A Combined Pattern-based and Distributional Approach for Automatic Hypernym Detection in Dutch

Gwendolijn Schropp
LT3, Language and Translation
Technology Team
University College Ghent
Gent, Belgium
`gwendolijn@gmail.com`

Els Lefever, Véronique Hoste
LT3, Language and Translation
Technology Team
Ghent University
Gent, Belgium
`Els.Lefever@ugent.be`
`Veronique.Hoste@ugent.be`

Abstract

This paper proposes a two-step approach to find hypernym relations between pairs of noun phrases in Dutch text. We first apply a pattern-based approach that combines lexical and shallow syntactic information to extract a list of candidate hypernym pairs from the input text. In a second step, distributional similarity information is used to filter the obtained list of candidate pairs. Evaluation of the system shows encouraging results and reveals that the distributional information particularly helps to improve the precision for context dependent hypernym pairs. The proposed hypernym module is considered an important step in building a semantic structure for automatically extracted terminology. As our approach does not require external lexical resources, it can be applied for any given Dutch input text and is particularly well suited for domain and user specific text.

1 Introduction

Recent work in knowledge-rich NLP tasks such as information retrieval, question answering, textual entailment and sentiment analysis have revealed a need for more structured data, where concepts are stored together with the semantic relationships that exist between these concepts and their corresponding surface forms. Structured lexical-semantic databases such as WordNet (Miller et al., 1990) or EuroWordNet (Vossen, 1998) have been deployed for a wide range of NLP tasks, but suffer from a number of shortcomings. Firstly, these manually crafted resources are very labour-intensive and costly to create. Secondly, existing lexical inventories contain more general vocabulary and have by consequence a low coverage for domain-specific terms. As a consequence,

researchers have started to investigate how semantic resources such as ontologies can be learned from text instead of being created manually. For an overview we refer to (Biemann, 2005).

In this paper, we focus on the detection of hypernym relations between nouns and noun phrases. Automatic extraction of nouns or noun phrases which are semantically related has been successfully achieved in prior research, for example using coordination and co-occurrence information (Oh et al., 2009; Cederberg and Widdows, 2003; Roark and Charniak, 1998; Widdows and Borow, 2002). However, automatically distinguishing exactly which semantic relationship exists between them is not that straightforward. One of these semantic relationships is the hypernym relation which can be seen as a set-subset relation. In the literature the following description is adopted the most: *a(n) NP0 is a (kind of) NP1*; where NP1 is the *hypernym* of NP0 (which is in turn the *hyponym*) and the relationship is reflexive and transitive but not symmetric (Miller et al., 1990; Hearst, 1992). Note the subtle difference with meronymy (Girju et al., 2003), which is the part-whole relationship, and synonymy (Lin et al., 2003), which expresses equality.

Automatic hypernym detection has been explored in multiple ways. A clear distinction can be made between the pattern-based approaches and the statistical approaches. The aim of the present research is to present a hybrid approach in which distributional information acts as a filter on the pattern-based output. Although our current focus is on hypernym detection of noun-noun pairs, the final goal of this research is to use the automatic hypernym detection system to obtain a hierarchically structured term list for any kind of input text. Prior research in hypernym detection suggested the extracted hypernym-hyponym pairs could be used to extend general thesauri like WordNet (Snow et al., 2006; Roark and Charniak,

1998) or EuroWordNet (Van der Plas and Bouma, 2005). Our aim, however, is to make a hypernym detection system that can be used to structure automatically obtained term lists from domain and user specific texts. These texts typically contain a wide variety of technical terms that do not occur in general-purpose inventories like WordNet.

In the following sections, we will first discuss relevant related research in Section 2, describe our hypernym detection system in Section 3 and present our results in Section 4. Section 5 concludes the paper with some prospects for future research.

2 Related Research

Two main approaches are used to learn hypernym relations from text: pattern-based (or rule-based) approaches and distributional approaches.

