

# Homonymy and Polysemy in Information Retrieval

Robert Krovetz

NEC Research Institute\*

4 Independence Way

Princeton, NJ. 08540

krovetz@research.nj.nec.com

## Abstract

This paper discusses research on distinguishing word meanings in the context of information retrieval systems. We conducted experiments with three sources of evidence for making these distinctions: morphology, part-of-speech, and phrases. We have focused on the distinction between homonymy and polysemy (unrelated vs. related meanings). Our results support the need to distinguish homonymy and polysemy. We found: 1) grouping morphological variants makes a significant improvement in retrieval performance, 2) that more than half of all words in a dictionary that differ in part-of-speech are related in meaning, and 3) that it is crucial to assign credit to the component words of a phrase. These experiments provide a better understanding of word-based methods, and suggest where natural language processing can provide further improvements in retrieval performance.

## 1 Introduction

Lexical ambiguity is a fundamental problem in natural language processing, but relatively little quantitative information is available about the extent of the problem, or about the impact that it has on specific applications. We report on our experiments to resolve lexical ambiguity in the context of information retrieval (IR). Our approach to disambiguation is to treat the information associated with dictionary

---

This paper is based on work that was done at the Center for Intelligent Information Retrieval at the University of Massachusetts. It was supported by the National Science Foundation, Library of Congress, and Department of Commerce under cooperative agreement number EEC-9209623. I am grateful for their support.

senses (morphology, part of speech, and phrases) as multiple sources of evidence.<sup>1</sup> Experiments were designed to test each source of evidence independently, and to identify areas of interaction. Our hypothesis is:

**Hypothesis 1** *Resolving lexical ambiguity will lead to an improvement in retrieval performance.*

There are many issues involved in determining how word senses should be used in information retrieval. The most basic issue is one of identity — what *is* a word sense? In previous work, researchers have usually made distinctions based on their intuition. This is not satisfactory for two reasons. First, it is difficult to scale up; researchers have generally focused on only two or three words. Second, they have used very coarse grained distinctions (e.g., ‘river bank’ v. ‘commercial bank’). In practice it is often difficult to determine how many senses a word should have, and meanings are often related (Kilgarriff 91).

A related issue is sense *granularity*. Dictionaries often make very fine distinctions between word meanings, and it isn’t clear whether these distinctions are important in the context of a particular application. For example, the sentence *They danced across the room* is ambiguous with respect to the word *dance*. It can be paraphrased as *They were across the room and they were dancing*, or as *They crossed the room as they danced*. The sentence is not ambiguous in Romance languages, and can only have the former meaning. Machine translation systems therefore need to be aware of this ambiguity and translate the sentence appropriately. This is a systematic class of ambiguity, and applies to all “verbs of translatory motion” (e.g., *The bottle floated under the bridge* will exhibit the same distinction (Talmy 85)). Such distinctions are unlikely to have an impact on information retrieval. However, there are

---

<sup>1</sup>We used the Longman Dictionary as our source of information about word senses (Procter 78).

also distinctions that are important in information retrieval that are unlikely to be important in machine translation. For example, the word *west* can be used in the context *the East versus the West*, or in the context *West Germany*. These two senses were found to provide a good separation between relevant and non-relevant documents, but the distinction is probably not important for machine translation. It is likely that different applications will require different types of distinctions, and the type of distinctions required in information retrieval is an open question.

Finally, there are questions about how word senses should be used in a retrieval system. In general, word senses should be used to supplement word-based indexing rather than indexing on word senses alone. This is because of the uncertainty involved with sense representation, and the degree to which we can identify a particular sense with the use of a word in context. If we replace words with senses, we are making an assertion that we are very certain that the replacement does not lose any of the information important in making relevance judgments, and that the sense we are choosing for a word is in fact correct. Both of these are problematic. Until more is learned about sense distinctions, and until very accurate methods are developed for identifying senses, it is probably best to adopt a more conservative approach (i.e., uses senses as a supplement to word-based indexing).

The following section will provide an overview of lexical ambiguity and information retrieval. This will be followed by a discussion of our experiments. The paper will conclude with a summary of what has been accomplished, and what work remains for the future.

