A WordNet-based Algorithm for Word Sense Disambiguation

Xiaobin Li*
Department of Computer Science
Concordia University
Montreal, Quebec
Canada H3G 1M8
xiaobin@cs.concordia.ca

Stan Szpakowicz and **Stan Matwin**
Department of Computer Science
University of Ottawa
Ottawa, Ontario
Canada K1N 6N5
{szpak,stan}@csi.uottawa.ca

Abstract

We present an algorithm for automatic word sense disambiguation, based on lexical knowledge contained in WordNet and on the results of surface-syntactic analysis. The algorithm is part of a system that analyzes texts in order to acquire knowledge in the presence of as little pre-coded semantic knowledge as possible. On the other hand, we want to make the best use of public-domain information sources such as WordNet. Rather than depend on large amounts of hand-crafted knowledge or statistical data from large corpora, we use syntactic information and information in WordNet and minimize the need for other knowledge sources in the word sense disambiguation process. We propose to guide disambiguation by semantic similarity between words and heuristic rules based on this similarity. The algorithm has been applied to the Canadian Income Tax Guide. Test results indicate that even on a relatively small text the proposed method produces correct noun meaning more than 72% of the time.

1 Introduction

This work is part of the project that aims at a synergistic integration of Machine Learning and Natural Language Processing. The long-term goal of the project is a system that performs machine learning on the results of text analysis to acquire a useful collection of production rules. Because such a system should not require extensive domain knowledge up front, text analysis is to be done in a knowledge-scant setting and with minimal user involvement. A domain-independent surface-syntactic parser produces an analysis of a text fragment (usually a sentence) that undergoes interactive semantic interpretation. By design, we only need the user to approve the system's findings or prompt it for alternatives. Also by design, we limit ourselves to information sources in the public domain: inexpensive dictionaries and other lexical sources, such as WordNet.

WordNet [Miller, 1990; Beckwith *et al.*, 1991] is a very rich source of lexical knowledge. Since most entries have multiple senses, we face a severe problem of ambiguity. The motivation for the work described here is the desire to design a word sense disambiguation (WSD) algorithm that satisfies the needs of our project (learning from text) without large amounts of hand-crafted knowledge or statistical data from large corpora. We concentrate on using information in WordNet to the fullest extent and minimizing the need for other knowledge sources in the WSD algorithm. Semantic similarity between words (defined in the next section) plays an important role in the algorithm. We propose several heuristic rules to guide WSD. Tested on an unrestricted, real text (the Canadian Income Tax Guide), this automatic WSD method gives encouraging results.

Word sense disambiguation is essential in natural language processing. Early symbolic methods [Hirst, 1987; Small and Rieger, 1982; Wilks, 1975] heavily rely on large amounts of hand-crafted knowledge. As a result, they can only work in a specific domain. To overcome this weakness, a variety of statistical WSD methods [Brown *et al.*, 1991; Gale *et al.*, 1992; Resnik, 1992; Schutze, 1992; Charniak, 1993; Lehman, 1994] have been put forward. They scale up easily and this makes them useful for large, unrestricted corpora. One of the most important steps in statistical WSD methods, however, is statistically motivated extraction of word-word relationships from corpora. As Resnik points out [1993], "the corpus may fail to provide sufficient information about relevant word/word relationships and word/word relationships, even those supported by the data, may not be the appropriate relationships to look at for some tasks". Most of the statistical methods suffer from this limitation. Resnik proposes constraining the set of possible word classes by using WordNet. Rather than improving statistical approaches [Resnik, 1993; Sussna, 1993; Voorhees, 1993], we propose a completely different WordNet-based algorithm for WSD.

*This work was done while the first author was with the Knowledge Acquisition and Machine Learning group at the University of Ottawa. The authors are grateful to the Natural Sciences and Engineering Research Council of Canada for having supported this work with a Strategic Grant No. STR0117764.

