

# Resolving Zero Pronouns in Texts using Textual Structure

Hiromi NAKAIWA\* Francis BOND† Takahiro UEKADO‡ Yayoi NOZAWA§

\*CCL, UMIST and NTT Communication Science Laboratories

`nakaiwa@nttkb.ntt.jp`

†NTT Communication Science Laboratories

`bond@nttkb.ntt.jp`

‡NTT Software Corp.

`uekado@rd.ntts.co.jp`

§NTT Advanced Technology Corp.

`nozawa@totsuka.ntt-at.co.jp`

## Abstract

This paper proposes a powerful method for the resolution of Japanese zero pronouns in a text with meta information which indicates the textual structure. This method uses not only semantic and pragmatic constraints but also information about the text structure. According to a window test for 480 zero pronouns in a specification with explicit structural information, all of the zero pronouns could be resolved consistently and correctly. Furthermore, 93% of zero pronouns with intersentential antecedents could be resolved correctly using only structural information in the text. This result indicates that, for a text with explicit structural information, the information can be used to effectively resolve zero pronouns with intersentential antecedents.

## 1 Introduction

In all natural languages, elements that can be easily deduced by the reader are frequently omitted from expressions in texts (Kuno 1978). This phenomenon causes big problems in any natural language processing system. For example, in a machine translation system, the system needs to recognize that elements not overtly indicated in the source language may become mandatory elements in the target language. In Japanese, the subject and object are often omitted, whereas they are often mandatory in English. Thus, in Japanese-to-English machine translation systems it is necessary to identify case elements omitted from the original Japanese (these are referred to as “zero pronouns”) for their translation into English expressions.

Recently the importance of electronic text has been recognized. Several kinds of meta languages (markup languages) have been developed which describe the overall information to be encoded within the target medium (not only text but also textual structure like chapters or sections and information on table structure or layout of the text). Furthermore, the computer

software to support the creation of documents coded using markup languages, such as HTML or SGML, has been commercialized and it has become easy to create electronic documents which include structural information about the text (Bond *et al.* 1996; Sperberg-McQueen and Burnard 1994). Recently, computerized documents (corpora) with meta information concerning textual structure, such as chapter and section, have become very common. This change in the nature of documents allows a new approach to the problem of resolving zero pronouns in text which has explicit structural markup.

Several methods have been proposed with regard to the problem of zero pronoun resolution (Kameyama 1986; Walker *et al.* 1990; Yoshimoto 1988; Dousaka 1994). Furthermore, several methods to resolve zero pronouns which consider applications for a practical machine translation system with an unlimited translation target area, have been proposed (Nakaiwa and Ikehara 1992; Nakaiwa and Ikehara 1995; Nakaiwa *et al.* 1996). But, these methods are not suitable for resolving zero pronouns within corpora with explicit structural markup because they do not take into account the use of the explicit structural markup.

In this paper, we propose a widely applicable method to determine the antecedents of Japanese zero pronouns using not only semantic and pragmatic constraints such as verbal semantic attributes, conjunction types and modal expressions but also the structural features of a text.

## 2 Appearance of Zero Pronouns in a structured Japanese Text

### 2.1 Test Sentences

To grasp the distribution of zero pronouns and their antecedents in Japanese text with textual structure, we examine which zero pronouns must be resolved and where their antecedents appear in a Japanese specification (for an international contract; 1856 sentences). The reasons for the selection of this text follow.

- This document is organized by section and chapter.  
(It has explicit structural markup)
- The purpose of the document is clearly described.

Figure 2.1 shows an example of text from the international specification contract.

### 2.2 Characteristics of the Textual Structure

In this section, we analyze the characteristics of textual structure in Japanese specifications, examining the types of features.

#### 2.2.1 Structural Role and Semantic Role

Sentences in documents with textual structure such as specifications and technical manuals have **semantic roles** depending on the textual structure. For example, “table of contents” and “index” each have semantic roles within the whole document. Furthermore, each sentence in the document has a **structural role**. For example, a sentence which includes a section number has a structural role as a TITLE OF SECTION and the title describes the summary of

#### 6.4.1 AV-shuuroku-setsubi (AV recording facility)

*AV-shuuroku-setsubi-ha tsugi-no-3tsu-no-setsubi-kara kousei-sare-masu*  
(AV recording facility consists of the following 3 facilities.)

