

Alternative Phrases and Natural Language Information Retrieval

Gann Bierner

Division of Informatics
University of Edinburgh
gbierner@cogsci.ed.ac.uk

Abstract

This paper presents a formal analysis for a large class of words called *alternative markers*, which includes **other(than)**, **such(as)**, and **besides**. These words appear frequently enough in dialog to warrant serious attention, yet present natural language search engines perform poorly on queries containing them. I show that the performance of a search engine can be improved dramatically by incorporating an approximation of the formal analysis that is compatible with the search engine's operational semantics. The value of this approach is that as the operational semantics of natural language applications improve, even larger improvements are possible.

1 Introduction

Consider the following examples discovered in a corpus of queries submitted to the **Electric Knowledge** search engine¹², the successor of the **On Point** natural language search system described in Cooper (1997). Each consists of a query, a response (not shown), and then a follow-up query.

- (1) *What is the drinking age in Afghanistan?*
*What is the drinking age in **other** countries?*
- (2) *Where can I find web browsers for download?*
*Where can I find **other** web browsers **than** Netscape for download?*
- (3) *Where can I find a list of all the shoe manufacturers in the world?*
*Where can I find shoes made by Buffalino, **such as** the Bushwackers?*
- (4) *Where are online auctions indexed?*
*Are there **other** auction search engines **besides** BidFind?*

¹formerly known as The Electric Monk

²<http://www.electricknowledge.com>

In each case, particular words are used to constrain the space of appropriate answers: e.g. **such (as)**, **other (than)**, and **besides**. I call these words, and others like them, *alternative markers*, and alternative markers along with their syntactic argument (e.g. *other countries*), I call *alternative phrases*. Alternative phrases that are closely bound to the noun phrase to which they refer, like those above, I call *connected alternative phrases* (modeled after similar terminology in Hoeksema (1995)). There are also *free alternative phrases*, such as *Other than Fido, every dog likes going for walks*, which are not discussed here but are discussed in depth in Bierner (2001). I have found queries containing all these forms despite the fact that no current NLIR system can handle them, either ignoring the alternative phrase or, worse, treating the queries as if the alternative marker were absent.

The fact that these phrases define the space of correct answers makes it absolutely necessary to correctly interpret them in an IR application. The user requires countries that are not Afghanistan, web browsers that are not Netscape, shoes with similar properties to Bushwackers, and auction search engines that are not BidFind. Answers that do not conform are wrong.

Another feature of alternative markers are their presuppositions. Through these presuppositions, alternative markers can provide a rich source of knowledge about the world, as Hearst (1992) has already recognized. For example, these queries imply that Afghanistan is a country, Netscape is a web browser, Bushwackers are shoes, and BidFind is an auction search engine. Anaphoric resolution can sometimes be critical for these inferences. In (1), for instance, *other countries* anaphorically depends on *Afghanistan* in the previous query. While not as obviously important to natural language information retrieval, this property of alternative phrases can be used to improve future queries (see Section 4.1).

The purpose of this paper is, first, to briefly provide a well-founded semantic analysis for alternative phrases that is amenable to computation. The main thrust is then to show that a motivated approximation of this analysis can dramatically improve the results of a practical natural language application—in particular, a natural language search engine. The value of this two-step approach is that as the operational semantics of practical applications are gradually extended, progressively more extensive approximations can be transparently incorporated.

2 Previous Work

Work on alternative phrases has previously been limited to Hearst’s work in connection with knowledge base construction and formal semantics by Hoeksema and von Fintel. Hearst (1992) demonstrates cases where pattern matching can be used to extract knowledge from some constructions with alternative phrases and gives the patterns in (5) as examples.

