Using context-window overlapping in synonym discovery and ontology extension

María Ruiz-Casado, Enrique Alfonseca and **Pablo Castells**

Department of Computer Science
Universidad Autónoma de Madrid
28049 Madrid
{Maria.Ruiz,Enrique.Alfonseca,Pablo.Castells}@uam.es

Abstract

This paper describes a new, unsupervised procedure called *Context-window overlapping* for calculating the semantic distance between two terms. It is based on the distributional semantics hypothesis, and, in particular, in the fact that synonym words should be interchangeable in every context, and hyponyms can be substituted by their hyperonyms in most contexts.

The procedure has been applied to synonym identification, and to ontology extension. In the first task, it has been evaluated with 80 synonym test questions from the TOEFL which already constitute a standard test set in this problem, and attains results similar to most other non-ensemble procedures. Interestingly, it clearly outperforms Latent Semantic Analysis, other procedure grounded on the Distributional Semantic hypothesis. Concerning ontology enrichment, the results obtained are promising, although they can still be much improved. Conclusions are drawn from this result, and we outline several possibilities for future work.

1 Introduction

There is much work concerning modelling semantic similarity between words. Some use statistical models, and other represent contexts using the vector space model, or make use of conceptual hierarchies (Banerjee & Pedersen 03; Budanitsky & Hirst 01; Resnik 99). Such metrics have very useful applications for both Information Retrieval and Automatic Annotation in the semantic web, as they have been used for disambiguating word senses inside documents (Agirre *et al.* 01), automatically extending conceptual ontologies (Alfonseca & Manandhar 02), and extending user queries with synonyms discovered automatically (Turney 01).

In this paper, we describe a new simple algorithm, also grounded on the Distributional Semantics hypothesis. The results obtained so far are very promising, when compared to most of the previous-mentioned procedures. The procedure has been evaluated in two different tasks: synonym identification, and automatic ontology enrichment, with encouraging results.

The paper is structured as follows: Section 2 describes the metric used to measure the similarity between terms; Sections 3 and 4 describe the two applications in which it has been evaluated. Finally, Section 5 draws some conclusions and describes open lines for future work.

2 Similarity metric with context-window overlapping

The Distributional Semantics (DS) hypothesis states that the meaning of a word w is highly correlated to the contexts where w appears (Rajman & Bonnet 92). From this assumption, it is possible to develop statistical computational tools for calculating similarities in word meanings, which have been applied to Information Retrieval (Rajman & Bonnet 92; Salton 89), Text Summarisation (Lin 97), word-sense disambiguation (Yarowsky 92; Agirre *et al.* 00), and word clustering (Lee 97; Faure & Nédellec 98).

This section starts with some commonly agreed definitions of two semantic relations that are very relevant for characterising word meaning: hyponymy and synonymy. Next, the new procedure proposed is described.

2.1 A definition of hyponymy and synonymy based on contexts

Hyponymy is a semantic relationship which relates a concept with more general concepts, such as *horse* with *animal*. It can be defined in the following way:

> **Definition 1a.** Hyponymy is a relation of meaning inclusion between linguistic expressions. A is a hyponym of B if B is true for any concept x whenever A is true for x.
>
> Hyperonymy is the inverse relation to hyponymy.

For example (Resnik 93), every single QUEEN is a WOMAN, and therefore QUEEN is a hyponym of WOMAN. This implies that any utterance about a queen x entails the same utterance where x is referred to as being a woman, e.g. (1a) entails (1b).

(1) a. The Prime Minister honoured the queen with his presence.

b. The Prime Minister honoured the woman with his presence.

The example leads us to the definition of hyponymy in terms of interchangeability of linguistic expressions:

> **Definition 1b.** A is a hyponym of B if and only if for every sentence S containing A, S entails the same sentence with A substituted for B, $S[A/B]$ (Lyons 61).

Word Meanings	Word forms horse	heroin	junk	debris
horse, Equus sp.	×			
horse, heroin (drug)	×	×	×	
junk (Chinese boat)			×	
debris, detritus			×	×
...				

Table 1: Example of lexical matrix, showing some words and the concepts they lexicalise.

