Evaluation of Automatic Hypernym Extraction from Technical Corpora in English and Dutch

Els Lefever, Marjan Van de Kauter and Véronique Hoste
LT[3], Language and Translation Technology Team
Ghent University
Groot-Brittanniëlaan 45, 9000 Ghent, Belgium
els.lefever, marjan.vandekauter, veronique.hoste@ugent.be

Abstract

In this research, we evaluate different approaches for the automatic extraction of hypernym relations from English and Dutch technical text. The detected hypernym relations should enable us to semantically structure automatically obtained term lists from domain- and user-specific data. We investigated three different hypernymy extraction approaches for Dutch and English: a lexico-syntactic pattern-based approach, a distributional model and a morpho-syntactic method. To test the performance of the different approaches on domain-specific data, we collected and manually annotated English and Dutch data from two technical domains, viz. the dredging and financial domain. The experimental results show that especially the morpho-syntactic approach obtains good results for automatic hypernym extraction from technical and domain-specific texts.

Keywords: terminology, semantic relations, hypernym extraction

1. Introduction

Due to globalization and the digital information explosion, it is no longer feasible for companies to manually create and manage their terminology and ontology bases. Existing mono- and multilingual databases such as WordNet (Miller et al., 1990) and EuroWordNet (Vossen, 1998) are freely available, but hardly contain any domain- or user-specific terminology. Other databases, such as EuroTermBank[1] and IATE[2], contain more technical terms, but lack user- and company-specific terminology and proper names. Therefore, researchers have started to investigate how terminological and semantically structured resources such as ontologies can be automatically extracted from text (Biemann, 2005).

In this research, we focus on the automatic extraction of hypernym relations from English and Dutch technical text, which should enable us to semantically structure automatically obtained term lists from domain- and user-specific texts. A hypernym relation can be described as a set – subset relation between two terms, and can be defined as follows: a lexical item L_0 is a *hyponym* of a lexical item L_1 if L_0 is a *kind of* L_1. L_1 is then the *hypernym* of L_0 and the relationship can be defined as reflexive and transitive, but not symmetric (Hearst, 1992).

Different approaches have been proposed for automatic hypernym detection in text. A first approach, inspired by the seminal work of Hearst (1992), uses a list of lexico-syntactic patterns to identify hyponymy relations. This approach has been further developed for English (Pantel and Ravichandran, 2004; Oakes, 2005; Pantel and Pennacchiotti, 2006) as well as for other languages such as French (Malaisé et al., 2004), Romanian (Mititelu, 2008) and Dutch (Tjong Kim Sang and Hofmann, 2007). Statistical and machine learning techniques have been used to automatically extend the list of lexico-syntactic patterns and to train hypernym classifiers (Snow et al., 2004; Ritter et al., 2009). A more flexible approach is presented by Navigli and Velardi (2010), who use word class lattices, or directed acyclic graphs, to develop a pattern generalization algorithm that is trained on a manually annotated training set, and is able to extract definitions and hypernyms from web documents.

Other researchers have applied a distributional approach to find hypernym pairs in text (Caraballo, 1999; Van der Plas and Bouma, 2005; Lenci and Benotto, 2012). These approaches start from the assumption that semantically related words tend to occur in similar lexical (or syntactical) contexts. Based on the distributional information, semantically similar words are clustered together. The hierarchical structure between the clusters then expresses the hypernym-hyponym relation. Caraballo (1999), for instance, uses bottom-up clustering techniques to cluster similar nouns and combines them by giving them a common parent. This results in a hierarchy of nouns and their hypernyms. Schropp et al. (2013) filter the output of their pattern-based approach for Dutch by applying a distributional model; only pairs of terms that occur in identical distributional clusters are kept as valid hypernym-hyponym pairs. Lenci and Benotto (2012) use directional (or asymmetric) similarity measures that rely on the distributional inclusion hypothesis to identify hypernyms.

Although these methods obtain a higher coverage for hypernym detection, they suffer from lower precision scores as they have problems to determine the exact nature of the semantic relationship between the terms (synonyms, antonyms, hypernyms, etc.).

Finally, researchers have also applied a morphological analysis of compounds that allows them to extend the list of hypernyms by considering the longest known suffix of the term as a valid hypernym of the compound term (Bosma and Vossen, 2010; Tjong Kim Sang et al., 2011).

