Two Languages Are More Informative Than One *

Ido Dagan
Computer Science Department
Technion, Haifa, Israel
and
IBM Scientific Center
Haifa, Israel
dagan@cs.technion.ac.il

Alon Itai
Computer Science Department
Technion, Haifa, Israel
itai@cs.technion.ac.il

Ulrike Schwall
IBM Scientific Center
Institute for Knowledge
Based Systems
Heidelberg, Germany
schwall@dhdibm1

Abstract

This paper presents a new approach for resolving lexical ambiguities in one language using statistical data on lexical relations in another language. This approach exploits the differences between mappings of words to senses in different languages. We concentrate on the problem of target word selection in machine translation, for which the approach is directly applicable, and employ a statistical model for the selection mechanism. The model was evaluated using two sets of Hebrew and German examples and was found to be very useful for disambiguation.

1 Introduction

The resolution of lexical ambiguities in non-restricted text is one of the most difficult tasks of natural language processing. A related task in machine translation is *target word selection* – the task of deciding which target language word is the most appropriate equivalent of a source language word in context. In addition to the alternatives introduced from the different word senses of the source language word, the target language may specify additional alternatives that differ mainly in their usages.

Traditionally various linguistic levels were used to deal with this problem: syntactic, semantic and pragmatic. Computationally the syntactic methods are the easiest, but are of no avail in the frequent situation when the different senses of the word show

the same syntactic behavior, having the same part of speech and even the same subcategorization frame. Substantial application of semantic or pragmatic knowledge about the word and its context for broad domains requires compiling huge amounts of knowledge, whose usefulness for practical applications has not yet been proven (Lenat et al., 1990; Nirenburg et al., 1988; Chodorow et al., 1985). Moreover, such methods fail to reflect word usages.

It is known for many years that the use of a word in the language provides information about its meaning (Wittgenstein, 1953). Also, statistical approaches which were popular few decades ago have recently reawakened and were found useful for computational linguistics. Consequently, a possible (though partial) alternative to using manually constructed knowledge can be found in the use of statistical data on the occurrence of lexical relations in large corpora. The use of such relations (mainly relations between verbs or nouns and their arguments and modifiers) for various purposes has received growing attention in recent research (Church and Hanks, 1990; Zernik and Jacobs, 1990; Hindle, 1990). More specifically, two recent works have suggested to use statistical data on lexical relations for resolving ambiguity cases of PP-attachment (Hindle and Rooth, 1990) and pronoun references (Dagan and Itai, 1990a; Dagan and Itai, 1990b).

Clearly, statistical methods can be useful also for target word selection. Consider, for example, the Hebrew sentence extracted from the foreign news section of the daily Haaretz, September 1990 (transcripted to Latin letters).

(1) Nose ze mana' mi-shtei ha-mdinot mi-lahtom 'al hoze shalom.

This sentence would translate into English as:

(2) That issue prevented the two countries from signing a peace treaty.

The verb 'lahtom' has four word senses: 'sign', 'seal', 'finish' and 'close'. Whereas the noun 'hoze' means both 'contract' and 'treaty'. Here the difference is not in the meaning, but in usage.

One possible solution is to consult a Hebrew corpus tagged with word senses, from which we would probably learn that the sense 'sign' of 'lahtom' appears more frequently with 'hoze' as its object than all the other senses. Thus we should prefer that sense. However, the size of corpora required to identify lexical relations in a broad domain is huge (tens of millions of words) and therefore it is usually not feasible to have such corpora manually tagged with word senses. The problem of choosing between 'treaty' and 'contract' cannot be solved using only information on Hebrew, because Hebrew does not distinguish between them.

The solution suggested in this paper is to identify the lexical relationships in corpora of the target language, instead of the source language. Consulting English corpora of 150 million words, yields the following statistics on single word frequencies: 'sign' appeared 28674 times, 'seal' 2771 times, 'finish' appeared 15595 times, 'close' 38291 times, 'treaty' 7331 times and 'contract' 30757 times. Using a naive approach of choosing the most frequent word yields

(3) *That issue prevented the two countries from closing a peace contract.

