Khoo, C., Kornfilt, J., Oddy, R., & Myaeng, S.H. (1998). Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary & Linguistic Computing*, 13(4), 177-186.

AUTOMATIC EXTRACTION OF CAUSE-EFFECT INFORMATION FROM NEWSPAPER TEXT WITHOUT KNOWLEDGE-BASED INFERENCING

By

Christopher S.G. Khoo
Centre for Advanced Information Systems
School of Applied Science
Nanyang Technological University
Republic of Singapore 639798.

Jaklin Kornfilt
Department of Linguistics
Syracuse University
Syracuse, New York 13244, USA

Sung Hyon Myaeng
Department of Computer Science
College of Natural Science
Chungnam National University
Taejon 305-764, Korea.

Robert N. Oddy
School of Information Studies
Syracuse University
Syracuse, New York 13244, USA

Correspondence should be addressed to:

Christopher Khoo
School of Applied Science
Nanyang Technological University
Blk N4 #2A-36
Nanyang Avenue
Republic of Singapore 639798.
Tel: (65) 7991276
Fax: (65) 7926559
Email: assgkhoo@ntu.edu.sg

Khoo, C., Kornfilt, J., Oddy, R., & Myaeng, S.H. (1998). Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary & Linguistic Computing*, 13(4), 177-186.

AUTOMATIC EXTRACTION OF CAUSE-EFFECT INFORMATION FROM NEWSPAPER TEXT WITHOUT KNOWLEDGE-BASED INFERENCING

Abstract

This study investigated how effectively cause-effect information can be extracted from newspaper text using a simple computational method (i.e. without knowledge-based inferencing and without full parsing of sentences). An automatic method was developed for identifying and extracting cause-effect information in Wall Street Journal text using linguistic clues and pattern-matching. The set of linguistic patterns used for identifying causal relations was based on a thorough review of the literature and on an analysis of sample sentences from Wall Street Journal. The cause-effect information extracted using the method was compared with that identified by two human judges. The program successfully extracted about 68% of the causal relations identified by both judges (the intersection of the two sets of causal relations identified by the judges). Of the instances that the computer program identified as causal relations, about 25% were identified by both judges, and 64% were identified by at least one of the judges. Problems encountered are discussed.

Khoo, C., Kornfilt, J., Oddy, R., & Myaeng, S.H. (1998). Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary & Linguistic Computing*, 13(4), 177-186.

AUTOMATIC EXTRACTION OF CAUSE-EFFECT INFORMATION FROM NEWSPAPER TEXT WITHOUT KNOWLEDGE-BASED INFERENCING[1]

Introduction

This study investigated how effectively cause-effect information can be extracted from newspaper text using a simple computational method. A method was developed to identify and extract cause-effect information in Wall Street Journal text automatically using pattern matching without full parsing of sentences. A set of linguistic patterns that usually indicate the presence of a causal relation was constructed and used for the pattern matching. No inferencing from commonsense knowledge or domain knowledge was used. Only linguistic clues were used to identify causal relations. The goal was to develop a method that was appropriate for use in an information retrieval system which catered to a heterogeneous user population with a wide range of subject interests.

The linguistic patterns were constructed based on an extensive review of the literature. General sources like the *Longman Dictionary of Contemporary English* and the *Roget International Thesaurus* were also consulted. The patterns were refined by repeatedly applying them to sample sentences taken from Wall Street Journal, modifying the patterns to eliminate the errors, and applying the patterns to a new sample of sentences. Though Wall Street Journal text was used in refining the linguistic patterns, it should be relatively easy to adapt the patterns to other kinds of text because the initial set of linguistic patterns was developed based on general sources.

It is important to know how accurately a computer program can extract cause-effect information from text using linguistic patterns alone. Previous studies made extensive use of domain knowledge that had to be coded by hand. Knowledge-based identification of causal relations is currently feasible only for very narrow domains. Some applications, e.g.

information retrieval, require more general approaches that are applicable to a large textual database covering a wide range of subjects. Also, applications such as information retrieval may not require a very high level of accuracy in identifying causal relations, and knowledge-based inferencing may not be necessary for the purpose.

Previous Studies

Studies on the automatic extraction of causal knowledge in text have focused on the use of knowledge-based inferencing.

Some researchers have attempted to identify causal relations expressed or implied in episodic or narrative text – text describing a series of related events involving human actions, e.g. a story (Bozsahin & Findler, 1992; Cullingford, 1978; Lebowitz, 1980; Mooney, 1990; Schank, 1982; Schubert & Hwang, 1989; and Wilensky, 1978 & 1983). These studies sought to find out what kind of knowledge and inferencing are needed to identify causal relations between events described in the text and to infer events that are implied in the text. These studies typically make little use of linguistic clues to identify causal relations. Presumably, explicit linguistic indications of cause and effect, such as *because*, *if ... then*, and *as a result of this*, do not occur often in episodic text.

A second group of studies have focused on identifying causal relations in short explanatory messages of the kind that a human expert on a particular subject might enter into the knowledge acquisition component of an expert system (Selfridge, Daniell & Simmons, 1985; Joskowsicz, Ksiezyk & Grishman, 1989). The approach taken in these studies has been to build a model of the system or domain. When there is ambiguity about whether a causal relation between two events is expressed in the text, the system uses the model of the domain to check whether a causal relation between the events is possible. Selfridge (1989) has reviewed the main issues involved in the automatic acquisition of causal knowledge from human experts.

The third group of studies dealt with expository text -- the kind of text found in textbooks. We found only two such studies dealing with English text: Kontos and Sidiropoulou (1991) and Kaplan and Berry-Rogghe (1991) both dealt with scientific text. They used linguistic patterns to identify causal relations, but all the information required for linguistic processing -- the grammar, the lexicon, and the patterns for identifying causal relations -- were hand-coded and were developed just to handle the sample texts used in the studies. In the study by Kaplan and Berry-Rogghe, the sample texts were parsed by hand. Knowledge-based inferencing was also used. The authors pointed out that substantial domain knowledge, which was hand-coded, was needed for the system to identify causal relations in the sample texts accurately. Scaling up is obviously a problem: the grammar, lexicon and patterns will not be usable in another subject area, and may not even be effective for other documents on the same subject.

