Enhancing Word Sense Disambiguation Using A Hybrid Knowledge-Based Technique

Eniafe Festus Ayetiran[1], Guido Boella[2], Luigi Di Caro[2], Livio Robaldo[2]

[1] CIRSFID, University of Bologna, Italy
eniafe.ayetiran2@unibo.it
[2] Department of Computer Science, University of Turin, Italy
{boella, dicaro, robaldo}@di.unito.it

Abstract. In this paper, we present a method for enhancing Word Sense Disambiguation through a hybrid system derived from the combination of Jiang & Conrath's similarity measure and a modified version of the original Lesk's algorithm by validating their results. The validation was carried out by computing the cumulative maximum overlap scores between the glosses of the senses returned by each of the algorithms and that of their semantically related senses in the WordNet lexical taxonomy with the original sentence containing the target word being disambiguated. Evaluation performed using the coarse-grained English all-words data of the SemEval 2007 showed that this method outperforms each of the two algorithms used alone and other independent state-of-the-art knowledge-based systems.

1   Introduction

Ambiguity is a fundamental characteristic of every language of which the English Language is not an exception. A considerable number of English words have more than one meaning. The meaning of word intended by a speaker or writer can be inferred considering the context of usage.

For example, consider the following sentences: (a) *My bank account yields a lot of interest annually* (b) *The children are playing on the bank of the river*. Based on the context of usage of the word "*bank*" in the two sentences above, we can infer that the first instance i.e sentence (a) is referring to a financial institution while the second instance, sentence (b) is referring to a sloping land beside a river. However, human identification of the right sense of a word as used in a statement is relatively easy but not with machines which need to process large unstructured textual information, carrying out complex computations in order to determine the right sense of the word as used in the statement.

The computational identification of the meaning of words in context is called Word Sense Disambiguation (WSD) also known as Lexical Disambiguation. Basically, the output of any word sense disambiguation system is a set of words sense-tagged with the right synonymous word (if any). Considering the instances in

the examples above, the sentences can be sense-tagged as follow: (a) *My bank/financial institution/banking concern* account yields a lot of interest annually. (b) *The children are playing on the **bank/sloppy land** of the river*.

Word Sense Disambiguation relies on knowledge. This means it uses a knowledge source or knowledge sources to associate the most appropriate senses with the words in context. Ideally, Word Sense Disambiguation is a means to an end but not usually the end itself, enhancing other tasks in different fields and application development such as parsing, semantic interpretation, machine translation, information retrieval and extraction, text mining, and lexical knowledge acquisition.

Approaches to word sense disambiguation may be knowledge-based (which depends on some knowledge dictionary or lexicon), supervised (machine learning techniques to train a system from labelled training sets) or unsupervised (based on unlabelled corpora, and do not exploit any manually sense-tagged corpus to provide a sense choice for a word in context). While knowledge-based and unsupervised approaches are automatic and do not involve manual training of systems, supervised approaches involve the herculean task of manually training systems with labelled datasets and often perform best on test sets from which part of the training sets have been taken.

The technique we used is a knowledge-based approach. We have employed English WordNet as the knowledge source and lexical resource. In the next section, we describe the English WordNet and discuss some relations in it that we have particularly employed for disambiguation in addition to those previously used in the literature. In the following sections, we also discuss related works, method, discussion of the results obtained and finally the conclusion.

2   WordNet

WordNet is a manually-constructed lexical system developed by George Miller (1990) and his colleagues at the Cognitive Science Laboratory at Princeton University and also made available in electronic form (Fellbaum, 1998). The basic object in WordNet is a set of strict synonyms called a synset. By definition, each synset in which a word appears is a different sense of that word. There are four main divisions in WordNet, one each for nouns, verbs, adjectives and adverbs. Within a division, synsets are organised by the lexical relations defined on them. For nouns, the lexical relations include antonymy, hypernymy/hyponymy (IS-A relation) and three different meronymy/holonymy (PART-OF) relations. The verb also include the antonymy, hypernymy/hyponymy, troponymy and other relations like entailment, causes etc. The IS-A relation is the dominant relation, and organizes the synsets into a set of hierarchies.