This may be improved upon if we use lexical relations. We consider word combinations and count how often they appeared in the same syntactic relation as in the ambiguous sentence. For the above example, among the successfully parsed sentences of the corpus, the noun compound 'peace treaty' appeared 49 times, whereas the compound 'peace contract' did not appear at all; 'to sign a treaty' appeared 79 times while none of the other three alternatives appeared more than twice. Thus we first prefer 'treaty' to 'contract' because of the noun compound 'peace treaty' and then proceed to prefer 'sign' since it appears most frequently having the object 'treaty' (the order of selection is explained in section 3). Thus in this case our method yielded the correct translation.

Using this method, we take the point of view that some ambiguity problems are easier to solve at the level of the target language instead of the source language. The source language sentences are considered as a noisy source for target language sentences, and our task is to devise a target language model that prefers the most reasonable translation. Machine translation (MT) is thus viewed in part as a recognition problem, and the statistical model we use specifically for target word selection may be compared with other language models in recognition tasks (e.g. Katz (1985) for speech recognition). In contrast to this view, previous approaches in MT typically resolved examples like (1) by stating various constraints in terms of the source language (Nirenburg, 1987). As explained before, such constraints cannot be acquired automatically and therefore are usually limited in their coverage.

The experiment conducted to test the statistical model clearly shows that the statistics on lexical relations are very useful for disambiguation. Most notable is the result for the set of examples for Hebrew to English translation, which was picked randomly from foreign news sections in Israeli press. For this set, the statistical model was applicable for 70% of the ambiguous words, and its selection was then correct for 92% of the cases.

These results for target word selection in machine translation suggest to use a similar mechanism even if we are interested only in word sense disambiguation within a single language! In order to select the right sense of a word, in a broad coverage application, it is useful to identify lexical relations between word senses. However, within corpora of a single language it is possible to identify automatically only relations at the word level, which are of course not useful for selecting word senses in that language. This is where other languages can supply the solution, exploiting the fact that the mapping between words and word senses varies significantly among different languages. For instance, the English words 'sign' and 'seal' correspond to a very large extent to two distinct senses of the Hebrew word 'lahtom' (from example (1)). These senses should be distinguished by most applications of Hebrew understanding programs. To make this distinction, it is possible to do the same process that is performed for target word selection, by producing all the English alternatives for the lexical relations involving 'lahtom'. Then the Hebrew sense which corresponds to the most plausible English lexical relations is preferred. This process requires a bilingual lexicon which maps each Hebrew sense separately into its possible translations, similar

to a Hebrew-Hebrew-English lexicon (like the Oxford English-English-Hebrew dictionary (Hornby et al., 1986)).

In some cases, different senses of a Hebrew word map to the same word also in English. In these cases, the lexical relations of each sense cannot be identified in an English corpus, and a third language is required to distinguish among these senses. As a long term vision, one can imagine a multilingual corpora based system, which exploits the differences between languages to automatically acquire knowledge about word senses. As explained above, this knowledge would be crucial for lexical disambiguation, and will also help to refine other types of knowledge acquired from large corpora[1].

2 The Linguistic Model

The ambiguity of a word is determined by the number of distinct, non-equivalent representations into which the word can be mapped (Van Eynde et al., 1982). In the case of machine translation the ambiguity of a source word is thus given by the number of target representations for that word in the bilingual lexicon of the translation system. Given a specific syntactic context the ambiguity can be reduced to the number of alternatives which may appear in that context. For instance, if a certain translation of a verb corresponds to an intransitive occurrence of that verb, then this possibility is eliminated when the verb occurs with a direct object. In this work we are interested only in those ambiguities that are left after applying all the deterministic syntactic constraints.

For example, consider the following Hebrew sentence, taken from the daily Haaretz, September 1990:

(4) Diplomatim svurim ki hitztarrfuto shell Hon Sun magdila et ha-sikkuyim l-hassagat hitqaddmut ba-sihot.

Here, the ambiguous words in translation to English are 'magdila', 'hitqaddmut' and 'sihot'. To facilitate the reading, we give the translation of the sentence to English, and in each case of an ambiguous selection all the alternatives are listed within curly brackets, the first alternative being the correct one.

---

[1]For instance, Hindle (1990) indicates the need to distinguish among senses of polysemic words for his statistical classification method.

(5) Diplomats believe that the joining of Hon Sun { increases | enlarges | magnifies } the chances for achieving { progress | advance | advancement } in the { talks | conversations | calls }.

