NAIST-IS-DT0161005

Doctor's Thesis

Acquiring Causal Knowledge from Text Using Connective Markers

Takashi INUI

February 6, 2004

Department of Information Processing
Graduate School of Information Science
Nara Institute of Science and Technology

Doctor's Thesis
submitted to Graduate School of Information Science,
Nara Institute of Science and Technology
in partial fulfillment of the requirements for the degree of
DOCTOR of ENGINEERING

Takashi INUI

Thesis Committee: Yuji MATSUMOTO, Professor
Shunsuke UEMURA, Professor
Katsumasa WATANABE, Professor
Kentaro INUI, Associate Professor

Acquiring Causal Knowledge from Text Using Connective Markers*

Takashi INUI

Abstract

One of the bottlenecks in developing natural language understanding systems is the prohibitively high cost of building and managing a comprehensive common-sense knowledge base. In this thesis, 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 of relations based on a variety of points of view. In this work, we consider 4 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 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.

In this work, focusing our attention on Japanese complex sentences including the word ため（because）, we consider the following topics:（1）What kinds and how much causal knowledge is present in the document collection, （2）How accurately can relation instances be identified, and （3）How can acquired causal knowledge be made available to applications.

First, we investigated the distribution of causal relation instances in Japanese newspaper articles. The main part of this investigation was conducted based on

*Doctor's Thesis, Department of Information Processing, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DT0161005, February 6, 2004.

human judgments using linguistic tests. Using approximately 1000 samples containing the word ため, we confirmed that it is possible to acquire causal relation instances from about 90% of samples, and if the volitionality of the subordinate and matrix clauses can be determined, samples can be classified into the four relation types — *cause*, *effect*, *precond* and *means* — with a precision of 85% or more.

Second, we attempted an experiment with the aim of assessing how accurately we can automatically acquire causal relation instances. By using machine leaning 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 one year of Japanese newspaper articles.

Third, we attempted to utilize causal knowledge for acquiring desirability lexical knowledge. From this investigation, it is clear that causal relation instances, at least instances of *cause* relations and *means* relations, are useful for acquiring desirability lexical knowledge.

Keywords:

causal relation, knowledge acquisition, connective marker, volitionality

接続標識に基づく文書集合からの因果関係知識獲得*

乾 孝司

内容梗概

人間のような深い言語理解能力を工学的に実現することを妨げている問題のひとつとして，計算機で利用可能な大量の常識的知識をいかに構築するかという問題がある．本論文では，常識的知識のうち，人工知能の分野をはじめ，多くの学術領域において関心が向けられている因果関係に関する知識に着目し，大規模な電子化文書集合から因果関係知識を自動的に獲得する方法について論じる．

因果関係とは，何らかの依存関係がある出来事間の関係を指し，例えば,「大雨が降ると洪水が発生する」や「列車に乗るので切符を買う」などがある．従来より，因果関係は様々な観点に従った幾つかの種類に分類されてきた．本論文では，談話理解研究で設計された Allen のプランオペレータに関わる４種類の因果関係（“原因”，“効果”，“前提条件” および “手段”）を取り扱う．

知識獲得の基本的アプローチは，接続助詞（「ため」や「ので」）などに代表される接続標識を手掛かりとすることにある．ただし，どの接続標識を含む言語表現であっても，常に必然性の高い因果関係を表現する保証はない．また，同一の接続標識を含む言語表現から，上述した因果関係のうち，２種類以上の因果関係知識が獲得できる可能性がある．そのため，接続標識以外の情報を利用することで獲得可能な因果関係を同定し，適切な因果関係知識を抽出する枠組みを検討する必要がある．

本論文では特に接続標識「ため」を含む複文に注目し,（１）対象（「ため」を含む複文）中に獲得すべき因果関係はどの程度現れるか,（２）対象中に現れた獲得すべき因果関係をどの程度の精度で同定でき，どの程度の量の知識が獲得できるか,（３）獲得された因果関係知識がどのような応用に貢献できるか，について議論している．

*奈良先端科学技術大学院大学 情報科学研究科 情報処理学専攻 博士論文, NAIST-IS-DT0161005, 2004 年 2 月 6 日.

まず,（１）では，言語テストを用いることにより,「ため」を含む複文から獲得すべき因果関係知識の比率を調査した．新聞記事から得た約 1000 件の「ため」を含む複文を対象とし，おおよそ９割の事例が獲得すべき因果関係を含んでいることを確認した.（２）では,「ため」を含む複文から４種類の因果関係知識の自動獲得課題を取り扱った．機械学習手法に基づく因果関係同定器を構築し，約 1000 件の未知データを用いたテストの結果，“原因結果”，“前提条件”，“手段” の各関係について，80 %の再現率で 95 %以上の獲得精度を達成した．また，“効果” 関係については，30 %の再現率で 90 %の獲得精度を達成した．新聞記事一年分の文書集合に本手法を適用した場合，27,000 件を超える因果関係知識が獲得できる見積もりを得た.（３）では，因果関係知識を利用して，出来事の望ましさに関する語彙的知識をブートストラップ的に獲得する手法を提案し，因果関係知識の有用性を示した.

キーワード

因果関係, 知識獲得, 接続標識，意志性

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Motivation

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 (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 [69, 23]. For example, knowledge about the preconditions and effects of actions is commonly used for discourse understanding based on plan recognition. Figure 1.1 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 [41] and OpenMind [74] 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 [16, 34, 17, 67]. 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

dry-laundry-in-the-sun($actor, $laundry)	
precondition:	weather(sunny)
effect:	get-dry($laundry)
decomposition:	hang($actor, $laundry)

Figure 1.1 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⟩)

Figure 1.2 An example of causal relation instances

kinds and how much causal knowledge they will be able to extract, how accurate the process can be made, and how useful acquired knowledge will be for language understanding. Motivated by this background, in this thesis we describe our approach to automatic acquisition of causal knowledge from a document collection.

1.2 Goal

We aim to acquire causal knowledge such as those in Figure 1.2 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 1.2 can be seen as constituent elements of the plan operator in Figure 1.1. The relations in Figure 1.2 therefore represent a decomposition of the plan operator in Figure 1.1.

We use resultative connective markers such as "because", and "but", and "if" as linguistic cues to acquire causal knowledge. For example, given the following sentences (1) , we may be able to acquire the causal knowledge given in Figure 1.2,

(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

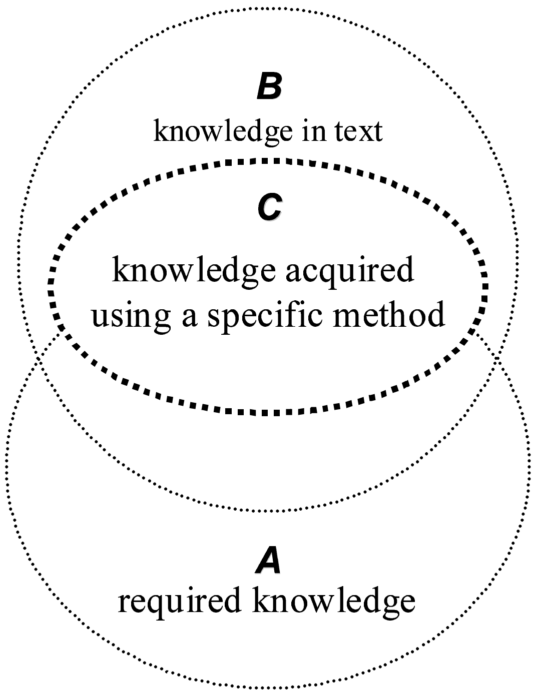


Figure 1.3 Sets of knowledge

not novel in itself. In this thesis however, we address the following subset of the unexplored issues, focusing on knowledge acquisition from Japanese texts:

1. What kinds and how much causal knowledge is present in the document collection?

2. How accurately can relation instances be identified?

3. How many relation instances can be acquired from currently available document collections?

4. How can acquired causal knowledge be made available to applications?

With the improvement of several natural language processing (NLP) techniques, involving a variety of disciplines including artificial intelligence, data mining, and computational linguistics, the new research field of knowledge acquisition from text has emerged. Suppose here that $\mathcal{A}$ in Figure 1.3 indicates the set of knowledge required to realize natural language understanding (NLU) systems, $\mathcal{B}$ indicates the set of knowledge potentially available for acquisition from all available text resources, and $\mathcal{C}$, which is a subset of $\mathcal{B}$, indicates the set of knowledge acquired from a document collection using a specific acquisition method. Although it is generally accepted among researchers that an enormous amount of knowledge is required to realize an NLU system, it is hard to answer the question of what kinds and how much knowledge are contained in $\mathcal{A}$. As a result of improvements in the information infrastructure, the volume of electronic text available is currently increasing exponentially. However, it is unclear whether the scale of $\mathcal{B}$ will ever be the same as that of $\mathcal{A}$, and how much overlap there is between them. The same applies to B & C: no one knows, given a document collection and a developed method, how much knowledge will be contained in $\mathcal{C}$?

In this work, focusing our attention on set $\mathcal{C}$, we attempt an investigation of the above mentioned four points.

1.3 Outline of the thesis

The thesis is organized as follows.

In Chapter 2, we describe the causal knowledge that we aim to acquire; the typology of causal relations we use; and 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 thesis.

In Chapter 3, we outline several previous research efforts on knowledge acquisition, in particular, attempts to develop methods for extracting causal knowledge from document collections.

NLP techniques enable us to utilize a text as a knowledge resource. In Chapter 4, we propose a parsing method which explores two directions: probabilistic partial parsing and committee-based decision making. This decision-making scheme enables a fine-grained arbitrary choice on the trade-off between accuracy and coverage. Such trade-off is important since there are various applications that require reasonably high accuracy even sacrificing coverage.

In Chapter 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.

Chapter 6 and Chapter 7 describe methods for automatically acquiring causal knowledge from text using a machine learning approach. In Chapter 6, we deal with volitionality estimation. Volitionality information is a useful feature for classifying samples in accordance with our typology of causal relations. Chapter 7 describes our method of automatic causal relation identification and experimental results using it.

In Chapter 8, we describe an application using causal knowledge acquired. We attempt to utilize causal knowledge in acquiring lexical knowledge for event desirability analysis.

Chapter 9 concludes the thesis and outlines future work.

Chapter 2

Causal knowledge acquisition from text

In this chapter, we describe the causal knowledge that we aim to acquire, and the issues involved in acquiring knowledge from text. In Section 2.1, we describe a typology of causal relations we deal with. In Section 2.2, we describe our approach based on cue phrases in texts; an important characterization of causal relation instances as knowledge in comparison with rhetorical expressions; and issues specifically discussed in this thesis.

2.1 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 et al. [46], Nagano [55] and Ichikawa [24]) and AI researchers (e.g. Allen [42, 2], Schank [70, 69] and Hobbs [21, 22]).

In this thesis, we focus mainly on a set of causal relations related to the Allen's plan operator [42, 2] which was proposed in the research field of discourse understanding. Figure 1.1 is a simple example of the plan operator. The reasons for the adoption of this scheme are:

- While some relations proposed by previous researchers are unclear as to

Table 2.1 Typology of causal relations

Causal relations	Meaning	Examples of templates used in linguistic tests
cause(SOA_1, SOA_2)	SOA_1 causes SOA_2	SOA_2 usually happens as a result of the fact that SOA_1 happens.
effect(Act_1, SOA_2)	SOA_2 is the effect of Act_1	SOA_2 usually happens as a result of the execution of Act_1.
precond(SOA_1, Act_2)	SOA_1 is a precondition of Act_2	Act_2 cannot be done unless SOA_1 holds/happens. If SOA_1 often holds/happens, one will often execute Act_2.
means(Act_1, Act_2) (same agents)	Act_1 is a means of executing Act_2	Someone executes often Act_1 in order to execute Act_2. If someone executes Act_1, then she can often execute Act_2.

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.

- Some previous works (e.g. [22]) 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 [1].

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 1.1, 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 role, *precondition*, *effect* and *decomposition*.

Based on the distinction between these relations, we have created a typology of causal relations as summarized in Table 2.1. 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 thesis 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 referred to 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 2.1 are a necessary condition for each causal relation class. In the table, Act_i denotes a volitional action and SOA_i denotes a non-volitional 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 non-volitional 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 non-volitional SOA has been remarked on already (e.g. [66]).

The *effect* relation in Table 2.1 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 sub-action 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 relate to the plan operator than other three relations, it often indicates typical causal relations such as "heavy rain causes flooding". Examples of each type of causal relation instances are demonstrated in Appendix A.

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 we use in linguistic tests are shown in Table 2.1. The details of linguistic tests are described in Section 5.4.1.

2.2 Approach and problem

2.2.1 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 <u>because</u> it was sunny.

b. The laundry dried well, <u>though</u> it was not sunny.

c. <u>If</u> it was sunny, the laundry could dry well.

d. The laundry dried well <u>because of</u> the sunny weather.

→ e. *cause*(⟨it is sunny⟩, ⟨laundry dries well⟩)

(3) a. Mary used a tumble dryer <u>because</u> she had to dry the laundry quickly.

b. Mary could have dried the laundry quickly <u>if</u> she had used a tumble dryer.

c. Mary used a tumble dryer <u>to</u> dry the laundry quickly.

d. Mary could have dried the laundry more quickly <u>with</u> a tumble dryer.

→ e. *means*(⟨using a tumble dryer⟩, ⟨drying laundry quickly⟩)

First, causal knowledge 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 minor 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.

The above discussion of English applies equally to Japanese. One could acquire the same causal relation instances from sentences with connective markers such as ため(because, in order to), が(but), and れば(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. 晴れ-てい-た-<u>ため</u> 洗濯物-が よく 乾い-た。
sunny-ing-PAST-<u>because</u> laundry-NOM well dry-PAST

b. 晴れ-てい-ない-<u>が</u> 洗濯物-は よく 乾い-た。
sunny-ing-not-<u>but</u> laundry-TOPIC well dry-PAST

c. 晴れ-てい-<u>れば</u> 洗濯物-が よく 乾い-た-のに。
sunny-ing-<u>if</u> laundry-NOM well dry-PAST-should

d. 晴天-<u>で</u> 洗濯物-が よく 乾い-た。
sunny-<u>because of</u> laundry-NOM well dry-PAST

→ e. *cause*(⟨ 晴れる ⟩, ⟨ 洗濯物がよく乾く ⟩)

(5) a. 洗濯物-を はやく 乾かす-<u>ため</u> 乾燥機-を 使っ-た。
laundry-NOM quickly dry-<u>because</u> dryer-nom use-PAST

b. 乾燥機-を 使っ-てい–<u>れば</u>
dryer-NOM use-ing-<u>if</u>
はやく 洗濯物-が 乾かせ-た-はずだ。
quickly laundry-nom dry-PAST-should

c. 乾燥機-を 使っ-た-の-は
dryer-NOM use-PAST-that-TOPIC
はやく 洗濯物-を 乾かし-たかった-<u>から</u>-だ。
quickly laundry-NOM want to dry-PAST-<u>because</u>

d. 乾燥機-<u>で</u> 洗濯物-を はやく 乾かし-た。
<u>with</u> dryer laundry-NOM quickly dry-PAST

→ e. *means*（⟨ 乾燥機を使う ⟩, ⟨ 洗濯物をはやく乾かす ⟩）

Thus, though 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 thesis.

2.2.2 Rhetorical and causal relations

The sentences exemplified in (6) represent a relationship between the volitional actions みかんの皮を剥く（peeling an orange） and みかんを食べる（eating an orange） expressed using various types of rhetorical expressions. Looking at these sentences, we are able to recognize different rhetorical relations among (6a_j) - (6c_j) , and (6a_e) - (6c_e) . In general, the sentences (6a_j) and (6a_e) can be interpreted as a PURPOSE rhetorical relation, (6b_j) and (6b_e) as a CONDITION rhetorical relation, and (6c_j) and (6c_e) as a CONTRAST rhetorical relation.

(6) a_j. みかんを食べる ためにu 皮を剥いた。

b_j. みかんは皮を剥かない と 食べられない。

c_j. みかんの皮を剥いた のに 食べることができなかった。

a_e. I peeled an orange to eat it.

b_e. If you do not peel an orange, you cannot eat it.

c_e. Though I peeled an orange, I could not eat it.

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（⟨ みかんの皮を剥く ⟩, ⟨ みかんを食べる ⟩）
peeling an orange　eating an orange

【meaning】

A volitional action みかんの皮を剥く（peeling an orange） is a means of executing another volitional action みかんを食べる（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.

(8) a_j. * みかんの皮を剥いた ために みかんを食べることができなかった。

b_j. * みかんの皮を剥いた のに 雨が降っている。

a_e. * I was not able to eat an orange to peel it.

b_e. * Though I peeled an orange, it rains.

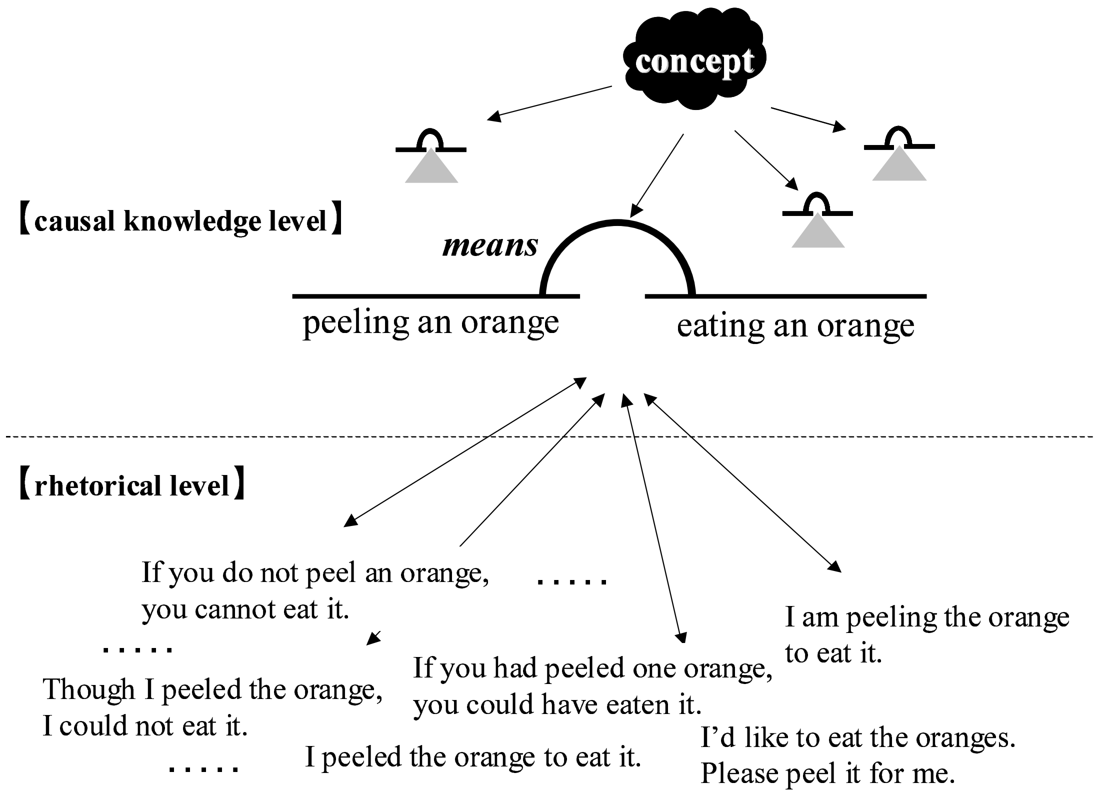


Figure 2.1 Rhetorical level and causal knowledge level

We aim to acquire causal knowledge like (7) which can work as the basis for recognition of rhetorical expressions such as in (6) as being coherent.

See Figure 2.1. We assume that the relationships between event instances indicated in a text are located at the rhetorical level. And we assume that when the event instances are abstracted from the rhetorical level to a higher abstract level which we call the causal knowledge level in which the abstracted classes are dealt with, 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:

- **Tense and aspect:** whether the event represented has already occurred or not.

  Ex. （after abstraction）
  みかんの皮を剥いた → みかんの皮を剥く
  I peeled the orange → I peel the orange

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

  Ex. （after abstraction）
  みかん-が 食べ-られ-た → みかん-を 食べる
  orange-NOM eat-PASSIVE-PAST → orange-ACC eat

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. Though some linguists take a different view, we deal with tense and aspect under the rubric of modality information in this work.

2.2.3 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[1] such as Figure 1.2, (2e) and (3e) instead of by any formal semantic representation language for the following reasons.

[1] In fact, our data is semi-structured in that we treat it as a surface case frame. For convenience, in our examples we omit this structure and express causal relation instances using plain natural language expressions. Note that it is important here that our causal relation instances are not abstracted as propositional information.

- 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 proven difficult to design a formal language that can fully represent the diverse meanings of natural language expressions.
- As discussed in [31], there has been a shift toward viewing natural language as the best means for knowledge representation.
- As discussed in detail in Section 7.3, 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 Sato's work [68]. We believe that acquired knowledge is sufficient for practical use without any abstraction process.

In fact, for example, Harabagiu et al. [18] 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 [73] is also represented by English sentences and Liu et al. [44] reported that Singh's database could be successfully used for textual affect sensing.

2.2.4 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 2.2.

> Given text segments, the process of acquiring causal knowledge forms two independent phases: proposition extraction and causal relation identification.
>
> **Proposition extraction:** Removing the modality expressions and extracting the propositional expressions from a text segment, normally a sentence. For example in Figure 2.2, two propositional expressions indicating different events 晴れる（it is sunny）

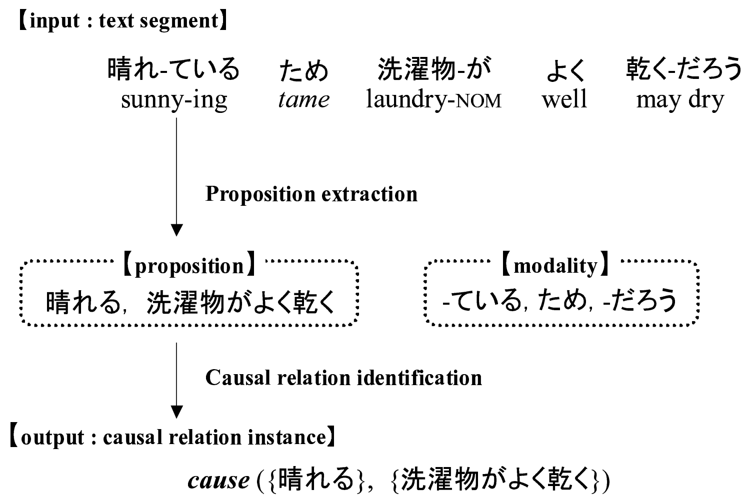


Figure 2.2 Knowledge acquisition workflow

and 洗濯物がよく乾く（laundry dries well） are extracted from the input sentence 晴れているため洗濯物がよく乾くだろう（the laundry may dry well because it is sunny）.

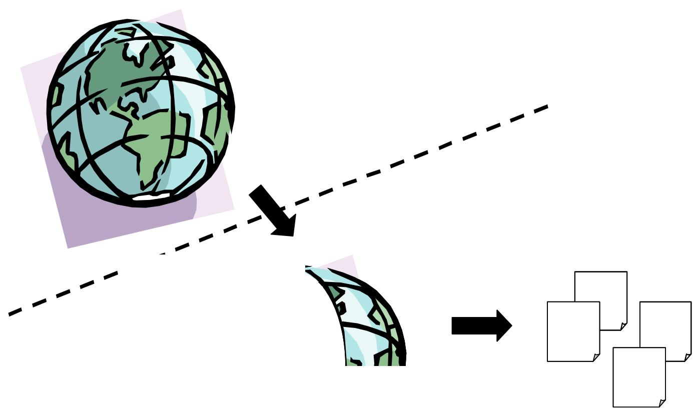
Causal relation identification: Identifying causal relations held between extracted proposition pairs. In Figure 2.2, the *cause* relation is identified as holding between the two propositional expressions 晴れる and 洗濯物がよく乾く.

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

2.3 Limited scalability

Dealing with text documents as sources of knowledge, we cannot neglect the fact that *only a fraction of all the events in the world are written in text* （Figure 2.3）.

For example, Lenat, one of the world's leading computer scientists, and head of the Cyc project, said in his paper [41] in 1995.



Only a fraction of all the events in the world are written in text

Figure 2.3 Limited Scalability

> ..... For example:
>
> - You have to be awake to eat.
> - You can usually see people's noses, but not their hearts.
>
> Such assertions are unlikely to be published in textbooks, dictionaries, magazines, or encyclopedias, even those designed for children. .....

In other words, an approach which acquires knowledge from text resources is restricted in the knowledge that can potentially be acquired. In this section, we describe our stance on this issue from the point of view of scalability.

Scalability

It is still unclear how much knowledge is needed to achieve natural language understanding（NLU） and applications using NLU techniques. However it is commonly accepted among the AI researchers that an enormous amount of knowledge is required. For example, Minsky, a leading AI researcher is quoted [43] as saying that:

> ..... somewhere on the order of 15 million pieces of knowledge may be needed in order to be comparable to what humans possess. .....

Language in the form of text is one of the most fundamental information media. Currently, the volume of electronic text is increasing exponentially, and this growth is expected to continue. We believe that dealing with text as a new source of knowledge presents a very attractive challenge. Although it is important to keep in mind the limitations of these resources as noted above, we believe that the study of acquiring knowledge from text has special meaning in that it should help clear up the issue of what kinds and what amount of knowledge can be acquired from currently available document collections.