We particularly focus on some of the important relations in WordNet which we have employed that had not been previously explored in word sense disambiguation. First is the derivationally related forms of a word which are generally thought to be highly regular and productive, and the addition of given affixes to their base forms produce new words whose meanings differ from that of the base words in a predictable way. Also used are the morphologically related nouns of adjectives and adverbs

and this is applicable to adjectives and adverbs formed from nouns. For instance, the following can be derived from the word "*editorial*": "*editorialize*", "*editorialist*", "editor" and the noun form "voice" can be derived from the adjective "*voca*l". The second property is the antonymy relation which is the opposite of the target word e.g *wet – dry*. The third is entailment (verbs). According to WordNet glossary, entailment is defined as: "A verb X entails Y if X cannot be done unless Y is, or has been done" e.g *divorce entail marry*. The last is causes (verbs) and this can defined as an action or actions (causative) that triggers another action (resultative) e.g. *give – have*.

3 The Original Lesk's Algorithm

A basic knowledge-based approach relies on the calculation of the word overlap between the sense definitions of two or more target words. Micheal Lesk (1986) invented this approach named *gloss overlap* or the *Lesk's* algorithm. It is one of the first algorithms developed for the semantic disambiguation of all words in unrestricted texts. The only resource required by the algorithm is a set of dictionary entries, one for each possible word sense, and knowledge about the immediate context where the sense disambiguation is performed. The idea behind the Lesk's algorithm represents the starting seed for today's corpus-based algorithms. Almost every supervised WSD system relies one way or the other on some form of contextual overlap, with the overlap being typically measured between the context of an ambiguous word and contexts specific to various meanings of that word, as learned from previously annotated data.

The main idea behind the original definition of the algorithm is to disambiguate words by finding the overlap among their sense definitions. Namely, given two words, W_1 and W_2, each with N_{W1} and N_{W2} senses defined in a dictionary, for each possible sense pair W_1^i and W_2^j, $i = 1..N_{W1}, j = 1..N_{W2}$, we first determine the overlap of the corresponding definitions by counting the number of words they have in common. Next, the sense pair with the highest overlap is selected, and therefore a sense is assigned to each word in the initial word pair. The Algorithm is summarized in Listing 1 below:

```
1. for each sense i of W₁
2. for each sense j of W₂
3. compute overlap(i,j), the number of words in common
   between the definitions of sense i and sense j
4. find i and j for which overlap(i,j) is maximized
5. assign sense i to W₁ and sense j to W₂
```

Listing 1: The Original Lesk's Algorithm

4 Related work

Since the seminal work of Michael Lesk (1986), several variations of the algorithm have been proposed. Each version of the algorithm's variation either (1) attempts to solve the combinatorial explosion of possible word sense combinations when more than two words are considered or (2) attempts to disambiguate, where each word in a given context is disambiguated individually by measuring the overlap between its corresponding dictionary definitions and the current sentential context and alternatives where the semantic space of a word meaning is augmented with definitions of semantically related words.

Cowie et al. (1992) worked on a variation called simulated Annealing. In this work, they define a function E that reflects the combination of word senses in a given text whose minimum should correspond to the correct choice of word senses. For a given combination of senses, all corresponding definitions from a dictionary are collected, and each word appearing at least once in these definitions receives a score equal to its number of occurrences. Adding all these scores together gives the *redundancy* of the text. The E function is then defined as the inverse of redundancy. The goal is to find a combination of senses that minimizes this function. To this end, an initial combination of senses is determined (e.g., pick the most frequent sense for each word), and then several iterations are performed, where the sense of a random word in the text is replaced with a different sense, and the new selection is considered as correct only if it reduces the value of the E function. The iterations stop when there is no change in the configuration of senses.

Banerjee & Pedersen (2002) developed a variation of the Lesk algorithm called the Adapted Lesk algorithm. Part of our technique builds on this, so we elaborate more on it than on other variations. Definitions of semantically related word senses in WordNet lexical hierarchy are used in addition to the definitions of the word senses themselves to determine the most likely sense for a word in a given context. Banerjee and Pedersen (2002) employed a function similar to the one defined by Cowie et al (1992) to determine a score for each possible combination of senses in a text and attempt to identify the sense configuration that leads to the highest score. While the original Lesk algorithm considers strictly the definition of a word meaning as a source of contextual information for a given sense, Banerjee and Pedersen extended this algorithm using WordNet. In their work, they employed WordNet synsets to obtain WordNet senses and their meanings (through their glosses). In addition, related concepts and the definitions of each word sense based on semantic lexical relations in WordNet were also used. These relations include hypernymy, hyponymy, meronymy etc. In their work, a limited window size of context words were used by considering only the words before and after the target word.

