Finding a Location for a New Word in WordNet

Paula Pääkkö
University of Helsinki
Helsinki, Finland
paula.paakko@helsinki.fi

Krister Lindén
University of Helsinki
Helsinki, Finland
krister.linden@helsinki.fi

Abstract

FinnWordNet is a Finnish wordnet which complies with the structure of the Princeton WordNet (Fellbaum, 1998). It was created by translating all the words in Princeton WordNet. It is open source and contains over 117 000 synsets. We are now testing different methods in order to improve and expand the content of FinnWordNet.

Since wordnets are structured ontologies, a location for a word in FinnWordNet can be pinpointed by its relations to other words. To us, finding a location for a word therefore means finding a hyperonym, a hyponym or a synonym for the word. This article describes some methods for finding a location for a new word in FinnWordNet. Our methods include searching for multiword terms, compounds and lexico-syntactic patterns. Testing shows that with a few simple methods, we were able to find an indicator of the location for 83.2 % of new words. Out of the new synonym pairs we tested, we were able to find an indication for 86.7 %.

1 Introduction

WordNet (Miller et al., 1990; Fellbaum, 1998) is a lexical database for English where words (adjectives, nouns, verbs, adverbs) are grouped into synonym sets, also called synsets. Each synset represents a concept. In addition to synonymy, WordNet also includes other types of semantic relations, for instance antonymy, hyponymy, meronymy and troponymy. The hyperonym/hyponym relation creates a hierarchical structure for nouns. Typically, wordnets are monolingual but lately multilingual wordnets have been under way, see (Vossen, 1998; Tufis et al., 2004; Pianta et al., 2002)

The Finnish WordNet (FinnWordNet or FiWN) is a Finnish version of the Princeton WordNet (Carlson and Lindén, 2010). There are three main approaches to creating wordnets (Carlson and Lindén, 2010): you can build a wordnet from scratch, translate an existing wordnet or translate the top ontology and extend it with local synonym dictionaries. FiWN was created by translating the word senses in Princeton WordNet 3.0 by using professional translators. The translation process was controlled with regard to quality, coverage, cost and speed of translation. FiWN with translations can also be used as a Finnish–English dictionary.

After translation, our next aim has been to improve and to expand the content of FiWN. One important part of this is to check that FiWN contains the most frequently used Finnish terms and concepts. The downside of creating a wordnet by translation is that the content tends to include terms specific to the source language, in this case terms related to English-speaking cultures, while some central concepts in the target language are possibly left missing.

Another goal is to make sure that the semantic relations as well as the translations are correct. The FiWN search interface[1] also includes a feedback and rating possibility. Crowdsourcing is one of the methods we are using to improve and expand FiWN (Lindén et al., 2012). One important method for enriching FiWN is finding new instances of semantic relations and words in corpora. Lindén et al. (2012) have already established some useful lexico-syntactic patterns for Finnish for finding instances of semantic relations, especially synonyms and hyponyms/hyperonyms.

The next stage of expanding the content of

[1] http://www.ling.helsinki.fi/cgi-bin/fiwn/search

FiWN is to add new words and relations to FiWN. To us, finding a location for a word in FiWN means finding hyponyms, hyperonyms or synonyms for that word. Hyponyms, hyperonyms and synonyms act as indicators for the location. In this article, we test a few simple methods for finding such indicators. We use synonym pairs found by Lindén et al. (2012) as our test set.

WordNet (Miller et al., 1990) describes hyponymy as follows: A concept represented by the synset {x, x',...} is said to be a hyponym of the concept represented by the synset {y, y',...} if native speakers of English accept sentences constructed from such frames as *An x is a (kind of) y*. Inversely, synset {y, y',...} is the hyperonym of synset {x, x',...}. Miller et al. (1990) also suggest the following definition for synonymy: two expressions are synonymous in a linguistic context C if the substitution of one for the other in C does not alter the truth value.

The article is divided as follows. Section 2 describes the related work for finding instances of semantic relations for English. Section 3 describes our methods in more detail. Section 4 describes the material and Section 5 outlines the evaluation and test results, whereas Section 6 discusses the results and future work. Finally, Section 7 draws the conclusions.

2 Related Work

There has been much research for finding word pairs which share a certain semantic relation. These approaches can be divided into two main categories: pattern-based and cluster-based approaches. The latter uses statistics and clustering algorithms to cluster similar words according to context. This approach is based on the *Distributional hypothesis* (Harris, 1968) which states that words that occur in a similar context tend to be similar in meaning. Works using this approach include for instance (Caraballo, 1999), (Lin, 1998) and (Pantel and Lin, 2002).