2.4 Natural language analysis for knowledge acquisition

We begin by applying the fundamental natural language processing（NLP）techniques of morphological analysis and dependency structure analysis in order to enable us to utilize text as a knowledge resource.

Since the cue phrases we use consist of words, usually a single word, we have to identify each word correctly. For example, the Japanese conjunctive particle ので（because） is one of the expressions used to signal a causal relation between events. If we identify the cue string ので in a text document without using morphological information extracted by a morphological analyzer, we may extract not only the target word in (9a) but also the non-word in (9b) by mistake.

(9) a. 晴れ-ている-<u>ので</u> 庭-に 干す
sunny-ing-because at garden dry

b. 庭-で 使う-も <u>の-で</u> 昔-から ある
at garden use-thing since early times exist

In this work we used *ChaSen*[2] [51] a well-known, highly accurate morphological analyzer.

Next, the causal knowledge we acquire has two arguments which refer to events. The events are expressed using a variety of linguistic forms; words, phrases, clauses, sentences and inter-sentential units. Among these forms, in

[2]Available from `http://chasen.aist-nara.ac.jp`.

this thesis we focus our attention on clauses in trying to capture events (the reasons for this decision is described in Section 5.2). Consequently, we need to analyze the dependency structure to identify the clauses in a sentence.

In the 1990's, there were a number of attempts to use statistical techniques to improve parsing performance. While this goal has been achieved to a certain degree due to the increasing availability of large tree banks, these would seem to be little room for significant further improvements using statistical techniques alone. We explored two directions for the statistical parsing: probabilistic partial parsing and committee-based decision making [30]. This decision-making scheme enables a fine-grained arbitrary choice on the trade-off between accuracy and coverage. Such trade-off is important since there are various applications that require reasonably high accuracy even sacrificing coverage. The details of the method is described in Chapter 4.

While we did not apply language analysis techniques at the semantic and contextual level for various reasons in the current knowledge acquisition model, it seems likely that anaphora resolution and named entity extraction would be needed to improve the performance of knowledge acquisition. Indeed, these two tasks have been studied extensively and levels of performance have increased substantially (e.g. Iida's study [25] on anaphora resolution and Yamada's study [82] on named entity extraction). Introducing anaphora resolution and named entity extraction will almost certainly increase the accuracy of knowledge acquisition.

Chapter 3

Related work

In this chapter, we outline several previous research efforts on knowledge acquisition[1]. First, we describe two well-known projects in Section 3.1. Their goal is to create a large-scale common-sense knowledge database. In Section 3.2, we describe electronic dictionaries composed of causal relation information. We review in Section 3.3 attempts to develop methods for extracting causal knowledge from document collections.

3.1 Projects to create commonsense knowledge databases

Cyc

The aim of Cyc project[2] [41] is to create the world's first true artificial intelligence with both a commonsense database and the ability reason about that knowledge. Cyc consists of three main components:

- A knowledge base（Cyc KB）

[1]Probabilistic modeling of causal relations [60, 61, 19] forms one sub-field of the causality research activity. However, since we focus in this thesis on knowledge acquisition rather than modeling, we do not describe this work.

[2]`http://www.cyc.com/cyc/`

- An environment: the interface editing/browsing tools, the multi-user knowledge server, etc.

- A representation language (CycL)

Work began on Cyc KB in 1984. The Cyc KB is built upon a core of over 1,500,000 hand-entered assertions (or rules) designed to capture a large portion of what we normally consider consensus knowledge about the world.

Cyc is now a working technology with applications to many real-world business problems such as improved speech recognition and semantic data mining.

However, it should be noted that:

- Cyc KB has to be handcrafted by engineers familiar with CycL

- So far it has taken over 15 years, at the cost of several tens of millions of dollars.

OMCS

The advent of the web has made it possible for thousands of people to collaborate to construct systems that no single individual or team could build alone. The Open Mind Initiative [74] was formed with the goal of studying whether a relatively small investment in a good collaborative tool for knowledge acquisition could support the distributed construction of a commonsense database by many people. As part of this initiative, the Open Mind Common Sense (OMCS) project[3] [72, 73] was born.

The goal of OMCS is to teach computers the myriad things which we all know and which underlie our general intelligence but which we usually take for granted. The project has moved since on to the next-generation version of the system, OMCS-2. We will briefly describe the first version of the system, OMCS-1.

OMCS-1 is a commonsense knowledge acquisition system targeted at the general public. It is a web site that gathers facts, rules, stories, and descriptions using a variety of simple elicitation activities. OMCS-1 has been running on the

[3] `http://commonsense.media.mit.edu/`

web since September 2000. As of August 2002, They have gathered about 456,000 pieces of commonsense knowledge from 9296 people.

They constructed a variety of elicitation activities. Each activity tries to make it simple for users to teach Open Mind a certain kind of knowledge. For example, one activity is an "Explain why" activity. In this activity, the system presents an item such as the following to the user:

Tigers are dangerous animals because [slot].

Users then answer in natural language via the web, for example:

Tigers are dangerous animals because [*tigers are large and carnivorous*].

A manual evaluation was performed on the OMCS-1 database to assess its quality and composition. About 3,000 items, which represents about 1% of all items in OCMS database, were distributed among 7 judges and were rated on a scale of 1 to 5 for the following attributes: generality (1=specific fact, 5=general truth), truth (1=false, 5=true), neutrality (1=biased, 5=neutral), and sense (1=makes no sense, 5=makes complete sense). The average rating was 3.26 for generality, 4.28 for truth, 4.42 for neutrality and 4.55 for sense. The following are sample items:

(10) a. *Birds often make nests out of grass.* (rated 5 for generality)

b. *Dew is wet* (rated 5 for generality)

c. *houses have many kinds of roofs* (rated 5 for truth)

d. *Eritrea is part of Africa* (rated 1 for generality)

e. *Men should do the laundry* (rated 1 for neutrality)

f. *cows can low quietly* (rated 1 for sense)

More work is needed to further increase the quality of entries, but the general approach is an appealing one which if successful, promises to deliver enormous benefits.

```
...   3ce68f  cause  3bde83 ...
                      [in the concept description dictionary]
```

```
...   3ce68f -> decease 死ぬ [シ・ヌ] ...
...   3bde83 -> bomb    爆撃する [バクゲキ・スル] ...
                      [in the the head-concept dictionary]
```

Figure 3.1 An example entry in the EDR dictionary

3.2 Enriched machine-readable dictionary

Several machine-readable dictionaries have been developed with the aim of achieving more intelligent information processing. The EDR dictionary and WordNet describe various kinds of information including words and concepts. One aspect of which is information about causal relations. These resources are used in frequently NLP research.

We will outline the two dictionaries and show examples of entries about causal relations.

EDR

The EDR Electronic Dictionary[4] [83] is a well-known large-scale machine-readable dictionary, which was developed for advanced processing of natural language by computers and is aimed at establishing an infrastructure for knowledge information processing.

The EDR dictionary is composed of eleven sub-dictionaries. One of the sub-dictionaries is a concept dictionary and consists of thesaurus-like concept classifications. The concept dictionary is composed of a number of sub-dictionaries. One of these sub-dictionaries describes relations, other than the super-sub relation, that exist between two concepts. There are eight types of concept relations. These relations are the various case relations in which a verbal concept governs a nominal concept.

[4]`http://www2.crl.go.jp/kk/e416/EDR/`

object *agent* *a-object* *place* *goal* *scene* *implement* *cause*

The *cause* relation is one of the case relations. Nearly 10,000 records including the *cause* relation as the case relation contained in the concept description dictionary. Figure 3.1 is an example of an entry with a causal relation.

WordNet

WordNet[5] [52, 12] is an on-line lexical reference system whose design is inspired by current psycholinguistic theories of human lexical memory. English nouns, verbs, and adjectives are organized into synonym sets, each representing a single underlying lexical concept.

WordNet currently contains approximately 95,600 different English word forms organized into some 70,100 word meanings, or sets of synonyms.

The verbs in WordNet consist of four types of lexical entailment shown in Figure 3.2. In logic, entailment is properly defined for propositions; a proposition P entails a proposition Q if and only if there is no conceivable state of affairs that could make P true and Q false. Entailment is a semantic relation because it involves reference to the states of affairs that P and Q represent. The term is generalized here to refer to the relation between two verbs V_1 and V_2 that holds when the sentence "Someone V_1." logically entails the sentence "Someone V_2."; this use of entailment can be called *lexical entailment*. Thus, for example, "snore" lexically entails "sleep" because the sentence "He is snoring." entails "He is sleeping"; the second sentence necessarily holds if the first one does.

Troponymy A Troponymy relation between two verbs V_1 and V_2 can be expressed by the formula "To V_1 is to V_2 in some particular manner". For example, troponyms of communication verbs often encode the speaker's intention or motivation for communicating, as in "examine", "confess", or "preach", or the medium of communication: "fax", "e-mail", "phone". The verbs categorized Troponymy take part in troponymic relations and one of the verbs always temporally includes the other.

+Troponymy +Troponymy is another troponymic relation. The activities referred to by a troponym and its more general superordinate are always

[5] `http://www.cogsci.princeton.edu/~wn/`

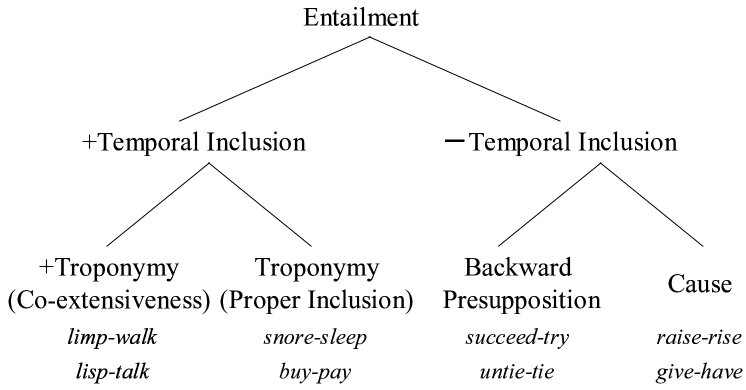


Figure 3.2 Lexical entailment in WordNet

temporally co-extensive. Consider, for example, the pair "limp-walk". The verbs in this pair are related by troponymy: "to limp is also to walk in a certain manner" and so limping and walking are always temporally co-extensive.

Backward Presupposition The activity denoted by the entailed verb V_1 always precedes the activity denoted by the entailing verb V_2 in time. These verbs are not related by temporal inclusion.

Cause If V_1 necessarily causes V_2, then V_1 also entails V_2. Like the backward presupposition relation, the entailment between verbs is characterized by the absence of temporal inclusion.

Information about causal relations is included in both dictionaries. However, relations are defined of the level of words only and only entailment relations are included.

3.3 Causal knowledge acquisition from text

In recent years, there have been several attempts at extracting causal expression patterns from document collections [16, 67, 34, 45, 17, 76, 77]. In this section, we introduce four studies, three of which make use of cue phrases as we do and one of which makes use of a statistical technique.

3.3.1 Cue phrase based approaches

Girju et al. [17]

Girju et al. [17] proposed a method for extracting causal expressions from text in English based on the triplet patterns as follows:

$\langle NP_1 \; clue \; NP_2 \rangle$

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 they collect all triplets matching the pattern above from the gloss of WordNet 1.7. The following is one such example.

(11) $\underline{\textit{extreme leanness}}_{NP_1}$ *usually caused by* $\underline{\textit{starvation or disease}}_{NP_2}$

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 determine automatically all patterns $\langle NP_1 \; verb/verbal\_expression \; NP_2 \rangle$ where NP_1-NP_2 is the pair under consideration. Third, to eliminate the patterns 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 [81] collection of texts which contains 3GB of news articles from Wall Street Journal, Financial Times, Financial Report, etc., they extracted about 1,300 patterns (causal expressions). Using 300 of the 1,300 patterns, the accuracy of their method, as evaluated by human subjects, was about 65%.

Terada [76]

Terada [76] proposed a method for extracting causal expressions similar to that of Girju et al. He used only a small set of cue phrases as follows:

- causative verb[6]
  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. $\underline{\textit{aircraft loss}}_{NP}$

- $NP + PP$ In addition to NP, a prepositional modifier of the NP is considered.
  ex. $\underline{\textit{aircraft loss}}_{NP}$ $\underline{\textit{of oil pressure}}_{PP}$

- N/A In addition to $NP + PP$, full clauses are considered.
  ex. $\underline{\textit{aircraft oil pressure lose}}_{clause}$

He applied a constraint based on the frequency of the patterns using a sequential pattern mining algorithm, PrefixSpan[63] instead of the semantic constraints Girju et al. applied. 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. [34]

Khoo et al. [34] acquired causal knowledge with manually created syntactic patterns specifically for the MEDLINE text database [59]. Figure 3.3 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

[6] While Girju et al. assigned verbs as cues, Terada restricted the word forms and prepositions allowed to follow a verb.

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

Figure 3.3 An example syntactic pattern proposed by Khoo et al.

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 abstract and had precision of about 60%.

The three studies described above share the characteristic of using cue phrases explicitly expressed in text as we do. However their approach is different direction from ours. While they focus 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 type of causal relations.

3.3.2 A statistics-based approach

Torisawa [77] proposed a statistical method for extracting commonsense inference rules from Japanese newspaper articles. In his method the degree of inevitability implicit in the relationship between events is measured by using statistics instead of explicit cue phrases as we do. The following are examples of rules extracted using his method.

(12) a. もし X-が ビール-を 飲む-ならば
if X-NOM beer-ACC drink-then
{ 普通 or しばしば } X-が ビール-に 酔う
usually or often X-NOM beer-DAT get drunk

- $Score(v, n, v_C) =$
  $MAX\{\{Score_v(v, n, v_C)\} \cup \{Score_{arg}(v, n, arg) | \langle arg, p_{arg}, r_{arg}\rangle \in Args_2(v, v_C) \wedge r_{arg} > \theta arg\}\}$
- $Score_v(v, n, v_C) = \sum_{\alpha \in Class}\{P_V(v_C|\alpha)P(\alpha|n)(P(\alpha|\langle v, \mathrm{WO}\rangle) + P_C(\alpha|\langle v, v_C\rangle))\}$
- $P_V(v_C|\alpha) = \sum_{p \in Rel} P(\langle v_C, p\rangle|\alpha)$ where $\alpha \in Class$
- $P_C(\alpha|\langle v, v_C\rangle)) =$
  $\frac{1}{Z_{v,v_C}} \sum_{\langle w, \mathrm{WO}, r\rangle \in Args_1(v, v_C)}\{P(\alpha|w) \cdot r\}$
  where $\alpha \in Class$ and $Z_{v,v_C} = \sum_{\alpha \in Class} \sum_{\langle w, \mathrm{WO}, r\rangle \in Args_1(v, v_C)}\{P(\alpha|w) \cdot r\}$
- $Score_{arg}(v, n, arg) =$
  $\sum_{\alpha \in Class}\{P(\alpha|n)\}\{P(\langle arg, \mathrm{NO}\rangle|\alpha)\}\{P(\alpha|\langle v, wo\rangle) + P_C(\alpha|\langle v, v_C\rangle)\}$

Figure 3.4 Score function

b. もし X-が 服-を 作る-ならば ｛普通 or しばしば｝ 服-が 売れる
   if X-NOM cloth-ACC produce-then usually or often cloth-NOM sell

In Torisawa's algorithm, parallel verb phrase pairs are first extracted from text such as:

(13) a. ビール-を 飲み $_{vp_1}$ 酔っ-た $_{vp_2}$
   beer-ACC drink get drunk

   b. ビール-を 飲み $_{vp_1}$ 車-を運転し-た $_{vp_2}$
   beer-ACC drink car-ACC drive-PAST

Next, inference rules are extracted automatically from the extracted verb phrase pairs based on the following hypothesis:

> If two language expressions e_A and e_B（which represent the two different events in a verb phrase pair）share the same object, then it is likely that they hold the relation "if e_A then e_B", else it is likely that they do not hold the relation "if e_A then e_B".

For example, given two pairs ビールを飲む（drinking a beer）/酔う（getting drunk）, and ビールを飲む（drinking a beer）/運転する（driving）, in the former pair, since the verb 酔う often co-occurs with the object ビール（beer）, the pair ビールを飲む and 酔う is likely to be extracted as an inference rule. On the other hand, in the latter case, since the verb 運転する seldom co-occurs with the object ビール, the pair ビールを飲む and 運転する is unlikely to be extracted as an inference rule.

Finally, every verb phrase pair is assigned a score. The verb phrase pair is extracted as an inference rule if its score is greater than a specified threshold. The score function $Score(v, n, v_C)$ is shown in Figure 3.4. It measures the degree of co-occurrence between two verb phrases. The variable v indicates a verb, n indicates an object of v, and v_C indicates a verb following v. WO indicates the particle を（ACC）. $Class$ indicates a subset of the semantic space which is identified using word clustering based on the EM algorithm. α is an element in $Class$. $Score(v, n, v_C)$ is equal to the value of $Score_v(v, n, v_C)$ or $Score_{arg}(v, n, arg)$ whichever is the larger. $Score_v(v, n, v_C)$ refers to the case where n depends directly on v_C via a particle p. $Score_{arg}(v, n, arg)$ refers to the case where a noun arg is included as an adnominal phrase between n and v_C.

Torisawa extracted about 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 uses parallel verb phrase pairs as its source. However, current reported coverage on extraction is not high enough to make it usable in applications such as inference systems.

3.4 Rhetorical parsing

Several linguistic theory of textual coherence have been proposed. Rhetorical Structure Theory（RST）[46] 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. Consider again the sentences exemplified in (6) represented here as (14). RST suggests that there are different rhetorical relations in each of (14a_j) - (14c_j) , and (14a_e) - (14c_e) . The sentences (14a_j) and (14a_e) can be interpreted as PURPOSE rhetorical relations, (14b_j) and (14b_e) as CONDITION rhetorical relations, and (14c_j) and (14c_e) as CONTRAST rhetorical relations.

(14) a_j. みかんを食べる <u>ために</u> 皮を剥いた。

b_j. みかんは皮を剥かない と 食べられない。

c_j. みかんの皮を剥いた のに 食べることができないかった。

a_e. I peeled an orange to eat it.

b_e. If you do not peel an orange, you cannot eat it.

c_e. Though I peeled an orange, I could not eat it.

The aim of rhetorical parsing [47, 48] is to correctly determine the type of rhetorical relation present in a given sentence, as seen above.

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 2.2.2, 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.

Chapter 4

Committee-based decision making in probabilistic partial parsing

4.1 Introduction

The causal knowledge we acquire has two arguments which refer to two distinguish events. The events are expressed using a variety of linguistic forms; words, phrases, clauses, sentences and inter-sentential units. Among these forms, as described in the next chapter, we focus our attention on clauses in trying to capture events. Consequently, we need to analyze a sentence's dependency structure to identify its clauses. In this chapter, we describe an improved method of dependency structure analysis.

4.2 Background

There have been a number of attempts to use statistical techniques to improve parsing performance. While this goal has been achieved to a certain degree given the increasing availability of large tree banks, the remaining room for the improvement appears to be getting saturated as long as only statistical techniques are taken into account. In this chapter, we explore two directions for the statistical parsing: probabilistic partial parsing and committee-based decision making.

Probabilistic partial parsing is a probabilistic extension of the existing notion of partial parsing (e.g. [32]) where a parser selects as its output only a part of the parse tree that are probabilistically highly reliable. This decision-making scheme enables a fine-grained arbitrary choice on the trade-off between accuracy and coverage. Such trade-off is important since there are various applications that require reasonably high accuracy even sacrificing coverage. Enabling such trade-off choice will make state-of-the-art parsers of wider application.

Committee-based decision making is to combine the outputs from several different systems (e.g. parsers) to make a better decision. Recently, there have been various attempts to apply committee-based techniques to NLP tasks such as part of speech tagging [79, 5], parsing [20], word sense disambiguation [62], machine translation [14], and speech recognition [13]. Those works empirically demonstrated that combining different systems often achieved significant improvements over the previous best system.

In order to couple those committee-based schemes with probabilistic partial parsing, however, one would still need to make a further extension. Aiming at this coupling, we consider a general framework of committee-based decision making that consists of a set of weighting functions and a combination function, and discuss how that framework enables the coupling with probabilistic partial parsing. To demonstrate how it works, we report the results of our parsing experiments on a Japanese tree bank.

4.3 Probabilistic partial parsing

4.3.1 Dependency probability

Here, we consider the task of determining the dependency structure of a Japanese input sentence such as Figure 4.1. A *bunsetsu* phrase (BP) consists of a content word (noun, verb, adjective, etc.) accompanied by some function words (particles, auxiliaries, etc.). A Japanese sentence can be analyzed as a sequence of BPs, which constitutes an inter-BP dependency structure.

Given an input sentence s as a sequence of *bunsetsu* phrases (BPs), $b_1\ b_2\ \ldots\ b_n$,

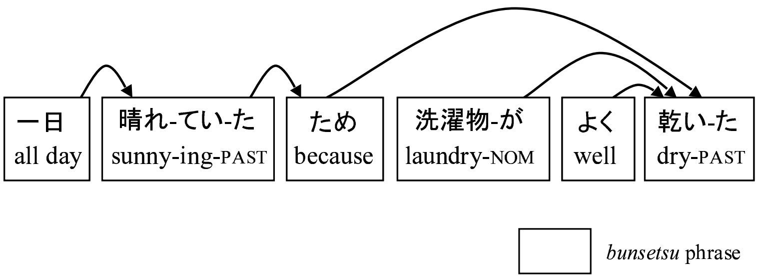


Figure 4.1 Dependency structure

our task is to identify their inter-BP dependency structure $R = \{r(b_i, b_j)|i = 1, \ldots, n\}$, where $r(b_i, b_j)$ denotes that b_i depends on (or modifies) b_j. Let us consider a *dependency probability* (DP), $P(r(b_i, b_j)|s)$, a probability that $r(b_i, b_j)$ holds in a given sentence s: $\forall i.\ \sum_j P(r(b_i, b_j)|s) = 1$.

4.3.2 Estimation of DPs

Some of the state-of-the-art probabilistic language models such as the bottom-up models $P(R|s)$ proposed by Collins [9] and Fujio et al. [15] directly estimate DPs for a given input, whereas other models such as PCFG-based top-down generation models $P(R, s)$ do *not* [8, 10, 71]. If the latter type of models were totally excluded from any committee, our committee-based framework would not work well in practice. Fortunately, however, even for such a model, one can still estimate DPs in the following way if the model provides the n-best dependency structure candidates coupled with probabilistic scores.

Let R_i be the i-th best dependency structure ($i = 1, \ldots, n$) of a given input s according to a given model, and let $\mathcal{R}_H$ be a set of R_i. Then, $P(r(b_i, b_j)|s)$ can be estimated by the following approximation equation:

$$P(r(b_i, b_j)|s) \approx \frac{P^r_{\mathcal{R}_H}}{P_{\mathcal{R}_H}} \tag{4.1}$$

where $P_{\mathcal{R}_H}$ is the probability mass of $R \in \mathcal{R}_H$, and $P^r_{\mathcal{R}_H}$ is the probability mass of $R \in \mathcal{R}_H$ that supports $r(b_i, b_j)$. The approximation error ϵ is given by $\epsilon \leq \frac{P_{\mathcal{R}} - P_{\mathcal{R}_H}}{P_{\mathcal{R}}}$, where $P_{\mathcal{R}}$ is the probability mass of all the dependency structure candidates for s (see [64] for the proof). This means that the approximation

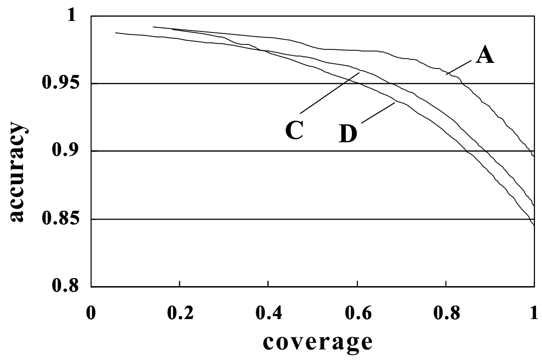


Figure 4.2 C-A curves

error is negligible if $P_{\mathcal{R}_H}$ is sufficiently close to $P_{\mathcal{R}}$, which holds for a reasonably small number n in most cases in practical statistical parsing.

