Hybrid acquisition of semantic relations based on context normalization in distributional analysis

Amandine Périnet[1,2]
[1] Lingua et Machina
c/o INRIA Rocquencourt
BP 105, 78153 Le Chesnay Cedex
ap@lingua-et-machina.com

Thierry Hamon[2]
[2] LIM&BIO (EA3969)
Université Paris 13, Sorbonne Paris Cité
74, rue Marcel Cachin, 93017 Bobigny
thierry.hamon@univ-paris13.fr

Abstract

Semantic relations between terms are important and usefull information for many applications that exploit specialized texts. In this paper we address the limits of semantic relation acquisition methods on such texts. Among these methods, distributional analysis is statistical and usually used with big amounts of data. But with low frequency words, improvements are still needed. To overcome this limit, we propose a hybrid method combining several approaches. We especially focus on the integration of three methods that acquire hyperonymy relations and morpho-syntactic variants in distributional contexts. We experiment the hybrid method on a corpus of nutrition, and evaluate the relations in terms of precision. The best hybrid model to acquire semantic relations appears to be the generalization of contexts with hyperonymy relations, for both nouns and terms as targets.

1 Introduction

Whatever the domain, specialized texts are characterized by terms and relations between terms. Identifying these relations is crucial in many applications in Natural Language Processing (NLP), such as information retrieval, question-answering systems, information extraction in search engines or specialized automatic translation. For instance, a semantic relation that links the terms *sucre (sugar)* and *saccharose* will allow to increase recall in a retrieval information system.

Those relations may be provided by terminologies, but usually those resources are not tuned to the targeted texts (Bourigault and Slodzian, 1999). Relations may also be automatically acquired from specialized corpora, through different strategies. We can take into account morphological (Grabar and Zweigenbaum, 2000), syntactic (Jacquemin, 1997) or semantic information (Jacquemin, 1999), define lexico-syntactic patterns through observation in corpora (Hearst, 1992; Morin, 1999; Auger and Barriere, 2008), use machine learning techniques (Snow et al., 2005) or distributional analysis (Habert and Zweigenbaum, 2002), etc. All the methods show various limits. Regarding the quality of the results, they can either get a low recall (methods are too restrictive) or a low precision (ambiguities or polysemy are not well identified). Furthemore, approaches usually aim at acquiring only one relation type (for eg., hyperonymy).

Let's take the example of two methods :

- Lexico-syntactic patterns allow to get a good precision, but are limited by their (very) low recall (Embarek and Ferret, 2008), because of the quite restricted contexts they use.
- On the contrary, distributional analysis (DA) is more flexible and allows to put many terms in relation, with a great diversity of relation types (Morlane-Hondère and Fabre, 2012), but without returning a type of relation. Indeed, DA do not seem to offer any obvious way to distinguish between syntagmatic (collocations, noun-verb relations) and paradigmatic relations (synonymy, hyperonymy) (Fabre and Bourigault, 2006).

Furthemore, methods based on DA are generally used with big amounts of data and tend to be less efficient with low frequency words (Caraballo, 1999). Results obtained with general language are promising, but improvement is still required with specialized texts, even if good results have already been achieved (Habert and Zweigenbaum, 2002).

As mentionned above, one of DA's limit is low frequency words. Indeed, for those words, similarity is computed from very little information (i.e. the one in contexts), that leads to generate poorer quality groupings of terms (Caraballo, 1999). We assume that this information could be increased with semantic information as the one contained in an existing resource or acquired by a relation acquisition method, as for example, using hyperonymy relations acquired with patterns. Following this idea, we intend to define a hybrid method that switches words in DA contexts for their hierarchical parent or morphosyntactic variant. This method normalizes contexts (Henneron et al., 2005), to increase their frequency.

We first present the related work, then our hybrid method and we finally describe the different experiments we led. The results we get are then evaluated in terms of precision.

