

# Acquiring Causal Knowledge from Text Using the Connective Marker *tame*

TAKASHI INUI

Research Fellow of the Japan Society for the Promotion of Science  
and

KENTARO INUI and YUJI MATSUMOTO

School of Information Science, Nara Institute of Science and Technology

---

In this paper, we deal with automatic knowledge acquisition from text, specifically the acquisition of *causal relations*. A causal relation is the relation existing between two events such that one event causes (or enables) the other event, such as “hard rain causes flooding” or “taking a train requires buying a ticket.” In previous work these relations have been classified into several types based on a variety of points of view. In this work, we consider four types of causal relations—*cause*, *effect*, *precond(ition)* and *means*—mainly based on agents’ volitionality, as proposed in the research field of discourse understanding. The idea behind knowledge acquisition is to use resultative connective markers, such as “because,” “but,” and “if” as linguistic cues. However, there is no guarantee that a given connective marker always signals the same type of causal relation. Therefore, we need to create a computational model that is able to classify samples according to the causal relation. To examine how accurately we can automatically acquire causal knowledge, we attempted an experiment using Japanese newspaper articles, focusing on the resultative connective “*tame*.” By using machine-learning techniques, we achieved 80% recall with over 95% precision for the *cause*, *precond*, and *means* relations, and 30% recall with 90% precision for the *effect* relation. Furthermore, the classification results suggest that one can expect to acquire over 27,000 instances of causal relations from 1 year of Japanese newspaper articles.

Categories and Subject Descriptors: I.2.6 [Artificial Intelligence]: Learning—Knowledge acquisition; H.3.1 [Information storage]: Content Analysis—Linguistic processing; I.2.7 [Artificial Intelligence]: Natural Language Processing—Language parsing and understanding; Text analysis

General Terms: Languages, Experimentation

Additional Key Words and Phrases: Causal relation, connective marker, volitionality

---

## 1. INTRODUCTION

The general notion of *causality* has been a topic of inquiry since the days of the ancient Greeks. From the early stages of research into artificial intelligence

---

Authors’ address: T. Inui, 4259, Nagatsuda, Midoriku, Yamato, 226-8503, Japan; K. Inui and Y. Matsumoto, School of Information Science, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept., ACM, Inc., 1515 Broadway, New York, NY 10036 USA, fax: +1 (212) 869-0481, or [permissions@acm.org](mailto:permissions@acm.org).

© 2005 ACM 1530-0226/05/1200-0435 \$5.00

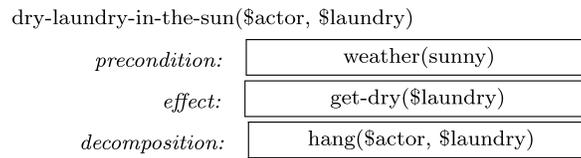


Fig. 1. An example of a plan operator.

*precond*((it is sunny), (dry the laundry in the sun))  
*effect*((dry the laundry in the sun), (the laundry gets dry))  
*means*((hang laundry), (dry the laundry in the sun))

Fig. 2. An example of casual relation instances.

(AI), many researchers have been concerned with common-sense knowledge, particularly cause–effect knowledge, as a source of intelligence. Relating to this interest, ways of designing and using a knowledge base of causality information to realize natural language understanding have also been actively studied [Schank and Abelson 1977; Heckerman et al. 1993]. For example, knowledge about the preconditions and effects of actions is commonly used for discourse understanding based on plan recognition. Figure 1 gives a typical example of this sort of knowledge about actions. An action consists of precondition and effect slots (in this case “weather” and “get-dry,” respectively) and is labeled with a header (“dry-laundry-in-the-sun”).

This knowledge-intensive approach to language understanding results in a bottleneck due to the prohibitively high cost of building and managing a comprehensive knowledge base. Despite the considerable effort put into the Cyc [Lenat 1995] and OpenMind [Stork 1999] projects, it is still unclear how feasible it is to try to build such a knowledge base manually. Very recently, on the other hand, several research groups have reported on attempts to automatically extract causal knowledge from a huge body of electronic documents [Garcia 1997; Khoo et al. 2000; Girju and Moldovan 2002; Satou et al. 1999]. While these corpus-based approaches to the acquisition of causal knowledge have considerable potential, they are still at a very preliminary stage in the sense that it is not yet clear what kinds and how much causal knowledge they will be able to acquire, how accurate the acquisition process can be made, and how useful acquired knowledge will be for language understanding.

Motivated by this background, in this paper, we describe our approach to automatic acquisition of causal knowledge from a document collection. We aim to acquire causal knowledge, such as those in Figure 2, which are binominal relations whose headings indicate the types of causal relation and whose arguments indicate the events involved in a causal relation. The causal relation instances in Figure 2 can be seen as constituent elements of the plan operator in Figure 1. The relations in Figure 2, therefore, represent a decomposition of the plan operator in Figure 1.

We use resultative connective markers as linguistic cues to acquire causal knowledge. For example, given the following sentences (1), we may be able to acquire the causal relation instances given in Figure 2,

- (1) a. Because it was a sunny day today, the laundry dried well.  
 b. It was not sunny today, but John could dry the laundry in the sun.

The idea of using these sorts of cue phrases to acquire causal knowledge is not novel in itself. In this paper, however, we address the following subset of the unexplored issues, focusing on knowledge acquisition from Japanese texts:

- What kinds and how much causal knowledge is present in the document collection?
- How accurately can relation instances be identified?
- How many relation instances can be acquired from currently available document collections?

In Section 2 and Section 3, we describe causal knowledge that we aim to acquire. In Section 2, we describe a typology of causal relations we deal with. In Section 3, we describe the importance of characterizing causal relation instances as knowledge rather than as rhetorical expressions. We then draw together issues in defining the problem to be tackled in this paper. In Section 4, we outline several previous research efforts on knowledge acquisition. In Section 5, we introduce the data on which we base our investigation and acquisition of causal knowledge. We describe an investigation into the distribution of causal relations in Japanese newspaper articles. The main part of the investigation is conducted based on human judgments using linguistic tests. Section 6 and Section 7 describe methods for automatically acquiring causal knowledge from text using machine-learning approaches. In Section 6, we deal with volitionality estimation. Volitionality information is a useful feature for classifying samples in accordance with our typology of causal relations. Section 7 describes our method of automatic causal relation identification and experimental results using it. Section 8 concludes this paper and outlines future work.

## 2. A TYPOLOGY OF CAUSAL RELATIONS

There have been various attempts to classify relations between textual segments, such as sentences and clauses. These range from rhetorical relations to higher-order abstract relations by linguistics [e.g. Mann and Thompson 1987; Nagano 1986; Ichikawa 1978] and AI researchers [e.g. Litman and Allen 1987; Allen 1995; Schank and Riesbeck 1981; Schank and Abelson 1977] and [Hobbs 1979, 1985].

In this work, we focus mainly on a set of causal relations related to the Allen's plan operator [Litman and Allen 1987; Allen 1995] which was proposed in the research field of discourse understanding. Figure 1 is a simple example of the plan operator. The reasons for the adoption of this scheme are:

1. While some relations proposed by previous researchers are unclear as to whether they represent the rhetorical level, the higher-order abstract level (discussed in detail later), or indeed a mixed level of both types of relation, Allen's relations definitely represent the higher-order abstract level.

Table I. Typology of Causal Relations

Causal Relations	Meaning	Examples of Templates Used in Linguistic Tests
$cause(SOA_1, SOA_2)$	SOA <sub>1</sub> causes SOA <sub>2</sub>	SOA <sub>2</sub> usually happens as a result of the fact that SOA <sub>1</sub> happens.
$effect(Act_1, SOA_2)$	SOA <sub>2</sub> is the effect of Act <sub>1</sub>	SOA <sub>2</sub> usually happens as a result of the execution of Act <sub>1</sub> .
$precond(SOA_1, Act_2)$	SOA <sub>1</sub> is a precondition of Act <sub>2</sub>	Act <sub>2</sub> cannot be done unless SOA <sub>1</sub> holds/happens. If SOA <sub>1</sub> often holds/happens, one will often execute Act <sub>2</sub> .
$means(Act_1, Act_2)$ (same agents)	Act <sub>1</sub> is a means of executing Act <sub>2</sub>	Someone executes often Act <sub>1</sub> in order to execute Act <sub>2</sub> . If someone executes Act <sub>1</sub> , then she can often execute Act <sub>2</sub> .

- Some previous works [e.g. Hobbs 1985] tend to involve various kinds of relations in addition to causal relations in order to explain all of the relations held in a text. On the other hand, Allen deals with causal relations in detail.
- It has been reported that Allen's relations are very useful in the research field of discourse recognition [Allen 1983].

The original plan operator consists of a frame-organized relationship between a central (core) event and its surrounding (marginal) situation. A core event, which is shown at the top of Figure 1, often represents an agent's volitional action. The marginal event surrounding the core event represents actions and states of affairs that occurred around the core event.

The idea of organizing a core event and its marginal events as an instance of a plan operator is an appealing one. One of the reasons for this is that it is assumed that a causal relation is held between a core event and each marginal event. Furthermore, it is assumed that different kinds of causal relations can be held between events, since the plan operator defines a number of different semantic roles, *precondition*, *effect*, and *decomposition*.

Based on the distinction between these relations, we have created a typology of causal relations as summarized in Table I. In our typology, we classify causal relations with respect to the *volitionality* of their arguments. The volitionality of an event distinguishes it as being an action or a state of affairs. In this paper, we call the set of elements which constitute the arguments of causal relations an *event*. An agent's volitional action, such as "drying laundry" is indicated as an *action* (abbreviated as Act) and all other kinds of non-volitional states of affairs such as "laundry drying" termed a *state of affairs* (abbreviated as SOA).

The volitionality combinations shown in the first column of Table I are a necessary condition for each causal relation class. In the table, Act<sub>*i*</sub> denotes a volitional action and SOA<sub>*i*</sub> denotes a nonvolitional state of affairs. For example,  $effect(Act_1, SOA_2)$  denotes that, if the *effect* relation holds between two arguments, the first argument must be a volitional action and the second must be a nonvolitional state of affairs. One of the main goals of discourse understanding is the recognition of the intention behind each volitional action appearing

in a given discourse. The importance of distinguishing volitional actions from nonvolitional SOA has been discussed already [e.g. Carberry 1990].

The *effect* relation in Table I represents the relationship between a core action and its effect on a state of affairs. The *precond* relation represents the relationship between a core action and its precondition state of affairs. The *means* relation represents the relationship between a core action and its marginal subaction, which is called *decomposition* in the plan operator. We impose the additional necessary condition on the *means* relation that the agents of the two argument actions must be identical. Because the two different cases obviously have a different intentional structure: the case where one agent executes two actions and the other case where two agents execute two different actions with independent intentions. The *cause* relation represents the relationship between two states of affairs. We decided to include this relation in this work, because although the *cause* relation is less related to the plan operator than the other three relations, it often indicates typical causal relations, such as “heavy rain causes flooding.”

It is not easy to provide rigorously sufficient conditions for each relation class. To avoid addressing unnecessary philosophical issues, we provide a set of linguistic tests for each relation class that loosely specify the sufficient condition. Some examples of the templates that we used in the linguistic tests are shown in Table I. Note that while the examples are expressed in English, our study is carried out on Japanese. The details of the linguistic tests are described in Section 5.2.

### 3. APPROACH AND PROBLEM

#### 3.1 Using Cue Phrases

Let us consider the following examples in English, from which one can obtain several observations about the potential sources of causal knowledge.

- (2) a. The laundry dried well today because it was sunny.  
 b. The laundry dried well, though it was not sunny.  
 c. If it was sunny, the laundry could dry well.  
 d. The laundry dried well because of the sunny weather.  
 → e. *cause*(⟨it is sunny⟩, ⟨laundry dries well⟩)
- (3) a. Mary used a tumble dryer because she had to dry the laundry quickly.  
 b. Mary could have dried the laundry quickly if she had used a tumble dryer.  
 c. Mary used a tumble dryer to dry the laundry quickly.  
 d. Mary could have dried the laundry more quickly with a tumble dryer.  
 → e. *means*(⟨using a tumble dryer⟩, ⟨drying laundry quickly⟩)

First, causal relation instances can be acquired from sentences with various connective markers. (2e) Is a *cause* relation instance that is acquired from subordinate constructions with various connective markers as in (2a–2d). Likewise, the other classes of relations are also acquired from sentences with various connective markers, as in (3). The use of several markers is advantageous for improving the recall of the acquired knowledge.