(1) sutajio-chousei-sitsu (Studio control room)

1) VTR (VTR)

*a. hon-souchi-ha TBC-kinou-wo yuu-shi, kaku-henshuu-shitsu-no edhitto-kontoroora-kara seigyo-kanou-nakoto*

(This facility is to be equipped with TBC function controllable by the edit controller in each editing room.)

*i ...*

2) oudhio-teepu-rekoodaa (Audio tape recorder)

*a. ika-no-taipu-no-teepu-rekoodaa-wo yuusuru-koto.*

(Following types of audio tape recorder are to be supplied.)

*i ...*

(2) shuuroku-sutajio (Recording studio)

...

Figure 1: Text Example in a Specification for an International Contract  
(English sentences in parenthesis indicate translation equivalents of Japanese sentences.)

the section in a simple phrase. Also, a sentence which includes symbols such as “*a.*” has a structural role as an LIST ITEM and the sentence describes the details of one item. These kinds of structural role can be used as conditions to determine the antecedents of zero pronouns.

Semantic and structural roles in the Japanese specification are summarized in Table 1.

### 2.2.2 Hierarchical Structure

Documents with textual structure have a hierarchical structure which indicates the scope of information stated by a sentence in the text. According to the results of the analysis of the specification, the use of cues, such as sentences with a section number, gives sufficient information to determine this type of hierarchical structure (levels of the structure) by comparing the number of the sections or subsections.

ex. Relationship between the number of section and the level of the structure

6 > 6.1 > 6.1.1 > (1) > 1) > a.

-> relative levels of each section/subsection

### 2.3 Tendency of Appearance of Zero Pronouns

The zero pronouns in the Japanese specification can be classified into 3 types; (1) zero pronouns with antecedents within the same sentence (intrasentential), (b) zero pronouns with antecedents elsewhere in the text (intersentential) and (c) zero pronouns with deictic reference (extrasentential). The results of the examination of zero pronouns and their referential elements in the specification are shown in Table 2. There were a total of 480 zero pronouns.

Semantic Role	Structural Role	Description of Text
Title	Title	Title of the document
Table of Contents	Table of Contents	Contents
	Item of Contents	Sentences of the items in contents
Body	Title of Section	Sentence with section number (ex. 5 Explanation 5.1 AV Production Facility) Sentence with subsection number such as “(1)”, “A.”. (ex. (1) Recording Studio)
	List item	Sentence which begin with symbol such as “_” “*”
	Supplementary Ex- planation of Item	Sentence following a LIST ITEM and be- ginning with a conjunction such as <i>nao</i> (further), <i>sarani</i> (furthermore)
	Body	Sentence which begins with a character (except for a LIST ITEM, or SUPPLEMEN- TARY EXPLANATION OF ITEM)
	Table	Table No.
Caption of Table		Caption of Table
Item in Table		Sentence about the title in each column
Body of Table		items within the table
Supplementary Ex- planation of Item		Supplementary explanation of items within the table like footnotes

Table 1: Semantic Roles and Structural Roles  
(Examples in this table are English equivalents of Japanese sentences in the Japanese specifications)

### 2.3.1 Intrasentential Anaphora

9% (41 out of 480 instances) of the antecedents were in the same sentence. Further study showed that 25 instances of zero pronouns with intrasentential antecedents refer to the theme with the postpositional particle, *ha* as shown in the following example.

*hon-souchi-ha, TBC-kinou-wo [ $\phi$ -ga] yuu-si, kontoroora-kara seigy-kanou-nakoto*  
 This-equipment-TOP TBC-function-OBJ  $\phi$ -SUBJ have-AND controller-from to-be-controllable  
 This equipment shall have a TBC function and be controllable by the controller.

This tendency of the appearance of zero pronouns and their antecedents in the same sentence is almost the same as the results discussed by Nakaiwa and Ikehara (1995). So, for the resolution of this kind of zero pronoun, semantic and pragmatic constraints such as the types of conjunctions, verbal semantic attributes (VSA) (Nakaiwa *et al.* 1994) and the types of modal expression seem to be effective. For example, in the previous Japanese sentence, the rule in Table 3 can be used to resolve the zero pronouns.