2 Related work

This work uses a DA method, based on the harrissian hypothesis that states that words appearing in a similar context tend to be semantically close (Harris, 1954). The DA principle has been automated in the 90's, and concepts and procedures used in distributional computations have been well defined (Sahlgren, 2006; Turney and Pantel, 2010; Baroni and Lenci, 2010). However, this area of research still represents some current issues concerning the building, the evaluation and the use of distributional resources[1]. We focus here on the building of distributional resources.

In that respect, during the past few years, research has shifted from using DA methods for modelling the semantics of words to tuning them for the semantics of larger units such as phrases or entire sentences (Hermann et al., 2013). Most approaches tackle the problem through vector composition. Mitchell and Lapata (2008) use linear algebraic vector operations, testing both additive and multiplicative models, and a combination of these models. Grefenstette and Sadrzadeh (2011) apply unsupervised learning of matrices for relational words to their arguments, in order to compute the meaning of intransitive and transitive sentences. Baroni and Zamparelli (2010) use matrices to model meaning, but only for adjective-noun phrases, whereas Grefenstette and Sadrzadeh (2011)'s work also applies to sentences containing combinations of adjectives, nouns, verbes and adverbs. Recently, the framework proposed by Grefenstette et al. (2013) combines both approaches.

An important issue in DA improvement focuses on distributional contexts, and more precisely on weighting contexts. Broda et al. (2009) consider that what matters is not the feature's exact frequency. They do not use these frequencies as simple weights but rank contexts and take into account this rank in DA. Influence on contexts may also be done by embedding additional semantic information. With a method based on bootstrapping, Zhitomirsky-Geffet and Dagan (2009) modify the weights of the elements in contexts relying on the semantic neighbours found with a distributional similarity measure. Based on this work, Ferret (2013) faces the problem of low frequency words by using a set of positive and negative examples selected in an unsupervised way from an original distributional thesaurus to train a supervised classifier. This classifier is then applied for reranking the semantic neighbours of the thesaurus selection. With the same purpose of solving the problem of data sparseness, other methods are based on dimensionality reduction, as Latent Semantic Analysis (LSA) in (Padó and Lapata, 2007), or on a bayesian approach of DA (Kazama et al., 2010). Above work exploits great collection of texts of general language. However, few works are also interested in applying DA to specialized domains where text collections are generally smaller and frequencies lower (Habert and Zweigenbaum, 2002; Embarek and Ferret, 2008). As presented previously, Ferret (2013) attempts to exploit machine learning approaches to face the problem of low frequency of words and contexts. In our work, we propose an approach that exploits relations acquired with linguistic approaches in order to normalize contexts and increase their frequency. As we work with specialized texts, our approach dif-

[1]See for instance, the recent workshop at ACL 2013 https://sites.google.com/site/cvscworkshop/ or at the TALN 2013 http://www.taln2013.org/ateliers/appel-atelier-semantique-distributionnelle/

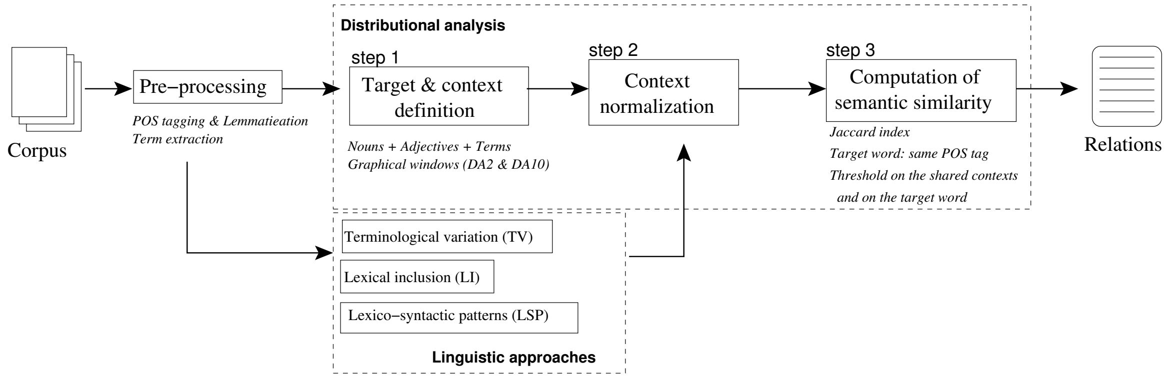

Figure 1: Processing steps

fers in considering nouns, adjectives and both simple and complex terms.