4.3.3 Coverage-accuracy curves

We then consider the task of selecting dependency relations whose estimated probability is higher than a certain threshold $\sigma(0 < \sigma \leq 1)$. When σ is set to be higher (closer to 1.0), the accuracy is expected to become higher, while the coverage is expected to become lower, and vice versa. Here, coverage C and accuracy A are defined as follows:

$$C = \frac{\#\text{ of the decided relations}}{\#\text{ of all the relations in the test set}} \quad (4.2)$$

$$A = \frac{\#\text{ of the correctly decided relations}}{\#\text{ of the decided relations}} \quad (4.3)$$

Moving the threshold σ from 1.0 down toward 0.0, one can obtain a coverage-accuracy curve (C-A curve). In probabilistic partial parsing, we evaluate the performance of a model according to its C-A curve. A few examples are shown in Figure 4.2, which were obtained in our experiment(see Section 4.5). Obviously, Figure 4.2 shows that model A outperformed the other two. To summarize a C-A curve, we use the 11-point average of accuracy (*11-point accuracy*, hereafter), where the eleven points are $C = 0.5,\ 0.55, \ldots,\ 1.0$. The accuracy of total parsing corresponds to the accuracy of the point in a C-A curve where $C = 1.0$. We call

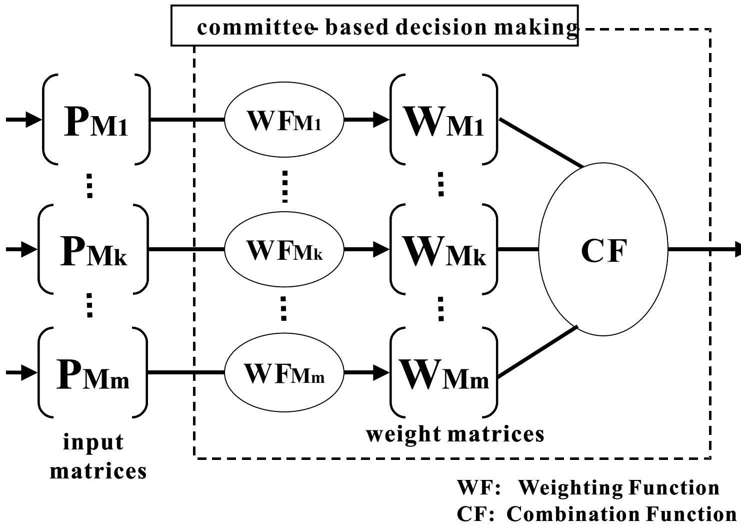


Figure 4.3 Committee-based probabilistic partial parsing

it *total accuracy* to distinguish it from 11-point accuracy. Note that two models with equal achievements in total accuracy may be different in 11-point accuracy. In fact, we found such cases in our experiments reported below. Plotting C-A curves enable us to make a more fine-grained performance evaluation of a model.

4.4 Committee-based probabilistic partial parsing

We consider a general scheme of committee-based probabilistic partial parsing as illustrated in Figure 4.3. Here we assume that each committee member M_k $(k = 1, \ldots, m)$ provides a DP matrix $P_{M_k}(r(b_i, b_j)|s)$ $(b_i,\ b_j \in s)$ for each input s. Those matrices are called input matrices, and are given to the committee as its input.

A committee consists of a set of weighting functions and a combination function. The role assigned to weighting functions is to standardize input matrices. The weighting function associated with model M_k transforms an input matrix given by M_k to a weight matrix W_{M_k}. The majority function then combines all the given weight matrices to produce an output matrix O, which represents

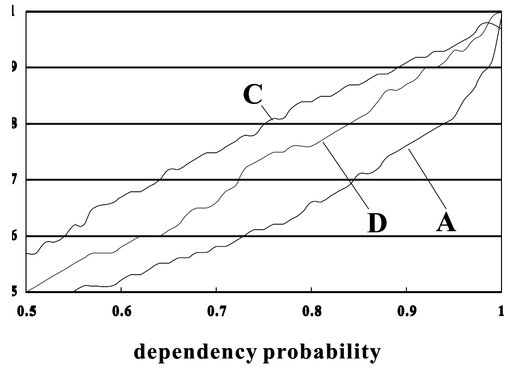


Figure 4.4 P-A curves

the final decision of the committee. One can consider various options for both functions.

4.4.1 Weighting functions

We have considered the following three options.

Simple The simplest option is to do nothing:

$$w_{ij}^{M_k} = P_{M_k}(r(b_i, b_j)|s) \tag{4.4}$$

where $w_{ij}^{M_k}$ is the (i, j) element of W_{M_k}.

Normal A bare DP may not be a precise estimation of the actual accuracy. One can see this by plotting probability-accuracy curves (P-A curves) as shown in Figure 4.4. Figure 4.4 shows that model A tends to overestimate DPs, while model C tends to underestimate DPs. This means that if A and C give different answers with the same DP, C's answer is more likely to be correct. Thus, it is not necessarily a good strategy to simply use given bare DPs in weighted majority. To avoid this problem, we consider the following weighting function:

$$w_{ij}^{M_k} = \alpha_i^{M_k} A_{M_k}(P_{M_k}(r(b_i, b_j)|s)) \tag{4.5}$$

where $A_{M_k}(p)$ is the function that returns the expected accuracy of M_k's vote with its dependency probability p, and $\alpha_i^{M_k}$ is a normalization factor. Such a function can be trained by plotting a P-A curve for training data. Note that training data should be shared by all the committee members. In practice, for training a P-A curve, some smoothing technique should be applied to avoid over-fitting.

Class The standardization process in the above option **Normal** can also be seen as an effort for reducing the averaged cross entropy of the model on test data. Since P-A curves tend to differ not only between different models but also between different problem classes, if one incorporates some problem classification into equation (4.5), the averaged cross entropy is expected to be reduced further:

$$w_{ij}^{M_k} = \beta_i^{M_k} A_{M_k C_{b_i}}(P_{M_k}(r(b_i, b_j)|s)) \tag{4.6}$$

where $A_{M_k C_{b_i}}(p)$ is the P-A curve of model M_k only for the problems of class C_{b_i} in training data, and $\beta_i^{M_k}$ is a normalization factor. For problem classification, syntactic/lexical features of b_i may be useful.

4.4.2 Combining functions

For combination functions, we have considered only simple weighted voting, which averages the given weight matrices:

$$o_{ij} = \frac{1}{m} \sum_{k=1}^{m} w_{ij}^{M_k} \tag{4.7}$$

where o_{ij} is the (i, j) element of O.

Note that the committee-based partial parsing framework presented here can be seen as a generalization of the previously proposed voting-based techniques in the following respects:

- A committee accepts probabilistically parameterized votes as its input.
- A committee accepts multiple voting (i.e. it allow a committee member to vote not only to the best-scored candidate but also to all other potential candidates).
- A committee provides a means for standardizing original votes.

- A committee outputs a probabilistic distribution representing a final decision, which constitutes a C-A curve.

For example, none of simple voting techniques for word class tagging proposed by van Halteren et al. [79] does not accepts multiple voting. Henderson and Brill [20] examined constituent voting and naive Bayes classification for parsing, obtaining positive results for each. Simple constituent voting, however, does not accept parametric votes. While Naive Bayes seems to partly accept parametric multiple voting, it does not consider either standardization or coverage/accuracy trade-off.

4.5 Experiments

4.5.1 Settings

We conducted experiments using the following five statistical parsers:

- KANA [11]: a bottom-up model based on maximum entropy estimation. Since dependency score matrices given by KANA have no probabilistic semantics, we normalized them for each row using a certain function manually tuned for this parser.
- CHAGAKE [15]: an extension of the bottom-up model proposed by Collins [9].
- Kanayama's parser [33]: a bottom-up model coupled with an HPSG.
- Shirai's parser [71]: a top-down model incorporating lexical collocation statistics. Equation equation (4.1) was used for estimating DPs.
- Peach Pie Parser [78]: a bottom-up model based on maximum entropy estimation.

Note that these models were developed fully independently of each other, and have significantly different characters (for a comparison of their performance, see Table 4.1). In what follows, these models are referred to anonymously.

For the source of the training/test set, we used the Kyoto corpus (ver.2.0) [40], which is a collection of Japanese newspaper articles annotated in terms of word boundaries, part of speech tags, BP boundaries, and inter-BP dependency relations. The corpus originally contained 19,956 sentences. To make the training/test sets, we first removed all the sentences that were rejected by any of the above five parsers (3,146 sentences). For the remaining 16,810 sentences, we next checked the consistency of the BP boundaries given by the parsers since they had slightly different criteria for BP segmentation from each other. In this process, we tried to recover as many inconsistent boundaries as possible. For example, we found there were quite a few cases where a parser recognized a certain word sequence as a single BP, whereas some other parser recognized the same sequence as two BPs. In such a case, we regarded that sequence as a single BP under a certain condition. As a result, we obtained 13,990 sentences that can be accepted by all the parsers with all the BP boundaries consistent [1]. We used this set for training and evaluation.

For closed tests, we used 11,192 sentences (66,536 BPs[2]) for both training and tests. For open tests, we conducted five-fold cross-validation on the whole sentence set.

For the classification of problems, we manually established the following twelve classes, each of which is defined in terms of a certain morphological pattern of modifying depending BPs:

1. nominal BP with a case marker "*wa* (TOPIC) "
2. nominal BP with a case marker "*no* (GEN) "
3. nominal BP with a case marker "*ga* (NOM) "
4. nominal BP with a case marker "*wo* (ACC) "
5. nominal BP with a case marker "*ni* (DAT) "
6. nominal BP with a case marker "*de* (LOC/...) "
7. nominal BP (residue)
8. adnominal verbal BP

[1]In the BP concatenation process described here, quite a few trivial dependency relations between neighboring BPs were removed from the test set. This made our test set slightly more difficult than what it should have been.

[2]This is the total number of BPs excluding the right-most two BPs for each sentence. Since, in Japanese, a BP always depends on a BP following it, the right-most BP of a sentence does not depend on any other BP, and the second right-most BP always depends on the right-most BP. Therefore, they were not seen as subjects of evaluation.

Table 4.1 Total and 11-point accuracies achieved by each model

	total	11-point
A	0.8974	0.9607
B	0.8551	0.9281
C	0.8586	0.9291
D	0.8470	0.9266
E	0.7885	0.8567

9. verbal BP（residue）
10. adverb
11. adjective
12. residue

4.5.2 Results and discussion

Table 4.1 shows the total/11-point accuracy of each individual model. The performance of each model widely ranged from 0.96 down to 0.86 in 11-point accuracy. Remember that A is the optimal model, and there are two second-best models, B and C, which are closely comparable. In what follows, we use these achievements as the baseline for evaluating the error reduction achieved by organizing a committee.

The performance of various committees is shown in Figure 4.5. The mark “ ＊” refers to a baseline accuracy for each committee. Our primary interest here is whether the weighting functions presented above effectively contribute to error reduction. In Figure 4.5, although the contribution of the function **Normal** is not very visible, the function **Class** consistently improved accuracy. These results are good evidence for the importance of weighting functions in combining parsers. While we performs problem classification manually in our experiment, automatic classification techniques are also obviously worth considering.

We then conducted another experiment to examine the effects of multiple voting. One can straightforwardly simulate a single-voting committee by replacing

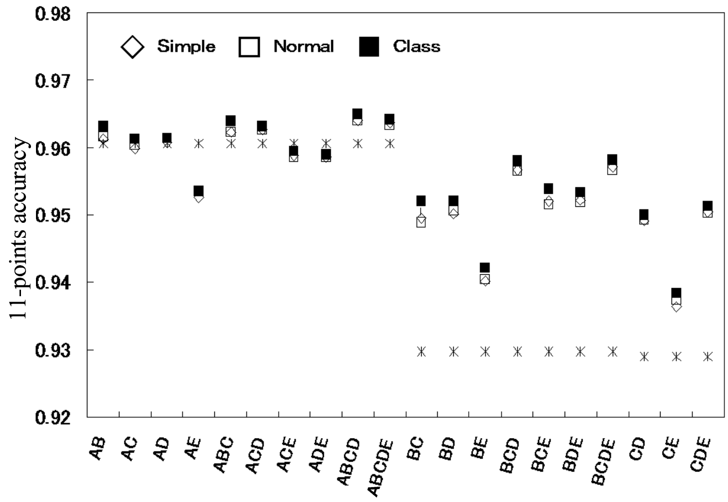


Figure 4.5 11-point accuracy

w_{ij} in equation equation (4.7) with w'_{ij} given by:

$$w'_{ij} = \begin{cases} w_{ij} & (\text{if } j = \arg\max_k w_{ik}) \\ 0 & (\text{otherwise}) \end{cases} \tag{4.8}$$

The results are shown in Figure 4.6, which compares the original multi-voting committees and the simulated single-voting committees. Clearly, in our settings, multiple voting significantly outperformed single voting particularly when the size of a committee is small.

The next issues are whether a committee always outperform its individual members, and if not, what should be considered in organizing a committee. Figure 4.5 show that committees not including the optimal model A achieved extensive improvements, whereas the merit of organizing committees including A is not very visible. This can be partly attributed to the fact that the competence of the individual members widely diversed, and A significantly outperforms the other models.

Given the good error reduction achieved by committees containing comparable members such as BC, BD and BCD, however, it should be reasonable to expect that a committee including A would achieve a significant improvement if another

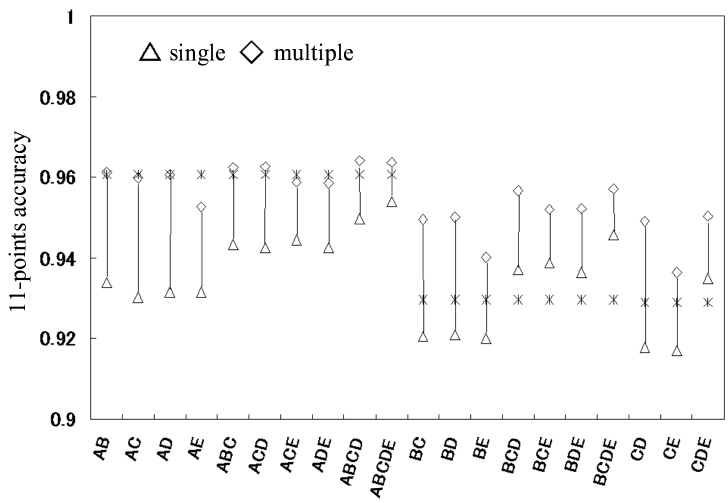


Figure 4.6 Single voting vs. multiple voting

nearly optimal model was also incorporated. To empirically prove this assumption, we conducted another experiment, where we add another parser KNP [39] to each committee that appears in Figure 4.5. KNP is most accurate model in total accuracy than the other models (0.9125 in total accuracy). However, it does not provide DP matrices since it is designed in a rule-based fashion — the current version of KNP provides only the best-preferred parse tree for each input sentence without any scoring annotation. We thus let KNP to simply vote its total accuracy. The results are shown in Figure 4.7. This time all the committees achieved significant improvements, with the maximum error reduction rate up to 31%.

As suggested by the results of this experiment with KNP, our scheme allows a rule-based non-statistical parser to play in a committee preserving its ability to output parametric DP matrices. To push the argument further, suppose a plausible situation where we have an optimal but rule-based non-statistical parser and several suboptimal statistical parsers. In such a case, our committee-based scheme may be able to organize a committee that can provide DP matrices while preserving the original total accuracy of the rule-based parser. To see this, we conducted another experiment, where we combined KNP with each of A and D,

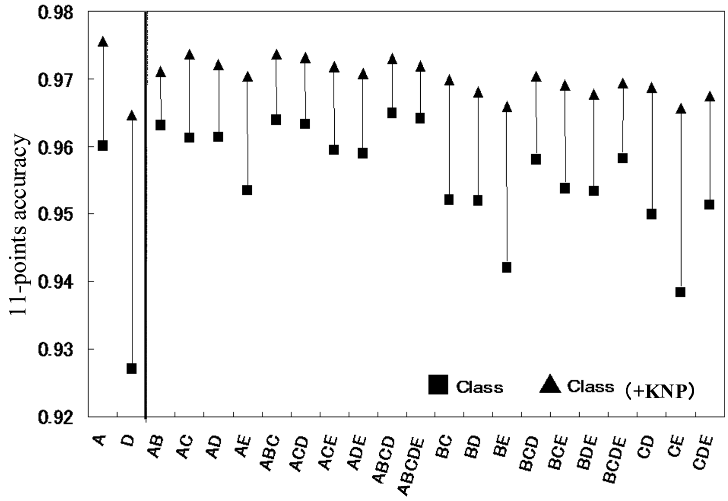


Figure 4.7 11-point accuracy (+KNP)

both of which are less competent than KNP. The resulting committees successfully provided reasonable P-A curves as shown in Figure 4.8,

4.6 Summary

In this chapter, we presented a general committee-based framework that can be coupled with probabilistic partial parsing. In this framework, a committee accepts parametric multiple votes, and then standardizes them, and finally provides a probabilistic distribution. We presented a general method for producing probabilistic multiple votes (i.e. DP matrices), which allows most of the existing probabilistic models for parsing to join a committee. Our experiments revealed that (a) if more than two comparably competent models are available, it is likely to be worthwhile to combine them, (b) both multiple voting and vote standardization effectively work in committee-based partial parsing, (c) our scheme also allows a rule-based non-statistical parser to make a good contribution.

While we empirically demonstrated that this framework contributed to more accurate decision-making than that based on any individual committee member,

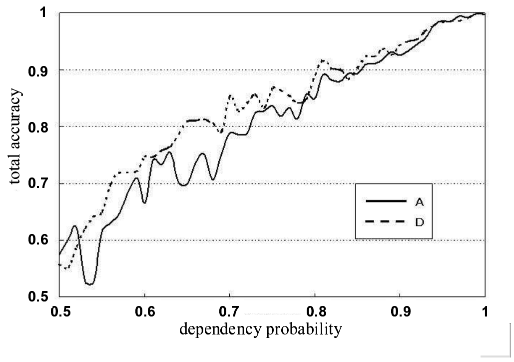


Figure 4.8 P-A curves（+KNP）

there are some drawbacks. One main weak point is the speed of analysis. Analysis within our framework is too slow for application to huge document collections because it requires running multiple parsers in parallel to act as the committee members in the decision-making.

In recent years, however, an accurate and fast dependency structure analyzer（parser）based on machine learning techniques [37, 38] has become available. We employed *CaboCha*[3] [38] as the dependency structure analyzer in the investigations described in this thesis.

[3] Available from `http://cl.aist-nara.ac.jp/~taku-ku/software/cabocha/`

Chapter 5

Data collection and analysis

5.1 Introduction

Linguists suggest that there are some expressions which 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 problems, in this chapter we describe the details of our investigation of the distribution of causal relations in Japanese newspaper articles.

In Section 5.2 we describe a variety of cue phrase expressions for causal relations. In Section 5.3, we introduce the data on which we base both our investigation, and the acquisition of causal knowledge. Section 5.4 and Section 5.5 describe the procedure adopted for data collection and the results of its analysis. The main component of the collection procedure was conducted based on human judgments using the linguistic tests described in Section 5.4.1. The data set created in this chapter was used as training data and evaluation data for supervised learning as described in Chapter 6 and Chapter 7.

5.2 Causal expressions in Japanese text

Causal relations can be expressed in various ways. For example, Altenberg [3] 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 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*）

In addition to these types of linkage, Altenberg mentions that there are two other types: one is implicit linkage and the others is a relation lexicalized as a verb, the causative verb. Girju et al. [17] used an extended version of Altenberg's scheme with a more detailed classification of causal expressions.

It can be assumed that a similar causal expression classification can be performed on Japanese text [50, 49] as has been performed on English. The following list shows Japanese causal expressions[1] categorized according to part of speech of the cue phrases.

- **Conjunction**

(15) ・なぜなら（*nazenara*）

私-は　眠い-です。 なぜなら　昨日-は 寝-てい-ない-から-です。
I-TOPIC feel sleepy *nazenara* yesterday-TOPIC sleep-ing-not

・だから（*dakara*）

風邪-を 引い-た。 だから 学校-を 休ん-だ。
catch a cold-PAST *dakara* absent from school-PAST

・というのは（*toiunoha*）

私-は 学校-を 休ん-だ。 というのは 熱-が あっ-た-から-です。
I-TOPIC absent from school *toiunoha* have a slight fever

[1] We call cue phrases which often implicitly induce causal relations between two events in a sentence *causal expressions*. Note that it is not guaranteed that causal relation instances can be extracted from sentences including causal expressions.

- **Conjunctive particle**

(16) ・ため（*tame*）

家-を	買う-ため	貯金する。
house-ACC	buy-*tame*	save（money）

・ので（*node*）

風邪な-ので	学校-を	休ん-だ。
cold-*node*	school-ACC	absent-PAST

・から（*kara*）

遅れる-から	そろそろ	出かけ-ます。
late-*kara*	now	go out

・あまり（*amari*）

急い-だ-あまり	転ん-で-しまっ-た。
hurry-PAST-*amari*	fall-PAST

・せいで（*seide*）

寝ぼうし-た-せいで	電車-に	乗り遅れ-た。
oversleep-PAST-*seide*	train	miss-PAST

- **Particle**

(17) ・に（*ni*）

酒-に	酔う。
alcohol-*ni*	get drunk

・から（*kara*）

一つ-の	失敗-から	低迷する。
one	failure-*kara*	downturn

・で（*de*）

風邪-で	学校-を	休む。
cold-*de*	school-ACC	absent

All of the expressions listed above induce causal relations explicitly. Just as in English, Japanese makes use of implicit causal expressions. Arita [4] suggests that the degree of inevitability expressed by implicit causal expressions is equivalent or higher than that of explicit causal expressions. In the sentences (18), the upper two sentences are examples of implicit causal expressions. These examples show that though no explicit linguistic expressions are used, causal relations are held between the sentences (18x) and (18y) which are the arguments of the causal relations in (18a) and (18b).

(18) a. お金が足りなくて困った。

b. お金が足りず困った。

x. お金-が 足り-ない
money-NOM enough-not

y. 困った
be in trouble-PAST

It can be assumed that implicit expressions of causal relation are used highly frequently in text. However, we excluded implicit expressions from our attention because it seems likely that the knowledge acquisition process for causal relations implicitly indicated in the text will be harder than that for causal relations explicitly indicated in the text. In this thesis, we therefore focus our attention on texts containing explicit causal expressions.

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

While it is preferable to tackle explicit causal relations first for the reasons given above, 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 case elements connected by particles

- Clauses are usually constructed using the same constituents as events, say, a predicate and case elements connected by particles

If good results for knowledge acquisition can be achieved using conjunctive particles, we will then expand to use of additional types of causal expressions.

5.3 Data

5.3.1 Selection of corpus

In the 1980's, several Japanese tagged corpora were constructed including the Kyoto corpus [40] and the EDR corpus [83] which are annotated with morphological and dependency structure information. Though this annotation would be useful for our goal, we deal with only the plain, non-tagged, corpus as a source of causal knowledge since annotated resources are less scalable due to the high cost of construction.

Currently, there are several kinds of electronic text resources available such as newspaper [28], dictionary [65], encyclopedia [6], novel, e-mail and web text. We selected text resources according to two criteria based on quantity and quality.

quantity criterion: The amount of knowledge acquired is dependent on the quantity of source text. Therefore, to acquire as much knowledge as possible, the more text available the better.

quality criterion: During the knowledge acquisition process, we should perform some pre-processing such as morphological analysis and syntactic dependency structure analysis as precisely as possible. If mistakes occur at the pre-processing stage, we can expect the knowledge acquired to also be incorrect. To acquire correct knowledge, the sentences included in the source text should be grammatical and be amenable to correct analysis at the pre-processing stage using the natural language analyzer.

However, in general, there is a trade-off between quantity and quality. Sentences in dictionaries and encyclopedias are of high quality since they are generated with care by experts. However the size of dictionaries and the encyclopedias is constrained and is unlikely to increase because of the long publication cycle and there are few new publications every years. On the other hand, text from e-mail and the web is increasing everyday, but the quality of these text may be very low because they are usually written without care and particular attention by ordinary people.

In this work, we selected newspaper articles as the source texts for causal knowledge acquisition because they more or less satisfy both criteria and both the morphological analyzer（*ChaSen*）and dependency structure analyzer（*CaboCha*）are optimized for the types of sentences found in newspaper articles. The web should be used as a source of knowledge in order to increase coverage in the future. However, it can be assumed that if we cannot achieve successful knowledge acquisition results using newspaper articles, we will be unable to do so using lower quality text documents such as web text.

5.3.2 Selection of cue phrases

We use explicit causal expressions such as だから（because）, ため（because）, and から（because of）. As mentioned in Section 5.2, causal relations are indicated in various ways in Japanese text.

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 cue phrase expressions. Table 5.1 shows the ten most frequent cue phrase expressions in the collection of Nihon Keizai Shimbun newspaper articles from 1990 [28]. This table was generated by counting all conjunctive particles after morphological analysis of the articles using *ChaSen*.

In this table, the word ため（because） and れば（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 ため based on the approximately 2000 examples used in Section 5.4, Section 6.3 and Section 7.2.