- (5) a. **such** NP as {NP,*} {**or|and**} NP
 b. NP {, NP}* {,} {**or|and**} **other** NP
 c. NP {,} **including** {NP,*} {**or|and**} NP
 d. NP {,} **especially** {NP,*} {**or|and**} NP

This technique is adequate for the purpose of acquiring hyponyms from large corpora because the goal is to take advantage of easily available information, not to handle a wide variation of linguistic phenomena. Handling a full range of natural language queries requires a deeper approach. For example, pattern matching alone cannot account for anaphoric reference such as in (1) or examples where discourse knowledge invalidates the hyponym relation as in (6).

- (6) John’s pet dog, Fido, was sick.
 So John and the other dogs went for a walk.

In contrast, von Fintel (1993) and Hoeksema (1995) provide in-depth semantic analyses of a limited class of alternative phrases; namely they focus entirely on *exceptive phrases* (ways of referring to exceptions) illustrated by the lexical items **but** and **except (for)**. The thrust of their analyses is directed towards how these words interact with quantifiers to determine the final set of entities. Since in queries, quantifiers rarely interact with alternative phrases in the manner discussed in Hoeksema and von Fintel’s work, their analyses have not been carried over into the present work.

3 Analysis

I present a formal approach to alternative phrases that is wider in scope than the alternatives reviewed in Section 2 (although less detailed in some respects than von Fintel and Hoeksema’s work).

3.1 Presupposition and Assertion

In my analysis of alternative phrases, I make use of the pragmatic view of presuppositions explored by Lewis (1979) and Stalnaker (1974) which, stated loosely, sees them as propositions that must be true for an utterance to make sense. (For an overview of presupposition, see Beaver (1997).) The semantics of lexical entries are separated into assertion and presupposition as in Stalnaker (1974) and Karttunen and Peters (1979). The idea is also used in Webber et al. (1999) to capture anaphoric (non-structural) links between discourse connectives and material derivable from previous discourse, and in Stone and Doran (1997) and Stone and Webber (1998) for natural language generation.

Lexical entries are written in the following form, where the semantic parameters scope both the assertion and presuppositions:

$$\text{word} \vdash \begin{cases} \text{syn} : & \text{syntactic category} \\ \text{sem} : & \lambda \dots \left\{ \begin{array}{l} \text{assert} : \text{proposition} \\ \text{presup} : \text{proposition}^* \end{array} \right. \end{cases}$$

3.2 Alternative Sets

The concept of alternative sets plays an important role in the semantics of alternative phrases. An alternative set is a set of propositions which differ with respect to how one or more arguments are filled. For example, the alternative set $\{\text{like}(\text{mary}, \text{jen}), \text{like}(\text{mary}, \text{bob}), \dots\}$ represents the entities that Mary likes. An early discussion of these structures is provided in Karttunen and Peters (1979) where an analysis is given for the focus particle **even**. Alternative sets are also used by Rooth (1985) and Rooth (1992) to develop a detailed account of focus, particularly with the focus particle **only**.

My analysis approximates this set of properties as a pair consisting of a set of entities (e.g. $\{\text{jen}, \text{bob}, \dots\}$) and the property they share (e.g. $\lambda x. \text{like}(\text{mary}, x)$). My analyses of alternative phrases uses the relation $\text{alts}(p, q)$ which, intuitively, specifies that the two sets of entities denoted by p and q can be found together in at least one alternative set in the knowledge base. The description component of the alternative set (i.e.

the property) need not be known. It is important to note that although here I focus on unifying such structures, I also make use of the fact that *alts* is a relation that is symmetric and reflexive, but not transitive.

The alternative phrases I have analyzed fall into two classes: those that assemble a set from elements and those that excise a set from a larger set (as in exceptive phrases). In either case, one particular set of elements is of interest, the *figure*. With assembly words, the figure is either admitted into the set or combined with a *complement* to form a set. With excision, the figure is explicitly excluded from the ground. The figure may derive from structurally-related constituents (as with **besides**), or it may be presupposed (as with **other**).