### 2.3.2 Intersentential Anaphora

In the case of zero pronouns with antecedents in other sentences in the specifications, they appear only in *ga*-cases ‘subject’ (74%; 356 instances) and they are the most common types of zero pronouns in the specifications. In the 356 instances, zero pronouns that refer to the title of a section are the most common: 301 instances. Further examination of these zero pronouns

Location of Zero Pronouns	Location of 'referential elements'													Sub Total
	Intrasentential				Intersentential							Extra-sentential		
	<i>ha</i> Top	<i>ga</i> Subj	<i>wo</i> Dir Obj	<i>ni</i> Ind Obj	Title of Section			List Item			Item of table	I	you	
					same level	one level up	two level up	one sent. before	two sent. before	three sent. before				
<i>ga</i>	12	6	0	0	293	7	1	17	2	2	34	39	34	447
<i>ga</i> (Embed.)	0	0	0	0	0	0	0	0	0	0	0	4	6	10
<i>wo</i>	13	2	7	1	0	0	0	0	0	0	0	0	0	23
sub total	41				356							83		480

Table 2: Distribution of Zero pronouns and their referential elements  
(Total Number of Zero Pronouns : 480 instances, 423 sentences)

Condition for sentence with zero pronoun			
Type of Sentence	Type of Conjunction	Sub Clause	
		<i>ga</i> -case	VSA
Complex Sent. <sup>1</sup>	<i>Renyo Chauusi</i> (=and)	$\phi$	POSSESSION

Condition for sentence with antecedent			Antecedent
Main Clause			
<i>ha</i> -case	VSA	Modal Exp.	
before $\phi$	BODILY ACTION	OBLIGATION	

Table 3: Rule for Resolving Zero Pronouns with Intrasentential Antecedents

indicates that the elements which are stated in the title of a section become the center of the discourse segment which makes up the section. Within the discourse segment (the section), the *ga*-case which refers to the center (the title of section) becomes a zero pronoun. This kind of zero pronoun often appears in the lower levels of the discourse segment, which consists of a hierarchical structure as explained in 2.2.2. This tendency indicates that the lower the level the stronger the relationship between zero pronouns in the level and the title of the subsection in the same level becomes. Furthermore, in the 301 instances, zero pronouns which refer to elements within the same level amount to 293 instances and those which refer to elements one level up amount to only 7 instances; there was only one element that referred two levels up. This result indicates that zero pronouns are more likely to have their antecedents within the same level than in upper levels.

In the case of zero pronouns in itemized list, almost all of the sentences with zero pronouns begin with conjunctions such as *mata* ‘and’ *sarani* ‘furthermore’, which are additional explanations of the previous sentence. In the case of zero pronouns in tables (34 instances), all the zero pronouns are in the body of the table and refer to elements within the table. There are no zero pronouns which refer to an element in a sentence outside the table. This result indicates that sentences in the table have no direct relationship with sentences outside the table.

### 2.3.3 Extrasentential Anaphora

Sentences in the specification have clear cues to identify the referential elements of zero pronouns because the specifications have a clear purpose to inform someone of something. This tendency is reflected in the distribution of zero pronouns and their referential elements. In the case of zero pronouns whose antecedents were not stated in the text, the referential elements were only either the writer (procurer) or reader (provider). The sentences with such zero pronouns have a modal expression which indicates the referential elements of zero pronouns. For example, the subject of a sentence with modal expression *site-kudasai* (REQUEST) or *-koto* (OBLIGATION) becomes a zero pronoun and refers to the reader (*You*). Furthermore, zero pronouns which are in the subject position and have the modal expression *-masu* (POLITE) in the sentence refer to the writer (*I*).

These kinds of cues are almost the same as the results discussed in Nakaiwa *et al.* (1996). So, for the resolution of these kinds of zero pronouns, semantic and pragmatic constraints such as the types of modal expressions, verbal semantic attributes (VSA) (Nakaiwa *et al.* 1994) and the types of conjunctions seem to be effective.

As shown in Table 2, zero pronouns with antecedents in other sentences in the document (intersentential) depend highly on the textual structure for resolution and the use of cues based on textual structure in the whole sentence for anaphora resolution seems very effective. So, when resolving zero pronouns in a text which is tagged with structural information in the text, this tagged information should be taken into account in the resolution process.

## 3 Resolution Method

Based on the information shown in the previous section 2, we propose a method to resolve Japanese zero pronouns within a text containing structural information. The process of the resolution can be divided as follows:

- determination of sentence structure based on explicit structural markup
- resolution of zero pronouns based on text structure

## 3.1 Determination of Discourse Structure using Textual Structure

### 3.1.1 Discourse Structure

We designed the information concerning discourse structure for each sentence as follows.