## 2 Lexical Ambiguity and Information Retrieval

### 2.1 Background

Many retrieval systems represent documents and queries by the words they contain. There are two problems with using words to represent the content of documents. The first problem is that words are ambiguous, and this ambiguity can cause documents to be retrieved that are not relevant. Consider the following description of a search that was performed using the keyword 'AIDS':

Unfortunately, not all 34 [references] were about AIDS, the disease. The references included "two helpful aids during the first three months after total hip replacement", and "aids in diagnosing abnormal voiding patterns". (Helm 83)

One response to this problem is to use phrases to reduce ambiguity (e.g., specifying 'hearing aids' if that is the desired sense). It is not always possible, however, to provide phrases in which the word occurs only with the desired sense. In addition, the requirement for phrases imposes a significant burden on the user.

The second problem is that a document can be relevant even though it does not use the same words as those that are provided in the query. The user is generally not interested in retrieving documents with exactly the same words, but with the concepts that those words represent. Retrieval systems address this problem by expanding the query words using related words from a thesaurus (Salton and McGill 83). The relationships described in a thesaurus, however, are really between word senses rather than words. For example, the word 'term' could be synonymous with 'word' (as in a vocabulary term), 'sentence' (as in a prison term), or 'condition' (as in 'terms of agreement'). If we expand the query with words from a thesaurus, we must be careful to use the right senses of those words. We not only have to know the sense of the word in the query (in this example, the sense of the word 'term'), but the sense of the word that is being used to augment it (e.g., the appropriate sense of the word 'sentence') (Chodorow et al 88).

### 2.2 Types of Lexical Ambiguity

Lexical ambiguity can be divided into homonymy and polysemy, depending on whether or not the meanings are related. The *bark* of a dog versus the *bark* of a tree is an example of homonymy; *review* as a noun and as a verb is an example of polysemy.

The distinction between homonymy and polysemy is central. Homonymy is important because it separates unrelated concepts. If we have a query about 'AIDS' (the disease), and a document contains 'aids' in the sense of a hearing aid, then the word *aids* should not contribute to our belief that the document is relevant to the query. Polysemy is important because the related senses constitute a partial representation of the overall concept. If we fail to group related senses, it is as if we are ignoring some of the occurrences of a query word in a document. So for example, if we are distinguishing words by part-of-speech, and the query contains 'diabetic' as a noun, the retrieval system will exclude instances in which 'diabetic' occurs as an adjective unless we recognize that the noun and adjective senses for that word are related and group them together.

Although there is a theoretical distinction between homonymy and polysemy, it is not always easy to tell

them apart in practice. What determines whether the senses are related? Dictionaries group senses based on part-of-speech and etymology, but as illustrated by the word *review*, senses can be related even though they differ in syntactic category. Senses may also be related etymologically, but be perceived as distinct at the present time (e.g., the 'cardinal' of a church and 'cardinal' numbers are etymologically related). We investigated several methods to identify related senses both across part of speech and within a single homograph, and these will be described in more detail in Section 3.2.1.

### 3 Experiments on Word-Sense Disambiguation

#### 3.1 Preliminary Experiments

Our initial experiments were designed to investigate the following two hypotheses:

**Hypothesis 2** *Word senses provide an effective separation between relevant and non-relevant documents.*

As we saw earlier in the paper, it is possible for a query about 'AIDS' the disease to retrieve documents about 'hearing aids'. But to what extent are such inappropriate matches associated with relevance judgments? This hypothesis predicts that sense mismatches will be more likely to appear in documents that are not relevant than in those that are relevant.

**Hypothesis 3** *Even a small domain-specific collection of documents exhibits a significant degree of lexical ambiguity.*

Little quantitative data is available about lexical ambiguity, and such data as is available is often confined to only a small number of words. In addition, it is generally assumed that lexical ambiguity does not occur very often in domain-specific text. This hypothesis was tested by quantifying the ambiguity for a large number of words in such a collection, and challenging the assumption that ambiguity does not occur very often.

To investigate these hypotheses we conducted experiments with two standard test collections, one consisting of titles and abstracts in Computer Science, and the other consisting of short articles from Time magazine.