3.3 A Grammar Formalism

I implement my analyses with Combinatory Categorical Grammar (Steedman, 1996; Steedman, 2000). CCG is a lexicalized grammar that encodes both the syntactic and semantic properties of a word in the lexicon. For the analyses presented in this paper, standard Categorical Grammar suffices.

A minor variation is that rather than having the basic categories N and NP, I simply use NP. Noun phrases with and without determiners are distinguished with the *bare* feature.

3.4 An Example

In this section, I provide an analysis of one syntactic form of **other** in order to illustrate the semantic technique described above. Discussion of alternate syntactic forms and other alternative markers can be found in Bierner (2001).

The semantic analysis below defines **other** as an excision word that excludes the figure from the ground. The figure is a free variable that must be provided from the common ground or discourse, as is the case in (1).

$$\mathbf{other} \vdash \begin{cases} \mathit{syn} : & \text{NP}_{\text{comp+}, \text{eq-}} / \text{NP}_{\text{bare+}} \\ \mathit{sem} : & \lambda g \begin{cases} \mathit{assert} : & \lambda x.g(x) \wedge \neg f(x) \\ \mathit{presup} : & \begin{cases} \forall x.f(x) \rightarrow g(x) \\ \mathit{alts}(f, \lambda x.g(x) \wedge \neg f(x)) \end{cases} \end{cases} \end{cases}$$

The analysis allows the derivation in Figure 1.

$$\frac{\frac{\mathbf{other}}{\text{NP/NP} : \lambda g \lambda x.g(x) \wedge \neg f(x)} \quad \frac{\mathbf{countries}}{\text{NP} : \mathbf{country}}}{\text{NP} : \lambda x.\mathbf{country}(x) \wedge \neg f(x)} \quad \text{presupposition set: } \begin{cases} \forall x.f(x) \rightarrow \mathbf{country}(x) \\ \mathit{alts}(f, \lambda x.\mathbf{country}(x) \wedge \neg f(x)) \end{cases}$$

Figure 1: Derivation for *other countries*

At this point, the semantics is dependent on the free variable *f*, the figure. This is reflected by the fact that, in isolation, *other countries* does not make sense. Although such anaphoric reference is difficult to resolve, in some constructions we can identify the figure without bringing full resolution techniques to bear—as we would have to in (1). Some of these constructions, as in Figure 2, are those that contain the word **than**, whose analysis is given below.

$$\mathbf{than} \vdash \begin{cases} \mathit{syn} : & \text{NP} \setminus \text{NP}_{\text{eq-}, \text{comp+}} / \text{NP} \\ \mathit{sem} : & \lambda x \lambda y \begin{cases} \mathit{assert} : & y \\ \mathit{presup} : & \mathit{alts}(x, y) \end{cases} \end{cases}$$

The presupposition set in Figure 2 is the union of the presuppositions of **other** and **than**, as bound during the derivation. The remaining variable, *f*, can be determined solely from the derivation’s presupposition set using the old AI planning heuristic “use existing objects” (Sacchetti, 1977) to avoid inventing new objects when others are already available. In particular, we can unify $\mathit{alts}(f, \lambda x.\mathbf{browser}(x) \wedge \neg f(x))$ and $\mathit{alts}(\mathbf{netscape}, \lambda x.\mathbf{browser}(x) \wedge \neg f(x))$, discovering that *f*, the figure, is **netscape**. This then instantiates the remaining presupposition, yielding $\forall x.\mathbf{netscape}(x) \rightarrow \mathbf{browser}(x)$: i.e. Netscape is a browser. Unifying logical forms to instantiate variables in this way follows the “interpretation as abduction” paradigm (Hobbs et al., 1993), where this merging is performed to exploit redundancy for “getting a minimal, and hence a best, interpretation.”

Similar analyses in terms of alternative sets have been developed for many other alternative phrases in Bierner (2001).

In the next section I show that practical applications such as natural language search engines can benefit from appropriate approximations of this kind of analysis.