2 WordNet and Semantic Similarity

WordNet is a lexical database with a remarkably broad coverage. One of its most outstanding qualities is a word sense network. A word sense node in this network is a group of strict synonyms called "synset" which is regarded as a basic object in WordNet. Each sense of a word is mapped to such a word sense node (i.e. a synset) in WordNet and all word sense nodes in WordNet are linked by a variety of semantic relationships, such as IS-A (hypernymy/hyponymy), PART-OF (meronymy/holonymy), synonymy and antonymy. The IS-A relationship is dominant – synsets are grouped by it into hierarchies. Our algorithm only accesses information about nouns and verbs, for which there exist such lexical hierarchies.

It has become common to use some measure of semantic similarity between words to support word sense disambiguation [Resnik, 1992; Schutze, 1992]. In this work, we have adopted the following definition: semantic similarity between words is inversely proportional to the semantic distance between words in a Wordnet IS-A hierarchy. By investigating the semantic relationships between two given words in WordNet hierarchies, semantic similarity can be measured and roughly divided into the following four levels:

Level 1: The words are strict synonyms according to WordNet: one word is in the same synset as the other word.

Level 2: The words are extended synonyms according to WordNet: one word is the immediate parent node of the other word in a IS-A hierarchy of WordNet. When used to get the synonyms of a word, WordNet not only produces the strict synonyms (a synset), but also the immediate parent nodes of this synset in a IS-A hierarchy of WordNet. Here, these immediate parent nodes are called "extended synonyms".

Level 3: The words are hyponyms according to WordNet: one word is a child node of the other word in a IS-A hierarchy of WordNet.

Level 4: The words have a coordinate relationship in WordNet: a word is a sibling node of the other word (i.e. both words have the same parent node) in a IS-A hierarchy of WordNet.

Level 1 concerns the semantic similarity between words inside a synset; Level 2 - Level 4 describe the semantic similarity between words that belong to different synsets (a synset which is composed of a group of strict synonyms is a node in the hierarchy of WordNet). The semantic similarity at Level 1 and Level 4 are symmetric, but the semantic similarity at Level 2 and Level 3 are not symmetric although both are about the relation between a word in a child synset and a word in its parent synset. In WordNet, a parent synset is considered as "extended synonyms" of its immediate child synset, but a child synset is not a synonym of its immediate child synset at all. For example, a parent synset "possession" is an "extended synonym" of its immediate child synset "property, belonging, holding, material possession", but a child synset "property, belonging, holding, material possession" is not a synonym of its immediate parent synset "possession" at all. Using the information about a parent synset to decide the intended meaning of a word in its immediate child synset is different from using the information about a child synset to decide the intended meaning of a word in its immediate parent synset. So here we have divided the semantic similarity between a parent synset and its immediate child synset into two levels (Level 2 and Level 3).

Although we have only applied the algorithm to WSD of noun objects in a text (i.e. nouns that are objects of verbs in sentences analyzed by the system), it can also be applied to other noun phrases in a sentence, in particular subjects.

In this approach, we must consider contexts that are relevant to our method and the semantic similarity in these contexts.

For all practical purposes, the possible senses of a word can be found in WordNet, but due to its extremely broad coverage, most words have multiple word senses. A context must be considered in order to decide a particular word sense. The notion of context and its use could differ widely across WSD methods. One may consider a whole text, a 100-word window, a sentence or some specific words nearby, and so on. In our work, we assume that a group of words co-occurring in a sentence will determine each other's sense despite each of them being multiply ambiguous. A simple and effective way is to consider as context the verbs that dominate noun objects in sentences. That is, we investigate verb-noun pairs to determine the intended meaning of noun objects in sentences.

In this work, then, we focus on investigating two aspects of semantic similarity:

- The semantic similarity of the noun objects
- The semantic similarity of their verb contexts

3 Heuristic Rules and Confidence of Results

We have set out to determine the intended meaning of a noun object in a given verb context: from all candidate word senses in WordNet of the noun object, select one sense that best fits the given verb context.