Second, it is also interesting to see that the source of knowledge could be extended to sentences with an adverbial clause or even a prepositional phrase as exemplified by (2d), (3c), and (3d). Note, however, that the acquisition of causal relation instances from such incomplete clues may require additional effort in order to infer elided constituents. To acquire a *means* relation instance (3e) from (3d), for example, one might need the capability to paraphrase the prepositional phrase “with a tumble dryer” into a subordinate clause, say, “if she had used a tumble dryer.”

Third, different kinds of instances can be acquired with the same connective marker. For example, the type of knowledge acquired from sentence (2a) is a *cause* relation, but that from (3a) is a *means* relation.

Here example (4) corresponds to (2) and example (5) corresponds to (3). For example, while (2a) is expressed in English and (4a) is expressed in Japanese, these sentences express the same meaning.

One could acquire the same causal relation instances from sentences with connective markers such as “tame (because, in order to),” “ga (but),” and “reba (if).” For example, a *cause* relation instance (4e) is acquired from subordinate constructions with various connective markers as in (4a–4d). A *means* relation instance (5e) is acquired from sentences, such as (5a–5d). Similarly, different kinds of instances can be acquired with the same connective marker. The type of knowledge acquired from sentence (4a) is a *cause* relation, but that acquired from (5a) is a *means* relation.

- (4) a. hareteiruta tame sentakumono ga yoku kawaita.  
 b. hareteinai ga sentakumono ga yoku kawaita.  
 c. haretei-reba sentakumono ga yoku kawaitanoni.  
 d. seiten de sentakumono ga yoku kawaita.  
 → e. *cause*((hareru), (sentakumono ga yoku kawaku))
- (5) a. sentakumono wo hayaku kawakasu tame kansouki wo tsukatta.  
 b. kansouki wo tsukattei-reba hayaku sentakumono ga kawakaseta hazuda.  
 c. kansouki wo tukatta no ha hayaku sentakumono wo kawakashitakatta kara da.  
 d. kansouki de sentakumono wo hayaku kawakashita.  
 → e. *means*((kansouki wo tsukau), (sentakumono wo hayaku kawakasu))

Thus, although the connective marker is useful for knowledge acquisition, as described above, there is a problem. There are no one-to-one correspondences between markers and causal relations. Therefore, we need to create a computational model that is able to classify and identify which type of causal relation can be acquired from a given sentence. This is the central issue addressed in this paper.

### 3.2 Rhetorical and Causal Relations

The sentences exemplified in (6) represent a relationship between the volitional actions “mikan no kawa wo muku (peeling an orange),” and “mikan wo taberu (eating an orange),” expressed using various types of rhetorical expressions. Looking at these sentences, we are able to recognize different rhetorical relations

among (6a<sub>j</sub>–6c<sub>j</sub>) and (6a<sub>e</sub>–6c<sub>e</sub>). In general, the sentences (6a<sub>j</sub>) and (6a<sub>e</sub>) can be interpreted as a PURPOSE rhetorical relation, (6b<sub>j</sub>) and (6b<sub>e</sub>) as a CONDITION rhetorical relation, and (6c<sub>j</sub>) and (6c<sub>e</sub>) as a CONTRAST rhetorical relation.

- (6) a<sub>j</sub>. mikan wo taberu tameni kawa wo muita.(PURPOSE)  
 b<sub>j</sub>. mikan ha kawa wo mukana to tabe rare nai. (CONDITION)  
 c<sub>j</sub>. mikan no kawa wo muita noni taberu koto ga dekinakatta.(CONTRAST)  
 a<sub>e</sub>. I peeled an orange to eat it.(PURPOSE)  
 b<sub>e</sub>. If you do not peel an orange, you cannot eat it.(CONDITION)  
 c<sub>e</sub>. Though I peeled an orange, I could not eat it.(CONTRAST)

Here, we assume that the reason that we are able to recognize all of these sentences as *coherent* is that the sentences conform to the causal knowledge presented in (7).

- (7) [knowledge]  
       *means*(⟨mikan no kawa wo muku⟩, ⟨mikan wo taberu⟩)  
                                   peeling an orange          eating an orange  
 [meaning]  
       A volitional action “mikan no kawa wo muku (peeling an orange),”  
       is a means of executing another volitional action “mikan wo taberu  
       (eating an orange)”.

This suggestion is supported by the following observation: we are able to recognize the sentences in (8) as *incoherent* because we do not possess any knowledge to which they can coherently conform. [We attached the symbol “\*” to the beginning of each example sentence in (8) to indicate that the sentence is *incoherent*. This is a convention in the linguistics community.]

- (8) a<sub>j</sub>. \* mikan no kawa wo muita tameni mikan wo taberu koto ga dekinakatta.  
 b<sub>j</sub>. \* mikan no kawa wo muita noni ame ga futteiru.  
 a<sub>e</sub>. \* I was not able to eat an orange to peel it.  
 b<sub>e</sub>. \* Though I peeled an orange, it rains.

We aim to acquire causal knowledge like (7), which can work as the basis for recognition of rhetorical expressions, such as in (8) as being coherent.

Figure 3 illustrates the relationship between the rhetorical level and the causal knowledge level. We assume that the relationships between event instances indicated in a text are located at the rhetorical level. We also assume that when the event instances are abstracted from the rhetorical level to a higher abstract level, the relationships between abstracted classes are located at the causal knowledge level.

For example, while the two events expressed in the sentence “Though I peeled the orange, I could not eat it” constitute the elements at the rhetorical level, the relationship between the abstracted classes “peeling an orange” and “eating an orange” is assumed at the causal knowledge level. Our proposed collection of causal relations should constitute a higher level of abstraction than mere rhetorical relations. At the very least, we must, therefore, abstract away modality information such as:

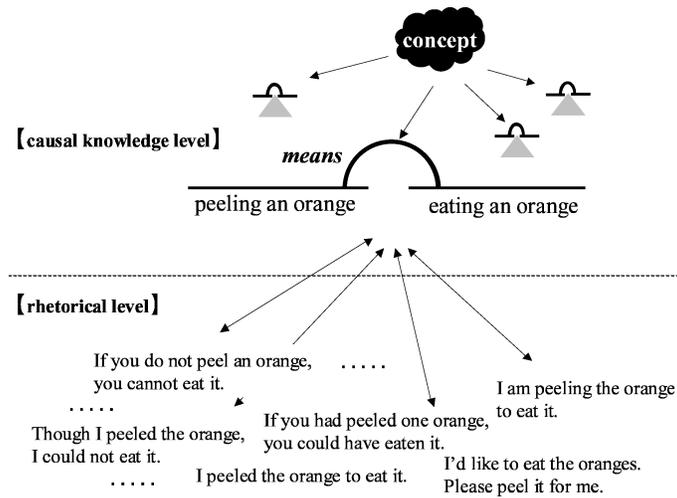


Fig. 3. Rhetorical level and causal knowledge level.

- Tense and aspect:<sup>1</sup> whether the event represented has already occurred or not.

(after abstraction)

mikan no kawa wo muita → mikan no kawa wo muku  
 I peeled the orange → I peel the orange

- Elements relating to the information structure: which element is focused on in a sentence and which element is new information.

(after abstraction)

mikan ga tabe rare ta → mikan wo taberu  
 The orange was eaten (by someone) → (someone) eats the orange

### 3.3 Representation of Arguments

Knowledge representation is one of the central issues in the field of AI. We represent arguments of causal relation instances by natural language expressions such as Figure 2, (2e), and (3e), instead of by any formal semantic representation language for the following reasons.

- It is still unclear whether abstraction is a necessary process in representing knowledge. Having decided to do abstraction, it is very hard to decide the most suitable level for the abstraction.
- It has proved difficult to design a formal language that can fully represent the diverse meanings of natural language expressions.

<sup>1</sup>In the linguistic community, in general, the constituents of a sentence can be separated into the proposition and the modality, according to the following definition:

- Proposition consists of the parts, which indicate objective facts, independent of the agent in the sentence.
- Modality consists of the remaining parts of sentence.

We follow the above definition. Although some linguists take a different view, we deal with tense and aspect under the rubric of modality information in this work.

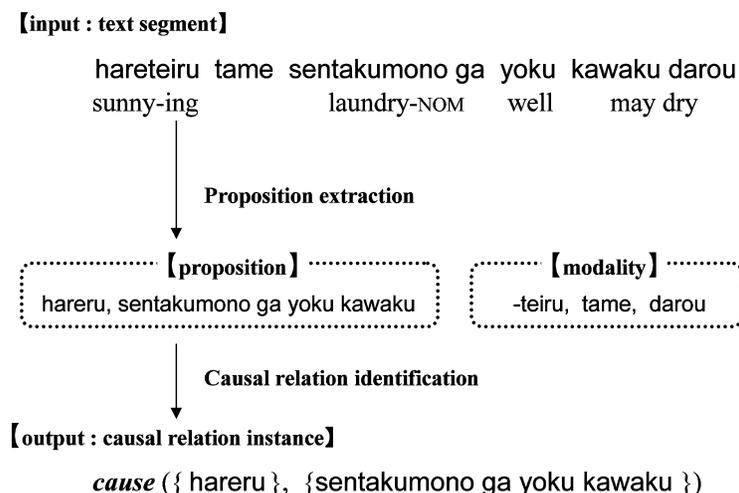


Fig. 4. Knowledge acquisition workflow.

- As discussed in Iwanska and Shapiro [2000], there has been a shift toward viewing natural language as the best means for knowledge representation.
- As discussed in detail in Section 7.4, we anticipate using acquired causal knowledge within the framework of case-based reasoning, which has been applied successfully in the machine translation community, for example Satou’s work [Satou and Nagao 1990]. We believe that acquired knowledge is sufficient for practical use without any abstraction process.

In fact, for example, Harabagiu and Moldovan [1997] proposed a text-based knowledge representation system, which applied a knowledge expression scheme based on natural language. All the knowledge in the Open-Mind Common Sense knowledge base organized by Singh et al. [2002] is also represented by English sentences and Liu et al. [2003] reported that Singh’s database could be successfully used for textual affect sensing.

### 3.4 Target Problem

On the basis of the above, to acquire causal knowledge from text, we use a simple procedure consisting of two main phases, shown in Figure 4.

Given text segments, the process of acquiring causal knowledge forms two independent phases: proposition extraction and causal relation identification.

1. Proposition extraction: Removing the modality expressions and extracting the propositional expressions from a text segment, normally a sentence. For example, in Figure 4, two propositional expressions indicating different events “hareru (it is sunny),” and “sentakumono ga yoku kawaku (laundry dries well),” are extracted from the input sentence “hareteiru tame sentakumono ga yoku kawaku darou (the laundry may dry well because it is sunny).”
2. Causal relation identification: Identifying causal relations held between extracted proposition pairs. In Figure 4, the *cause* relation is identified as

holding between the two propositional expressions “hareru” and “sentakumono ga yoku kawaku.”

In this paper, we aim to develop an implementation of the latter phase, causal relation identification, because the former phase, proposition extraction, can be simply resolved using current natural language processing techniques.

Simply speaking, the proposition extraction phase is run as follows. First, the morphological analysis is executed to get part-of-speech information for each word in the input sentence. Next the dependency structure analysis is executed to extract chunks that express events. For each chunk extracted, some modality elements, which are identified by part-of-speech information, are removed. Finally, the remaining elements in the chunk are extracted as a proposition.

#### 4. RELATED WORK

In recent years, there have been several attempts at automatically acquiring causal knowledge from document collections [Garcia 1997; Satou et al. 1999; Khoo et al. 2000; Low et al. 2001; Girju and Moldovan 2002; Terada 2003; Torisawa 2003]. First, in this section, we introduce four studies of causal knowledge acquisition, three of which make use of cue phrases as we do and one of which makes use of a statistical technique. Next, we describe the study of rhetorical parsing and compare it to our work. Probabilistic modeling of causal relations, such as Pearl [1988, 2000] and Heckerman et al. [1997] forms one subfield of the causality research activity. However, since in this paper, we focus on automatic knowledge acquisition from text rather than modeling, we do not describe the studies of probabilistic modeling.

##### 4.1 Cue Phrase-Based Approaches

Girju and Moldovan [2002] proposed a method for acquiring causal knowledge from text in English based on the triplet patterns as follows:

$$\langle NP_1 \textit{ clue } NP_2 \rangle$$

where *clue* is a causative verb, and  $NP_1$  and  $NP_2$  are noun phrases. Causative verbs express a causal relation between the subject and object (or prepositional phrase of the verb), such as “cause” and “force.”