- serial number of each sentence
- explicit structural markup of each sentence
- hierarchical level
- structural role
- semantic role
- special semantic role

We assume that the serial number and explicit structural markup of each sentence is already tagged for each sentence of Japanese input before processing. The explicit structural markup is equivalent to meta information, such as chapter or section, in mark up language. The other four kinds of information: hierarchical level, structural role, semantic role and special semantic role, correspond to the types of textual structure. All the explicit structural markup corresponds to the structural roles, semantic roles and special semantic roles respectively and some textual structural information corresponds to the hierarchical levels one by one. The hierarchical level indicates the depth of the level of semantic structure, which will be explained in Section 3.1.2. Level 1 of the hierarchical level indicates the top level of the document and the deeper the level becomes the higher the number. The structural role acts as a trigger to create a new level for the next sentence. Special semantic roles are used to distinguish discourse segments which have different individual structures corresponding to the types of special semantic roles which must be treated differently. The reason for the use of special semantic roles is as follows.

- Sentences in tables or figures in a document must be treated as different types of sentence because even if one sentence which is the last sentence in a table or figure, is followed by another sentence which is not in the table, these two sentences, whilst adjacent physically (in layout), have no direct relationship between them. So, when determining the discourse structure of the new sentence based on the discourse structure of the previous sentence, if the previous sentence has a special semantic role and the new sentence has not, the system must refer to a sentence just before the table or figure.

In the process of the determination of discourse structure, sentences which have the same semantic role comprise one semantic discourse structure. Each semantic discourse structure has a hierarchical substructure which consists of discourse segments. All the semantic discourse structures consist of discourse segments.

### 3.1.2 Determination Process

The process for the determination of discourse structure is as follows. Basically, the appropriate position for a new sentence in the discourse structure of the whole document is determined by using the state of discourse structure after the previous sentence was processed and the explicit structural markup of the new sentence.

(Step 1) Select a target sentence to compare with the new sentence

To determine the discourse segment of the new sentence, the system looks for a target sentence with which to compare it. The target sentence is normally the previous sentence except for the following case.

- If the previous sentence has a special semantic role and the new sentence does not have a special semantic role, the sentence which is just before the sentence set which has the special semantic roles (table or figure), becomes the target sentence.

(Step 2) Compare the target sentence and the new sentence.

If these two sentences have different semantic roles, make a new semantic discourse structure which begins with the new sentence.

Else, determine the discourse segment of the new sentence depending on the the types of structural role of the new sentence and the target sentence.

- If the new sentence has a structural role which makes it the first sentence of a new discourse segment, compare the hierarchical level of the new sentence (marked in the text structural information) and the hierarchical level of the target sentence. Then, set the new discourse segment in the appropriate hierarchical level.
- If the new sentence has a structural role which does not become the first sentence of a new discourse segment, set the new sentence to the discourse segment corresponding to the target sentence.

The discourse structure which is generated by this process is similar to the discourse structure which was proposed by [Grosz and Sidner \(1986\)](#). So, the discourse structure which is proposed is quite standard and reliable. Furthermore the method proposed in this section only uses information which corresponds to meta information coded by a mark up language (explicit structural markup) for the determination of discourse structure. So, this method is very simple and is easy to convert for other types of meta information.

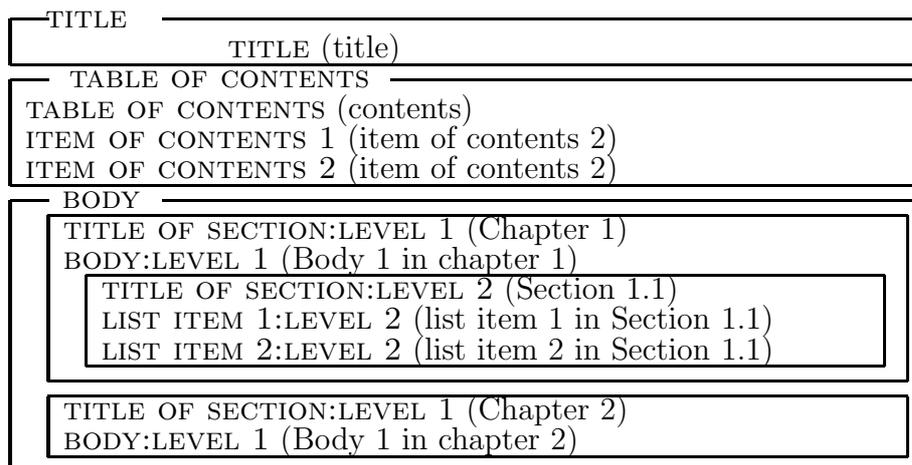
### 3.1.3 Example

By using the method, the following text, in which each sentence has its structural information in the text,

```
Title
Table of Contents
Content 1
```

Content 2  
 Chapter 1  
 Body 1 in Chapter 1  
 Section 1.1  
 List Item 1 in Section 1.1  
 List Item 2 in Section 1.1  
 Chapter 2  
 Body 1 in Chapter 2

is analyzed into this discourse structure.