We use the term a *lexical relation* to denote the cooccurrence relation of two (or possibly more) specific words in a sentence, having a certain syntactic relationship between them. Typical relations are between verbs and their subjects, objects, complements, adverbs and modifying prepositional phrases. Similarly, nouns are related also with their objects, with their modifying nouns in compounds and with their modifying adjectives and prepositional phrases.

The relational representation of a sentence is simply the list of all lexical relations that occur in the sentence. For our purpose, the relational representation contains only those relations that involve at least one ambiguous word. The relational representation for example (4) is given in (6) (for readability we represent the Hebrew word by its English equivalent, prefixed by 'H' to denote the fact that it is a Hebrew word):

(6) a. (subj-verb: H-joining H-increase)
    b. (verb-obj: H-increase H-chance)
    c. (verb-obj: H-achieve H-progress)
    d. (noun-pp: H-progress H-in H-talks)

The relational representation of a source sentence is reflected also in its translation to a target sentence. In some cases the relational representation of the target sentence is completely equivalent to that of the source sentence, and can be achieved just by substituting the source words with target words. In other cases, the mapping between source and target relations is more complicated, as is the case for the following German example:

(7) Der Tisch gefaellt mir. — I like the table.

Here, the original subject of the source sentence becomes the object in the target sentence. This kind of mapping usually influences the translation process and is therefore encoded in components of the translation program, either explicitly or implicitly, especially in transfer based systems. Our model assumes that such a mapping of source language relations to target language relations is possible, an assumption that is valid for many practical cases.

When applying the mapping of relations on one lexical relation of the source sentence we get several

alternatives for a target relation. For instance, applying the mapping to example (6-c) we get three alternatives for the relation in the target sentence:

(8) (verb-obj: achieve progress)
    (verb-obj: achieve advance)
    (verb-obj: achieve advancement)

For example (6-d) we get 9 alternatives, since both 'H-progress' and 'H-talks' have three alternative translations.

In order to decide which alternative is the most probable, we count the frequencies of all the alternative target relations in very large corpora. For example (8) we got the counts 29, 5 and 1 respectively. Similarly, the target relation 'to increase chance' was counted 20 times, while the other alternatives were not observed at all. These counts are given as input to the statistical model described in the next section, which performs the actual target word selection.

3 The Statistical Model

Our selection algorithm is based on the following statistical model. Consider first a single relation. The linguistic model provides us with several alternatives as in example (8). We assume that each alternative has a theoretical probability p_i to be appropriate for this case. We wish to select the alternative for which p_i is maximal, provided that it is significantly larger than the others.

We have decided to measure this significance by the odds ratio of the two most probable alternatives $\rho = p_1/p_2$. However, we do not know the theoretical probabilities, therefore we get a bound for ρ using the frequencies of the alternatives in the corpus.

Let $\hat{p}_i$ be the probabilities as observed in the corpus ($\hat{p}_i = n_i/n$, where n_i is the number of times that alternative i appeared in the corpus and n is the total number of times that all the alternatives for the relation appeared in the corpus).

For mathematical convenience we bound $\ln \rho$ instead of ρ. Assuming that samples of the alternative relations are distributed normally, we get the following bound with confidence $1 - \alpha$:

$$\ln\left(\frac{p_1}{p_2}\right) \geq \ln\left(\frac{\hat{p}_1}{\hat{p}_2}\right) - Z_{1-\alpha}\sqrt{var\left[\ln\left(\frac{\hat{p}_1}{\hat{p}_2}\right)\right]} \quad,$$

where Z is the confidence coefficient. We approximate the variance by the delta method (e.g. John-

son and Wichern (1982)):

$$\ln\left(\frac{\hat{p}_1}{\hat{p}_2}\right) \approx \ln\left(\frac{p_1}{p_2}\right) + (\hat{p}_1 - p_1)\left[\frac{\partial}{\partial x_1}\ln\left(\frac{x_1}{x_2}\right)\right]_{p_1,p_2}$$
$$+ (\hat{p}_2 - p_2)\left[\frac{\partial}{\partial x_2}\ln\left(\frac{x_1}{x_2}\right)\right]_{p_1,p_2}$$
$$= \ln\left(\frac{p_1}{p_2}\right) + \frac{\hat{p}_1 - p_1}{p_1} - \frac{(\hat{p}_2 - p_2)}{p_2}$$
$$= \ln\left(\frac{p_1}{p_2}\right) + \frac{\hat{p}_1}{p_1} - \frac{\hat{p}_2}{p_2}.$$