More recently, Garcia (1997) developed a computer program to extract cause-effect information from French technical text without using domain knowledge. He focused on causative verbs and reported a precision rate of 85%.

Our study did not make use of knowledge-based inferencing to identify causal relations, but relied entirely on linguistic clues. Knowledge-based inferencing of causal relations require a detailed knowledge of the domain. The studies referred to in this section dealt with very narrow domains, and most of the systems developed were demonstration prototypes working with a very small amount of text. In contrast, our study dealt with a realistic full-text database comprising about five years of *Wall Street Journal* articles. Though the *Wall Street Journal* is business oriented, it covers a very wide range of topics and the articles are non-technical. Since the purpose of this study was to develop a method that could be used by an information retrieval system dealing with a heterogeneous database, it was not possible to manually encode domain knowledge for all the subject areas covered by the database.

Khoo, C., Kornfilt, J., Oddy, R., & Myaeng, S.H. (1998). Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary & Linguistic Computing*, 13(4), 177-186.

Expression of Causal Relations in Text

Many of the causal relations in text are implicit and are inferred by the reader using general knowledge. The focus of this study is not on implicit causal relations but on cause and effect that is explicitly indicated in written English.

From a review of the linguistic literature, we identified the following ways of explicitly expressing cause-effect:

1. using *causal links* to link two phrases, clauses or sentences
2. using *causative verbs*
3. using *resultative* constructions
4. using *conditionals*, i.e. "if ... then ..." constructions
5. using *causative adverbs and adjectives.*

Causal Links

Altenberg (1984) classified causal links into four main types:

a. the adverbial link, e.g. *so, hence, therefore*

b. the prepositional link, e.g. *because of, on account of*

c. subordination, e.g. *because, as, since*

d. the clause-integrated link, e.g. *that's why, the result was*.

He presented a detailed typology of causal links and an extensive list of such linking words compiled from several sources, including Greenbaum (1969), Halliday and Hasan (1976) and Quirk, Greenbaum, Leech and Svartvik (1972).

Causative Verbs

Causative verbs are verbs the meanings of which include a causal element. Examples include the transitive form of *break* and *kill*. The transitive *break* can be paraphrased as *to cause to break*, and the transitive *kill* can be paraphrased as *to cause to die*.

It is important to distinguish causative verbs from other transitive words that are not causative, e.g. *hit*, *kick*, *slap* and *bite* (Thompson, 1987). We adopted the following criterion (adapted from Szeto, 1988) for distinguishing causative verbs from other transitive verbs:

> *a causative verb is a transitive verb that specifies the result of an action, event or state, or the influence of some object*.

Some action verbs specify the action but not the result of the action. Causative verbs include some action verbs like *kill*, as well as some transitive verbs like *amaze* which are not action verbs but nevertheless specify the impact of some object or event.

We also made use of the rule (adapted from Thompson, 1987) that the subject of a causative verb must be separable from the result. This is to exclude words like *mar*, *surround* and *marry*, for which the subject of the verb is an integral part of the effect specified by the verb, as in the following examples from the *Longman Dictionary of Contemporary English* (2nd ed.):

(1) The new power station *mars* the beauty of the countryside.

(2) A high wall *surrounds* the prison amp.

(3) Will you *marry* me?

To obtain a comprehensive list of causative verbs, the above criteria were applied to the first two senses of all verb entries in the *Longman Dictionary of Contemporary English* (2nd ed.). To help us make consistent decisions on whether a verb was causative or not, we classified the results specified by the verbs into 47 result types, listed in Appendix 1. A verb was accepted as causal if it specified one of the listed types of results. Verbs that do not belong to one of the

47 types but are nevertheless clearly causal are listed in a "miscellaneous" category. A total of 2082 verbs were identified as causative verbs. They are listed in Khoo (1995).

Resultative Constructions

A resultative construction is a sentence in which the object of a verb is followed by a phrase describing the state of the object as a result of the action denoted by the verb. The following examples are from Simpson (1983):

(4a) I painted the car *yellow*.

(4b) I painted the car *a pale shade of yellow*.

(4c) I cooked the meat *to a cinder*.

(4d) The boxer knocked John *out*.

In example (4a), the adjective *yellow* is the "resultative phrase" describing the result of the action of painting the car.

In this study, we make use of the syntactic pattern *V-NP-Adj* to identify resultative sentences in which the resultative phrase is an adjective. Simpson (1983) said that this is the most common kind of resultative.

Conditionals

If-then constructions often indicate that the antecedent (the *if* part) causes the consequent (the *then* part). This study uses *if-then* constructions as an indication of a causal relation.

Causative Adverbs and Adjectives

Some adverbs and adjectives have a causal element in their meanings (Cresswell, 1981). One example is the adverb *fatally*:

(5) Brutus *fatally* wounded Caesar.

(6) Catherine *fatally* slipped.

These can be paraphrased as:

(7) In wounding Caesar, Brutus caused Caesar to die.

(8) Catherine slipped, and that caused her to die.

The adjective *fatal* also has a causal meaning:

(9) Caesar's wound was *fatal*.

(10) Guinevere's *fatal* walk ...

This study did not make use of causal adverbs and adjectives because they are not well studied, and a comprehensive list of such adverbs and adjectives has not been identified.

Linguistic Patterns for Identifying Causal Relations

Based on the types of causal expressions described in the previous section, we constructed a set of linguistic patterns that could be used by a computer program to identify causal relations *within a sentence*, as well as *between adjacent sentences*. The patterns are listed in Khoo (1995).

To identify causal relations in a document, a computer program locates all parts of the document that match with any of the linguistic patterns. "Slots" in a linguistic pattern indicate which part of the text is the *cause* and which the *effect*. For example, the pattern

[effect] *is the result of* [cause]

indicates that the part of the sentence following the phrase "is the result of" represents the *cause* and the part of the sentence preceding the phrase represents the *effect*. Each pattern is thus a template for expressing cause and effect, and is equivalent to a finite state transition network.

Each pattern consists of a sequence of tokens separated by a space. Each token indicates one of the following:

- a particular word
- a word having a particular part-of-speech label (e.g. an adjective)
- a particular type of phrase (e.g. noun phrase)
- a set of subpatterns (as defined in a *subpatterns file*)
- any verb from a particular group of verbs (as defined in a *verb groups file*)
- a slot to be filled by one or more words representing the cause or the effect
- any word or phrase (i.e. a wild card symbol).