The algorithm takes as input an example or an instance in which target word occurs, and it will produce the sense for the word based on information about it and few immediately surrounding words. The crux of the work is the adaptation of WordNet semantic relations to the original Lesk's algorithm and computation of similarity

based on maximum overlap of words (taking into account the number of times each word appears in the glosses).

5 Semantic Relatedness and Similarity

For a discourse to be coherent, words in the discourse must be related in meaning (Halliday and Hasan, 1976). This is a natural property of human languages and at the same time one of the most powerful constraints used in automatic word sense disambiguation. Words that share a common context are usually closely related in meaning (Harris, 1954), and therefore the appropriate senses can be selected by choosing those meanings found within the smallest semantic distance (Rada et al., 1989). While this kind of semantic constraint is often able to provide unity to an entire discourse, its scope has been usually limited to a small number of words found in the immediate vicinity of a target word, or to words connected by syntactic dependencies with the target word. These methods target the *local context* of a given word, and do not take into account additional contextual information found outside a certain window size. Similar to the Lesk's algorithm, these similarity methods become extremely computationally-intensive when more than two words are involved.

Patwardhan et al (2003) proposed that words in a discourse must be related in meaning for the discourse to be coherent. They distinguish between semantic relatedness and similarity following Budanitsky and Hirst (2004). They described semantic similarity as a kind of relatedness between two words that defines resemblance and semantic relatedness as a broader relationship between concepts that includes similarity as well as other relations such as is-a-kind-of, is-a-part-of as found in WordNet. Semantic similarity measures include Wu and Palmer (1994), Reisnik (1995), Agirre and Rigau (1996), Leacock et al. (1998), Hirst and St-Onge (1998), Lin (1998), Mihalcea and Moldovan (1999). Next, we discuss the Jiang and Conrath's measure which we used in our technique. We have adopted Jiang & Conrath's similarity measure for our final experiment because in our various experiments with each of the semantic similarity measures, we found it to be the best performing similarity measure.

6 Jiang & Conrath's Similarity Measure

Jiang & Conrath's similarity measure (Jaing & Conrath, 1998) is a similarity measure derived from corpus statistics and WordNet lexical taxonomy. It is a combined model that is derived from the edge-based notion by adding the information content as a decision factor. The model is based on the lexical taxonomy of the lexicon and statistics in the information content. In particular, attention is given to the determination of the link strength of an edge that links a parent node to a child node. Jiang & Conrath uses the difference in the information content of the two concepts to indicate their similarity. They used the information content defined by Resnik (1995) and augmented it with the notion of path length between concepts. This approach

includes the information content of the concepts themselves along with the information content of their lowest common subsumer. A lowest common subsumer is a concept in a lexical taxonomy which has the shortest distance from the two concepts compared. The resulting formula can be expressed in Equation 1 below:

$$Similarity = \frac{1}{(IC(s1) + IC(s2) - 2 * IC(lcs))} \tag{1}$$

Where *IC* is the information content, s_1 and s_2 are the first and second senses, respectively and *lcs* is the lowest common subsumer.

7   Method

In dealing with a raw text, we split the text into a set of sentences and the set of sentences split into a set of words called tokens. That is, for a text T, we split T into a set of finite sentences S_i, S ϵ T and tokenize each sentence S ϵ T into set of tokens W_i, Wϵ S. Each W ϵ S are then tagged into a part of speech. For monosemous words, the sense is returned as disambiguated based on the part of speech. For polysemous words, we followed the Adapted Lesk's approach of Bernajee and Pederson (2002) but we used the whole sentence as the window size. According to Heng Ji (2010), the antonymy relation is usually not exploited in WSD. However, if two context words are antonyms and belong to the same semantic cluster, they tend to represent the alternative attributes for the target word. Also, we opined that if two words are antonyms, there is possibility of mentioning the name of the opposite word in the gloss of the other opposite word, that of their semantically related senses or their examples and vice versa. Therefore, we also employed the *antonyms* in addition to hypernyms, hyponyms, meronyms etc used by Bernerjee and Pederson (2002). Also, for verbs we have used the *entailment* and *causes* relations. For adjectives and adverbs, we added the morphologically related nouns in computing the overlap score. In like manner, we disambiguated each target word in a sentence using the Jiang & Conrath's similarity measure using the whole sentence as the window size.