The first experiment was concerned with determining how often sense mismatches occur between a query and a document, and whether these mismatches indicate that the document is not relevant. To test this hypothesis we manually identified the senses of the words in the queries for two collections (Computer Science and Time). These words

were then manually checked against the words they matched in the top ten ranked documents for each query (the ranking was produced using a probabilistic retrieval system). The number of sense mismatches was then computed, and the mismatches in the relevant documents were identified.

The second experiment involved quantifying the degree of ambiguity found in the test collections. We manually examined the word tokens in the corpus for each query word, and estimated the distribution of the senses. The number of word types with more than one meaning was determined. Because of the volume of data analysis, only one collection was examined (Computer Science), and the distribution of senses was only coarsely estimated; there were approximately 300 unique query words, and they constituted 35,000 tokens in the corpus.

These experiments provided strong support for Hypotheses 2 and 3. Word meanings are highly correlated with relevance judgements, and the corpus study showed that there is a high degree of lexical ambiguity even in a small collection of scientific text (over 40% of the query words were found to be ambiguous in the corpus). These experiments provided a clear indication of the potential of word meanings to improve the performance of a retrieval system. The experiments are described in more detail in (Krovetz and Croft 92).

#### 3.2 Experiments with different sources of evidence

The next set of experiments were concerned with determining the effectiveness of different sources of evidence for distinguishing word senses. We were also interested in the extent with which a difference in form corresponded to a difference in meaning. For example, words can differ in morphology (authorize/authorized), or part-of-speech (diabetic [noun]/diabetic [adj]), or in their ability to appear in a phrase (database/data base). They can also exhibit such differences, but represent different concepts, such as author/authorize, sink[noun]/sink[verb], or stone wall/stonewall. Our default assumption was that a difference in form is associated with a difference in meaning unless we could establish that the different word forms were related.

##### 3.2.1 Linking related word meanings

We investigated two approaches for relating senses with respect to morphology and part of speech: 1) exploiting the presence of a variant of a term within its dictionary definition, and 2) using the overlap of the words in the definitions of suspected variants.

For example, *liable* appears within the definition of *liability*, and this is used as evidence that those words are related. Similarly, *flat* as a noun is defined as ‘a flat tire’, and the presence of the word in its own definition, but with a different part of speech, is taken as evidence that the noun and adjective meanings are related. We can also compute the overlap between the definitions of *liable* and *liability*, and if they have a significant number of words in common then that is evidence that those meanings are related. These two strategies could potentially be used for phrases as well, but phrases are one of the areas where dictionaries are incomplete, and other methods are needed for determining when phrases are related. We will discuss this in Section 3.2.4.

We conducted experiments to determine the effectiveness of the two methods for linking word senses. In the first experiment we investigated the performance of a part-of-speech tagger for identifying the related forms. These related forms (e.g., *flat* as a noun and an adjective) are referred to as instances of zero-affix morphology, or functional shift (Marchand 63). We first tagged all definitions in the dictionary for words that began with the letter ‘W’. This produced a list of 209 words that appeared in their own definitions with a different part of speech. However, we found that only 51 (24%) were actual cases of related meanings. This low success rate was almost entirely due to tagging error. That is, we had a false positive rate of 76% because the tagger indicated the wrong part of speech. We conducted a failure analysis and it indicated that 91% the errors occurred in idiomatic expressions (45 instances) or example sentences associated with the definitions (98 instances). We therefore omitted idiomatic senses and example sentences from further processing and tagged the rest of the dictionary.<sup>2</sup>

The result of this experiment is that the dictionary contains at least 1726 senses in which the headword was mentioned, but with a different part of speech, of which 1566 were in fact related (90.7%). We analyzed the distribution of the connections, and this is given in Table 1 ( $n = 1566$ ).

However, Table 1 does not include cases in which the word appears in its definition, but in an inflected form. For example, ‘cook’ as a noun is defined as ‘a person who prepares and cooks food’. Unless we recognize the inflected form, we will not capture all of the instances. We therefore repeated the procedure, but allowing for inflectional variants. The result is given in Table 2 ( $n = 1054$ ).