There are studies with this approach for Finnish as well. In his Ph.D. thesis Piitulainen (2011) did a case study on the distributional similarity of words. He studied nearly twenty thousand nouns that occur often in a Finnish newspaper corpus. Lindén and Piitulainen (2004) introduced similarity recalculation after context clustering to find subclusters and they used translation dictionaries for evaluating the rate of synonymy found in word clusters.

Pattern-based approaches use lexico-syntactic patterns in order to find the context of a relation. Hearst (1992; 1998) conducted one of the first studies that used patterns for hyponymy relation. Hearst's patterns included for example:

- *NP such as* {*NP, NP..., (and|or)*} *NP*
  "The bowlute, such as the Bambara ndang"
  hyponym("Bambara ndang", "bowlute")

- *NP* {*, NP*}* {,} *or other NP*
  "Bruises, wounds, broken bones or other injuries"
  hyponym("bruise", "injury"),
  hyponym("wound", "injury"),
  hyponym("broken bone", "injury)

These manually created patterns usually have good precision but low recall (Hearst, 1998). Hearst also introduced her Lexico-Syntactic Pattern Extraction method (LSPE) which many later methods are based on. She also tried this method for meronymy, but the results were not as promising. The patterns for meronymy were not unique enough but also found other instances of semantic relations. Berland and Charniak (1999) on the other hand got better results since they used more refined statistical measures for ranking the output.

Many automatic and semi-automatic approaches to finding patterns have later been proposed. Many of them are based on Hearst (1992). They usually have the same idea: First gather seed instances of the desired relation and find those occurrences in text. From this, the context determines a new pattern which is then used to find new instances. New instances in turn can be used to find new patterns. Methods differ in how new patterns are evaluated and selected.

Instead of manually created patterns, some methods use machine learning to learn new patterns. These approaches include for example (Snow et al., 2005). They used a supervised learning algorithm to create a hyperonym classifier. Other works are (Girju et al., 2003), (Girju et al., 2006) for finding instances of part-whole relation (meronymy/holonymy) and (Agichtein and Gravano, 2000) for instances of any kind of semantic relation.

Different approaches also use general patterns. These broad coverage noisy patterns have high recall but low precision. Works include (Girju et al., 2003), (Girju et al., 2006) and (Pantel and Pennacchiotti, 2006). These patterns produce both right and wrong instances and the methods need to filter out the wrong ones. Many methods use statistical evaluation to estimate the accuracy of generic patterns. Works include for example (Brin, 1999), (Agichtein and Gravano, 2000), (Agichtein et al., 2001), (Ravichandran and Hovy, 2002) and (Pantel et al., 2004)

Statistics-based methods are mainly applicable to medium- or high-frequency words, whereas pattern-based methods are applicable also to low-frequency words, because even one occurence of a pattern is often sufficient. Since most of the remaining words not in FiWN are low-frequency, we will focus on pattern-based methods.

Many have compared their results with wordnet, see e.g. (Hearst, 1992; Lin and Pantel, 2002; Snow et al., 2005). This can be seen as the first step to add words to wordnet because it requires finding a location for an existing word. However, they have not directly addressed the issue of where to add the words that were not in wordnet.

3 Method

The purpose of our study is to find an indication of the location in FiWN for a new word. The indications are hyponyms, hyperonyms and of course, synonyms. Our test set includes synonym pairs for which we try to find a place in FiWN. The synonyms in our material may be single words, compounds or multiword terms. The methods focus on finding locations for nouns or noun phrases.

We have chosen to test and evaluate the following simple methods:

(A) Categorize the word pairs into five groups:
  1) Pairs where both words are in FiWN and:
    (a) in the same synset
    (b) not in the same synset
  2) Pairs with one word already in FiWN and:
    (a) the other word can be added to one of the synsets
    (b) the other word cannot be added to one of the synsets
  3) Pairs where neither of the words are in FiWN

(B) Check compound words.

(C) Check main word of a multiword term.

(D) Use patterns.

In method A, the synonym pairs can be mechanically compared with the synsets in FiWN and catecorized into groups 1, 2 or 3 accordingly. Separating group 2 into groups 2a and 2b needs to be judged by a human.

