自然言語処理 116−8
（1996. 11. 18）

PP AMBIGUITY RESOLUTION THROUGH SEMANTIC MATCHING

Jiri Stetina Makoto Nagao

Department of Electrical Engineering II
Kyoto University, Yoshida Honmachi, Kyoto 606, Japan
stetina@pine.kuee.kyoto-u.ac.jp, nagao@pine.kuee.kyoto-u.ac.jp

ABSTRACT

One of the crucial ambiguity problems in natural language processing is the prepositional phrase (PP) attachment resolution. In order to decide whether a prepositional phrase is adjectival or adverbial, we need to simultaneously employ different types of knowledge. This paper shows how a robust semantic hierarchy can facilitate the process of correct prepositional phrase attachment. We propose two semantic based methods both of which utilise such a semantic hierarchy.

The first method is based on a calculation of the semantic distance between an unseen candidate for PP attachment and the matching example in the training corpus. The second method belongs to the family of inductive learning algorithms and derives a decision tree which then serves as an universal mechanism for the PP attachment resolution. We present the results comparing these methods with other corpus based approaches and discuss their feasibility in solving other NLP related problems.

1. INTRODUCTION

The problem with successful resolving of PP attachment is that we need to employ various types of knowledge. Consider, for example, the following sentence:

1. *Buy books for children.*

The PP "for children" can be either adjectival and attach to the object noun "books" or adverbial and attach to the verb "buy", leaving us with an ambiguity of two possible syntactic structures:

adj) VP(VB=buy, NP(NNS=books, PP(IN=for,NP(NN=children))))
adv) VP(VB=buy, NP(NNS=books), PP(IN=for,NP(NN=children))).

It is obvious that without any additional information we cannot disambiguate such a sentence correctly. We need some sort of world knowledge that would tell us, which situations are more likely than the others. Sometimes, however, world knowledge on its own is not sufficient for disambiguation. The reason is, that both of the syntactic structures may be plausible. The correct disambiguation also depends on the context in which the sentence appears, or on combination of both. Unfortunately, up to date there is not enough theoretical linguistic background to determine what is contextually related to the current ambiguity and what type of general knowledge and at what level of inference has to be applied to disambiguate correctly. Therefore, we will focus only on stand-alone situations and disregard any contextual dependency. The problem we are facing can be then defined as deciding whether a PP is attached to the verb or to the noun, given the quadruple:

(v)verb (o)object (p)preposition (d)description.

Even without any surrounding context many PP's can still be correctly attached. The human performance, which is about 88.2% accurate on stand-alone quadruples [RRR84], can be taken as the upper limit of any algorithm performing the same task[1]. Each preposition has certain percentage of occurrences for each attachment, relying on which would put us at around 72.7% of correct attachments [C&B95]. We are therefore aiming at the range between 72.7 and 88.2%, in order to consider our approach intelligent.

2. BACKGROUND

Altman and Steedman [A&S88] have shown that in many cases PP can be attached correctly only if the context of the current discourse is used. Using the discourse context is, however, extremely difficult because we do not have enough theoretical background to decide which bits of context are needed to correctly disambiguate and which are irrelevant[2]. There have been numerous attempts to substitute context by superficial knowledge extracted from a large corpus [H&R93]. Most of the methods are based on statistical frequencies and require storing of a huge table of probabilities which turns computationally very expensive at run-time. Promising approach is the transformation-based rule derivation presented by Brill and Resnik in [B&R94] which is a simple learning algorithm that derives a set of transformation rules. These rules are then used for PP attachment and therefore, unlike the statistical methods, it is not necessary to store huge frequency tables.

BACKED-OFF MODEL

[C&B95] proposed a new statistical method for PP attachment disambiguation where a decision on the attachment depends on the frequency of similar situations for both cases in a training corpus. Using a huge set of classified quadruples, the algorithm first finds the frequencies at which a full quadruple match[3] occurs for each attachment class (adjectival or adverbial). The PP is then attached to the class with a higher frequency of matches. If a full match is not found, which happens more likely, the algorithm looks for the partial matches, i.e. backs-off to matches on three words. Occurrences of combined matches of (v-n-p),(v-p-d) and (n-p-d) are counted for each attachment class and the PP is attached to the class with a higher match frequency. If again, no match is found in neither class, the algorithm backs-off to doubles, i.e. combined occurrences of matches of (v-p),(n-p),(p-d) and assigns the class accordingly. If a match is not found again, the PP is attached according to the class probability of the given preposition. If a

[1] Average human performs about 93.2% correctly on full sentences which shows that sometimes we also need the context beyond the one sentence.