Based on the above discussion, we selected the word **ため** as our main target for further exploration. The reasons for this choice are:

- ため is used relatively frequently in our corpus

Table 5.1 Frequency distribution of connective markers

が（*ga*）	（but）	131,164
ため（*tame*）	（because）	76,087
と（*to*）	（if/when）	56,549
れば（*reba*）	（if）	48,606
ながら（*nagara*）	（while）	13,796
から（*kara*）	（because）	10,209
ので（*node*）	（because）	9,994
なら（*nara*）	（if）	7,598
たら（*tara*）	（if）	6,027
のに（*noni*）	（but）	2,917

Table 5.2 Frequency distribution of ため in intra-sentential contexts

Type	Freq.	Examples
（a）adverbial verb phrase	42,577	晴れ-た-ため 洗濯物-が よく 乾い-た。 fine-PAST-*tame* laundry-NOM well dry-PAST
（b）other types	33,510	これ-は 旅行者-の-ため-の 乾燥器-です。 this-TOPIC tourist-*tame*-GEN tumble dryer-COPULA

- ため typically indicates causal relations more explicitly than other markers

5.3.3 Selection of linguistic unit

Table 5.2 shows the frequency distribution of the intra-sentential contexts in which ため appears in the same newspaper article corpus used in Section 5.3.2. The sentences classified into the "adverbial verb phrase" type are defined as follows:

adverbial verb phrase: The dependency structure of each sentence including ため is analyzed with *CaboCha* and each modifier *bunsetsu* phrase of ため indicated by "A" in Figure 5.1 and modified *bunsetsu* phrase of ため indicated by "B" in Figure 5.1 is identified. A sentence is classified as being

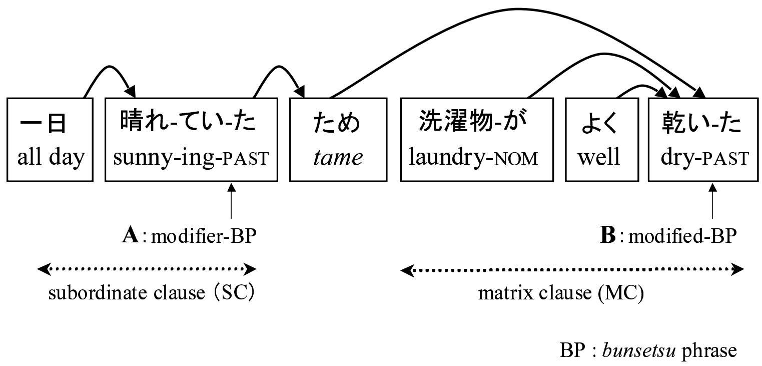


Figure 5.1 Structure of a ため-complex sentence

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
“動詞-自立 (verb-independent)” as defined in *ChaSen*’s dictionary[2].
e.g. 乾く(dry)

c2.) Includes a morpheme whose part of speech is
“形容詞-自立 (adjective-independent)” as defined in *ChaSen*’s dictionary.
e.g. 強い(strong)

c3.) Includes a morpheme whose part of speech is
“名詞-形容動詞語幹 (nominal adjectival stem)” as defined in *ChaSen*’s dictionary.
e.g. 好調だ(satisfactoriness)

c4.) Includes a morpheme whose part of speech is any noun category except “名詞-形容動詞語幹 (nominal adjectival stem)”, and does not include the morpheme “no” whose part of speech is “助詞- 連体化 (particle-adnominal)”
e.g. 見通しだ(prospect)

[2] http://chasen.aist-nara.ac.jp/stable/ipadic/ipadic-2.6.3.tar.gz

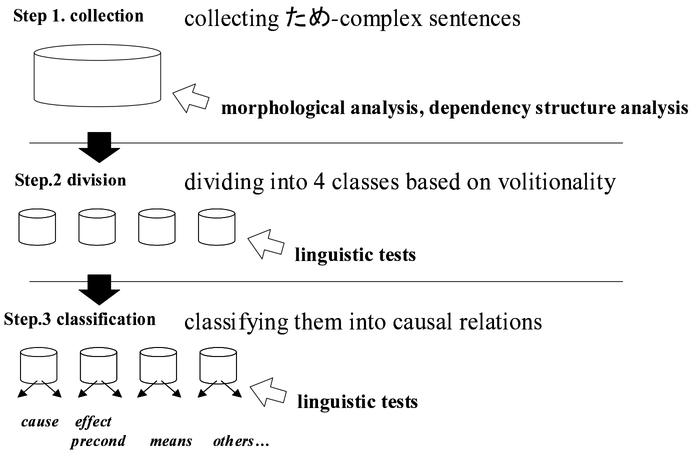


Figure 5.2 Workflow for investigating frequency distribution

From Table 5.2, we can see that the word ため is most frequently used as an adverbial connective marker accompanying a verb phrase that constitutes an adverbial subordinate clause (see Table 5.2-(a)). Hereafter, sentences including such clauses will be referred to as **ため-complex sentences**. We were pleased to observe this tendency because, as argued in Section 2.2.1, the acquisition from complex sentences with adverbial subordinate clauses is expected to be easier than from sentences with other types of clues such as nominal phrases (see Table 5.2-(b)). Based on this preliminary survey, we restrict our attention to ため-complex sentences.

5.4 Procedure

We assembled a collection of data for examining the distribution of implicitly held causal relations in ため-complex sentences as follows (see also Figure 5.2):

Step 1: collection. We first took random samples from a newspaper article corpus of 1000 sentences that were automatically categorized into ため-complex sentences. Removing interrogative sentences and sentences from which a subordinate-matrix clause pair was not properly extracted due to preprocessing (morphological analysis and dependency structure analysis)

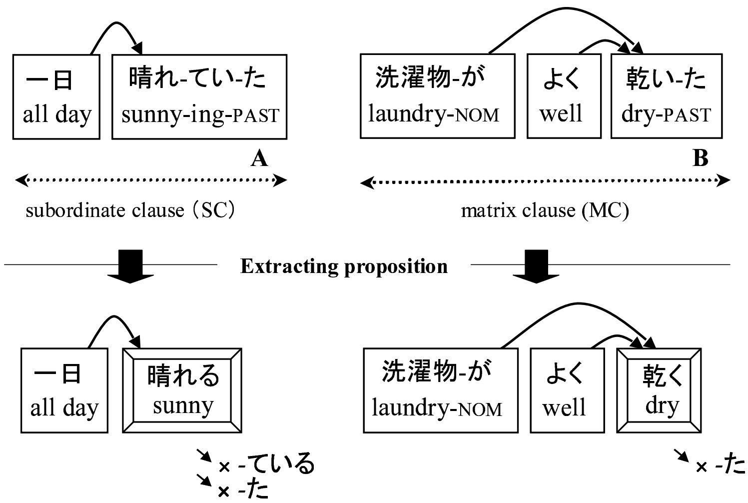


Figure 5.3 Proposition extraction

errors, we were left with 994 remaining sentences. These 994 sentences fulfill the condition of c1 for the adverbial verb phrase. We refer to this set of sentences as $\mathcal{S}_1$.

Step 2: division. We extracted the proposition from each subordinate and matrix clause of sentences in $\mathcal{S}_1$. We removed the modality elements attached to the end of the head verb which is the verb in *bunsetsu* phrase "A" or "B". An example is shown in Figure 5.3. The remaining elements were considered the proposition. By this operation, some modality information such as tense or passive voice information was erased. Hereafter, we refer to 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 non-volitional SOA) in the subordinate and matrix clauses. Volitionality was judged using the linguistic tests described in the next section. The frequency distribution of the four classes (A – D) is shown in the left-hand side of Table 5.3. 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

Table 5.3. Distribution of causal relations in ため-complex sentences in $\mathcal{S}_1$ （SC denotes the subordinate clause and MC denotes the matrix clause. Act_s and SOA_s denote an event referred to by the SC, and Act_m and SOA_m denote an event referred to by the MC.）

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)

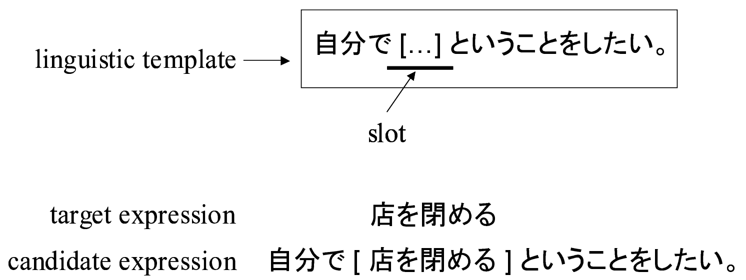


Figure 5.4 Linguistic template

into the class D fulfill the necessary condition for the *means* relation[3].

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 exemplified in Table 2.1. The details of the linguistic tests for judging the causal relations will be described in more detail in the next section.

5.4.1 Linguistic tests

[3]It should be noted that it is not always possible to confirm that the same agents condition is fulfilled.

A linguistic test is a method for judging whether a linguistic expression, normally a sentence, conforms to a given set of rules. We call the sentence that we wish to judge a *target* expression. The rules are realized as a linguistic template, a linguistic expression including several slots as shown in Figure 5.4.

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.

Linguistic tests for judging volitionality

We prepared 4 templates for volitionality judgments as follows:

vol_t1 自分-で　[...] と-いう-こと-を　し-たい。
by oneself　the thing that [...] -ACC　want to do

vol_t2 自分-で　[...] と-いう-こと-を　する-つもり-です。
by oneself　the thing that [...] -ACC　will do

vol_t3 自分-で　[...] と-いう-こと-を　する-こと-に-する。
by oneself　the thing that [...] -ACC　will do

vol_t4 自分-では　[...] と-いう-こと-を　し-たく-は-ない。
by oneself　the thing that [...] -ACC　not want to do

The square brackets indicate the slot 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 non-volitional 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.

(19) a. 店-を 閉める。
store-ACC close

b. 自分で [店を閉める] ということをするつもりです。

c. volitional action

(20) a. 株式-市場-が 低迷する。
stock market-NOM downturn

b. * 自分で [株式市場が低迷する] ということをしたい。

c. * 自分で [株式市場が低迷する] ということをするつもりです。

d. * 自分で [株式市場が低迷する] ということをすることにする。

e. * 自分で [株式市場が低迷する] ということをしたくはない。

f. non-volitional SOA

Linguistic tests for judging causal relations

We prepared 3 to 7 templates for each causal relation as follows:

- ***cause***

cau_t1 [SOA] （という） こと-が 起こる その-結果-として
[SOA] （that） thing-NOM happen as a result of
しばしば [SOA] （という） こと-が 起こる。
usually [SOA] （that） thing-NOM happen

cau_t2 [SOA] （という） 状態-で-は、
[SOA] （that） state-TOPIC
しばしば [SOA] （という） 状態-と-なる。
usually [SOA] （that） become

cau_t3 [SOA]（という）状態-に-な-れば、それに伴い、
[SOA]（that） become-if it-based on
しばしば [SOA]（という） 状態-に-なる。
usually [SOA] （that） become

- ***effect***

eff_1 [Act]（という） こと-を する-と、
[Act] （that） thing-ACC execute
しばしば [SOA]（という） こと-が 起こる。
usually [SOA] （that） thing-NOM happen

eff_2 [Act]（という） こと-を する-と、 その-結果、
[Act] （that） thing-ACC execute-when as a result of
通常 [SOA]（という） 状況-に-なる。
usually [SOA] （that） state-DAT-happen

eff_3 [Act]（という） こと-を する-こと-は、
[Act] （that） thing-ACC execute-thing-TOPIC
ふつう [SOA]（という） 状態-を 保つ。
usually [SOA] （that） state-ACC keep

- ***precond***

pre_1 [SOA]（という） 状況-で-は、
[SOA] （that） state-TOPIC
しばしば [Act]（という） こと-を する。
usually [Act] （that） thing-ACC execute

pre_2 [SOA]（という） 状態-で-は、
[SOA] （that） state-TOPIC
しばしば [Act]（という） こと-を する。
usually [Act] （that） thing-ACC execute

pre_3 [SOA]（という） 状態-に-なる-場合、
[SOA] （that） become-when
しばしば [Act]（という） こと -を する。
usually [Act] （that） thing-ACC execute

- ***means***

mea_1 X-が [Act] （という）こと-を 実現する その 手段-として、
X-NOM [Act] （that）thing-ACC realize it by means of
X-が [Act]（という）こと-を する-の-は もっとも-で-ある。
X-NOM [Act]（that） thing-ACC execute-thing-TOPIC plausible

mea_2 X-が [Act] （という） こと-を 実現する その 手段-として、
X-NOM [Act] （that） thing-ACC realize it by means of
X-が [Act]（という） こと-を 行なう-の-は もっとも-で-ある。
X-NOM [Act] （that） thing-ACC execute-thing-TOPIC plausible

mea_3 X-が [Act] （という） こと-を 実現する その 手段-として、
X-NOM [Act] （that） thing-ACC realize it by means of
X-が [Act] の-は もっとも-で-ある。
X-NOM [Act] thing-TOPIC plausible

mea_4 X-が [Act] （という） こと-を する-こと-によって、
X-NOM [Act] （that） thing-ACC execute-as a result of
X-が [Act]（という） こと-を する。
X-NOM [Act] （that） thing-ACC execute

mea_5 X-が [Act] （という） こと-を する-こと-によって、
X-NOM [Act] （that） thing-ACC execute-as a result of
X-が [Act] の-で-ある。
X-NOM [Act] thing-COPULA

mea_6 X-が [Act] （という） こと-を する-こと-によって、
X-NOM [Act] （that） thing-ACC execute-as a result of
X-が [Act]（という） こと-が できる。
X-NOM [Act] （that） thing-ACC can execute

mea_7 X-が [Act]（という） こと-の 一貫-として、
X-NOM [Act] （that） thing-GEN as part of
X-が [Act]（という）こと-を する-の-は もっとも-で-ある。
X-NOM [Act] （that） thing-ACC execute plausible

We embed the subordinate clause and matrix clause 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 しばしば（often） or 普通（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 (21a) . However, we are able to recognize the sentence in (21b) which contains the expression 普通（usually） as incorrect, since the relation possesses a very low degree of inevitability.

(21) a. 宝くじ-を 買ったら 一等-が 当たっ-た。
lottary-ACC buy-PAST-when first prize-NOM win-PAST.

b. * 宝くじ-を 買ったら <u>普通</u> 一等-が 当たる。
lottary-ACC buy-when <u>usually</u> first prize-NOM win.

This constraint affect 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.

(22) a. 市場-が 低迷し-た-<u>ため</u> 店-を 閉める。
market-NOM downturn-PAST-<u>*tame*</u> store-ACC close

b. [市場が低迷する] 状況ではしばしば [店を閉める] ことをする。

c. *precond* relation

(23) a. 海岸線-を 開発する-<u>ために</u> 調査-を 終え-ている。
coastline-ACC exploit-<u>*tameni*</u> survey finish

b. * [海岸線を開発する] ということをすると、その結果、通常 [調査を終える] 状況になる。

c. ...

d. （no templates are correct）

e. *others* relation

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. We calculated the κ statistical measure using 200 annotated samples. The κ value was 0.93 for volitionality, 0.88 for causal relations. This means that the reliability judgments of both volitionality and causal relations is sufficiently high.

5.5 Analysis

5.5.1 The marker ため

The right-hand side of Table 5.3 shows the most abundant relation and its ratio for each class A – D. For example, given a ため-complex sentence, if the subordinate clause refers to a volitional action and the matrix clause refers to a non-volitional SOA（namely, class B）, they are likely to conform to the relation *effect*(Act$_s$, SOA$_m$) with a probability of 0.93（149/161）.

The following are examples of the most abundant relation in a given class（for further examples of causal relation instances, see Appendix A）.

(24) タイ-で マングローブ-を
in Thailand mangrove-ACC
破壊した-ため、 大水害-が 発生した。
destroy-PAST-*tame* flooding-NOM occur-PAST

Act$_s$: タイでマングローブを破壊する
SOA$_m$: 大水害が発生する

→ *effect*(⟨ タイでマングローブを破壊する ⟩, ⟨ 大水害が発生する ⟩)

(25) 北京-への 切符-を 買う-<u>ため</u> 切符売場-に 行った。
for Beijing ticket-ACC buy-<u>*tame*</u> to ticket office go-PAST

Act_s: 北京への切符を買う
Act_m: 切符売場に行く

→ *means*(⟨ 切符売場に行く ⟩, ⟨ 北京への切符を買う ⟩)

The following are examples of cases where the most abundant relation in a given class did not hold. Hereafter, we refer to these as *others* relations.

(26) a. 企業-の 成長-が 見込める-よう-に-なっ-た-<u>ため</u>
company-GEN growth-NOM can be expected-<u>*tame*</u>
機運-が 高まっ-ている。
growing tendency

（Although this sentence fulfills the necessary condition for the *cause* relation, it is rejected for all of the templates in linguistic test.）

b. たくさん 収容する-<u>ため</u> ホール-は 扇形- に なっている。
many（audience） contain-<u>*tame*</u> hall-TOPIC sector be

（Although this sentence fulfills the necessary condition for the *effect* relation, it is rejected for all of the templates in linguistic test.）

c. 古楽器-が 舞台-の 温度-に なれる-<u>ために</u>
old instrument-NOM stage-GEN temperature-DAT accustom-<u>*tameni*</u>
オーケストラ-は 十分-あまり 調整し-た。
orchestra about 10 min. key-PAST

（Although this sentence fulfills the necessary condition for the *precond* relation, it is rejected for all of the templates in linguistic test.）

The distribution shown in Table 5.3 is quite suggestive. As far as ため-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 ため-complex sentences into the four relations — *cause*, *effect*, *precond* and *means* — with precision of 85% or more. Motivated by this observation, in the next chapter we first address the issue of automatic estimation of clausal volitionality before moving onto the issue of automatic classification of causal relations.

5.5.2 Other markers

We attempted the same procedure outlined in Section 5.4 using the other five cue phrases on a small sample set. The cue phrases and data sizes used in this investigation are as follows:

cue phrase		data size
ので	（because）	1000
れば	（if）	200
たら	（if）	200
が	（but）	200
のに	（but）	200

In order to apply the linguistic tests to the cue phrases が and のに, we developed a minor variation of candidate sentence generation as described below.

- If a target expression located in the matrix clause includes a negative expression, we remove the negative expression from the target expression. We then embed the resulting target expression in the slots of the templates to form the candidate sentences.

(27) 父親-が 亡くなっ-た-のに 告別式-に 駆けつけ-ない
father-NOM die-PAST-*noni* to funeral rush-not

Act$_m$: 告別式に 駆けつけ-ない
to funeral rush-not
→ 告別式に 駆けつける
to funeral rush

- If a target expression located in the matrix clause does not include a negative expression, we add a negative expression onto it. We then embed the resulting target expression in the slots of the templates to form the candidate sentences.

(28) 日-が さす-のに 雨-が 降る
sun-NOM shine-*noni* rain
SOA_m: 雨が 降る
rain
→ 雨が 降ら-ない
rain-not

The results for each cue phrase are shown in Table 5.4 to Table 5.8. Looking at the tables, it is clear that these five cue phrases are of less use than ため due to the fact that almost half of the samples were not classifiable within our typology of causal relations. The word ので has a relatively similar distribution to ため as compared to the other four cue phrases. However, no samples were classified as the *means* relations.

Based on the above results, in this thesis we do not use these five cue phrases in the experiments on knowledge acquisition described in Chapter 6 and Chapter 7.

The following (29) are sample sentences from which extracted subordinate and matrix clause pairs were identified as either (a) causal relations or (b) non-causal relations within our typology. The samples are grouped according to cue phrase.

(29) ・ので (*node*)

a. 寒く-なる-ので 健康状態-を 心配する。
cold-become-*node* health condition-ACC worry
(*precond* relation)

b. 植物-は 環境-の 変化-で 痛む-ので
plant-TOPIC environment-GEN change-cause blight-*node*
頻繁な 場所替え-は 避ける。
frequent resite-TOPIC avoid
(*others* relation)

・れば（*reba*）

a. 不況-が 深刻化す-れば 債務負担-が 増加する。
economic downturn-NOM become serious-*reba* debt burden-NOM increase
（*cause* relation）

b. 良い 時-も あ-れば 悪い 時-も ある。
good condition-case-also be-*reba* bad condition-case-also be
（*others* relation）

・たら（*tara*）

a. 副作用-が で-たら 服用-を 中止する。
side effect-NOM experience-*tara* take medicine-ACC stop
（*precond* relation）

b. 本-の 整理-を し-てい-たら おもしろい もの-が でてきた。
book tidy up-ACC doing-*tara* interesting thing-NOM find-PAST
（*others* relation）

・が（*ga*）

a. 金利-高-で 受注-が 伸び悩む-が
high interest rates-cause order entry-NOM low growth-*ga*
フル生産-を 続行する。
full-produce-ACC continue
（*precond* relation）

b. 殺人-は 減っ-た-が 強盗-の 件数-が 増え-た。
murder-TOPIC decrease-*ga* rob-GEN number--nom increase-PAST
（*others* relation）

・のに（*noni*）

a. 料金-を　倍-に　した-<u>のに</u>　客-が　増え-た。
fee-ACC　double　do-PAST-<u>*noni*</u>　visitor-NOM　increase-PAST
（*effect* relation）

b. 会社-に　行く-<u>のに</u>　団地-から　バス-に　乗る。
to office　go-<u>*noni*</u>　from apartment　bus-ACC　get on
（*others* relation）

5.6 Summary

In this chapter we described the data collection procedure, and the results of investigating the distribution of causal relations in Japanese newspaper articles. As shown in Table 5.3, the distribution is quite suggestive. The subordinate clause, matrix clause pairs extracted from ため-complex sentences indicate events and can be classified into the four relation types — *cause*, *effect*, *precond* and *means* — with a precision of 85% or more.

Based on these results, we attempt to automatically acquire causal knowledge from ため-complex sentences as described in Chapter 7. Note that the difficulty of acquiring causal knowledge depends on the characteristics of the target cue phrases. In the case of ため 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 ため-complex sentences. However, when focusing on cue phrases with a lower degree of inevitability than ため such as たら（if）, in addition to classifying causal relations, it is necessary to take account of the framework that determines the degree of inevitability.

Table 5.4 Distribution of causal relations in sentences including *node*

class	SC	MC	frequency	Most frequent relation and its ratio	
A	SOA	SOA	337	*cause*(SOA_s, SOA_m)	0.88 (297/337)
B	Act	SOA	180	*effect*(Act_s, SOA_m)	0.93 (160/180)
C	SOA	Act	310	*precond*(SOA_s, Act_m)	0.81 (251/310)
D	Act	Act	151	—	0 (0/151)
		total	978		0.72 (708/978)

Table 5.5 Distribution of causal relations in sentences including *reba*

class	SC	MC	frequency	Most frequent relation and its ratio	
A	SOA	SOA	55	*cause*(SOA_s, SOA_m)	0.73 (40/55)
B	Act	SOA	80	*effect*(Act_s, SOA_m)	0.33 (26/80)
C	SOA	Act	25	*precond*(SOA_s, Act_m)	0.56 (14/25)
D	Act	Act	22	—	0 (0/22)
		total	182		0.44 (80/182)

Table 5.6 Distribution of causal relations in sentences including *tara*

class	SC	MC	frequency	Most frequent relation and its ratio	
A	SOA	SOA	45	*cause*(SOA_s, SOA_m)	0.38 (17/45)
B	Act	SOA	65	*effect*(Act_s, SOA_m)	0.34 (22/65)
C	SOA	Act	44	*precond*(SOA_s, Act_m)	0.50 (22/44)
D	Act	Act	27	—	0 (0/27)
		total	181		0.34 (61/181)

Table 5.7 Distribution of causal relations in sentences including *ga*

class	SC	MC	frequency	Most frequent relation and its ratio		
A	SOA	SOA	86	*cause*(SOA_s, SOA_m)	0.17	(15/86)
B	Act	SOA	44	—	0	(0/44)
C	SOA	Act	28	*precond*(SOA_s, Act_m)	0.32	(9/28)
D	Act	Act	31	—	0	(0/31)
		total	189		0.13	(24/189)

Table 5.8 Distribution of causal relations in sentences including *noni*

class	SC	MC	frequency	Most frequent relation and its ratio		
A	SOA	SOA	105	*cause*(SOA_s, SOA_m)	0.40	(42/105)
B	Act	SOA	27	*effect*(Act_s, SOA_m)	0.37	(10/27)
C	SOA	Act	31	*precond*(SOA_s, Act_m)	0.74	(23/31)
D	Act	Act	24	—	0	(0/24)
		total	187		0.40	(75/187)

Chapter 6

Estimation of volitionality

6.1 Introduction

As discussed above, to acquire causal knowledge, it is important to estimate the volitionality (volitional action or non-volitional SOA) of the clauses. Hereafter, we call volitionality of the clause *clausal volitionality.* In the fields of linguistics, works have 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 chapter, we present our approach to estimating clausal volitionality.

6.2 Feature discovery

What factors are used to characterize clausal volitionality? In this section, before describing our estimation method, we describe some of these factors.

6.2.1 Contextual ambiguity

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 $\mathcal{S}_1$ 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 (30a) and the matrix clause of (30b) contain the same verb 拡大する(expand), however (30a) refers to a volitional action and (30b) refers to a non-volitional SOA.

(30) a. 生産能力-を 拡大する-ため 設備投資する。
production ability-ACC expand-*tame* make plant investment

b. 管理費-が 削減した ため、 営業利益-が 拡大した。
cost-NOM reduce-PAST-*tame* profit-NOM expand-PAST

6.2.2 Verbal volitionality

Here we introduce some work related to verbal volitionality.