Group 1a is unproblematic. Group 1b indicates that some meanings are not yet covered in the existing FiWN. As both words in group 1b are already in FiWN, finding a location for the words might seem unnecessary. But now we are actually finding a location for a new meaning of a word. For this group, we can add some assumptions, because we are most likely finding a less frequent or more specialized meaning of one of the words. We assume that either word of the word pair can be added to a synset of the other word or the word can be added as a hyponym of the other word.

Group 2a is straightforward. The synonym pair's meaning is represented by one of the synsets which contains the known word of the word pair. Since different synsets represent different concepts (ergo meanings) it is likely that the new word is added to only one of the synsets. This is reinforced by the fact that most of the high-frequency words are already included in FiWN. This means that if a word is missing from FiWN, it is most likely a rare word, which usually has only one specific meaning.

Group 2b is similar to 1b, but group 2b also implies that the synonym pair's meaning is somehow new. We have two views on this. First, we can assume that the meaning is more precise so that the new word should be added as a hyponym. On the other hand, it is also possible that we have found a new meaning for a word that is already in FiWN.

Method A is used on word pairs whereas methods B-D are used on words. Since our test set consists of synonyms pairs, it is enough that we get an indicator for at least one of the words in a word pair. The other word gets its hyperonym or hyponym via its synonym.

Method B and C rely on a language's innate mechanism of coining new words and terms based on established ones. In Finnish, compound words are written together and the last word is always the main word. This differs from multiword terms, because the main word is not always the last word. For example a *White-tailed Tropicbird* is a kind of bird and a *role-playing game* is a kind of game.

Finding noun compound words is relatively easy in Finnish based on the compound word border. Additionally, proper nouns are excluded from the results to make the results more useful. This leaves out compound person names, for example the surname *Tois#kallio* (*neighbor-hill*). Compound names may reflect concepts, but they are no longer perceived and used that way. In our annotated corpus a hash sign (#) indicates a word boundary in compound words. In addition, we also check if the word includes a dash (-), which is sometimes used as an explicit compound word delimiter, for instance *MIRV-ohjus* (*MIRV missile*).

For the multiword terms we took similar steps as for the compound words, e.g. proper nouns were discarded to exclude person names. Our program chooses the head word of the NP as the hyperonym. If no head word is annotated, then the last noun of the NP is chosen.

Method D uses hand-made patterns. Lindén et al. (2012) established some useful patterns for Finnish which we are using to find a hyperonym or a hyponym for a word in a corpus. These patterns are based on Hearst's patterns (1992; 1998).

4 Material

Lindén et al. (2012) evaluated a few lexico-syntactic patterns. From their subsection of Finnish Wikipedia[2] articles they found 1405 occurrences of pattern *eli* (*a.k.a.*). They manually checked 1100 and evaluated 583 (53.3 %) as useful (that is to say the occurrence of the pattern produced a word pair with a known semantic relation). Relations were categorized as synonyms, translations (which can be seen as kind a of synonymy) and glosses. From this, we got a test set of 459 synonym pairs.

Since the purpose of these tests is to find a hyponym or hyperonym for the words, we only look at noun pairs. This gives us a test set of 422 unique synonym pairs and 594 unique new words (strings)[3].

[2]http://fi.wikipedia.org/wiki/Wikipedia:Etusivu

[3]The word for a special species of dolphin, i.e. *inia*, was left out as it was incorrectly given the baseform *in* and would have affected the search results of method D by retrieving English text fragments.

4.1 Preprocessing

The corpus we used for finding patterns was Finnish Wikipedia[4]. This corpus was cleaned from the Wikipedia tagging in order to get only the text. The size of the cleaned corpus was 379.4 MB. This corpus was then annotated with Connexor's[5] fi-fdg -tool[6]. Below is a sample sentence of *Times are hard* in Finnish:

```
1 Ajat aika    subj>2  @NH N PL NOM (Times)
2 ovat olla    main>0  @MAIN V ACT IND
PRES PL P3 (are)
3 kovia kova comp>2 @NH A PL PTV (hard)
4 . .
5 <s><s>
```

Our test words were also annoted with the fi-fdg -tool. The annotated words have one word per line with the information for each word in tsv:

1. Word number
2. Surface form of the word
3. Baseform of the word, word boundaries marked with a hash sign (#)
4. Role of the word, e.g. main word or attribute etc.
5. Other annotations, e.g. class, case etc.