In this research, we investigated three different hypernymy extraction approaches for Dutch and English, viz. a pattern-based, distributional and morpho-syntactic method, and tested their performance on domain-specific data. We col-

[1] http://www.eurotermbank.com/
[2] iate.europa.eu/

lected and manually annotated English and Dutch data from two technical domains, viz. the dredging and financial domain. We will discuss the construction of our gold standard dataset in Section 2, describe our hypernym detection system in Section 3 and present the results of the evaluation in Section 4. Section 5 concludes the paper with some prospects for future research.

2. Data annotation

The performance of hypernym extraction systems is usually measured on the basis of the hypernym relations enclosed in WordNet or EuroWordNet. Technical texts, however, typically contain a wide variety of specialized terms that do not occur in general-purpose inventories. Using WordNet as a gold standard for these types of data necessarily results in very low precision and recall figures, that are not really representative of the quality of the automatic hypernym extraction. Therefore, we decided to create a dedicated gold standard for our technical data.

We collected highly specialized data for two different domains, viz. the dredging and financial domain, and two languages, being Dutch and English. The dredging dataset consists of year reports obtained from a Belgian dredging company. For the financial domain, we collected news articles from the newspapers De Tijd and The Financial Times. The entire corpus was first automatically preprocessed. We performed sentence splitting, tokenization, part-of-speech tagging and lemmatization by means of the LeTs Preprocess Toolkit (Van de Kauter et al., 2013). Subsequently, a gold standard for the evaluation of automatic hypernym extraction was manually created by linguistic annotators through the following steps:

1. manual identification of domain-specific terms
2. manual identification of named entities
3. manual identification of hypernym relations between the identified terms and/or named entities
4. manual identification of synonym relations between the identified terms and/or named entities[3]

The manual annotation process was performed using the brat rapid annotation tool (Stenetorp et al., 2012), which allows for the marking of entities as well as relations between these entities, as exemplified in Figure 1. It is important to note that the annotators could identify hypernym and synonym relations between terms occurring within the same sentence, as well as between terms occurring in more distant sentences in the text.

The manual annotation effort resulted in term, named entity, hypernym and synonym lists, jn which each lexical item is enriched with its PoS tag and lemma as produced by the LeTs Preprocess Toolkit. Subsequently, we extended the lists of annotated hypernym relations by means of the following transitivity rules:

- If L_0 is a hyponym of L_1 and L_1 is a hyponym of L_2, then L_0 is a hyponym of L_2
- If L_0 is a hyponym of L_1 and L_1 is a synonym of L_2, then L_0 is a hyponym of L_2
- If L_0 is a synonym of L_1 and L_1 is a hyponym of L_2, then L_0 is a hyponym of L_2

Table 1 gives an overview of the size of the resulting gold standard datasets (number of tokens) and the amount of terms, named entities, hypernyms and synonyms identified. The final five columns represent the number of unique lemmas/relations between unique lemmas.

3. Hypernym Extractor

For the automatic detection of hypernym relations , we applied three different approaches: a lexico-syntactic pattern-based, a distributional and a morpho-syntactic approach. For each of these approaches, we took the gold standard term and named entity lists as a starting point, since we want to detect domain-specific hypernym relations. In future work, we will add a terminology extraction module and named entity recognizer to the system and start from the automatically generated domain-specific terms and named entities for the hypernym extraction.

3.1. Pattern-based model

Our pattern-based approach is inspired by the lexico-syntactic patterns defined by Hearst (1992) and Mititelu (2008). We extended these patterns and translated them into their Dutch equivalents. In case such equivalents did not logically exist in Dutch (e.g. *not least* as in *countries , not least Germany*), we looked for a similar existing pattern instead. This resulted in a list of regular expressions that match both on part-of-speech tags (e.g. *noun*) as well as on complete chunk tags (e.g. *noun phrase*). The regular expressions were fine-tuned by means of the SoNaR corpus (Oostdijk et al., 2012) for Dutch, and the BNC corpus[4] for English. Take the following regular expressions:

1. NP (zo|even)als NP {, NP}* {(en|of) NP}
2. NP (like|such as) NP {, NP}* {(and|or) NP}

NP is shorthand for at least one noun (PoS tag *noun*), but can also be a complex noun phrase (compound noun, noun preceded by a determiner and an adjective, etc.). Examples of extracted hypernym pairs that were matched by this regular expression are listed in example (1) for Dutch and example (2) for English.