An agreement between the results produced by each of the two algorithms means the word under consideration has been rightly disambiguated and the sense on which they agreed is returned as the right sense. Wherever one method produces result and the other does not, the sense produced by the producing algorithm is returned as the right sense for the word being considered. Finally, in a situation where they produced different results, we validated the results returned by each of them. We did this by first obtaining the derivationally related forms of the words in context of the one under consideration and then add it to the original sentence to produce a new word list. Then for each of the senses produced by Modified Lesk and Jiang & Conrath algorithms, we obtain the maximum overlap between their glosses and that of their hypernyms, antonyms, hyponyms etc with the new word list. The sense with highest maximum overlap score is returned as the right sense. The intuition behind this notion of validation is that that the glosses of a word sense, and that of their semantically related ones in the WordNet lexical taxonomy should share words in common as much as

possible with words in the original sentence. Adding the derivationally related forms of the words in the sentence increases the chances of overlap where there are changes in the word morphology. Finally, where the validation of senses could not arrive at a result due to lack of overlap, the result of the Modified Lesk is chosen as the right sense since it performs better than the Jiang & Conrath's similarity measure.

We illustrate the technique discussed above with the first sentence of the Semeval 2007 coarse-grained English all-words data. The target words in this sentence with their parts of speech are: *editorial - noun, ill - adjective, homeless - noun, refer - verb, research - noun, six - adjective, colleague - noun, report - verb, issue - noun*. In this case, we do not have to tokenize and tag into parts of speech since the data is a pre-processed one. The first word in the sentence which is a noun is monosemous because only a single noun sense exists for the word "*editorial*", therefore the single sense is returned as the right sense. The second word is "*ill*" and is an adjective. It is polyse-mous because the adjective form of it has five senses. To disambiguate using the Modified Lesk, for each of the senses of "*ill*", we obtained their glosses and that of their semantically related senses in the WordNet taxonomy. We also obtained the glosses and glosses of semantically related senses of "*editorial*", "*homeless*"," *refer*", "*research*", "*six*", "*colleague*", "*report*" and "*issue*" and computed the cumulative maximum overlap among them. The sense with the highest cumulative overlap score is returned as the right sense. To disambiguate the other words in the sentence, the same process was applied.

Furthermore, the Jiang & Conrath measure is not applicable to "*ill*" in this case because it requires the hierarchical relation of lowest common subsumer (lcs) in WordNet which adjectives and adverbs do not have. Using the next word "*homeless*" to illustrate this, for each of the two noun senses of "*homeless*", we compute the Jiang & Conrath similarity between each sense and that of each noun senses of "*editorial*", "*research*", "*colleague*", and "*issue*". This is because the lowest common subsumer can be measured for senses with the same part of speech. We also computed the cu-mulative similarity score and returned the sense with the highest score as the right sense of "*homeless*". Also, to disambiguate the other words in the sentence using Jiang & Conrath, the same process was applied.

Finally, the Modified Lesk's method returned sense "*homeless%1:14:00::*" as the right sense with a gloss of "someone without a housing" while the Jiang & Con-rath method returned "*homeless%1:18:00*" with a gloss of "poor people who unfortu-nately do not have a home to live in". We added the derivationally related forms of each of the words in context of the target word to the original sentence and we have the following new word list: [*'editorial', 'editorialize', 'editorialist', 'editor', 'Ill', 'ill-ness', 'Homeless', 'homeless', 'homeless', 'homeless', 'homelessness', 'homeless', 'refer', 'reference', 'reference', 'reference', 'referee', 'referee', 'reference', 'referral', 'refer-ence', 'reference', 'reference', 'reference', 'research', 'research', 'research', 'research', 'researcher', 'research', 'researcher', 'research', 'research', 'six', 'colleague', 'collegi-al', 'report', 'report', 'report', 'report', 'report', 'report', 'report', 'report', 'report', 're-port', 'report', 'reporter', 'report', 'report', 'reportage', 'reporting', 'issue', 'issue', 'is-sue', 'issue', 'issue', 'issue', 'issue', 'issue', 'issuer', 'issue', 'issuer', 'issuing', 'issue', 'issue', 'issue', 'issue'*].

Like we did for the Modified Lesk, we obtained the glosses each of "*home-less%1:18:00::*" and "*homeless%1;14;00::*", that of their hypernyms, hyponyms,

meronyms, antonyms etc and computed the maximum overlap between them and the new word list. The final validation experiment returned "*homeless%1:18:00*" as the right sense. The same technique was applied for all the target words where the results of Modified Lesk and Jiang & Conrath algorithms do not agree.