The second lexical relationship described in this section is **synonymy**, which relates words that convey the same meaning. In (Miller 95), synonymy is characterised as a matrix that relates word meanings to word forms. Word forms are typically sequences of characters delimited by spaces. However, in some contexts special symbols may be considered words and, as (Resnik 93) points out, provision must be made as well for multi-word expressions. Word meanings refer to "the lexicalised concept that a form can be used to express" (Miller 95). A particular example with some concepts and word forms is shown in Table 1. Here, {horse, heroin, junk} is a set of synonyms (a *synset*) that represents the concept *heroin* as a drug.

Some semanticists argue that the denotational meaning of a word is fully realised in contexts. As Firth (57, pg. 7) says, "The complete meaning of a word is always contextual, and no study of meaning apart from a complete context can be taken seriously", a theory agreed also by Cruse (86, p. 270) when he says that "natural languages abhor absolute synonyms just as nature abhors a vacuum". Under this premise, it is rare that two words have exactly the same meaning and are exchangeable in every possible context. Edmonds & Hirst (02) argue that many words are not absolute synonyms, but near-synonyms (also called *plesionyms*).

Even so, for practical purposes, we often use the relationship of synonymy between words, for instance, when explaining the meaning of a word in a context by giving other words which can be used in the same place (Resnik 93). In this way, we could define synonym words as words that convey the same meaning. Therefore, we can write parallel definitions to (1a) and (1b), using the fact that synonym words must be interchangeable in every context.

> **Definition 2a. Synonymy** is a relation of meaning identity between linguistic expressions. A and B are synonyms if and only if B is true for a concept x whenever A is true for x **and vice versa**.
>
> **Definition 2b.** Two word forms w_1 and w_2 are **synonyms** if and only if for every sentence S containing A, then S entails $S[A/B]$, and for every sentence T containing B, then T entails $T[B/A]$.

Corollary 2c comes straightforwardly from definition 2b. If two word forms w_1 and w_2 are exchangeable in every sentence where any one of them appears, then they can be used in exactly the same contexts in language:

> **Corollary 2c.** If two word forms w_1 and w_2 are synonym, then they can appear in exactly the same contexts, preserving the truth value.

Finally, we can define synonymy in terms of hyperonymy as in the following definition. It can be seen that, if we use definition (2d), then (2a) and (2b) can be derived from (1a) and (1b).

> **Definition 2d.** Two word forms w_1 and w_2 are synonyms if and only if both w_1 is a hyperonym of w_2 and w_2 is a hyperonym of w_1.

This said, we should also bear in mind that, although the notion of synonymy may be useful for practical purposes, it is very rare to find two words that are completely interchangeable. The difference in meaning may just be stylistic, or due to dialectal variations, but even in these cases we can expect that the sets of contexts in which we shall be able to find the two words will not be absolutely coincident.

2.2 Measuring similarity between contexts

These definitions of hyperonymy and synonymy give us ground to define metrics for semantic similarity between word forms which are based on the similarity between contexts. A popular technique to encode contexts and measure their similarity is the Vector Space Model (VSM), given a word w which appears in a corpus, we first define a context length (e.g. the words in the same sentence, or the words in a window of width L), and next we collect all the words in the context of every occurrence of w inside a bag. That *bag of words* will represent the meaning of w, and several semantic similarity metrics can be defined between the bags corresponding to two words. VSM can be extended with Latent Semantic Analysis (LSA), a dimensionality reduction procedure (Landauer & Dumais 97). We should note that, in VSM, there is much information lost, as all the words are put together in the bag and the syntactic dependences between the contextual terms will not be stored in the model.

A different approach, Pointwise Mutual Information (PMI) (Turney 01) is grounded on the slightly different assumption that two words with similar meanings will tend to appear near each other:

$$PMI(w_1, w_2) = \frac{hits(w_1\ NEAR\ w_2)}{hits(w_1)hits(w_2)}$$

So, for instance, many documents about *cars* are expected to contain the synonym word *automobile* as well.

one_way_similarity(w_1, w_2)
1. return count(w_1, w_2)

two_ways_similarity(w_1, w_2)
1. return count(w_1, w_2) + count(w_2, w_1)

count(w_1, w_2)
1. Collect, in S_1, $N_{snippets}$ Google snippets where w_1 appears
2. Set $n = 0$
3. For each snippet s_i in S_1,
   3.1. ctx = window of width L around w_1 in s_i.
   3.2. Remove the words from ctx if there is a sentence ending between them and w_1.
   3.3. If number of open-class words in $ctx < \theta$, *continue*.
   3.4. If ctx has already been seen, *continue*.
   3.5. Substitute w_1 by w_2 in ctx.
   3.6. Search in Google for ctx.
   3.7. If found any result, increment n.
4. Return n.