1. Dutch examples for NP (zo|even)als NP {, NP}* {(en|of) NP}

(1) *matched text*: ['grondstoffen', 'zoals', 'steenkool', ',', 'ijzererts']
hypernym pairs:
(grondstoffen, steenkool)
(grondstoffen, ijzererts)

(2) *matched text*: ['functie', ',', 'zoals', 'project', 'manager']
hypernym pairs:
(functie, project manager)

[3]In this paper, we focus on automatic hypernym detection, not on synonymy.

[4]http://www.natcorp.ox.ac.uk/

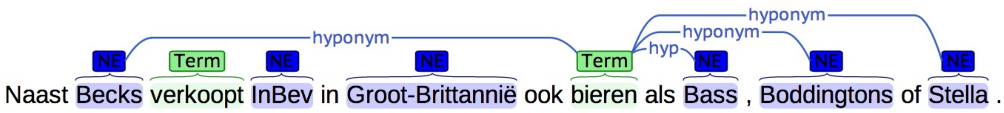

Figure 1: Annotation process in brat

language	domain	tokens	terms	named entities	synonym relations	hypernym relations	extended hypernym relations
English	financial	5108	630	127	14	357	418
	dredging	9761	631	252	14	427	554
Dutch	financial	5053	407	169	29	282	344
	dredging	8893	645	250	27	343	692

Table 1: Properties of the gold standard datasets for hypernym detection

2. English examples for NP (like|such as) NP {, NP}* {(and|or) NP}

(3) *matched text*: ['primary', 'raw', 'materials', 'such', 'as', 'coal', ',', 'iron', 'ore']
hypernym pairs:
(primary raw materials, coal)
(primary raw materials, iron ore)

(4) *matched text*: ['major', 'realisations', 'such', 'as', 'the', 'concert', 'hall']
hypernym pairs:
(major realisations, concert hall)

All hypernym relations linking lexical items that were not in the gold standard term or named entity lists were filtered out. Because of the strict syntactic constraints imposed by the predefined patterns and the fact that hypernym relations can be expressed in multiple different ways in natural language, the coverage of this approach is assumed to be rather low (Navigli and Velardi, 2010).

3.2. Distributional model

Therefore, as a second approach to detect hypernym tuples, we created a distributional semantic model for English and Dutch, in which no restrictions are defined a priori on the syntactic patterns of the hypernym relationship. For the creation of the distributional models, the following steps were applied: (1) we first built a *word-context matrix* for all words occurring in a reference corpus and then converted this matrix into *context vectors* and (2) we clustered these context vectors by means of an agglomerative clustering technique. The resulting clusters thus contain words occurring in similar lexical contexts and are supposed to be semantically related according to the distributional hypothesis (Harris, 1954).

For Dutch, we constructed a semantic distributional model for part of the SoNaR corpus (about 327 million tokens), whereas for English, a model was built using the British National Corpus (about 100 million tokens). Both corpora were tokenized using the LeTs Preprocess Toolkit. In order to build the models for our Dutch and English reference corpus, we first constructed a *word-context frequency matrix* storing how many times each word in the reference corpus occurs in a certain context. To define the context, we used co-occurring words within a window of 5 words. In a second step, we applied Pointwise Mutual Information (Church and Hanks, 1990) as a weighting function to discover informative semantic similarity relations between words. As we only want to consider contexts with a high semantic discrimination value, we smoothened the matrix by removing stop words and low frequent words (occurring less than 3 times in the corpus) from the context features. Each row in the resulting matrix is considered a co-occurrence vector for the word associated with that row. Subsequently, we represent the contexts to be clustered using *second order context vectors* (Schütze, 1998). This is done by replacing each word in the context with its associated vector, and then averaging together all these word vectors, which results in a single vector representing the overall context. The matrix and vector construction was performed with the SenseClusters Package (Pedersen and Purandare, 2004).

Finally, we used the CLUTO clustering toolkit (Karypis, 2002) to group semantically related words into clusters. Similarity between the context vectors was computed by taking their cosine, the cosine of the angle between two vectors being the inner product of the vectors. We used a K-means clustering algorithm and ran experiments with a varying number of output clusters[5].