8   Results and Discussion

In evaluating a Word Sense Disambiguation system, coverage is usually an important factor because it affects recall and F1 of the system. For systems that do not have a full coverage, back-off strategy is often used to supplement the results. It is a strategy where usually the Most Frequent Sense (MFS) of the target words are returned as the right sense wherever the system could not produce a result. However, the Most Frequent Sense (MFS) may not always help in all tasks as the accuracy varies from tasks to tasks and may not be ideal in practical application of Word Sense Disambiguation.

Our technique is strictly knowledge-based and independent as no back-off strategy or any form of supervision was employed. We evaluated our system with the coarse-grained English all-words data of the SemEval 2007 (Navigli et al., 2007) and presented the results obtained in Table 1 below:

Table 1. Analysis of Results by Part of Speech

Method	Precision				Recall				F1			
	Noun	Verb	Adj	Adverb	Noun	Verb	Adj	Adverb	Noun	Verb	Adj	Adverb
Modified Lesk	77.514	66.269	77.901	68.750	76.534	65.820	77.901	68.750	77.044	66.043	77.901	68.750
Jiang & conrath	76.748	66.263	-	-	76.263	64.805	-	-	76.504	65.525	-	-
KHybrid	82.581	68.020	-	-	82.581	68.020	-	-	82.581	68.020	-	-

There are several knowledge-based and unsupervised word sense disambiguation systems in the literature but we pick a handful with the state-of-the-art results using the same evaluation dataset like our system. We present the final result of our hybrid technique in comparison with other knowledge-based/unsupervised systems using the same evaluation data in Table 2 below:

Table 2. Final Result and Results of other State-of-the-Art Knowledge-Based/Unsupervised Systems

Method	Precision	Recall	F1
TKB-UO	70.207	70.207	70.207
Treematch	73.650	73.650	73.650
**KHybrid	**76.509**	**76.509**	**76.509**
ExtLesk (with back off)	79.1	79.1	79.1
Degree (with back-off)	81.7	81.7	81.7

TKB-UO (Sánchez et al., 2007) was the best performing unsupervised system at the SemEval 2007 in the coarse-grained English all-words task. Treematch (Chen et al., 2009), ExtLesk (Ponzento & Navigli, 2012) and Degree (Ponzeto & Navigli, 2012) emerged after the workshop. Degree incorporates weak supervision in its implementation. KHybrid is our hybrid method. ExtLesk without back-off for nouns has precision of 82.7, recall of 69.2 and F1 of 75.4 while Degree without back-off for nouns has a precision of 87.3, recall of 72.7 and F1 of 79.4. The overall results of both ExtLesk and Degree without back-off were not made available.

9   Conclusion

Implementing both the Modified Lesk and Jiang & Conrath's algorithms using sentential window size are computationally intensive. However, they produce better results than using specific limited window sizes. Obtaining similarity between glosses of senses produced by each of the algorithms and the original sentences containing the target word under consideration improves the overall performance. In the future, we will try to experiment with this technique on several word sense disambiguation methods to evaluate and compare its effect on them.

References

Banerjee, S. & Pedersen, T. (2002) An adapted Lesk algorithm for Word Sense Disambiguation using WordNet. In: Gelbukh, A. (ed) *Proceedings of the 3rd International Conference on Computational Linguistics and Intelligent Text Processing (CICLING),* Mexico City, Mexico, 17 – 23 February, 2002, pp. 136–145

Cowie, J., Guthrie, J. & Guthrie, L. (1992) Lexical Disambiguation Using Simulated Annealing. In: Boitet, C. (ed) *Proceedings of the 14th International Conference on Computational Linguistics (COLING),* Nantes, France, 23 – 28 August, 1992, pp. 157–161

Miller, G. (1990) WordNet: An Online Lexical Database. *International Journal of Lexicography, 3(4): 235 - 244*

Jiang, J. J. & Conrath, D. W. (1998) Semantic similarity Based on Corpus Statistics and Lexical Taxonomy. In: Chen, K., Huang, C. and Sproat, R. (eds) *Proceedings of the 10th International Conference on Research in Computational Linguistics,* Taipei, Taiwan, 2 – 4 August 1998, pp. 19 – 33