Figure 1: Pseudocode of the *Context-window overlapping* algorithm. $N_{snippets}$ is the number of snippets obtained from Google; L is the context width; θ is the minimum number of open-class words to consider a context.

2.3 Context-window overlapping

A possible drawback of the VSM technique is that much information is lost when all the sentences are reduced to a bag-of-words representation. In this operation:

- We lose information of which terms appeared in which contexts, as all the contexts are merged in a single vector.
- We lose information of the word-order and the phrasal structures inside each context.

If we want to calculate the similarity between two words, w_1 and w_2, ideally, it should be better to keep, for each of them, the complete contexts in which they can appear, and to compare the two sets of contexts, without any other transformation. If we consider sentences as contexts, we could describe the ideal procedure for calculating the similarity between two words in the following way:

1. Collect in S_1 every possible sentence in which w_1 can appear.
2. Collect in S_2 every possible sentence in which w_2 can appear.
3. Calculate the percentage of sentences in S_1 in which we can substitute w_1 for w_2 to obtain a sentence from S_2, and vice versa.

This procedure has two problems which stem from the current limitations of the technology:

- The number of possible sentences in which any word can appear can be arbitrarily large. If the sentences are collected from a textual corpus, we will necessarily have a sparse data problem. To overcome or reduce this problem, rather than collecting full sentences, we restrict the length of the context to a narrow *window*.
- It is highly unlikely that any corpus, apart from the Internet, will be large enough to let us collect enough contexts for both words. Therefore, we shall be forced to use the Internet. In this case, if we collect the contexts using a search engine, the time needed to get all the contexts in which w_1 appears (which may be hundreds of millions) will be so high that the procedure will not be usable at all.

  This problem might be reduced if (a) we collect a limited number of contexts for the first word, w_1, and (b) we directly substitute w_1 for w_2 inside those same contexts, to estimate the size of the intersection of S_1 and S_2.

Figure 1 shows the pseudocode of the *Context-window overlapping* algorithm. In a few words, it collects a list of contexts where w_1 appears, and it counts in how many of them it is possible to substitute w_1 with w_2, using the Internet as the reference corpus. In the version called *two_way_similarity*, the same is repeated exchanging the roles of w_1 and w_2.

The following sections describe the application of this algorithm for two different tasks: identification of synonym words, and ontology extension.

3 Synonym discovery

A particular application of semantic similarity metrics is the automatic identification of synonym words. Reported approaches for to solve this problem include LSA (Landauer & Dumais 97), PMI (Turney 01), metrics of proximity in documents combined with patterns of incompatibility (Lin & Zhao 03), thesaurus-based methods (Jarmasz & Szpakowicz 03), corpus-based similarity metrics (Terra & Clarke 03), and a combination of various procedures (Turney *et al.* 03). Several of the previous methods are grounded on the DS hypothesis.

All the systems reported here have been tested on a TOEFL test. The data set consists of 80 words, and for each of these words there are four possible

synonym candidates. The purpose of the task is to decide which of those candidates is the actual synonym. For instance, the first term in the data set is *enormously*, with candidate synonyms *appropriately, uniquely, tremendously* and *decidedly*. The system has to decide that *tremendously* is the synonym of the word. Recently, (Freitag *et al.* 05) have proposed a procedure for automatically generating TOEFL questions from WordNet.

In this approach, the two-way context-window overlapping procedure is used. As stated in section 2.1, the general idea is that, if two words are synonyms, then they are exchangeable in every context. Following the procedure introduced before, we can collect some snippets for the first word, substitute it by each of the candidate synonyms, and look how many of the context windows, with the original word substituted by the candidate, are also indexed by Google. Next, the same process is repeated by substituting the candidate synonym by the original word. The candidate that maximises the number of context windows in which we can interchange the two words will be selected.

