A Method for Refining Automatically-Discovered Lexical Relations: Combining Weak Techniques for Stronger Results

Gregory Grefenstette
Department of Computer Science
210 MIB
University of Pittsburgh
Pittsburg, PA 15232
grefen@cs.pitt.edu

Marti A. Hearst
Computer Science Division
571 Evans Hall
University of California, Berkeley
Berkeley,CA 94720
marti@cs.berkeley.edu

Abstract

Knowledge-poor corpus-based approaches to natural language processing are attractive in that they do not incur the difficulties associated with complex knowledge bases and real-world inferences. However, these kinds of language processing techniques in isolation often do not suffice for a particular task; for this reason we are interested in finding ways to combine various techniques and improve their results.

Accordingly, we conducted experiments to refine the results of an automatic lexical discovery technique by making use of a statistically-based syntactic similarity measure. The discovery program uses lexico-syntactic patterns to find instances of the hyponymy relation in large text bases. Once relations of this sort are found, they should be inserted into an existing lexicon or thesaurus. However, the terms in the relation may have multiple senses, thus hampering automatic placement. In order to address this problem we tried to make a term-similarity determination technique choose where, in an existing thesaurus, to install a lexical relation. The union of these two corpus-based methods is promising, although only partially successful in the experiments run so far. Here we report some preliminary results, and make suggestions for how to improve the technique in future.

1 Introduction

Knowledge-poor corpus-based approaches to natural language processing are attractive in that they do not incur the difficulties associated with complex knowledge bases and real-world inferences. However, coarse-level language processing techniques in isolation often do not suffice for a particular task; for this reason we are interested in finding ways to combine various approaches and improve their results.

Accordingly, we conducted experiments to refine the results of an automatic lexical discovery technique by making use of a statistically-based syntactic similarity measure, and integrating them with an existing knowledge structure. The discovery program uses lexico-syntactic patterns to find instances of the hyponymy relation in large text bases. Once relations of this sort are found, they should be inserted into an existing lexicon or thesaurus. However, the terms in the relation may have multiple senses, thus hampering automatic placement. In order to address this problem we tried to make a term-similarity determination technique choose where, in an existing thesaurus, to install a lexical relation. The union of these two corpus-based methods is promising, although only partially successful in the experiments run so far.

These ideas are related to other recent work in several ways. We make use of restricted syntactic information as do Brent's (Brent 1991) verb subcategorization frame recognition technique and Smadja's (Smadja & McKeown 1990) collocation acquisition algorithm. The work reported here attempts to find semantic similarity among terms based on the contexts they tend to occur in; (Church & Hanks 1990) uses frequency of co-occurrence of content words to create clusters of semantically similar words, (Hindle 1990) uses both simple syntactic frames and frequency of occurrence of content words to determine similarity among nouns, and (Calzolari & Bindi 1990) use corpus-based statistical association ratios to determine lexical information such as prepositional complementation relations, modification relations, and significant compounds. This paper presents an attempt to combine knowledge-poor techniques; (Wilks 1990) discusses the potential power behind combining weak methods, and in (Wilks *et al.*

1990) describes advances achieved using this paradigm.

The next section describes in detail the problem being addressed and the two existing coarse-level language processing techniques that are to be combined. Section 3 describes how the similarity calculations are done, and Section 4 presents the results of applying these calculations to several examples and the difficulties that arise in each case. Section 5 sketches some solutions to the difficulties mentioned in the previous section and discusses one aspect of the integration of statistical techniques and a thesaurus, and Section 7 presents a brief conclusion.

2 The Problem: Integration of Lexical Relations

(Hearst 1992) reports a method for the automatic acquisition of the hyponymy lexical relation from unrestricted text. In this method, the text is scanned for instances of distinguished lexico-syntactic patterns that indicate the relation of interest.

For example, consider the lexico-syntactic pattern

... NP {, NP} {,} or other NP ...*

When a sentence containing this pattern is found (with some restrictions on the syntax to the left and the right of the pattern) it can be inferred that the NP's on the left of *or other* are hyponyms of the NP on the right (where NP indicates a simple noun phrase). From the sentence

```
Bruises, wounds, broken bones or other
injuries are common.
```

we can infer:

hyponym(bruise, injury)
hyponym(wound, injury)
hyponym(broken bone, injury)

This pattern is one of several that have been identified as indicating the hyponymy relation.