The input for the distributional model is a list containing all possible combinations of the domain-specific terms and named entities annotated in the gold standard dataset as described in Section 2. A lookup is performed for both terms/named entities in the candidate hypernym pair, and only in case both terms/named entities appear in the same cluster, the pair is considered a valid hypernym pair. Two further remarks have to be made concerning the implementation of the distributional approach. To start with, we performed a lowercased lookup of the terms as they occur in the running text (full forms) in the distributional model, but we output the lemmatized version of the terms in the resulting hypernym tuples. Second, in case of multiword terms in the input, we only considered the last word of the term for lookup in the distributional model, which was built on

[5]When applying an agglomerative clustering technique, the number of desired output clusters has to be predefined.

the basis of co-occurrence information for isolated words. Table 2 and Table 3 list the number of hypernym pairs generated by the distributional models for the Dutch and English dredging and financial data. It is clear from the number of generated hypernym pairs that considerably more terms are found by the English distributional model, although it was trained on less data than the Dutch one. We assume that this is due to the different compounding strategy in both languages; in Dutch, different compound parts are glued together in one orthographic unit, whereas in English, they are separated by spaces. As a result, the last word of a compound term in English is more frequently retrieved by the single word lookup of the distributional model than a complete compound term in Dutch. This observation will also have consequences for the recall figures for both languages, as will be discussed in Section 4.

Dutch		
number of output clusters	nr of hypernym pairs	
	dredging	financial
200	5866	4672
300	4812	3899
500	4688	3912
750	4464	3713
1000	3293	2557

Table 2: Number of hypernym pairs generated by the Dutch distributional model for the dredging and financial data

English		
number of output clusters	nr of hypernym pairs	
	dredging	financial
200	27093	37225
300	24117	32378
500	19420	21547
750	11018	17683
1000	8371	10203

Table 3: Number of hypernym pairs generated by the English distributional model for the dredging and financial data

3.3. Morpho-syntactic model

As stated by Bosma (2010), a substantial amount of hypernym relations in Dutch and English can be identified by determining whether a term is part of a second (multi-word or compound) term. The inspiration for this approach is the head-modifier principle. Sparck Jones (1985) already pointed out that in a compound noun, the linear arrangement of the compound parts expresses the kind of information being conveyed. The head then expresses the more general semantic category, while the modifiers restrict the sense of the compound term. The complete compound can then be considered as a hyponym of the head term, which constitutes the hypernym in these cases, as in (footstep, step). Three different rules were implemented that, for each of the domains and languages, compared each term in the gold standard term list to every other item in the term list and, additionally, the named entity list.

Single-word noun phrase If lexical item L_0 is a suffix string of lexical item L_1, we consider L_0 to be a hypernym of L_1. Examples of hypernym pairs identified by means of this approach are *biervolume – volume* (beer volume – volume) for Dutch, and *oilfield – field* for English.
An additional restriction for this rule is that the remaining part of L_1, after stripping off L_0, should contain at least three characters. This way, we filter out pairs such as *soil – oil*. This problem could be tackled in future research by adding a dedicated decompounding module to the system.

Multi-word noun phrases If lexical item L_0 is the head term of lexical item L_1, we consider L_0 to be a hypernym of $L_1$1. Both in English and Dutch, the head of a nominal phrase appears at the right edge, so we always consider the last constituent to be the head of the compound, and by consequence, to be the hypernym of the complete term. Examples of hypernym pairs retrieved with this rule are *offshore pijpleiding – pijpleiding* (offshore pipeline – pipeline) for Dutch and *maintenance contract – contract* for English.

Noun phrase + prepositional phrase If lexical item L_0 is the first part of a term L_1 containing a noun phrase + preposition (English: of, for, in, before, from, to, on and Dutch: van, op, in, uit) + noun phrase, we consider L_0 to be a hypernym of L_1. In case of a prepositional compound phrase, the head is situated at the left edge of the compound term. Examples of such hypernym pairs are *saneren van verontreinigde bodems – saneren* (remediation of contaminated soils – remediation) for Dutch and *obstruction of justice – obstruction* for English.

4. Evaluation

We calculated precision and recall scores by comparing the output of each method to the manually annotated as well as the extended gold standard hypernym lists. The scores obtained by comparing the output to the extended lists can be found between brackets. Since the extended hypernym lists contain a larger number of hypernym relations, evaluating the approaches of the hypernym detection methods on these lists generally results in lower recall scores. Note that we only took into account relations between unique lemmas.