## 3.2 Resolution of Japanese Zero Pronouns

In this section, we explain a method to resolve Japanese zero pronouns in text which has explicit structural markup. The zero pronouns can be divided into three types, (1) intrasentential anaphora, (2) intersentential anaphora and (3) extrasentential anaphora.

### 3.2.1 Resolution Rules

The rules to resolve zero pronouns can be classified into the following three parts.

- (1) Conditions for a sentence which contains zero pronouns (omitted sentence) can be classified as follows:
  - Conditions for structural features of the text to identify omitted sentence
  - Conditions for the syntactic and semantic structures to identify the zero pronouns in the omitted sentence
- (2) Conditions for a sentence which contains antecedents of the zero pronouns (referred sentence) can be classified as follows:
  - Conditions for structural features of the text to identify the referred sentence
  - Conditions for the syntactic and semantic structures to identify zero pronouns in the referred sentence

(3) Information to control the rule matching process is as follows:

- priority for each rule
- reapplicability of one rule in the same sentence (Whether or not the rule can be applied to the substructure in the same sentence after the rule was already applied to another substructure in the same sentence)

### 3.2.2 Intrasentential Resolution

Intrasentential resolution of Japanese zero pronouns is conducted using rules which are based on semantic and pragmatic constraints such as the type of conjunction, verbal semantic attributes and the type of modal expressions. Rules to resolve this type of zero pronoun consist of conditions for one sentence only and the structural information in the text because, in this case, the omitted sentence and the referred sentence are the same. The method to resolve Japanese zero pronouns with intrasentential antecedents uses basically the same types of constraints as were proposed by [Nakaiwa and Ikehara \(1995\)](#) in their work on intrasentential resolution of Japanese zero pronouns for a machine translation system.

### 3.2.3 Intersentential Resolution

Intersentential resolution of Japanese zero pronouns uses structural information in the text to narrow down the applicable sentences for each rule. The resolution process is conducted based on resolution rules.

#### (1) Determination of Zero Pronouns

First of all, in the process of the determination of zero pronouns which must be resolved, we match structural information of the omitted sentence with conditions for structural information in each rule and narrow down the applicable rules from whole rules as in the following example:

- If the structural role of a sentence which contains a zero pronoun is BODY and if the sub rule to determine the omitted sentence in one rule has a condition that the structural role should be BODY then the sentence is determined as a matched omitted sentence. But if the sub rule has a condition that structural role should be LIST ITEM, the rule is eliminated from the candidates of matchable rules.

After this matching process, each of the matchable rules is applied to the syntactic/semantic structure of the sentence and if some rule matches with the structure, the zero pronouns, which must be resolved, are determined. If there are no matching rules, the resolution process finishes.

#### (2) Determination of Antecedents

When determining the antecedents of zero pronouns, we do not check the syntactic/semantic structure of previous sentences one by one from adjacent sentences, because zero pronouns often refer to elements not in the adjacent sentence but in the title of the section in which the omitted sentence appears (Section 2).

First of all, select the possible referred sentence whose structural information matches with structural information in each rule by checking in the following order:

- matching semantic role
- matching hierarchical level and special semantic role
- matching structural role

For the syntactic/semantic structures of each possible referred sentence being considered, conditions for syntactic/semantic structure are applied. If a match is found, then it determines the antecedent. If there is no suitable referred sentence or all conditions for syntactic/semantic structure are unmatched, the process finishes.

For example, in the following text in the specification, the *ga*-case ‘subject’ of the fourth sentence becomes a zero pronoun and refers to the second sentence, *sutajio-chousei-sitsu* ‘Studio control room’. Each sentence has explicit structural markup and, according to the process of determination of discourse structure, discourse structure of the sentences was determined using the structural markup as shown in each sentence of the following example.