We also conducted an experiment to determine

<sup>2</sup>Idiomatic senses were identified by the use of font codes.

the effectiveness of capturing related senses via word overlap. The result is that if the definitions for the root and variant had two or more words in common,<sup>3</sup> 93% of the pairs were semantically related. However, of the sense-pairs that were actually related, two-thirds had only one word in common. We found that 65% of the sense-pairs with one word in common were related. Having only one word in common between senses is very weak evidence that the senses are related, and it is not surprising that there is a greater degree of error.

The two experiments, tagging and word overlap, were found to be to be highly effective once the common causes of error were removed. In the case of tagging the error was due to idiomatic senses and example sentences, and in the case of word overlap the error was links due to a single word in common. Both methods have approximately a 90% success rate in pairing the senses of morphological variants if those problems are removed. The next section will discuss our experiments with morphology.

### 3.2.2 Experiments with Morphology

We conducted several experiments to determine the impact of grouping morphological variants on retrieval performance. These experiments are described in detail in (Krovetz 93), so we will only summarize them here.

Our experiments compared a baseline (no stemming) against several different morphology routines: 1) a routine that grouped only inflectional variants (plurals and tensed verb forms), 2) a routine that grouped inflectional as well as derivational variants (e.g., -ize, -ity), and 3) the Porter stemmer (Porter 80). These experiments were done with four different test collections which varied in both size and subject area. We found that there was a significant improvement over the baseline performance from grouping morphological variants.

Earlier experiments with morphology in IR did not report improvements in performance (Harman 91). We attribute these differences to the use of different test collections, and in part to the use of different retrieval systems. We found that the improvement varies depending on the test collection, and that collections that were made up of shorter documents were more likely to improve. This is because morphological variants can occur within the same document, but they are less likely to do so in documents that are short. By grouping morphological variants, we are helping to improve access to the shorter documents. However, we also found improvements even

<sup>3</sup>Excluding closed class words, such as *of* and *for*.

in a collection of legal documents which had an average length of more than 3000 words.

We also found it was very difficult to improve retrieval performance over the performance of the Porter stemmer, which does not use a lexicon. The absence of a lexicon causes the Porter stemmer to make errors by grouping morphological "false friends" (e.g., *author/authority*, or *police/policy*). We found that there were three reasons why the Porter stemmer improves performance despite such groupings. The first two reasons are associated with the heuristics used by the stemmer: 1) some word forms will be grouped when one of the forms has a combination of endings (e.g., *-ization* and *-ize*). We empirically found that the word forms in these groups are almost always related in meaning. 2) the stemmer uses a constraint on the form of the resulting stem based on a sequence of consonants and vowels; we found that this constraint is surprisingly effective at separating unrelated variants. The third reason has to do with the nature of morphological variants. We found that when a word form *appears* to be a variant, it often *is* a variant. For example, consider the grouping of *police* and *policy*. We examined all words in the dictionary in which a word ended in 'y', and in which the 'y' could be replaced by 'e' and still yield a word in the dictionary. There were 175 such words, but only 39 were clearly unrelated in meaning to the presumed root (i.e., cases like *policy/police*). Of the 39 unrelated word pairs, only 14 were grouped by the Porter stemmer because of the consonant/vowel constraints. We also identified the morphological "false friends" for the 10 most frequent suffixes. We found that out of 911 incorrect word pairs, only 303 were grouped by the Porter stemmer.

Finally, we found that conflating inflectional variants harmed the performance of about a third of the queries. This is partially a result of the interaction between morphology and part-of-speech (e.g., a query that contains *work* in the sense of *theoretical work* will be grouped with all of the variants associated with the the verb - *worked*, *working*, *works*); we note that some instances of *works* can be related to the singular form *work* (although not necessarily the right meaning of *work*), and some can be related to the untensed verb form. Grouping inflectional variants also harms retrieval performance because of an overlap between inflected forms and uninflected forms (e.g., *arms* can occur as a reference to weapons, or as an inflected form of *arm*). Conflating these forms has the effect of grouping unrelated concepts, and thus increases the net ambiguity.