This approach differs from statistical techniques in an interesting way. Both require as their primary resource a large text collection, but whereas statistical techniques try to find correlations and make generalizations based on the data of the entire corpus, only a single instance of a lexico-syntactic pattern need be found in order to have made a discovery.

Once the relation is found, it is desirable to integrate it as a part of an existing network of lexical relations. We want to develop a means to correctly insert an instance of the hyponymy relation into an existing hyponymically-structured network (hyponymy is reflexive, symmetric and transitive).

For our experiments we use the manually constructed thesaurus WordNet (Miller *et al.* 1990). In WordNet, word forms with synonymous meanings are grouped into sets, called synsets. This allows a distinction to be made between senses of homographs. For example, the noun *"board"* appears in the synsets {*board, plank*} and {*board, committee*}, and this grouping serves for the most part as the word's definition. In version 1.1, WordNet contains about 34,000 noun word forms, including some compounds and proper nouns, organized into about 26,000 synsets. Noun synsets are organized hierarchically[1] according to the hyponymy relation with implied inheritance and are further distinguished by values of features such as meronymy. WordNet's coverage and structure are impressive and provide a good basis for an automatic acquisition algorithm to build on.

Now, assuming we have discovered the relation *hyponym(X, Y)*, indicating that X is a kind of Y, we wish to enter this relation into the WordNet network.

If the network is sufficiently mature, as is WordNet, we can assume that most of the highly ambiguous words are already present and appear in higher levels of the network. Therefore, most of the time we will be trying to insert a rather specific term that itself does not need to be disambiguated (i.e., it has only one main sense) as a hyponym of a term that can have one or more senses. If we assume this is indeed the case, then there are two scenarios to consider:

(Scenario 1) Each sense of Y has several child subtrees and the task is to determine which subtree X shares context with. This in turn implies which sense of Y the hyponym relation refers to.

(Scenario 2) Y has more than one sense or else one or more of its senses has no children. Thus there are no subtrees to compare X against.

There are two considerations associated with Scenario 1:

(1a) X is not a direct descendent of Y, but is two or more levels down.

(1b) X belongs in a subtree of its own, even though the correct sense of Y has child subtrees.

In the work described here we address only the situation

[1] Although WordNet's hyponyms are structured as a directed network, as opposed to as a tree, for the purposes of this paper, we treat it as if it were a tree.

associated with Scenario 1, since our technique uses the child subtrees to determine which sense of Y to associate X with.

In order to place the hyponym relation into the network, we propose the following:

> Similarity Hypothesis: the context of a hyponym X is significantly more similar to the contexts of the sense of Y that is its parent hypernym than to the contexts of the senses of Y that are not its parent hypernym. Furthermore, these contexts can be determined from the contexts that surround the child subtrees of each sense of Y.

(Grefenstette 1992b) has developed a weak technique, embodied in a program called SEXTANT, which, given a target word, determines which other words in a corpus are most similar to it based on their syntactic usage patterns. For scenario (1) we will employ this pairwise-similarity determination method on all the children of each sense of Y, in an effort to determine which subtree X is most similar to. Bearing consideration (1a) in mind, we may be matching against sub-sub-trees, down several levels, although presumably the more detailed the divisions, the more difficult it will be to distinguish among them using context alone.

3 Similarity Determination

This section describes the mechanism by which the syntactic similarity determination is done, using one running example. The results of determining syntactic similarity for several examples appear in the following section.

3.1 Generating the corpus

One of the relations extracted from an on-line encyclopedia (Grolier 1990) using the technique described in (Hearst 1992), is *hyponym(Harvard, institution)*. As described above, if we wish to insert this relation into a hierarchical structure such as WordNet, we have to decide which sense of 'institution' is appropriate. Below we list an abbreviated version of the WordNet synsets associated with the various senses of 'institution.' Each sense is followed by the hyponymic subtrees associated with it, indicated by an arrow.

```
Institution: (hyponyms)

constitution, establishment,
formation, initiation, founding,
foundation, institution, origination,
creation, instauration
      => colonization, settlement

institution, institute
      => orphanage, orphans' asylum
      => penal institution

institution, establishment
      => religion, faith, church
      => school
          => college
          => university
          => conservatory
      => academy
      => foundation
      => bank
```

Our goal is to see if, by examining the syntactic contexts of these terms in a corpus of text, we can decide under which synset to place 'Harvard.'