Suppose that the algorithm seeks the intended meaning of a noun object NOBJ in its verb context VC (that is, the intended meaning of NOBJ in a verb-object pair (VC, NOBJ)). NOBJ has n candidate word senses in WordNet. s(k) means the kth word sense of NOBJ in WordNet, $1 \leq k \leq n$. SS means the semantic similarity between words in a IS-A hierarchy of WordNet. An arrow is used to describe the relationship between words or between a word and its word sense. The heuristic rules adopted in the WSD algorithm are as follows:

Heuristic Rule 1 (HR1): Search the text for a verb-object pair (VC, NOBJ') in which NOBJ' has the semantic similarity with a word sense of NOBJ in WordNet and then consider this word sense of NOBJ as a result (see Fig. 1 and the example in

STEP 2 and STEP 3 of the WSD algorithm in the next section).

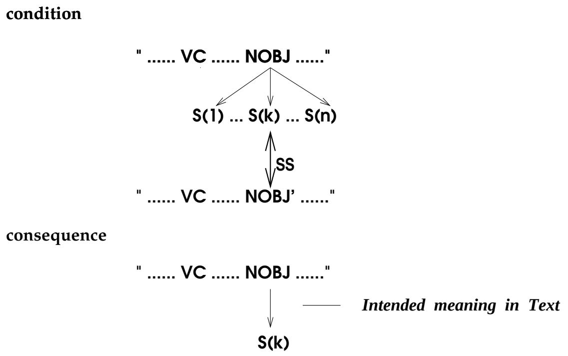

Figure 1: Principle of Heuristic HR1

Heuristic Rule 2 (HR2): Search the text for a verb-object pair (VC', NOBJ) in which VC' has the semantic similarity with VC. If the word sense of NOBJ in verb context VC' has already been acquired, then consider this word sense as a result (see Fig. 2 and the example in STEP 4 of the WSD algorithm).

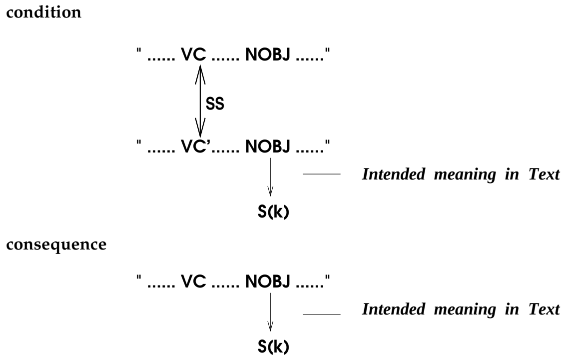

Figure 2: Principle of Heuristic HR2

Heuristic Rule 3 (HR3): Search the text for a verb-object pair (VC', NOBJ') in which VC' has the semantic similarity with VC and NOBJ' has the semantic similarity with a word sense of NOBJ in WordNet. Then consider this word sense as a result (see Fig. 3 and the example in STEP 5 and STEP 6 of the WSD algorithm).

Obviously, HR2 is not always feasible and it just contributes to the inference of further results based on the previously acquired results. HR3 infers results with a weaker constraint on the semantic similarity between verb-object pairs. HR1 is the main heuristic rule in the algorithm.

Besides the heuristic rules based on the semantic similarity between verb-object pairs, special heuristic rules related to the syntactic structure have also been formulated to assist word sense disambiguation; see [Grefenstette and Hearst, 1992] for a similar approach.

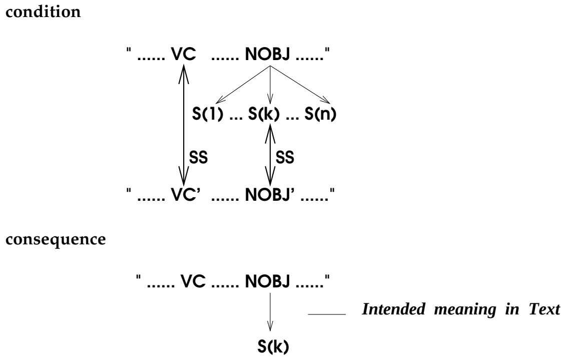

Figure 3: Principle of Heuristic HR3

Heuristic Rule 4 (HR4): Search the text for a "such as" structure which follows the verb-object pair (VC, NOBJ) in the text. The object NOBJ' of "such as" certainly has a synonym or hyponym relationship with a word sense of NOBJ in WordNet. Then consider this word sense as a result (see Fig. 4 and the example in STEP 7 of the WSD algorithm in the next section).