There are three parameters have been set empirically:

- The window width (L) that has been taken is 5 words (the word under study, and two at each side). If this size is incremented, then the program returns a score of 0 for most of the candidates, because the windows would be too large and the probability of finding the same context window with the candidate synonym is very small.
- The threshold to consider that a context is informative is $\theta = 2$ (see step 3.3 in Figure 1). In this way, if Google has returned a context that is too small, for instance, because the original word is starting and ending a sentence, or because the context mainly contains closed-class words, then it will be ignored. With this threshold, context windows such as *for the WORD of the* will not be considered, because all the words at the left and at the right sides are closed-class words.
- Concerning the number of snippets to download, $N_{snippets}$, we have tried with several values, and we discovered that some of the candidates are more frequent than others. Hence, with a fixed number of snippets, it may be the case that all the candidates receive a similarity of 0. Therefore, $N_{snippets}$ is chosen dynamically to ensure that, from the several candidates, at least one of them reaches a count greater than 30. If there is a draw, more snippets are collected until it is untied. Note that these restrictions may require the collection of more than 1000 snippets from Google in some cases.

For instance, in the example mentioned above, *tremendously* had a score of 31, *uniquely* had a score of 5, *appropriately* had a score of 2, and *decidedly* had a score of 0. Therefore the first one was chosen as the candidate synonym for *enormously*.

Procedure	Acc.	95% conf.
(Landauer & Dumais 97)	64.40%	52.90–74.80%
non-native speakers	64.50%	53.01–74.88%
(Turney 01)	73.75%	62.71–82.96%
(Jarmasz & Szpakowicz 03)	78.75%	68.17–87.11%
(Terra & Clarke 03)	81.25%	70.97–89.11%
(Lin & Zhao 03)	81.25%	70.97–89.11%
CW overlapping	**82.50%**	72.38–90.09%
(Turney *et al.* 03)	97.50%	91.26–99.70%

Table 2: Results obtained (accuracy), and other published results on the TOEFL synonym results, from Turney *et al.* (03).

Results Table 2 shows the results obtained, compared to other published results on the TOEFL data set[1]. As can be seen, it outperforms all the previous approaches (although there is a statistical tie with some of them) except (Turney *et al.* 03). However, compared to this, our approach has the advantage that it does not require training, as it is fully unsupervised, and it is much more simple to implement.

4 Ontology extension

Ontologies are often described as "explicit specifications of a conceptualisation" (Gruber 93). They have proved to be a useful tool for knowledge representation. In many cases, ontologies are structured as hierarchies of concepts, by means of the hyperonymy relationship. Given the large cost of building and maintaining ontologies, there is already much work on procedures for automatically structuring concepts in ontologies, and for extending existing ontologies with new terms, and for populating an ontology with instances of its concepts. These tasks are usually called *ontology building*, *ontology enrichment* and *ontology population*, respectively. We may classify current approaches for ontology enrichment from text in the following groups:

- Systems based on distributional properties of words: they use some kind of distance metric based on co-occurrence information. This metric can be applied for clustering (Lee 97; Faure & Nédellec 98), for Formal Concept Analysis (Cimiano & Staab 04) or for classifying words inside existing ontologies (Hastings 94; Hahn & Schnattinger 98; Pekar & Staab 03; Alfonseca & Manandhar 02) or supersense categories (Curran 05).
- Systems based on pattern extraction and matching: these rely on lexical or lexicosemantic patterns to discover ontological and non-taxonomic relationships between concepts in unrestricted text. They may be based on manually defined regular expressions of words, (Hearst 92; Hearst 98; Berland & Charniak 99) or may learn such

[1]Obtained from Landauer and Praful Chandra Mangalath.

findHyperonyms(Word w)
1. Initialise a list *Candidates* with the top node.
2. While the list *Candidates* has changed in the previous iteration:
    2.1. Extend the list *Candidates* with the hyponyms of all the nodes that are already inside it.
    2.2. For every node n in *Candidates* (which is a set of synonym words),
        2.2.1. Initialise *n.score* to 0.
        2.2.2. For every synonym word s in that node,
            2.2.2.1. *n.score+ = one-way-similarity(w,s).*
    2.3. *Candidates* $\leftarrow$ the N nodes with the best scores.
3. Return *Candidates*.