Lesk, M. **(**1986) Automatic Sense Disambiguation Using Machine Readable Dictionaries: How to Tell a Pine Cone from an Ice Cream Cone. In: Debuys, V. (ed) *Proceedings of the 5th ACM-SIGDOC Conference,* Toronto, Canada, 8 – 11 June 1986, pp. 24–26

Heng J. (2010) One Sense per Context Cluster: Improving Word Sense Disambiguation Using Web-Scale Phrase Clustering. In: Zhou, A. and Nakamura, S. (eds) *Proceedings of the 4th Universal Communication Symposium (IUCS),* Beijing, China, 18-19 October 2010, pp. 181 -184

Halliday, M. A. K. and Hasan, R. (1976) *Cohesion in English. English Language Series.* Longman, *New York*

Budanitsky, A., & Hirst, G. (2004) Evaluating WordNet-based measures of lexical semantic relatedness. *Computational Linguistics, 1(1): 1-49.*

Patwardhan, S., Banerjee, S. & Pedersen, T. (2003) Using Measures of Semantic Relatedness for Word Sense Disambiguation. In: Gelbukh, A. (ed) *Proceedings of the 4th International Conference on Intelligent Text Processing and Computational Linguistics,* Mexico City, Mexico, 16 – 22 February 2003, pp. 241–257

Resnik, P. (1995) Using Information Content to Evaluate Semantic Similarity. In: Mellish, C. S. (ed) *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI),* Montreal, Canada, 20 – 25 August 1995, pp. 448–453

Anaya- Sánchez, H., Pons-Porrata, A., & Berlanga-Llavori, R (2007) *TKB-UO:* Using sense clustering for Word Sense Disambiguation. In: Agirre, E., Màrquez, L., Wicentowski, R. (eds) *Proceedings of the 4th International Workshop on Semantic Evaluations (SemEval),* Prague, Czech Republic, 23 – 24 June 2007, pp. 322 - 325

Chen, P., Ding, W., Bowes, C. & Brown, D. (2009) A Fully Unsupervised Word Sense Disambiguation Method Using Dependency Knowledge. In: Doran, C., Ringger, E. (eds) *Proceedings of the 10th Annual Conference of the North American Chapter of the ACL,* Boulder, Colorado, 31 May – 5 June 2009, pp. 28–36

Ponzeto, S. P. & Navigli, R. (2010) Knowledge-rich Word Sense Disambiguation Rivaling Supervised Systems. In: Hajic, J. (ed) *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics,* Uppsala, Sweden, 11 – 16 July 2010, pp. 1522–1531.

Wu, Z. & Palmer, M. (1994) Verb semantics and lexical selection. In: Pustejovsky, J. (ed) *Proceedings of 32nd Annual Meeting of the Association for Computational Linguistics,* Las Cruces, New Mexico 27 – 30 June 1994, pp. 133–138.

Navigli, R., Litkowski, K. C. & Hargraves, O (2007*) SemEval-2007 Task 07: Coarse-Grained English All-Words Task. In: Agirre, E., Màrquez, L., Wicentowski, R. (eds) Proceedings of the 4th International Workshop on Semantic Evaluations,* Prague, 23 – 24 June 2007, pp. 30–35.

Lin, D. (1998*) An information theoretic definition of similarity. In: Jude, W. S (ed) Proceedings of the 15th International Conference on Machine Learning,* Madison, U.S.A., 24 – 27 July 1998, pp. 296–304.

Hirst, G. & St-Onge, D. (1998) Lexical chains as representations of context in the detection and correction of malaprorisms. In: Fellbaum, C. (ed), *WordNet: An Electronic Lexical Database,* pp. 305–332. MIT Press, Massachusetts, U.S.A.

Fellbaum, C. (ed) (1998) *WordNet: An Electronic Lexical Database.* MIT Press.

Agirre, E. & Rigau, G. (1996*) Word sense disambiguation using conceptual density. In: Tsujii, J. (ed) Proceedings of the 16th International Conference on Computational Linguistics (COLING),* Copenhagen, Denmark, pp. 16–22.

Mihalcea, R., & Moldovan, D. I. (1999) A method for word sense disambiguation of unrestricted text. In: Robert Dale and Ken Church (eds) *Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics,* Maryland, USA, 20 – 26 June 1999, pp. 152-158.

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

Rada, R., Mili, H., Bicknell, E., & Blettner, M. (1989) Development and application of a metric on semantic nets. *IEEE Transactions on Systems,* Man and Cybernetics, 19(1): 17-30.