Given a large enough text sample SEXTANT can tell us what words are used the most similar to 'Harvard.' In order to generate this text, we took words from the above list, and extracted all the sentences from (Grolier 1990) that contained those terms. This generated a corpus of 4 Million characters (730,000 words, 26,000 sentences). Sample sentences are:

```
Both the form of the letter and its position
at the beginning of the alphabet are from
the Latin alphabet , which derived it from
the Greek by way of the Etruscan.

Zwingli's Sixty-seven Articles (1523)
for disputation became a basic doctrinal
document for the Swiss reformed church.
```

3.2 Syntactic Analysis of the Corpus

These sentences were grammatically analyzed using a robust system developed for CLARIT (Evans *et al.* 1991). At this stage, the following type of information is output, associating one grammatical category and a normalized word form for each word:

```
"both" quant both
"the" d the
"form" sn form
"of" prep of
"the" d the
"letter" sn letter
```

```
"and" cnj and
"its" gen its
"position" sn position
"at" prep at
"the" d the
"beginning" vb-prog begin
"of" prep of
"the" d the
"alphabet" sn alphabet
"are" auxb be
"from" prep from
"the" d the
"latin" adj latin
"alphabet" sn alphabet
...
```

The SEXTANT system takes this output and uses a number of simple robust algorithms (Grefenstette 1992a) to divide the sentences of the corpus into complex noun phrases (NP) and verb phrases (VP) as shown below:

```
NP     both
NP     the form of the letter and its
       position at the begin of the
       alphabet
VP     be
NP     from the latin alphabet
--     , which
VP     derive
NP     it from the greek by way of
       the  etruscan
--     .
```

Then, the modifiers are connected to head nouns and the verbs are connected to their subjects and objects, producing connections like the following:

head	*modifier*	*type*
position	letter	NN
form	position	NNPREP
position	beginning	NNPREP
alphabet	begin	SUBJ
alphabet	beginning	IOBJ
beginning	alphabet	NNPREP
form	be	SUBJ
alphabet	latin	ADJ
form	derive	SUBJ
derive	greek	DOBJ
derive	way	IOBJ
way	etruscan	NNPREP

3.3 Syntactically-Derived Similarity

From these relations we retain a simplified version of head-modifier pairs which becomes the input to a similarity calculation. Using measures derived in the social sciences for comparing two individuals, each described by a number of attributes (Romesburg 1984), we compare each of the nouns in the corpus derived from the words in the list above. The attributes of each word are those other words found in syntactic relation to it by the simple syntactic processing of SEXTANT. The data at this stage are a list of words and modifying attributes:

```
position letter
form position
position beginning
alphabet begin-SUBJ
beginning alphabet-IOBJ
beginning alphabet
alphabet latin
rm derive-SUBJ
greek derive-DOBJ
way derive-IOBJ
way etruscan
```

We use, as the similarity measure, the Jaccard coefficient, which is roughly the sum of shared attributes divided by the sum of unique attributes in the two objects being compared. We use a weighted version of the Jaccard coefficient; each attribute is weighted between 0 and 1 as a function of how many different words it modifies. See (Grefenstette 1992a) for details.

4 Several Examples

4.1 Example 1

As described in the preceeding section, we processed 4 megabytes of text for the 'Harvard' example, performing morphological analysis, dictionary look-up, rough grammatical disambiguation, division into noun and verb phrases, parsing, and extraction of lexical attributes for each noun in the corpus.

Many of the terms in the institution synsets were found to have reasonable associations using these rough techniques, as indicated in the table below. The terms at the bottom of this table have low frequency of occurrence in the corpus as a whole and so the terms associated with them tend to be less representative than those of more frequent terms.

term	*freq*	*closest term*
school	2130	college
university	1954	college
church	1866	school
college	1202	school
settlement	1006	church
formation	880	creation
bank	751	government
foundation	746	program
religion	743	faith
constitution	665	government
creation	652	establishment
establishment	597	creation
academy	565	college
faith	368	religion
founding	225	history
colonization	113	colony
conservatory	108	academy
harvard	91	cambridge
initiation	80	rite
asylum	64	peace
orphan	11	mottel
orphanage	7	vigilantism
origination	2	coorigination

Note that the terms related to the third WordNet subtree, the one containing the school sense, did not interact with the terms of the first and second subtrees. However, although the association between 'church' and 'school' indicates the correct subtree, it focuses on the wrong sub-sub-tree.