4 Natural Language IR

There are a variety of techniques for allowing natural language queries in information retrieval systems. The simplest approach is simply to remove the “function words” from the query and use the remaining words in a standard keyword search (Alta Vista). In more complex approaches, pattern matching (the EK search engine), parsing (Ask Jeeves), and machine learning (Zelle and Mooney, 1993) techniques can support the association of more appropriate keywords with a query.

I will concentrate on the pattern matching technique of the Electric Knowledge search engine and

other	web browsers	than	Netscape
$\text{NP}_{\text{comp}+}/\text{NP} :$ $\lambda g \lambda x. g(x) \wedge \neg f(x)$	$\text{NP}_{\text{comp}-} :$ browser	$(\text{NP} \setminus \text{NP}_{\text{comp}+})/\text{NP} :$ $\lambda x \lambda y. y$	$\text{NP} :$ netscape
$\text{NP}_{\text{comp}+} : \lambda x. \text{browser}(x) \wedge \neg f(x)$		$\text{NP} \setminus \text{NP}_{\text{comp}+} : \lambda y. y$	
$\text{NP} : \lambda x. \text{browser}(x) \wedge \neg f(x)$			

presupposition set: $\left\{ \begin{array}{l} \forall x. f(x) \rightarrow \text{browser}(x) \\ \text{alts}(f, \lambda x. \text{browser}(x) \wedge \neg f(x)) \\ \text{alts}(\text{netscape}, \lambda x. \text{browser}(x) \wedge \neg f(x)) \end{array} \right.$

Figure 2: Derivation for *other web browsers than Netscape*

shown how a theory of alternative phrases can drastically improve results.

4.1 The EK search engine’s Operational Semantics

The Electric Knowledge search engine uses pattern recognition to transform a natural language question into a series of increasingly more general Boolean queries that can be used with standard back-end retrieval techniques. The question is filtered through a hierarchy of regular expressions, and hand-crafted rules are used to extract information from the question into Boolean expressions. The regular expression matching is aided by an ISA hierarchy such that generalizations of keywords can be captured. As mentioned in Section 1, the fact that the presuppositions of alternative phrases encode hyponym information can be useful in augmenting this aspect of systems like the EK search engine.

This technique suffers from the fact that, in order to be tractable, this set of patterns is limited and important information in the query can be lost. In particular, the Electric Knowledge search engine does not have patterns that attempt to associate alternative phrases with appropriate pieces of boolean query.

To overcome this, an appropriate approximation of the semantic result of my analysis that is compatible with the back end search system must be found. For the Electric Knowledge search engine (similar approaches are certainly possible for other NLIR systems), a hybrid query has been introduced to account for alternative phrases, which combines a natural language query with further restrictions added in a system-specific language. The syntax is shown in (7).

- (7) Query :|: ANSWER NOT NEAR (| word list)
- (8) What are some web browsers? :|: ANSWER NOT NEAR (| netscape)

The natural language query is separated from the restrictions by the :|: symbol. The restrictions

specify that the answer to the query must not be near to certain words.

The hybrid query in (8), for example, is a transformation of the original query *What are some other web browsers than Netscape?*. The EK search engine uses the natural language part of the query to initially locate possible answering documents. The rest of the query is used when gathering evidence for including a document in the final results. The EK search engine finds a location in the document that should answer the query and then compares it against the criteria appended to the end of the query. If it does not meet the criteria (that it not be near the word Netscape), another location is tried. If there are no more possible answers, the document is rejected.

This is, of course, not exactly what the original query meant. However, it is superior to queries like ‘‘browsers’’ AND NOT ‘‘netscape’’ which rejects all pages containing Netscape, even if they also contain other browsers. The evaluation in Section 5 shows that this operational semantics is sufficient to dramatically improve the results for queries with alternative phrases.

