herbert Clark. *Bridging*, pages 441–420. Cambridge University Press, 1977.
7. Oren Etzioni, Michael Cafarella, Doug Downey, Ana-Maria Popescu, Tal Shaked, Stephen Soderland, Daniel S. Weld, and Alexander Yates. Unsupervised named-entity extraction from the web: an experimental study. *Artif. Intell.*, 165, 2005.
8. Claire Gardent and Hélène Manuélian. Création d'un corpus annoté pour le traitement des descriptions définies. *Traitement Automatique des Langues*, 46-I, 2005.
9. Marti A. Hearst. Automatic acquisition of hyponyms from large text corpora. In *Proc. of COLING*, pages 539–545, 1992.
10. Raphael Hoffmann, Congle Zhang, and Daniel S. Weld. Learning 5000 relational extractors. In *Proc. of ACL*, pages 286–295, 2010.
11. Georges Kleiber. *L'anaphore associative*. PUF, 2001.

12. Katja Markert and Malvina Nissim. Using the web for nominal anaphora resolution. In *Proc. of EACL Workshop on the Computational Treatment of Anaphora*, pages 39–46, 2003.
13. Katja Markert and Malvina Nissim. Comparing knowledge sources for nominal anaphora resolution. *Comput. Linguist.*, 31(3):367–402, 2005.
14. Ryan T. McDonald, Koby Crammer, and Fernando C. N. Pereira. Online large-margin training of dependency parsers. In *Proc. of ACL*, 2005.
15. Ruslan Mitkov. *Anaphora resolution*. Pearson Education, 2002.
16. Vivi Nastase and Michael Strube. Combining collocations, lexical and encyclopedic knowledge for metonymy resolution. In *Proc. of EMNLP*, pages 910–918, 2009.
17. Vivi Nastase, Michael Strube, Benjamin Boerschinger, Caecilia Zirn, and Anas Elghafari. Wikinet: A very large scale multi-lingual concept network. In *Proc. of LREC*, Valletta, Malta, 2010.
18. Vincent Ng. Supervised noun phrase coreference research: The first fifteen years. In *Proc. of ACL*, Uppsala, Sweden, July 2010.
19. Massimo Poesio, Tomonori Ishikawa, Sabine Schulte im Walde, and Renata Vieira. Acquiring lexical knowledge for anaphora resolution. In *Proc. of LREC*, pages 1220–1224, 2002.
20. Massimo Poesio, Rahul Mehta, Axel Maroudas, and Janet Hitzeman. Learning to resolve bridging references. In *Proc. of ACL*, page 143, 2004.
21. Massimo Poesio and Renata Vieira. A corpus-based investigation of definite description use. *Computional Linguistics*, 24(2):183–216, 1998.
22. Massimo Poesio, Renata Vieira, and Simone Teufel. Resolving bridging references in unrestricted text. In *Proc. of Workshop on Operational Factors in Practical, Robust Anaphora Resolution for Unrestricted Texts*, pages 1–6. Association for Computational Linguistics, 1997.
23. Benoît Sagot and Darja Fišer. Construction d'un wordnet libre du français à partir de ressources multilingues. In *Proc. of TALN*, 2008.
24. Rion Snow, Daniel Jurafsky, and Andrew Y. Ng. Learning syntactic patterns for automatic hypernym discovery. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Proc. of NIPS*, pages 1297–1304. MIT Press, 2005.
25. Wee Meng Soon, Hwee Tou Ng, and Daniel Chung Yong Lim. A machine learning approach to coreference resolution of noun phrases. *Computional Linguistics*, 27(4):521–544, 2001.
26. Michael Strube and Simone Paolo Ponzetto. Wikirelate! computing semantic relatedness using wikipedia. In *Proc. of AAAI*, pages 1419–1424, 2006.
27. Yannick Versley. Using the web to resolve coreferent bridging in german newspaper text. In *Proc. of GLDV-Frühjahrstagung 2007*, 2007.
28. Renata Vieira and Massimo Poesio. Corpus-based development and evaluation of a system for processing definite descriptions. In *Proc. of COLING*, 2000.
29. Renata Vieira, Susanne Salmon-Alt, and Emmanuel Schang. Multilingual corpora annotation for processing definite descriptions. In *Proc. of ANLP*, pages 249–258, London, UK, 2002. Springer-Verlag.
30. Morton R. Winston, Roger Chaffin, and Douglas Herrmann. A taxonomy of part-whole relations. *Cognitive Science 11*, pages 417–444, 1987.
31. Fei Wu and Daniel S. Weld. Open information extraction using wikipedia. In *Proc. of ACL*, pages 118–127, 2010.
32. Jun Zhu, Zaiqing Nie, Xiaojiang Liu, Bo Zhang, and Ji-Rong Wen. Statsnowball: a statistical approach to extracting entity relationships. In *Proc. of WWW*, pages 101–110, 2009.