From this data, the similarity comparison of SEXTANT finds that the two words which are most similar to 'Harvard' are 'Cambridge' and 'Oxford.' For example, 'Harvard' and 'Oxford' are found to be similar because they are both modified by the following words at different places throughout the corpus:

```
study-SUBJ university educate-DOBJ
college graduate-SUBJ establish-SUBJ
include-SUBJ serve-SUBJ teach-SUBJ
return-SUBJ return-IOBJ train-SUBJ
enter-DOBJ study-IOBJ educate-IOBJ
educate-SUBJ leave-SUBJ oxford
```

This result points out one difficulty with our task. This particular portion of WordNet does not contain listings of specific instances of university names (the unabbreviated version of the network beneath 'institution' does contain names of specific religious groups, however). Note that the terms found closest to 'Cambridge' and 'Oxford' are 'Oxford' and 'Cambridge', respectively. It may well be the case that the distributional properties of proper nouns are different than those of other terms, but in any case, there are no terms currently in the network that are at the same level of description as these names. If 'Oxford' had been correctly assigned as a child of one sense of 'institution,' we could reasonably assign 'Harvard' to that same subtree.

Although, in reality, any hierarchy containing 'Oxford' and 'Cambridge' would most probably include 'Harvard', this kind of result should be useful in non-static domains in which new terms are introduced over time (e.g., in the medical domain). If a knowledge structure exists for the domain, new text will produce new terms that must be integrated into this structure. This experiment provides an indication that such integration may be automated to a large degree using existing lexicons, grammars, and text corpora that cover the relevant phenomena well.

4.2 Example 2

As another example, the acquisition algorithm discovered the lexical relation between 'rice' and 'cereal'. WordNet has two senses of 'cereal,' one as a grain, and one as a breakfast product. We extracted a large 260,000 characters of text involving the following strings:

```
frumenty kasha grits hominy
grits farina millet
oatmeal hasty pudding mush
burgoo flummery gruel loblolly
porridge rice oat cornmeal meal
corn wheat buckwheat barley
cold cereal puffed wheat
puffed rice wheatflakes cornflakes
granola
```

In Table 1, we see that 'rice' is found to be closest to 'wheat,' and vice versa. This table also shows the results involving the breakfast product terms (some did not occur in the corpus); note that the frequency of these terms is too low for valid assessments to be made. For example, the cold cereal terms that were unambiguous, such as 'wheatflakes,' 'cornflakes,' and 'granola,' as well as all of the hot cereal items are underrepresented in the corpus. This fact reduces the possibility that 'rice' could be found similar to the breakfast product terms, although the fact that it is strongly related to the 'wheat' sense does lend some validity to the result.

So this example highlights another difficulty; underrepresentation of data in the corpus can make the position assignment technique unapplicable.

word	*freq*	*closest terms*
wheat	234	rice, corn
rice	217	wheat, corn
corn	182	rice, wheat
crop	179	wheat, cultivation, rice, grain, production
meal	129	corn grain food, export, product, rice, hay,
product	102	export, production, industry, cattle
grain	101	export, growing, plant, cereal, producer
production	93	cultivation, growing
area	90	region, farm, land
food	89	export, grain, crop
plant	84	variety grain, production, seed
center	79	distribution, production, product, export, farmer
cultivation	71	growing, production,
farming	65	field, farm, export
	...	
millet	59	francois, sorghum, barley
barley	46	potatoe, cotton, rye
oat	44	barley, vegetable, potatoe, cotton
cereal	41	producer, starch, flour,
buckwheat	10	bread, sugarcane, pineapple, sorgum
grit	10	wing, farina, estrilidae
loblolly	8	grade drought, pine
gruel	3	second, conglomerate
mush	3	cornmeal
oatmeal	3	quaker, gascony
porridge	3	symbol, ale, roll
cornmeal	1	counterpart, sports-writer, literature
farina	1	embayment, peel
pudding	1	NO RELATIONS