Our experiments with morphology support our ar-

gument about distinguishing homonymy and polysemy. Grouping related morphological variants makes a significant improvement in retrieval performance. Morphological false friends (*policy/police*) often provide a strong separation between relevant and non-relevant documents (see (Krovetz and Croft 92)). There are no morphology routines that can currently handle the problems we encountered with inflectional variants, and it is likely that separating related from unrelated forms will make further improvements in performance.

### 3.2.3 Experiments with Part of Speech

Relatively little attention has been paid in IR to the differences in a word's part of speech. These differences have been used to help identify phrases (Dillon and Gray 83), and as a means of filtering for word sense disambiguation (to only consider the meanings of nouns (Voorhees 93)). To the best of our knowledge the differences have never been examined for distinguishing meanings within the context of IR.

The aim of our experiments was to determine how well part of speech differences correlate with differences in word meanings, and to what extent the use of meanings determined by these differences will affect the performance of a retrieval system. We conducted two sets of experiments, one concerned with homonymy, and one concerned with polysemy. In the first experiment the Church tagger was used to identify part-of-speech of the words in documents and queries. The collections were then indexed by the word tagged with the part of speech (i.e., instead of indexing 'book', we indexed 'book/noun' and 'book/verb').<sup>4</sup> A baseline was established in which all variants of a word were present in the query, regardless of part of speech variation; the baseline did not include any morphological variants of the query words because we wanted to test the interaction between morphology and part-of-speech in a separate experiment. The baseline was then compared against a version of the query in which all variations were eliminated except for the part of speech that was correct (i.e., if the word was used as a noun in the original query, all other variants were eliminated). This constituted the experiment that tested homonymy. We then identified words that were related in spite of a difference in part of speech; this was based on the data that was produced by tagging the dictionary (see Section 3.2.1). Another version of the queries was constructed in which part of speech variants were retained if the meaning was related,

<sup>4</sup>In actuality, we indexed it with whatever tags were used by the tagger; we are just using 'noun' and 'verb' for purposes of illustration.

and this was compared to the previous version.

When we ran the experiments, we found that performance decreased compared with the baseline. However, we found many cases where the tagger was incorrect.<sup>5</sup> We were unable to determine whether the results of the experiment were due to the incorrectness of the hypothesis being tested (that distinctions in part of speech can lead to an improvement in performance), or to the errors made by the tagger. We also assumed that a difference in part-of-speech would correspond to a difference in meaning. The data in Table 1 and Table 2 shows that many words are related in meaning despite a difference in part-of-speech. Not all errors made by the tagger cause decreases in retrieval performance, and we are in the process of determining the error rate of the tagger on those words in which part-of-speech differences are also associated with a difference in concepts (e.g., *novel* as a noun and as an adjective).<sup>6</sup>

### 3.2.4 Experiments with Phrases

Phrases are an important and poorly understood area of IR. They generally improve retrieval performance, but the improvements are not consistent. Most research to date has focused on syntactic phrases, in which words are grouped together because they are in a specific syntactic relationship (Fagan 87), (Smeaton and Van Rijsbergen 88). The research in this section is concerned with a subset of these phrases, namely those that are *lexical*. A lexical phrase is a phrase that might be defined in a dictionary, such as *hot line* or *back end*. Lexical phrases can be distinguished from phrases such as *sanctions against South Africa* in that the meaning of a lexical phrase cannot necessarily be determined from the meaning of its parts.

Lexical phrases are generally made up of only two or three words (overwhelmingly just two), and they usually occur in a fixed order. The literature mentions examples such as *blind venetians* vs. *venetian blinds*, or *science library* vs. *library science*, but these are primarily just cute examples. It is very rare that the order could be reversed to produce a different concept.

Although dictionaries contain a large number of phrasal entries, there are many lexical phrases that are missing. These are typically proper nouns (*United States*, *Great Britain*, *United Nations*) or technical concepts (*operating system*, *specific heat*,

*due process*, *strict liability*). We manually identified the lexical phrases in four different test collections (the phrases were based on our judgement), and we found that 92 out of 120 phrases (77%) were not found in the Longman dictionary. A breakdown of the phrases is given in (Krovetz 95).