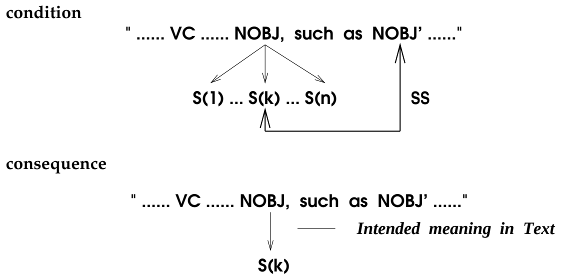

Figure 4: Principle of Heuristic HR4

Heuristic Rule 5 (HR5): Search the text for a VP coordinate structure such as either "VC and VC' " or "VC or VC' " which has the noun object NOBJ. If the word sense of NOBJ in the verb context VC' has already been acquired, then consider this word sense as a result (see Fig. 5 and the example in STEP 8 of the WSD algorithm).

HR4 only aims at one kind of special cases and its coverage is limited. HR5, like HR2, also depends on previously acquired results, but its coverage is more limited.

Because the results acquired by applying different heuristic rules have different accuracy, we have to define the measure of confidence in the result ("CF" for short) according to different heuristic rules and different semantic similarity adopted in these rules. In our algorithm, the assignment of CF values intuitively corresponds to our notion of decreasing semantic similarity (the test of the viability of this intuitive assignment has been shown in section 5). The CF values are numbers between 0 and 1. The assignment of CF values to levels is arbitrary, but it has served us well so far. Because

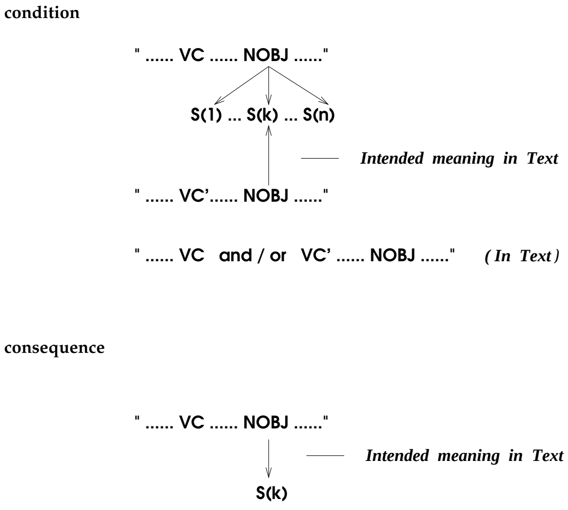

Figure 5: Principle of Heuristic HR5

the WSD process here is never completed only based on Level 1 semantic similarity between words, we have defined the CF values as numbers between 0 and 1 ($0 \leq$ CF ≤ 1). The assignment of CF values is shown as follows:

- **CF = 1.0:** The results acquired by applying HR1; NOBJ has only one word sense in WordNet.
- **CF = 0.9:** The results acquired by
  - applying HR1 and considering the synonymy or hyponymy relationship between NOBJ and NOBJ' in HR1.
  - applying HR4.
- **CF = 0.8:** The results acquired by applying HR1 and considering the coordinate relationship between NOBJ and NOBJ' in HR1.
- **CF = 0.7:** The results acquired by applying HR2 and considering the previously obtained results whose CF is 0.9 or 0.8 in HR2.
- **CF = 0.6:** The results acquired by applying HR3 and considering the synonym or the hyponym relationship between NOBJ and NOBJ' in HR3.
- **CF = 0.5:** The results acquired by applying HR3 and considering the coordinate relationship between NOBJ and NOBJ' in HR3.