In their method, first all triplets matching the pattern above are collected from the gloss of WordNet 1.7 [Fellbaum 1998]. The following is one such example:

(9) *extreme leanness*  $NP_1$  *usually caused by* *starvation or disease*  $NP_2$

Next, for each pair of nouns determined as above, they search for sentences containing the noun pairs in a document collection. From these sentences, they automatically determine all patterns  $\langle NP_1 \textit{ verb/verbal expression } NP_2 \rangle$  where  $NP_1$ - $NP_2$  is the pair under consideration. For example, “*extreme leanness*  $NP_1$  *associated with* *starvation or disease*  $NP_2$ ” is an extracted pattern corresponding to the noun pairs in the previous example. In addition, “*associated with*” is a new extracted expression that suggests causality. Third, to eliminate the patterns,

```

[*]-
  &(v-ch) -> (subj) -> [T:cause.object]
  (cc|cnd) -> [result] +->
    (loc) +-> [in] +-> (pcomp)
    ->[T:effect.event]
    (phr) -> [in] -> (pcomp)
    ->[T:effect.object],,

```

Fig. 5. An example syntactic pattern proposed by Khoo et al.

which do not represent causal relation, they apply semantic constraints, which are mainly based on the semantic categories defined in WordNet. Finally, they ranked the remaining patterns according to the ambiguity of the sense for the verb, and its frequency. From the evaluation using the TREC-9 [Voorhees and Harman 2001] collection of texts which contains 3GB of news articles from Wall Street Journal, Financial Times, Financial Report, etc., they extracted about 1300 patterns (causal expressions). Using 300 of the 1300 patterns, the accuracy of their method, as evaluated by human subjects, was about 65%.

Terada [2003] proposed a method for acquiring causal knowledge similar to that of Girju et al. He used only a small set of cue phrases as follows:

- causative verb  
*cause causing caused by result in resulting in result from lead to*
- prepositional phrase  
*because of due to*

While Girju et al. used only noun phrases as contextual information around the cue phrases, Terada used three types of contextual information.

- *NP* Noun phrase (equivalent to Girju et al.)  
ex. aircraft loss *NP*
- *NP+PP* In addition to *NP*, a prepositional modifier of the *NP* is considered.  
ex. aircraft loss *NP of oil pressure* *PP*
- *N/A* In addition to *NP+PP*, full clauses are considered.  
ex. aircraft oil pressure lose *clause*

He applied a constraint based on the frequency of the patterns using a sequential pattern-mining algorithm, PrefixSpan [Pei et al. 2001] instead of the semantic constraints applied by Girju et al. In an evaluation using aviation safety reports handled by the Aviation Safety Reporting System in National Aeronautics and Space Administration, a collection containing 24,600 documents (15.4 words/document), he extracted very few causal expressions. In the case of using *N/A* contextual information, he extracted 23 expressions, which was the largest number of causal expressions extracted for any of the contextual information classes.

Khoo et al. [2000] acquired causal knowledge with manually created syntactic patterns specifically for the MEDLINE text database [MEDLINE 2001]. Figure 5 shows an example of syntactic patterns where square brackets refer to character strings and round brackets refer to syntactic or semantic role. In their method, an input sentence is first parsed and a syntactic structure is built.

Next, if the syntactic structure matches any of the syntactic patterns, the element in the “cause” slot is extracted as the cause element of the causal relation and the element in the “effect” slot is extracted as the effect element of the same causal relation. In all, 68 syntactic patterns were constructed. Their method was evaluated using 100 MEDLINE abstracts and had precision of about 60%.

In three studies described above, explicit cue phrases were used; this approach is adopted in our method. Girju et al. and Terada focused on causative verbs, and Khoo et al. on syntactic patterns. However their approach is different from ours. While they focused mainly on discovering and creating new patterns for acquiring causal expressions to cover more instances of causal relations, we focus on classifying and identifying the types of causal relations using a particular cue phrase (described in the remaining sections in this paper).

#### 4.2 A Statistics-Based Approach

Torisawa [2003] proposed a statistical method for extracting common-sense inference rules, which are identical with our causal knowledge from Japanese newspaper articles. While the above-mentioned previous studies focused on explicit cue phrases to acquire knowledge of causal relation, in Torisawa’s method, in addition to the expressions with cue phrases, the expressions without cue phrases are also dealt with as knowledge source such as:

- (10)  $\frac{\text{biiru wo nomi}}{\text{beer-ACC drink}}_{VP_1} \quad \frac{\text{yotta}}{\text{get drunk}}_{VP_2} .$

The key idea of Torisawa’s method for knowledge acquisition is to construct a statistical model for causal relations using several kinds of expressions, including those with/without cue phrases based on an assumption that is related to the sharing of nouns by two expressions.

**object-sharing assumption:** If two expressions  $e_1$  and  $e_2$  (which represent two different events in a verb phrase pair) share the same object, then it is likely that they hold the relation “if  $e_1$  then  $e_2$ ,” otherwise it is likely that they do not hold the relation “if  $e_1$  then  $e_2$ .”

Torisawa extracted approximately 200 inference rules using 33 years of newspaper articles. The extracted rules were all of high quality. His approach, based on statistics, may have wider coverage than cue phrase-based approaches, since it can use as its source parallel verb phrase pairs without cue phrases. However, current reported coverage on extraction is not high enough to make it usable in applications, such as inference systems.

#### 4.3 Rhetorical Parsing

Several linguistic theories of textual coherence have been proposed. Rhetorical Structure Theory (RST) [Mann and Thompson 1987] is one such theory. In RST, every text segment, or more precisely clause, has a relationship to another text segment. This relationship is known as a rhetorical relation.

The aim of rhetorical parsing [Marcu 1997, 2002] is to correctly determine the type of rhetorical relation in a given document. However, note that our typology

of causal relations is not just a simple subset of common rhetorical relations as proposed in RST. That is, identifying causal relations is fundamentally different from rhetorical parsing. As described in Section 3.2, our proposed collection of causal relations constitute a higher level of abstraction than mere rhetorical relations. While rhetorical parsing can make clear which types of coherence relations are presented in linguistic expressions, causal knowledge provides the basis for explaining how a rhetorical relation can be recognized as coherent.

## 5. DATA COLLECTION AND ANALYSIS

Linguists suggest that there are some expressions that act as cue phrases for causal relations. However, it is not clear what proportion of linguistic expressions containing cue phrases actually implicitly include causal relations and what kinds of causal relations these are. To resolve these issues, we investigated the distribution of causal relations in Japanese text documents.

Currently, there are several kinds of electronic text resources available, such as newspapers, dictionaries [e.g. RWC 1998], encyclopedias [e.g. Britannica 1998], novels, e-mails, and web documents. We selected newspaper articles as text resources in this study. One of the reasons for this decision is that both the morphological analyzer and dependency structure analyzer we used in preprocessing are optimized for the types of sentences found in newspaper articles. To acquire correct knowledge, the sentences included in the source text should be amenable to correct analysis at the preprocessing stage.

### 5.1 Causal Expressions in Japanese Text

Causal relations can be expressed in various ways. For example, Altenberg [1984] attempted to make an inventory of various causal expressions in spoken and written British English. In Altenberg's work, four major types of causal expression are taken into consideration, which are defined on the basis of the link between the events:

- Adverbial linkage (e.g., *so*, *hence*, *therefore*)
- Prepositional linkage (e.g., *because of*, *on account of*)
- Subordination linkage (e.g., *because*, *as*, *since*)
- Clause-integrated linkage (e.g., *that's why*, *the result was*)

It can be assumed that a similar causal expression classification can be performed on Japanese text [Masuoka and Takubo 1992; Masuoka 1997] as has been performed on English. The following examples show Japanese causal expressions categorized according to part of speech of the cue phrases.

(11) —Conjunction

kaze wo hiita. dakara gakkou wo yasunda.

(someone) caught a cold. Therefore (someone) was absent from school.

—Conjunctive particle

ie wo kau tame chokinsuru.

(someone) saves money in order to buy her house.

Table II. Frequency Distribution of Conjunctive Particles

ga	(but)	131,164
tame	(because)	76,087
to	(if/when)	56,549
reba	(if)	48,606
nagara	(while)	13,796
kara	(because)	10,209
node	(because)	9,994
nara	(if)	7,598
tara	(if)	6,027
noni	(but)	2,917

—Particle

kaze de gakkou wo yasumu.

(someone) is absent from school with a cold.

The categories of explicit cue phrases described above have a relaxed correspondence with their arguments, which are marked by the follow categories of cue phrase.

cue phrase	argument
conjunction	sentence
conjunctive particle	clause
particle	phrase

Of the three types of cue phrases, we focus here on the conjunctive particle by which clauses are most likely to be indicated since:

- In general, an event is described in text concisely using a predicate and some case elements connected by particles
- Clauses are usually constructed using the same constituents as events, say, a predicate and case elements connected by particles

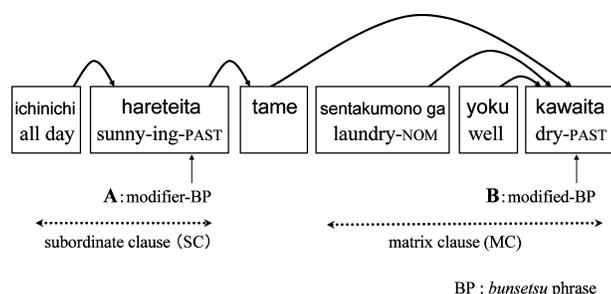
We consider that if a cue phrase expression occurs with high frequency and no ambiguous uses, it is a suitable expression for use in knowledge acquisition. Here, we examined the frequency distribution of conjunctive particles. Table II shows the ten most frequent conjunctive particles in the collection of Nihon Keizai Shimbun newspaper articles from 1990 [Nikkei 1990]. This table was generated by counting all conjunctive particles after morphological analysis of the articles using *ChaSen*<sup>2</sup> [Matsumoto et al. 1999].

In this table, the word “tame (because),” and “reba (if),” have a pragmatic constraint on the inevitability implicit in the relationship between two arbitrary events. The relations signaled by these words usually involve a high degree of inevitability and, therefore, indicate less ambiguous relations. We manually confirmed this pragmatic constraint on the inevitability implicitly in *tame*, based on the approximately 2000 examples. Based on the above discussion, we selected the word *tame* as our main target for further exploration.

<sup>2</sup>Available from <http://chasen.aist-nara.ac.jp/hiki/ChaSen/>.

Table III. Frequency Distribution of *Tame* in Intrasentential Contexts

Type	Freq.	Examples
(a) Adverbial verb phrase	42,577	hareta <i>tame</i> sentakumono ga yoku kawaita. The laundry dried well because it was sunny.
(b) Other types	33,510	kore ha ryokousha no <i>tame</i> no kansouki desu. This is a tumble drier for the tourist.

Fig. 6. Structure of a *tame*-complex sentence.

Next, Table III shows the frequency distribution of the intra-sentential contexts in which *tame* appears in the same newspaper article corpus used above. The sentences classified into the “adverbial verb phrase” type are extracted as follows:

*adverbial verb phrase.* First, for each sentence including *tame*, the dependency structure between *bunsetsu*-phrases is automatically constructed with *CaboCha*<sup>3</sup> [Kudo and Matsumoto 2003]. The *bunsetsu*-phrase is one of the fundamental units in Japanese, which consists of a content word (noun, verb, adjective, etc.) accompanied by some function words (particles, auxiliaries, etc.). Next, each modifier *bunsetsu*-phrase of *tame* indicated as “A” in Figure 6 and modified *bunsetsu*-phrase of *tame* indicated as “B” in Figure 6 is identified. A sentence is classified as being of the “adverbial verb phrase” type if both modifier *bunsetsu*-phrase and modified *bunsetsu*-phrase fulfill one of the following morphological conditions:

- c1. Includes a morpheme, whose part of speech is “doushi-jiritsu (verb-independent),” as defined in *ChaSen*’s dictionary.<sup>4</sup>  
e.g., kawaku (dry)
- c2. Includes a morpheme whose part of speech is “keiyoushi-jiritsu (adjective-independent),” as defined in *ChaSen*’s dictionary.  
e.g., tsuyoi (strong)
- c3. Includes a morpheme, whose part of speech is “meishi-keiyoudoushigokan (nominal adjectival stem),” as defined in *ChaSen*’s dictionary.  
e.g., kouchouda (satisfactoriness)

<sup>3</sup>Available from <http://chasen.org/~taku/software/cabocho/>.

<sup>4</sup><http://chasen.aist-nara.ac.jp/stable/ipadic/ipadic-2.6.3.tar.gz>.

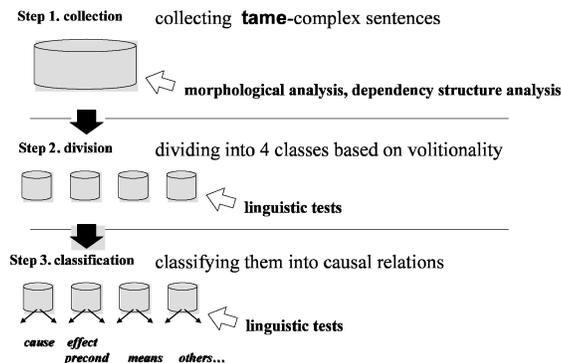


Fig. 7. Workflow for investigating frequency distribution.