Figure 2: Pseudo-code of the program for finding candidate hyperonyms for a given word. N is the beam width of the search.

Step	Top-5 Candidates	Score
1	(a) unit, whole, whole thing	31
	(b) location	17
	(c) body of water, water	17
	(d) building block, unit	16
	(e) part, piece	13
2	(a) unit, whole, whole thing	31
	(b) point	30
	(c) part, region	29
	(d) region	19
	(e) line	19
3	(a) area, country	34
	(b) point	16
	(c) district, territory	15
	(d) place, spot, topographic point	13
	(e) unit, whole, whole thing	11
4	(a) center, centre, eye, heart, middle	33
	(b) area, country	20
	(c) district, territory	13
	(d) place, spot, topographic point	9
	(e) point	9
5	(a) center, centre, eye, heart, middle	33
	(b) area, country	20
	(c) district, territory	13
	(d) place, spot, topographic point	9
	(e) point	9

Figure 3: Example showing the classification of Colchester

patterns from text (Finkelstein-Landau & Morin 99; Ruiz-Casado *et al.* 05). Navigli & Velardi (04) incorporates terminology extraction and ontology construction.

- Systems based on dictionary definitions analysis (Wilks *et al.* 90; Rigau 98; Richardson *et al.* 98) take advantage of the particular structure of dictionaries in order to extract hyperonymy relationships with which to arrange the concepts in an ontology. Concept definitions and glosses have been found very useful, as they are usually concise descriptions of the concepts and include the most salient information about them (Harabagiu & Moldovan 98).

This section describes a procedure for automatically extending WordNet with new terms using the context-window overlapping algorithm. If we follow definition (1b), we can assume that a term, in a sentence, can, in principle, be substituted by any of its hyperonyms. On the other hand, the inverse does not necessarily hold. Therefore, in this case, the procedure used is the one-way overlapping.

The algorithm is a top-down beam search procedure, in which we start at the top node in the ontology, and we proceed downwards, considering that node and all its children as candidate hyperonyms. The process is described in Figure 2.

Evaluation and results For the moment, the algorithm has been tested using the taxonomy of *entities* from WordNet, and 23 terms from the Simple English Wikipedia which did not appear in WordNet. The choice of these resources was done because our final purpose is to apply these techniques in a project about automatic knowledge acquisition from the Wikipedia. To choose the terms for the experiment, we first identified all the terms in the Wikipedia which were not in WordNet (around 600). Next, we removed from the beginning of the list, manually, those which were not hyponyms of *entity*. The 23 terms were chosen in order to have a representation of several kinds of concepts: persons, animals, locations and objects. Furthermore, in order to speed up the process, WordNet has been pruned to the 483 synsets that have an Information Content less than 7 (Resnik 99).

As an example, Figure 3 shows the classification performed for the concept Colchester. In the first step, *Colchester* is compared to the WordNet synset *entity*, and all its hyponyms, and the five ones with the highest scores (1a-1d) are kept for the next iteration. In the second step, Colchester is compared with these five synsets and all their hyponyms. This procedure is repeated until the set of five hyponyms does not change in one interation. In the example, this happens in iteration 5, moment in which they are returned to the user as candidates. In this example, the proposed hyperonyms at the end are *centre*, *area*, *district*, *place* and *point*.