Most of the pattern-based approaches were inspired by the seminal work of Hearst (1992) in which she identified a set of lexico-syntactic patterns for the identification of hyponymy relations in English text. Subsequently, various researchers continued working with this pattern-based approach for English (Cederberg and Widdows, 2003; Pantel and Ravichandran, 2004; Riloff and Shepherd, 1997; Roark and Charniak, 1998) as well as for other languages such as French (Malaisé et al., 2004) or Romanian (Mititelu, 2008). The patterns were further extended through translation and manually searching through texts (Kozareva et al., 2008), or by using more sophisticated methods of clustering related terms, starting from known hypernym pairs and features (Snow et al., 2006; Lin, 1998) or lists of seed words known to have the desired relationship (Roark and Charniak, 1998; Riloff and Shepherd, 1997; Widdows and Borow, 2002). Pantel and Pennacchiotti (2006) used generic patterns (broad coverage noisy patterns) to extract semantic relations and subsequently apply refining techniques to deal with the wide variety of such relations. Similar approaches that combine pattern extraction with post-processing techniques to enrich the system and improve the results have been investigated, for example, with Support Vector Machines and Hidden Markov Models (Ritter et al., 2009). A different approach has been used by Navigli et al. (2010), that use word class lattices, or directed acyclic graphs, to develop a pattern generalization algorithm that is able to extract definitions and hypernyms from web documents.

For Dutch, several methods have been investigated. Tjong Kim Sang et al. (2011; 2007) have tried to extract hypernymy information from text in three ways: comparing extraction of one pattern from the web with extraction from multiple patterns from a corpus, extraction with and without word sense tagging, and finally they also investigated the impact of using deep syntactic information for hypernym extraction. Bosma et al. (2010; 2011) applied different relation extraction methods in a way that the results of one method are used as input for another method, aiming to find the complete terminology of domain specific texts. In addition to applying a pattern-based and distributional approach, they also perform a morpho-syntactic analysis of compound terms and consider the longest known suffix of the term as a valid hypernym of the compound term. Van der Plas and Bouma (2005) present a searching method for semantically similar words on the basis of a parsed corpus of Dutch text and used these relations to boost the performance of an open-domain question answering system.

Other researchers have applied a distributional approach to automatically extract hypernym pairs from text. The latter approaches start from the distributional hypothesis, stating that words that occur in similar contexts tend to be semantically similar (Harris, 1968). In order to define the *context* of a given target word, both cooccurrence and syntactic information can be extracted from the surrounding words. Unsupervised learning methods like clustering to obtain taxonomies, definitions and semantically similar words have been applied by (Widdows, 2003; Pereira et al., 1993; Van de Cruys, 2010). Clustering has also shown to be a valid approach to automatically detect hypernym relations between terms. By clustering words according to their contexts in text and assigning a label to each cluster, it is then also possible to extract $is - a$ relations between each cluster member and the cluster label. Caraballo (1999) uses syntactic dependency features (such as conjunction and apposition) to automatically build noun clusters. Pantel and Ravichandran (2004) extended his

work by including all syntactic dependency relations for each considered noun.

More recent distributional approaches rely on the *Distributional Inclusion Hypothesis*, according to which semantically narrower terms include a significant number of distributional features of their hypernyms (Lenci and Benotto, 2012).

The main advantage of the distributional approaches is that they allow to find semantically related terms, even when they do not explicitly occur in predefined patterns in text. The main disadvantage, however, is that these clustering approaches have difficulties to determine the exact semantic relationship (synonymy, antonymy, hyponymy) between the semantically related concepts.

In order to improve on precision for the automatic hypernym detection, we decided to combine the lexico-syntactic pattern-based approach with a distributional approach that filters candidate hypernym pairs containing noun pairs that are not semantically related (and that by consequence are not contained by the same sense cluster).

3 Dutch hypernym finder system

3.1 Pattern-based module

For our pattern detection system, we used the patterns from Hearst (1992), complemented with those from Mititelu (2008), and translated them into their Dutch equivalents. This resulted in a list of 42 patterns. If such equivalents did not logically exist in Dutch (e.g. *not least* and *become*), we either left them out or took a similar existing pattern instead. A few examples are the following:

English	Dutch
like	NP, zoals NP {, NP}* {(en|of) NP}
and/or other	NP {, NP} {,} (en|of) andere NP
(e)specially	NP, (voornamelijk|vooral|speciaal) NP {, NP}* {(en|of) NP}
including	NP, inclusief {NP, }* {(en|of) NP}
is a	NP is (een) NP
are	NP {, NP}* {(en|of) NP} zijn NP
for example	NP {,} bijvoorbeeld NP {, NP}* {(en|of) NP}
and/or similar	NP {, NP}* (en|of) soortgelijk(e)|dergelijk(e) NP

Some of the patterns were not as likely to occur in their Dutch translation as they might be in English (e.g. *in common with other*), but we decided to test all the patterns to get an idea which patterns would yield the correct noun pairs and which would more often result in false positives.