- c4. Includes a morpheme, whose part of speech is any noun category except “meishi-keiyoudoushigokan (nominal adjectival stem),” and does not include the morpheme *no* whose part of speech is “joshi-rentaika (particle-adnominal),”  
 e.g., *mitoushida* (prospect)

In Table III, we see that more than half of sentences are classified into type (a). In these sentences, the word *tame* is used as an adverbial connective marker accompanying a verb phrase that constitutes an adverbial subordinate clause. We are pleased to observe this tendency, because the knowledge acquisition from this type of sentence is expected to be easier than from the other types of sentences shown in Table IIIb. As described in Section 3.1, one may need to infer elided constituents in order to acquire causal relation instances from the sentences classified into type b. Based on this preliminary survey, we restrict our attention to the sentences classified into type a in Table III. Hereafter, this type of sentence will be indicated as **tame-complex sentences**.

## 5.2 Data Collection Procedure

We assembled a collection of data for examining the distribution of causal relation in *tame-complex sentences* as follows (see also Figure 7):

*Step 1: Collection.* First we randomly took 1000 samples that were automatically categorized into *tame-complex sentences* from the same newspaper article corpus mentioned in Section 5.1. We used *ChaSen* and *CaboCha* again for preprocessing. Next, from the 1000 samples, we removed sentences which are inappropriate as knowledge resource. We removed interrogative sentences and sentences from which a subordinate-matrix clause pair was not properly extracted due to preprocessing (morphological analysis and dependency structure analysis) errors. In the preliminary analysis, we found that causal relation instances have not been acquired from interrogative sentences. As a result we were left with 994 sentences. These 994 sentences fulfill the condition of *c1* for the adverbial verb phrase. We indicate this set of sentences as  $S_1$ .

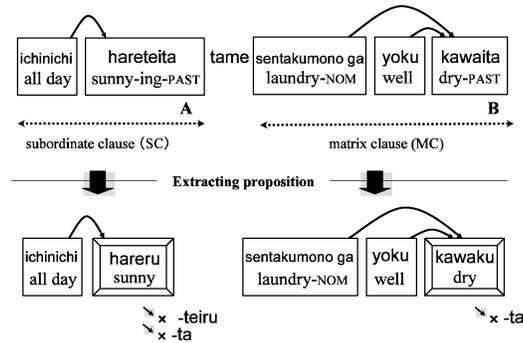


Fig. 8. Proposition extraction.

 Table IV. Distribution of Causal Relations in Tame Complex Sentences in  $S_1^a$ 

Class	SC	MC	Frequency	Most Frequent Relation and its Ratio
A	SOA	SOA	229	$cause(SOA_s, SOA_m)$ 0.96 (220/229)
B	Act	SOA	161	$effect(Act_s, SOA_m)$ 0.93 (149/161)
C	SOA	Act	225	$precond(SOA_s, Act_m)$ 0.90 (202/225)
D	Act	Act	379	$means(Act_m, Act_s)$ 0.85 (323/379)
	total		994	0.90 (894/994)

<sup>a</sup>SC denotes the subordinate clause and MC denotes the matrix clause.  $Act_s$  and  $SOA_s$  denote an event indicated by the SC, and  $Act_m$  and  $SOA_m$  denote an event indicated by the MC.

*Step 2: Division.* We extracted the proposition from each subordinate and matrix clause of sentences in  $S_1$ . Figure 8 shows an example of proposition extraction process. For each sentence in  $S_1$ , we first removed the modality elements attached to the end of the head verb, which is the verb in *bunsetsu*-phrase “A” or “B” from the sentence. The remaining elements in the sentence were then regarded as the proposition. By this operation, some modality information, such as tense or passive voice information, was erased. Hereafter, we indicate the extracted proposition as the *clause*. Next, we manually divided the 994 samples into four classes depending on the combination of volitionality (volitional action or nonvolitional SOA) in the subordinate and matrix clauses. Volitionality was judged using the linguistic tests described in Section 5.2.1. The frequency distribution of the four classes (A–D) is shown in the left-hand side of Table IV. The clause pairs classified into the class A fulfill the necessary conditions for the *cause* relation, the clause pairs classified into the classes B and C fulfill the necessary conditions for the *effect* relation and the *precond* relation, and the clause pairs classified into the class D fulfill the necessary condition for the *means* relation. (It should be noted that it is not always possible to confirm that the same agents’ condition for class D is fulfilled.)

*Step 3: Classification.* We then examined the distribution of the causal relations we could acquire from the samples of each class using the linguistic tests. The details of the linguistic tests for judging the causal relations are also described in Section 5.2.1.

**5.2.1 Linguistic Tests.** A linguistic test is a method for judging whether a linguistic expression, normally a sentence, conforms to a given set of rules. We

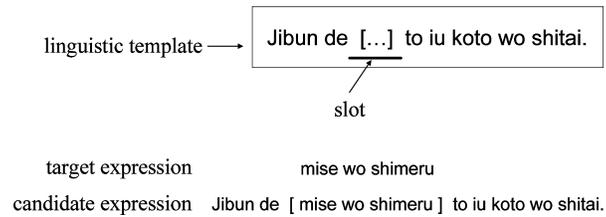


Fig. 9. Linguistic template.

call the sentence to be judged a *target* expression. The rules are realized as a linguistic template, which is linguistic expression including several slots. An example of linguistic templates is shown in Figure 9.

In practice, a linguistic test is usually applied using the following steps:

1. Preparing the templates.
2. Embedding the target expression in the slots of the template to form a *candidate* sentence.
3. If the candidate sentence is judged to be correct syntactically and semantically, the target expression is judged to conform to the rules. If the candidate sentence is determined to be incorrect, the target is judged nonconforming.

**5.2.2 Linguistic Tests for Judging Volitionality.** We prepared four templates for volitionality judgments as follows:

- vol\_t1*    jibun de    [ ... ] to iu koto wo        shitai.  
           by oneself    the thing that [ ... ]-ACC    want to do
- vol\_t2*    jibun de    [ ... ] to iu koto wo        suru tsumori desu.  
           by oneself    the thing that [ ... ]-ACC    will do
- vol\_t3*    jibun de    [ ... ] to iu koto wo        suru kotoni suru.  
           by oneself    the thing that [ ... ]-ACC    will do
- vol\_t4*    jibun deha [ ... ] to iu koto wo        shitaku ha nai.  
           by oneself    the thing that [ ... ]-ACC    do not want to do

The square brackets indicate the slots in which the target expression is embedded. If a candidate sentence is determined to be correct by a human subject, the embedded target is judged to be a volitional action. On the other hand, if the candidate sentence is incorrect, this template is rejected, and another is tried. If all templates are tried without success, the target expression is judged to be a nonvolitional SOA. The following are examples of this process. In each case, the first item shows a target expression, the middle item or items show candidate sentences, and the last item shows the final judgment.

- (12) a. mise wo shimeru.  
           (someone) closes the store.  
       b. jibun de [ mise wo shimeru ] to iu koto wo suru tsumori desu.  
       c. volitional action
- (13) a. kabushiki shijou ga teimeisuru.  
           The stock market downturn occurs.

- b. \* jibun de [ kabushiki shijou ga teimeisuru ] to iu koto wo shitai.
- c. \* jibun de [ kabushiki shijou ga teimeisuru ] to iu koto wo suru tsumori desu.
- d. \* jibun de [ kabushiki shijou ga teimeisuru ] to iu koto wo suru kotoni suru.
- e. \* jibun de [ kabushiki shijou ga teimeisuru ] to iu koto wo shitaku ha nai.
- f. nonvolitional SOA

5.2.3 *Linguistic Tests for Judging Causal Relations.* We prepared three to seven templates for each causal relation. The following are some examples. All templates are shown in the appendix.

—*cause*

[SOA] (to iu) koto ga okoru sono kekka toshite  
 [SOA] (that) thing-NOM happen as a result of  
 shibashiba [SOA] (to iu) koto ga okoru.  
 usually [SOA] (that) thing-NOM happen

—*effect*

[Act] (to iu) koto wo suru to  
 [Act] (that) thing-ACC execute  
 shibashiba [SOA] (to iu) koto ga okoru.  
 usually [SOA] (that) thing-NOM happen

—*precond*

[SOA] (to iu) joukyou deha  
 [SOA] (that) state-TOPIC  
 shibashiba [Act] (to iu) koto wo suru.  
 usually [Act] (that) thing-ACC execute

—*means*

X ga [Act] (to iu) koto wo jitsugensuru sono shudan toshite  
 X-NOM [Act] (that) thing-ACC realize its by means of  
 X ga [Act] (to iu) koto wo suru no ha mottomo de aru.  
 X-NOM [Act] (that) thing-ACC execute-thing-TOPIC plausible

We embed the subordinate and matrix clauses in the slots of the templates to form candidate sentences. If a candidate sentence is determined to be correct, the causal relation corresponding to the particular template used is assumed to hold between the clauses. If the candidate sentence is incorrect, this template is rejected, and another is tried. If all templates are tried without success, the candidate sentence contains a relation unclassifiable within our typology and is assigned to the class *others*.

The expressions “shibashiba (often)” or “futsuu (usually)” in templates indicate a pragmatic constraint on the inevitability of the relationship between any two events: that is, the relations indicated by these words usually have a high degree of inevitability. For example, a causal relation can be said to exist between two events shown in (14a). However, we are able to recognize the sentence in (14b) which contains the expression *futsuu* as incorrect, since the relation possesses a very low degree of inevitability.

- (14) a. takara kuji wo kattara ittou ga atatta.  
When (someone) bought a lottery ticket, (someone) won a first prize.
- b. \*takara kuji wo kau to futsuu ittou ga ataru.  
When (someone) buys a lottery ticket, (someone) usually wins a first prize.

This constraint affects the judgments, which are made based on inevitability. Therefore, causal relations with a very low degree of inevitability can be rejected.

The following are examples of judgment process. In each case, the first item shows a sentence including target expressions, the middle item or items show candidate sentences, and the last item shows the final judgment.

- (15) a. shijou ga teimeishita tame mise wo shimeru.  
(someone) closes the store because the stock market downturn occurs.
- b. [ shijou ga teimeisuru ] joukyou deha  
shibashiba [ mise wo shimeru ] koto wo suru.
- c. *precond* relation
- (16) a. kaigansen wo kaihatsusuru tameni chousa wo oeteiru.  
The survey for exploiting the coastline has finished.
- b. \* [ kaigansen wo kaihatsusuru ] to iukoto wo suru to sono kekka tsuujou  
[chousa wo oeru ] joukyou ni naru.
- c. ...
- d. (no templates are correct)
- e. *others* relation

5.2.4 *Reliability of Judgments.* Volitionality and causal relations were judged using the linguistic tests. To estimate the reliability of judgments, two human subjects majoring in computational linguistics annotated the texts with both volitionality and causal relation information. For a measure of reliability of judgments, we used the  $\kappa$  statistic, which is one of the well-known statistics for a measure of reliability of judgments' agreement between two subjects originally proposed by Cohen [1960]. It is claimed by Krippendorff [1980] that if  $\kappa > 0.8$ , it is highly reliable.

We calculated the  $\kappa$  statistic using 200 annotated samples. The  $\kappa$  value was 0.93 for volitionality and 0.88 for causal relations. This means that the reliability of judgments of both volitionality and causal relations are sufficiently high.

### 5.3 Analysis

5.3.1 *The Marker tame.* The right-hand side of Table IV shows the most abundant relation and its ratio for each class A–D. For example, given a *tame*-complex sentence, if the volitionality of the subordinate clause is a volitional action and the volitionality of the matrix clause is a nonvolitional SOA (namely, class B), they are likely to conform to the relation *effect*(Act<sub>s</sub>, SOA<sub>m</sub>) with a probability of 0.93 (149/161). The following are examples of the most abundant relation in a given class.

- (17) Tai de manguroubu wo hakaishita tame dai suigai ga hasseishita.  
 Serious flooding occurred because mangrove swamps were destroyed in Thailand.  
 Act<sub>s</sub>: Tai de manguroubu wo hakaisuru  
 SOA<sub>m</sub>: dai suigai ga hasseisuru  
 → *effect*(⟨Tai de manguroubu wo hakaisuru⟩, ⟨dai suigai ga hasseisuru⟩)
- (18) Pekin eno kippu wo kau tame kippu uriba ni itta.  
 (someone) went to the ticket office in order to buy a ticket to Beijing.  
 Act<sub>s</sub>: Pekin eno kippu wo kau  
 Act<sub>m</sub>: kippu uriba ni iku  
 → *means*(⟨kippu uriba ni iku⟩, ⟨Pekin eno kippu wo kau⟩)

The following are examples of cases where the most abundant relation in a given class did not hold.