[Insert Table 1 about here.]

Table 1 gives, as examples, some of the patterns involving the word *because*. *[1]* and *[2]* in the patterns represent slots to be filled by the first and second member of the relation – the first member of the causal relation being the cause and the second member the effect. The type of phrase or word (e.g. noun phrase) that may fill a slot may also be indicated.

The symbol *&* followed by a label refers to a set of subpatterns (usually a set of synonymous words or phrases). For example, *&AUX* in patterns (3) to (6) of Table 1 refers to auxiliary verbs like *will*, *may*, and *may have been*.

For each set of patterns, the patterns are tried in the order listed in the set. Once a pattern is found to match a sentence (or some part of a sentence), all the words that match the pattern (except for the words filling the slots) are flagged, and these flagged words are not permitted to match with tokens in any subsequent pattern. So, the order in which patterns are listed in the set is important. As a rule, a more "specific" pattern is listed before a more "general" pattern. A

pattern is more specific than another if it contains all the tokens in the other pattern as well as additional tokens not in the other pattern.

Consider the following three patterns:

1. [1] and because of this , [2]
2. because [1] , [2]
3. [2] because [1]

Pattern 1 is more specific than patterns 2 and 3, and pattern 2 is more specific than pattern 3. All the sentences that pattern 1 will match, patterns 2 and 3 will match also. For example, all three patterns will match the following sentence:

(11) It was raining heavily and because of this, the car failed to brake in time.

Note that a pattern does not need to match the whole sentence for a match to occur. A pattern needs to match just some part of the sentence for a causal relation to be identified. So, pattern 2 does not require the word *because* to appear at the beginning of the sentence.

Only pattern 3 will match the sentence:

(12) The car failed to brake in time because it was raining heavily.

Pattern 1 will not match sentence (12) because the sentence does not contain the phrase *and because of this*. Pattern 2 will not match the sentence because pattern 2 requires that there be a comma after the word *because*. So, pattern 3 is more general than patterns 1 and 2 in the sense that pattern 3 contains fewer constraints.

Although all three patterns will match sentence (11), only pattern 1 will correctly identify the cause and the effect in the sentence. Applying pattern 1 to sentence (11), we obtain:

cause: it was raining heavily

effect: the car failed to brake in time

Applying, pattern 2 to sentence (11), we obtain:

cause: of this

effect: the car failed to brake in time

which, although not wrong, is not as informative as the result of applying pattern 1. On the hand, applying pattern 3 to sentence (11) yields the incorrect result:

cause: of this, the car failed to brake in time.

effect: it was raining heavily and

Because pattern 1 is listed before patterns 2 and 3, pattern 1 will be applied to the sentence first and the words *and because of this* are flagged in the sentence so that they are not permitted to match with any of the non-slot tokens in patterns 2 and 3.[2] In particular, the word *because* is flagged and is not permitted to match with the token *because* in patterns 2 and 3.

The linguistic patterns constructed in this study assume that the text has been pre-processed in the following ways:

- the beginning and end of sentences have been identified, and each sentence is placed on a separate line.
- words in the text have been tagged with part-of-speech labels
- the boundaries of phrases (e.g. noun phrases) have been marked with brackets.

Sentence and phrase boundary identification was done using text processing programs developed in the DR-LINK project (Liddy & Myaeng, 1993; Liddy & Myaeng, 1994). Part-of-speech tagging was performed using the POST tagger (obtained from BBN Systems and Technologies) which uses 36 part-of-speech tags (Meteer, Schwartz, & Weischedel, 1991).

Evaluation

The evaluation was based on a random sample of 509 pairs of adjacent sentences and 64 single sentences (1082 sentences in all) taken from about four months of Wall Street Journal articles. The effectiveness of the computer program in identifying and extracting cause-effect information from Wall Street Journal using the patterns was evaluated by comparing the output

of the computer program against the judgments of two human judges (identified as Judge A and B), who were asked to identify causal relations in the sample sentences. The judges were "trained" using a training set of 200 pairs of sentences randomly selected from Wall Street Journal.

The evaluation is divided into two parts. Part 1 of the evaluation focuses on whether the computer program can identify the presence of a causal relation and the direction of the causal relation. Part 2 evaluates how well the computer program can identify the "scope" of the causal relation, i.e. can correctly extract all the words in the text that represent the cause and all the words that represent the effect. Since a *cause* and *effect* can comprise more than one word, there will be instances where the computer program extracts more words or fewer words than is appropriate.

Evaluation Part 1: Identifying the Presence of a Causal Relation

[Insert Table 2 about here.]

The performance measures used are *recall* and *precision*. *Recall*, in this context, is the proportion of the causal relations identified by the human judges that are also identified by the computer program. *Precision* is the proportion of causal relations identified by the computer program that are also identified by the human judges. *Recall* measures how comprehensive the identification of causal relations is, whereas *precision* measures what proportion of the causal relations identified by the computer program is in fact correct. The results are given in Table 2. We highlight the more important results.

Judge A identified many more causal relations than judge B (615 for judge A and 174 for judge B). Why such a big difference between the two judges? One possible reason was that the judges had a different understanding of causal relations or a different understanding of the instructions. However, a closer look at the results showed that most of the causal relations

picked out by judge B (91%) were also identified by judge A. The causal relations identified by judge B were largely a subset of the relations identified by judge A. This suggests a high degree of consistency between the two judgments. It is just that judge A picked out a lot more causal relations. We feel that this is a nice result – though not everyone will agree. Judge B's list of causal relations probably represents the more obvious causal relations. Judge A spent much more time on the task than judge B (about three or four times more) and went over the sample sentences a few times. So, judge A's judgments were more thorough and probably more liberal than B's.

Judge A and judge B had 161 causal relations in common. We shall refer to the causal relations identified by both A and B as the intersection set, and the causal relations identified by either A or B as the union set. In calculating the *recall* and *precision*, we compared the judgments made by the computer program with the intersection set, which was made up of causal relations identified by both human judges. There is some amount of subjectivity involved in identifying causal relations in text -- especially in deciding whether the causal relation is explicitly expressed or merely implied. Taking the intersection set of two judgments eliminates idiosyncratic judgments by either judge, and ensures that the causal relations used to evaluate the effectiveness of the computer program are those that are clearly expressed in the text. The intersection set probably also represents the more obvious causal relations.