4.1. Pattern-based approach

As is shown in Table 4, the pattern-based approach obtains very high precision scores. Incorrectly extracted hypernym pairs are mainly the result of general syntactic patterns that overgenerate. As an example, we can cite the sentence fragment *a leap in long-term bonus as Vodafone exceeds targets*, where the pattern [NP (such) as NP] fires and wrongly results in the hypernym pair (Vodafone, long-term bonus).
On the other hand, as already mentioned in Section 3.1. and illustrated for our approach in Table 4, pattern-based approaches often obtain low recall figures since the hypernym pairs have to match very strict syntactic constraints. An additional consequence is that hypernym relations can only be found for terms occurring close to one another and never beyond the sentence-level. In addition, incorrect part-of-speech tags can also cause a mismatch between hypernym patterns and the input text. An example from our data

set is the noun phrase *maintenance dredging*, which was not matched because *dredging* was sometimes inaccurately tagged as a verb instead of a noun. To improve these low recall figures, researchers have tried to add additional patterns by using various bootstrapping strategies (Reiplinger et al., 2012; Pantel and Pennacchiotti, 2006). Adding additional patterns improves recall but does not solve the recall issue, as the patterns stay very strictly and locally defined.

Dutch		
	dredging	financial
precision	90% (97%)	78.13% (78.13%)
recall	7.87% (4.2%)	8.87% (7.27%)
English		
	dredging	financial
precision	72.73% (72.73%)	60% (70%)
recall	1.87% (1.44%)	1.68% (1.67%)

Table 4: Dutch and English precision and recall figures for the pattern-based approach

4.2. Distributional approach

Tables 5 and 6 give an overview of the precision and recall scores for Dutch and English, calculated on both the manually annotated as well as the extended (figures between brackets) gold standards for a varying number of output clusters. The results show good recall scores for English and more moderate recall scores for Dutch. As already introduced in Section 3.2., the different compounding strategy in both languages results in a lower coverage for the Dutch distributional model. In addition, the results clearly show very low precision scores for the distributional approach. This is mainly due to the fact that the model succeeds in detecting semantically similar words, but is not able to make a distinction between different types of semantic relations (e.g. synonymy, antonymy, meronymy, hypernymy, etc.). In future work, we will consider using a method capable of finding hierarchical relationships between these terms, as is done for instance by Caraballo (1999), who use hierarchical clustering to decide which of the two terms has a broader meaning.

The recall figures of the distributional approach could be improved by extending the distributional model itself. A shallow error analysis has revealed two shortcomings of the current semantic model, being (1) words that are semantically related but that appear in different clusters (e.g. *ship* and *vessel*, *ship* and *tanker*) and (2) words that are simply missing in the distributional model, due to the technical nature of the vocabulary (e.g. *replenishment, leveling, causeways*). In order to extend and improve the quality of the semantic model, domain-specific data could be added.

4.3. Morpho-syntactic approach

The high recall scores obtained by the morpho-syntactic approach (as seen in Table 7) show that the technical datasets used for the evaluation contain a large amount of compound and multi-word terms.

When looking at the errors made by the morpho-syntactic approach, we noticed that a high number of hypernym pairs

Dutch		
	dredging	financial
precision	39.20% (42.52%)	66.48% (67.58%)
recall	34.40% (18.50%)	42.91% (35.76%)
English		
	dredging	financial
precision	68.47% (72.44%)	66.48% (67.58%)
recall	56.44% (46.03%)	65.55% (61.24%)

Table 7: Dutch and English precision and recall figures for the morpho-syntactic approach

not contained in the gold standard hypernym lists are actually valid hypernym pairs. We therefore manually validated the output of the morpho-syntactic approach and calculated the scores again. This resulted in precision scores between 85% and 92%, and recall scores between 51% and 69%. Particularly for Dutch, the precision obtained was substantially higher. This can be explained by the fact that the number of single-word compound terms in Dutch is higher than in English, and hypernyms of these compounds are not always situated in the local context of these compounds. As a result, especially terms that occur in very distant sentences in the gold standard corpus are not always identified as hypernym pairs by the linguistic annotators. Another problem encountered by the morphological approach is that suffix strings of terms are not always hypernyms of that term. The term *port* for example is not a hypernym of *transport*. This is a problem that could be solved in future work by means of decompounding. However, even a decompounding module is not sufficient to determine that *bank* is not a hypernym of *rechtbank*.