- (19) a. kigyō no seichō ga mikomeru youni natta tame kiun ga takamatteiru.  
 The growing tendency appeared because the company growth is expected.  
 (Although this sentence fulfills the necessary condition for the *cause* relation, it is rejected for all of the templates in linguistic test.)
- b. takusan shuuyousuru tame houru ha oogigata ni natteiru.  
 The hall contains a large audience owing to the shape of a sector.  
 (Although this sentence fulfills the necessary condition for the *effect* relation, it is rejected for all of the templates in linguistic test.)
- c. kogakki ga butai no ondo ni nareru tameni oukesutora ha juppun amari chouseishita.  
 The orchestra keyed their old instruments for about 10 minutes for having them accustomed to stage temperature.  
 (Although this sentence fulfills the necessary condition for the *precond* relation, it is rejected for all of the templates in linguistic test.)

Table IV shows the quite suggestive result for causal knowledge acquisition. As far as *tame*-complex sentences are concerned, if one can determine the value of the volitionality of the subordinate and matrix clauses, one can classify samples, that is, subordinate and matrix clauses pairs indicating each different event extracted from *tame*-complex sentences into the four relations—*cause*, *effect*, *precond*, and *means*—with precision of 85% or more. Motivated by this observation, in the next section we first address the issue of automatic estimation of volitionality before moving onto the issue of automatic classification of causal relations. As shown in Table IV, we cannot completely classify the samples into correct causal relation classes using only the value of the volitionality and syntactic information. Therefore, we need to use further rich information for more precise classification. The details will be described in Sections 6 and 7.

5.3.2 *Other Markers*. We attempted the same procedure outlined in Section 5.2 using the other five cue phrases. The cue phrases and data sizes used in this investigation are shown in Table V.



Table VI. Distribution of Causal Relations in Sentences Including *node*

Class	SC	MC	Frequency	Most Frequent Relation and its Ratio	
A	SOA	SOA	337	<i>cause</i> (SOA <sub>s</sub> , SOA <sub>m</sub> )	0.88 (297/337)
B	Act	SOA	180	<i>effect</i> (Act <sub>s</sub> , SOA <sub>m</sub> )	0.93 (160/180)
C	SOA	Act	310	<i>precond</i> (SOA <sub>s</sub> , Act <sub>m</sub> )	0.81 (251/310)
D	Act	Act	151	—	0 (0/151)
total			978	0.72 (708/978)	

Table VII. Distribution of Causal Relations in Sentences Including *reba*

Class	SC	MC	Frequency	Most Frequent Relation and its Ratio	
A	SOA	SOA	55	<i>cause</i> (SOA <sub>s</sub> , SOA <sub>m</sub> )	0.73 (40/55)
B	Act	SOA	80	<i>effect</i> (Act <sub>s</sub> , SOA <sub>m</sub> )	0.33 (26/80)
C	SOA	Act	25	<i>precond</i> (SOA <sub>s</sub> , Act <sub>m</sub> )	0.56 (14/25)
D	Act	Act	22	—	0 (0/22)
total			182	0.44 (80/182)	

Table VIII. Distribution of Causal Relations in Sentences Including *tara*

Class	SC	MC	Frequency	Most Frequent Relation and its Ratio	
A	SOA	SOA	45	<i>cause</i> (SOA <sub>s</sub> , SOA <sub>m</sub> )	0.38 (17/45)
B	Act	SOA	65	<i>effect</i> (Act <sub>s</sub> , SOA <sub>m</sub> )	0.34 (22/65)
C	SOA	Act	44	<i>precond</i> (SOA <sub>s</sub> , Act <sub>m</sub> )	0.50 (22/44)
D	Act	Act	27	—	0 (0/27)
total			181	0.34 (61/181)	

Table IX. Distribution of Causal Relations in Sentences Including *ga*

Class	SC	MC	Frequency	Most Frequent Relation and its Ratio	
A	SOA	SOA	86	<i>cause</i> (SOA <sub>s</sub> , SOA <sub>m</sub> )	0.17 (15/86)
B	Act	SOA	44	—	0 (0/44)
C	SOA	Act	28	<i>precond</i> (SOA <sub>s</sub> , Act <sub>m</sub> )	0.32 (9/28)
D	Act	Act	31	—	0 (0/31)
total			189	0.13 (24/189)	

Table X. Distribution of Causal Relations in Sentences Including *noni*

Class	SC	MC	Frequency	Most Frequent Relation and its Ratio	
A	SOA	SOA	105	<i>cause</i> (SOA <sub>s</sub> , SOA <sub>m</sub> )	0.40 (42/105)
B	Act	SOA	27	<i>effect</i> (Act <sub>s</sub> , SOA <sub>m</sub> )	0.37 (10/27)
C	SOA	Act	31	<i>precond</i> (SOA <sub>s</sub> , Act <sub>m</sub> )	0.74 (23/31)
D	Act	Act	24	—	0 (0/24)
total			187	0.40 (75/187)	

(b) noncausal relations within our typology. The samples are grouped according to cue phrase.

(22)—*node*

- a. samuku naru node kenkou joutai wo shinpaisuru.  
(someone) worries about her own health since it becomes cold.  
(*precond* relation)
- b. shokubutsu ha kankyo no henka de itamu node hinpan na basho gae ha sakeru.

You should avoid resiting your garden plants because the blight is caused by changes of the environment.  
(*others* relation)

—reba

- a. fukyou ga shinkokuka su-reba saimu futan ga zoukasuru.  
The debt burden increases if the economic downturn becomes serious.  
(*cause* relation)
- b. yoi toki mo a-reba warui toki mo aru.  
There are ups and downs.  
(*others* relation)

—tara

- a. fukusayou ga de-tara fukuyou wo chuushisuru.  
You should stop taking medicines when you experience any side effects from them.  
(*precond* relation)
- b. hon no seiri wo shitei-tara omoshiroi mono ga detekita.  
I found an interesting thing when I tidied up my bookshelf.  
(*others* relation)

—ga

- a. kinri daka de juchuu ga nobinayamu ga furu seisan wo zokkousuru.  
Although the order entry is low in consequence of high interest rates, we continue full production.  
(*precond* relation)
- b. satsujin ha hetta ga goutou no kensuu ga fueta.  
While the number of murders decreases, the number of robberies increases.  
(*others* relation)

—noni

- a. ryoukin wo bai ni shita noni kyaku ga fueta.  
The customers increased although we doubled our usual price.  
(*effect* relation)
- b. kaisha ni iku noni danchi kara basu ni noru.  
I get on the bus near my apartment to go to office.  
(*others* relation)

Based on these results, we attempt to automatically acquire causal knowledge from *tame*-complex sentences, as described in Section 7. Note that the difficulty of acquiring causal knowledge depends on the characteristics of the target cue phrases. In the case of *tame*, we can focus on classifying causal relations without worrying about the degree of inevitability implicit in relations between events thanks to the pragmatic constraint on inevitability in *tame*-complex sentences. However, when focusing on cue phrases with a lower degree of inevitability than *tame*, such as *tara*, in addition to classifying causal relations, it is necessary to take account of the framework that determines the degree of inevitability.

## 6. ESTIMATION OF VOLITIONALITY

In the previous section, to acquire causal knowledge, it is important to estimate the volitionality (volitional action or nonvolitional SOA) of the clauses. Hereafter, we call volitionality of the clause *clausal volitionality*. In the field of linguistics, work has been done on volitionality of verbs as one of the mood attributes (we call volitionality of the verb *verbal volitionality*). However, no work has been done on volitionality over larger linguistic segments, such as phrases and clauses. In this section, we investigated experimentally how accurately clausal volitionality of clauses can be estimated using Support Vector Machines—an accurate binary classification algorithm.

### 6.1 Preliminary Analyses

Clausal volitionality depends mostly on the verb in a clause, more precisely, it depends on verbal volitionality. That is, if certain clauses contain the same verb, volitionality of these clauses also tends to be the same. For example, in the set  $S_1$  which includes 1988 clauses, there are 720 different verbs, 299 of which occur 2 times or more. Of these 299 verbs, 227 occurred exclusively in clauses sharing the same clausal volitionality.

Nevertheless, there are some counterexamples. Some samples were found to depend on other contextual factors. For example, both the subordinate clause of (23a) and the matrix clause of (23b) contain the same verb “*kakudaisuru* (expand).” However although the clausal volitionality of the former case is a volitional action, the clausal volitionality of the latter case is a nonvolitional SOA.

- (23) a. *seisan nouryoku wo kakudaisuru tame setsubi touchisuru.*  
 (A company) will make plant investments to expand production ability.
- b. *kanrihi ga sakugenshita tame eigyourieki ga kakudaishita.*  
 Business profit expanded as a result of management costs being reduced.

In order to estimate clausal volitionality with high accuracy, we need to incorporate additional factors. As a result of analyzing the *tame*-complex sentences in  $S_1$ , we found other factors in addition to the verb that help determine clausal volitionality.

- A clause tends to be nonvolitional SOA when the agent is not a person or an organization.
- The volitionality value of a clause tends to change depending on whether it appears as a subordinate clause or a matrix clause.
- The volitionality value of a clause tends to change based on modality, such as tense.

### 6.2 Experimental Conditions

**6.2.1 Learning Algorithm.** We applied Support Vector Machines (SVMs) as learning algorithm. SVMs are binary classifiers, originally proposed by Vapnik [Vapnik 1995]. SVMs have performed with high accuracy in various tasks, such

as text categorization [Joachims 1998], bioinformatics [Vert 2002], and face identification [Jonsson 2000].

SVMs are one of the large-margin classifiers. Based on the large-margin criterion, the training procedure finds a hyperplane, that not only separates the samples in one class from those in the other, but also for which the separation is as large as possible.

Although we omit details of the algorithm here, we describe some advantages of SVMs. In most cases, the generalization performance of SVMs does not depend on the dimensionality of the feature space. Furthermore, SVMs can use *kernel* functions, which can deal with nonlinear classification by means of an implicit mapping of the original feature space into a high-dimensional space. We used the quadratic polynomial kernel as the kernel function. The polynomial kernel function makes it possible to deal with any combination of features.

In this experiment, using only intraclause information, we created a separate classifier for each clause type, since we found little evidence of a correlation between the clausal volitionality of a matrix clauses and a subordinate clause.

In this work, we used the TinySVM software package.<sup>5</sup>

**6.2.2 Features.** Based on the results of our preliminary analyses mentioned in Section 6.1, we used several kinds of information for the features to represent the clauses in the sentences. Table XI shows all the features we used. Fortunately, the verbal features, which provide the most important information, can be easily extracted from the currently obtainable dictionaries: the EDR concept dictionary, the dictionary incorporated in the ALT-J/E translation system, and NTT Goi-Taikai. The case and modality information can also be extracted using simple pattern-matching rules. While almost all features can be extracted automatically, it is not so easy to extract agent information. Phrases to represent agents usually do not appear overtly in Japanese complex sentences. In the field of natural language processing, ellipsis resolution is well-known as a very difficult task. In this experiment, we implemented a simple agent feature extractor with a precision of about 60%, instead of attempting to implement an ellipsis resolution component. This means agent information extraction is not perfect due to ellipsis. The effect of this source of error will be described later.

**6.2.3 Data.** We used all the sentences in  $S_1$  as training samples and a new set of tame-complex sentences,  $S_2$ , as test samples. The set  $S_2$  includes 985 tame-complex sentences. This set was created using the same procedure as  $S_1$ . We first sampled 1000 random sentences from a newspaper article corpus issued in a different year than  $S_1$ . We then removed interrogative sentences and sentences from which a subordinate-matrix clause pair was not properly extracted due to preprocessing errors, leaving us with 985 test samples. The frequency distributions of clausal volitionality for both  $S_1$  and  $S_2$  are shown in Table XII.

<sup>5</sup>Available from <http://chasen.org/~taku/software/TinySVM/>.