Of the causal relations in the intersection set, 109 were picked up by the computer program, giving a recall of 68% (109/161), with a 95% confidence interval of ±7%. Of the causal relations in the intersection set, 63 involved causal links and 98 involved causative verbs. For causal links the recall was 78% (49/63), whereas for causative verbs the recall was 61% (60/98).

Of the 437 causal relations identified by the program, only 25% (109/437) (precision) were picked out by both judges (i.e. were in the intersection set). The 95% confidence interval

was ±4%. For causal links the precision was 42% (49/117), whereas for causative verbs the precision was 19% (60/320). Clearly, it is much more difficult to identify causal relations expressed using causative verbs than using causal links.

Of the 75% of the instances identified by the program as causal relations but were not in the judges' intersection set, not all of them were clearly wrong. If we take a more liberal approach and consider as correct the instances when *either* judge picked out the causal relations (i.e. use the union set for calculating precision), then the precision was 64% (280/437) with a 95% confidence interval of ±5%. For causal links the precision calculated in this way was 74% (86/117), whereas for causative verbs the precision was 61% (194/320).

Evaluation Part 2: Determining the Scope of the Causal Relation

[Insert Table 3 about here.]

This section evaluates how accurately the computer program can determine what part of the text is the cause and what part is the effect. For this evaluation, we examined each causal relation that was identified by the computer program as well as by either of the human judges. (In other words, this evaluation is done using only those instances where the computer program correctly identified the *presence* of a causal relation.) We compared the words that were extracted by the computer program as the *cause* with the words that were identified by a human judge as the *cause*, and calculated the measures of recall and precision -- *recall* being the proportion of words extracted by the human judge that were also extracted by the computer program, and *precision* being the proportion of words extracted by the computer program that were also extracted by the human judge. The recall and precision measures were also calculated for the *effect* part of the relation. The recall and precision figures were then averaged across all the causal relations. The results are given in Table 3.

For the *cause* part of the relation, the average recall was 98% for causal links and 93% for causative verbs. The average precision was 96% for causal links and 94% for causative verbs. For the *effect* part, the average recall was 96% for causal links and 86% for causative verbs. The average precision was 91% for causal links and 98% for causative verbs.

Sources of Error

The errors made by the computer program in identifying causal relations were examined to see why the errors occurred. For each of these, we discuss both errors of commission (instances where the computer program indicated there was a causal relation when in fact there wasn't) and errors of omission (causal relations that the computer program failed to identify).

Errors Involving Causal links

[Insert Table 4 about here.]

The reasons for the errors involving causal links are summarized in Table 4. Most of the errors of commission were due to lexical ambiguity – the same words and sentence constructions that are used to indicate cause-effect can be used to indicate other kinds of relations as well.

The sentence pattern that gave rise to the highest number of errors of commission was the pattern

[effect] by [present participle phrase: cause]

which accounted for 7 errors. This pattern was constructed to identify causal relations in sentences like:

(13) [$_{effect}$ Japan has become a major economic power] mainly by [$_{cause}$ exporting to the U.S.]

However, this sentence construction can also be used to indicate the manner in which something is done, as in the following examples:

(14) Secretary Baker has done a service just *by* mentioning the word in public.

(15) Senator Proxmire challenged the nominee *by* disputing economic forecasts he had made during the Ford administration.

In sentence (14), "mentioning the word in public" was how Secretary Baker did a service, not why he did it. Similarly, in sentence (15), "disputing economic forecasts ..." was the manner Senator Proxmire challenged the nominee, rather than the reason he challenged the nominee.

The conjunction "as" accounted for 4 of the errors of commission, and "if .. then" constructions accounted for 3 errors.

Most of the errors of omission were due to particular kinds of linking words or sentence constructions not included in our list of patterns. Many of these linking words and sentence constructions are seldom used for indicating cause and effect. Below are 2 examples of sentences that contain causal relations not picked up by the computer program:

(16) [$_{effect}$ Crop conditions improved considerably in several states] *with* [$_{cause}$ widespread rains in June.]

(17) It's such a volatile stock -- [$_{cause}$ the slightest thing goes wrong] *and* [$_{effect}$ the stock takes a nosedive.]

For the above sentences, inferencing from general knowledge is needed to identify the causal relations.

Errors Involving Causative Verbs

[Insert Table 5 about here.]

The reasons for the errors in identifying causal relations involving causative verbs are summarized in Table 5. Some of the reasons listed in the table require an explanation.

Reason C3 refers to sentences such as the following:

(18) Forest products segment sales increased 11.6% to $157.6 million.

The noun phrase following the verb is not assigned a "patient" role by the verb, i.e. the noun phrase "11.6%" does not refer to the object affected by the verb. Rather, it indicates the magnitude of the process denoted by the verb. It was the subject of the verb, "forest products segment sales", that increased.

Reason C5 refers to instances where the *cause* was not specified in the sentence but the computer program nevertheless extracted one part of the sentence as the cause. In some cases, the computer program was confused by the complex sentence structure. These errors can be avoided if an accurate parser is used. For some of the sentences, it is difficult to tell from the sentence structure alone whether the cause is specified or not. The following pairs of sentences illustrate this difficulty. The sentences labeled (a) do not specify the *cause*, whereas the sentences labeled (b) having the same syntactic structure do specify the *cause*:

(19a) Friends have suggested pouring [$_{\text{effect}}$ vermouth into the soap dispenser.]

(19b) [$_{\text{cause}}$ Friends] have admitted pouring [$_{\text{effect}}$ vermouth into the soap dispenser.]

(20a) Measures are being taken to make [$_{\text{effect}}$ the loan more attractive.]

(20b) [$_{\text{cause}}$ Low interest rates] are being offered to make [$_{\text{effect}}$ the loan more attractive.]

The most common reason for the errors of commission was word sense ambiguity. A word can have several senses, some senses having a causal meaning and others not. No lexical disambiguation was attempted in this study.

We now discuss the errors of omission. Reasons D1 to D3 (Table 5) together accounted for the highest number of errors. These three types of errors can be reduced by using an accurate parser.

Reason D4 refers to sentences like the following:

(21) The flaps on each wing help *provide lift* for a jetliner to get off the ground.