5. Conclusion

In this paper, we presented three automatic hypernym detection approaches that were developed on the basis of newspaper data and that do not require any external knowledge at run time: a lexico-syntactic pattern-based, a distributional and a morpho-syntactic approach. In order to evaluate the different hypernym detection models, we collected and manually annotated Dutch and English data from two technical domains, viz. the automotive and financial domain.

The experimental results show that the pattern-based and especially the morpho-syntactic approach achieve good performance on the technical domain data, demonstrating that these general purpose hypernym detection modules are portable to other domain- and user-specific data.

In future research, we will investigate how we can further improve the performance of the presented hypernym detection modules and evaluate additional hypernym detection approaches. One possible extension of the presented approaches could be a multilingual model, which incorporates information from other languages to detect hypernym relations in parallel data. We will also build a meta learning system that combines the different hypernym detection modules and uses voting to extract hypernym tuples from text corpora.

Dutch				
number of output clusters	**dredging domain**		**financial domain**	
	precision	recall	precision	recall
200	1.18% (1.49%)	18.66% (11.71%)	2.18% (2.23%)	33.33% (27.91%)
300	1.27% (1.56%)	16.33% (9.97%)	2.32% (2.35%)	29.43% (24.42%)
500	1.47% (1.86%)	18.66% (11.71%)	2.43% (2.48%)	30.85% (25.87%)
750	1.52% (1.86%)	18.37% (11.13%)	2.56% (2.62%)	30.85% (25.87%)
1000	1.98% (2.24%)	17.78% (9.97%)	3.66% (3.71%)	30.85% (25.58%)

Table 5: Dutch precision and recall figures obtained by the distributional method for a varying number of output clusters for the dredging and financial domain

English				
number of output clusters	**dredging domain**		**financial domain**	
	precision	recall	precision	recall
200	1% (1.15%)	53.63% (47.83%)	0.71% (0.80%)	63.59% (61.24%)
300	1.12% (1.29%)	53.63% (47.83%)	0.82% (0.93%)	63.59% (61.24%)
500	1.34% (1.54%)	53.63% (47.47%)	1.25% (0.14%)	63.31% (61.00%)
750	2.32% (2.52%)	52.46% (43.86%)	1.53% (0.17%)	63.31% (61.00%)
1000	3.04% (3.3%)	51.99% (43.5%)	2.58% (0.29%)	62.18% (60.05%)

Table 6: English precision and recall scores obtained by the distributional method for a varying number of output clusters for the dredging and financial domain

6. References

C. Biemann. 2005. Ontology Learning from Text: A Survey of Methods. *LDV Forum*, 20(2):75–93.

W.E. Bosma and P. Vossen. 2010. Bootstrapping language neutral term extraction. In *Proceedings of the 7th international conference on Language Resources and Evaluation (LREC2010)*.

S. Caraballo. 1999. Automatic acquisition of a hypernym-labeled noun hierarchy from text. In *Proceedings of ACL-99*, pages 120–126, Baltimore, MD.

K. Church and P. Hanks. 1990. Word Association Norms, Mutual Information, and Lexicography. *Computational Linguistics*, 16(1):22–29.

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

M. Hearst. 1992. Automatic Acquisition of Hyponyms from Large Text Corpora. In *Proceedings of the International Conference on Computational Linguistics*, pages 539–545.

G. Karypis. 2002. CLUTO - a clustering toolkit. Technical Report 02-017, University of Minnesota, Department of Computer Science.

A. Lenci and G. Benotto. 2012. Identifying hypernyms in distributional semantic spaces. In *Proceedings of the first Joint conference on Lexical and Computational Semantics (*SEM)*, pages 75–79, Montréal, Canada.

V. Malaisé, P. Zweigenbaum, and B. Bachimont. 2004. Detecting Semantic Relations between Terms in Definitions. In *the CompuTerm workshop 2004: 3rd International Workshop on Computational Terminology*, pages 55–62.

G.A. Miller, R. Beckwith, C. Fellbaum, D. Gross, and K.J. Miller. 1990. Introduction to wordnet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–244.