Table XI. Feature Set Used for Volitionality Estimation<sup>a</sup>

<b>class</b> —descriptions
<b>V Word</b> —a base form item
<b>V EDR</b> —Four features indicating the verb class given by the EDR concept dictionary [Yokoi 1995]: (1) true if the verb is “idou (movement),” or “kou (action),” false otherwise; (2) true if the verb is “joutai (state),” “henka (change),” or “genshou (phenomenon),” false otherwise; (3) true if both (1) and (2) are true; (4) true if none of the above is true.
<b>V ALT-J/E</b> —A set of binary features indicating the verb class given by the dictionary incorporated in the ALT-J/E translation system [Ikehara et al. 1991; 1997]: “joutai (state),” “keizoku (continuous situation),” “shunkan (momentary situation),” “jidou (intransitive),” “tadou (transitive),” “hojo (auxiliary),” “kanou (potential),” “jihatsu (spontaneous),” “shieki (causative),” “ukemi (passivizable),” “ukemi (higai) (indirect-passivizable).”
<b>V Goi-Taikai</b> —Verbal semantic attributes in NTT Goi-Taikai [Ikehara et al. 1997].
<b>C Marker</b> —“ga (NOM),” “wo (ACC),”
<b>C Element</b> —The concept of case element described in NTT Goi-Taikai [Ikehara et al. 1997].
<b>M Tense</b> —“-ru (PRESENT),” form or “-ta (PAST),” form.
<b>M Aspect</b> —“-teiru (-ing),” form or not.
<b>M Voice</b> —“-reru (PASSIVE),” form or not. “-seru (CAUSATIVE),” form or not.
<b>M Potential</b> —“-dekiru (POSSIBLE),” form or not.
<b>M Negative</b> —“-nai (NEG),” form or not.
<b>Agent</b> —Whether or not the agent is a human or an organization.

<sup>a</sup> V: Verb C: Case M: Modality.

Table XII. Ratio of Volitionality for Each Clause Type

		Frequency	
		Act / SOA	Total
$S_1$	Subordinate clause	539 / 455	994
	Matrix clause	603 / 391	994
$S_2$	Subordinate clause	613 / 372	985
	Matrix clause	650 / 335	985

### 6.3 Results

Table XIII shows the results. The values in Table XIII denote accuracy, which is calculated by dividing the number of correct samples by the total number of samples. The row “Sc & Mc” is the accuracy when estimating both subordinate and matrix, clauses in the same sentence. The column “SVM2” denotes accuracy in the case where a classifier is trained using all the features shown in Table XI. The column “SVM1” denotes accuracy in the case where a classifier is trained using all the features shown in Table XI, except agent information. The columns “BL1” and “BL2” denote accuracy for the baseline. “BL1” is very simple but the most commonly used model which outputs volitional action all the time because the frequency of appearance of a clause with volitional action is greater than the frequency with nonvolitional SOA (see Table XII). “BL2” denotes the accuracy achieved by applying a simple classification strategy as follows:

Table XIII. Accuracy of Volitionality Estimation

	BL1	BL2	SVM1	SVM2
Subordinate clause	0.622	0.803	0.841	0.885
Matrix clause	0.660	0.864	0.879	0.888
Sc & Mc	0.474	0.686	0.784	0.867

*BL2: voting strategy.* (a) if a verb of an input clause appeared in the training set, the clause is classified by a majority vote, and (b) if the voting is even or the verb is not present in the training set, the clause is classified as a volitional action by default.

From the preliminary analyses mentioned in Section 6.1, it is found that clausal volitionality is mostly dependent on verb information. Therefore, the approach based on verb information is used as the second baseline model.

Table XIII offers several insights. First, the accuracy of “BL2” is an improvement on the accuracy of “BL1.” This result suggests that verb information is useful in performing clausal volitionality estimation. Second, the results obtained using SVMs are an improvement on the baseline accuracy. The case feature and the modality feature are responsible for this improvement because the accuracy of both “SVM1” and “SVM2” is greater than that of “BL2.” Comparison of the accuracy of “SVM1” and “SVM2” shows that agent information also contributes to improved accuracy, although the agent information extraction is not perfect in this work.

Next, we introduced a reliability metric to further improve accuracy. When the reliability of estimated volitionality is known, the accuracy of automatic classification of causal relations can be improved by removing samples where the reliability value is low.

To estimate the reliability, we used the absolute values of the discriminant function (the distances from the hyperplane) output by the SVMs. We set a reliability threshold value  $\sigma$ , thus ensuring that a judgment would only be output for a given sample when the reliability was greater than  $\sigma$ .

We applied this metric to the results of “SVM2.” By varying  $\sigma$ , we obtained the coverage-accuracy curves of Figure 10 where:

$$\text{Coverage} = \frac{\# \text{ of samples output by the model}}{\# \text{ of number of all samples}}$$

$$\text{Accuracy} = \frac{\# \text{ of samples } \textit{correctly} \text{ output by the model}}{\# \text{ of samples output by the model}}$$

Figure 10 indicates that when the threshold value is increased, the number of samples output decreases, and coverage is low. The accuracy, in the case of low coverage, is higher than that in the case of high coverage. These results confirm that it is possible to produce clausal volitionality estimates with a very high confidence level.

## 7. AUTOMATIC ACQUISITION OF CAUSAL KNOWLEDGE

In this section, we describe an experiment designed to assess how accurately causal relation instances can be acquired automatically. As shown in Figure 4

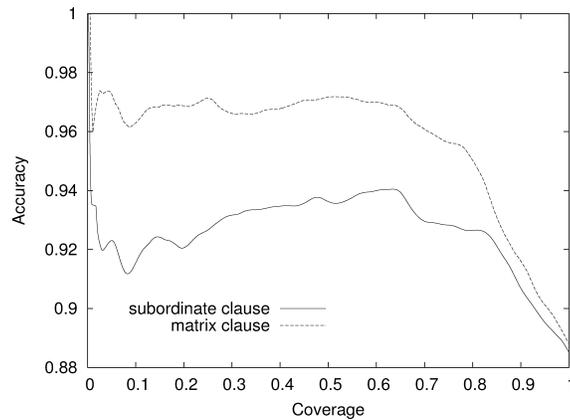


Fig. 10. Coverage-accuracy curves for clausal volitionality estimation.

in Section 3.4, the process consists of two main phases. We implemented a high precision rule based proposition extractor for the first phase using existing natural language processing techniques. In this section we describe our approach to automatic identification (classification) of causal relations.

In Section 5, we described the distribution of causal relations in Japanese newspaper articles. It was demonstrated that the pairs of subordinate and matrix clauses in the *tame*-complex sentences can be classified by hand into the causal relation classes within our typology with a precision of 85% or more. In Section 6, we showed that it is possible to achieve clausal volitionality estimation with good accuracy using a machine-learning approach.

Based on these findings, we attempted to identify causal relation instances contained in *tame*-complex sentences. Clausal volitionality was used as a feature in performing classification. We used five classes—*cause*, *effect*, *precond*, *means*, and *others* relations. The *others* class contains relation instances, which were categorized into the *others* class during the investigation of distribution of causal relations described in Section 5. We again used SVMs as the classifier.

### 7.1 Identification Procedure

Clausal volitionality is a major factor used in classifying the sample. A sample therefore could be incorrectly identified if the degree of confidence in clausal volitionality estimation is low. Based on this assumption, we use the following procedure for the identification of causal relations.

*Step A: Clausal Volitionality Estimation.* A subordinate clause/matrix clause pair is extracted from an input *tame*-complex sentence. The clausal volitionality for each clause is also estimated with a given degree of confidence. Hereafter, the degree of confidence of a subordinate clause is denoted by  $conf_s$  and that of a matrix clause is denoted by  $conf_m$ .

*Step B: Filtering.* Suppose that a threshold for the degree of confidence of subordinate clauses is  $\sigma_s$  and a threshold for the degree of confidence of matrix

Table XIV. Distribution of Causal Relations in Tame Complex Sentences in  $S_2$ 

Class	Most Frequent Relation and its Ratio	
A	<i>cause</i> (SOA <sub>s</sub> , SOA <sub>m</sub> )	0.98 (193/196)
B	<i>effect</i> (Act <sub>s</sub> , SOA <sub>m</sub> )	0.78 (108/139)
C	<i>precond</i> (SOA <sub>s</sub> , Act <sub>m</sub> )	0.94 (166/176)
D	<i>means</i> (Act <sub>m</sub> , Act <sub>s</sub> )	0.79 (375/474)
		0.85 (842/985)

clauses is  $\sigma_m$ . If  $conf_s \geq \sigma_s$  and  $conf_m \geq \sigma_m$  then go on to Step C, otherwise classify into the *others* relation class deterministically and the procedure is finished. Although this step reduces recall, our hope is that it increases precision.

*Step C: Classification.* Classify the sample into a causal relation class using SVMs. The standard SVMs can apply only to binary classification. Therefore, the one-versus-rest method was used so that we could apply SVMs to multiple classifications. The one-versus-rest method is one of common extensions to the multiple class case. See, for example, Rifkin and Klautau [2004] for details. For the N-ary classification problem, N different binary classifiers are constructed, one classifier for each class. The classifier for a class  $i$  classifies the samples into the class  $i$  or not- $i$ . In the classification phase the classifier with the maximal output defines the estimated class.

## 7.2 Experimental Conditions

7.2.1 *Features.* The features we used are as follows:

- f1.* The clausal volitionality estimated by the technique described in the previous section;
- f2.* All the features except “V word” shown in Table XI;
- f3.* Whether the agents of the two clauses in the sentence are the same.

The third, agent correspondence feature can be automatically extracted with a high level of precision by using the technique described in Nakaiwa and Ikehara [1995]. However, in this experiment, we were unable to implement this method because of the unavailability of the detailed linguistic rules. Instead, a simple rule-based extractor was used. For example, we decided that a sample has the same agent when the agent is represented by the case marker *ga* in the matrix clause and by ellipses in the subordinate clause.

7.2.2 *Data.* The data are the same as those in Section 6. We used the sentences in  $S_1$  as training samples and  $S_2$  as test samples. We first estimated the clausal volitionality and its reliability using all the training data in  $S_1$ . We then removed about 20% of the samples by applying the reliability metric. The remaining samples were then used to train the classifiers for the causal relations.

The distribution of the causal relations in tame-complex sentences in  $S_2$  is shown in Table XIV. The notations in this table follow those in Table IV.

Table XV. Three-Point Averaged Precision (Step B Skipped)

Feature	Three-point Averaged Precision			
	<i>cause</i>	<i>effect</i>	<i>precond</i>	<i>means</i>
R1: f1	0.883	0.556	0.781	0.869
R2: f2 + f3	0.769	0.588	0.943	0.722
R3: f1 + f2 + f3	0.993	0.854	0.992	0.972
R3': f1 + f2' + f3'	0.995	0.864	0.992	0.975

Table XVI. Three-Point Averaged Precision (Step B Included)

Feature	Three-point Averaged Precision			
	<i>cause</i>	<i>effect</i>	<i>precond</i>	<i>means</i>
R1: f1	0.905*	0.593*	0.799*	0.879*
R2: f2 + f3	—	—	—	—
R3: f1 + f2 + f3	0.992*	0.859*	0.989*	0.984*
R3': f1 + f2' + f3'	0.996*	0.882*	0.993*	0.988*

7.2.3 *Evaluation Measure.* Classification performance is evaluated using recall and precision, where, for each causal relation R:

$$\text{Recall} = \frac{\text{\# of samples correctly classified as R}}{\text{\# of all samples holding the target relation R}}$$

$$\text{Precision} = \frac{\text{\# of samples correctly classified as R}}{\text{\# of all samples output as being R}}$$

The three-point averaged precision is also used as a summary of the recall-precision curves to evaluate each causal relation. This value is the average of precision at the three points corresponding to recall values of 0.25, 0.50, and 0.75. The recall-precision curves are calculated using the same procedure as described for reliability in Section 6.3, where the reliability was defined as:

$$s_1 + (s_1 - s_2)$$

where  $s_1$  is the maximum discriminant function value obtained through the one-versus-rest method, and  $s_2$  is the second highest value.

### 7.3 Results

The results are shown in Tables XV, and XVI, and Figure 11.<sup>6</sup> Table XV shows the three-point averaged precision of the causal relation classification where Step B (filtering) is skipped. On the other hand, Table XVI shows the results where Step B is included. Figure 11 shows the recall-precision curves corresponding to R3 in Table XVI.

We discuss the results from the viewpoint of the effect of clausal volitionality on the classification. First, we consider the results of Table XV. In Table XV, R1

<sup>6</sup>Note that these results were obtained by running both stages of the acquisition process: proposition extraction and causal relation identification. However, since the first phase is highly accurate, these results are almost the same as those of automatic classification alone.

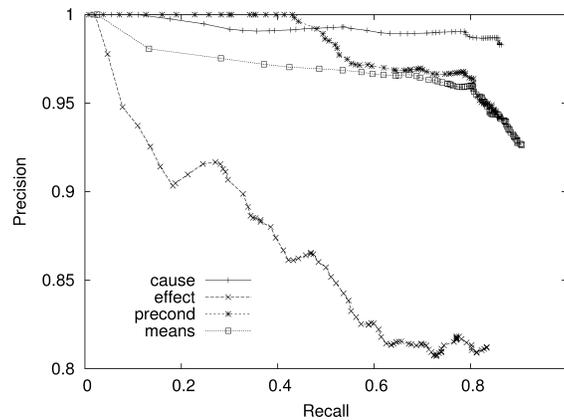


Fig. 11. Recall-precision curves for causal relation classification.

denotes the results in the case where the classifiers were trained using only  $f_1$ , clausal volitionality as features. In this case, no information except clausal volitionality was used. Therefore, we applied a minor variation of the classification process, “Step C-forR1,” instead of the one presented in Section 7.1.