3 Hybrid method

The contexts in which occurs a target word have associated frequencies which may be used to form probability estimates. The goal of our hybrid method is to influence those distributional context frequencies by normalizing contexts. Indeed, normalization tends to decrease diversity in contexts in order to increase contexts' frequency. Our hybrid method follows the scheme presented in figure 1.

Target and context definition During Step 1, we define target words and contexts. Through the literature, syntactic analysis is mainly used to get dependency relations. But as it is time-consuming and heavy, we choose to use instead graphical windows within a sentence and around the target word. As we work on specialized texts, we also identify terms with the term extractor YaTeA (Aubin and Hamon, 2006).

We define the following parameters:

- Target words: words are in relation when they have the same POS tag; restricted to adjectives, nouns and terms.
- Distributional contexts: contexts are made of words that co-occur in a graphical window. In contexts, we don't take into account non-content words (determiners, conjunctions, adverbs, etc.) and keep only adjectives, nouns, verbs and terms.
- Fixed window size: we tested two different sizes described in section 4.
- Word form: for both contexts and target words we use the lemmas.

Linguistic approaches During the normalization process described below, we use three existing linguistic approaches: two methods that aim at acquiring hyperonymy relations and one that allows to get morphosyntactic variants.

- Lexico-syntactic Patterns (LSP): we use the patterns defined by (Morin and Jacquemin, 2004):
  1. {some | several etc.} NP : LIST.
  2. {other}? NP such as LIST.

  where NP is a noun phrase and LIST a list of noun phrases.
- Lexical Inclusion (LI): uses the syntactic analysis of the terms. Based on the hypothesis that if a term (ex: *épice (spice)*) is lexically included in another (ex: *épice aromatique (aromatic spice)*), there is a hyperonymy relation between the two terms generally (Bodenreider et al., 2001).
- Terminological Variation (TV): uses rules that define a morpho-syntactic transformation. This transformation may be an insertion, as the insertion of the adjective *aromatic* in *épice asiatique (asian spice) - épice aromatique asiatique (asian aromatic spice)* (Jacquemin, 1996).

Context normalization Once targets and contexts are defined comes the core of the hybrid method with context normalization. During

Step 2, we normalize contexts with the relations acquired by the three linguistic approaches we mentionned.

The relations are integrated in contexts in the following way: a word in context is replaced by its hyperonym or its morphosyntactic variant. We define two rules :

- If the word in context matches with only one hyperonym, context is replaced by this hyperonym. For example, if LSP give the relation *matière grasse (fat)/beurre (butter)*, *beurre (butter)* is replaced by *matière grasse (fat)*.
- If the context matches with several hyperonyms or variants, we take the hyperonym's or variant's frequency into account, and choose the one that is the most frequently in relation with the word in context. For example, if LSP give the following supposed hyperonyms: *matière grasse (fat)*, *pâte feuilletée (falky pastry)*, *béchamel*, *casserole (sauce pan)*, the one that enters the most frequently in relation with *beurre (butter)* is selected and used to replace this word in context.

We normalize contexts with each method separetly and sequentially: the first normalization is processed on all contexts before the second normalization starts, and so on.

Computation of semantic similarity When contexts have been normalized, similarity between two target words of the same POS tag is computed. As we decrease diversity in contexts during the normalization step, we choose among the existing measures (Weeds et al., 2004) a measure that favors words appearing in similar contexts compared to words appearing in many different contexts.