For the phrase experiment we not only had to identify the lexical phrases, we also had to identify any related forms, such as *database/data base*. This was done via brute force — a program simply concatenated every adjacent word in the database, and if it was also a single word in the collection it printed out the pair. We tested this with the Computer Science and Time collections, and used those results to develop an exception list for filtering the pairs (e.g., do not consider 'special ties/specialties'). We represented the phrases using a proximity operator<sup>7</sup> and tried several experiments to include the related form when it was found in the corpus.

We found that retrieval performance *decreased* for 118 out of 120 phrases. A failure analysis indicated that this was due to the need to assign partial credit to individual words of a phrase. The component words were always related to the meaning of the compound as a whole (e.g., *Britain* and *Great Britain*).

We also found that most of the instances of open/closed compounds (e.g., *database/data base*) were related. Cases like 'stone wall/stonewall' or 'bottle neck/bottleneck' are infrequent. The effect on performance of grouping the compounds is related to the relative distribution of the open and closed forms. *Database/data base* occurred in about a 50/50 distribution, and the queries in which they occurred were significantly improved when the related form was included.

### 3.2.5 Interactions between Sources of Evidence

We found many interactions between the different sources of evidence. The most striking is the interaction between phrases and morphology. We found that the use of phrases acts as a filter for the grouping of morphological variants. Errors in morphology generally do not hurt performance within the restricted context. For example, the Porter stemmer will reduce *department* to *depart*, but this has no effect in the context of the phrase 'Justice department'.

<sup>5</sup>See (Krovetz 95) for more details about these errors.

<sup>6</sup>There are approximately 4000 words in the Longman dictionary which have more than one part-of-speech. Less than half of those words will be like *novel*, and we are examining them by hand.

<sup>7</sup>The proximity operator specifies that the query words must be adjacent and in order, or occur within a specific number of words of each other.

## 4 Conclusion

Most of the research on lexical ambiguity has not been done in the context of an application. We have conducted experiments with hundreds of unique query words, and tens of thousands of word occurrences. The research described in this paper is one of the largest studies ever done. We have examined the lexicon as a whole, and focused on the distinction between homonymy and polysemy. Other research on resolving lexical ambiguity for IR (e.g., (Sanderson 94) and (Voorhees 93)) does not take this distinction into account.

Our research supports the argument that it is important to distinguish homonymy and polysemy. We have shown that natural language processing results in an improvement in retrieval performance (via grouping related morphological variants), and our experiments suggest where further improvements can be made. We have also provided an explanation for the performance of the Porter stemmer, and shown it is surprisingly effective at distinguishing variant word forms that are unrelated in meaning. The ex-

periment with part-of-speech tagging also highlighted the importance of polysemy; more than half of all words in the dictionary that differ in part of speech are also related in meaning. Finally, our experiments with lexical phrases show that it is crucial to assign partial credit to the component words of a phrase. Our experiment with open/closed compounds indicated that these forms are almost always related in meaning.

The experiment with part-of-speech tagging indicated that taggers make a number of errors, and our current work is concerned with identifying those words in which a difference in part of speech is associated with a difference in meaning (e.g., *train* as a noun and as a verb). The words that exhibit such differences are likely to affect retrieval performance. We are also examining lexical phrases to decide how to assign partial credit to the component words. This work will give us a better idea of how language processing can provide further improvements in IR, and a better understanding of language in general.

	V	N	Adj	Adv	Proportion
V		1167 (95%)	57 (4.6%)	3 (0.4%)	77.8%
N	63 (32.6%)		126 (65.3%)	4 (2.0%)	12.2%
Adj	15 (15.2%)	82 (82.8%)			6.3%
Adv		23 (41.8%)	31 (56.4%)		3.3%

Table 1: Distribution of zero-affix morphology within dictionary definitions

	V	N	Adj	Adv	Proportion
V		239 (97%)	7 (3.0%)		23%
N	486 (85%)		87 (15%)	1 (0.1%)	54%
Adj	15 (14%)	87 (81%)		4 (3.7%)	10%
Adv	2 (2%)	4 (3%)	119 (95%)		12%

Table 2: Distribution of zero-affix morphology (inflected)

## Acknowledgements

I am grateful to Dave Waltz for his comments and suggestions.