[2] At this point we can also view general knowledge as a kind of a context which stands above the context of the discourse.

[3] Full quadruple match is a situation where all four words in the currently classified quadruple are the same as all four words in the example from the training set.

preposition does not occur in the training set, the algorithm chooses adjectival attachment as a default, because it has been statistically proven that about 59% out of all PP attachments are adjectival.
[C&B95] tested the algorithm using 20801 sentences for training and 3097 sentences for testing. 84.5% of the testing sentences were classified correctly which is appreciably better result than any other method tested on the same data. A particularly surprising was the significance of low count matches in data but at the same time the method performs badly when there is no match found in the training data. Another problem is that for each PP to be attached, the program has to either go through an entire training set in order to count the match frequencies (which causes inevitable delays at run-time), or has to store a huge frequency table in memory.

RULE-BASED APPROACH

Brill and Resnik [B&R94] proposed another corpus-based approach, that partially avoids these deficiencies. Their rule-based method derives a set of transformation rules that are then used for classification of unseen quadruples. As the method gave promising results, let us describe it briefly.
Transformation based learning is a simple algorithm that has been applied to a number of natural language problems [BR92], [BR93], [BR94]. In the first step, all samples from the training set are classified as adjectival as would be predicted by right association [KI73]. The allowable transformations are in the form:

CHANGE THE ATTACHMENT FROM X TO Y IF:
v = Wln = Wlp = Wld = Wlv = W1 and n = W2 ... etc.,

where 'X to Y' indicates 'adjectival to adverbial' or vice-versa. In the next step, each possible transformation is tried on entire training set and a reduction/increase of the error rate is calculated. The best scoring transformation becomes the first derived transformation in the transformation list. It is applied to the training set and learning continues on the modified corpus. The process stops when no rule that would reduce the error rate is found. In this way, Brill and Resnik had derived a total of 471 transformations from a training corpus of 12766 sentences. These rules tested on 500 randomly selected test sentences provided the accuracy of 80.8% (both training and testing data were different from [C&B95]. Similarly to the backed-off model, this method suffers from sparse data problem as only exactly same words can be matched. A modification which allows matches on words that belong to the same semantic class, rather than exactly the same words(strings of characters), has lowered the number of transformations to 266 and improved the accuracy to 81.8%. Its main advantage, however, is the fact that only a small set of transformation rules is needed for future classifications.

2.1 HIERARCHICAL APPROACH

Both of the above described algorithms suffer from sparse data problem. Both are based on matching the words from the analysed sentence against the words in the training set. The problem is that only exact matches are allowed. The backed-off model showed an overall accuracy of 84.5%, but the accuracy of full quadruple matches was 92.6%. Due to the sparse data problem the full quadruple matches were quite rare and contributed to the result only in 4.8% of cases. Brill and Resnik had allowed for matches of the words belonging to the same semantic class as defined in WordNet [MI90], but in many situations such a generalisation was not sufficient.
An obvious solution to the data sparseness problem would be to allow for matches of words that are not only synonyms, but also of the semantically close words. The question is, however, what is the limit distance for two concepts to be matched. Many nouns in the WordNet hierarchy share the same root (entity) and there is a danger of over-generalisation. Another problem is that most of the words are semantically ambiguous and unless disambiguated, it is difficult to establish distances among them. Let us return to these problems later.

2.2 UTILISATION OF SEMANTIC HIERARCHY

We will now discuss the issues connected with matching two different words based on their semantic distance. Employing the notion of semantic distance, we have to address several problems. At first, we have to specify the semantic hierarchy. Second, we have to tackle the problem of semantic ambiguity. And, finally, we have to determine how to calculate a distance between two different concepts in the hierarchy.

2.3 SEMANTIC HIERARCHY

The hierarchy we chose for semantic matching is the semantic network of WordNet [MI90], [MI93]. WordNet is a network of meanings connected by a variety of relations. WordNet presently contains approximately 95,000 different word forms organised into 70,100 word meanings, or sets of synonyms. It is divided into four categories (nouns, verbs, adjectives and adverbs), out of which we will be using only verbs and nouns. Nouns are organised as 11 topical hierarchies, where each root represents the most general concept for each topic. Verbs, which tend to by more polysemous and can change their meanings depending on the kind of the object they take, are formed into 15 groups and have altogether 337 possible roots. Verb hierarchies are more shallow than those of nouns, which is because nouns tend to be more easily organised by the is-a relation while it is more difficult for verbs[4].