3.1.1 Datasets

The corpus used in the experiments is a one-million subcorpus of the 500-million word balanced reference corpus for contemporary (1954-present) Dutch texts: SoNaR (Oostdijk et al., 2012). It consists of 38 text types coming both from Flanders (1/3) and the Netherlands (2/3). The SoNaR-corpus was tokenized, lemmatized, Part-of-Speech-tagged and chunk-tagged using a preprocessing toolkit that was developed in-house (reference omitted). In order to develop and test our pattern detection system, we divided the one-million corpus in two parts: a development set of 250.000 words and a test set of 750.000 words. The development set was used to fine-tune the hypernym patterns and optimize the distributional model (See section 3.2).

3.1.2 Pattern-Based Approach

In order to define the patterns, a set of regular expressions was designed to match on both Part-of-Speech as well as chunk tags. Take for example the pattern *NP zoals NP* ((a(n) NP like NP). This is the simplified version of *NP (zo\|even)als NP {, NP}* {(en\|of) NP}*, in which NP is shorthand for at least one noun (PoS-tag 'N'). NP can also be a compound noun: a noun preceded by either another noun, an adjective (PoS-tag 'ADJ', chunk tag 'I-NP') or a verbal adjective (PoS-tag 'WW', chunk tag 'I-NP'). This allows us for example to capture hyponym/hypernym relations between phrases such as 'automatic gearbox', 'manual gearbox' and 'gearbox'. As patterns were often interrupted by adverbial phrases, we ignored adverbs (PoS-tag 'BW').

The detection system returns both the pattern matches (containing lemmas) as well as the hypernym-hyponym pairs themselves, as exemplified below:

```
['sector', 'als', 'biotechnologie', ',',
'farmacie']¹
(sector, biotechnologie)
(sector, farmacie)
```

3.2 Distributional semantic module

Vector space models (VSMs) have been widely used for semantic processing of text (Turney and Pantel, 2010). These VSMs use statistical patterns of human word usage to build up an artificial

---

¹English: 'sector', 'such as', 'biotechnology', ',', 'pharmacy'

understanding of a given text. In order to post-process the pattern-based hypernym pairs, we created a distributional semantic model for Dutch by applying following steps:

1. build a large *word-context matrix* for all words occurring in a Dutch reference corpus and convert this matrix into *context vectors*

2. *cluster* the resulting context vectors

The resulting clusters contain Dutch words occurring in similar lexical contexts and can by consequence be used to filter hypernym pairs that show little semantic relatedness.

We constructed a semantic model for part of the Twente News Corpus (TwNC), a multifaceted Dutch corpus that contains material from different sources such as national newspapers, television subtitles, broadcast news transcripts, etc. (Ordelman et al., 2007). The corpus was tokenized and contains around twenty million tokens. In order to build a VSM model for our Dutch reference corpus, we first built a **word-context frequency matrix** storing for every word in the Dutch corpus how many times it occurred in a certain context. To define the context, we used cooccurring 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. Finally, the cooccurrence matrix was converted into a vector of context features per target word. The matrix and vector construction was performed with the SenseClusters Package (Pedersen and Purandare, 2004).

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. The impact of the desired number of output clusters is discussed in section 4.

3.3 Filtering module

The filtering module uses distributional evidence to remove candidate hypernym pairs that are not semantically related; nouns that are considered to have a hypernym relationship (resulting from the pattern-based module) and that do not figure in the same semantic cluster (distributional semantic module) are removed from the hypernym pair list. In case one of the nouns does not appear in the clusters at all – because the word occurred less than three times in the reference corpus – we do not filter the given hypernym-hyponym pair. As our clusters are composed of single word terms, we only consider the last word of the hypernym/hyponym in case the pair contains multiword terms[2]. If we take for instance the hyponym *eerstelijns zorgverstrekker* [English: primary care provider], we only consider the last word "zorgverstrekker" [English: care provider] for comparison with the clustering output.

4 Experimental results

4.1 Experimental set-up

To evaluate the performance of both the pattern-based and combined approach, we extracted a test set from the Sonar corpus that contains 750.000 tokens. The output of the system was manually labeled by two annotators using the following labels:

- **strict**: correct hypernym-hyponym pair

- **context-specific**: there is context-specific hypernym relation between both noun phrases.