The Jaccard Index (Grefenstette, 1994) normalizes the number of contexts shared by two words by the total number of contexts of those two words.

$$sim{-}JACCARD_{mn} = \frac{|ctxt(w_m) \bigcap ctxt(w_n)|}{|ctxt(w_m) \bigcup ctxt(w_n)|}$$

Parameter: threshold We filter the relations according to three parameters, two of them applied on the contexts and the third one on the target.

- Number of shared contexts: number of lemmatized contexts. For example, if two words share *crème (cream)*, *battre (shake)*, *poivre (pepper)*, *sel (salt)*, *crèmes (creams)*, *battant (shaking)*, the number of shared contexts is 4.
- Frequency of the shared contexts: number of occurrences of the same lemma when shared in the context position of two target words. In the previous example, frequencies are *crème (cream)-2*, *battre (shake)-2*, *sel (salt)-1* and *poivre (pepper)-1*.
- Frequency of the target words: number of occurrences of the lemma in the target position.

For each parameter, a threshold is automatically computed, according to the corpus. It corresponds to the mean of the values taken by each parameter on the whole corpus.

4 Experiments

In order to evaluate the contribution and influence of relations acquired by the three methods, we define several sets of experiments and evaluate the relations acquired on existing resources.

4.1 Corpus

We use the merging of the two corpora provided by DEFT 2013 French challenge[2]: the training corpus (2,388,731 words) and the test corpus (1,539,927 words). They are both French corpora and contain cooking recipes. Each text of the corpus is made of a title, ingredients and the body of the recipe, and we use all the information.

We pre-process the corpus within the Ogmios platform (Hamon et al., 2007). We perform morphosyntactic tagging and lemmatization with Tree Tagger (Schmid, 1994), and use the term extractor YATEA(Aubin and Hamon, 2006).

4.2 Parameters and models of hybridization

In these different sets we vary two main kinds of parameters (cf. table 1): window size and models of hybridization.

We test two window sizes. With a large one of 20 words around the target (10 before, 10 after, henceforth W10) we may take into account the highest number of possible relations, because the average size of a sentence in French is 20 words

[2] http://deft.limsi.fr/2013/

Window size	4 (W2) and 20 (W10) words around the target
Hybridization	none: DAonly one method: DA/LSP, DA/LI, DA/TV two method combination: DA/LI+LSP, DA/LSP+LI, DA/TV+LSP, DA/LSP+TV three method combination: DA/LI+LSP+TV, DA/LSP+LI+TV, DA/TV+LSP+LI

Table 1: Parameters

and we restricted the relation acquisition to the sentence level. But such a large window may face a lack of specificity and get too much noise. We also test a window of 4 words (2 before, 2 after, henceforth W2). Such a size applied after removing the function words is comparable to a 8 word window applied to the original texts (Rapp, 2003).

We test different models of hybridization. We first use DA on its own, without normalizing the contexts (DAonly). This set is a reference to which compare the hybridization sets. As for the models of hybridization, we first separately evaluate the contribution of each method (LSP, LI, TV) in distributional context, and then different types of combinations of the methods integrated in DA. Within these combinations, we first exploit two methods together and then three. Our goal is to evaluate the impact of the order of the methods and the contribution of each method.

4.3 Comparison with existing resources

In order to evaluate the quality of the acquired relations, we compare our relations with three different resources: Agrovoc[3], of 75,222 relations [AGRO], and UMLS[4].

With the UMLS resource, we build two different resources: one more general [UMLS] of 2,325,006 relations, and a more specific one restricted to terms belonging to the *Food* concept (semantic type T168) [UMLS/Food] of 1,843 relations.

We only use the relations for the nouns and terms of our corpus, because adjectives were not represented in the resources. In that respect, we evaluate our work with 1,551 ([AGRO]), 1,800 ([UMLS]) and 871 ([UMLS/Food]) relations.