IPA verb dictionary [7]

The IPA verb dictionary is a manually annotated resource including information about verbal volitionality. All information in this dictionary is very accurate, and is frequently used in linguistic analysis. However, it has one major drawback which is its small scale. Figure 6.1 shows the relationship between number of sentences from newspaper articles, and number of verb entries in the dictionary, for verbs appearing in the sentences. Triangles in the graph refer to the results using *ChaSen*'s dictionary. Squares refer to the results when using the IPA verb dictionary. We can see from this graph that the IPA verb dictionary has very few entries in comparison with the number of verb entries that appear in the newspaper articles. Furthermore, some verbs in the IPA dictionary are assigned more than one volitionality value leading to ambiguity. Nakagawa et al. [56] tried to achieve disambiguation of verbal volitionality in the IPA verb dictionary using simple heuristic rules. However their rules are not accurate because no

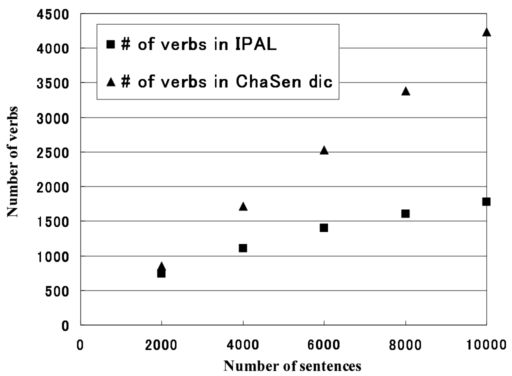


Figure 6.1. The number of verbs in the IPAL dictionary and *ChaSen*'s dictionary

consideration of context is taken at all. Based on the above factors, the IPA verb dictionary is likely to be of little use in estimating clausal volitionality.

Adverbs of subject aspect [58]

Nitta [58] tried to analyze and systematically describe adverbial modifiers. He asserts that there is a category of adverbs specifically related to the volitionality of verbs, called adverbs of subject aspect, such as あえて（dare）, しぶしぶ（reluctantly）, てきぱき（promptly） and 精いっぱい（to the best of one's ability). These adverbs have the characteristic of often co-occurring with a volitional verb in the sentence. The following are example sentences.

(31) a. 太郎-は 日曜-に <u>しぶしぶ</u> 働い-た。
Taro-TOPIC on Sunday <u>reluctantly</u> work-PAST

b. * 花子-は <u>しぶしぶ</u> 風邪-を 引い-た。
Hanako-TOPIC <u>reluctantly</u> cold catch-PAST

That means that if one of these adverbs of subject aspect occurs in a sentence, the modified verb is volitional.

Unfortunately, these adverbs, is though appearing to offer promising cues, are entirely absent from the sample sentences. This reflects the nature of newspaper articles which generally adopt on objective point of view.

Transitivity

Transitivity of verbs is one factor that can be used to characterize verbs. Verbs can be divided into two types based on their transitivity. One type is transitive verbs such as 食べる（eat）, as in the sentence 太郎が野菜を食べる（Taro eats vegetables）. Transitive verbs take a direct object, which in this case is 野菜（vegetables）. The other type is intransitive verbs such as 上がる（rise）, as in the sentence 東から太陽が上がる（The sun rises in the east）. Intransitive verbs do not take any object.

Transitive verbs usually express a volitional action. However, it is by no means always the case that there is a correspondence between the transitivity of a verb and verbal volitionality. Some verbs are transitive, but non-volitional such as なくす（lose）, as in the sentence お金をなくした（(I) lost some money）. Other verbs are intransitive, but volitional such as 歩く（walk）, as in the sentence 公園の周りを歩く（(I) walk around the park）.

The IPA dictionary and adverbs of subject aspect are unsuitable for our purposes, and while transitivity is useful, it does not offer a complete solution. Therefore, we need to incorporate additional factors in order to estimate clausal volitionality with high accuracy. As a result of analyzing the ため-complex sentences in $\mathcal{S}_1$, we found other factors in addition to the verb that help determine clausal volitionality.

- A clause tends to be non-volitional 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.

6.3 Estimation of volitionality using SVMs

Based on the above findings, we investigated experimentally how accurately clausal volitionality（volitional action or non-volitional SOA）of clauses can be estimated using Support Vector Machines – an accurate binary classification algorithm.

6.3.1 Experimental conditions

Support vector machines

Support Vector Machines (SVMs) are binary classifiers, originally proposed by Vapnik [80]. SVMs have performed with high accuracy in various task, such as text categorization, bio-informatics and face identification. In this work, we use the TinySVM[1] software package.

In this experiment, using only intra-clause 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. We used the quadratic polynomial kernel as the kernel function.

Though we omit here details of the algorithm, note that in most cases the generalization performance of SVMs does not depend on the dimensionality of the feature space. And, the kernel function we use can deal with non-linear classification by means of an implicit mapping of the original feature space into a high dimensional space. In particular, the polynomial kernel function makes it possible to deal with any combination of features.

Features

Table 6.3 (at the end of this chapter) shows the features we used to represent the clauses in the sentences. Fortunately, the verbal features which provide the most important information, can be easily extracted from the dictionaries: the EDR concept dictionary, the dictionary incorporated in the ALT-J/E translation system, and NTT Goi-Taikei. 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 such as the sentences shown in (30) presented in Section 6.2.1. In the field of NLP, ellipsis resolution is well-known as a very difficult task. In this experiment,

[1] Available from `http://cl.aist-nara.ac.jp/~taku-ku/software/TinySVM/`

Table 6.1 Ratio of volitionality for each clause type

		frequency	
		Act / SOA	total
$\mathcal{S}_1$	Subordinate clause	539 / 455	994
	Matrix clause	603 / 391	994
$\mathcal{S}_2$	Subordinate clause	613 / 372	985
	Matrix clause	650 / 335	985

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 is described in the next section.

Data

We used all the sentences in $\mathcal{S}_1$ (described in a previous chapter) as training samples and a new set of ため-complex sentences, $\mathcal{S}_2$ as test samples. The set $\mathcal{S}_2$ includes 985 ため-complex sentences. This set was created using the same procedure as $\mathcal{S}_1$. We first sampled 1000 random sentences from a newspaper article corpus issued in a different year than $\mathcal{S}_1$. 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. We extracted subordinate and matrix clauses from $\mathcal{S}_1$ and $\mathcal{S}_2$ respectively using a set of heuristic rules. The frequency distributions of clausal volitionality for both $\mathcal{S}_1$ and $\mathcal{S}_2$ are shown in Table 6.1.

6.3.2 Results

Table 6.2 shows the results. The values in Table 6.2 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 6.3.

Table 6.2 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

The column "SVM1" denotes accuracy in the case where a classifier is trained using all the features shown in Table 6.3 except agent information. The columns "BL1" and "BL2" denote accuracy for the baseline. "BL1" is a very simple 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 non-volitional SOA (see Table 6.1). "BL2" denotes the accuracy achieved by applying a simple classification strategy as follows:

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 volitional action by default.

Table 6.2 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.

Introducing a reliability metric

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

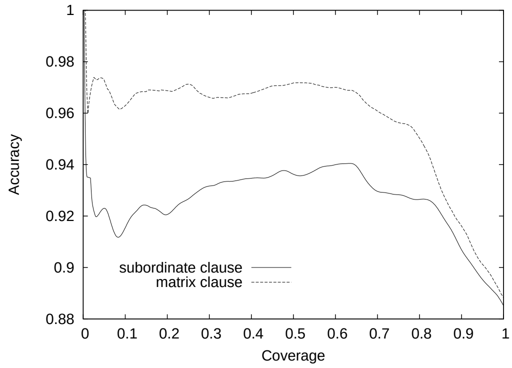


Figure 6.2 Coverage-accuracy curves for clausal volitionality estimation

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 σ, thus ensuring that a judgment would only be output for a given sample when the reliability was greater than σ.

We applied this metric to the results of "SVM2". By varying σ, we obtained the coverage-accuracy curves of Figure 6.2 where:

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

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

Figure 6.2 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.

6.4 Summary

In this chapter, we described volitionality estimation. Although clausal volitionality often corresponds to verbal volitionality, it is slightly different due to the presence of additional contextual information. By using a machine learning approach to include this contextual information, we achieved promising results for clausal volitionality estimation.

Table 6.3 Feature set used for volitionality estimation
V:Verb C:Case M:Modality

class — descriptions
V Word — a base form item
V EDR — Four features indicating the verb class given by the EDR concept dictionary [83]: （1）true if the verb is “移動(movement）” or “行為(action）”, false otherwise; （2）true if the verb is “状態（state)”, “変化（change)” or “現象（phenomenon)”, false otherwise; （3）true if both（1）and（2）are true; （4）true if none of the above is true.
V ALT-J/E — A set of binary features indicating the verb class given by the dictionary incorporated in the ALT-J/E translation system [27, 26]: “状態動詞（state）”, “継続動詞（continuous situation）”, “瞬間動詞（momentary situation）”, “自動詞(intransitive）”, “他動詞(transitive）”, “補助動詞(auxiliary）”, “可能動詞（potential）”, “自発動詞（spontaneous）”, “使役動詞（causative）”, “受身（passivizable）”, “受身（被害）（indirect-passivizable）”.
V Goi-Taikei — Verbal semantic attributes in NTT Goi-Taikei [26].
C Marker — が（NOM）, を（ACC） **C Element** — The concept of case element described in NTT Goi-Taikei [26].
M Tense — -る（PRESENT）form or -た（PAST）form. **M Aspect** — -ている（-ing）form or not. **M Voice** — -れる（PASSIVE）form or not. -せる（CAUSATIVE）form or not. **M Potential** — -できる（POSSIBE）form or not. **M Negative** — -ない（NEG）form or not.
Agent — Whether or not the agent is a human or an organization.

Chapter 7

Automatic acquisition of causal knowledge

7.1 Introduction

In this chapter, we describe an experiment designed to assess how accurately causal relation instances can automatically be acquired. As shown in Figure 2.2, represented here as Figure 7.1, the process consists of two main phases. We implemented a high precision rule based proposition extractor for the first phase using existing NLP techniques. In this chapter we describe our approach to automatic identification（classification）of causal relations.

In Chapter 5, we described the distribution of causal relations in Japanese newspaper articles. It was demonstrate that the pairs of subordinate and matrix clauses in the ため-complex sentences can be classified by hand into the causal relation classes within our typology with a precision of 85% or more. In Chapter 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 ため-complex sentences. Clausal volitionality was used as a feature in performing classification. We used five classes — *cause*, *effect*, *precond*, *means* and *others* relation classes. The *others* class contains relation instances which were categorirzed into the *others* class during the investigation of distribution of causal relations described in Chapter 5. We again used SVMs as the classifier.

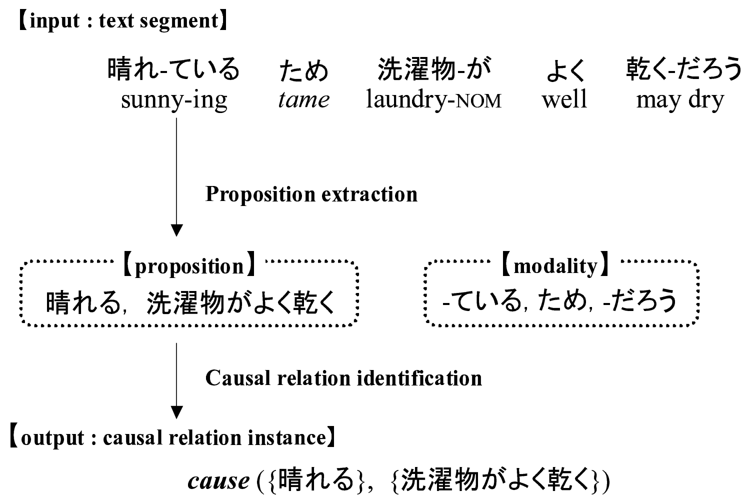


Figure 7.1 Knowledge acquisition workflow

7.2 Automatic identification using SVMs

7.2.1 Procedure

It seems likely that when the degree of confidence in clausal volitionality estimation is low a sample could be incorrectly identified since clausal volitionality is a major factor used in classifying the sample. 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 ため-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 referred as $conf_s$ and that of a matrix clause is referred as $conf_m$.

Step B: filtering Suppose that a threshold for the degree of confidence of subordinate clauses is σ_s, and a threshold for the degree of confidence of matrix clauses is σ_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 increase precision.

Step C: classification Classify the sample into a causal relation class using SVMs. The one-versus-rest method was used so that we could apply SVMs to multiple classifications. When the discriminant function value acquired from two or more classifiers with this method was positive, the classifier with the maximum function value was ultimately selected.

7.2.2 Experimental conditions

Features

The features we used are as follows:

f1. The clausal volitionality estimated by the technique described in the previous chapter,

f2. All the features except "V word" shown in Table 6.3, and

f3. Whether the agents of the two clauses in the sentence are the same.

The third, agent correspondence feature can be automatically extracted by using the technique described in Nakaiwa et al. [57] with a high level of precision. However, in this experiment, we were unable to implement this method due to the unavailabling of the detailed linguistic rules required for implementation of this technique. Instead, a simple rule-based extractor was used.

Data

The data are the same as those in Section 6.3. We used the sentences in $\mathcal{S}_1$ as training samples and $\mathcal{S}_2$ as test samples. We first estimated the clausal volitionality and its reliability using all the training data in $\mathcal{S}_1$. 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 ため-complex sentences in $\mathcal{S}_2$ is shown in Table 7.1. The notations in this table follow those in Table 5.3.

Table 7.1 Distribution of causal relations in ため-complex sentences in $\mathcal{S}_2$

class	Most frequent relation and its ratio	
A	$\mathit{cause}(\mathrm{SOA}_s,\ \mathrm{SOA}_m)$	0.98 (193/196)
B	$\mathit{effect}(\mathrm{Act}_s,\ \mathrm{SOA}_m)$	0.78 (108/139)
C	$\mathit{precond}(\mathrm{SOA}_s,\ \mathrm{Act}_m)$	0.94 (166/176)
D	$\mathit{means}(\mathrm{Act}_m,\ \mathrm{Act}_s)$	0.79 (375/474)
		0.85 (842/985)

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 3-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 3 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.

Table 7.2 3-point averaged precision (Step B skipped)

feature		3-point averaged precision			
		cause	*effect*	*precond*	*means*
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 7.3 3-point averaged precision (Step B included)

feature		3-point averaged precision			
		cause	*effect*	*precond*	*means*
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 Results

The results are shown in Table 7.2, Table 7.3, and Figure 7.2[1]. Table 7.2 shows the 3-point averaged precision of the causal relation classification where Step B (filtering) is skipped. On the other hand, Table 7.3 shows the results where Step B is included. Figure 7.2 shows the recall-precision curves corresponding to Table 7.3.

In this section, we discuss the results from the viewpoint of the effect of clausal volitionality on the classification. First, we consider the results of Table 7.2. In Table 7.2, R1 refers to the results in the case where the classifiers were trained using only f1, clausal volitionality as features. In this case, no information except

[1]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.

clausal volitionality was used. Therefore, we applied a minor variation of the classification process, "Step C-forR1" presented later, instead of the one presented in Section 7.2.1. R2 refers to the results in the case where the classifiers were trained using f2 and f3. R3 refers to the results in the case where the classifiers were trained using all the features described in Section 7.2.2.

Step C-forR1:

- If the clausal volitionality of a subordinate clause is estimated to be **a non-volitional SOA** and the clausal volitionality of a matrix clause is estimated to be **a non-volitional 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 non-volitional SOA**, classify the sample into the ***effect*** relation class.

- If the clausal volitionality of a subordinate clause is estimated to be **a non-volitional 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.

  where the degree of confidence in a causal relation classification is calculated as $conf_s + conf_m$.

The accuracy 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 accuracy 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 accuracy 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 different between R3 and R3' is small. This means that even if the feature extractor is improved, the significant improvements in the causal relation classification will not be achieved.

Next, we discuss the effect of step B, the sample filtering process. Table 7.3 shows the results when Step B is included. The up-arrow in Table 7.3 indicates that the performance improves as a result of the sample filtering process. In this work, we set the threshold such that 20% of input samples are classified into the *others* relation class due to having a lower score than the threshold. 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.

7.3 Discussion

How much causal knowledge could be acquired from resources?

Let us estimate the amount of knowledge one can acquire from ため-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*

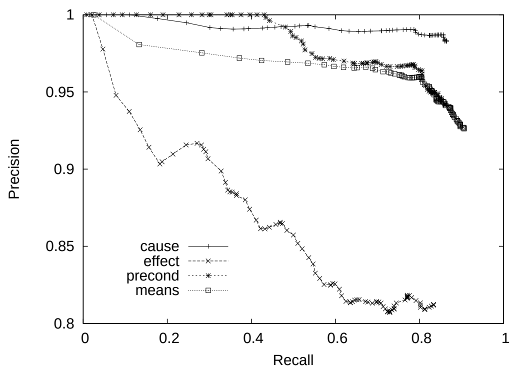


Figure 7.2 Recall-precision curves for causal relation classification

relations, and 90% for the *effect* relations. First, it can be seen from Figure 7.2 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 ため-complex sentences are as given in Table 7.1. 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: $0.19_{\mathrm{FR}}^{cause} \times 0.79_{\mathrm{REC}}^{cause} + 0.11_{\mathrm{FR}}^{effect} \times 0.30_{\mathrm{REC}}^{effect} + 0.17_{\mathrm{FR}}^{precond} \times 0.82_{\mathrm{REC}}^{precond} + 0.38_{\mathrm{FR}}^{means} \times 0.83_{\mathrm{REC}}^{means} = 0.64$.

Finally, since we collected about 42,500 ため-complex sentences from one year of newspaper articles (see Table 5.2), 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 [73] and Marcu's results [48].

How can the acquired knowledge be applied?

We anticipate using acquired causal knowledge within the framework of case-based reasoning. Our causal relation instances, as discussed in Section 2.2.3, 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:

- 花子が洗濯物を乾かす（Hanako dries the laundry）
- 洗濯物が乾く（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 (32) , whose arguments are similar to the input samples, is included in the memory-base.

(32) *effect*(⟨太郎が洗濯物を乾かす⟩, ⟨洗濯物が乾く⟩)
Taro dries the laundry the laundry dries

Similarly, when an input is only partial e.g. 花子が洗濯物を乾かす and the causal relation instance (32) is in the memory-base, an event 洗濯物が乾く could be inferred to happen as a result of 花子が洗濯物を乾かす.

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 (33) would include the constituents タイで（in Thailand） because the location where the mangrove is destroyed and where flooding occurred should be the same.

(33) *effect*(⟨タイで マングローブを 破壊する⟩,
（someone）destroy mangrove swamps in Thailand
⟨（ϕ = タイで）大水害が 発生する⟩)
serious flooding occur（in Thailand）

Failing to correct for these constituents is likely to result in incorrect inferences. We therefore 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 冷たい（cool） in （34a） may be required to construct the *cause* relation instance （34a）. On the other hand, 秋口に（in early autumn） in （34b） may be optional.

(34) a. *cause*(⟨冷たい 空気が流れ込む⟩, ⟨ 最低気温が平年を下回る ⟩)
cool air is coming   lowest temperature is lower than in a normal year

b. *cause* (⟨秋口に 出荷頭数が増える ⟩,
the number shipped increases in early autumn
⟨ 豚価は下がる ⟩)
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.

7.4 Summary

In this chapter, we presented accuracy figures for the automatic acquisition of causal relation instances using our method. By employing machine leaning 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（see Figure 7.2）. Furthermore, the experimental results suggest that one can expect to acquire over 27,000 instances of causal relations from one year of Japanese newspaper articles.

Chapter 8

Acquiring desirability lexical knowledge from causal knowledge

8.1 Introduction

In this chapter, we describe an application in which we use previously acquired causal knowledge. The typology of causal relations we employ was originally proposed in the field of discourse understanding research. It might be assumed that the causal knowledge acquired using the method discussed in the previous chapter would be of immediate use in a discourse understanding system. Unfortunately however, we do not have available an implementation of a discourse understanding system. Instead we attempt to utilize causal knowledge in an alternative application, desirability lexical knowledge acquisition.

8.1.1 Motivation

Automatic extraction of human opinion from text media has attracted a certain amount of research interest. In particular, one requirement is for a technique for estimating whether events are considered desirable or undesirable.

Desirability of an event, that is, whether an event is desirable or undesirable, is sometimes explicitly indicated with cue phrases such as よい（good） as in (35a), which express a subject's evaluation of the event. On the other hand, it is sometimes expressed using objective constituents only, such as the sentence

(35b). In this case, though no explicit cue phrases expressing the subject's evaluation are used, we are able to judge that the event expressed in (35b) is undesirable.

(35) a. 美味しくて　よかっ-た
delicious　good-PAST

b. 遠足-なのに　雨-が　降っ-ている
picnic-but　it rains

It seems that word dictionaries with information about desirability such as "よい(good)/desirable" and "悪い(bad)/undesirable" could be useful for estimating desirability of sentences. However, since desirability depends on a number of factors such as domain information and the viewpoints of subjects, a low cost framework for the construction of dictionaries is required.

Given this background, we propose a bootstrap method for acquiring lexical knowledge about desirability with no human cost, and estimating desirability of sentences, based on equivalence relations between sentences (with equal desirability values). The method consists of three processes: a construction process, a transition process, and a decomposition process (described in detail later). In the transition process, the causal relation instances discussed in previous chapters are used.

In this chapter, we describe our method for the construction of dictionaries and experimental results from our investigation into the feasibility of our method.

8.2 Definition of desirability

Desirability of events usually depends on several factors such as how subjects are involved in the events. We define desirability of events as follows:

> Given a specific goal, when subjects achieve the goal or make progress towards the goal, the event in question is a desirable event (we call a desirable event a *positive* event). On the other hand, if subjects deviate or retreat from the goal, the current event is an undesirable event (we call an undesirable event a *negative* event). Events which

cannot be classed as either desirable or undesirable are considered *neutral*.

In this work, we deal with text documents which discuss recycling. We set the goal to be environmental protection. In this case, events which relate to recycling such as garbage reduction, paper recycling, and conservation of resources are positive events. Events which prevent recycling are negative events. The following are example sentences from documents about recycling where ⟨*positive*⟩/⟨*negative*⟩ at the end of sentences refers to the desirability of those sentences.

(36) a. スチール缶-の　再-資源-化-率-を　引き上げる　⟨*positive*⟩
steel can　recycling-rate-ACC　raise

b. 空き-缶-の　デポジット-制-が　定着する　⟨*positive*⟩
empty can　deposit system-NOM　establish

c. エコ-バッグ-を　使い　店-の　包装-紙-を　使わ-ない　⟨*positive*⟩
eco-bag-NOM　use　store-GEN　package paper-NOM　use-not

(37) a. OA-化-が　進み　紙-ゴミ-が　急増する
office automation-NOM　develop　paper-garbage-NOM　explosion
⟨*negative*⟩

b. イメージダウン-を　恐れて　簡易-包装-に-は　消極的だ
image down-ACC　fear　simple-package-DAT-TOPIC　passive
⟨*negative*⟩

c. 一般-の　回収-業者-で-は　引き取り-手-が　少ない　⟨*negative*⟩
regular　collection-agency-TOPIC　adopt-NOM　few

8.3 *pn* dictionary

8.3.1 Characteristics of *pn* dictionary

We call a set of word/desirability pairs a *pn* dictionary. Table 8.1 shows example entries in our *pn* dictionary for the recycling domain. One of our aims is

Table 8.1 Examples of *pn* dictionary entries

positive entries
ゴミ-減量(garbage reduction) リサイクル(recycle) 分別(separate)
再生-品(reprocessed goods) 分別-作業(garbage separation)
低-公害-車(low pollution car) 回収(collection) 優れる(exceed)
回収-ペットボトル(recovered PET bottle) 好調(favorableness)
negative entries
使用-済み-ペットボトル(used PET bottle) 捨てる(throw away)
深刻-化(escalation) 恐れる(fear) 過剰-包装(excess packaging)
環境-破壊(environmental destruction) 不純物(insufficient)
無駄(waste) 回収-コスト(collection cost) ゴミ-問題(garbage problem)

to acquire word/desirability pairs and so increase entries in the dictionary. The entries in the *pn* dictionary have the following characteristics.

- **domain dependent entries**

  Some entries are independent of domain such as 楽しい(happy) and 恐れる(fear). On the other hand, other entries are domain dependent such as リサイクル(recycle) and 回収(recover).

- **object entries**

  While single verbs are able to represent an event, nouns are usually not able to do so in isolation but instead represent an object. To enable us to interpret nouns as events, we extend the meanings of nouns as follows:

  – OBJECT(NOUN)_ENTRY-が ある (ゴミ-が ある)
    OBJECT(NOUN)_ENTRY-NOM exists (the garbage exists)

  – OBJECT(NOUN)_ENTRY-が 増える (ゴミ-が 増える)
    OBJECT(NOUN)_ENTRY increases (*garbage increases*)

  – OBJECT(NOUN)_ENTRY-が たくさん ある (ゴミ-がたくさん ある)
    There is a lot of OBJECT(NOUN)_ENTRY (there is a lot of garbage)

  In this example, the noun ゴミ(garbage) is interpreted, for example, as ゴミがある(the garbage exists). Consequently, since it seems that ゴミ

がある is a negative event, the pair, ゴミ/*negative* is registered in the *pn* dictionary.

- **compound word entries**

  It was found that though the same head word occurs sometimes in two different compound words, the desirability of those compound words is not necessarily the same, as shown in (38) .

  (38) a. 使用-済み-ペットボトル ⟨*negative*⟩
  used PET bottle

  b. 回収-ペットボトル ⟨*positive*⟩
  recovered PET bottle

  We register both the compound words, and their constituent words in the *pn* dictionary.

8.3.2 *pn* operators

Nagae et al. [54] mention that there are words which cannot specify desirability in themselves. We treat these words as *pn* operators which form patterns. The patterns contain a slot into which other words can be placed（Figure 8.1）. When a word fills the slot in the pattern, desirability of the combination of the *pn* operator and the words filling the slot is activated. For example, the word 高い（high）is an example of a *pn* operator. 高い in the sentence (39a) and (39b) cannot specify desirability in itself. However, when a word such as リサイクル率（proportion recycled）or リサイクルコスト（recycling cost）fills the slot in the pattern, the desirability of the sentence is specified.