V. Mititelu. 2008. Hyponymy Patterns. Semi-automatic Extraction, Evaluation and Inter-lingual Comparison. *Text, Speech and Dialogue: Lecture Notes in Computer Science*, 5246:37–44.

R. Navigli and P. Velardi. 2010. Learning Word-Class Lattices for Definition and Hypernym Extraction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pages 1318–1327, Uppsala, Sweden.

M. Oakes. 2005. Using hearst's rules for the automatic acquisition of hyponyms for mining a pharmaceutical corpus. In *Proceedings of the Workshop Text Mining Research*.

N. Oostdijk, M. Reynaert, and I. Schuurman. 2012. The construction of a 500-million-word reference corpus of contemporary written Dutch. In P. Spyns and J. Odijk, editors, *Essential Speech and Language Technology for Dutch, Theory and Applications of Natural Language Processing*. Springer.

P. Pantel and M. Pennacchiotti. 2006. Espresso: Leveraging generic patterns for automatically harvesting semantinc relations. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual meeting of the Association for Computational Linguistics*, pages 113–120.

P. Pantel and D. Ravichandran. 2004. Automatically labeling semantic classes. In *Proceedings of HLT/NAACL-04*, pages 321–328, Boston, MA.

T. Pedersen and A. Purandare. 2004. SenseClusters - Finding Clusters that Represent Word Senses. In *Proceedings of Fifth Annual Meeting of the North American Chapter of the Association for Computational Linguistics (NAACL-04)*, pages 26–29, Boston, .A.

M. Reiplinger, U. Schäfer, and M. Wolska. 2012. Extracting glossary sentences from scholarly articles: A comparative evaluation of pattern bootstrapping and deep analysis. In *Proceedings of the ACL-2012 Special Workshop on Rediscovering 50 years of Discoveries*, pages 55–65, Jeju Island, Korea.

A. Ritter, S. Soderland, and O. Etzioni. 2009. What is this, anyway: Automatic hypernym discovery. In *Proceedings of Association for Advancement of Artificial Intelligence Spring Symposium on Learning by Reading and Learning to Read*, pages 88–93.

G.Y.R. Schropp, E. Lefever, and V. Hoste. 2013. A combined pattern-based and distributional approach for automatic hypernym detection in Dutch. In *Proceedings of the 9th International Conference on Recent Advances in Natural Language Processing (RANLP 2013)*, pages 593–600, Hissar, Bulgaria.

H. Schütze. 1998. Automatic word sense discrimination. *Computational Linguistics*, 24(1):97–123.

R. Snow, D. Jurafsky, and A.Y. Ng. 2004. Learning syntactic patterns for automatic hypernym discovery. In *Proceedings of Advances in Neural Information Processing Systems*, pages 1297–1304.

K. Sparck Jones. 1985. Compound noun interpretation problems. *Computer Speech Processing*, pages 363–381.

P. Stenetorp, S. Pyysalom, G. Topić, T. Ohta, S. Ananiadou, and J. Tsujii. 2012. brat: a Web-based Tool for NLP-Assisted Text Annotation. In *Proceedings of the Demonstrations Session at EACL 2012*, Avignon, France.

E. Tjong Kim Sang and K. Hofmann. 2007. Automatic extraction of Dutch hypernym-hyponym pairs. In *Proceedings of CLIN-2006*.

E. Tjong Kim Sang, K. Hofmann, and M. De Rijke. 2011. Extraction of hypernymy information from text. In A. Van den Bosch and G. Bouma, editors, *Interactive multi-modal question-answering*, Series: Theory and Applications of Natural Language Processing, pages 223–245. Springer-Verlag Berlin Heidelberg.

M. Van de Kauter, G. Coorman, E. Lefever, B. Desmet, L. Macken, and V. Hoste. 2013. LeTs Preprocess: the Multilingual LT3 Linguistic Preprocessing Toolkit. *Computational Linguistics in the Netherlands Journal*.

L. Van der Plas and G. Bouma. 2005. Automatic acquisition of lexico-semantic knowledge for question answering. In *Proceedings of the IJCNLP Workshop on Ontologies and Lexical Resources*, Jeju Island, Korea.

P. Vossen, editor. 1998. *EuroWordNet: a multilingual* *database with lexical semantic networks*. Kluwer Academic Publishers, Norwell, MA, USA.