We use those three resources because of availability. The comparison with UMLS/Food and Agrovoc is justified by the presence of relations between food terms in both resources. But in cooking recipes, we may find other types of relations, as the relation between a food term and a term belonging to another semantic class. The comparison with the whole UMLS may allow to detect other relations than ingredient relations. Even if we can not expect an important overlap between these resources and the corpus, the comparison of our results to the relations issued from these resources gives an indication of the contribution of each proposed hybridization model.

We compute precision for each target term: semantic neighbours (acquired by our method) found in the resource by the semantic neighbours acquired by our method. For each target term, we sorted the semantic neighbours we obtained according to their similarity measure, and apply four thresholds: precision after examining 1 (P@1), 5 (P@5), 10 (P@10) and 100 (P@100) neighbours.

5 Results and discussion

We proceed to the analysis and discussion of the results we obtain with our hybrid method. Regarding the relations provided by the terminologies, we present here the results obtained for nouns and terms only.

We evaluate precison after examining four groups of neighbours. The best results are obtained with P@1, and decrease when we consider more neighbours: the more neighbours we consider, the lower precision is. For instance, for nouns-W10, precision decreases from 0.089 for P@1 to 0.009 for P@100, when compared with Agrovoc. We make similar observations on all the sets of results. Best results in first position means that the values of the measures rank quite correctly the proposed relations, and therefore that the choice of the measure was a good choice.

The table 2 presents the results for P@1, given the two window sizes (W10 and W2). We describe here only those results. The relations produced by DA (DAonly) are considered as our baseline. The low precision of our results was expected and

[3] http://aims.fao.org/standards/agrovoc/about
[4] http://www.nlm.nih.gov/research/umls/

Resources	Context definition and window size	DAonly	DA/TV	DA/LI	DA/LSP	DA/TV+LSP	DA/LI+LSP	DA/LSP+TV	DA/LSP+LI	DA/TV+LSP+LI	DA/LI+LSP+TV	DA/LSP+LI+TV
Agrovoc	Noun-W2	0.024	0.024	0.024	**0.073**	**0.072**	0.067	**0.073**	0.039	0.039	0.067	0.039
	Noun-W10	0.089	0.089	0.071	**0.109**	**0.109**	**0.111**	**0.109**	0.071	0.056	**0.111**	0.071
	Term-W2	0.023	0.024	0.000	**0.034**	**0.034**	0.000	**0.034**	0.000	0.000	0.000	0.000
	Term-W10	0.010	0.010	**0.031**	0.000	0.000	0.000	0.000	**0.047**	**0.047**	0.000	**0.047**
UMLS	Noun-W2	0.098	0.098	**0.139**	0.074	0.074	0.034	0.074	0.051	0.051	0.034	0.051
	Noun-W10	0.088	0.088	0.086	0.077	0.077	0.038	0.077	0.086	0.081	0.038	0.086
	Term-W2	0.094	**0.100**	0.000	**0.115**	**0.120**	0.000	**0.115**	0.000	0.000	0.000	0.000
	Term-W10	0.037	**0.042**	0.000	**0.042**	**0.043**	0.000	**0.042**	0.000	0.000	0.000	0.000
UMLS/food	Noun-W2	0.059	0.059	**0.094**	**0.083**	**0.080**	0.034	**0.083**	0.054	0.054	0.034	0.054
	Noun-W10	0.075	0.075	**0.095**	0.074	0.074	0.038	0.074	**0.095**	**0.102**	0.038	**0.095**
	Term-W2	0.070	0.071	0.000	**0.097**	**0.097**	0.000	**0.097**	0.000	0.000	0.000	0.000
	Term-W10	0.026	0.026	0.000	**0.032**	**0.031**	0.000	**0.032**	0.000	0.000	0.000	0.000

Table 2: Precision of the results against each resource after examining the first neighbour (P@1)

can be explained by the fact that even if the resources are relevant for our corpus, they are not fully adapted. However, the comparison of the precision values gives important information on the usefulness of the hybridization models.