In this sentence, the causative verb *lift* is used in a noun form. The sentence may be paraphrased as

(22) The flaps on each wing help *lift* a jetliner off the ground.

Nominalized verbs, i.e. verbs used in noun form, were not handled in this study.

Conclusion

This study investigated how effectively cause-effect information can be extracted from Wall Street Journal (a newspaper) using simple pattern matching without knowledge-based inferencing and without full parsing of sentences. The results indicate that for Wall Street Journal text, about 68% of the causal relations that are clearly expressed within a sentence or between adjacent sentences can be correctly identified and extracted using the linguistic patterns developed in this study. Of the instances that the computer program identified as causal relations, about 25% were identified by both judges, and 64% were identified by at least one of the judges.

Most of the errors made by the computer program are due to

- complex sentence structure
- lexical ambiguity
- absence of inferencing from world knowledge.

This study makes use of a phrase bracketer for identifying phrase boundaries (e.g. noun phrases), but not a full parser. If an accurate parser is used, the maximum recall that can be attained is around 83% (assuming no error due to sentence structure), and the maximum precision attainable is about 82%. Much of the complexity of the linguistic patterns constructed in this study is due to the need to handle different sentence structures. If a parser is used, the linguistic patterns can be made much simpler, and fewer patterns need be used.

Accurate word sense disambiguation, especially for verbs, can also substantially reduce errors. Inferencing from world knowledge will also help, but it is possible to implement this only for very narrow domains.

How well will the approach used in this study work for other corpora? It depends, of course, on the corpus. Using linguistic patterns for identifying causal relations will be effective to the extent that the corpus satisfies the following conditions:

1. Most of the causal relations in the text are explicitly expressed using linguistic means. The reader is seldom required to infer cause-effect from general knowledge or domain knowledge.
2. Most of the sentences are simple and straightforward.
3. The subject content of the corpus is limited to a narrow subject area so that word sense ambiguity is not a problem.

We surmise that the approach will work well with databases containing abstracts of journal articles in a particular subject area -- particularly abstracts reporting results of empirical research. Causal relations will probably be explicitly stated in such abstracts. We expect the approach to fare poorly with episodic text -- text describing a series of related events (e.g. a story). For this kind of text, causal relations between events usually have to be inferred by the reader using extensive knowledge about the types of events described in the text (Cullingford, 1978; Schank, 1982; Schank & Abelson, 1977; Wilensky, 1978).

The automatic method for identifying causal relations was used in an experimental document retrieval system to identify and match causal relations expressed in documents with causal relations expressed in users' queries. Causal relation matching was found to yield a small but significant improvement in retrieval results when the weights used in combining different sources of evidence were customized for each query (Khoo, 1995).

Future work will explore other possible uses of the automatic method for extracting causal information from text.

Acknowledgements

Our grateful thanks to *Longman Group UK Ltd* for use of the machine-readable file of the *Longman Dictionary of Contemporary English* (new ed.), *Dow Jones & Co.* for use of the *Wall Street Journal* text, the *U.S. National Institute of Standards and Technology* for use of the Tipster/TREC information retrieval test collection, *BBN Systems and Technologies* for use of the POST part-of-speech tagger.

Khoo, C., Kornfilt, J., Oddy, R., & Myaeng, S.H. (1998). Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary & Linguistic Computing*, 13(4), 177-186.

Notes

1. This study was part of the PhD dissertation research of the first author, and was funded in part by a Syracuse University Fellowship.

2. Punctuation marks are not flagged when they match with a token in a pattern. This is because a punctuation mark does not have a meaning the way a word has. Punctuation marks only help to indicate the syntactic structure of the sentence. In the linguistic patterns constructed in this study, punctuation marks are used not so much to identify causal relations as to identify where the cause or effect phrase begins or ends in the sentence. It is necessary to use punctuation marks in the patterns only because sentences are not parsed in this study.

References

Altenberg, B. (1984). Causal Linking in Spoken and Written English, *Studia Linguistica*, 38(1): 20-69.

Bozsahin, H. C. and Findler, N. V. (1992). Memory-Based Hypothesis Formation: Heuristic Learning of Commonsense Causal Relations from Text, *Cognitive Science*, 16(4): 431-54.

Cresswell, M. J. (1981). Adverbs of Causation. In H.-J. Eikmeyer and H. Rieser (eds.), *Words, Worlds, and Contexts: New Approaches in Word Semantics*, Berlin, pp. 21-37.

Cullingford, R. E. (1978). *Script Applications: Computer Understanding of Newspaper Stories*. Technical Report No. 116, Yale University, Department of Computer Science, New Haven.

Garcia, D. (1997). COATIS, an NLP System to Locate Expressions of Actions Connected by Causality Links, *Knowledge Acquisition, Modeling and Management, 10th European Workshop, EKAW '97 Proceedings*, Sant Feliu de Guixols, Catalonia, Spain, October 1997..

Greenbaum, S. (1969). *Studies in English Adverbial Usage.* Longman, London.

Halliday, M. A. K. and Hasan, R. (1976). *Cohesion in English.* Longman, London.

Joskowsicz, L., Ksiezyk, T. and Grishman, R. (1989). Deep Domain Models for Discourse Analysis, *The Annual AI Systems in Government Conference*, Washington, D.C., March, 1989.

Kaplan, R. M. and Berry-Rogghe, G. (1991). Knowledge-Based Acquisition of Causal Relationships in Text, *Knowledge Acquisition*, 3(3): 317-37.

Khoo, C. S. G. (1995). Automatic Identification of Causal Relations in Text and Their Use for Improving Precision in Information Retrieval. Ph.D. thesis, Syracuse University.

Kontos, J. and Sidiropoulou, M. (1991). On the Acquisition of Causal Knowledge from Scientific Texts with Attribute Grammars, *Expert Systems for Information Management*, 4(1): 31-48.

Lebowitz, M. (1980). *Generalization and Memory in an Integrated Understanding System*. Technical Report No. 186, Yale University, Department of Computer Science, New Haven.

Liddy, E. D. and Myaeng, S. H. (1993). DR-LINK's Linguistic-Conceptual Approach to Document Detection, *The First Text REtrieval Conference (TREC-1)*, Gaithersburg, Maryland, November 1992.