Words get their baseform where word boundaries of compound words are also marked. In addition, the main word is marked in multiword terms. These are needed in the methods we are testing.

An example annotation of *urethritis* i.e. *inflammation of the urethra*:

```
1 virtsaputken virtsa#putki attr>2
@PREMOD N SG GEN     (of the urethra)
2 tulehdus tulehdus main>0
@NH N SG NOM         (inflammation)
3 <p><p>
```

5 Results

Results for our methods can be seen in Table 1 and 2.

Group 1 was divided into two groups using FiWN. As we assumed, one word of each word pair in group 1b fit in the same synset as the other word, but not necessarily the otherway around.

[4]http://dumps.wikimedia.org/fiwiki/, downloaded January 2011

[5]http://www.connexor.com/nlplib/

[6]http://www.csc.fi/english/research/software/fi-fdg

Group	1a	1b	2a	2b	3	Altogether
Word Pairs	46	14	88	31	243	422
Percentage	10.9 %	3.3 %	20.8 %	7.3 %	57.6 %	100.0 %

Table 1: Results for method A (counting word pairs).

Method	Words	Applicable	Useful	Accuracy
Method B	479	151	119	78.8 %
Method C	479	116	82	70.7 %
Method D	479	4250	601	14.1 %

Table 2: Results for methods B, C and D (counting words).

Adding one word to the other words's synset required knowing the meaning of the words in order to decide which should be added to which synset, so this was done by hand.

Group 2 was manually checked and divided into two groups. Group 2a were clear cases where the unknown word could be added to one of the synsets of the other word. Some of these words were for instance Finnish terms for international words. For example *kudosoppi* can be added to synset {*histologia*} (*histology*).

Group 2b implied that the word already in FiWN has some new meaning which is why the new word cannot simply be added to an existing synset. This is the case for example in a synonym pair *aurinkokoira/hevonen* (*sun dog/horse*). Since *sun dog* is a native american term for *horse*, it is more precise and should be added as a hyponym. In group 2b, most of the new words were best suited as hyponyms of the known words.

Only four words from groups 1b and 2b were not suitable as a hyponym or in the same synset of the word in FiWN. For these words we instead used methods B, C and D and found indicators for all of them.

Group 3 is the largest group, which produces most of the new words for FiWN. This means that there is still room for expanding FiWN. On the other hand, at this point we did not exclude proper names from the test words. Test words might include words we are not currently interested in adding to FiWN. These include for example various person names and sport associations.

Excluding pairs from groups 1 and 2, we are left with 243 pairs from group 3, where neither word is in FiWN. The pairs contain 479 unique words. On these words, we used methods B, C and D.

Among the unique words from group 3 our program found 151 compounds. Of these, 119 were evaluated as having a good hyperonym which gives an accuracy of 78.8 %. There were some clear cases, e.g. *begonia#kasvi* (*begonia plant*) which is a plant. Some were a mere interpretation, for example a legal term *oikeus#olettama* (*legal presumption*) is a certain type of *olettama* (*presumption).*

In group 3 our program found 116 multiword terms and 82 were evaluated as having a good hyperonym. This gives an accuracy of 70.7 %. Words include for example *neoklassinen musiikki* (*neoclassical music*) which is a kind of *musiikki* (*music*). Out of the good hyperonyms, 19 were compound words from the previous method.

The last method was to use patterns from (Lindén et al., 2012). Patterns included:

- *kuten, kuten esimerkiksi* (as/like/such as, as for example)
- *ja/tai muu* (and/or...other)
- *NP(nom)...on/ovat/oli/olivat..NP* (NP(nominative form)...is/are/was/were...NP)

First we searched for sentences which contained the specific pattern. With the first pattern we found 42360 sentences, with pattern *and...other* 33482 sentences and with pattern *or...other* 5171 sentences. Sentences having the last pattern were 964911.

From these senteces we searched those which contained the test words from group 3. The first pattern matched 340 sentences. With pattern *and...other* we found 266 sentences and with pattern *or...other* 26 sentences. With the last pattern we added a restriction that the word should be in nominative form. This resulted in 3618 sentences.

We manually evaluated which ones indicated a hyperonym or a hyperonym. Most of the good results came from the last pattern. Useful patterns in Table 2 indicate how many of the results produced a hyponym or hyperonym for the given word. Testing for compounds and multiword terms resulted in 297 words with no indication of a hyperonym or hyponym. With patterns, we were able to find indications for 142 of those words. We also found new hyperonyms or hyponyms for some of the compound words and multiword terms.