$$var\left[\ln\left(\frac{\hat{p}_1}{\hat{p}_2}\right)\right] = var\left[\ln\left(\frac{\hat{p}_1}{\hat{p}_1}\right) - \ln\left(\frac{p_1}{p_2}\right)\right]$$
$$\approx var\left[\frac{\hat{p}_1}{p_1} - \frac{\hat{p}_2}{p_2}\right]$$
$$= \frac{1}{p_1^2}\frac{p_1(1-p_1)}{n} + \frac{1}{p_2^2}\frac{p_2(1-p_2)}{n} + 2\frac{p_1 p_2}{n p_1 p_2}$$
$$= \frac{1}{n p_1} + \frac{1}{n p_2} \approx \frac{1}{n\hat{p}_1} + \frac{1}{n\hat{p}_2} = \frac{1}{n_1} + \frac{1}{n_2}.$$

Therefore we get that with probability at least $1 - \alpha$,

$$\ln\left(\frac{p_1}{p_2}\right) \geq \ln\left(\frac{\hat{p}_1}{\hat{p}_2}\right) - Z_{1-\alpha}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \quad.$$

We denote the right hand side (the bound) by $B_\alpha(n_1, n_2)$.

In sentences with several relations, we consider the best two alternatives for each relation, and take the relation for which B_α is largest. If this B_α is less than a specified threshold then we do not choose between the alternatives. Otherwise, we choose the most frequent alternative to this relation and select the target words appearing in this alternative. We then eliminate all the other alternative translations for the selected words, and accordingly eliminate all the alternatives for the remaining relations which involve these translations. In addition we update the observed probabilities for the remaining relations, and consequently the remaining B_α's. This procedure is repeated until all target words have been determined or the maximal B_α is below the threshold.

The actual parameters we have used so far were $\alpha = 0.05$ and the bound for B_α was -0.5.

To illustrate the selection algorithm, we give the details for example (6). The highest bound for the odds ratio ($B_\alpha = 1.36$) was received for the relation 'increase-chance', thus selecting the translation 'increase' for 'H-increase'. The second was $B_\alpha = 0.96$,

133

for 'achieve-progress'. This selected the translations 'achieve' and 'progress', while eliminating the other senses of 'H-progress' in the remaining relations. Then, for the relation 'progress-in-talks' we got $B_\alpha = 0.3$, thus selecting the appropriate translation for 'H-talks'.

4 The Experiment

An experiment was conducted to test the performance of the statistical model in translation from Hebrew and German to English. Two sets of paragraphs were extracted randomly from current Hebrew and German press. The Hebrew set contained 10 paragraphs taken from foreign news sections, while the German set contained 12 paragraphs of text not restricted to a specific topic.

Within these paragraphs we have (manually) identified the target word selection ambiguities, using a bilingual dictionary. Some of the alternative translations in the dictionary were omitted if it was judged that they will not be considered by an actual component of a machine translation program. These cases included very rare or archaic translations (that would not be contained in an MT lexicon) and alternatives that could be eliminated using syntactic knowledge (as explained in section 2)[2]. For each of the remaining alternatives, it was judged if it can serve as an acceptable translation in the given context. This a priori judgment was used later to decide whether the selection of the automatic procedure is correct. As a result of this process, the Hebrew set contained 105 ambiguous words (which had at least one unacceptable translation) and the German set 54 ambiguous words.

Now it was necessary to identify the lexical relations within each of the sentences. As explained before, this should be done using a source language parser, and then mapping the source relations to the target relations. At this stage of the research, we still do not have the necessary resources to perform the entire process automatically[3], therefore we have approximated it by translating the sentences into English and extracting the lexical relations using the English Slot Grammar (ESG) parser (mc-

Cord, 1989)[4]. Using this parser we have classified the lexical relations to rather general classes of syntactic relations, based on the slot structure of ESG. The important syntactic relations used were between a verb and its arguments and modifiers (counting as one class all objects, indirect objects, complements and nouns in modifying prepositional phrases) and between a noun and its arguments and modifiers (counting as one class all noun objects, modifying nouns in compounds and nouns in modifying prepositional phrases). The success of using this general level of syntactic relations indicates that even a rough mapping of source to target language relations would be useful for the statistical model.