(39) a. $\boxed{\text{リサイクル-率}_{positive}}$-が 高い$_{plus}$ ⟨*positive*⟩
proportion recycled-NOM high

b. $\boxed{\text{リサイクル-コスト}_{negative}}$-が 高い$_{plus}$ ⟨*negative*⟩
recycling cost-NOM high

c. $\boxed{\text{環境負荷}_{negative}}$-が 少ない$_{minus}$ ⟨*positive*⟩
environmental-load-NOM small

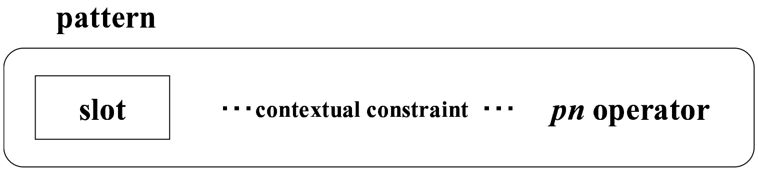


Figure 8.1 Pattern

Table 8.2 Examples of *pn* operators

plus	高い(high) 増える(increase) 乗り出す(start) 進める(advance)
minus	減らす(decrease) 少ない(few) なくす(remove) 抑える(curb)

d. $\boxed{株価_{neutral}}$ -が 高い$_{plus}$ ⟨*neutral*⟩
stock price-NOM high

We use as *pn* operators a selection of words with the meanings of *degree*, *start*, and *continuance*. Examples are shown in Table 8.2. We assign *pn* operators an attribute value of *plus* or *minus*, which is registered in the *pn* dictionary in the desirability field. When a word fills the slot in a pattern which includes a *pn* operator whose attribute value is *plus*, desirability of the combination of *pn* operator and the word filling the slot is equal to that of the word that fills the slot. On the other hand, if a word fills the slot in a pattern which includes a *pn* operator whose attribute value is *minus*, the desirability of the combination is the inverse value to that of the word filling the slot of the pattern. That is, if the word filling in the slot is *positive*, the desirability value of the combination will be *negative*, and if a word filling the slot is *negative*, the desirability value of the combination will be *positive*. The example sentence (39c) illustrates the inverse case. When desirability of a word filling the slot is *neutral*, desirability of the combination is specified as *neutral* independent of the attribute value of the *pn* operator. When a contextual constraint on the pattern (described later) which restricts the syntactic structures that can appear between a *pn* operator and the pattern's slot is not fulfilled, the *pn* operator is treated as an ordinary *pn* entry whose attribute value is *neutral*. The contextual constraint allows only particles between the *pn* operator and the pattern's slot. The patterns we use are as follows:

$$\boxed{\text{slot}}\ (\text{が}(\textsc{nom})|\text{を}(\textsc{acc})|\text{に}(\textsc{dat})|\text{の}(\textsc{gen})|\text{は}(\textsc{topic})) \quad (8.1)$$
$$\rightarrow\ pn\_\text{operator}$$
$$(\text{“}\rightarrow\text{” refers to a dependency relation})$$

In this thesis, we call a body of lexical knowledge such as Table 8.1 and Table 8.2 a *pn* dictionary. However, it seems that *pn* operators may be shared among different domains since most *pn* operators have general meanings and are not domain dependent. Therefore, we aim to develop a method for acquiring ordinary entries whose attribute values are *positive*, *negative* or *neutral*, and ignore acquisition of *pn* operators.

8.4 A bootstrap method

In this section, we describe a bootstrap method for acquiring *pn* entries and estimating desirability of sentences, using equivalence relations on event desirability. Causal relation instances are used to identify equivalence relations between sentences. Figure 8.2 shows the method's workflow. The method consists of the following three processes:

pn **construction:** Desirability of a sentence is estimated based on construction rules and the *pn* dictionary.

pn **transition:** Desirability of a sentence is given by another sentence using an equivalence relation based on equality of desirability of both sentences.

pn **decomposition:** *pn* entries are estimated based on desirability of a sentence and a *pn* decomposition scheme.

First, when a sentence is composed of words with *pn* entries such as リサイクル運動を推し進める（the recycling movement is gaining momentum）, desirability

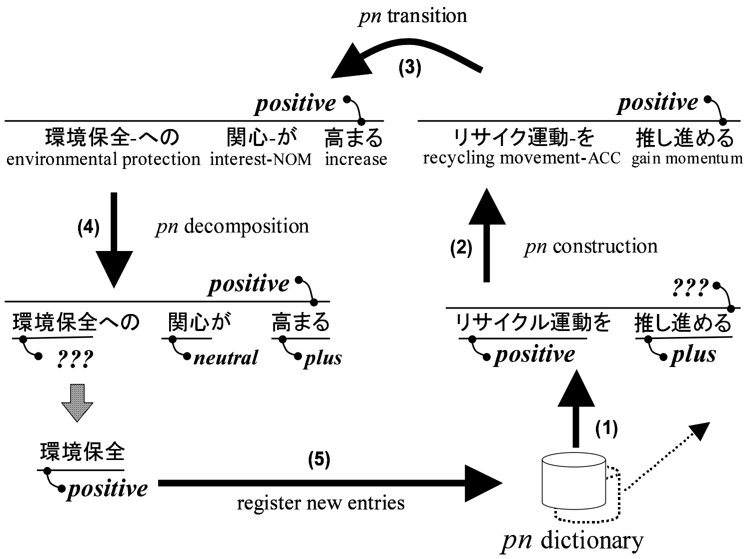


Figure 8.2 Workflow for *pn* entry acquisition

of the sentence is estimated using that of its component words (hereafter, referred to as a *pn-specified* sentence), as shown at (1) and (2) in Figure 8.2. Next, if it is clear that the desirability of the *pn*-specified sentence is equivalent to that of another sentence which contains words without *pn* entries (hereafter referred to as a *pn-unspecified* sentence) such as 環境保全への関心が高まる (interest in environmental protection increases), the desirability of the *pn*-unspecified sentence is assumed to be equal to that of the specified sentence via an equivalence relation on desirability, shown at (3). The desirability of the words without *pn* entries can then be estimated based on the desirability of the sentence as a whole and that of the words which do have *pn* entries. New *pn* entries are registered, increasing the size of the *pn* dictionary, shown at (4) and (5). The revised *pn* dictionary is then used to estimate the desirability of further *pn*-unspecified sentences. Alternating between estimating the desirability of sentences, and then words, the cycle is repeated until no new *pn* entries can be registered.

In the remainder of this section, every component of the above cycle, that is, *pn* construction, *pn* transition, and *pn* decomposition are described in detail. Realizing these three components is a novel challenge. We report current experi-

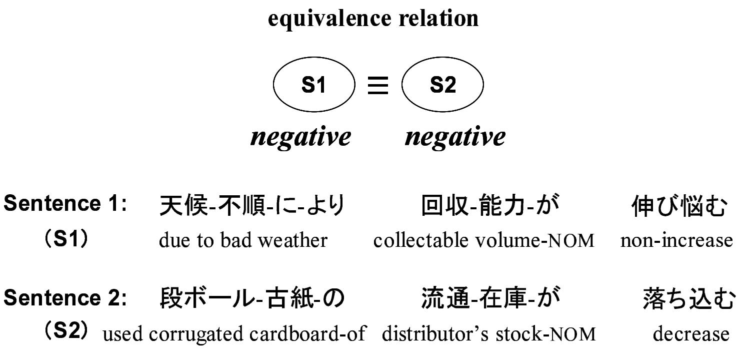


Figure 8.3 Equivalence relation on desirability

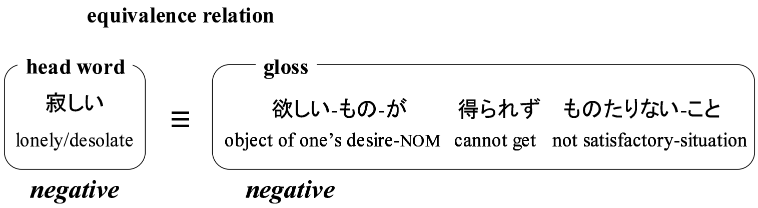


Figure 8.4. A desirability equivalence relation between a head word and its gloss

mental results for the feasibility of each component.

8.4.1 *pn* transition

In the *pn* transition process, the desirability of a *pn*-unspecified sentence is given by another *pn*-specified sentence based on an equivalence relation of desirability. Figure 8.3 shows an example of a desirability equivalence relation between two sentences. The problem to be solved is how to identify the equivalence of desirability between a *pn*-specified sentence and a *pn*-unspecified sentence.

Kobayashi et al. [36] proposed a method for acquiring *pn* entries using an ordinary dictionary. In her approach, *pn* entries are acquired based on the assumption that a desirability equivalence relation exists between a head word and its gloss (Figure 8.4). Since every head word in the dictionary is used in the gloss of another word, given a set of words that are registered in the initial *pn* dictionary, the desirability of entries can be gradually propagated to cover all entries. Although this approach is interesting, the number of *pn* entries that can be acquired using her method is limited.

To resolve this problem, we deal with open text instead of dictionaries as a source for identifying desirability equivalence relations between sentences. For this purpose, we use causal relation instances. As a first approximation, it seems reasonable to assume that:

> a positive event causes another positive event and a negative event causes another negative event.

In other words, when a causal relation exists between two events, the desirability of those events would be equivalent. If this assumption is correct, relation instances whose arguments are in a desirability equivalence relation could be automatically acquired in large quantities using publicly available documents, because causal relation instances can be automatically acquired using the method described in the previous chapter.

Investigation

We investigated how many causal relation instances have arguments with a desirability equivalence relation using 539 causal relation instances from newspaper articles in the recycling domain [28, 29]. We refer to this set as $\mathcal{S}_3$. This set was handcrafted using the same procedure described in Section 5.4 to ensure only correct causal relation instances were used in the investigation. The following are examples of causal relation instances in $\mathcal{S}_3$.

(40) a. *cause* (⟨天候-不順-により 回収-量-が 伸び悩む⟩,
due to bad weather collectable-volume-NOM non-increase
⟨段ボール-古紙-の 流通-在庫-が 落ち込む⟩)
used corrugated cardboard-of distributor's stock-NOM decrease

b. *effect* (⟨ごみ-分別-収集-に 取り組む⟩,
garbage-separating collection-DAT deal with
⟨使用-済み-缶-の 集荷-率-が 高まる⟩)
used can collect-rate-NOM increase

Table 8.3 Number of equivalence relations per causal relation type

relation type	equivalent	total
cause	89	107
effect	30	37
precond	28	70
means	316	325

c. *precond* (⟨ 大量-の 生-ごみ-が 発生する ⟩,
a lot of raw garbage-NOM occur
⟨ 堆肥-に 変えて リサイクル-を 促進する ⟩)
compost change recycle-ACC promote

d. *means* (⟨ 再生-紙-利用-の ガイドライン-を 作成する ⟩,
recycle paper-use guideline-ACC implement
⟨ 古紙-の リサイクル-を 促進する ⟩)
used paper recycle-ACC promote

Results

The total number of causal relation instances for each causal relation type, and number of those instances held in an equivalence relation are shown in Table 8.3. From these results, it is clear that two events held in a *cause* relation or a *means* relation tend also to be held in a desirability equivalence relation. The *effect* relation is very similar. On the other hand, half of the *precond* relation instances do not contain a desirability equivalence relation. There are a few linguistic patterns which appear when two events are not held in a desirability equivalence relation. For example, "someone executes a *positive* volitional action to improve a *negative* non-volitional SOA" is one such pattern, illustrated in (40c) .

As discussed in the previous chapter, we can acquire causal relation instances with a high degree of accuracy from newspaper articles. We should be able to use causal relation instances, at least those with a *cause* relation or a *means* relation, in order to identify desirability equivalence relations between events.

8.4.2 *pn* construction

In this section, we describe an algorithm for sentence desirability estimation using simple construction rules, and a *pn* dictionary. The construction rules estimate the desirability of sub-parts of a sentence based on the dependency structure. Although some work has been done on extracting sentences expressing opinions from documents [35, 53, 75], semantic orientation of the sentences such as *positive* or *negative* has not been considered in detail.

In our method, a sentence is first parsed by the dependency structure analyzer *CaboCha*. Next, desirability of all *bunsetsu* phrases in a sentence is given by that of the head words in each *bunsetsu* phrase using the *pn* dictionary. Function words are ignored. The construction rules are then applied to each dependency relation in turn. Application of the rules to the final dependency relation produces the desirability of the sentence as a whole.

Construction rules: Construction rules were designed based on the insight, obtained from a preliminary analysis, that desirability of sentences depends heavily on their head phrases. For example, consider the sentences shown in (41) where desirability of the sentences is shown in brackets, and the *positive* and *negative* labels to the right of words, refer to a word's desirability. It is assumed that the desirability of the sentences depends on that of the underlined *bunsetsu* phrases.

(41) a. ごみ$_{negative}$-の → $\underline{\text{リサイクル-施設}}$$_{positive}$-を → 建設する$_{neutral}$
garbage / recycling center-ACC / build
⟨*positive*⟩

b. 深刻-化$_{negative}$-する → ごみ-問題$_{negative}$-に → $\underline{\text{対処する}}$$_{positive}$
escalation / garbage problem / address
⟨*positive*⟩

c. 再生$_{positive}$-が → $\underline{\text{難しい}}$$_{negative}$ → 材料$_{neutral}$-を → 含む$_{neutral}$
recycle-NOM / hard / material-ACC / include
⟨*negative*⟩

（"→" indicates to a dependency relation）

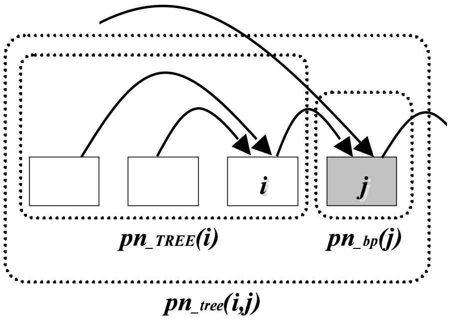


Figure 8.5 $pn_{\_bp}(j)$, $pn_{\_TREE}(i)$, $pn_{\_tree}(i,j)$ in the construction process

Table 8.4 pn construction table

p: *positive* n: *negative* e: *neutral*

		$pn_{\_bp}(j)$				
		p	n	$plus$	$minus$	e
	p	p	n	p	n	p
$pn_{\_TREE}(i)$	n	p	n	n	p	n
	e	p	n	e	e	e

Given a dependency relation between two *bunsetsu* phrases i and j, the insight described above suggests that the desirability of a dependency relation depends strongly on the modified *bunsetsu* phrase j rather than modifier *bunsetsu* phrase i. We designed the construction rules based on this insight, as shown in equation (8.2).

$$pn_{\_tree}(i,j) = \begin{cases} pn_{\_bp}(j) & \text{if } pn_{\_bp}(j) = \ positive \mid negative \\ pn_{\_TREE}(i) & \text{otherwise} \end{cases} \tag{8.2}$$

where *bunsetsu* phrase i modifies j (referred to as $rel(i,j)$) , $pn_{\_bp}(j)$ refers to the desirability of j, $pn_{\_TREE}(i)$ refers to the desirability of the largest possible sub-tree whose root node is i, and $pn_{\_tree}(i,j)$ refers to the desirability of a sub-tree which consists of *bunsetsu* phrase j and the largest possible sub-tree whose root node is i (see also Figure 8.5). To summarize, foreach i (the beginning of the sentence ... the end of the sentence)

① Estimate $pn_{\_TREE}(i)$ using the function $Eval_{\_TREE}(i)$

② If i is located at the end of the sentence then exit

③ Get *bunsetsu* phrase j which is modified by *bunsetsu* phrase i

④ Get $pn_{\_bj}(j)$ using the *pn* dictionary

⑤ Get $pn_{\_tree}(i,j)$ using the *pn* construction table and register $pn_{\_tree}(i,j)$ in the list $pn_{\_list}(i)$.

end

Figure 8.6 *pn* construction algorithm

function $Eval_{\_TREE}(i)$
  if $pn_{\_list}(i) = \phi$ then get $pn_{\_bj}(i)$ using the *pn* dictionary and return it
  else return the tail element which is either positive or negative in $pn_{\_list}(i)$
  end
end

Figure 8.7 $Eval_{\_TREE}(i)$

the desirability of a sub-tree $pn_{\_tree}(i,j)$ can be estimated from the combination of $pn_{\_TREE}(i)$ and $pn_{\_bp}(j)$ shown in Table 8.4, We call Table 8.4 a *pn* construction table.

Algorithm: Using the *pn* dictionary and the *pn* construction rules (table) described above, the desirability of a sentence is estimated according to the algorithm shown in Figure 8.6. In step ⑤ in Figure 8.6, the desirability value $pn_{\_tree}(k,i)$ ($\{\forall k | rel(k,i)\}$) are stored in $pn_{\_list}(i)$ in the same order the *bunsetsu* phrases appear in the sentence. $pn_{\_list}(i)$ is used in evaluating $Eval_{\_TREE}(i)$. We use the function $Eval_{\_TREE}(i)$ shown in Figure 8.7. After running the algorithm, the output returned from $Eval_{\_TREE}$(the end of the sentence) is the desirability of the input sentence. The algorithm is very simple, however, it can handle input sentences of arbitrary length and is robust since it is based only on the dependency structures of input sentences.

Table 8.5 Accuracy of *pn* construction

	Recall		Precision	
positive	0.90	(656/725)	0.99	(656/665)
negative	0.87	(104/119)	0.79	(104/132)
neutral	0.60	(12/ 20)	0.18	(12/ 67)

Evaluation

The details of the evaluation of our construction algorithm were as follows.

- The data were the 432 samples（864 phrases）in $\mathcal{S}_3$ used in Section 8.4.1, which contain either a *cause* relation or a *means* relation.
- We used the *pn* dictionary which contains all the words in the 864 phrases.
- Two negative expressions（“-ない” and “-にくい”）were decomposed into independent *bunsetsu* phrases.

The results are shown in Table 8.5. It is observed that our algorithm for *pn* construction performed well except in the case of *neutral*. In particular, precision for *positive* is almost perfect. Despite our algorithm being simple, it should be possible to apply *pn* construction to sentences in newspaper articles.

Error analysis revealed several types of sentences which are not compatible with our algorithm. One such is sentences including ellipses such as sentence (42) . In the sentence (42) the *bunsetsu* phrase 値段-の（price）is an elliptical expression whose desirability value is negative. Another is sentences including adnominal phrases such as sentence (43) . In sentence (43) the *bunsetsu* phrase 再-使用-可能-な（reuse-possible）is an adnominal phrase. The fact that these elements, elliptical phrases and adnominal phrases, are not considered in the current construction algorithm results in errors.

- ellipses

(42) 再生-紙$_{positive}$-の（$\underline{\text{値段}_{negative}\text{-の}}$）方$_{neutral}$-が 上質-紙$_{neutral}$-より 高い$_{plus}$

recycling paper price than quality paper high cost
correct=⟨negative⟩ output=⟨positive⟩

- adnominal phrases

(43) 再-使用-可能$_{positive}$-な 包装-紙$_{negative}$-の 普及$_{plus}$-を 目指す
reuse-possible packaging paper popularize-ACC aim
correct=⟨positive⟩ output=⟨negative⟩

8.4.3 *pn* decomposition

In the decomposition process, desirability of words which are not registered in the *pn* dictionary（we call these target words）are estimated using the desirability of the sentence as a whole and of words in the sentence with *pn* entries. New *pn* entries can then be registered and the size of the *pn* dictionary increases.

This process is the reverse of the construction process. We use Table 8.6 as a decomposition table in place of Table 8.4. If *bunsetsu* phrase i is the nearest modifying *bunsetsu* phrase to *bunsetsu* phrase j , we estimate $pn_{\_bp}(i)$ from $pn_{\_TREE}(j)$ and $pn_{\_bp}(j)$ （see also Figure 8.8）. Here, we assume that:

$$pn_{\_bp}(i) = pn_{\_TREE}(i) \tag{8.3}$$

The decomposition rules shown in Table 8.6 are applied to each *bunsetsu* phrase in turn starting from the end of the sentence and working backwards towards the beginning. If the value of $pn_{\_TREE}(i)$ cannot be evaluated using the decomposition table, the process cannot proceed and is therefore terminated.

Evaluation

The results of our experiment are shown in Table 8.7 and Table 8.8. The left-hand side of Table 8.7 shows decomposition accuracy for each desirability type, and Table 8.8 shows examples of these. From Table 8.7, it is found that our algorithm makes a lot of mistakes when estimating the desirability of words whose correct value are *neutral*. Then we try to run the *pn* decomposition process with additional constraint below:

Table 8.6 pn decomposition table

p: *positive* n: *negative* e: *neutral*

		$pn_{\_bp}(j)$				
		p	n	*plus*	*minus*	e
	p	–	–	p	n	p
$pn_{\_TREE}(j)$	n	–	–	n	p	n
	e	–	–	e	e	e

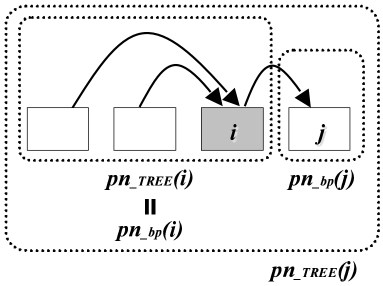


Figure 8.8 $pn_{\_TREE}(j)$, $pn_{\_bp}(j)$, $pn_{\_TREE}(i)$ in the decomposition process

> The pn construction process is run using the *neutral* desirability value of a target word instead of the estimated desirability value from the decomposition process. If the desirability of the sentence obtained is correct, the target word is rejected（i.e. not registered in the pn dictionary）.

The right-hand side is the accuracy which is obtained using the above modified procedure. The modified procedure has lower coverage but higher accuracy than the original procedure. The following sentence （44） shows an incorrect sample. The word 家庭-ゴミ is the target word.

(44) 回収$_{positive}$-を 家庭-ゴミ【$_{target}$】-に 広げる$_{neutral}$ ⟨*positive*⟩
collection-ACC domestic garbage extend
correct=⟨negative⟩ output=⟨positive⟩

Compared to the construction process, the decomposition process is less ac-

Table 8.7 Accuracy of the decomposition process

	original		modified procedure considered	
positive	0.53	169/ 314	0.67	147/218
negative	0.67	48/ 72	0.82	42/ 51
neutral	0.73	11/ 15	0.75	3/ 4

Table 8.8 Examples of acquired *pn* entries

positive	分別-回収	再-利用-率	循環-型-社会
	separate collection	proportion recycled	recycling-based society
negative	伸び悩む	生-ごみ-廃棄-物	過剰-包装
	non-increase	raw garbage	over packaging

curate. In order to realize the *pn* entry acquisition cycle, the accuracy of the decomposition process must be increased.

8.5 Summary

We are currently attempting to utilize causal knowledge for desirability estimation. Our aim is to acquire lexical knowledge about desirability using equivalence relations based on desirability of events. The causal relation instances discussed in previous chapters are used to obtain these equivalence relations. The decision to adopt this approach is based on the assumption that a positive event causes another positive event, and a negative event causes another negative event.

From this investigation, it is clear that two events held in a *cause* relation or a *means* relation tend also to be held in an equivalence relation based on desirability. As discussed in previous chapters, we can acquire causal relation instances with high accuracy from newspaper articles. We expect to be able to use causal relation instances, at least instances of *cause* relations and *means* relations, in order to obtain equivalence relations based on desirability between events.

Chapter 9

Conclusion

9.1 Summary

In this thesis, we described our approach to automatic knowledge acquisition of *causal relations* from a document collection. We considered 4 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 including the word ため led to the following findings:

- Accurate parsing is required in order to correctly identify different structures in complex sentences. Using existing dependency structure analyzers, our general committee-based framework improved performance against that of any of the individual committee members （Chapter 4）.
- The pairs of subordinate and matrix clauses indicating each different event extracted from ため-complex sentences can be classified into the four relation types — *cause*, *effect*, *precond* and *means* — with a precision of 85% or more （Chapter 5）.
- 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

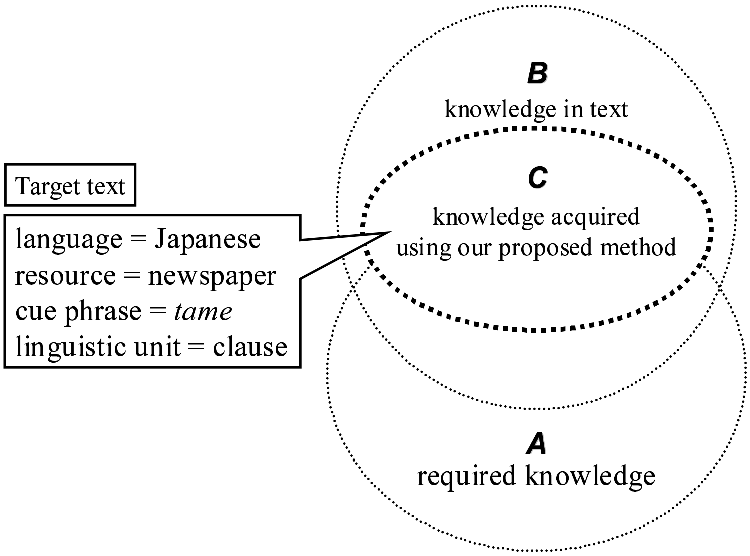


Figure 9.1 Target set dealt with in this work

relations could potentially be acquired from one year of Japanese newspaper articles（Chapter 7）.