Results are better for nouns (between 0.056 and 0.111 for nouns-W10 and between 0.024 and 0.073 for nouns-W2, with Agrovoc) than for terms (between 0 and 0.047 for terms-W10, and between 0 and 0.34 for terms-W2, with Agrovoc). This is not surprising because terms do not match easily with other terms in resources. This can be due to two main factors: terms are less frequent and it is difficult to match terms from the terminological resources in the corpus. As for the window size, we observe that generally W10 gives good results for nouns and W2 is better with terms. But when we look more in details, we observe that the quality of the results depends on the resource used for comparison. For nouns, with Agrovoc and UMLS/food, W10 gives the best results, but when compared with UMLS results are better with W2. The difference is similar with terms, but in this case results are better with W2 when compared with UMLS and UMLS/food, and better with W10 when compared with Agrovoc.

Linguistic approaches Considering the three methods individually, TV seems to have no influence on the computation of semantic similarity; the results obtained with DAonly and DA/TV are identical, except for terms W2 and W10 when compared with UMLS. Also, in all the hybrid sets, exploiting TV in the distributional contexts doesn't influence the results, except for DA/LI+LSP+TV with nouns-W10 when compared with Agrovoc and UMLS/food, and DA/TV+LSP with term-W2 when compared with UMLS. This may be because of the small number of relations used and our current way of DA hybridization with TV.

On the contrary, LSP is the method that most influences the results: most of the time they give better results than DA. The best hybridization model for terms is the normalization with LSP, whereas for nouns the combination of LI and LSP is the best choice. The order of the methods also matters, but results also differ according to the resource; DA/LSP+LI (and DA/LSP+LI+TV) give better results when compared with UMLS/food, and DA/LI+LSP (and DA/LI+LSP+TV) give better results when compared with Agrovoc. What emerge from these results is that generalization with hyperonyms is the best configuration, for both terms and nouns, and that the quality of the hyperonymy relation is important as well. Lexical inclusion used after patterns does not seem to bring new relations but allow to rule out noisy relations. By noisy relations, we mean relations not found in the resource. But these relations may be interesting and may be domain relations.

Resources and relation types The relations found by our method in UMLS/Food are co-hyponyms (eg: *ail (garlic)/oignon*), those found in Agrovoc are both hyperonyms (eg: *épice (spice)/poivre (pepper)*) and meronyms (eg: *miel (honey)/sucre (sugar)*). Relations found in the whole UMLS are the same as those found in UMLS/food. The identification of terms allow to find more relations, between simple terms and complex terms. For instance, in UMLS/food, our method found the co-hyponyms *poivre blanc (white pepper)/poivre noir (black pepper)* and *miel (honey)/fruit* that are not identified by taking into account nouns only.

6 Conclusion

In this work, we present our hybrid method based on normalization of distributional contexts. Our method aims at acquiring semantic relations from specialized texts, and is adapted to low frequency words. We normalize contexts with relations acquired by three linguistic approaches; two methods of hyperonymy relation acquisition and a method of morpho-syntactic variant acquisition. We focus on relations between nouns and terms. We tested our method on a French corpus composed of cooking recipes, varying one parameter in our DA method, the window size, and testing different models of normalization. Normalization obtains the best results when realized with hyperonyms and also depends on the quality of the hyperonymy relations. In our method, the hyperonym used for normalization is the one with the highest frequency. Even if precision values presented in this work are currently low and results differ according to the resource used for evaluation, it emerges that the best parameters are for nouns a W10 with LSP, and for terms a W2 with LSP and LI. This set of parameters is to be used for classical types of relations. But other types may be acquired with our DA++ method, especially domain specific relations. In order to have a better knowledge of the influence of each hybridization model, quality of the results has to be analyzed more deeply by manually checking with the validation of a subset of relations, and with a study of relations that are in common or not between the various results sets. For future work, we plan to investigate other strategies of normalization by assigning a weight to the relations proposed by the linguistic methods, or taking into account the level in the hierarchy. In that latter approach, the choice of the hyperonym used for the normalization could be guided by a distance (Resnik, 1995; Leacock and Chodorow, 1998). Relations used for normalization can also be issued from terminological resources. Furthemore, we will intend to combine the methods before normalization and exploit other similarity measures.