The statistics for the alternative English relations in each sentence were extracted from three corpora: The Washington Post articles (about 40 million words), Associated Press news wire (24 million) and the Hansard corpus of the proceedings of the Canadian Parliament (85 million words). The statistics were extracted only from sentences of up to 25 words (to facilitate parsing) which contained altogether about 55 million words. The lexical relations in the corpora were extracted by ESG, in the same way they were extracted for the English version of the example sentences (see Dagan and Itai (1990a) for a discussion on using an automatic parser for extracting lexical relations from a corpus, and for the technique of acquiring the statistics). The parser failed to produce any parse for about 35% of the sentences, which further reduced the actual size of the corpora which was used.

5 Evaluation

Two measurements, *applicability* and *precision*, are used to evaluate the performance of the statistical model. The applicability denotes the proportion of cases for which the model performed a selection, i.e. those cases for which the bound B_α passed the threshold. The precision denotes the proportion of cases for which the model performed a correct selection out of all the applicable cases.

We compare the precision of the model to that of the "word frequencies" procedure, which always selects the most frequent target word. This naive "straw-man" is less sophisticated than other methods suggested in the literature but it is useful as a common benchmark (e.g. Sadler (1989)) since it can

---

[2]Due to some technicalities, we have also restricted the experiment to cases in which all the relevant translations of a word consists exactly one English word, which is the most frequent situation.

[3]We are currently integrating this process within GSG (German Slot Grammar) and LMT-GE (the German to English MT prototype).

[4]The parsing process was controlled manually to make sure that we do not get wrong relational representation of the examples due to parsing errors.

be easily implemented. The success rate of the "word frequencies" procedure can serve as a measure for the degree of lexical ambiguity in a given set of examples, and thus different methods can be partly compared by their degree of success relative to this procedure.

Out of the 105 ambiguous Hebrew words, for 32 the bound B_α did not pass the threshold (applicability of 70%). The remaining 73 examples were distributed according to the following table:

Hebrew-English		Word Frequencies	
		correct	incorrect
Relations	correct	45	22
Statistics	incorrect	2	4

Thus the precision of the statistical model was 92% $(67/73)$[5] while relying just on word frequencies yields 64% $(47/73)$.

Out of the 54 ambiguous German words, for 22 the bound B_α did not pass the threshold (applicability of 59%). The remaining 32 examples were distributed according to the following table:

German-English		Word Frequencies	
		correct	incorrect
Relations	correct	16	8
Statistics	incorrect	2	6

Thus the precision of the statistical model was 75% $(24/32)$, while relying just on word frequencies yields 53% $(18/32)$. We attribute the lower success rate for the German examples to the fact that they were not restricted to topics that are well represented in the corpus.

Statistical analysis for the larger set of Hebrew examples shows that with 95% confidence our method succeeds in at least 86% of the applicable examples (using the parameters of the distribution of proportions). With the same confidence, our method improves the word frequency method by at least 18% (using confidence interval for the difference of proportions in multinomial distribution, where the four cells of the multinomial correspond to the four entries in the result table).

In the examples that were treated correctly by our

---

[5] An a posteriori observation showed that in three of the six errors the selection of the model was actually acceptable, and the a priori judgment of the human translator was too severe. For example, in one of these cases the statistics selected the expression 'to begin talks' while the human translator regarded this expression as incorrect and selected 'to start talks'. If we consider these cases as correct then there are only three selection errors, getting a 96% precision.

method, such as the examples in the previous sections, the statistics succeeded to capture two major types of disambiguating data. In preferring 'sign-treaty' upon 'seal-treaty', the statistics reflect the relevant semantic constraint. In preferring 'peace-treaty' upon 'peace-contract', the statistics reflect the lexical usage of 'treaty' in English which differs from the usage of 'hoze' in Hebrew.

6 Failures and Possible Improvements

A detailed analysis of the failures of the method is most important, as it both suggests possible improvements for the model and indicates its limitations. As described above, these failures include either the cases for which the method was not applicable (no selection) or the cases in which it made an incorrect selection. The following paragraphs list the various reasons for both types.