- It is clear that two events held in a *cause* relation or a *means* relation tend also to be held in an equivalence relation based on desirability. This relationship is useful for acquiring desirability lexical knowledge（Chapter 8）.

9.2 Future work

As shown in Figure 9.1, in this work we have dealt with only a small subset of all the textually encoded knowledge potentially available in the world. The research effort must be continued in order to clarify the scales and overlaps of the sets shown in Figure 9.1.

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 inter-sentential 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 thesis is not specific to Japanese. We want to investigate applications to other languages such as English. As described in Section 2.2.3, 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 as mentioned in Section 5.3.1. 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.

Co-reference and unneccesary modifiers: As described in Section 7.3, there are some constituents in causal relation instances such as ellipses and pronouns which render the instances incomplete. Techniques for co-reference (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.

References

[1] J. F. Allen. *Recognizing intentions from natural language utterances. In M. Brady and R.C. Berwick（Eds.）, Computational models of discourse.* MIT Press, 1983.

[2] J. F. Allen. *Natural Language Understanding.* Benjamin/Cummings, 1995.

[3] B. Altenberg. Causal linking in spoken and written English. *Studia Linguistica*, 38:1, 1984.

[4] S. Arita. Linguistics on causality. *Gekkan Gengo*, 25(5):20–23, 1996.（in Japanese）.

[5] E. Brill and J. Wu. Classifier combination for improved lexical disambiguation. In *Proc. of The 17th COLING*, pages 191–195, 1998.

[6] Britannica. Britannica CD98 multimedia edition, 1998.

[7] Information Technology Promotion Agency Technical Center. *IPAL: IPA Lexicon of the Japanese Language for Computers.* Information-technology Promotion Agency, 1986.

[8] E. Charniak. Statistical parsing with a context-free grammar and word statistics. In *Proc. of The AAAI*, pages 598–603, 1997.

[9] M. J. Collins. A new statistical parser based on bigram lexical dependencies. In *Proc. of The 34th ACL*, pages 184–191, 1996.

[10] M. J. Collins. Three generative, lexicalised models for statistical parsing. In *Proc. of The 35th ACL*, pages 16–23, 1997.

[11] T. Ehara. Estimating the consistency of Japanese dependency relations based on maximum entropy modeling. In *Proc. of The 4th Annual Meeting of The Association of Natural Language Processing*, pages 382–385, 1998.（in Japanese）.

[12] C. Fellbaum. *WordNet: An Electronic Lexical Database.* The MIT Press, 1998.

[13] J. G. Fiscus. A post-processing system to yield reduced word error rates: Recognizer output voting error reduction（ROVER）. In *EuroSpeech*, 1997.

[14] R. Frederking and S. Nirenburg. Three heads are better than one. In *Proc. of The 4th ANLP*, 1994.

[15] M. Fujio and Y. Matsumoto. Japanese dependency structure analysis based on lexicalized statistics. In *Proc. of The 3rd EMNLP*, pages 87–96, 1998.

[16] D. Garcia. COATIS, an NLP system to locate expressions of actions connected by causality links. In *Proc. of The 10th European Knowledge Acquisition Workshop*, pages 347–352, 1997.

[17] R. Girju and D. Moldovan. Mining answers for causation questions. In *Proc. The AAAI Spring Symposium on Mining Answers from Texts and Knowledge Bases*, 2002.

[18] S. M. Harabagiu and D. I. Moldovan. Textnet - a text-based intelligent system. *Natural Language Engineering*, 3:171–190, 1997.

[19] D. Heckerman, C. Meek, and G. Cooper. A Bayesian approach to causal discovery. Technical report, Microsoft Research Advanced Technology Division, Microsoft Corporation, Technical Report MSR-TR-97-05, 1997.

[20] J. C. Henderson and E. Brill. Exploiting diversity in natural language processing: Combining parsers. In *Proc. of The 1999 Joint SIGDAT Conference on EMNLP and VLC*, pages 187–194, 1999.

[21] J. R. Hobbs. Coherence and co-reference. *Cognitive Science*, 1:67–82, 1979.

[22] J. R. Hobbs. On the coherence and structure of discourse. Technical report, Technical Report CSLI-85-37, Center for The Study of Language and Information, 1985.

[23] J. R. Hobbs, M. Stickel, D. Appelt, and P. Martion. Interpretation as abduction. *Artificial Intelligence*, 63:69–142, 1993.

[24] T. Ichikawa. *Introduction to tyle theory for Japanese education.* Education, 1978.（in Japanese）.

[25] R. Iida, K. Inui, H. Takamura, and Y. Matsumoto. Incorporating contextual cues in trainable models for coreference resolution. In *Proc. of The EACL Workshop: The Computational Treatment of Anaphora*, 2003.

[26] S. Ikehara, M. Miyazaki, S. Shirai, A. Yokoo, H. Nakaiwa, K. Ogura, Y. Ooyama, and Y. Hayashi. *Goi-Taikei - A Japanese Lexicon.* Iwanami Shoten, 1997.

[27] S. Ikehara, S. Shirai, A. Yokoo, and H. Nakaiwa. 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*, pages 101–106, Washington DC, 1991.

[28] Nihon Keizai Shimbun Inc. Nihon Keizai Shimbun CD-ROM version.

[29] Nihon Keizai Shimbun Inc. Nikkei Sangyo Shimbun CD-ROM version.

[30] T. Inui and K. Inui. Committee-based decision making in probabilistic partial parsing. In *Proc. of The 18th International Conference on Computational Linguistics*, pages 347–354, 2000.

[31] L. M. Iwanska and S. C. Shapiro. *Natural Language Processing and Knowledge Representation - Language for Knowledge and Knowledge for Language.* The MIT Press, 2000.

[32] K. Jensen, G. E. Heidorn, and editors. S. D. Richardson. *natural language processing: The PLNLP Approach.* Kluwer Academic Publishers, 1993.

[33] H. Kanayama, K. Torisawa, Y. Mitsuishi, and J. Tsujii. Statistical dependency analysis with an HPSG-based Japanese grammar. In *Proc. of The NLPRS*, pages 138–143, 1999.

[34] C. S. G. Khoo, S. Chan, and Y. Niu. Extracting causal knowledge from a medical database using graphical patterns. In *Proc. of The 38th. Annual Meeting of The Association for Computational Linguistics*（*ACL2000*）, pages 336–343, 2000.

[35] N. Kobayashi, K. Inui, Y. Matsumoto, K. Tateishi, and S. Fukushima. Collecting evaluative expressions by a text mining technique. In *IPSJ SIG NOTE* (*2002-NL-154*), 2003. (in Japanese).

[36] N. Kobayashi, T. Inui, and K. Inui. Dictionary-based acquisition of the lexical knowledge for p/n analysis. In *Japanese Society for Artificial Intelligence, SLUD-33*, 2001. (in Japanese).

[37] T. Kudo and Y. Matsumoto. Japanese dependency analysis based on Support Vector Machines. In *Empirical Methods in Natural Language Processing and Very Large Corpora*, pages 18–25, 2000.

[38] T. Kudo and Y. Matsumoto. Japanese dependency analysis using cascaded chunking. In *Proc. of The 6th. Conference on Natural Language Learning* (*CoNLL*), 2003.

[39] S. Kurohashi and M. Nagao. KN Parser : Japanese dependency/case structure analyzer. In *Proc. of The International Workshop on Sharable Natural Language Resources*, pages 48–55, 1994.

[40] S. Kurohashi and M. Nagao. Building a Japanese parsed corpus while improving the parsing system. In *Proc. of NLPRS*, pages 151–156, 1997.

[41] D. Lenat. Cyc: A large-scale investment in knowledge infrastructure. *Communications of the ACM*, 38(11). See also, `http://www.cyc.com/cyc/`.

[42] D. J. Litman and J. F. Allen. A plan recognition model for subdialogues in conversations. *Cognitive Science*, 11:163–200, 1987.

[43] H. Liu, H. Lieberman, and T. Selker. GOOSE: A goal-oriented search engine with commonsense. In *Proc. of The 2nd. International Conference on Adaptive Hypermedia and Adaptive Web Based Systems* (*AH2002*), pages 253–263, 2003.

[44] H. Liu, H. Lieberman, and T. Selker. A model of textual affect sensing using real-world knowledge. In *Proc. of The International Conference on Intelligent User Interfaces*, pages 125–132, 2003.

[45] B. T. Low, K. Chan, L. L. Choi, M. Y. Chin, and S. L. Lay. 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*）, pages 114–123, 2001.

[46] W. C. Mann and S. A. Thompson. Rhetorical structure theory: A theory of text organization. In *USC Information Sciences Institute, Technical Report ISI/RS-87-190*, 1987.

[47] D. Marcu. The rhetorical parsing of natural language texts. In *Proc. of ACL97/EACL97*, pages 96–103, 1997.

[48] D. Marcu. An unsupervised approach to recognizing discourse relations. In *Proc. of The 40th. Annual Meeting of The Association for Computational Linguistics*（*ACL2002*）, pages 368–375, 2002.

[49] T. Masuoka. *Complex sentence*. Kuroshio, 1997.（in Japanese）.

[50] T. Masuoka and Y. Takubo. *Fundamental Japanese grammar(revised version*）. Kuroshio, 1992.（in Japanese）.

[51] Y. Matsumoto, A. Kitauchi, T. Yamashita, Y. Hirano, H. Matsuda, and M. Asahara. *Japanese Morphological Analyzer ChaSen Users Manual version 2.0.* Technical Report NAIST-IS-TR990123, Nara Institute of Science and Technology Technical Report, 1999.

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

[53] S. Murano and S. Sato. Automatic extraction of subjective evaluative sentences using phrase templates. In *Proc. of The 9th Annual Meeting of The Association for Natural Language Processing*, pages 67–70, 2003.（in Japanese）.

[54] T. Nagae, G. Mochizuki, K. Shirai, and A. Shimazu. Analyses of the relationship between a product concept and evaluative sentences describing the

product. In *Proc. of The 8th Annual Meeting of The Association for Natural Language Processing*, pages 583–586, 2002. (in Japanese).

[55] M. Nagano. *A reveiw of style theory*. Asakura, 1986. (in Japanese).

[56] Y. Nakagawa and T. Mori. Pragmatic analysis of aspect morphemes in manual sentences in Japanese. *Natural Language Processing*, 2(4):19–35, 1995. (in Japanese).

[57] H. Nakaiwa and S. Ikehara. Intrasentential resolution of Japanese zero pronouns in a machine translation system using semantic and pragmatic constraints. In *Proc. of The 6th TMI*, pages 96–105, 1995.

[58] Y. Nitta. *Aspects of adverbial expressions*. Kuroshio, 2002. (in Japanese).

[59] U.S. National Library of Medicine. The MEDLINE database, 2001. See also, `http://www.ncbi.nlm.nih.gov/PubMed/`.

[60] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, 1988.

[61] J. Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge Universiy Press, 2000.

[62] T. Pedersen. A simple approach to building ensembles of naive Bayesian classifiers for word sense disambiguation. In *Proc. of The NAACL*, pages 63–69, 2000.

[63] J. Pei, J. Han, B. Mortazavi-Asi, H. Pinto, Q. Chen, U. Dayal, and M. C. Hsu. PrefixSpan: mining sequential patterns efficiently by prefix projected pattern growth. In *Proc. of 1st. Conference of Data Enginnering* (*ICDE2001*) , pages 215–226, 2001.

[64] D. Poole. Average-case analysis of a search algorithm for estimating prior and posterior probabilities in Bayesian networks with extreme probabilities. In *Proc. The 13th IJCAI*, pages 606–612, 1993.

[65] RWC. RWC text corpus 2nd edition, Iwanami Japanese dictionary tagged/morphological data 5th edition, 1998.

[66] Carberry S. *PLAN RECOGNITION IN NATURAL LANGUAGE DIALOGUE*. The MIT Press, 1990.

[67] H. Satou, K. Kasahara, and K. Matsuzawa. Rertrieval [sic] of simplified causal knowledge in text and its application. In *Technical report of IEICE, Thought and Language*, 1999. (in Japanese).

[68] S. Satou and M. Nagao. Toward memory-based translation. In *Proc. of The 13th. International Conference on Computational Linguistics* (*COLING90*), pages 247–252, 1990.

[69] R. Schank and R. Abelson. *Scripts, Plans, Goals and Understanding: An Inquiry into Human Knowledge Structures*. Lawrence Erlbaum Associates, 1977.

[70] R. Schank and C. Riesbeck. *Inside Computer Understanding: Five Programs Plus Minitures*. Lawrence Erlbaum Associates, 1981.

[71] K. Shirai, K. Inui, T. Tokunaga, and H. Tanaka. An empirical evaluation on statistical parsing of Japanese sentences using lexical association statistics. In *Proc. of The 3rd EMNLP*, pages 80–87, 1998.

[72] P. Singh. The public acquisition of commonsense knowledge. In *Proc. of AAAI Spring Symposium on Acquiring Linguistic Knowledge for Information Access*, 2002.

[73] P. Singh, T. Lin, E. T. Mueller, G. Lim, T. Perkins, and W. L. Zhu. 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*, 2002.

[74] D. G. Stork. Character and document research in the Open Mind Initiative. In *Proc. of International Conference on Document Analysis and Recognition*, pages 1–12, 1999.

[75] K. Tateishi, Y. Ishiguro, and S. Fukushima. Opinion information retrieval from the Internet. In *IPSJ SIG NOTE* (*2001-NL-144*), pages 75–82, 2002. (in Japanese).

[76] A. Terada. *A Study of Text Mining Techniques Using Natural Language Processing*. PhD thesis, Tokyo Institute of Technology, 2003.
（in Japanese）.

[77] K. Torisawa. Automatic extraction of 'commonsense' inference rules from corpora. In *Proc. of The 9th Annual Meeting of The Association for Natural Language Processing*, pages 318–321, 2003.（in Japanese）.

[78] K. Uchimoto, S. Sekine, and H. Isahara. Japanese dependency structure analysis based on maximum entopy models. In *Proc. of The 13th EACL*, pages 196–203, 1999.

[79] H. van Halteren, J. Zavrel, and W. Daelemans. Improving data driven word-class tagging by system combination. In *Proc. of The 17th COLING*, 1998.

[80] V. N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 1995.

[81] E. M. Voorhees and D. K. Harman. The Ninth Text Retrieval Conference（TREC9）, 2001. See also, `http://trec.nist.gov`.

[82] H. Yamada, T. Kudo, and Y. Matsumoto. Japanese named entity extraction using Support Vector Machines. *Transactions of IPSJ*, 43(1):43–53, 2002.

[83] T. Yokoi. The EDR electronic dictionary. *Communications of The ACM*, 38(11):42–44, 1995.

Appendix

A Examples of causal relation instances

The following are additional examples of causal relation instances to supplement to those shown in Chapter 5. The relation instances are grouped according to the connective marker used in each original sentence.

ため

1. *cause*(⟨ ラッシュアワーと重なる ⟩, ⟨ 通勤、通学客計六万五千人の足に影響が出る ⟩)
2. *cause*(⟨ 悪天候が続く ⟩, ⟨ 予定より五日遅れる ⟩)
3. *cause*(⟨ 上空に冷たい空気が流れこむ ⟩, ⟨ 各地で朝の最低気温が軒並み平年を下回る ⟩)
4. *cause*(⟨ 気温が平年より高めに推移する ⟩, ⟨ コートなど重衣料の売り上げが伸び悩む ⟩)
5. *cause*(⟨ カードに被害届けが出る ⟩, ⟨ この時は、未遂に終わる ⟩)
6. *cause*(⟨ フィルムや半導体など日米間の摩擦の火種がくすぶり続ける ⟩, ⟨ 米側の出方に警戒感も出る ⟩)
7. *cause*(⟨ スチール住宅は割安で耐震性に優れる ⟩, ⟨ 米国で需要が急増する ⟩)
8. *cause*(⟨ バッテリーも小型電池で済む ⟩, ⟨ 電子手帳などへのポケベル機能組み込みも容易になる ⟩)
9. *cause*(⟨ 気温が平年より高めに推移する ⟩, ⟨ コートなど冬物衣料の売り上げが伸び悩む ⟩)
10. *cause*(⟨ 原料価格は依然として高水準にある ⟩, ⟨ メーカー各社の採算は厳しくなる ⟩)
11. *cause*(⟨ 合理化が遅れる ⟩, ⟨ 販売費・一般管理費が増加気味なことも利益の足を引っ張る ⟩)
12. *cause*(⟨ 阪神大震災の被災地中心に住宅のリフォームが増える ⟩, ⟨ 住居に対する支出は、増加する ⟩)
13. *cause*(⟨ 気温が比較的高めに推移する ⟩, ⟨ 衣料品の売り上げが伸び悩む ⟩)
14. *cause*(⟨ 消防車三十四台が出動、約六時間燃え続ける ⟩, ⟨ 付近は一時騒然となる ⟩)

15. $cause$(⟨ 食料品は、猛暑の影響でビールや清涼飲料水の出荷が増加する ⟩, ⟨ 生産指数は一〇九・二と前月比九・九%上昇する ⟩)
16. $cause$(⟨ 新築物件の平均価格は四千万円強で弱含みで推移する ⟩, ⟨ 中古物件に割高感が出る ⟩)
17. $cause$(⟨ 生産見通しは産地の降雨や気温、自然災害の発生などによって目まぐるしく変化する ⟩, ⟨ 天候相場期には変動幅が大きくなる ⟩)
18. $cause$(⟨ 秋口にかけては出荷頭数が増える ⟩, ⟨ 豚価は下がる ⟩)
19. $cause$(⟨ 船価は回復基調にある ⟩, ⟨ 押し目には個人などの買いも入る ⟩)
20. $cause$(⟨ 診療報酬は治療項目ごとに医療保険の支払額が決まる ⟩, ⟨ 病院経営の決め手はコストの管理にある ⟩)
21. $cause$(⟨ 発信機の取り付けを誤る ⟩, ⟨ 盗聴は失敗する ⟩)
22. $cause$(⟨ 出荷が伸び悩む ⟩, ⟨ 在庫は減らない ⟩)
23. $cause$(⟨ 年後半は短期金利が低水準で安定推移する ⟩, ⟨ 取引が伸び悩む ⟩)
24. $cause$(⟨ 不動産部門で、採算割れ物件の処理が一巡する ⟩, ⟨ 経常損益は黒字転換する ⟩)
25. $cause$(⟨ 利食い売りも断続的に出る ⟩, ⟨ ダウ工業株三十種平均は伸び悩む ⟩)
26. $cause$(⟨ 夕方のラッシュ時に重なる ⟩, ⟨ 帰宅途中の会社員ら約七万五千人が影響を受ける ⟩)
27. $effect$(⟨ ＣＡＴＶ網を活用する ⟩, ⟨ 市内電話部分での定額制導入など安価な料金でのサービスが可能となる ⟩)
28. $effect$(⟨ これまでは基本設計データを千葉から郵送でボンベイに送る ⟩, ⟨ 時間と手間がかかる ⟩)
29. $effect$(⟨ 子会社の業務範囲を当初はかなり限定する ⟩, ⟨ 収益見通しは厳しくなる ⟩)
30. $effect$(⟨ 短期金融市場で日銀が緩めの資金調節をする ⟩, ⟨ 下げ渋る ⟩)
31. $effect$(⟨ 証券会社が利食いの円買いを断続的に入れる ⟩, ⟨ 円は上昇に転じる ⟩)
32. $effect$(⟨ 駅やバス通りに近く、人通りの多い路面に出店する ⟩, ⟨ 店舗面積は小ぶりになる ⟩)
33. $effect$(⟨ ドイツ系企業がマルクを買う ⟩, ⟨ ドルの上昇は限る ⟩)
34. $effect$(⟨ ドル建てで出荷する ⟩, ⟨ 円高も重くのしかかる ⟩)

35. *effect*(⟨ 映画の音をそのまま用いる ⟩, ⟨ 妙な臨場感がある ⟩)
36. *effect*(⟨ 皮革製品の材料として乱獲する ⟩, ⟨ 絶滅の危機にひんする ⟩)
37. *effect*(⟨ 工場でパネルをあらかじめ組み立てる ⟩, ⟨ 工期も約三十日短縮する ⟩)
38. *effect*(⟨ 海上輸送運賃に割高な国内のトラック運賃を上乗せする ⟩, ⟨ 荷主の物流コストがかさむ ⟩)
39. *effect*(⟨ 仮想企業は案件ごとに社外スタッフと契約を結ぶ ⟩, ⟨ 人件費など固定費負担が少なくて済む ⟩)
40. *effect*(⟨ 海外生活の場数を踏む ⟩, ⟨ 交渉術にもたける ⟩)
41. *effect*(⟨ 各社が一斉に増産する ⟩, ⟨ ミニコンポは世界的な供給過剰に陥る ⟩)
42. *effect*(⟨ 観光客は下呂や伊勢など周辺の観光地に泊まる ⟩, ⟨ 名古屋は通過点になる ⟩)
43. *effect*(⟨ 原材料ごとに仕入れ先を見直す ⟩, ⟨ 原価率が改善する ⟩)
44. *effect*(⟨ 宗教を隠れみのにする ⟩, ⟨ 宗教法人法の改正が浮上する ⟩)
45. *effect*(⟨ 初心者は力任せに引き抜こうとする ⟩, ⟨ 軟らかい根元部分が切れる ⟩)
46. *effect*(⟨ 石油会社など輸入企業も先物のドル買い予約を入れる ⟩, ⟨ 円は伸び悩む ⟩)
47. *effect*(⟨ 南アフリカ共和国の有力アルミ精錬所が増産に入る ⟩, ⟨ 原料のアルミナ需給は引き締る ⟩)
48. *effect*(⟨ 農家はより収益性の高い作物に転作する ⟩, ⟨ 供給不足が深刻となる ⟩)
49. *precond*(⟨ ＰＨＳの利用者が伸び悩む ⟩, ⟨ 契約を当面、棚上げする ⟩)
50. *precond*(⟨ ドルが対マルクで急落する ⟩, ⟨ 円をすぐに買い戻す ⟩)
51. *precond*(⟨ 中国への進出計画が相次ぐ ⟩, ⟨ 上海に拠点を設ける ⟩)
52. *precond*(⟨ 配管室の白煙が増加する ⟩, ⟨ 手動で炉を停止する ⟩)
53. *precond*(⟨ 仙台地区を中心に高水準の売り上げを維持する ⟩, ⟨ 新規出店を加速する ⟩)
54. *precond*(⟨ 入会者が毎月五、六十人にとどまる ⟩, ⟨ 新カードで顧客拡大を目指す ⟩)
55. *precond*(⟨ 関連会社の脱税も発覚する ⟩, ⟨ 税務当局はマブロジ氏を逮捕、起訴する ⟩)

56. $precond$(⟨ パソコンの個人需要が急増する ⟩, ⟨ パソコン売り場を集客の核として位置付ける ⟩)
57. $precond$(⟨ ビル関連の需要が激減する ⟩, ⟨ 住宅向けの需要を各社で奪い合う ⟩)
58. $precond$(⟨ 機械化で事務量が減少する ⟩, ⟨ 一般職の採用を六六%減らす ⟩)
59. $precond$(⟨ 引き合いが活発で半年近い受注残を抱える ⟩, ⟨ 月産能力を倍増する ⟩)
60. $precond$(⟨ 環境装置部門が順調に伸びる ⟩, ⟨ 需要の多い首都圏に本格的な生産拠点を置く ⟩)
61. $precond$(⟨ 旧施設が老朽化、手狭になる ⟩, ⟨ 移転建築する ⟩)
62. $precond$(⟨ 琵琶湖の水位が低下、冬季渇水の懸念が強まる ⟩, ⟨ 建設省近畿地方建設局は渇水対策本部を設置する ⟩)
63. $precond$(⟨ 先安観が後退する ⟩, ⟨ 東南アジア諸国も徐々に買い付けを増やし始める ⟩)
64. $precond$(⟨ 断水時は水圧が下がる ⟩, ⟨ タンク内の水を手動ポンプでくみ上げる ⟩)
65. $precond$(⟨ 店は無事だったが神戸の自宅が全壊する ⟩, ⟨ 北区鹿の子台の仮設住宅に移り住む ⟩)
66. $precond$(⟨ 同市は平たん地が狭く、中心部の交通渋滞が慢性化する ⟩, ⟨ バス利用の促進、駐車場案内システムの整備を進める ⟩)
67. $precond$(⟨ 髪質が違う ⟩, ⟨ 外国人と日本人で染料も変える ⟩)
68. $precond$(⟨ 来訪者が高齢化する ⟩, ⟨ 慰霊塔前の階段を一部手すり付きのスロープに改修する ⟩)
69. $means$(⟨ 通信インフラ整備を急ピッチで進める ⟩, ⟨ ＴＯＴは慢性的な電話不足を解消する ⟩)
70. $means$(⟨ 幹線道路や鉄道など緊急時の輸送ルートを県市別に規定する ⟩, ⟨ 救援物資の輸送をスムーズに行う ⟩)
71. $means$(⟨ 歩道に花の苗を設置する ⟩, ⟨ 街並みを美しくする ⟩)
72. $means$(⟨ パトロールを強化する ⟩, ⟨ ゲレンデでの衝突事故を防ぐ ⟩)
73. $means$(⟨ 榧根教授はスリランカ西方の海水表面の温度の変化を調査する ⟩, ⟨ この降雨量の変化の原因を探る ⟩)
74. $means$(⟨ 企業も様々な努力をする ⟩, ⟨ 集客力を高める ⟩)