In the algorithm, HR1 is always applied first, then HR2 and finally HR3. Both HR4 and HR5 are applied later than the first three heuristic rules because both HR4 and HR5 are only applicable in few cases. The high semantic similarity between words is always considered earlier than low semantic similarity. The results with higher CF values are usually acquired by the algorithm early.

4 WSD algorithm

In this section, we give a detailed description of the WSD algorithm and illustrate it with examples.

Suppose a noun Wn has n word senses in WordNet. In a text, we are going to decide the intended meaning of the noun object Wn in its verb context Wv or in a verb-object pair (Wv, Wn).

Here, sense(Wn, k) means the kth sense of Wn in WordNet ($0 \leq k \leq 1$); synonymy between words means that one word is in the same synset as another word or is a immediate parent node of another word in WordNet's IS-A hierarchy. Hyponymy between words means that one word is a child node of another word. A coordinate relationship means that one word is a sibling node of another word. The confidence in a result is defined as a number CF ($0 \leq$ CF ≤ 1).

<u>WSD Algorithm</u>

The algorithm attempts to disambiguate a given word in eight ways by going through eight steps. If any step succeeds, the remaining ones are skipped.

STEP 1

Search for Wn in WordNet. If Wn only has one sense, sense(Wn, 1), in WordNet, the meaning of Wn in its verb context Wv is sense(Wn, 1). The confidence in this result is 1.

For example, suppose Wn = ”income”. It only has one sense in WordNet, denoted sense(income, 1); this sense is given as ”financial gain”. The meaning of ”income” in any of its verb contexts is sense(income, 1). The confidence in this result is 1.

STEP 2

Find a verb-object pair (Wv, Wn') in the parsed text and Wn' is synonymous or hyponymous with sense(Wn, k) $1 \leq k \leq n$. The meaning of Wn in its verb context Wv is sense(Wn, k). The confidence in this result is 0.9.

For example, suppose Wn = ”contribution” and Wv = ”calculate”. ”contribution” has 5 senses in WordNet. A verb-object pair ”calculate amount” can be found in the text and ”amount” is synonymous with sense(contribution, 3); this sense is given as ”an amount of money contributed”. The meaning of ”contribution” in its verb context ”calculate” is sense(contribution, 3).

STEP 3

The same as STEP 2, but with a coordinate relationship instead of synonymy or hyponymy between Wn' and sense(Wn, k) $1 \leq k \leq n$. The confidence in this result is 0.8.

For example, suppose Wn = ”expense” and Wv = ”subtract”. ”expense” has 2 senses in WordNet. A verb-object pair ”subtract grant” can be found in the text and ”grant” has a coordinate relationship with sense(expense, 1); this sense is given as ”a possession whose ownership has changed”. The meaning of ”expense” in its verb context ”subtract” is sense(expense, 1).

STEP 4
Find in the parsed text a verb-object pair (Wv', Wn) in which Wv' has a synonymy, hyponymy or coordinate relationship with Wv and the word sense of Wn in its verb context Wv' has already been acquired (suppose it is sense(Wn, k) $1 \le k \le n$). The meaning of Wn in its verb context Wv is also sense(Wn, k). The confidence in this result is 0.7.

For example, suppose Wn = "contribution" and Wv = "prorate". "contribution" has 6 senses in WordNet. A verb-object pair "calculate contribution" can be found in the text and "calculate" has a synonym relationship with "prorate". The meaning of "contribution" in its verb context "calculate" has already been acquired (it is sense(contribution, 3), see the example in STEP 2). The meaning of "contribution" in its verb context "prorate" is also sense(contribution, 3).