6.1 Inapplicability

Insufficient data. This was the reason for nearly all the cases of inapplicability. For instance, none of the alternative relations 'an investigator of corruption' (the correct one) or 'researcher of corruption' (the incorrect one) was observed in the parsed corpus. In this case it is possible to perform the correct selection if we used only statistics about the cooccurrences of 'corruption' with either 'investigator' or 'researcher', without looking for any syntactic relation (as in Church and Hanks (1990)). The use of this statistic is a subject for further research, but our initial data suggests that it can substantially increase the applicability of the statistical method with just a little decrease in its precision.

Another way to deal with the lack of statistical data for the specific words in question is to use statistics about similar words. This is the basis for Sadler's *Analogical Semantics* (1989) which has not yet proved effective. His results may be improved if more sophisticated techniques and larger corpora are used to establish similarity between words (such as in (Hindle, 1990)).

Conflicting data. In very few cases two alternatives were supported equally by the statistical data, thus preventing a selection. In such cases, both alternatives are valid at the independent level of the lexical relation, but may be inappropriate for the specific context. For instance, the two alternatives of 'to take

a job' or 'to take a position' appeared in one of the examples, but since the general context concerned with the position of a prime minister only the latter was appropriate. In order to resolve such examples it may be useful to consider also cooccurrences of the ambiguous word with other words in the broader context. For instance, the word 'minister' seems to cooccur in the same context more frequently with 'position' than with 'job'.

In another example both alternatives were appropriate also for the specific context. This happened with the German verb 'werfen', which may be translated (among other options) as 'throw', 'cast' or 'score'. In our example 'werfen' appeared in the context of 'to throw/cast light' and these two correct alternatives had equal frequencies in the corpus ('score' was successfully eliminated). In such situations any selection between the alternatives will be appropriate and therefore any algorithm that handles conflicting data will work properly.

6.2 Incorrect Selection

Using the inappropriate relation. One of the examples contained the Hebrew word 'matzav', which two of its possible translations are 'state' and 'position'. The phrase which contained this word was: 'to put an end to the {state | position} of war ...'. The ambiguous word is involved in two syntactic relations, being a complement of 'put' and also modified by 'war'. The corresponding frequencies were:

(9) verb-comp: put-position 320
      verb-comp: put-state 18

      noun-nobj: state-war 13
      noun-nobj: position-war 2

The bound of the odds ration (B_α) for the first relation was higher than for the second, and therefore this relation determined the translation as 'position'. However, the correct translation should be 'state', as determined by the second relation.

This example suggests that while ordering the involved relations (or using any other weighting mechanism) it may be necessary to give different weights to the different types of syntactic relations. For instance, it seems reasonable that the object of a noun should receive greater weight in selecting the noun's sense than the verb for which this noun serves as a complement.

Confusing senses. In another example, the Hebrew word 'qatann', which two of its meanings

are 'small' and 'young', modified the word 'sikkuy', which means 'prospect' or 'chance'. In this context, the correct sense is necessarily 'small'. However, the relation that was observed in the corpus was 'young prospect', relating to the human sense of 'prospect' which appeared in sport articles (a promising young person). This borrowed sense of 'prospect' is necessarily inappropriate, since in Hebrew it is represented by the equivalent of 'hope' ('tiqva'), and not by 'sikkuy'.

The reason for this problem is that after producing the possible target alternatives, our model ignores the source language input as it uses only a monolingual target corpus. This can be solved if we use an aligned bilingual corpus, as suggested by Sadler (1989) and Brown et al. (1990). In such a corpus the occurrences of the relation 'young prospect' will be aligned to the corresponding occurrences of the Hebrew word 'tiqva', and will not be used when the Hebrew word 'sikkuy' is involved. Yet, it should be brought in mind that an aligned corpus is the result of manual translation, which can be viewed as a manual tagging of the words with their equivalent senses in the other language. This resource is much more expensive and less available than the untagged monolingual corpus, while it seems to be necessary only for relatively rare situations.

Lack of deep understanding. By their nature, statistical methods rely on large quantities of shallow information. Thus, they are doomed to fail when disambiguation can rely only on deep understanding of the text and no other surface cues are available. This happened in one of the Hebrew examples, where the two alternatives were either 'emigration law' or 'immigration law' (the Hebrew word 'hagira' is used for both subsenses). While the context indicated that the first alternative is correct, the statistics preferred the second alternative. It seems that such cases are quiet rare, but only further evaluation will show the extent to which deep understanding is really needed.