- **no**: not a correct hypernym pair

We included the context-specific class to cover hypernym relations between automatically extracted terms from domain or user-specific corpora, which is the ultimate goal of our work. Such a class can also cover domain-specific relations including proper names. Theoretically, proper names do not occur in a hypernym relation, since whether or not they would be considered correct is highly dependent on the context of the document: the hypernym pair (priest, John) can be correct in a text where John is in fact a priest, but if another non-priest John is referred to this pair would be

---

[2]In Dutch, the last word is usually the most meaningful part of a given multiword term.

incorrect. There are, however, pairs where the hypernym is more specific, which makes the pair less ambiguous. As an example, we can cite the pair '(queen, Beatrix)' or '(queen, Elisabeth)', which one might consider to be a correct hypernym pair. As many proper nouns occur in domain specific and technical texts, we decided to consider them as potential terms in a hypernym-hyponym relationship. Examples extracted from our corpus are: '(buurland, Nederland)' [English: neighboring country, The Netherlands] and '(concurrent, Inbev)' [English: competitor, Inbev].

Inter-annotator agreement We calculated inter-annotator agreement using Kappa on a subset of the test data containing 1000 hypernym-hyponym pairs (Carletta, 1996). The Kappa statistic was 0.687 for on the strict labeling task and 0.678 on the context-specific hyponyms. In addition, we also calculated inter-annotator agreement by measuring precision, recall and their harmonic mean $F1$ (van Rijsbergen, 1979). F-scores were calculated by taking one annotator as the gold standard and scoring the annotations of the other for precision and recall. This yields the same results as averaging the precision or the recall scores of both annotators, when using the other as a gold standard. A $F1$ score of 89% was obtained on the strict labeling task, whereas a 87% agreement was obtained on the labeling task in which also context-specific hypernyms were indicated.

Evaluation metrics In order to assess the performance of our hypernym extraction module, we calculated **Precision** by dividing the number of correct hypernym pairs by the total number of predicted hypernym pairs:

$$Precision = \frac{strict}{predicted} \qquad (1)$$

We also measured the **Relaxed Precision** (*RelaxedP*) that measures the system performance on the context-specific hypernym relations:

$$RelaxedP = \frac{strict + context\ specific}{predicted} \qquad (2)$$

4.2 Results of the pattern-based module

In the complete corpus, 13 patterns were found. As is shown in Table 1, there is a striking difference between the strict and relaxed precision.

Pattern	# tuples	Relaxed Precision	Precision
als NP zijn NP	1	0	0
NP, zoals NP	874	0.57	0.38
NP, inclusief NP	11	0.45	0.09
NP is (een) (soort (van)) NP	849	0.31	0.11
NP en gelijke / andere NP	8	0.875	0.875
NP, anders dan NP	7	0.57	0.14
NP, d.w.z. NP	2	0.5	0.5
NP, met uitzondering van NP	1	0	0
NP, ofwel NP	6	0.5	0.17
NP, genaamd NP	8	0	0
NP, die (een) NP zijn	32	0.19	0.06
NP, een NP	946	0.36	0.14
NP, maar niet NP	1	1	0

Table 1: Precision and Relaxed precision scores per pattern.

The relaxed scores are comparable to the 40% reported by Cederberg and Widdows (2003) and our *'zoals'*-pattern performs even better than the 52% reported by Hearst (1992) for the English version (*'such as'*). When comparing our results to those obtained for Dutch by Tjong Kim Sang and Hofmann (2009), several things can be noted. They report a 57,5% precision for the pattern *'such as'* on a Wikipedia corpus, whereas it only scored 25,1% on a Newspaper corpus. Our Sonar test corpus consists of both kinds of texts and others still, and also scored 57%. The other patterns we can compare with are *'N be N'*, scoring 22,9%, and *'N be a N'*, scoring 40,8%, which are both contained in our pattern *'NP is (een) (soort (van)) NP'*, scoring 31%.
Tjong Kim Sang et al. (2011) examined the effect of two text preprocessing approaches on the task of extracting hypernymy information, i.e. a pattern-based approach and a dependency parsing approach. Their pattern-based approach scores 43% precision on a newspaper corpus and 63,4% precision on a Wikipedia corpus.