2.4 SEMANTIC AMBIGUITY

In order to determine the position of a word in the semantic hierarchy, we have to know what sense or what meaning of the word we are talking about. For example, the noun '*bank*' can take any of the nine meanings defined in WordNet (*financial institution, building, ridge, container, slope*, etc.). It is not a trivial problem and has been approached by many researchers. We will show that word sense disambiguation can be accompanied by PP attachment resolution simultaneously and that they complement each other. At the same time we would like to note, that PP attachment and sense disambiguation are heavily context dependent problems, therefore we know in advance that without incorporation of context the full disambiguation will be never reached.

2.5 SEMANTIC DISTANCE

The traditional method of evaluating semantic distance between two meanings based on the length of the path between the nodes representing them, does not work well in WordNet. It is because the distance of two meanings depends also on other factors. For example, the root entity is directly followed by the concept of *life_form*, while a *sedan*, a kind of a car, is in terms of path more distant from the concept of *express_train*, although they are both vehicles and therefore semantically closer. The distance in this case also depends on the depth at which the concepts appear. In the case of verbs, the situation is even more complex, because many verbs do not share the same hierarchy and therefore there is no direct path between the concepts they represent. There has been numerous attempts to define a measure for semantic distance of WordNet contained concepts [RE96], [SU95], [SU96], etc. We have based our semantic distance calculation on a combination of the path distance between two nodes and their depth. Having found the nearest common ancestor in the hierarchy, the distance is calculated as:

$$D = L1/D1 + L2/D2$$

where L1, L2 are the lengths of paths between the concepts and the nearest common ancestor, and D1, D2 are the depths of each concept in the hierarchy (the distance to the root). The more abstract the nearest common ancestor (the higher in the hierarchy) is, the bigger the distance. If the concepts do not have a common ancestor, the distance calculation fails.

3. SEMANTIC MATCHING

Having defined the semantic distance, we tried to improve the backed-off algorithm [C&B95] by increasing the number of matches on full quadruples. The original form of the backed-off algorithm, suffering from a sparse data problem, found a full quadruple match in only 4.8% of tested quadruples. Because the percentage of correct attachments among theses matches was the highest (92.6%), we tried to raise the number of matched quadruples by matching also words with small semantic distance.

Because both the training and the testing quadruple can have multiple meanings for each content word (verb, noun, description), we tried to match all possible senses of each word with all possible senses of its correspondent word in the sample[5]. If the distance of any of two senses of the two words was below a pre-set threshold, the words were matched. If all three content words were matched with their counterpart in the training set, a full match was scored. Using the same training and testing data supplied by [C&B95], we managed to raise the number of full matches substantially. The number of matched quadruples depended on the empirically set distance threshold and is shown in Table1.

When the algorithm didn't find a match on all four words, similarly to its original counterpart, it backed-off to a match on less words. Our expectation was that by raising the number of fully matched quadruples, the overall rate of correctly attached PP's would surpass the original method. However, although close, the overall result, did not exceed the correct percentage of the version that allowed for word-strings matches

---

[4] Alhough people are comfortamble with statements like 'a horse is an animal', they are likely to reject such statements as 'mumbling is talking'.

[5] Because the preposition is particularly important to the attachment decison, only matches of quadruples with the same preposition were allowed.

only.

Table 1. *Match count threshold dependency and overall accuracy*

THRESHOLD	FULL MATCHES	PERCENT	OF THOSE CORRECT	OVERALL ACCURACY
same words	148	4.8 %	92.6 %	84.5 %
0 (synonyms)	487	15.8 %	91.8 %	84.4 %
0.3	734	23.2 %	89.7 %	83.1 %
0.5	900	29.1 %	88.9 %	82.3 %
1.0	1246	40.2 %	84.3 %	81.1 %

We explain this in the following way:

- Both the testing and the training data contained many unknown words, names, abbreviations, etc., which are not registered in WordNet. Such situations contributed to lower match count[6].
- Most of the words in both the testing and the training data were ambiguous. Since a match was scored when the distance of any of the senses of the two words was below a pre-set threshold, even inappropriate senses contributed to the match. We therefore tried to modify the method by choosing only the senses that appeared in the top 5% of the matches for the given quadruple. This partially eliminated this deficiency, but what is more important, also provided us with sense disambiguation of the words in the tested quadruple[7].
- Another problem with multiple senses is in its combinational ambiguity. An unseen quadruple v1-o1-p-d1 could be matched with both adjectival and adverbial training sample, when in each case, different senses of its words are selected.
- Setting the same distance threshold for all the words in the quadruple obviously does not provide a good selective measure. In one case, the distance of the verbs may be more important than that of nouns, in the other it may be the similarity of descriptions that is more important. Similarly, in one case a concept can be replaced by a more general substituent, in another case it is the deep specialisation that determines the attachment. As we believe that this is the most important point in resolving the PP attachment, we have devised the following method that automatically uncovers how and when each concept can be generalised.

4. LEARNING HOW TO ATTACH PP

An algorithm that seems to overcome the above mentioned problems belongs to the family of inductive learning algorithms [KO88]. Using a huge training set of classified examples it uncovers the importance of the individual attributes (in our case:verb, noun, preposition and description) and creates a decision tree that is later used for classification of unseen examples[8]. The algorithm, a modification of the ID3 algorithm [BR91], creates the decision tree in the following way:

Let T be a training set of classified quadruples.
1. If all the examples in T belong to the same class then result is a leaf labelled with this class, **else**
2. Select the most informative attribute A among verb, noun and description.
3. For each possible value Av of the selected attribute construct recursively a subtree Sv calling the same algorithm on a set of quadruples for which A=Av.
4. Return a tree whose root is A and whose subtrees are Sv and links between A and Sv are labelled Av.

Let us briefly explain each step of the algorithm.
1. If the examples belong to the same class *Cr* (set **T** is homogenous), the tree expansion terminates. Such

[6] Eliminating the words that are not present in WordNet would certainly raise the accuracy, but would not allow for the comparison with the original method.
[7] Whether the top 5% of matches is an appropriate figure has to be determined by further testing. Preliminary results, however, are very promising and leave an open space for future experiments.
[8] Classification in this case means deciding whether the PP is adjectival or averbial.

situation is, however, very unlikely due to the non-perfect training data. Therefore, we relaxed the complete homogenity condition by terminating the expansion if more than 90% of the examples in the set belonged to the same class. If the set T is still hetereogenous and there are no more attribute values to divide with, it is labelled by SEARCH, tree is terminated and statistical classification has to be applied.

2. We consider the most informative attribute to be the one which splits the set **T** into possibly most homogenous subsets, i.e. subsets with either high percentage of samples with adjectival attachments and low of adverbial ones, or vice-versa. The optimal split would be such that all the subsets would contain only samples of one attachment type. For each attribute **A** we split the set into subsets each associated with attribute value **Av** and containing samples which were unifiable with the value **Av** (see the next point). Then, we calculated the overall heterogenity (OH) of all these subsets as a weighted sum of their expected information:

$$OH = -\sum_{v} p(A = Av) \sum_{r} p(Cr|A = Av) \log_2(p(Cr|A = Av))$$

The attribute with the lowest overall heterogenity was selected.

3. In our case, the attribute is either a verb, noun, or a description[9]. The values of these attributes are the concept identificators (synsets) associated with each attribute (word). In case of multiple meanings, attribute values are the lists of all possible synsets for the given word. When splitting the set by the attribute **A** according to its values **Av**, the emerging subsets contain those quadruples whose attribute **A** value of which was lower in the WordNet hierarchy, i.e. belonged to the same class[10].

4.1 CLASSIFICATION

As soon as the decision tree is induced, classifying an unseen quadruple is quite simple. A path is traversed in the tree starting at its root and ending at a leaf. At each internal node, we follow the branch labelled by the attribute value of the quadruple, or the value more general in terms of WordNet hierarchy. The quadruple is assigned the class associated with the leaf, i.e. either adjectival or adverbial attachment. If no match is found for the attribute value of the quadruple at any given node, the closest attribute value is chosen for selection (using the semantic distance calculation described above). If the semantic distance calculation fails, i.e. if the attribute of the quadruple does not have a corresponding value in the tree, classificator descends one level in the tree and tries to match according to the attribute at the next level. If a match is not found for any attribute or when a leaf labelled by SEARCH is reached, the quadruple is classified according to the statistic probability of its preposition.