STEP 5
Find in the parsed text a verb-object pair (Wv', Wn') in which Wv' has a synonymy, hyponymy or coordinate relationship with Wv and Wn' has a synonym or hyponym relationship with sense(Wn, k) $1 \le k \le n$. The meaning of Wn in its verb context Wv is sense(Wn, k). The confidence in this result is 0.6.

For example, suppose Wn = "deduction" and Wv = "calculate". "deduction" have seven senses in WordNet. A verb-object pair "subtract allowance" can be found in the text. "subtract" is synonymous with "calculate" and "allowance" is synonymous with sense(deduction, 3); this sense is given as "an amount or percentage deducted". The meaning of "deduction" in its verb context "calculate" is sense(deduction, 3).

STEP 6
The same as STEP 5, but with a coordinate relationship instead of synonymy or hyponymy between Wn' and sense(Wn, k) $1 \le k \le n$. The confidence in this result is 0.5.

For example, Wn = "investment" and Wv = "list". "investment" has 3 senses in WordNet. A verb-object pair "enter credit" can be found in the text. "enter" has a coordinate relationship with "list" and "credit" has a coordinate relationship with sense(investment, 1); this sense is given as "any valuable or useful possession". The meaning of "investment" in its verb context "list" is sense(investment 1).

STEP 7
Find in the parsed text a "such as" structure that follows the verb-object pair (Wv, Wn) in the text and the object Wn' of this "such as" structure is synonymous or hyponymous with sense(Wn, k) $1 \le k \le n$. The meaning of Wn in its verb context Wv is sense(Wn, k). The confidence in this result is 0.9.

For example, suppose Wn = "property" and Wv = "sell". "property" has 5 senses in WordNet. A structure "such as real estate" can be found in the text which follows "sell property" in the text (that is, "... sell property, such as real estate ...") and the object "real estate" of this "such as" structure is hyponymous with sense(property, 1); this sense is given as "any tangible possession that is owned by someone". The meaning of "property" in its verb context "sell" is sense(property, 1).

STEP 8
Find in the parsed text a coordinate verb phrase structure "Wv and Wv' " or "Wv or Wv' " whose noun object is Wn and the meaning of Wn in its verb context Wv' has already been acquired (suppose it is sense(Wn, k) $1 \le k \le n$). The meaning of Wn in its verb context Wv is also sense(Wn, k). The confidence in this result is the same as for the verb context Wv'.

For example, suppose Wn = "property" and Wv = "dispose of". "property" has 5 senses in WordNet. In the text, a VP coordinate structure " dispose of or sell" whose noun object is "property" has been found (that is, there is "... dispose of or sell property ..." in the text) and the meaning of "property" in its verb context "sell" has already been acquired (it is sense(property, 1) with the confidence 0.9, see the example in STEP 7), then the meaning of "property" in its verb context "dispose of" is also sense(property, 1). The confidence in this result is also 0.9.

5 Evaluation

The WSD algorithm has been implemented in Prolog and tested on the Canadian Income Tax Guide. This text contains 1797 sentences in which there are 593 different verb-object pairs. These pairs employ 173 different noun-objects, so the average noun-object appears in the text in more than three verb contexts. There are 70 noun-objects which have three or more verb contexts in the text. Among such 70 noun-objects, 60 noun-objects have more than one sense in WordNet.

We have carried out a post hoc evaluation of the results, based on a manual assessment of the algorithm's output. Another approach, which we intend to use in further tests of the WSD algorithm in addition to post hoc analysis, is to mark nouns in a portion of text in advance and measure the number of matches.

The evaluation of the WSD algorithm has been done on a test of 397 cases: we determined the intended meanings of 60 noun-objects in 397 different verb contexts.

The WSD program may give five possible results:

One correct solution – one reasonable meaning of the noun object has been selected by the WSD algorithm.

For example, for the pair "report loss", only one sense among 6 candidate word senses in WordNet, sense(loss, 2), has been selected by the WSD algorithm as the intended meaning of "loss" in the verb context "report".