#### Step C-forR1:

- If the clausal volitionality of a subordinate clause is estimated to be a **nonvolitional SOA** and the clausal volitionality of a matrix clause is estimated to be a **nonvolitional SOA**, classify the sample into the **cause** relation class.
  - If the clausal volitionality of a subordinate clause is estimated to be a **volitional action** and the clausal volitionality of a matrix clause is estimated to be a **nonvolitional SOA**, classify the sample into the **effect** relation class.
  - If the clausal volitionality of a subordinate clause is estimated to be a **nonvolitional SOA** and the clausal volitionality of a matrix clause is estimated to be a **volitional action**, classify the sample into the **precond** relation class.
  - If the clausal volitionality of a subordinate clause is estimated to be a **volitional action** and the clausal volitionality of a matrix clause is estimated to be a **volitional action**, classify the sample into the **means** relation class.
- (The degree of confidence in a causal relation classification is calculated as  $conf_s + conf_m$ .)

R2 denotes the results in the case where the classifiers were trained using  $f_2$  and  $f_3$ . R3 denotes the results in the case where the classifiers were trained using all the features described in Section 7.2.

The three-point averaged precision of R1 for each causal relation class is lower than that of R3. In R1, the error in clausal volitionality estimation is directly responsible for the error in causal relation classification. The three-point averaged precision of R2 is also lower than that of R3. Comparing R2 and R3, it

is clear that clausal volitionality plays an important role in classifying causal relations. The results for R1 and R2 demonstrate that clausal volitionality is one important factor for causal relation classification, however, the samples cannot be classified with high precision by considering clausal volitionality alone. This remark is supported by the observation that the three-point averaged precision of R3 is higher than that of both R1 and R2.

R3' represents the current upper bound of causal relation classification. These are the results in the case where the classifiers were trained with the feature information for the two primitive features, the agent feature and the agent correspondence feature, using a human judge instead of our simple feature extractor in an effort to avoid machine-induced errors in input data. In comparison between R3 and R3' our results (R3) do not reach the current upper bound R3'. However, the difference between R3 and R3' is small. This means that even if the feature extractor is improved, significant improvements in the causal relation classification will not be achieved.

Next, we discuss the effect of step B, the sample filtering process. Table XVI shows the results when Step B is included. The up arrow in Table XVI indicates that the performance improves as a result of the sample filtering process. The thresholds for the degree of confidence  $\sigma_s$  and  $\sigma_m$  are set by focusing on the ratio of the input samples that have the confidence score lower than the threshold. We had a series of experiments, varying the value of threshold. We set the threshold such that 20% of input samples are classified into the *others* relation class, because we achieved the best accuracy when using this ratio. For the most part the results suggest that the sample filtering process contributes to improving classification. The effect of filtering is especially strong in R1. In this work, Step B was implemented using a very simple algorithm. It can be assumed that classification precision can be further increased by using a more refined measure of the degree of confidence.

Finally, to statistically confirm the difference between methods (“Step B skipped” vs. “Step B included”), we applied the Wilcoxon matched-pairs signed-ranks test. The null hypothesis is that the difference between results of two methods has median value zero. As a result, the null hypothesis was rejected at 0.05 significance level ( $p$ -value  $\leq 0.03906$ ).

## 7.4 Discussion

### 7.4.1 How Much Causal Knowledge Could Be Acquired from Resources?

Let us estimate the amount of knowledge one can acquire from *tame*-complex sentences in a collection of one year of newspaper articles with a total of approximately 1,500,000 sentences.

Suppose that we want to acquire causal relation instances with a precision of, say, 99% for the *cause* relations, 95% for the *precond* relations and the *means* relations, and 90% for the *effect* relations. First, it can be seen from Figure 11 that we achieved 79% recall (REC) for the *cause* relations, 30% for the *effect* relations, 82% for the *precond* relations, and 83% for the *means* relations. Second, we assume that the frequency ratios (FR) of these relations to all the *tame*-complex sentences are as given in Table XIV. In this case, the frequency

ratio of the *cause* relations class for example was  $193/1000 = 19\%$ . From this, it can be seen that we achieved 64% recall:

$$\begin{aligned} & 0.19_{\text{FR}}^{\text{cause}} \times 0.79_{\text{REC}}^{\text{cause}} \\ & + 0.11_{\text{FR}}^{\text{effect}} \times 0.30_{\text{REC}}^{\text{effect}} \\ & + 0.17_{\text{FR}}^{\text{precond}} \times 0.82_{\text{REC}}^{\text{precond}} \\ & + 0.38_{\text{FR}}^{\text{means}} \times 0.83_{\text{REC}}^{\text{means}} = 0.64 \end{aligned}$$

Finally, since we collected about 42,500 tame-complex sentences from 1 year of newspaper articles (see Table III), we expect to acquire over 27,000 instances of causal relations ( $\simeq 42,500 \times 0.64$ ). This number accounts for 1.8% of all sentences (1,500,000 sentences), and is not small in comparison to the number of causal instances included in the Open Mind Common Sense knowledge base [Singh et al. 2002] and Marcu’s results [Marcu 2002].

**7.4.2 How Can the Acquired Knowledge Be Applied?** We anticipate using acquired causal knowledge within the framework of case-based reasoning. The reason for this decision is as below. Our causal relation instances, as discussed in Section 3.3, are not abstracted as propositional information. A reasonable concern is that when one applies causal relation instances as inference rules, there will be few input samples matched to the inference rules. However, even if there are no input samples that strictly match any of the inference rules, case-based reasoning can estimate a similarity value between an input sample and a sample in the memory-base, i.e., a causal relation instance, by initiating a dynamic abstraction process in order to discover an appropriate inference rule.

For example, suppose the input samples are:

- Hanako ga sentakumono wo kawakasu (Hanako dries the laundry)
- sentakumono ga kawaku (the laundry dries)

In this case, it could be inferred that the *effect* relation holds between the input samples when a causal relation instance such as (24), whose arguments are similar to the input samples, is included in the memory-base.

(24) *effect* ( $\langle$ Taro ga sentakumono wo kawakasu $\rangle$ ,  $\langle$ sentakumono ga kawaku $\rangle$ )  
                   Taro dries the laundry                   the laundry dries

Similarly, when an input is only partial, e.g., “Hanako ga sentakumono wo kawakasu” and the causal relation instance (24) is in the memory-base, an event “sentakumono ga kawaku” could be inferred to happen as a result of “Hanako ga sentakumono wo kawakasu.”

**7.4.3 Feasibility of the Extension to Another Resource.** We describe the feasibility of the extension of our proposed method to another resources. To do this, we focus on two points: topic and style.

**TOPIC.** We used only newspaper articles on economics in this study. Indeed, while one-half of sentences in the articles are about economics, the other half of sentences are about noneconomics, such as sentences (17) and (18)

in Section 5. We can expect that given document articles with some another topics as knowledge resources, the similar results to ours described in this paper, is obtained.

*STYLE*. On the other hand, we have to pay attention to extend the resources into texts which have different styles. Since we need to execute the morphological analysis and the dependency structure analysis for preprocessing, the quality of acquired knowledge is dependent on the performance of linguistic analyzers. For newspaper articles, we can use the highly accurate linguistic analyzers (*ChaSen* for morphological analysis and *CaboCha* for dependency structure analysis). However, for another text documents such the web data, we require additional effort in order to handle preprocessing errors. Fortunately, recently, the natural language processing techniques for the web data have been studied extensively and levels of performance have substantially increased [Suyama 2005].

### 7.5 Problems

There are some problems still to be resolved in order to refine the acquired knowledge. Indeed, there are some constituents in the instances, such as ellipses and pronouns, which render the instances incomplete. For example, the second arguments of the instance (25) would include the constituents “Tai de (in Thailand),” because the location where the mangrove is destroyed and where flooding occurred should be the same.

- (25) *effect* ((Tai de manguroubu wo hakaisuru),  
 (someone) destroys mangrove swamps in Thailand  
 (( $\phi$  = Tai de) daisuigai ga hasseisuru)  
 serious flooding occur (in Thailand))

Therefore we need to insert these various missing constituents in order to refine the acquired causal relation instances.

Furthermore, acquired instances currently contain unnecessary modifiers. These constituents should be removed in order to refine the acquired knowledge. Unfortunately, however, it is hard to determine which constituents are required to construct a causal relation instance and which constituents are optional. For example, the constituent “tsumetai (cool),” in (26a) may be required to construct the *cause* relation instance (26a). On the other hand, “akiguchi ni (in early autumn),” in (26b) may be optional.

- (26) a. *cause* ((tsumetai kuuki ga nagarekomu),  
cool air is coming  
 (saitei kion ga heinen wo shitamawaru))  
 lowest temperature is lower than in a normal year  
 b. *cause* ((akiguchi ni shukka tousuu ga fueru),  
 the number shipped increases in early autumn  
 (tonka ha sagaru))  
 cost of pork falls

To utilize acquired knowledge in applications such as inference systems, we, therefore, need to be careful to remove unnecessary modifiers only.

## 8. CONCLUSION

We described our approach to automatic knowledge acquisition of *causal relations* from a document collection. We considered four types of causal relations based on agents' volitionality, as proposed in the research field of discourse understanding. The idea behind knowledge acquisition is to use resultative connective markers as linguistic cues.

Our investigation of Japanese complex sentences led to the following findings:

- The pairs of subordinate and matrix clauses indicating each different event extracted from tame-complex sentences can be classified into the four relation types—*cause*, *effect*, *precond* and *means*—with a precision of 85% or more.
- Using SVMs, automatic causal knowledge acquisition can be achieved with high accuracy: 80% recall with over 95% precision for the *cause*, *precond*, and *means* relations, and 30% recall with 90% precision for the *effect* relation.
- The experimental results suggest that over 27,000 instances of causal relations could potentially be acquired from 1 year of Japanese newspaper articles.

In this work, we have dealt with only a small subset of all the textually encoded knowledge potentially available in the world. In order to increase the volume and refine the quality of the causal knowledge acquired, the following issues will need to be addressed.

- *Linguistic forms expressing events*: Events are expressed using a variety of linguistic forms: words, phrases, clauses, sentences, and intersentential units. In this work, we focused on clauses in trying to capture events. Other linguistic units should be exploited in order to increase coverage.
- *Other languages*: The framework discussed in this paper is not specific to Japanese. We want to investigate applications to other languages such as English. As described in Section 3.3, the arguments of causal relation instances are represented in natural language (in this case, Japanese) rather than in any formal semantic representation language. It will be interesting to investigate the compatibility of causal relation instances acquired from different source languages.
- *Other resources*: There is always a trade-off between quantity and quality of text documents. We used newspaper articles as a source of knowledge in this work. As preprocessing modules (morphological analysis and dependency structure analysis) are improved, we anticipate the incorporation of additional types of source texts such as e-mail and web pages in order to increase coverage.
- *Coreference and unnecessary modifiers*: As described in Section 7.5, there are some constituents in causal relation instances, such as ellipses and pronouns, which render the instances incomplete. Techniques for coreference (ellipses

and anaphora) resolution will need to be incorporated in our framework in the form of a preprocessing module. Similarly, unnecessary modifiers should be removed to refine acquired knowledge.

## APPENDIX

The following are all linguistic templates for judging causal relations to supplement to those shown in Section 5.2.3.