75. $means(\langle$ 中間車両で約二十トン、先頭車両でも約三十トンと超軽量にする $\rangle$, $\langle$ 車体をまるごと浮かせる $\rangle)$
76. $means(\langle$ 同農協は施設内に新洗浄設備を設置する $\rangle$, $\langle$ ミネラル水洗浄を始める $\rangle)$
77. $means(\langle$ 職場での根回しには気を配る $\rangle$, $\langle$ 仕事を円滑に続ける $\rangle)$
78. $means(\langle$ ごみの回収・処理の有料化を検討する $\rangle$, $\langle$ 事業者、一般家庭にごみ減量化・資源化の努力を促す $\rangle)$
79. $means(\langle$ 特殊コーティングも採用する $\rangle$, $\langle$ 偽造を防ぐ $\rangle)$
80. $means(\langle$ 屋上にはヘリポートを設置する $\rangle$, $\langle$ 救命救急センターとしての機能を強化する $\rangle)$
81. $means(\langle$ 自動化投資を進める $\rangle$, $\langle$ 残業を減らす $\rangle)$
82. $means(\langle$ 水田整備率も五五%から六〇%に引き上げる $\rangle$, $\langle$ 生産基盤整備を進める $\rangle)$
83. $means(\langle$ インターネットにホームページを開設する $\rangle$, $\langle$ 県内の水産業情報をＰＲする $\rangle)$
84. $means(\langle$ 条件付き一般競争入札制度を導入する $\rangle$, $\langle$ 長野市は公共工事の発注で不正を防ぐ $\rangle)$
85. $means(\langle$ 試食会への一般参加者を募集する $\rangle$, $\langle$ 同振興会では味や商品化へのアドバイスを聞く $\rangle)$
86. $means(\langle$ 建設中の南房総広域水道が、通水を開始する $\rangle$, $\langle$ 南房総地域の慢性的な水不足を解消する $\rangle)$
87. $means(\langle$ 各店ともに知恵を絞る $\rangle$, $\langle$ 売上高を増やす $\rangle)$
88. $means(\langle$ 製造工程を見直す $\rangle$, $\langle$ 品質管理を徹底する $\rangle)$
89. $others(\langle$ 五十一社が市内での移転を希望する $\rangle$, $\langle$ 市工連は研究会結成に踏み切る $\rangle)$
90. $others(\langle$ 高橋容疑者はその後も依頼を続ける $\rangle$, $\langle$ 鈴木容疑者は承諾する $\rangle)$
91. $others(\langle$ 新進党が拒否する $\rangle$, $\langle$ 与党は本会議への改正案上程を断念する $\rangle)$
92. $others(\langle$ 男が上着から短銃のようなものを出す $\rangle$, $\langle$ 遠山支店長ら男性行員数人で、取り押さえる $\rangle)$
93. $others(\langle$ 六和機械が台湾で鋳物素材を生産する $\rangle$, $\langle$ 新会社では鋳物素材の内製化も検討する $\rangle)$

94. *others*(⟨ 米国はキューバと国交を断絶する ⟩, ⟨ ビザ発給に難色を示す ⟩)
95. *others*(⟨ 球界に籍をおく ⟩, ⟨ この間の事情には口を閉ざす ⟩)
96. *others*(⟨ 払い込みで国内とユーロ市場で同時に二百二十万株の公募増資をする ⟩, ⟨ 株主への利益配分を増やす ⟩)
97. *others*(⟨ 食品を扱う ⟩, ⟨ 特に衛生面には気を使う ⟩)
98. *others*(⟨ 複数の映画館やホテルなども併設する ⟩, ⟨ 年間千三百万人の集客を見込む ⟩)
99. *others*(⟨ 預金はすべて福井銀に引き継ぐ ⟩, ⟨ 元金、利息は完全に保護する ⟩)
100. *others*(⟨ 和菓子と洋菓子両方を取り扱う ⟩, ⟨ その製造ノウハウを新商品に活用する ⟩)

ので

1. *cause*(⟨ 低コストはここにも響く ⟩, ⟨ 教師陣はつねに不足する ⟩)
2. *cause*(⟨ゴルフ場の標高によって、農薬の使用量が違う ⟩, ⟨ 画一的な基準では現実的でないとの指摘もある ⟩)
3. *cause*(⟨ 加工ミスもない ⟩, ⟨ 住宅の質の向上にもつながる ⟩)
4. *cause*(⟨ 原油高、高金利はコストアップ要因としてはねかえる ⟩, ⟨ 借入金依存の高い中小企業に影響は出る ⟩)
5. *cause*(⟨ エビ餃子や焼売は売値の割に材料費が高く、手間がかかる ⟩, ⟨ コスト割れする ⟩)
6. *effect*(⟨ それらがいっせいに自分たちの主張を始める ⟩, ⟨ 議事もとどこおる ⟩)
7. *effect*(⟨ 兼高帰省もお盆や年末年始に一斉に帰る ⟩, ⟨ 交通がパンクする ⟩)
8. *effect*(⟨ 両ソフトの操作性を引き継ぐ ⟩, ⟨ 導入時のトレーニングが簡単で済む ⟩)
9. *effect*(⟨ それらがいっせいに自分たちの主張を始める ⟩, ⟨ 議事もとどこおりがちになる ⟩)
10. *precond*(⟨ 診療行為が多ければ多いほど収入も多くなる ⟩, ⟨ 薬の多用や過剰診療の誘発といった欠点も指摘する ⟩)
11. *precond*(⟨ 路線価の見直しだけでは課税強化になる ⟩, ⟨ 課税最低限の引き上げなども同時に実施する ⟩)
12. *precond*(⟨ Wさんの状態は座骨神経にも刺激がある ⟩, ⟨ 整形外科の医師にも診察と検査を依頼する ⟩)

13. *precond*(⟨ キリスト教徒もいる ⟩, ⟨ 月一回は金沢の教会に連れて行く ⟩)
14. *precond*(⟨ 主婦ドライバーや初心者などユーザー側からの問い合わせが目立つ ⟩, ⟨ 国内乗用車向けの販売にも力を入れる ⟩)
15. *precond*(⟨ 伝統的な広東人の家庭に育つ ⟩, ⟨ 昼と晩の二食は米を主食にする ⟩)
16. *precond*(⟨ 今年度の予算枠四千八百四十万円に余裕がある ⟩, ⟨ 引き続き助成希望団体を募集する ⟩)
17. *precond*(⟨ 特に便秘はニキビ悪化の原因になる ⟩, ⟨ 日ごろの食事には気をつける ⟩)
18. *others*(⟨ トマトなどは腐る ⟩, ⟨ 保冷庫つきの船が必要となる ⟩)
19. *others*(⟨ 患者の多くが夏は軽井沢に避暑に行く ⟩, ⟨ 七、八月は軽井沢に聖路加診療所を開設する ⟩)
20. *others*(⟨ ウィーン・フィルのバイオリニストが金沢に来る ⟩, ⟨ コンサートに行く ⟩)

れば

1. *cause*(⟨ ひとたび経済成長のプロセスをたどる ⟩, ⟨ 貧困問題もいずれ解消する ⟩)
2. *cause*(⟨ このまま活躍する ⟩, ⟨ 日米欧のゴルフの距離感が縮まる ⟩)
3. *cause*(⟨ 一歩対応を誤る ⟩, ⟨ 自民党内から海部批判が噴き上げる ⟩)
4. *cause*(⟨ 不況が深刻化する ⟩, ⟨ 倒産や破産がさらに増える ⟩)
5. *cause*(⟨ 失敗する ⟩, ⟨ 経営の根幹を揺るがす ⟩)
6. *cause*(⟨ 県南と静岡県西部を結ぶ高規格幹線道路、三遠南信道路ができる ⟩, ⟨ 清水港向けなどの取扱量が増える ⟩)
7. *cause*(⟨ 先進工業国側で景気後退が広がる ⟩, ⟨ 途上国の輸出収入は落ち込む ⟩)
8. *effect*(⟨ その時に公定歩合を下げる ⟩, ⟨ 三カ月小口定期預金や流動性預金の実質金利はマイナスとなる ⟩)
9. *effect*(⟨ 品種を減らす ⟩, ⟨ ライバルに負ける ⟩)
10. *effect*(⟨ 日本で数年働く ⟩, ⟨ ブラジルで家が買える ⟩)
11. *precond*(⟨ 銀行の証券子会社が倒産しそうになる ⟩, ⟨ 救済する ⟩)
12. *precond*(⟨ 株主総会で反対がない ⟩, ⟨ 公正取引委員会などの審査も受けたうえで正式な本認可を申請する ⟩)
13. *precond*(⟨ 為替相場の安定が続く ⟩, ⟨ 各社とも二カ月連続で卸値を下げる ⟩)

14. *others*(⟨ ミネバリという木を使って髪をすく ⟩, ⟨ 頭痛が治る ⟩)
15. *others*(⟨ 言い換える ⟩, ⟨ ナノメートル（１００万分の１ミリ）の単位で鏡面の平滑性を確保する ⟩)
16. *others*(⟨ 相撲界で二十年前の「トレーニングウエアとタキシード」に類することを探す ⟩, ⟨ 高砂部屋のけいこ場風景にとどめを刺す ⟩)
17. *others*(⟨ 今年、何が興味深かったかという ⟩, ⟨ 基本的にはニュース番組だったと思う ⟩)
18. *others*(⟨ 良い時もある ⟩, ⟨ 悪い時もある ⟩)
19. *others*(⟨ アキノ大統領が立候補する ⟩, ⟨ 私は辞退する ⟩)
20. *others*(⟨ 遊び型もいる ⟩, ⟨ 勉強熱心な者もいる ⟩)

たら

1. *cause*(⟨ ふだんの食生活が乱れる ⟩, ⟨ 効果が半減する ⟩)
2. *cause*(⟨ 難民が大量発生する ⟩, ⟨ もちろん、困る ⟩)
3. *cause*(⟨ すき間がある ⟩, ⟨ 雨が漏る ⟩)
4. *cause*(⟨ 飲酒運転がバレる ⟩, ⟨ 困る ⟩)
5. *cause*(⟨ 果実だけでなく木まで凍結する ⟩, ⟨ 被害は来年にも及ぶ ⟩)
6. *effect*(⟨ 設立して十年になったのを機会に、一部コートを屋内施設にする ⟩, ⟨ スクール生が二倍に増える ⟩)
7. *effect*(⟨ 全員が同じ時期に休暇を取る ⟩, ⟨ 仕事がストップする ⟩)
8. *effect*(⟨ ソ連軍が介入する ⟩, ⟨ その後の東欧史は大きく変わる ⟩)
9. *precond*(⟨ 契約で定めた期間が終わる ⟩, ⟨ 必ず土地を返す ⟩)
10. *precond*(⟨ 副作用が出る ⟩, ⟨ 早めに服用を中止する ⟩)
11. *precond*(⟨ シーズンを過ぎる ⟩, ⟨ 飾りを外し、観葉植物とする ⟩)
12. *precond*(⟨ 鉢の表土がやや乾く ⟩, ⟨ 水はたっぷり与える ⟩)
13. *precond*(⟨ 保有株式の時価と簿価の差が拡大する ⟩, ⟨ ある段階で実現益をいったん計上して買い直す ⟩)
14. *others*(⟨ 本の整理をする ⟩, ⟨ 面白いものが出る ⟩)
15. *others*(⟨ ソ連が経済封鎖という強硬措置に出る ⟩, ⟨ 我々も受けて立つ ⟩)
16. *others*(⟨ また行けと言う ⟩, ⟨ 行く ⟩)

17. *others*(⟨ コロラドから一人旅とはどういうことだろうと思う ⟩, ⟨ 少女が言う ⟩)
18. *others*(⟨ 改札口を出る ⟩, ⟨ そのままゴンドラに乗って五分でゲレンデに到着する ⟩)
19. *others*(⟨ 仕事が一段落する ⟩, ⟨ 妻とともにハワイか米国本土へ行く ⟩)
20. *others*(⟨ 柄についていろいろ相談する ⟩, ⟨ その店員が声をひそめる ⟩)

が

1. *cause*(⟨ 市民の暴力追放運動も盛り上がりをみせる ⟩, ⟨ 抗争終結のめどは立つ ⟩)
2. *cause*(⟨ 三十五万台分の増産がでる ⟩, ⟨ これで不足が生じない ⟩)
3. *precond*(⟨ ソ連が国内に不安定要素を抱える ⟩, ⟨ 内戦に突入するような状況とは思う ⟩)
4. *precond*(⟨ 一時的な特殊要因や外部のマイナス要因が増えて調整局面にある ⟩, ⟨ 個人消費は悲観視する ⟩)
5. *precond*(⟨ 造船は中東紛争の長期化や金利高で受注が伸び悩む ⟩, ⟨ 受注残が多くフル生産を続行しない ⟩)
6. *others*(⟨ 円高になる二年ほど前から輸出を始める ⟩, ⟨ 円高で採算が合う ⟩)
7. *others*(⟨ このときは不成立となる ⟩, ⟨ また同じような法案が提出される可能性もある ⟩)
8. *others*(⟨ これまでは駅中心に整備する ⟩, ⟨ これからは自動車交通に対応した街づくりが必要である ⟩)
9. *others*(⟨ 自由で普遍的な人間的価値をもった精神を求める ⟩, ⟨ そのために天皇制「国家」の論理との深い矛盾をはらむ ⟩)
10. *others*(⟨ この男はメガネをかけていない ⟩, ⟨ 同本部では防犯カメラに写った際はだてメガネをかけていない ⟩)
11. *others*(⟨ これまでも独身寮や借り上げ社宅の整備を進める ⟩, ⟨ 福利厚生を一層充実する必要があると判断しない ⟩)
12. *others*(⟨ その後もたびたび外国人労働者が工場を訪れる ⟩, ⟨ 鵜沢さんはその場でみな断らない ⟩)
13. *others*(⟨ イラクにはまだ三十六人の在留邦人が残る ⟩, ⟨ 少なくとも三十三人がこの政府チャーター機で帰国しない ⟩)

14. *others*(⟨ スキー場の新設・増設など設備投資が盛んになる ⟩, ⟨ これに合わせてゲレンデ整備車両の需要も増加しない ⟩)
15. *others*(⟨ 遺体からは現金三十六円や自転車などのカギ三個がみつかる ⟩, ⟨ 遺書や犯行に使った凶器などは発見する ⟩)
16. *others*(⟨ こたつが少しずれる ⟩, ⟨ 居間は室内を物色したり激しく争った様子は見えない ⟩)
17. *others*(⟨ 今年度は市内の大手酒店が一手に販売を引き受ける ⟩, ⟨ 来年度は複数の卸・小売りルートに流す ⟩)
18. *others*(⟨ 女の子らしい遊びをする ⟩, ⟨ 近くの原っぱで男の子相手によくチャンバラなんかもしない ⟩)
19. *others*(⟨ 人工関節や人工骨の需要は高齢化社会を迎えて増える ⟩, ⟨ 骨とのなじみが悪く骨との間に膜ができて人工関節がはずれやすいなど問題がない ⟩)
20. *others*(⟨ 同社には約百人の従業員が勤務する ⟩, ⟨ 夜間は無人でない ⟩)

のに

1. *cause*(⟨ ＧＮＰの成長率は鈍る ⟩, ⟨ 賃金は上がらない ⟩)
2. *cause*(⟨ 演出も俳優も違う ⟩, ⟨ 味わいはそう大きくは変わる ⟩)
3. *cause*(⟨ 日が差す ⟩, ⟨ 雨が降らない ⟩)
4. *cause*(⟨ 街には音楽があふれる ⟩, ⟨ 大ヒット曲が現れる ⟩)
5. *cause*(⟨ 国土の五二％は積雪地帯で二千万人以上が住む ⟩, ⟨ 雪についての研究は遅れない ⟩)
6. *cause*(⟨ 誰でも名を知る ⟩, ⟨ 実体はよく知る ⟩)
7. *effect*(⟨ 二十数回も見合いを重ねる ⟩, ⟨ そんな男性に出会う ⟩)
8. *effect*(⟨ 家族向けの中華レストランにする ⟩, ⟨ 二十歳前後の女の子でいつもいっぱいでない ⟩)
9. *effect*(⟨ 今年度は定員を二百五十人に増やす ⟩, ⟨ すぐ満杯とならない ⟩)
10. *effect*(⟨ 料金は従来の倍にする ⟩, ⟨ 客は増えない ⟩)
11. *precond*(⟨ 仕事はとっくに終わる ⟩, ⟨ 職場から離れる ⟩)
12. *precond*(⟨ 四歳と七ケ月である ⟩, ⟨ 母親に抱っこをねだらない ⟩)
13. *precond*(⟨ 実の父親が亡くなる ⟩, ⟨ 告別式にも駆けつける ⟩)
14. *others*(⟨ 一枚の絵を仕上げる ⟩, ⟨ 二、三日かからない ⟩)

15. $others(\langle$ 一人前の職人になる $\rangle, \langle$ 最低三年はかからない $\rangle)$
16. $others(\langle$ 欧米文学の翻訳書は書店にあふれる $\rangle, \langle$ 東南アジア諸国の文学は全く無視しない $\rangle)$
17. $others(\langle$ 会社に行く $\rangle, \langle$ 団地から最寄りの駅までバスに乗らない $\rangle)$
18. $others(\langle$ 今日もここへ出席する $\rangle, \langle$ クギを刺さない $\rangle)$
19. $others(\langle$ 一つのものの謎を解く $\rangle, \langle$ 私はルネ・デカルトの方法に従わない $\rangle)$
20. $others(\langle$ 目の疲れがひどく、一個作る $\rangle, \langle$ 休み休みで数日はかからない $\rangle)$

List of Publications

Journal Paper

[1] T. Inui and K. Inui. Committee-based decision making in probabilistic partial parsing. Journal of the Information Processing Society of Japan, Vol. 42, No. 12, pp. 3160–3172, 2001. (in Japanese).

[2] T. Inui, K. Inui, and Y. Matsumoto. Acquiring causal knowledge from text using the connective "tame". Journal of the Information Processing Society of Japan, Vol. 45, No. 3, 2004. (in Japanese).

International Conferences (Reviewed)

[1] T. Inui and K. Inui. Committee-based decision making in probabilistic partial parsing. In *Proc. of The 18th International Conference on Computational Linguistics*, pp. 347–354, 2000.

[2] T. Inui, K. Inui, and Y. Matsumoto. What kinds and amounts of causal knowledge can be acquired from text by using connective markers as clues? In *The 6th. International Conference on Discovery Science*, Vol. 2843 of *Lecture Notes in Artificial Intelligence*, pp. 180–193, Springer-Verlag, 2003.

Other Publications

[1] T. Inui, K. Inui, and Y. Matsumoto. Acquiring lexical knowledge for event desirability analysis. In *Proc. of The 10th Annual Meeting of The Association for Natural Language Processing*, 2004. (in Japanese) .

[2] K. Tateishi, S. Fukushima, N. Kobayashi, M. Wade, T. Takahashi, T. Inui, A. Fujita, K. Inui and Y. Matsumoto. Opinion information extraction from web document and summary generation based on viewpoint. In *Proc. of The 10th Annual Meeting of The Association for Natural Language Processing*, 2004. (in Japanese) .

[3] T. Inui, K. Inui, and Y. Matsumoto. A study of automatic identification of causal knowledge in text: the case of the resultative connective "tame". In *Proc. of The 9th Annual Meeting of The Association for Natural Language Processing*, pp. 707–710, 2003.（in Japanese）.

[4] T. Inui, K. Inui, and Y. Matsumoto. Extracting causal knowledge from text: the case of the resultative connective "tame". In *IPSJ SIG NOTE*（*2002-NL-150*）, pp. 171–178, 2002.（in Japanese）.

[5] N. Kobayashi, T. Inui and K. Inui. Dictionary-based acquisition of lexical knowledge for p/n analysis. In *Proc. of Japanese Society for Artificial Intelligence, SLUD-33*, 2001.（in Japanese）.

[6] T. Inui and K. Inui. Statistical dependency structure analysis based on probabilistic generalized committee-based decision making. In *Proc. of The 8th students lecture meeting, Kyushu section, IEICE*, 2000.（in Japanese）.

[7] T. Inui and K. Inui. Committee-based decision making in probabilistic partial parsing. In *Proc. of The 6th Annual Meeting of The Association for Natural Language Processing*, 2000.（in Japanese）.

[8] T. Inui and K. Inui. An application of probabilistic partial parsing - detection of syntactic-tag errors in treebanks - In *IPSJ SIG NOTE*（*2002-NL-134*）, 1999.（in Japanese）.

[9] T. Inui, A. Kimura and K. Inui. Behavior of a statistical partial parser. Post-conference Workshop in Conjunction with The 5th Annual Meeting of The Association for Natural Language Processing "Parsing: current analysis and future directions ", pp. 171–178, 1999.（in Japanese）.

Awards

- Lecture Prize. The 8th students lecture meeting, Kyushu section, IEICE（2000）.

  T. Inui and K. Inui. Statistical dependency structure analysis based on probabilistic generalized committee-based decision making.

- Excellence of presentation Prize. The 9th Annual Meeting of The Association for Natural Language Processing（2003）.
  T. Inui, K. Inui, and Y. Matsumoto. A study of automatic identification of causal knowledge in text: the case of the resultative connective “tame”.

Acknowledgements

(I am unable to fully express in English my gratitude to everyone who supported me in my research at NAIST. I therefore offer my sincere thanks here in Japanese.)

本研究を進めるにあたり，多くの方々の御支援，御協力を賜わりました．特に次の諸氏に感謝いたします．

まず，本研究を行なう機会を提供して下さった松本裕治教授に感謝致します．数多くの学生を抱え，多忙な研究活動を送られているも関わらず，相談に伺った際には常に快く議論の時間を割いて頂き，本研究の方向性を決定する上での示唆に富む多くの助言を頂きました．ありがとうございました．

乾健太郎助教授は，九州工業大学在籍当時，私が自然言語処理の研究を始めるきっかけを与えてくれた恩師であります．研究当初から終始，議論に参加して頂き，筆者の誤った意思決定をしばしば改め，正しい方向へと導いて頂きました．本論文は単独著者による著作物としてのみ受領されますが，同氏はまちがいなく本論文の第2著者に位置します．

植村俊亮教授，渡邊勝正教授には，審査委員として本論文を査読して頂き，貴重なコメントを頂きました．この場を借りて両氏に御礼申し上げます．

本研究では，諸機関の方々から言語データや言語処理ソフトウェアを提供して頂きました．毎日新聞社，日本経済新聞社からは，電子記事データの使用を許可して頂きました．日本電子化辞書研究所からはEDR電子化辞書を，またNTTコミュニケーション科学基礎研究所からは日本語語彙大系, ALT-J/E翻訳システム用辞書を利用させて頂きました．

4章では，次の方々の協力によって実験を実施しました．快く実験に協力して頂いたNHK放送技術研究所の江原暉将氏，奈良先端科学技術大学院大学 (現在，日立中央研究所) の藤尾正和氏，東京大学 (現在，IBM東京基礎研究所) の金山博氏，東京工業大学 (現在，北陸先端科学技術大学院大学) の白井清昭氏，通信総合研究所の内元清貴氏および京都大学 (現在，東京大学) の黒橋禎夫氏の諸氏に深く感謝いたします．5章で取り扱った言語データは，自然言語処理学講座の小林のぞみさん，徳永泰浩氏，縄田浩三氏の協力のもと，アノテイション作業を実施しました．諸氏に感謝します．

Edson T. Miyamoto助手，新保仁助手をはじめ，自然言語処理学講座の諸氏からは研究会等を通じ，多くのご意見を頂きました．特に，世界知識獲得勉強会の飯田龍氏，SVM勉強会の山田寛康氏，高村大也氏からは，多岐にわたる御助

言を頂きました．また，工藤拓氏からは，SVM を中心とした機械学習の諸概念について御教示いただくとともに，SVM 学習パッケージを提供して頂きました．同氏に感謝します．自然言語処理学講座の登淑恵秘書，大川基子秘書には事務手続きの面でお世話になりました．

井佐原均氏をはじめ，通信総合研究所けいはんな情報通信融合研究センター自然言語グループの皆様には，研究発表会等を通じて示唆に富む貴重な助言を頂きました．特に，高梨克也氏，竹内和広氏，大塚（乾）裕子女史とは多くの議論の場を持たせて頂きました．深く感謝します．

藤田篤氏，高橋哲朗氏，徳久良子さんには，九州工業大学在籍当時から，研究活動における様々な面で相談に乗って頂きました．諸氏に心より感謝します．

森本芳弘氏，與口悠一氏には，計算機環境の面で色々と相談に乗って頂きました．英語論文の校正では，Campbell Hore 氏，Eric Nichols 氏に大変お世話になりました．改めて御礼申し上げます．

最後になりましたが，これまでの学業期間を終始温かく見守ってくれた家族に感謝します．