Liddy, E. D., & Myaeng, S. H. (1994). DR-LINK: A System Update for TREC-2, *The Second Text REtrieval Conference (TREC-2)*, Gaithersburg, Maryland, August-September 1993.

Longman Dictionary of Contemporary English. (1987). 2nd ed. Longman, Harlow, Essex.

Meteer, M., Schwartz, R. and Weischedel, R. (1991). POST: Using Probabilities in Language Processing, *IJCAI-91: Proceedings of the Twelfth International Conference on Artificial Intelligence*.

Mooney, R. J. (1990). Learning Plan Schemata from Observation: Explanation-Based Learning for Plan Recognition, *Cognitive Science*, 14(4): 483-509.

Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1972). *A Grammar of Contemporary English*. Longman, London.

Schank, R. C. (1982). *Dynamic Memory*. Cambridge University Press, New York.

Schank, R. C. and Abelson, R. P. (1977). *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Erlbaum, Hillsdale, New Jersey.

Schubert, L. and Hwang, C. H. (1989). An Episodic Knowledge Representation for Narrative Texts, *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, Toronto, 1989.

Selfridge, M. (1989). Toward a Natural Language-Based Causal Model Acquisition System, *Applied Artificial Intelligence*, 3(2-3): 107-128.

Selfridge, M., Daniell, J. and Simmons, D. (1985). Learning Causal Models by Understanding Real-World Natural Language Explanations, *The Second Conference on Artificial Intelligence Applications: The Engineering of Knowledge-Based Systems*, Miami Beach, Florida, December, 1985.

Simpson, Jane. (1983). Resultatives. In L. Levin, M. Rappaport, and A. Zaenen (eds.), *Papers in Lexical-Functional Grammar*, Bloomington, Indiana, pp. 143-57.

Szeto, Y.-K. (1988). The Semantics of Causative and Agentive verbs, *Cahiers Linguistiques d'Ottawa*, 16: 1-51.

Thompson, J. J. (1987). Verbs of Action, *Synthese*, 72(1): 103-22.

Wilensky, R. W. (1978). *Understanding Goal-Based Stories*. Technical Report No. 140, Yale University, Department of Computer Science, New Haven.

Wilensky, R. W. (1983). *Planning and Understanding: A Computational Approach to Human Reasoning*. Addison-Wesley, Reading, Mass.

Khoo, C., Kornfilt, J., Oddy, R., & Myaeng, S.H. (1998). Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary & Linguistic Computing*, 13(4), 177-186.

Appendix 1. Classification of Causative Verbs by Type of Result

Major Classes

A. Verbs that mean *to cause something*

B. Verbs that mean *to be caused by something*

C. Verbs that mean *to prevent something from happening*

D. Verbs that mean *to affect something* without specifying in what way

A. Verbs that mean *to cause something*

1a. Verbs that are primarily causal in meaning, and where the subject of the verb can be an event (e.g. *cause, lead to, precipitate, result in, trigger*)

1b. Verbs that are primarily causal in meaning, and where the subject of the verb cannot be an event but has to be a state, an object or an agent (e.g. *engineer, foment, get (+adj), render (+adj), wreak*)

2. Verbs that mean to force (someone) to (do something) (e.g. *coerce (+to-v;+into), compel (+to-v;+into), force (+out of;+from; +to-v;+into), evict, muzzle*)

3. To persuade or cause (someone) to (do something) (e.g. *con (+into),entice (+to-v;+prep;particle), goad (+into;+to-v) inspire (+to-v), persuade (+to-v;+into)*)

4a. To let or allow (someone) to (do something) (*e.g. admit, allow (+prep;particle;+to-v), enable (+to-v), let (+v; +into; down; in; off; out), permit (+to-v)*)

4b. To let or allow (an event or state) to happen or to continue to happen, or to make (an event or state) possible (e.g. *allow, enable, permit, tolerate*)

5a. To cause (an event) to start (e.g. *commence, ignite, initiate, set (+v-ing), start*)

5b. To bring (something) into existence, or to produce (something) (e.g. *build, create, establish, form, produce*)

6. To cause (an event or state) to continue, or to maintain or preserve (something) (e.*g. continue, keep (+v-ing), maintain, perpetuate, preserve*)

7. To cause (something) to operate or to become more active, or to cause (something) to come back into use or existence (e.*g. activate, arouse, reactivate, revive, wake*)

8a. To put (something) out of existence, or to destroy (something) (e.*g. annihilate, assassinate, dismantle, extinguish, kill*)

8b. To cause (an event or state) to come to an end, or to stop (an event or state) that has been happening, or to cause (something) to fail (e.g. *cancel, disable, discontinue, eliminate, end, eradicate*)

8c. To cause (something) to have no effect (e.g. *deactivate, decommission, invalidate, neutralize, nullify*)

9. To cause (something) to be performed or to succeed (e.g. *bring (off), complete, effectuate, implement, push (through)*)

10. To cause (something) to be removed, or to cause (something) to have something removed from it (e.g. *comb (out), debone, dehumidify, delete, detoxify, remove*)

11. To cause (something) to make a sound (e.g. *chime, clatter, hoot, ring, rustle*)

12a. To cause (something) to have a physical feature (e.g. *blister, breach, dent, equip, glaze, retread, upholster*)

12b. To cause (something) to contain something (e.g. *include (+in), inject (+with), poison, salt, sweeten*)

12c. To cause (something) to be covered with something, or to cause (something) to cover something (e.g. *bandage, cover, plaster (+on;+over;+with), shower (+on), shower (+with), sprinkle (+prep), sprinkle (+with)*)

12d. To cause (something) to be filled with something, or to cause (something) to fill something (e.g. *brick (up), pack (+with), refill, replenish, saturate (+with)*)

12e. To cause (something) to be decorated with (something) (e.g. *adorn, decorate (+with), emboss (+on;+with), imprint (+on), tattoo*)

12f. To cause (something) to have certain color(s) (e.g. *blacken, blanch, bleach, color (+adj), paint (+adj)*)

13. To cause (someone) to possess (something), or to cause (something) to be in the possession of (someone) (e.g. *arm, bequeath (+n), bequeath (+to), confer (+on;+upon), empower, give (+to;+n), give (back;+back to)*)

14a. To cause (someone) to have a certain feeling or to be in a certain state of mind (e.g. *agitate, alarm, amaze, bias, convince, evoke, inspire (+in;+to;+with), interest (+in)*)