Table 3 shows the results obtained when classifying the 23 terms inside WordNet. In the table, the candidates which are correct appear in bold-font, and

Term	Candidate 1	Candidate 2	Candidate 3	Candidate 4	Candidate 5
A Brief History of Time	human	piece	**whole**	body of water	organism
Alanis Morisette	**human**	**adult**	animal	**being**	part
Alaskan Native	**human**	*female person*	whole	*male person*	*woman*
Alpha male	**male person**	**human**	*male child*	**adult male**	**chief**
Angelina Jolie	**human**	animal	**adult**	part	compeer
Audrey Hepburn	**human**	animal	part	unit	**adult**
Bangalore	human	part	flora	compeer	line
Basque Country	*centre*	human	**area, country**	**region**	animal
Brad Pitt	**human**	animal	flora	**adult**	*friend*
Breakfast sausage	human	**whole**	part	body of water	friend
Britney Spears	part	body of water	unit	whole	cell
Brixton	human	whole	**part, region**	*line*	body of water
Burnham-on-Sea	**place, stop**	**location**	**part, region**	*line*	whole
Buzz Aldrin	**human**	animal	flora	part	whole
Caenorhabditis elegans	human	whole	**being**	cell	flora
Carl Sagan	**human**	animal	unit	**compeer**	part
Chorizo	human	animal	**whole**	body of water	part
Christina Ricci	part, region	whole	body of water	part	line
Christmas cracker	human	part	region	compeer	body of water
Christopher Columbus	**human**	part	unit	compeer	body of water
Coca-Cola	body of water	unit	part	line	region
Colchester	*centre*	**area, country**	**district**	**place**	*point*
Crewkerne	**place**	**part, region**	whole	*point*	part

Table 3: Results obtained for each of the the 23 terms.

those which are near correct hyperonyms in the ontology appear in italics. There are two terms which appear underlined; they correspond to the cases in which it was possible to identify automatically that the classification had not been successful, because the five candidates were all very far apart from each other in WordNet. Note that none of the candidates is located far too deep in the hierarchy; that is due to the pruning performed to WordNet.

It is possible to draw interesting observations from the results obtained:

- In some cases, the name chosen, in the ontology, for the hyperonym, is not easily associated to the word that we want to classify. This is the case of most objects and artifacts, which should be classified as *whole,unit* in order to proceed with the classification. Things such as book titles can probably be substituted by the term *book*, but they are hardly exchangeable by *whole* or *unit*. In these cases, the algorithm usually remains in the upper parts of the ontology and does not reach the most specific candidate hyperonym, as has happened with most objects in our experiment.
  This problem might improve if we modify the classification algorithm to force it proceed down deeper in the ontology.
- In some of the nodes, there are some synonyms terms which are used most of the times with a different sense. For instance, there is a node in WordNet, which is a hyponym of *location*, with the synonym terms {*center, centre, middle, heart, eye*}. The score for this node is inflated because there are many pages in the Internet containing the words *heart* and *eye*, but used with a different sense (as body parts).
  A possible solution might be to start by pruning the words in all the synsets to remove the meanings of the words that are rarely used.
  A clear weakness of this algorithm is its inability to treat polisemous words. This has been seen in the example with *eye*, but it would also happen with examples such as *horse* in Table 1 (meaning both Equus and heroin).
- Some terms are more common in the Internet that others. For instance, the words in the synset {*person, individual, someone, somebody, mortal, human, soul*} appear, as indexed by Google, with a frequency that is one order of magnitude higher than the words in many of the other synsets. Therefore, it will be more probable to find context windows with these words, just because they are more common. In fact, *person* was one of the five hyperonym candidates in 17 out of the 23 cases.
  This may indicate that it should be useful to adjust the frequencies using a statistical test, such as the χ^2 or the log likelihood.

5 Conclusions and future work

In this paper, we describe a procedure for calculating a semantic similarity metric between terms, based on their interchangeability in textual contexts. The metric has been tested on two different tasks: synonym discovery, consisting on identifying amongst four candidates, which one is the synonym of a given word; and ontology enrichment with new terms. The results for synonym detection are very good, being either equal or higher than all the other unsupervised methods. In the case of ontology enrichment, the results seem promising for the moment, and we also describe several ways in which we believe that the algorithm can be improved.

Some lines open for future work include to study more in-depth how the performance changes if we vary the parameters L, θ and $N_{snippets}$; and to check whether this procedure also outperforms VSM or LSA

in other problems.

Concerning ontology enrichment, we believe that the system could be much improved if we apply the solutions proposed in Section 4: to modify the algorithm to search deeper in the ontology; to work with weights calculated with an statistical test, rather than working with frequencies, and to remove, from each synset, the words which are generally used with a different sense, for instance, using Semcor to calculate the frequency of each sense.