- (1) *sutajio-chousei-sitsu* (Studio control room) : TITLE OF SECTION:LEVEL 4
- 1) *VTR* (VTR) : TITLE OF SECTION:LEVEL 5
- 2) *oudhio-teepu-rekoodaa* (Audio tape recorder) : TITLE OF SECTION:LEVEL 5
- a. *ika-no-taipu-no-teepu-rekoodaa-wo phi-ga yuusuru-koto.* : LIST ITEM:LEVEL 5  
( $\phi$  should have the following types of audio tape recorder.)
- i ...

The zero pronoun can be resolved using the following rule:

Condition for sentence with zero pronoun (omitted sentence)					
Structural Role	Hierarchical Level	Type of Sentence	<i>ga</i> -case	VSA	Modal Exp.
LIST ITEM	5	Unit sent.	$\phi$	POSSESSION	OBLIGATION

Condition for sentence with its antecedents			Antecedent
Structural Role	Hierarchical Level	Condition for element which is not its antecedent	
TITLE OF SECTION	same level	same as the head of <i>wo</i> -case in omitted sentence	The title
TITLE OF SECTION	one level up	—————	The title

Table 4: Rule for Resolving Zero Pronouns with Intersentential Antecedents

### 3.2.4 Extrasentential Resolution

Extrasentential resolution of Japanese zero pronouns is conducted using rules which are based on semantic and pragmatic constraints such as the type of modal expressions, verbal semantic attributes and the type of conjunction. The method to resolve this type of zero pronouns uses basically the same types of constraints which were proposed by Nakaiwa et al (Nakaiwa *et al.* 1996).

The entire algorithm for resolving zero pronouns is summarized in Figure 2. The method to resolve zero pronouns in a corpus with structural information in the text was implemented in the Japanese-to-English machine translation system **ALT-J/E** (Ikehara *et al.* 1991). In each step in the algorithm, when the referential element within or without the text is determined, the system checks not only the conditions in the following algorithm, but also the semantic conditions that verbs impose on zero pronouns in the case elements in each pattern of the Japanese to English transfer pattern dictionaries.

1. Determination of omitted sentence, zero pronouns  
narrow down applicable rule candidates based on text structural information
2. Application of intrasentential rules
3. Application of intersentential rules  
Determination of referred sentence, antecedents  
narrow down matchable sentences based on text structural information
  - narrow down matchable sentences using semantic role
  - narrow down matchable sentences using hierarchical level and special semantic role
  - narrow down matchable sentences using structural roleapply conditions for syntactic/semantic structure
4. Apply extrasentential rules

Figure 2: Zero Pronoun Resolution using Textual Structure

## 4 Evaluation

### 4.1 Evaluation Method

In this section, we show the results of evaluating the method proposed above. The criteria for the evaluation and procedures used were as follows.

#### 4.1.1 Resolution Target

The target was to resolve successfully the following three types of zero pronouns: These are the 85 zero pronouns (54 sentences) with intrasentential antecedents; the 356 zero pronouns (329 sentences) with intersentential antecedents and the 41 zero pronouns (40 sentences) with extrasentential referents found within the 480 zero pronouns in the 1856 sentences in the specification for an international contract.

### 4.1.2 Rules to Resolve Zero Pronouns

The rules to resolve three types of 480 zero pronouns were created by examining these zero pronouns using the constraints discussed in section 3 (116 rules)<sup>2</sup>. Table 5 shows the types of rules depending on the types of anaphora and the types of zero pronouns and their antecedents.

Types of anaphora	Types of zero pronoun and their antecedents	# of rules
Intrasentential	$\phi$ at subject position	8
	$\phi$ at object position	9
Intersentential	refer to TOPIC OF SECTION in the same level	55
	refer to TOPIC OF SECTION two levels higher	1
	refer to ITEM IN TABLE	11
	refer to LIST ITEM (1–3 sentences before)	7
Extrasentential	refer to writer <i>I</i>	10
	refer to reader <i>you</i>	15
Sum		116

Table 5: Types of Rules for Zero Pronoun Resolution

### 4.1.3 Tests for the Evaluation

To examine the relationship between conditions of resolution and accuracy of resolution, we examined the accuracy of resolution depending on the types of conditions in anaphora resolution such as hierarchical level, other text structural information and constraints for syntactic/semantic structure such as modal expressions, verbal semantic attributes and conjunctive expressions. We evaluated the accuracy as further levels of constraints were applied.

- using conditions of hierarchical level only
- using conditions of hierarchical level and other structural information
- using conditions of hierarchical level, other structural information and constraints on the syntactic/semantic structure.