14b. To cause (someone) to change his/her mind or to have certain beliefs (e.g. *brainwash, disabuse, enlighten, inculcate (+in;+into;+with), lead (on), persuade*)

14c. To cause the body to be in a certain physical state or to experience something (e.g. *anesthetize, deafen, enervate, overcome, starve*)

15. To cause (something) to increase in amount, speed, etc (e.g. *accelerate, boost, fatten, increase, multiply*)

16. To cause (something) to decrease in amount, speed, etc (e.g. *alleviate, cut (back), decrease, ease, narrow, weaken*)

17. To cause (something) to improve or to be in a better state (e.g. *civilize, elevate, improve, optimize, reform*)

18. To cause (something) to worsen or to be in a worse state (e.g. *demote, pollute, soil, tarnish, worsen*)

19. To cause (something) to be restricted in some way (e.g. *imprison, inhibit, limit, localize, restrict*)

20. To cause (someone) to be injured (e.g. *beat (up), cripple, harm, hurt, injure*)

21. To cause (something) to become closed or blocked (e.g. *block, close, obstruct, sew (up)*)

22a. To cause (someone or something) to move (e.g. *bounce, deflect, fly, hurl, move, rotate, transfer*)

22b. To cause (someone or something) to fall or move down (e.g. *cut (down), drop, lower, topple, trip, unhorse*)

22c. To cause (something) to come out (e.g. *gouge (out), pluck, shed, spew (+prep;particle), spray*)

22d. To cause (something) to rise or move up (e.g. *hoist, jack (up), levitate, lift, raise*)

22e. To cause (someone or something) to be located at a certain place (e.g. *berth, bottle, center, inject (+into), place (+prep;particle)*)

22f. To cause (something) to hang from some place (e.g. *dangle, drape, hang, suspend (+from;particle)*)

23. To cause (someone or something) to be become the thing specified (e.g. *enslave, install (+as), knight, make (+n), martyr*)

24a. To cause (something) to be joined or connected (e.g. *connect, fuse, interlace, join, network*)

24b. To cause (something) to be fastened (e.g. *button, fasten, glue, harness, staple*)

24c. To cause (something) to be twisted together (e.g. *braid, entangle, twist (together), weave (+prep;particle)*)

25a. To cause (something) to be unfastened (e.g. *disconnect, loose, unbutton, unfasten, unlock*)

25b. To cause (something) to open or to be opened (e.g. *open (up), unfurl, unstop*)

26a. To cause (something) to separate or break into smaller pieces (e.g. *break, disperse, dissipate, dissolve, separate, smash, snap*)

26b. To cause (something) to be physically damaged (e.g. *break, burst, damage, fracture, vandalize*)

27. To cause (someone) to be set free from something (e.g. *disentangle, free, ransom, rescue*)

28. To cause (something) to be safe, or to protect (something) (e.g. *immunize, protect, save, secure, shelter*)

29. To cause (an event) to be delayed (e.g. *defer, delay, hold (over), postpone*)

30. To cause (someone) to lose something (e.g. *deprive (+of), dispossess, lose (+n), relieve (+of)*)

31. To cause (people) to gather, unite or form a group (e.g. *assemble, cluster, convene, merge, unite*)

32. To cause (someone) to wear something (e.g. *dress (+prep), garland, saddle*)

33. To cause (something) to be put right or to be back in good working order (e.g. *correct, mend, rehabilitate, repair, restore*)

34. To cause (someone or some animal) to be castrated (e.g. *alter, castrate, geld, spay*)

35. To cause (something) to be legal (e.g. *legalize, legitimize, ratify, validate*)

36. To cause (something) to change physically or chemically (e.g. *atomize, ionize, thaw, transmute*)

37. To cause (something) to change in some unspecified way (e.g. *adjust, alter, change, modify, transform*)

38. To cause (something) to be aligned or arranged in a particular way (e.g. *aim (+at), align (+with), arrange, jumble, transpose*)

39. To cause (something) to have a different shape (e.g. *bend, coil, curl, fold, straighten*)

40. To cause (something) to be revealed or uncovered (e.g. *conjure, reveal, unearth, unveil*)

41. To cause (something) to be concealed or hidden from view (e.g. *blot (out), bury, conceal, screen*)

42. Miscellaneous causal verbs where the effect can be described with an adjective (e.g. *bankrupt, beautify, empty, impoverish, perfect*)

43. Other miscellaneous causal verbs (e.g. *balance, calibrate, computerize, detonate, disguise, endanger, expedite, flavor, float, rename*)

B. Verbs that mean *to be caused by something* B. Verbs that mean *to be caused by something*

Examples: *proceed from, result from, stem from*

C. Verbs that mean *to prevent something*

1. Verbs that mean to prevent (an event), or to prevent (something) from coming into existence (e.g. *avert, cancel, forestall, prevent, ward (off)*). This is to be distinguished from *to stop something happening*. *To stop something* means to cause an ongoing event to come to an end. *To prevent something* means to cause something that would otherwise happen to not happen.

2. To prevent or stop (someone) from doing (something) (e.g. *detain, foil, prevent (+v-ing), silence, stop (+from; +v-ing)*)

3. To persuade (someone) not to (do something) (e.g. *con (+out of), dissuade, persuade (+out of), reason (+out of), talk (+out of)*)

D. Verbs that mean *to affect (something)* without specifying in what way

Examples: *act (+on/upon), affect, condition, impact, impinge (+on)*

Khoo, C., Kornfilt, J., Oddy, R., & Myaeng, S.H. (1998). Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary & Linguistic Computing*, 13(4), 177-186.

Khoo, C., Kornfilt, J., Oddy, R., & Myaeng, S.H. (1998). Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary & Linguistic Computing*, 13(4), 177-186.

TABLES

Khoo, C., Kornfilt, J., Oddy, R., & Myaeng, S.H. (1998). Automatic extraction of cause-effect information from newspaper text without knowledge-based inferencing. *Literary & Linguistic Computing*, 13(4), 177-186.