4.2 The Algorithm

Instead of using the EK search engine’s pattern matching on the initial question, the algorithm does a post-analysis of the syntactic structure produced by parsing the question with my analysis.

The algorithm recursively descends the derivation, searching for semantic forms that result from alternative phrases. The information from the alternative phrase is removed and then appended to the end of the hybrid query in a different form. In Figure 2, for example, the information *other than Netscape* is removed, leaving only *web browsers*, to form the hybrid query in (8).

	Physics		EE		Total
	Stud	Tut	Stud	Tut	
in addition to	0	0	0	3	3
besides	1	3	0	2	6
another	56	124	10	38	228
especially	0	1	0	0	1
except	0	17	1	1	19
other	107	484	36	59	686
in particular	0	6	0	0	6
such	2	18	3	10	33
unlike	0	0	0	1	1
Total	166	653	50	114	983

Alternative Phrases per Dialog					
Dialogs	203	203	66	66	269
AP/dialog	0.82	3.22	0.76	1.73	3.65

Alternative Phrases in Queries per Dialog					
Query APs	51	261	16	86	414
per dialog	0.25	1.29	0.24	1.3	1.54

Table 1: Frequency of Alternative Phrases in Dialog

5 Evaluation

5.1 Frequency of Alternative Phrases

First, it is useful to determine how many queries contain alternative phrases in order to judge how large a problem this really is. Unfortunately, this is complicated by the fact that users, in general, *know* that such constructions are not understood by search engines, so they avoid them. In fact, even in NLIR systems, users often use keywords even though doing so performs worse than asking natural language questions. They do this because they do not trust the system. Work will be necessary to improve users' awareness of NL capabilities through advertising and by implementing new user interfaces. Work will also be needed to take more account of the fact that search is often an iterative and even interactive process. As noted by Lewis and Sparck-Jones (1996), the results of large-scale document-retrieval competitions such as the Text Retrieval Conference (TREC, 2000) do not necessarily reflect the experience many users have with retrieval systems.

In the meantime, I have attempted to find a baseline for this number by considering two corpora of human/human dialogs. The corpora are both from tutorial situations where a mentor helps a student through problems in the subject of Physics for one corpus (VanLehn et al., 1998; VanLehn et al., in press), and Basic Electricity and Electronics (Rose et al., 1999), for the other. Tutoring dialogs are an ideal place to look for data relevant to NLIR because they consist entirely of one party attempting to elicit information from

the other. In some cases, it is the tutor eliciting information from the student, and in others it is the other way around.

Table 1 shows the frequencies of some alternative phrases. A more illuminating statistic is how often alternative phrases appear in a single dialog. I consider a dialog to be a single *session* between the student and tutor where the discussion of each problem is considered a separate session. The table shows that in 269 total dialogs, each dialog contained, on average, 3.65 alternative phrases. If one only considers alternative phrases that occur in the context of a question, there are, on average, 1.54 alternative phrases per dialog. I consider an alternative phrase to be in a question context if it is in the same dialog turn as a question. Because tutors ask questions in order to lead the student to the answer, it is perhaps better to consider just the student data, where a quarter of the dialogs contained question contexts with alternative phrases.

This data is not meant to be considered a rigorous result, but it is a strong indication that any query-answering system is likely to confront an alternative phrase during the course of interacting with a user. Furthermore, the data shows that during the course of the interaction it will be appropriate for the system to *respond* using an alternative phrase, which is important when considering more responsive search engines than are available today.

It is also interesting to note that a wide variety of alternative phrases occur in this data. (9) contains some examples. Because excision words, especially **other**, are by far the most frequent, I have only considered them in the evaluation presented here.

- (9)
- The battery is the green cylinder right? I don't see anything negative **other than** #5.
 - And what do you have **in addition to** voltage?
 - Are there any **other** forces on that knob **besides** that one you've labeled W1?