Table 1: Results of Cereal Experiment

4.3 Example 3

Another relation we considered is *hyponym(logarithm, calculation)*. WordNet records 'logarithm' as a kind of 'exponent' which in turn is a kind of 'mathematical notation'. However, the sense of logarithm as a calculation should also be considered. The WordNet structure for 'calculation' is:

```
calculation, computation, figuring, reckoning
       => extrapolation
       => interpolation
       => estimate, estimation
           => guess, guesswork, shot, dead
              reckoning
           => approximation
       => integral
           => indefinite integral
           => definite integral

calculation, computation
       => mathematical process, operation
           => differentiation
           => division
           => integration
           => multiplication
           => subtraction
           => summation, addition
           => exponentiation, involution
```

Although there is no common one-word term for computing a logarithm, there should be an entry for 'logarithm' alongside 'exponentiation' in the second subtree (or perhaps 'taking a logarithm'). Our results found 'logarithm' to be closest to 'extrapolation' and 'multiplication', one term from each subtree, and thus eluding correct classifiction. (It isn't clear, however, that the first subtree is entirely well-defined. Why is 'integral' associated with figuring, as opposed to listing 'integration' in the second subtree?)

This example shows that difficulties arise when the shades of differences among the terms are subtle, fine-grained, or somewhat arbitrary. The original goal was to make coarser-level distinctions, but because WordNet is so well-developed, it turns out that any decision will require choosing between finely divided subtrees.

4.4 Other Examples

The previous three examples have shown various difficulties associated with this approach. Foremost among them is the fact that WordNet senses have been manually created, and correspond to a human conception of what attributes of two words or concepts are saliently similar. It may well be that two concepts considered to be similar in the WordNet hierarchy do not display the kinds of similarities that our methods recognize in a particular corpus.

There are other difficulties as well. For example, in one instance we found *hyponym(granite,rock)*. In WordNet there are very fine differences among the senses of 'rock' and in fact 'granite' appears in more than one of its subtrees. Other difficulties were: the hypernym senses have no child subtrees to compare against, the hypernym is a very general term and thus has hundreds of children, and the hyponym does not appear frequently enough in the corpus to take statistics on.

5 Discussion

Although none of the examples here yield perfect results, the implications, especially in the 'Harvard' case, are promising. In this section we discuss some ways to improve the approach and then point out an effect that arises from the integration of statistical techniques with an existing knowledge structure.

5.1 Suggestions for Improvements

As seen from the previous section, there are many cases in which the technique is not applicable, but we have some ideas for improvements.

First, we might be able to improve the results by combining the similar terms from each subtree and comparing them to the term we are interested in placing, as opposed to performing pairwise comparisons only. This might help in the cases where there are only fine distinctions among the senses.

Second, currently only individual words are compared, whereas in WordNet as well as in common text, many of the interesting concepts (or even less interesting ones such as 'puffed rice') appear as multi-word phrases. For example, when we tried to classify 'Einstein' under the proper class of 'genius', 'Einstein' was found to be closest to 'Albert' and 'theory'. (This was also another case in which there were no terms at the correct level of description – names of individual geniuses do not currently appear in the network.) We are hoping to integrate the phrase extraction mechanisms of CLARIT (Evans *et al.* 1991) in order to treat these properly.

Third, the uses of only syntactic contexts gives an interesting but restricted idea of what a word means. We plan to add to a word's context other easily extractable contexts such as, for example, looking at the head nouns appearing in the same sentence as the word.

5.2 WordNet as a Corpus Partitioner

Although neither of the language processing techniques described in this paper require knowledge bases for their use, in these experiments we have made use of the WordNet thesaurus in an interesting way.

We assumed that the WordNet synsets could be used for disambiguation since each sense of a polysemous word is associated with its own distinguishing hypernyms and hyponyms. This assumption allows us to extract a corpus that is coherent with respect to a semantic subdomain. For the 'Harvard' example discussed earlier, we extracted sentences that contained the terms found in the subtrees beneath 'institution.' We showed which terms from the corpus were found to be semantically closest to these institution terms. However, many words that are not found in the subtrees of 'institution' nevertheless occur frequently in these sentences. Furthermore, these terms tend to be highly ambiguous.

In Table 2, we show the results of performing our similarity analysis on these terms "once-removed." From this table, it appears that the institutional sense of these polysemous words has been identified, e.g., *member-president, right-freedom, service-program,* and *union-association.* In other words, it seems that by taking a WordNet synset as a filter through a large text, we have extracted a rather coherent corpus. In this case, the corpus is one with a bent towards the notion of institution. If we were to use a corpus associated with terms beneath the 'sports' synset, for example, we would expect different associations for 'service', 'movement', and 'court'.