### —*cause*

*cau.t1.* [SOA] (to iu) koto ga okoru sono kekka to shite  
[SOA] (that) thing-NOM happen as a result of  
shibashiba [SOA] (to iu) koto ga okoru.  
usually [SOA] (that) thing-NOM happen

*cau.t2.* [SOA] (to-iu) joutai deha  
[SOA] (that) state-TOPIC  
shibashiba [SOA] (to iu) joutai to naru.  
usually [SOA] (that) become

*cau.t3.* [SOA] (to iu) joutai ni nareba sore ni tomonai  
[SOA] (that) become-if it-based on  
shibashiba [SOA] (to iu) joutai ni naru.  
usually [SOA] (that) become

### —*effect*

*eff.1.* [Act] (to iu) koto wo suru to  
[Act] (that) thing-ACC execute  
shibashiba [SOA] (to iu) koto ga okoru.  
usually [SOA] (that) thing-NOM happen

*eff.2.* [Act] (to iu) koto wo suru to sono kekka  
[Act] (that) thing-ACC execute-when as a result of  
tsuujou [SOA] (to iu) joukyou ni naru.  
usually [SOA] (that) state-happen

*eff.3.* [Act] (to iu) koto wo suru koto ha  
[Act] (that) thing-ACC execute-thing-TOPIC  
futsuu [SOA] (to iu) joutai wo tamotu.  
usually [SOA] (that) state-ACC keep

### —*precond*

*pre.1.* [SOA] (to iu) joukyou de ha  
[SOA] (that) state-TOPIC  
shibashiba [Act] (to iu) koto wo suru.  
usually [Act] (that) thing-ACC execute

*pre.2.* [SOA] (to iu) joutai de ha  
[SOA] (that) state-TOPIC  
shibashiba [Act] (to iu) koto wo suru.  
usually [Act] (that) thing-ACC execute

*pre\_3.* [SOA] (to iu) joutai ni naru baai  
 [SOA] (that) become-when  
 shibashiba [Act] (to iu) koto wo suru.  
 usually [Act] (that) thing-ACC execute

—*means*

*mea\_1.* X ga [Act] (to iu) koto wo jitsugensuru sono shudan toshite  
 X-NOM [Act] (that) thing-ACC realize its by means of  
 X ga [Act] (to iu) koto wo suru no ha mottomo de aru.  
 X-NOM [Act] (that) thing-ACC execute-thing-TOPIC plausible

*mea\_2.* X ga [Act] (to iu) koto wo jitsugensuru sono shudan toshite  
 X-NOM [Act] (that) thing-ACC realize its by means of  
 X ga [Act] (to iu) koto wo okonau no ha mottomo de aru.  
 X-NOM [Act] (that) thing-ACC execute-thing-TOPIC plausible

*mea\_3.* X ga [Act] (to iu) koto wo jitsugensuru sono shudan toshite  
 X-NOM [Act] (that) thing-ACC realize its by means of  
 X ga [Act] (to iu) no ha mottomo de aru.  
 X-NOM [Act] thing-TOPIC plausible

*mea\_4.* X ga [Act] (to iu) koto wo suru koto niyotte  
 X-NOM [Act] (that) thing-ACC execute-as a result of  
 X ga [Act] (to iu) koto wo suru.  
 X-NOM [Act] (that) thing-ACC execute

*mea\_5.* X ga [Act] (to iu) koto wo suru koto niyotte  
 X-NOM [Act] (that) thing-ACC execute-as a result of  
 X ga [Act] (to iu) no de aru.  
 X-NOM [Act] thing-COPULA

*mea\_6.* X ga [Act] (to iu) koto wo suru koto niyotte  
 X-NOM [Act] (that) thing-ACC execute-as a result of  
 X ga [Act] (to iu) koto ga dekiru.  
 X-NOM [Act] (that) thing-ACC can execute

*mea\_7.* X ga [Act] (to iu) koto no ikkan toshite  
 X-NOM [Act] (that) thing-GEN as part of  
 X ga [Act] (to iu) koto wo suru no ha mottomo de aru.  
 X-NOM [Act] (that) thing-ACC execute plausible

#### ACKNOWLEDGMENTS

We would like to express our special thanks to the creators of Nihongo-Goi-Taikei and several of the dictionaries used in the ALT-J/E translation system at NTT Communication Science Laboratories, and the EDR electronic dictionaries produced by Japan Electronic Dictionary Research Institute. We would also like to thank Nihon Keizai Shimbun, Inc. for allowing us to use their newspaper articles. We are grateful to the reviewers for their suggestive comments, Taku Kudo for providing us with his dependency analyzer and SVM tools, and Eric Nichols and Campbell Hore for proofreading.

## REFERENCES

- ALLEN, J. F. 1983. Recognizing intentions from natural language utterances. In M. Brady and R.C. Berwick (Eds.), *Computational models of discourse*. MIT Press, Cambridge, MA.
- ALLEN, J. F. 1995. *Natural Language Understanding*. Benjamin/Cumming, New York.
- ALTENBERG, B. 1984. Causal linking in spoken and written English. *Studia Linguistica* 38, 1.
- BRITANNICA. 1998. Britannica CD98 multimedia edition.
- CARBERRY, S. 1990. *Plan Recognition in Natural Language Dialogue*. MIT Press, Cambridge, MA.
- COHEN, J. 1960. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement* 20, 37–46.
- FELLBAUM, C. 1998. *WordNet: An Electronic Lexical Database*. MIT Press, Cambridge, MA.
- GARCIA, D. 1997. COATIS, an NLP system to locate expressions of actions connected by causality links. In *Proc. of The 10th European Knowledge Acquisition Workshop*. 347–352.
- GIRJU, R. AND MOLDOVAN, D. 2002. Mining answers for causation questions. In *Proc. The AAAI Spring Symposium on Mining Answers from Texts and Knowledge Bases*.
- HARABAGIU, S. M. AND MOLDOVAN, D. I. 1997. Textnet—a text-based intelligent system. *Natural Language Engineering* 3, 171–190.
- HECKERMAN, D., MEEK, C., AND COOPER, G. 1997. A Bayesian approach to causal discovery. Tech. rep., Microsoft Research Advanced Technology Division, Microsoft Corporation, Technical Report MSR-TR-97-05.
- HOBBS, J. R. 1979. Coherence and co-reference. *Cognitive Science* 1, 67–82.
- HOBBS, J. R. 1985. On the coherence and structure of discourse. Tech. rep., Technical Report CSLI-85-37, Center for The Study of Language and Information.
- HOBBS, J. R., STICKEL, M., APPEL, D., AND MARTIN, P. 1993. Interpretation as abduction. *Artificial Intelligence* 63, 69–142.
- ICHIKAWA, T. 1978. *Introduction to tyle theory for Japanese education*. Education (in Japan).
- IKEHARA, S., MIYAZAKI, M., SHIRAI, S., YOKOO, A., NAKAIWA, H., OGURA, K., OYAMA, Y., AND HAYASHI, Y. 1997. *Goi-Taikei—A Japanese Lexicon*, Iwanami Shoten.
- IKEHARA, S., SHIRAI, S., YOKOO, A., AND NAKAIWA, H. 1991. Toward an MT system without pre-editing—effects of new methods in ALT-J/E-. In *Proc. of the Third Machine Translation Summit: MT Summit III*, Washington DC. 101–106.
- IWANSKA, L. M. AND SHAPIRO, S. C. 2000. *Natural Language Processing and Knowledge Representation—Language for Knowledge and Knowledge for Language*. MIT Press, Cambridge, MA.
- JOACHIMS, T. 1998. Text categorization with support vector machines: learning with many relevant features. In *Proceedings of ECML-98, 10th European Conference on Machine Learning*, C. Nédellec and C. Rouveirol, Eds. Number 1398. Springer Verlag, New York. 137–142.
- JONSSON, K. 2000. Robust correlation and support vector machines for face identification. Ph.D. thesis, University of Surrey.
- KHO, C. S. G., CHAN, S., AND NIU, Y. 2000. Extracting causal knowledge from a medical database using graphical patterns. In *Proc. of The 38th. Annual Meeting of The Association for Computational Linguistics (ACL2000)*. 336–343.
- KRIPPENDORF, K. 1980. *Content analysis: An introduction to its methodology*. Sage, Thousand Oaks, CA.
- KUDO, T. AND MATSUMOTO, Y. 2003. Japanese dependency analysis using cascaded chunking. In *Proc. of The 6th. Conference on Natural Language Learning (CoNLL)*.
- LENAT, D. 1995. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM* 38, 11.
- LITMAN, D. J. AND ALLEN, J. F. 1987. A plan recognition model for subdialogues in conversations. *Cognitive Science* 11, 163–200.
- LIU, H., LIEBERMAN, H., AND SELKER, T. 2003. A model of textual affect sensing using real-world knowledge. In *Proc. of The International Conference on Intelligent User Interfaces*. 125–132.
- LOW, B. T., CHAN, K., CHOI, L. L., CHIN, M. Y., AND LAY, S. L. 2001. Semantic expectation-based causation knowledge extraction: A study on Hong Kong stock movement analysis. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*. 114–123.
- MANN, W. C. AND THOMPSON, S. A. 1987. Rhetorical structure theory: A theory of text organization. In *USC Information Sciences Institute, Technical Report ISI/RS-87-190*.

- MARCU, D. 1997. The rhetorical parsing of natural language texts. In *Proc. of ACL97/EACL97*. 96–103.
- MARCU, D. 2002. An unsupervised approach to recognizing discourse relations. In *Proc. of The 40th. Annual Meeting of The Association for Computational Linguistics (ACL2002)*. 368–375.
- MASUOKA, T. 1997. *Complex sentence*. Kuroshio (in Japan).
- MASUOKA, T. AND TAKUBO, Y. 1992. *Fundamental Japanese grammar(revised version)*. Kuroshio. (in Japan).
- MATSUMOTO, Y., KITAUCHI, A., YAMASHITA, T., HIRANO, Y., MATSUDA, H., AND ASAHARA, M. 1999. *Japanese Morphological Analyzer ChaSen Users Manual version 2.0*. Technical Report NAIST-IS-TR990123, Nara Institute of Science and Technology Technical Report.
- MEDLINE. 2001. The MEDLINE database. See also, <http://www.ncbi.nlm.nih.gov/PubMed/>.
- NAGANO, M. 1986. *A reveiw of style theory*. Asakura (in Japan).
- NAKAIWA, H. AND IKEHARA, S. 1995. Intrasentential resolution of Japanese zero pronouns in a machine translation system using semantic and pragmatic constraints. In *Proc. of The 6th TMI*. 96–105.
- NIKKEI. 1990. Nihon Keizai Shimbun CD-ROM version.
- PEARL, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.
- PEARL, J. 2000. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, London.
- PEI, J., HAN, J., MORTAZAVI-ASI, B., PINTO, H., CHEN, Q., DAYAL, U., AND HSU, M. C. 2001. Prefix Span: mining sequential patterns efficiently by prefix projected pattern growth. In *Proc. of 1st. Conference of Data Engineering (ICDE2001)*. 215–226.
- RIFKIN, R. AND KLAUTAU, A. 2004. In defense of one-vs-all classification. *Journal of Machine Learning Research* 5, 101–141.
- RWC. 1998. RWC text corpus 2nd edition, Iwanami Japanese dictionary tagged/morphological data, 5th edn.
- SATOU, H., KASAHARA, K., AND MATSUZAWA, K. 1999. Retrieval [sic] of simplified causal knowledge in text and its application. In *Technical report of IEICE, Thought and Language*. (in Japan).
- SATOU, S. AND NAGAO, M. 1990. Toward memory-based translation. In *Proc. of The 13th. International Conference on Computational Linguistics (COLING90)*. 247–252.
- SCHANK, R. AND ABELSON, R. 1977. *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Lawrence Erlbaum Assoc., Mahwah, NJ.
- SCHANK, R. AND RIESBECK, C. 1981. *Inside Computer Understanding: Five Programs Plus Miniatures*. Lawrence Erlbaum Assoc., Mahwah, NJ.
- SINGH, P., LIN, T., MUELLER, E. T., LIM, G., PERKINS, T., AND ZHU, W. L. 2002. Open Mind Common Sense: Knowledge acquisition from the general public. In *Proc. of The 1st. International Conference on Ontologies, Databases and Applications of Semantics for Large Scale Information Systems*.
- STORK, D. G. 1999. Character and document research in the Open Mind Initiative. In *Proc. of International Conference on Document Analysis and Recognition*. 1–12.
- SUYAMA, A. 2005. Sentence identification in HTML documents with machine learning. M.S. thesis, Tokyo Institute of Technology (in Japan).
- TERADA, A. 2003. A study of text mining techniques using natural language processing. Ph.D. thesis, Tokyo Institute of Technology (in Japan).
- TORISAWA, K. 2003. Automatic extraction of ‘commonsense’ inference rules from corpora. In *Proc. of The 9th Annual Meeting of The Association for Natural Language Processing*. 318–321 (in Japan).
- VAPNIK, V. N. 1995. *The Nature of Statistical Learning Theory*. Springer, New York.
- VERT, J. 2002. Support vector machine prediction of signal peptide cleavage site using a new class of kernels for strings. In *Proceedings of the Pacific Symposium on Biocomputing 2002*. 649–660.
- VOORHEES, E. M. AND HARMAN, D. K. 2001. The Ninth Text Retrieval Conference (TREC9). See also, <http://trec.nist.gov>.
- YOKOI, T. 1995. The EDR electronic dictionary. *Communications of the ACM* 38, 11, 42–44.

Received April 2005; revised July 2005; accepted August 2005