5.2 Potential Improvement

Of interest is the potential for improving various search engines through a better way of treating alternative phrases. To start with, given a set of queries containing alternative phrases, it is important to see how well these systems perform without enhancement. Table 2 shows the performance of Alta Vista, Ask Jeeves, and the EK search engine on eight excision examples taken from the

	AV	Jeeves	Monk
Cancer	2/9, 6	2/5, 3	2/3, 0
BidFind	2/7, 5	1/6, 0	0/0, 0
Jobs	2/9, 0	2/5, 0	3/9, 6
Warts	0/9, 6	1/4, 2	0/1, 1
Drinking	4/10, 0	0/5, 1	0/2, 0
Browsers	2/8, 6	3/5, 2	0/9, 8
Witches	0/10, 10	0/5, 4	0/0, 0
Hondo	0/7, 6	1/6, 3	0/0, 0
Avr. Precision	17%	25%	20%
Avr. % false positives due to searching for the figure	67%	53%	58%

Table 2: Potential Improvement for NLIR Systems

corpus of EK search engine queries. For a data entry “ $x/y, z$ ”, y is the total number of returned documents and x are those which contain an answer to the query. z are the number of answers that are wrong because they are about the subject that was explicitly being excluded in the query.

As the data shows, none of the search engines fare particularly well. The precision for all three is around 20%. That is, only about one in five of the responses contain an answer to the query. Furthermore, from a half to two thirds of the incorrect responses were specifically about the subject the query wanted to exclude, displaying little or no understanding of excision alternative phrases.

The point is not to draw any conclusions about the relative merits of the search engines from this test. Rather, it is that each NLIR system shows room for improvement. Since I will demonstrate that improvement only for the Electric Knowledge search engine, next, this shows that that improvement is not due to exceptionally bad prior performance by the EK search engine.

5.3 Enhancing the EK search engine

Table 3 shows the results of asking the EK search engine questions in three different forms: without an alternative phrase, with an alternative phrase that has not been translated, and with the alternative phrase translated as described in Section 4.2. The first row of the table, for instance, refers to the questions in (10). The remaining sentences can be found in Bierner (2001). Although an implementation exists that is capable of performing the translation in Section 4.2, this was done by hand for this evaluation to abstract away from parsing issues.

(10) What are some works by Edgar Allan Poe?

What are some works by Edgar Allan Poe other than the Raven?

What are some works by Edgar Allan Poe?
: |: ANSWER NOT NEAR (| raven)

Unfortunately, at the time of this evaluation, Electric Knowledge had taken down their public portal in favor of providing search for the web pages of specific clients. This means that the large index used to process the queries in Table 2 was no longer available, and I therefore used different indices and different queries for this evaluation. I used indices of pages about American history and literature—each about 11,000 pages. These new, more specialized, indices have the benefit of abstracting this evaluation away from coverage issues (explaining the differences in precision between Table 2 and Table 3).

I created the questions in two ways. For several, I began by asking a question without an alternative phrase. I then added an alternative phrase in order to remove some responses I was not interested in. For example, when I asked **Who are the romantic poets?**, all responses were about female romantic poets. I therefore used the query **Who are the romantic poets not including women?** in the evaluation. Some queries were made without first trying the non-alternative phrase version: **What are some epics besides Beowulf?**, for example. This variation reflects the fact that it is unclear which is more common, excluding clutter from a set of responses or *a priori* excluding cases that the questioner is not interested in. The queries also vary in their syntactic structure, information requested, and alternative phrase used. The purpose of varying the queries in these ways is to ensure that the results do not simply reflect a quirk in the EK search engine’s implementation.