When examining the accuracy without all conditions of resolution (using conditions of hierarchical level only or using conditions of hierarchical level and other text structural information), accuracy was calculated when applying rules which ignore subconditions which are not in the conditions to be examined. The rules which did not use all conditions of resolution for intersentential resolution can be summarized as follows. Each rule has a priority which affects the order in which it should be applied.

- Rules using conditions on the hierarchical level only
  1. Refer to the subject<sup>3</sup> in the nearest previous sentence in the same level
  2. Refer to the subject in the nearest previous sentence in one level up
  3. Refer to the subject in the nearest previous sentence in two levels up

---

<sup>2</sup>At the moment, it is difficult to use sentences which were not already successfully syntactically and semantically analyzed for the evaluation of our method. So, to evaluate the technical limitations of the proposed method, we evaluated the resolution accuracy in the sentences which were examined to make the 116 rules (window test). We will conduct blind tests after we have finished debugging the whole system.

<sup>3</sup>If the ‘sentence’ is a noun phrase (as in the title of a section) consider it to be its own subject.

- Rules using conditions of hierarchical level and other text structure
  1. Rules for antecedents within the nearest previous sentence in the same hierarchical level
    - If the zero pronoun is in the sentence whose structural role is BODY OF TABLE, refer to the whole sentence whose structural role is ITEM IN TABLE
    - If the zero pronoun is in the sentence whose structural role is SUPPLEMENTARY EXPLANATION OF ITEM, refer to the subject of the sentence whose structural role is LIST ITEM
    - If the zero pronoun is in the sentence whose structural role is LIST ITEM or SUPPLEMENTARY EXPLANATION OF ITEM, refer to the whole sentence whose structural role is TITLE OF SECTION
  2. Rules for antecedents within the nearest previous sentence in one level up
    - If the zero pronoun are in the sentence whose structural role is SUPPLEMENTARY EXPLANATION OF ITEM, refer to the subject of the sentence whose structural role is LIST ITEM
    - If the zero pronoun is in the sentence whose structural role is BODY OF TABLE, refer to the whole sentence whose structural role is ITEM IN TABLE
    - If the zero pronoun is in the sentence whose structural role is LIST ITEM or SUPPLEMENTARY EXPLANATION OF ITEM, refer to the whole sentence whose structural role is TITLE OF SECTION
  3. Rules for antecedents within the nearest previous sentence in two levels up
    - If the zero pronoun is in the sentence whose structural role is BODY OF TABLE, refer to the whole sentence whose structural role is ITEM IN TABLE
    - If the zero pronoun is in the sentence whose structural role is SUPPLEMENTARY EXPLANATION OF ITEM, refer to the subject of the sentence whose structural role is LIST ITEM
    - If the zero pronoun is in the sentence whose structural role is LIST ITEM or SUPPLEMENTARY EXPLANATION OF ITEM, refer to the whole sentence whose structural role is TITLE OF SECTION

#### 4.1.4 Evaluation Factor

We used recall to evaluate the accuracy of resolution. Recall is the ratio of the number of zero pronouns that can be successfully resolved to the number of zero pronouns that need to be resolved. Recalls were calculated for each type of anaphora (intrasentential, intersentential and extrasentential) separately.

## 4.2 Resolution Accuracy for Conditions of Resolution

To examine how the resolution accuracy varied according to the conditions of resolution used, we tested the accuracy of the proposed method at the three different conditions specified in 4.1.3. Table 6 shows the results of the resolution depending on the types of constraint applied.

The results indicate that, for a highly structured text such as a specification, with explicit textual markup, textual information is very effective to resolve zero pronouns with intersentential antecedents. In fact, the accuracy achieved for intersentential anaphora was as high as 93%, even without using the constraints for syntactic/semantic structure.