4.2 TRAINING AND TESTING DATA

Because the method is based on a supervised training, first we have to obtain a representative set of training quadruples. The PP's of these quadruples have to be marked either adjectival or adverbial, in order to enable learning. We therefore needed a syntactically analysed corpus from which we extracted the training quadruples. For this purpose we used the analysed Brown Corpus of the Penn Treebank [MA93] from which we extracted 11,892 training samples. We extracted the testing quadruples also from the Penn Treebank, but this time from its Wall Street Journal corpus to ensure that there was no implicit training of the method on the test set itself. In this way we obtained 4,032 test samples of known attachment used for future accuracy evaluation. We converted each word in the testing and the training set into its base form and looked up its corresponding set of possible senses (synsets) in WordNet. The training samples with words not in WordNet were discarded because such samples would have only a limited value for learning which is very dependent on the representativness of the training set. Then, we tried to eliminate inappropriate senses and thus reducing the semantic ambiguity by applying semantic constraints as described briefly in the next section.

---

[9] We induce the decison tree separately for each preposition.

[10] If some quadruples had the attribute value equal to the values of A, additional subset was added but its further splitting by the same attribute was prohibited.

4.3 SEMANTIC AMBIGUITY REDUCTION

Both during the construction of the decision tree and when classifying unseen quadruples we were faced with the sense ambiguity. The more meanings a word has the bigger the ambiguity and the more difficult and computationally expensive the processing becomes. Therefore, we have tried to reduce the space of word meanings whenever possible using the following semantic constraints: As mentioned before, verbs tend to by more polysemous and can change their meanings depending on the kind of the object they take. As the first step, we have dismissed all the intransitive verb senses. Next, using the frame information of WordNet[11] we have constrained the senses of the verbs according to the object in the quadruple. Some senses take only human objects, some only non living things. By matching the verb-object pairs with the frame information, we have substantially reduced the sense ambiguity of verbs and their objects.

5. RESULTS AND COMPARISON WITH OTHER WORK

For testing purposes, we selected the preposition *'for'* because it has almost equal distribution over both attachment classes. We have used 1815 quadruples (922 adjectival and 893 adverbial) containing the preposition *'for'* extracted from the Brown Corpus for training and 341 quadruples for testing.
Although we applied semantic constraints to limit the semantic ambiguity of individual words, many of them remained quite ambiguous. In order to overcome this ambiguity problem, we applied two approaches. At first, we used only the most frequent senses of the words which led to a decision tree with nodes marked by single synsets. In the second approach, we interpreted each word as a list of its possible meanings, which led to a tree whose nodes were marked by synset groups. These groups were the conjunctions of attribute values of all the quadruples belonging to the given node. Classification in the first case was straightforward since all that had to be done was checking whether the attribute in the classified quadruple belongs to the same class (in WordNet terms) as the attribute value of each node on the classification path. In the second case, however, we had to check every sense in the quadruple attribute against every sense in the node attribute. During the decision tree construction, this led to situations when the number of the samples in the subsets was higher than the number of the samples in the originating node[12]. The decision on the PP attachment therefore became dependent on which sense of the word was selected at each node. This may seem as a failure to determine the correct attachment, however, the algorithm unveiled the importance of the sense of the given attribute for the attachment. This shows how closely connected the PP attachment is with the sense disambiguation and that one cannot be achieved without the other. This mechanism would have a great value in a more complex natural language processing system where information from PP attachment module would be used for sense disambiguation and vice-versa.
Table 2 shows the accuracy results for both approaches compared with the statistical backed-off algorithm and its modification with semantic matching tested on the same data:

Table 2. *Accuracy comparisons*

METHOD	TOTAL NUMBER OF QUADRUPLES	Non-ambiguously/ Ambiguously Attached	CORRECTLY ATTACHED	PERCENT CORRECT
First sense only	341	341/0	245	71.8
Multiple senses	341	140/201	112/(159)	30.8 [80/79[79.5
Backed-off	341	341/0	235	69.4
Semantic matching	341	341/0	227	66.6

As the table shows, selecting the correct word sense has a crucial importance on the correctness of the PP attachment. In the multiple senses approach, only 140 quadruples were attached unambiguously, the correct attachment percentage of which was by far the best (112=80%) We tried to reduce the overall low accuracy of this approach (30.8%) by manually tagging the senses of the words in the quadruples that were attached ambiguously, thus gaining unambiguous attachment for these as well. After manual assignment of senses these remaining 201 quadruples were attached with 79% accuracy (159), thus gaining the best overall accuracy of all tested methods (79.5%).