For each query, Table 3 shows *total*, the number of documents returned; *good*, the number of true positives; and *top 5*, the number of true positives in the top five returned documents. A true positive was given to a document if it contained an answer to the question, and half a point was given to a document that contained an obvious link to a document with an answer to the question. Precision is computed for all documents and for the top five. It is important to note that the scores for the queries without alternative phrases are still computed with respect to the alternative phrase. That is, documents only about the “Raven” are considered false positives. In this way, we can view these scores as a baseline—what would have happened had the system simply removed the alternative phrase. This should be taken with a grain

	Baseline			Original Query			Hybrid											
	Tot	Good	Top 5	Tot	Good	Top 5	Tot	Good	Top 5									
Poe	12	6.5	3	10	0.5	0.5	10	5.5	2.5									
Romantics	10	0	0	15	0	0	10	3	3									
Witch Hunts	10	8	3	14	2	1	10	8	5									
US Wars	15	12	2	0	0	0	16	13	4									
Sonnets	15	10	5	10	2	0	10	8	4									
Presidents	15	2	2	15	0	0	15	2	2									
Epics	10	7	4	10	5	3	10	7	4									
Dec of Ind	10	2	0	0	0	0	10	5.5	2									
Avr. Precision	48%			47.5%			14.9%			15%			58%			66.3%		

Table 3: Evaluation of Improvement for NLIR Systems

of salt because, in many cases, I chose the query *because* there were documents to remove (as in the *Romantics* example). However, a concern was that the transformed query would cause numerous false negatives. This is not the case as seen by the fact that the precision of the transformed query is not lower than the baseline. In fact, in no example was the precision less than the baseline, and at worse, the precision remained the same.

Performance on questions containing alternative phrases was quite poor, with an average of 15% precision. This is significantly worse than the transformed query, and even the baseline. The performance drop is due to the fact that the complex syntax of the query confuses the EK search engine’s analysis. The EK search engine is forced to reduce the query to a simple set of keywords including the figure, which we were trying to avoid.

Thus as predicted in the discussion of potential improvement, not accounting for alternative phrases can greatly increase the number of false positives containing the figure (FPF), the very thing the query is attempting to exclude. Table 4 shows that in the baseline case, where there is no alternative phrase, on average for 28% of the returned documents the only answer was the one we wanted to exclude. Adding the alternative phrase has the opposite of the intended effect as the percentage of FPFs increases to 44% for reasons described above. Transforming the query, on the other hand, causes the desired effect, more than halving the percentage of FPFs of the baseline.

6 Conclusions

In this paper, I briefly presented a formal analysis for a large class of words called *alternative markers*. The analysis elegantly captures the desired phenomena by separating assertion and presupposition and making use of alternative phrases and the “use existing objects” heuristic. A simple pattern matching approach would fail to

capture examples requiring knowledge and those with anaphora. Furthermore, sophisticated noun phrase detection would be required for many examples.

I then show that alternative phrases appear frequently enough in dialog to warrant serious attention, yet present natural language search engines perform poorly on queries containing alternative phrases. However, by approximating my semantic analysis into a form understood by a natural language search engine, I showed that the performance of that search engine improved dramatically. As the operational semantics of natural language applications improves, even larger improvements are possible.

Further improvement is also possible by approaching the problem from the other direction. Alternative phrases can be used to help characterize documents beyond the simple word indices used in many current systems. This richer data can be used in the retrieval process to more effectively identify appropriate documents.

References

- David Beaver. 1997. Presupposition. In Johan van Benthem and Alice ter Meulen, editors, *Handbook of Logic and Language*, pages 939–1008. North Holland, Amsterdam.
- Gann Bierner. 2001. *Alternative phrases: theoretical analysis and practical applications*. Ph.D. thesis, Division of Informatics, University of Edinburgh.
- Edwin Cooper. 1997. The On Point system: A natural language search engine for the world wide web. Master’s thesis, University of Chicago.
- Marti Hearst. 1992. Automatic acquisition of hyponyms from large text corpora. In *Proceedings of the Fourteenth International Conference*