## References

- Chodorow M and Y Ravin and H Sachar, "Tool for Investigating the Synonymy Relation in a Sense Disambiguated Thesaurus", in *Proceedings of the Second Conference on Applied Natural Language Processing*, pp. 144-151, 1988.
- Church K, "A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text", in *Proceedings of the Second Conference on Applied Natural Language Processing*, pp. 136-143, 1988.
- Dagan I and A Itai, "Word Sense Disambiguation Using a Second Language Monolingual Corpus", *Computational Linguistics*, Vol. 20, No. 4, 1994.
- Dillon M and A Gray, "FASIT: a Fully Automatic Syntactically Based Indexing System", *Journal of the American Society of Information Science*, Vol. 34(2), 1983.
- Fagan J, "Experiments in Automatic Phrase Indexing for Document Retrieval: A Comparison of Syntactic and Non-Syntactic Methods", PhD dissertation, Cornell University, 1987.
- Grishman R and Kittredge R (eds), *Analyzing Language in Restricted Domains*, LEA Press, 1986.
- Halliday M A K, "Lexis as a Linguistic Level", in *In Memory of J. R. Firth*, Bazell, Catford and Halliday (eds), Longman, pp. 148-162, 1966.
- Harman D, "How Effective is Suffixing?", *Journal of the American Society for Information Science*, Vol 42(1), pp. 7-15, 1991
- Helm S., "Closer Than You Think", *Medicine and Computer*, Vol. 1, No. 1., 1983
- Kilgarriff A, "Corpus Word Usages and Dictionary Word Senses: What is the Match? An Empirical Study", in *Proceedings of the Seventh Annual Conference of the UW Centre for the New OED and Text Research: Using Corpora*, pp. 23-39, 1991.
- Krovetz R and W B Croft, "Lexical Ambiguity and Information Retrieval", *ACM Transactions on Information Systems*, pp. 145-161, 1992.
- Krovetz R, "Viewing Morphology as an Inference Process", in *Proceedings of the Sixteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 191-202, 1993.
- Krovetz R, "Word Sense Disambiguation for Large Text Databases", PhD dissertation, University of Massachusetts. 1995.
- Marchand H, "On a Question of Contrary Analysis with Derivational Connected but Morphologically Uncharacterized Words", *English Studies*. Vol. 44, pp. 176-187, 1963.
- Popovic M and P Willet, "The Effectiveness of Stemming for Natural Language Access to Slovene Textual Data", in *Journal of the American Society for Information Science*, Vol. 43(5), pp. 384-390, 1992.
- Porter M, "An Algorithm for Suffix Stripping", *Program*, Vol. 14 (3), pp. 130-137, 1980.
- Proctor P., *Longman Dictionary of Contemporary English*, Longman, 1978.
- Salton G., *Automatic Information Organization and Retrieval*, McGraw-Hill, 1968.
- Salton G. and McGill M., *Introduction to Modern Information Retrieval*, McGraw-Hill, 1983.
- Sanderson M, "Word Sense Disambiguation and Information Retrieval", in *Proceedings of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 142-151, 1994.
- Small S., Cottrell G., and Tannenhaus M. (eds). *Lexical Ambiguity Resolution*, Morgan Kaufmann, 1988.
- Smeaton A and C J Van Rijsbergen, "Experiments on Incorporating Syntactic Processing of User Queries into a Document Retrieval Strategy", in *Proceedings of the Eleventh Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 31-51, 1988.
- Talmy L, "Lexicalization Patterns: Semantic Structure in Lexical Forms", in *Language Typology and Syntactic Description. Volume III: Grammatical Categories and the Lexicon*, T Shopen (ed), pp. 57-160, Cambridge University Press, 1985.
- Van Rijsbergen C. J., *Information Retrieval*, Butterworths, 1979.
- Voorhees E, "Using WordNet to Disambiguate Word Senses for Text Retrieval", in *Proceedings of the Sixteen Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pp. 171-180, 1993.
- Yarowsky D, "Word Sense Disambiguation Using Statistical Models of Roget's Categories Trained on Large Corpora", in *Proceedings of the 14th Conference on Computational Linguistics, COLING-92*, pp. 454-450, 1992.