Sense 2
loss – (failure to retain possession; "the company

wrote off its losses")
⇒ transferred property, transferred possession

Correct multiple solutions – more than one reasonable meaning of the noun object has been selected by the WSD algorithm. Because some verbs cannot put strong restrictions on their objects, it is possible that several word senses of a noun object are all reasonable in the verb context.

For example, for the verb-object "claim charge", among 13 candidate word senses in WordNet, two word senses of "charge" have been selected by the algorithm as the intended meaning of "charge" in the verb context "claim": sense(charge, 1) and sense(charge, 2). In fact, both are reasonable in the verb context "claim".

Sense 1
charge – (a financial liability; such as a tax)
⇒ liability, financial obligation, indebtedness, pecuniary obligation

Sense 2
charge – (the price charged for some article or service)
⇒ cost, price, terms, damage

Partially correct solutions – more than one word sense of the noun object has been selected by the WSD algorithm. Among those, at least, there is the correct word sense that is reasonable in the verb context.

For example, for the verb-object pair "attach receipt", among three candidate word senses, two word senses of "receipt" have been selected by the WSD algorithm as the intended meaning of "receipt" in the verb context "attach": sense(receipt,2) and sense(receipt,3). In fact, only sense(receipt, 2) is reasonable in the verb context "attach".

Sense 2
receipt, acknowledgment, acknowledgement
⇒ commercial document, commercial instrument

Sense 3
reception, receipt, receiving – (the act of receiving)
⇒ acquisition, acquiring, taking

Wrong solution – the meanings of the noun object selected by the WSD algorithm are unreasonable for its verb context.

No solution – no result has been acquired by the algorithm. For some cases, if no relevant information for supporting the WSD process of these cases can be found in the text (e.g. there is no sufficient information in the text), the algorithm would not produce any solution.

The following statistical data have been obtained by running the WSD program to deal with the 397 cases:

all cases (397 different verb-object pairs)	397	
one correct solution	225	57%
multiple solutions	59	15%
partially correct solutions	20	5%
wrong solution	17	4%
no solution	76	19%

This test result for the WSD algorithm is encouraging. With human's assessment, the word sense disambiguation has a high accuracy of up to 72% (one correct solution and correct multiple solutions). Among the remaining 28% cases, there are only 4% wrong solutions. The 5% cases with partially correct solutions show how word sense ambiguity has been improved by the WSD algorithm.

Of course, on the other hand, with 15% correct multiple solution cases, this WSD algorithm is obviously limited. In these cases, a verb context, for example, "include" or "have", is not strong enough to limit the multiple word senses of its noun object to the only intended meaning. In addition, even for some single-solution cases, the only intended meaning obtained by the algorithm is suitable in the verbal context considered by our algorithm, but is not necessarily appropriate for the whole text. In order to produce the only intended meaning of the noun objects in the whole text, contexts other than just the verb context should be exploited in the future versions of our algorithm.

There are 19% no-solution cases in the test. This means that 593 background verb-object pairs in the text are not sufficient to support word sense disambiguation of these cases. The WSD algorithm depends on the relevant information in a text, so the more relevant information in the text, the better the WSD result. If a text is short of relevant information for supporting WSD process, a good idea is that instead of using relevant information from the text, the WSD algorithm works based on some supporting information from a large corpus (that is, with a large amount of verb-object pairs from a large corpus as its background knowledge).

The sensitivity of results to the specific setting of CFs used has been noticed in the experiment as well. The following results indicate that indeed the quality decreases with the decreasing values although maybe the actual setting could be changed, as long as its values are in roughly the same decreasing order for HR1 - HR5.

	total	wrong cases	error rate
$0 \leq$ CF < 1	397	17	4%
CF=0.9	138	1	0.7%
CF=0.8	78	4	5.1%
CF=0.7	83	9	10.8%
CF=0.6	11	0	0%
CF=0.5	11	3	27%

These data have very reasonably shown that a solution with low confidence generally is more likely to be a wrong solution. This means that the assignment of confidence values in the WSD algorithm is quite acceptable.