	Baseline		Original Query		Hybrid	
	FPF	FPF/Tot	FPF	FPF/Tot	FPF	FPF/Tot
Poe	1	.08	1	.1	1	.1
Romantics	8	.8	15	1	2	.2
Witch Hunts	2	.2	1	.07	0	0
US Wars	3	.2			3	.19
Sonnets	5	.33	8	.8	2	.2
Presidents	5	.33	10	.67	2	.13
Epics	0	0	0	0	0	0
Dec of Ind	3	.3			2	.2
	28.1%		44%		12.8%	

Table 4: False Positives Containing Figure

- on *Computational Linguistics*, Nantes, France, July.
- Jerry R. Hobbs, Mark E. Stickel, Douglas E. Appelt, and P. Martin. 1993. Interpretation as abduction. *Artificial Intelligence*, 63(1-2):69-142, October.
- Jacob Hoeksema. 1995. The semantics of exception phrases. In Jaap van der Does and Jan van Eijck, editors, *Quantifiers, Logic, and Language*, chapter 6, pages 145-177. Cambridge University Press.
- Lauri Karttunen and Stanley Peters. 1979. Conventional implicature. In Choon-Kyu Oh and David Dinneen, editors, *Syntax and Semantics 11: Presupposition*. New York, Academic Press.
- David Lewis and Karen Sparck-Jones. 1996. Natural language processing for information retrieval. In *Communications of the ACM*, volume 39, pages 92-101, January.
- David Lewis. 1979. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8:339-359.
- Mats Rooth. 1985. *Association with Focus*. Ph.D. thesis, University of Massachusetts, Amherst.
- Mats Rooth. 1992. A theory of focus interpretation. *Natural Language Semantics*, 1:75-116.
- C. P. Rose, B. Di Eugenio, and J. D. Moore. 1999. A dialogue based tutoring system for basic electricity and electronics. In *Proceedings of AI in Education*.
- Earl Sacerdoti. 1977. *A Structure for Plans and Behavior*. Elsevier/North-Holland, Amsterdam.
- Robert Stalnaker. 1974. Pragmatic presuppositions. *Semantics and Philosophy*, pages 129-214.
- Mark Steedman. 1996. *Surface Structure and Interpretation*. MIT Press, Cambridge Mass. Linguistic Inquiry Monograph, 30.
- Mark Steedman. 2000. *The Syntactic Process*. The MIT Press, Cambridge Mass.
- Matthew Stone and Christine Doran. 1997. Sentence planning as description using tree adjoining grammar. In *Proceedings of the 35th Annual Meeting of the Association for Computational Linguistics and the 8th Conference of the European Association for Computational Linguistics, Madrid, July*, pages 198-205.
- Matthew Stone and Bonnie Webber. 1998. Textual economy through close coupling of syntax and semantics. In *International Workshop on Natural Language Generation*, pages 178-187, Niagara-on-the-Lake, Canada, August.
- TREC. 2000. <http://trec.nist.gov/>.
- K. VanLehn, S. Siler, C. Murray, and W.B. Baggett. 1998. What makes a tutorial event effective? In M. A. Gernsbacher and S. Derry, editors, *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, pages 1084-1089, Hillsdale, NJ.
- K. VanLehn, S. Siler, C. Murray, T. Yamauchi, and W.B. Baggett. in press. Human tutoring: Why do only some events cause learning? *Cognition and Instruction*.
- Kai von Fintel. 1993. Exceptive constructions. *Natural Language Semantics*, 1(2):123-148.
- Bonnie Webber, Alistair Knott, Matthew Stone, and Aravind Joshi. 1999. Discourse relations: A structural and presuppositional account using lexicalised tag. In *Proceedings of the 37th Conference of the Association for Computational Linguistics*, pages 41-48, College Park, MD.
- John M. Zelle and Raymond J. Mooney. 1993. Learning semantic grammars with constructive inductive logic programming. In *Proceedings of the 11th National Conference on Artificial Intelligence*, pages 817-823, Menlo Park, CA, USA. AAAI Press.