Methods A-D result in only 155 words with no indication of a location in FiWN. On the other hand, since our test set consisted of synonym pairs, we only need an indication for one of the words. This lowered the words with no indication to 100 (50 synonym pairs). Since we had 594 new words (strings), this means that for 83.2 % of all of the new words we found some indication. Out of the 376 new synonym pairs, i.e. excluding group 1a already in FiWN from the total, we were able to find an indication for 86.7 % of the pairs that were new to FiWN.

6 Discussion and Future Work

The tests gave us some insight into which methods are useful in order to find hyperonyms and hyponyms for a given word. Interestingly, finding compound words and multiword terms is simple and the accuracy is far better than using patterns. On the other hand, we currently only tested that a sentence contained both the pattern and the given word. To get better accuracy for patterns, we need to more accurately check that the pattern itself contains the given word.

Categorizing synonym pairs into different groups allows us to concentrate on specific synonym pairs. With a small amount of manual work, we were able to cover groups 1 and 2 and focus solely on group 3 with our remaining methods.

Using both methods B and C for the words allows us to create multilevel hierarchies. For example, the word *laskennallinen virtaus#mekaniikka* (*Computational Fluid Dynamics*) is a multiword term giving us a hyperonym *virtaus#mekaniikka* (*Fluid Dynamics*). In addition, this main word is also a compound word, giving us hyperonym *mekaniikka*[7] (*Dynamics*).

One problem was the quality of the fi-fdg -tool. Some words were incorrectly annotated, for example the wrong word was annotated as the main word. That is why some compounds and multiword terms from methods B and C did not show up in the results.

In group 1b of our synonym pairs, both words were included in FiWN, but they were not in the same synset. We concluded that one of the words in the word pair can be added to the synset of the other word. Deciding on which synset the other word should be added into was done by hand. Later on, this should be automated. For example, if we know which article produced the synonym pair, we can check what the article is about and deduce which synset is the best based on its hyperonyms. For example, the synonym pair *moira/kohtalotar* (*Moirai, Moirae/Norn,weird sister*) was found in an article talking about Greek gods. *Moirai/Mairae* has a hyperonym *Greek deity* whereas *Norn/weird sister* has a hyperonym *Norn deity*.

Groups 1b and 2b produced new meanings for words already in FiWN. Since those word pairs cover only 10.6 % of our test pairs, we can infer that most of the meanings of the words in FiWN are covered in the current version. We also conclude that categorizing word pairs is a useful method to discover new meanings.

The obvious thing to point out about these methods is that they can be iterated. Even if the hyperonym or hyponym for a word is missing from FiWN, we can use these same methods to find a location for the new hyperonym/hyponym. We can also use the pattern *a.k.a* to find synonyms for the new word.

7 Conclusions

We have described some simple methods for finding a location for a new word in FiWN. This means finding a synonym, a hyponym or a hyperonym for a new word. These indicate where a word can be added. As a side-effect of our first method, we can also find new meanings for some of the known words in FiWN.

Testing showed that for a test set of 594 totally new words (strings) for FiWN, we were able to find an indication of a location for 83.2 % of the words. Generally, exploiting the usual way a language coins new words based on established ones, e.g. compounding and multiword terms, has a better accuracy than general lexico-syntatic patterns. All in all, we were able to find an indication for

[7] Literal translation is *mechanics*.

86.7 % of the pairs that were new to FiWN.

Acknowledgements

We thank the FINCLARIN and METANORD projects for funding our research. We also thank the FinnWordNet participants and the anonymous reviewers for useful comments on the manuscript.

References

Eugene Agichtein and Luis Gravano. 2000. Snowball: Extracting relations from large plain-text collections. In *Proceedings of the 5th ACM International Conference on Digital Libraries*, pages 85–94.

Eugene Agichtein, Luis Gravano, Jeff Pavel, Viktoriya Sokolova, and Aleksandr Voskoboynik. 2001. Snowball: a prototype system for extracting relations from large text collections. In *Proceedings of the 2001 ACM SIGMOD international conference on Management of data*, SIGMOD '01, pages 612–, New York, NY, USA. ACM.

Matthew Berland and Eugene Charniak. 1999. Finding parts in very large corpora. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 57–64, Morristown, NJ, USA. Association for Computational Linguistics.