Table1 Examples of linguistic patterns for identifying the cause-effect relation

NO.	RELATION	PATTERN
(1)	C	[1] &AND because of &THIS[1],[2] &._ Example: It was raining heavily and because of this the car failed to brake in time.
(2)	-	&NOT because Example: It was not because of the heavy rain that the car failed to brake in time.
(3)	C	it &AUX &ADV_ because of [1] that [2] &._ Example: It was because of the heavy rain that the car failed to brake in time.
(4)	C	it &AUX &ADV_ because [1] that [2] &._ Example: It was because the rain was so heavy that the car failed to brake in time.
(5)	C	&C2_ &[2](AND_THIS) &AUX &ADV_ because of [1] &._ Example: The car failed to brake in time and this was because of the heavy rain.
(6)	C	&C2_ &[2](AND_THIS) &AUX &ADV_ because [1] &._ Example: The car failed to brake in time and this was because it was raining heavily.
(7)	C	&C because of &[N:1],[2] Example: John said that because of the heavy rain , the car failed to brake in time.
(8)	C	&C2_ [2] because of [1] &._ Example: The car failed to brake in time because of the heavy rain.
(9)	C	&C because &[C:1],[2] Example: Because it was raining so heavily , the car failed to brake in time.
(10)	C	&C2_ [2] because [1] &._ Example: The car failed to brake in time because it was raining so heavily.

Notes

"C" in the second column indicates that the pattern can be used to identify a cause-effect relation. The symbol "-" in the second column indicates a null relation, i.e. the pattern does not identify the presence of any relation.

[1] and *[2]* in the patterns represent slots to be filled by the first and second member of the relation respectively, the first member of the causal relation being the cause and the second member the effect. The type of phrase or word that may fill a slot may also be indicated. The symbol *[N:1]* indicates that the slot for *cause* is to be filled by a noun phrase, whereas *[n:1]*

indicates that the slot is to be filled by a noun. *[C:1]* indicates that the slot is to be filled by a clause.

The symbol *&* followed by a label in uppercase refers to a set of subpatterns (usually a set of synonymous words or phrases). For example, *&AUX* in patterns (3) to (6) refers to auxiliary verbs like *will*, *may*, and *may have been*. *&C* and *&C2_* in patterns (5) to (10) refer to subpatterns that indicate the beginning of a clause. *&._* refers to a set of subpatterns that indicate the end of a clause or sentence, and this of course includes the period.
&[2](AND_THIS) in patterns (5) and (6) refers to the following set of three subpatterns:

[2] &AND &THIS/IT
[2] &AND &THIS [1]
[2]

The first two subpatterns above contain the tokens *&AND*, *&THIS/IT* and *&THIS*, each referring to a set of subpatterns. The example illustrates that a subpattern can contain tokens that refer to a set of subpatterns.

Table 2 Number of causal relations identified by the computer program and the human judges

Causal relations identified by 2 human judges

Number of causal relations identified by judge A: 615

Number of causal relations identified by judge B: 174

Number of causal relations identified by both A and B (intersection of judgments A and B): 161

Number of causal relations identified by either A or B (union of judgments A and B): 628

Causal relations identified by computer program

Total number of causal relations identified by computer program: 437

Number involving a causal link: 117
Number involving a causative verb: 320

Comparison between human judgments and judgments by computer program

Number of causal relations identified by both computer program and judge A: 279

Number involving causal links: 86
Number involving causative verbs: 193

Number of causal relations identified by both computer program and judge B: 110

Number involving causal links: 49
Number involving causative verbs: 61

Number of causal relations identified by computer program and both human judges: 109

Number involving causal links: 49
Number involving causative verbs: 60

Table 3 Evaluation of how accurately the computer program can identify the scope of the cause and the effect

Identifying the scope of the cause

For causal links (averaged over 86 causal relations):

Average recall = 0.98
Average precision = 0.96

For causative verbs (averaged over 194 causal relations):

Average recall = 0.93
Average precision = 0.94

Identifying the scope of the effect

For causal links (averaged over 86 causal relations)

Average recall = 0.96
Average precision = 0.91

For causative verbs (averaged over 194 causal relations)

Average recall = 0.86
Average precision = 0.98

Table 4 Analysis of errors made by computer program in identifying causal relations involving causal links

A. Errors of commission

No. of instances identified by the computer program to be a causal relation involving a causal link, but not identified by either of the human judges: 31

Reasons why errors occurred:

A1. No. of these instances that, in my opinion, can be considered to be correct: 6
A2. Unexpected sentence structure resulting in the wrong part of the sentence extracted as the cause or the effect: 1
A3. Unexpected sentence structure resulting in the program identifying a causal relation where there is none: 2
A4. Linking words not used in a causal sense: 22

B. Errors of omission

No. of causal relations *not* identified by the program: 14

Reasons why errors occurred:

B1. Unexpected sentence structure resulting in the causal relation not picked up by the system: 2
B2. Unexpected sentence structure resulting in the wrong part of the sentence extracted as the cause or the effect: 1
B3. Causal link is not in the list of patterns: 11

Table 5 Analysis of errors make by computer program in identifying causal relations involving causative verbs

C. Errors of commission

No. of instances identified by the computer program to be a causal relation involving a causative verb, but not identified by either of the human judges: 126

Reasons why errors occurred:

C1. No. of instances that can be considered to be correct: 27
C2. Error in part-of-speech tagging (a word is incorrectly tagged as verb): 4
C3. The noun phrase occupying the object position is not the "patient" of the verb: 8
C4. Unexpected sentence structure resulting in the wrong part of the sentence extracted as the cause or the effect: 15
C5. Unexpected sentence structure where the cause is not specified in the sentence: 10
C6. The sentence having the syntactic pattern *V-NP-Adj* is not a resultative sentence: 4
C7. The verb is not used in its causal sense: 58

D. Errors of omission

No. of causative verbs identified by both judges but not identified by program: 38

Reasons why errors occurred:

D1. Error in part-of-speech tagging: 3
D2. Error in phrase bracketing: 5
D3. Unexpected sentence structure resulting in the causal relation not picked up by the program: 13
D4. Causative verb is used in nominalized form: 2
D5. Resultative construction not handled: 1
D6. Verb is not in my list of causative verbs: 14
    6 of the verbs involve an unusual sense of the verb.
    8 of the verbs can, arguably, be included in the list of causative verbs. (The 8 verbs are: benefit, bolster, design, drive down, require, credit (somebody) for, highlight, and ban (somebody) from.)