We also calculated recall and precision of our hypernym pairs in comparison with the synsets of the Dutch part of EuroWordNet (EWN). Recall was 0.12 and precision 0.03. The reason for these low scores is mainly a coverage problem of the Dutch EWN. This caused a lot of correct pairs to be found incorrect (nonexistent) in EWN. As an example, the pair '(land, Nederland)' [English: country, The Netherlands] was considered correct since 'Nederland' is part of EWN, whereas the

pair '(land, Rusland)' [English: country, Russia] was considered incorrect due to the fact that 'Rusland' is not incorporated in EWN.

We encountered some issues that are characteristic for a pattern detection system, such as words disturbing the pattern and preventing it from being matched, patterns that overgenerate and do not always indicate a hypernymy relationship (e.g. [NP, a NP]), or mistakes from preprocessing (e.g. nouns being tagged as verbs, or vice versa) yielding incorrect pairs or preventing correct ones from being matched. Furthermore, in running text, semantic relations are often left implicit, while a pattern-based approach can only handle the explicit instances. Were we to test on a text wherein conceptual relationships are explicit, like an encyclopedia, the system would probably perform better.

4.3 Results of the filtering module

The list of hypernym-hyponym pairs that resulted from the pattern-detection module was filtered by means of the distributional semantic module discussed in Section 3.2. By filtering hypernym pairs that do not appear in the same semantic cluster, we expect to partially solve the problem of overgeneration that is caused by very general patterns matching term pairs that are not semantically related.

Figure 1 confirms our hypothesis: although the strict precision is similar between the two methods, the combined system clearly improves the relaxed precision that also considers the context-specific hypernym pairs. The improved relaxed precision can be observed for all tested numbers of output clusters, but as can be expected slightly increases when grouping the nouns into smaller and thus semantically more narrow clusters[3].

Inspection of the results from the combined system revealed a couple of issues. First, the semantic model only covers part of the terms that appear in the hypernym-hyponym pairs. A matched hypernym pair such as for instance '(afvalproduct, stro)' [English: waste product, straw] is not filtered because the nouns are not contained in the semantic model. Second, we observed that semantically related words do not always appear in the same cluster. As a consequence, correct hypernym-hyponym pairs not occurring in the same clus-

---

[3] A larger number of output clusters results into a smaller list of words contained by each cluster, and by consequence tighter semantic relations between these terms.

ters are erroneously eliminated by the filtering module. We detected for instance that the nouns in '(land, Rusland)' [English: country, Russia] are contained by different clusters (and are subsequently filtered by the distributional module), whereas the words in '(land, Nederland)' [English: country, Netherlands] do occur in the same cluster.

A possible explanation for both problems could be the modest size of our reference corpus (20 million words) where low frequent terms were filtered as well. We expect by consequence to solve these issues by using a much bigger reference corpus that allows us to store more contexts and examples for a broader range of words. In addition, we will also perform lemmatization and parsing of the reference corpus, in order to experiment with different kinds of features.

5 Conclusion and Future Research

We presented a first set of experiments for a Dutch hypernym detection system that combines a lexico-syntactic pattern-based and distributional approach. The experimental results show the effectiveness of the filtering step; adding a distributional model clearly improves the relaxed precision of the system.

Analysis of the test results revealed a number of shortcomings of the current approach that will be tackled in future research. Since at one hand the pattern detector purely matches on surface-syntactic forms, and on the other hand these patterns can also occur without actually representing a hypernym relation, we believe that a more flexible and sense-orientated approach is needed to amplify our pattern detector. Further experiments with a larger reference corpus are also needed to improve the semantic model for Dutch. Additional research is also needed to determine the best context representation (lexical or syntactic context, window size of the context) and clustering parameters (desired number of output clusters, clustering algorithm, etc.).

In future research, we will also develop gold standard corpora for different domains and different languages, in order to measure both precision and recall on technical and user specific data.

References

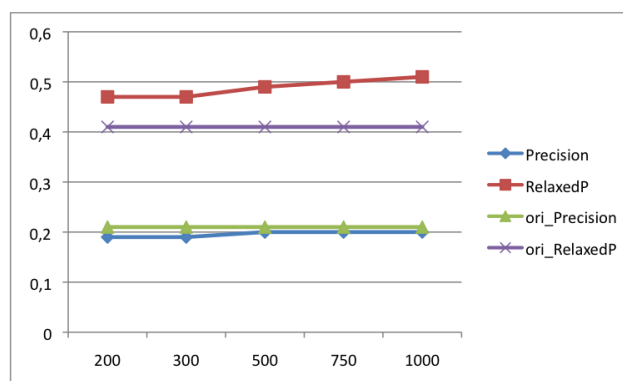Chris Biemann. 2005. Ontology Learning from Text: A Survey of Methods. *LDV Forum*, 20(2):75–93.