Sergey Brin. 1999. Extracting patterns and relations from the world wide web. In *Selected papers from the International Workshop on The World Wide Web and Databases*, pages 172–183, London, UK. Springer-Verlag.

Sharon A. Caraballo. 1999. Automatic construction of a hypernym-labeled noun hierarchy from text. In *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics*, pages 120–126, Morristown, NJ, USA. Association for Computational Linguistics.

Lauri Carlson and Krister Lindén. 2010. FinnWordNet - WordNet på finska via översättning. *In LexicoNordica - Nordic Journal of Lexicography 2010, vol 17, ISSN 0805-2735.*

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.

Roxana Girju, Adriana Badulescu, and Dan Moldovan. 2003. Learning semantic constraints for the automatic discovery of part-whole relations. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology - Volume 1*, NAACL '03, pages 1–8, Morristown, NJ, USA. Association for Computational Linguistics.

Roxana Girju, Adriana Badulescu, and Dan Moldovan. 2006. Automatic discovery of part-whole relations. *Computational Linguistics*, 32:83–135, March.

Zellig S. Harris. 1968. *Mathematical Structures of Language*. Wiley, New York, NY, USA.

Marti A. Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the 14th conference on Computational linguistics*, pages 539–545, Morristown, NJ, USA. Association for Computational Linguistics.

Marti A. Hearst. 1998. Automated discovery of WordNet relations. In *C. Fellbaum, WordNet: An Electronic Lexical Database*, pages 131–153. MIT Press.

Dekang Lin and Patrick Pantel. 2002. Concept discovery from text. In *Proceedings of the 19th international conference on Computational linguistics - Volume 1*, COLING '02, pages 1–7, Stroudsburg, PA, USA. Association for Computational Linguistics.

Dekang Lin. 1998. Automatic retrieval and clustering of similar words. In *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics - Volume 2*, ACL-36, pages 768–774, Morristown, NJ, USA. Association for Computational Linguistics.

Krister Lindén and Jussi Piitulainen. 2004. Discovering synonyms and other related words. In *Proceedings of CompuTerm 2004, 3rd International Workshop on Computational Terminology*, Geneva, Switzerland, August. Coling.

Krister Lindén, Mirka Hyvärinen, Jyrki Niemi, and Paula Pääkkö. 2012. Translating WordNet – and then what? *Language Resources and Evaluation*. Submitted.

George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J. Miller. 1990. Introduction to WordNet: An on-line lexical database. *International Journal of Lexicography*, 3(4):235–244.

Patrick Pantel and Dekang Lin. 2002. Discovering word senses from text. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, KDD '02, pages 613–619, New York, NY, USA. ACM.

Patrick Pantel and Marco Pennacchiotti. 2006. Espresso: leveraging generic patterns for automatically harvesting semantic relations. In *Proceedings of the 21st International Conference on Computational Linguistics and the 44th annual meeting of the Association for Computational Linguistics*, ACL-44, pages 113–120, Morristown, NJ, USA. Association for Computational Linguistics.

Patrick Pantel, Deepak Ravichandran, and Eduard Hovy. 2004. Towards terascale knowledge acquisition. In *Proceedings of the 20th international conference on Computational Linguistics (COLING-04)*, pages 771–777, Geneva, Switzerland. Association for Computational Linguistics.

Emanuele Pianta, Luisa Bentivogli, and Christian Girardi. 2002. MultiWordNet: developing an aligned multilingual database. In *Proceedings of the First International Conference on Global WordNet*, January.

Jussi Piitulainen. 2011. *Explorations in the distributional and semantic similarity of words*. Ph.D. thesis, University of Helsinki.

Deepak Ravichandran and Eduard Hovy. 2002. Learning surface text patterns for a question answering system. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, ACL '02, pages 41–47, Morristown, NJ, USA. Association for Computational Linguistics.

Rion Snow, Daniel Jurafsky, and Andrew Y. Ng. 2005. Learning syntactic patterns for automatic hypernym discovery. In Lawrence K. Saul, Yair Weiss, and Léon Bottou, editors, *Advances in Neural Information Processing Systems 17*, pages 1297–1304. MIT Press, Cambridge, MA.

D. Tufis, D. Cristea, and S. Stamou. 2004. Balkanet: Aims, methods, results and perspectives. a general overview. *Romanian Journal of Information Science and Technology, Vol. 7, Nos. 1-2*, pages 9–43.

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