References

(Agirre *et al.* 00) E. Agirre, O. Ansa, E. Hovy, and D. Martinez. Enriching very large ontologies using the www. In *Ontology Learning Workshop, ECAI*, Berlin, Germany, 2000.

(Agirre *et al.* 01) E. Agirre, O. Ansa, D. Martínez, and E. Hovy. Enriching wordnet concepts with topic signatures. In *Proceedings of the NAACL workshop on WordNet and Other lexical Resources: Applications, Extensions and Customizations*, 2001.

(Alfonseca & Manandhar 02) E. Alfonseca and S. Manandhar. Extending a lexical ontology by a combination of distributional semantics signatures. In *Knowledge Engineering and Knowledge Management*, volume 2473 of *Lecture Notes in Artificial Intelligence*, pages 1–7. Springer Verlag, 2002.

(Banerjee & Pedersen 03) P. Banerjee and T. Pedersen. Using measures of semantic relatedness for word sense disambiguation. In *Proceedings of the Fourth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico City, February 2003.

(Berland & Charniak 99) M. Berland and E. Charniak. Finding parts in very large corpora. In *Proceedings of ACL-99*, 1999.

(Budanitsky & Hirst 01) A. Budanitsky and G. Hirst. Semantic distance in wordnet: An experimental, application-oriented evaluation of five measures. In *Workshop on WordNet and Other Lexical Resources, Second meeting of NAACL*, Pittsburgh, 2001.

(Cimiano & Staab 04) P. Cimiano and S. Staab. Clustering concept hierarchies from text. In *Proceedings of LREC-2004*, 2004.

(Cruse 86) D. A. Cruse. *Lexical Semantics*. Cambridge University Press, 1986.

(Curran 05) J. Curran. Supersense tagging of unknown nouns using semantic similarity. In *Procs. of ACL'05*, pages 26–33, 2005.

(Edmonds & Hirst 02) P. Edmonds and G. Hirst. Near synonymy and lexical choice. *Computational Linguistics*, 2002.

(Faure & Nédellec 98) D. Faure and C. Nédellec. A corpus-based conceptual clustering method for verb frames and ontology acquisition. In *LREC workshop on Adapting lexical and corpus resources to sublanguages and applications*, Granada, Spain, 1998.

(Finkelstein-Landau & Morin 99) M. Finkelstein-Landau and E. Morin. Extracting semantic relationships between terms: supervised vs. unsupervised methods. In *Proceedings of the International Workshop on Ontologial Engineering on the Global Information Infrastructure*, 1999.

(Firth 57) J. Firth. *Papers in Linguistics 1934-1951*. Oxford University Press, London, 1957.

(Freitag *et al.* 05) D. Freitag, M. Blume, J. Byrnes, E. Chow, S. Kapadia, R. Rohwer, and Z. Wang. New experiments in distributional representations of synonymy. In *Proceedings of CoNLL-2005*, pages 25–32, 2005.

(Gruber 93) T. R. Gruber. A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2):199–220, 1993.

(Hahn & Schnattinger 98) U. Hahn and K. Schnattinger. Towards text knowledge engineering. In *AAAI/IAAI*, pages 524–531, 1998.

(Harabagiu & Moldovan 98) A. M. Harabagiu and D. I. Moldovan. *Knowledge Processing. In (C. Fellbaum (Ed.) WordNet: An Electronic Lexical Database*, pages 379–405. MIT Press, 1998.

(Hastings 94) P. M. Hastings. *Automatic acquisition of word meaning from context*. University of Michigan, Ph. D. Thesis, 1994.

(Hearst 92) M. A. Hearst. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of COLING-92*, Nantes, France, 1992.

(Hearst 98) M. A. Hearst. *Automated Discovery of WordNet Relations. In Christiane Fellbaum (Ed.) WordNet: An Electronic Lexical Database*, pages 132–152. MIT Press, 1998.

(Jarmasz & Szpakowicz 03) M. Jarmasz and S. Szpakowicz. Roget's thesaurus and semantic similarity. In *Proceedings of RANLP-03*, 2003.