Figure 1: Precision and Relaxed precision scores for the pattern-based (*ori_Precision* and *ori_RelaxedP*) and combined module (*Precision/RelaxedP*) with a varying number of output clusters.

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)*, May.

W. E. Bosma, P. Vossen, and H. van der Vliet. 2011. Termextractie in kyoto. In *Terminologie in het Nederlandse Taalgebied*, volume 2010.

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

Jean Carletta. 1996. Assessing Agreement on Classification Tasks: The Kappa Statistic. *Computational Linguistics*, 22(2):249–254.

S. Cederberg and D. Widdows. 2003. Using lsa and noun coordination information to improve the precision and recall of automatic hyponymy extraction. In *Proceedings of the 7th CONLL at HLT-NAACL 2003*, volume 4, pages 111–118.

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

R. Girju, A. Badulescu, and D. Moldovan. 2003. Learning semantic constraints for the automatic discovery of part-whole relations. In *Proceedings of HLT-NAACL*, volume 1, pages 1–8.

Zelig Sabbetai Harris. 1968. *Mathematical structures of language*. Wiley.

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.

Z. Kozareva, E. Riloff, and E. Hovy. 2008. Semantic class learning from the web with hyponym pattern linkage graphs. In *Proceedings of the 46th Annual Meeting of the Association for Computational Linguistics (ACL)*, pages 1048–1056, Columbus, Ohio, USA.

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.

D. Lin, S. Zhao, L. Qin, and M. Zhou. 2003. Identifying synonyms among distributionally similar words. In *Proceedings of the 18th International Joint Conference on Artificial Intelligence*, pages 1492–1493.

D. Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of Coling-ACL*, pages 768–774.

V. Malaisé, P. Zweigenbaum, and B. Bachimont. 2004. Detecting semantic relations between terms in definitions. In *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 online 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.

J. Oh, K. Uchimoto, and K. Torisawa. 2009. Bilingual co-training for monolingual hyponymy-relation acquisition. In *Proceedings of ACL-09: IJCNLP*, pages 432–440.

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.

R. Ordelman, F. de Jong, A. Hessen, and H. Hondorp. 2007. TwNC: a Multifaceted Dutch News Corpus. *ELRA Newsletter*, 12(3-4).

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.

F. Pereira, N. Tishby, and L. Lee. 1993. Distributional clustering of English words. In *Proceedings of the 31st annual meeting on Association for Computational Linguistics*, pages 183–190.

E. Riloff and J. Shepherd. 1997. A corpus-based approach for building semantic lexicons. In *Proceedings of the 2nd Conference on Empirical Methods in NLP*, pages 117–124.

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.

B. Roark and E. Charniak. 1998. Nound-phrase co-occurrence statistics for semiautomatic semantic lexicon construction. In *Proceedings of COLING and ACL*, volume 2, pages 1110–1116.

R. Snow, D. Jurafsky, and A.Y. Ng. 2006. Semantic taxonomy induction from heterogenous evidence. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th Annual meeting of the Association for Computational Linguistics*, pages 801–888.

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

E.F. Tjong Kim Sang and K. Hofmann. 2009. Lexical patterns or dependency patterns: which is better for hypernym extraction? In *Proceedings of the Thirteenth Conference on Computational Natural Language Learning*.

Tjong Kim Sang, E. and Hofmann, K. and De Rijke, M. 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.

P. Turney and P. Pantel. 2010. From frequency to meaning: Vector space models of semantics. *J. Artif. Intell. Res. (JAIR)*, 37:141–188.

Tim Van de Cruys. 2010. *Mining for Meaning. The Extraction of Lexico-Semantic Knowledge from Text*. Ph.D. thesis, University of Groningen, The Netherlands.

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

C.J. van Rijsbergen. 1979. *Information Retrieval*. Buttersworth, London.

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

D. Widdows and B. Borow. 2002. A graph model for unsupervised lexical acquisition. In *Proceedings of the 19th International Conference on Computational Linguistics*, volume 1, pages 1–7.

D. Widdows. 2003. Unsupervised methods for developing taxonomies by combining syntactic and statistical information. In *Proceedings of HLT-NAACL*, pages 197–204.