---

[11] WordNet divides the verb senses into 35 frame categories [MI93].

[12] We are currently testing different tree pruning techniques in order to improve the accuracy and reduce the number of ambiguous attachments.

6. CONCLUSIONS

Apart from achieving higher accuracy than any other known method, the main advantage of our approach lays in unveiling the importance of individual attributes for the PP attachment in the form of a decision tree. Compared with the statistical methods, once the decision tree is induced, the classification of unseen quadruples is easy and does not require storing of huge frequency tables. Therefore, it is more useful for real-time systems where a decision on the attachment is computationally very easy. The method is less sensitive to spare data problem (thanks to the generalisation of words through WordNet hierarchy), but partly suffers from dependency on sense disambiguation. Also, it is more dependent on the quality of the training examples. We believe that a manual selection of testing samples would gain more predictive decision trees. Similarly to the backed-off model, even our learning algorithm enables to evaluate the certainty of decision by calculating the heterogenity of the terminal nodes. Another merit is in the possibility of incorporating the decision trees in a complete language processing system and use its result in constraining other types of ambiguities (e.g. in sense disambiguation, POS tagging, anaphora resolution, etc.). Apart from noting the advantages of our method, we have to note that so far we have tested only one preposition(for). Induction of decision trees for other prepositions and their subsequent testing is still necessary.

7. FUTURE WORK

At this moment it is obvious that further testing has to be done to improve the accuracy of the decision trees. More prepositions have to be tested on more testing data and more representative training data selected. Particularly, we would like to test the method on a semantically tagged corpus, once available. Also, different techniques to handle ambiguity and decision tree pruning have to be experimented with. We would also like to examine the possibility of incorporating the wider sentential context in terms of more attributes participating on the decision tree construction (verb subjects, adverbs, adjectives of the object nouns and of the descriptions, etc.).

The algorithm described above works in the top down pattern. We would like to experiment with bottom-up approach, starting with individual samples as individual nodes and joining them upwards by generalising over their attributes.

Our ultimate goal is the implementation of the PP attachment module within a parser based on a simultaneous disambiguation through processing constraint from other modules.

REFERENCES

[A&S88] G. Altmann, M. Steedman. Interaction with Context During Human Sentence Processing. *Cognition, 30:191-238,* 1998

[H&R93] D. Hindle, M. Rooth. Structural Ambiguity and Lexical Relations. *Computational Linguistics, 19(1)103-120,* 1993.

[C&B95] M. Collins, J. Brooks. PP Attahcment Through a Backed-Off Model. *VLC 1995*

[RRR94] A. Ratnaparkhi. A Maximum Entrophy Model for PP Attachment. In *Procedings of the ARPA Workshop on Human Language Technology,* 1994.

[BR92] E. Brill. A Simple Rule-Based POS Tagger. In *Proceedings of the 3rd Conference on Applied Natural Language Processing,* 1992

[BR93] E. Brill. Automatic Grammar Induction and Parsing Free Text: A Transformation Based Approach. In *Proceedings of the 31st Meeting of the ACL,* 1993.

[BR94] E. Brill, P. Resnik. A Rule Based Approach to PP Attachment Disambiguation. In *Proceedings of COLING* 1994.

[KI73] J. Kimball. Seven Principles of Surface Structure Parsing in Natural Language. *Cognition,* 2, 1973.

[MI90] G. Miller. Wordnet: an On-Line Lexical Database. *International Journal of Lexicography.*

[MI93] G. Miller. Introduction to WordNet: An On-line Lexical Database. Princeton University. *ftp://clarity.princeton.edu*

[RE96] P. Resnik. Disambiguating Noun Groupings with Respect to WordNet Senses. *Sun Microsystems Laboratories.*

[SU95] M. Sussna. Information Retrieval using Semantic Distance in WordNet. *sussna@cs.ucsd.edu*

[SU96] M. Sussna. Word Sense Disambiguation for Free-text Indexing Using a MAssive Semantic Network. *sussna@cs.ucsd.edu*

[KO88] Y. Kodratoff. Introduction to Machine Learning. *Pitman, London,* 1988

[BR91] I. Bruha. Machine Learning: Empirical Methods. *Sofsem 91,* 1991

[MA93] M. Marcus. Building a Large Annotated Corpus of English: The Penn Treebank. *Association for Computational Linguistics,* 1993