(Landauer & Dumais 97) T. K. Landauer and S. T. Dumais. A solution to plato's problem: The latent semantic analysis theory of acquisition, induction and representation of knowledge. *Psychological Review*, 104(2):211–240, 1997.

(Lee 97) L. Lee. *Similarity-Based Approaches to Natural Language Processing*. Ph.D. thesis. Harvard University Technical Report TR-11-97, 1997.

(Lin & Zhao 03) D. Lin and S. Zhao. Identifying synonyms among distributinally similar words. In *Proceedings of the IJCAI-2003 Conference*, pages 1492–1493, 2003.

(Lin 97) C.-Y. Lin. *Robust Automated Topic Identification*. Ph.D. Thesis. University of Southern California, 1997.

(Lyons 61) J. Lyons. *A structural theory of semantics and its applications to lexical sub-systems in the vocabulary of Plato*. Ph. D. thesis, University of Cambridge, England. Published as *Structural Semantics*, No. 20 of the Publications of the Philological Society, Oxford, 1963, 1961.

(Miller 95) G. A. Miller. WordNet: A lexical database for English. *Communications of the ACM*, 38(11):39–41, 1995.

(Navigli & Velardi 04) R. Navigli and P. Velardi. Learning domain ontologies from document warehouses and dedicated websites. *Computational Linguistics*, 30(2), 2004.

(Pekar & Staab 03) V. Pekar and S. Staab. Word classification based on combined measures of distributional and semantic similarity. In *Proceedings of Research Notes of the 10th Conference of the European Chapter of the Association for Computational Linguistics*, Budapest, Hungary, 2003.

(Rajman & Bonnet 92) M. Rajman and A. Bonnet. Corpora-based linguistics: new tools for natural language processing. In *1st Annual Conference of the Association for Global Strategic Information*, Germany, 1992. Bad Kreuznach.

(Resnik 93) P. Resnik. *Selection and Information: A Class-Based Approach to Lexical Relationships*. Ph.D. thesis. Dept. of Computer and Information Science, University of Pennsylvania, 1993.

(Resnik 99) P. S. Resnik. Semantic similarity in a taxonomy: An information-based measure and its application to problems of ambiguity in natural language. *Journal of Artificial Intelligence Research*, 11:95–130, 1999.

(Richardson *et al.* 98) S. D. Richardson, W. B. Dolan, and L. Vanderwende. MindNet: acquiring and structuring semantic information from text. In *Proceedings of COLING-ACL'98*, volume 2, pages 1098–1102, Montreal, Canada, 1998.

(Rigau 98) G. Rigau. *Automatic Acquisition of Lexical Knowledge from MRDs*. PhD Thesis, Departament de Llenguatges i Sistemes Informàtics.– Universitat Politècnica de Catalunya. – Barcelona, 1998.

(Ruiz-Casado *et al.* 05) M. Ruiz-Casado, E. Alfonseca, and P. Castells. Automatic extraction of semantic relationships for wordnet by means of pattern learning from wikipedia. In *Proceedings of NLDB-05*, 2005.

(Salton 89) G. Salton. *Automatic text processing*. Addison-Wesley, 1989.

(Terra & Clarke 03) E. Terra and C. L. A. Clarke. Frequency estimates for statistical word similarity measures. In *Proceedings of HLT/NAACL-2003*, pages 244–251, 2003.

(Turney 01) P. D. Turney. Mining the web for synonyms: PMI-IR versus LSA on TOEFL. In *Proceedings of the 12th European Conference on Machine Learning (ECML-2001)*, pages 491–502, 2001.

(Turney *et al.* 03) P. D. Turney, M. L. Littman, J. Bigham, and V. Schnayder. Combining independent modules to solve multi-choice synonym and analogy problems. In *Proceedings of RANLP-03*, 2003.

(Wilks *et al.* 90) Y. Wilks, D. Fass, C. Ming Guo, J. McDonald, T. Plate, and B. Slator. Providing machine tractable dictionary tools. *Journal of Computers and Translation*, 2, 1990.

(Yarowsky 92) D. Yarowsky. Word-Sense Disambiguation using statistical models of Roget's categories trained on large corpora. In *Proceedings of COLING-92*, pages 454–460, France, 1992.