References

Sophie Aubin and Thierry Hamon. 2006. Improving term extraction with terminological resources. In *Advances in Natural Language Processing*, number 4139 in LNAI, pages 380–387.

Alain Auger and Caroline Barriere. 2008. Pattern-based approaches to semantic relation extraction: A state-of-the-art. *Terminology*, 14(1):1–19.

Marco Baroni and Alessandro Lenci. 2010. Distributional memory: A general framework for corpus-based semantics. *Computational Linguistics*, 36(4):673–721.

Marco Baroni and Roberto Zamparelli. 2010. Nouns are vectors, adjectives are matrices: representing adjective-noun constructions in semantic space. In *Proceedings of EMNLP 2010*, pages 1183–1193, Stroudsburg, PA, USA.

Olivier Bodenreider, Anita Burgun, and Thomas Rindflesch. 2001. Lexically-suggested hyponymic relations among medical terms and their representation in the umls. In *Proceedings of TIA 2001*, pages 11–21, Nancy, France.

Didier Bourigault and Monique Slodzian. 1999. Pour une terminologie textuelle. *Terminologies Nouvelles*, 19:29–32.

Bartosz Broda, Maciej Piasecki, and Stan Szpakowicz. 2009. Rank-based transformation in measuring semantic relatedness. In *Canadian Conference on AI*, volume 5549, pages 187–190.

Sharon A. Caraballo. 1999. Automatic construction of a hypernym-labeled noun hierarchy from text. In *ACL*, pages 120–126.

Mehdi Embarek and Olivier Ferret. 2008. Learning patterns for building resources about semantic relations in the medical domain. In *Proceedings of LREC 2008*, Marrakech, Morocco. ELRA.

Cécile Fabre and Didier Bourigault. 2006. Extraction de relations sémantiques entre noms et verbes au-delà des liens morphologiques. In *TALN 2006*, pages 121–129, Leuven.

Olivier Ferret. 2013. Sélection non supervisée de relations sémantiques pour améliorer un thésaurus distributionnel. In *TALN 2013*, pages 48–61, Les Sables d'Olonne, France.

Natalia Grabar and Pierre Zweigenbaum. 2000. Automatic acquisition of domain-specific morphological resources from thesauri. In *RIAO*, pages 765–784.

Edward Grefenstette and Mehrnoosh Sadrzadeh. 2011. Experimental support for a categorical compositional distributional model of meaning. In *Proceedings of EMNLP 2011*, pages 1394–1404.

Edward Grefenstette, Georgiana Dinu, Yao-Zhong Zhang, Mehrnoosh Sadrzadeh, and Marco Baroni. 2013. Multi-step regression learning for compositional distributional semantics. *Proceedings of IWCS 2013*.

Gregory Grefenstette. 1994. Corpus-derived first, second and third-order word affinities. In *Sixth Euralex International Congress*, pages 279–290.

Benoit Habert and Pierre Zweigenbaum, 2002. *Contextual Acquisition of Information Categories: what has been done and what can be done automatically?*, pages 203–231. Nevin (Bruce), Amsterdam.

Thierry Hamon, Adeline Nazarenko, Thierry Poibeau, Sophie Aubin, and Julien Derivière. 2007. A robust linguistic platform for efficient and domain specific web content analysis. In *Proceedings of RIAO*, Pittsburgh, USA.

Zellig Harris. 1954. Distributional structure. *Word*, 10(23):146–162.

Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *International Conference on Computational Linguistics*, pages 539–545, Nantes, France.

Gérard Henneron, Rosalba Palermiti, and Yolla Politi. 2005. *L'organisation des connaissances, approches conceptuelles*. Harmattan.

Karl Moritz Hermann, Edward Grefenstette, and Phil Blunsom. 2013. "not not bad" is not "bad": A distributional account of negation. In *Proceedings of the Workshop on Continuous Vector Space Models and their Compositionality*, pages 74–82, Sofia, Bulgaria.

Christian Jacquemin. 1996. A symbolic and surgical acquisition of terms through variation. In *CoRR*, pages 425–438.

Christian Jacquemin. 1997. Variation terminologique : Reconnaissance et acquisition automatique de termes et de leurs variantes en corpus. Mémoire d'HDR en informatique, Université de Nantes.

Christian Jacquemin. 1999. Syntagmatic and paradigmatic representations of term variation. In *Proceedings of ACL 1999*, pages 341–348, University of Maryland.

Jun'ichi Kazama, Stijn De Saeger, Kow Kuroda, Masaki Murata, and Kentaro Torisawa. 2010. A bayesian method for robust estimation of distributional similarities. In *In proceedings of ACL 2010*, pages 247–256, Stroudsburg, PA, USA.

Claudia Leacock and Martin Chodorow. 1998. Combining local context and wordnet similarity for word sense identification. In *Proceedings of MIT Press*, pages 265–283, Cambridge, Massachusetts.

Jeff Mitchell and Mirella Lapata. 2008. Vector-based models of semantic composition. In *ACL-08: HLT*, pages 236–244, Columbus, Ohio.

Emmanuel Morin and Christian Jacquemin. 2004. Automatic Acquisition and Expansion of Hypernym Links. *Computers and the Humanities*, 38(4):363–396.

Emmanuel Morin. 1999. *Extraction de liens sémantiques entre termes à partir de corpus de textes techniques*. Thèse de doctorat, Institut de recherche en informatique de Nantes.

François Morlane-Hondère and Cécile Fabre. 2012. Étude des manifestations de la relation de méronymie dans une ressource distributionnelle. In *TALN'2012*, pages 169–182, Grenoble, France.

Sebastian Padó and Mirella Lapata. 2007. Dependency-based construction of semantic space models. *Comput. Linguist.*, 33(2):161–199.

Reinhard Rapp. 2003. Word sense discovery based on sense descriptor dissimilarity. In *Proceedings of MT Summit'2003*, pages 315–322.

Philip Resnik. 1995. Using information content to evaluate semantic similarity in a taxonomy. In *Proceedings of IJCAI'1995*, pages 448–453, San Francisco, CA, USA.

Magnus Sahlgren. 2006. *The Word-Space Model: Using Distributional Analysis to Represent Syntagmatic and Paradigmatic Relations between Words in High-Dimensional Vector Spaces*. Ph.D. thesis, Stockholm University, Stockholm, Sweden.

Helmut Schmid. 1994. Probabilistic part-of-speech tagging using decision trees. In *New Methods in Language Processing*, pages 44–49, Manchester, UK.

Rion Snow, Daniel Jurafsky, and Andrew Y. Ng. 2005. Learning syntactic patterns for automatic hypernym discovery. In *Proceedings of NIPS 17*, pages 1297–1304.

Peter D. Turney and Patrick Pantel. 2010. From frequency to meaning: Vector space models of semantics. *Journal of artificial intelligence research*, 37:141–188.

Julie Weeds, David Weir, and Diana McCarthy. 2004. Characterising measures of lexical distributional similarity. In *Proceedings of COLING'2004*, Stroudsburg, PA, USA.

Maayan Zhitomirsky-Geffet and Ido Dagan. 2009. Bootstrapping distributional feature vector quality. *Computational Linguistics*, 35(3):435–461.