7 Conclusions

The method presented takes advantage of two linguistic phenomena: the different usage of words and word senses among *different* languages and the importance of lexical cooccurrences within syntactic relations. The experiment shows that these phenomena are indeed useful for practical disambiguation.

We suggest that the high precision received in the experiment relies on two characteristics of the am-

136

biguity phenomena, namely the *sparseness* and *redundancy* of the disambiguating data. By sparseness we mean that within the large space of alternative interpretations produced by ambiguous utterances, only a small portion is commonly used. Therefore the chance of an inappropriate interpretation to be observed in the corpus (in other contexts) is low. Redundancy relates to the fact that different informants (such as different lexical relations or deep understanding) tend to support rather than contradict one another, and therefore the chance of picking a "wrong" informant is low.

The examination of the failures suggests that future research may improve both the applicability and precision of the model. Our next goal is to handle inapplicable cases by using cooccurrence data regardless of syntactic relations and similarities between words. We expect that increasing the applicability will lead to some decrease in precision, similar to the tradeoff between *recall* and *precision* in information retrieval. Pursuing this tradeoff will improve the performance of the method and reveal its limitations.

8   Acknowledgments

References

[1] Brown, P., Cocke, J., Della Pietra, S., Della Pietra, V., Jelinek, F., Mercer, R.L. and Rossin P.S., A statistical approach to language translation, *Computational Linguistics*, vol. 16(2), 79-85 (1990).

[2] Chodorow, M. S., R. J. Byrd and G. E. Heidron, Extracting Semantic Hierarchies from a Large On-Line Dictionary. Proc. of the 23rd Annual Meeting of the ACL, 299-304 (1985).

[3] Church, K. W., and Hanks, P., Word association norms, mutual information, and Lexicography, *Computational Linguistics*, vol. 16(1), 22-29 (1990).

[4] Dagan, I. and A. Itai, Automatic Acquisition of Constraints for the Resolution of Anaphora References and Syntactic Ambiguities, COLING 1990, Helsinki, Finland.

[5] Dagan, I. and A. Itai, A Statistical Filter for Resolving Pronoun References, Proc. of the 7th Israeli Sym. on Artificial Intelligence and Computer Vision, 1990.

[6] Hindle, D. Noun Classification from Predicate-Argument Structures, Proc. of the 28rd Annual Meeting of the ACL, (1990).

[7] Hindle D. and M. Rooth, Structural Ambiguity and Lexical Relations, Proc. of the Speech and Natural Language Workshop, (DARPA), June 1990.

[8] Hornby, A. S., C. Ruse, J. A. Reif and Y. Levy, *Oxford Student's Dictionanary for Hebrew Speakers*, Kernerman Publishing Ltd, Lonnie Kahn & Co. Ltd. (1986).

[9] Johnson, R. A. and D. W. Wichern, *Multivariate Statistical Analysis*, Prentice-Hall, 1982.

[10] Katz, S., Recursive *m*-gram language model via a smoothing of Turing's formula, IBM Tech. Disclosure Bull. , 1985.

[11] Lenat, D. B., R. V. Guha, K. Pittman, D. Pratt and M. Shepherd, Cyc: toward programs with common sense, *Comm. ACM*, vol. 33(8), 1990.

[12] McCord, M. C., A new version of slot grammar, *Research Report RC 14506*, IBM Research Division, Yorktown Heights, NY, 1989.

[13] Nirenburg, S., (ed.), *Machine Translation*, Cambridge University Press (1987).

[14] Nirenburg, S., I. Monarch, T. Kaufmann, I. Nirenburg and J. Carbonell. Acquisition of Very Large Knowledge Bases: Methodology, Tools and Applications, Center for Machine Translation, Carnegie-Mellon, CMU-CMT-88-108, (1988).

[15] Sadler, V., *Working with analogical semantics: disambiguation techniques in DLT*, Foris Publications, 1989.

[16] Van Eynde, F. et al.: The Task of Transfer vis-a-vis Analysis and Generation. Eurotra Final Report ET-10-B/NL, (1982).

[17] Wittgenstein, L. *Philosophical Investigations*, Oxford (1953).

[18] Zernik U., and P. Jacobs, Tagging for Learning: Collecting Thematic Relations from Corpus. Proc. COLING 1990.