Although the specification has various kinds of structural information in the text as described in Section 2.2, the resolution of intrasentential anaphora and extrasentential anaphora, was possible using only rules and constraints for syntactic/semantic structure. All the zero pronouns which refer to intrasentential or extrasentential antecedents can be resolved correctly with these rules. This result shows that even for a corpus with structural information in the text, the rules for zero pronouns with intrasentential or extrasentential antecedents in the corpus can be created independent of text structural information. <sup>4</sup>

Resolution Condition	Accuracy		
	Intrasentential	Intersentential	extrasentential
hierarchical level	24% (10/41)	74% (265/356)	0% (0/83)
hierarchical level + other textual information	24% (10/41)	93% (330/356)	0% (0/83)
hierarchical level + other textual information + constraints for syntactic semantic structure	100% (41/41)	100% (356/356)	100% (83/83)

Table 6: Resolution Accuracy for Resolution Condition

## 5 Conclusion

This paper has proposed a powerful method for the resolution of Japanese zero pronouns in a corpus with meta information which indicates structural information in the text using not only constraints for syntactic/semantic structure which were used in previously proposed methods but also the explicit structural markup in the text. According to a window test for 480 zero pronouns in a specification with structural information in the text, all zero pronouns could be resolved consistently and correctly. Furthermore, it was found that 93% of zero pronouns with intersentential antecedents could be resolved correctly using only structural information in the text. This result indicates that, for a corpus with explicit structural markup, the use of information about textual structure is very effective to resolve zero pronouns with intersentential antecedents.

In the future, we will examine a method for determining discourse structure using several kinds of corpora with meta information coded by markup language such as HTML, SGML and examine the universality of the rules that have been discussed in this paper by applying them to these corpora. Furthermore, we will also examine a process to control the priorities of each rule for anaphora resolution efficiently and effectively.

---

<sup>4</sup>In the results of Table 6, although the rules for intrasentential resolution only use the constraints for syntactic/semantic structure, the accuracy of intrasentential resolution only using hierarchical level achieves 24%. This is because of the cases where antecedents exist within the same sentence and also in previous sentences so that the rules for intersentential resolution using only hierarchical level can determine the correct antecedents from the previous sentences.

## References

- BOND, FRANCIS, YAMATO TAKAHASHI, SETSUO YAMADA, and MAKIKO NISHIGAKI. 1996. Still tagging an aligned Japanese/English corpus. In *2nd Annual Meeting of the Association for Natural Language Processing*, 205–208.
- DOUSAKA, KOUJI. 1994. Identifying the referents of Japanese zero-pronouns based on pragmatic condition interpretation. *Transactions of Information Processing Society of Japan* 35.768–778. (in Japanese).
- GROSZ, BARBARA J., and CANDACE L. SIDNER. 1986. Attention, intentions, and the structure of discourse. *Computational Linguistics* 12.175–204.
- IKEHARA, SATORU, SATOSHI SHIRAI, AKIO YOKOO, and HIROMI NAKAIWA. 1991. Toward an MT system without pre-editing – effects of new methods in **ALT-J/E**–. In *Third Machine Translation Summit: MT Summit III*, 101–106, Washington DC. (<http://xxx.lanl.gov/abs/cmp-lg/9510008>).
- KAMEYAMA, MEGUMI. 1986. A property-sharing constraint in centering. In *24th Annual Conference of the Association for Computational Linguistics*, 200–206.
- KUNO, SUSUMU. 1978. *Danwa no Bunpoo*. Tokyo: Taishukan Publ. Co. (in Japanese).
- NAKAIWA, HIROMI, and SATORU IKEHARA. 1992. Zero pronoun resolution in a Japanese to English machine translation system using verbal semantic attributes. In *3rd Conference on Applied Natural Language Processing: ANLP-92*, 201–208.
- , and —. 1995. Intrasentential resolution of Japanese zero pronouns in a machine translation system using semantic and pragmatic constraints. In *Sixth International Conference on Theoretical and Methodological Issues in Machine Translation: TMI-95*, 96–105, Leuven.
- , SATOSHI SHIRAI, and SATORU IKEHARA. 1996. Anaphora resolution of Japanese zero pronouns with deictic reference. In *16th International Conference on Computational Linguistics: COLING-96*, 812–817, Copenhagen.
- , AKIO YOKOO, and SATORU IKEHARA. 1994. A system of verbal semantic attributes focused on the syntactic correspondence between Japanese and English. In *15th International Conference on Computational Linguistics: COLING-94*, 672–678, Kyoto.
- SPERBERG-MCQUEEN, C. M., and LOU BURNARD (eds.) 1994. *Guidelines for Electronic Text Encoding and Interchange*. Oxford: Chicago.
- WALKER, MARILYN, MASAYO IIDA, and SHARON COTE. 1990. Centering in Japanese discourse. In *13th International Conference on Computational Linguistics: COLING-90*.
- YOSHIMOTO, KEI. 1988. Identifying zero pronouns in Japanese dialogue. In *12th International Conference on Computational Linguistics: COLING-88*, 779–784.