This observation helps substantiate the claim that we can use existing knowledge structures to help coarse-level analysis techniques – in this case the thesaurus helps find a semantic partition on unrestricted textual data.

6 Conclusion

This experiment shows that two techniques for extracting information from text: one that extracts hyponymic relations, and another that determines word similarity, have the potential to be combined (with limited success) to provide cleaner information than either technique separately. Both techniques are fully operational and work better as the corpora they treat get larger.

word	*freq*	*closest terms*
state	771	college, constitution, university
system	751	program, government, education
group	545	movement, organization, number
education	530	program, training, student, study
art	505	architecture, education, academy
power	502	authority, control, right, government
society	400	association, art, education, academy
movement	374	group, organization, leader, education
faith	368	religion, tradition, form, constitution
member	344	president, year, director, student
company	323	center, association, bank, theater, city
theory	322	study, idea, foundation, curriculum, view
life	321	society, history, development, organization
professor	299	department, director, degree, history
service	295	program, education, organization, center
court	294	constitution, government, law, authority
people	288	student, society, child, country, area
practice	284	system, tradition, education, reform
tradition	284	history, architecture, religion, art
council	283	organization, constitution, movement
museum	283	library, institute, art, center, site
right	277	freedom, power, constitution, government
control	273	power, administration, authority, rule
world	234	culture, tradition, community, country
union	189	association, party, movement, history
level	162	course, curriculum, education, training

Table 2: Terms similar to terms "once-removed" from 'institution'

Despite the difficulites associated with the task at hand, we are confident that such robust knowledge-poor techniques hold promise for the future of automated natural language processing, and for intelligent treatment of large text collections.

Acknowledgements. Hearst's research was sponsored in part by the University of California and Digital Equipment Corporation under Digital's flagship research project Sequoia 2000: Large Capacity Object Servers to Support Global Change Research, and in part by an internship at Xerox Palo Alto Research Center.

References

Brent, M. R. (1991). Automatic acquisition of subcategorization frames from untagged, free-text corpora. In *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics.*

Calzolari, N. & R. Bindi (1990). Acquisition of lexical information from a large textual italian corpus. In *Proceedings of the Thirteenth International Conference on Computational Linguistics*, Helsinki.

Church, K. & P. Hanks (1990). Word association norms, mutual information, and lexicography. *American Journal of Computational Linguistics*, 16(1):22–29.

Evans, D. A., S. K. Henderson, R. G. Lefferts, & I. A. Monarch (1991). A summary of the CLARIT project. Technical Report CMU-LCL-91-2, Laboratory for Computational Linguistics, Carnegie-Mellon University.

Grefenstette, G. (1992a). SEXTANT: extracting semantics from raw text implementation details. Technical Report CS92-05, University of Pittsburgh, Computer Science Dept.

Grefenstette, G. (1992b). Use of syntactic context to produce term association lists for text retrieval. In *Proceedings of SIGIR '92*, Copenhagen, Denmark.

Grolier (1990). *Academic American Encyclopedia.* Grolier Electronic Publishing, Danbury, Connecticut.

Hearst, M. A. (1992). Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the Fourteenth International Conference on Computational Linguistics*, Nantes, France.

Hindle, D. (1990). Noun classification from predicate-argument structures. *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, pages 268–275.

Miller, G. A., R. Beckwith, C. Fellbaum, D. Gross, & K. J. Miller (1990). Introduction to wordnet: An on-line lexical database. *Journal of Lexicography*, 3(4):235–244.

Romesburg, H. C. (1984). *Cluster Analysis for Researchers.* Lifetime Learning Publications, Belmont, CA.

Smadja, F. A. & K. R. McKeown (1990). Automatically extracting and representing collocations for language generation. In *Proceedings of the 28th Annual Meeting of the Association for Computational Linguistics*, pages 252–259.

Wilks, Y. (1990). Combining weak methods in large-scale text processing. In P. S. Jacobs, editor, *Text-Based Intelligent Systems: Current Research in Text Analysis, Information Extraction, and Retrieval.* GE Research & Development Center, TR 90CRD198.

Wilks, Y. A., D. C. Fass, C. ming Guo, J. E. McDonald, T. Plate, & B. M. Slator (1990). Providing machine tractable dictionary tools. *Journal of Machine Translation*, 2.