6 Conclusions

In this paper, we have presented a WordNet-based algorithm for word sense disambiguation. The algorithm is designed to support text analysis with minimal pre-coded knowledge. Although the algorithm is assumed to aim at word sense disambiguation of noun objects in a text, in fact it can be easily transformed to cover some other parts of speech in a text.

The approach focuses on two parts: the full utilization of the important relationships between words in WordNet and the exploration of WSD heuristic rules based on the semantic similarity between words. The experiment described here demonstrates the viability of the WSD algorithm, the rationality of the confidence value assignment and show the potential of a WordNet-based method. The experiment also shows the limitations of considering only verb contexts in the WSD process.

Future work on WSD will focus on investigating the possibility of the involvement of more complex context in the WSD process and considering the effective combination between the current result with verb contexts and the possible future result with some other contexts.

References

[Beckwith *et al.*, 1991] Beckwith, R., C. Fellbaum, D. Gross and G. Miller, "WordNet: A Lexical database Organized on Psycholinguistic Principles", In *Lexical Acquisition: Exploiting On-line Resources to Build a Lexicon.* Lawrence Erlbaum Associates, 1991.

[Brown *et al.*, 1991] Brown, P. F., S. A. Della Pietra, V. J. Della Pietra and R. L. Mercer, "Word-Sense Disambiguation Using Statistical Methods", In *Proc ACL Meeting*, Berkeley 1991, 264-270.

[Charniak, 1993] Charniak, E., *Statistical Language Learning.* MIT Press, 1993.

[Gale *et al.*, 1992] ale, W. A., K. W. Church and D. Yarowsky, "A Method for Disambiguating Word Senses in a Large Corpus", In *Computers and the Humanities*, 1992.

[Grefenstette and Hearst, 1992] Grefenstette, G. and M. Hearst, "A Method for Refining Automatically Discovered Lexical Relations: Combining Weak Techniques for Stronger Results", In *AAAI Workshop on Statistically-Based NLP Techniques*, San Jose, July 1992.

[Hirst, 1987] Hirst, G., *Semantic interpretation and the Disambiguation of Ambiguity* Cambridge University Press, England, 1987.

[Lehman, 1994] Lehman, J. F., "Toward the Essential Nature of Statistical Knowledge in Sense Disambiguation", In *Proc AAAI-94*, 1994, 734-741.

[Miller, 1990] Miller, G., "WordNet: An On-line Lexical Database", In *International Journal of Lexicography*, Vol. 3, No. 4, 1990.

[Resnik, 1992] Resnik, P., "WordNet and Distributional Analysis: A Class-based Approach to Lexical Discovery", In *AAAI Workshop on Statistically-Based NLP Techniques*, San Jose, July 1992.

[Resnik, 1993] Resnik, P., "Semantic Classes and Syntactic Ambiguity", In *Proc. of the ARPA Workshop on Human Language Technology.* Princeton, 1993.

[Schutze, 1992] Schutze, H., "Word Sense Disambiguation With Sublexical Representations", In *AAAI Workshop on Statistically-Based NLP Techniques*, San Jose, July 1992, 109-113.

[Small and Rieger, 1982] Small, S. L. and Rieger, G. J., "Parsing and Comprehending with Word Experts (a theory and its realization)", W. G. Lehnert and M. H. Ringle, (ed.), In *Strategies for Natural Language Processing*, Lawrence Erlbaum Associates, 1982, 89-147.

[Sussna, 1993] Sussna, M., "Word Sense Disambiguation for Free-text Indexing Using a Massive Semantic Network", In *CIKM'93*, 1993.

[Voorhees, 1993] Voorhees, E. M., "Using WordNet to Disambiguate Word Sense for Text Retrieval", In *Proc ACM SIGIR'93*, Pittsburgh, 1993, 171-180.

[Wilks., 1975] Wilks, Y. A., "A Preferential, Pattern-seeking Semantics for Natural Language Inference", In *Artificial Intelligence.* Vol. 6, 1975.