

# Pronoun Resolution in Japanese Sentences Using Surface Expressions and Examples

Masaki Murata    Hitoshi Isahara  
Communications Research Laboratory  
588-2, Iwaoka, Nishi-ku, Kobe, 651-2401, Japan  
{murata, isahara}@crl.go.jp  
TEL: +81-78-969-2181, FAX: +81-78-969-2189

Makoto Nagao  
Kyoto University  
Sakyo, Kyoto 606-8501, Japan  
nagao@pine.kuee.kyoto-u.ac.jp

## Abstract

In this paper, we present a method of estimating referents of demonstrative pronouns, personal pronouns, and zero pronouns in Japanese sentences using examples, surface expressions, topics and foci. Unlike conventional work which was semantic markers for semantic constraints, we used examples for semantic constraints and showed in our experiments that examples are as useful as semantic markers. We also propose many new methods for estimating referents of pronouns. For example, we use the form "X of Y" for estimating referents of demonstrative adjectives. In addition to our new methods, we used many conventional methods. As a result, experiments using these methods obtained a precision rate of 87% in estimating referents of demonstrative pronouns, personal pronouns, and zero pronouns for training sentences, and obtained a precision rate of 78% for test sentences.

## 1 Overview

This paper describes how to resolve the referents of pronouns: demonstrative pronouns, personal pronouns, and zero pronouns. Pronoun resolution is especially important for machine translation. For example, if the system cannot resolve zero pronouns<sup>1</sup>, it cannot translate sentences containing them from Japanese into English. When the word order of sentences is changed and the pronominalized words are changed in translation into English, the system must detect the referents of the pronouns.

A lot of work has been done in Japanese pronoun resolution (Kameyama 86) (Yamamura et al. 92) (Walker et al. 94) (Takada & Doi 94) (Nakaiwa & Ikehara 95). The main distinguishing features of our work are as follows:

- In conventional pronoun resolution methods, semantic markers have been used for semantic constraints. On the other hand, we use examples for semantic constraints and show in our experiments that examples are as useful as semantic markers. This is an important result because the cost of constructing the case frame using semantic markers is generally higher than the cost of constructing the case frame using examples.
- We use examples in the form "X no Y" (Y of X) for estimating referents of demonstrative adjectives.

<sup>1</sup>Omitted noun phrases are called *zero pronouns*.

Condition  $\Rightarrow$  {Proposal Proposal ..}  
Proposal := (Possible-Antecedent Points)

Figure 1: Form of Candidate enumerating rule

Condition  $\Rightarrow$  (Points)

Figure 2: Form of Candidate judging rule

- We deal with the case when a demonstrative refers to elements that appear later.
- We resolve a personal pronoun in a quotation by determining who is the speaker and who is the listener.

In this work, we used almost all the potentials of conventional methods and also propose a new method.

## 2 The Framework for Estimating the Referent

Prior to the pronoun resolution process, sentences are transformed into a case structure by a case structure analyzer (Kurohashi & Nagao 94). The antecedents of pronouns are determined by heuristic rules from left to right. Using these rules, our system assigns points to possible antecedents, and judges that the one having the maximum total score is the desired antecedent.

Heuristic rules are classified into two kinds: Candidate enumerating rules and Candidate judging rules. Candidate enumerating rules are used in enumerating candidate antecedents and giving them points (which represent the plausibility of being the correct antecedent). Candidate judging rules are used in giving points to the candidate antecedents selected by Candidate enumerating rules. These rules are shown in Figures 1 and 2. Surface expressions, semantic constraints, referential properties, etc. are written as conditions in the *Condition* part. Possible antecedents are written in the *Possible-Antecedent* part. *Points* means the plausibility of the possible antecedent.

An estimation of the referent is performed using the total scores of possible antecedents given by Candidate enumerating rules and Candidate judging rules. First, the system applies all Candidate enumerating rules to the anaphor and enumerates candidate antecedents having points. Next, the system applies all Candidate judging rules to all the candidate antecedents and sums the scores of all the candidate antecedents. Consequently,

Table 1: The weight in the case of topic

Surface expression	Example	Weight
Pronoun/zero-pronoun <i>ga/wa</i>	( <u>John</u> <i>ga</i> (subject)) <i>shita</i> (done)	21
Noun <i>wa/niwa</i>	<u>John</u> <i>wa</i> (subject) <i>shita</i> (do)	20

Table 2: The weight in the case of focus

Surface expression	Example	Weight
Pronoun/zero-pronoun <i>wo</i> (object)/ <i>ni</i> (to)/ <i>kara</i> (from)	( <u>John</u> <i>ni</i> (to)) <i>shita</i> (done)	16
Noun <i>ga</i> (subject)/ <i>mo/da/nara</i>	<u>John</u> <i>ga</i> (subject) <i>shita</i> (do)	15
Noun <i>wo</i> (object)/ <i>ni/, /</i>	<u>John</u> <i>ni</i> (object) <i>shita</i> (do)	14
Noun <i>he</i> (to)/ <i>de</i> (in)/ <i>kara</i> (from)	<u>gakkou</u> (school) <i>he</i> (to) <i>iku</i> (go)	13

the system judges the candidate antecedent having the best score to be the proper antecedent. If several candidate referents have the best score, the candidate referent selected first in order<sup>2</sup> is judged to be the correct antecedent.

We made 50 Candidate enumerating rules and 10 Candidate judging rules for analyzing demonstratives, 4 Candidate enumerating rules and 6 Candidate judging rules for analyzing personal pronouns, and 19 Candidate enumerating rules and 4 Candidate judging rules for analyzing zero pronouns. Some of the rules are described in the following sections.

### 3 Heuristic Rules for Demonstratives

We made heuristic rules for demonstratives by consulting the papers (NLRI 81) (Hayashi 83) (Takahashi et al. 90) (Kinsui & Takubo 92) and by examining Japanese sentences by hand. Demonstratives have three categories: demonstrative pronouns, demonstrative adjectives, and demonstrative adverbs. In the following sections, we explain the rules for analyzing demonstratives.

#### 3.1 Rule for Demonstrative Pronouns

##### Rule in the case when the referent is a noun phrase

###### Candidate enumerating rule 1

When a pronoun is a demonstrative pronoun or “*sono* (of it) / *kono* (of this) / *ano* (of that)”,  
 {(A topic which has weight  $W$  and distance  $D$ ,  
 $W - D - 2$ )  
 (A focus which has weight  $W$  and distance  $D$ ,  $W - D + 4$ )}

This bracketed expression represents the lists of proposals in Figure 1. The definition and weight  $W$  of the topic and focus are shown in Tables 1 and 2. The distance ( $D$ ) is the number of topics and foci between the demonstrative and the possible referent. Since a demonstrative more often refers to foci than a zero pronoun does, we add the coefficient  $-2$  or  $+4$  as compared with the heuristic rules in zero pronoun resolution.

The score (in other words, the certification value) of a candidate referent depends on the weight of topics/foci and the physical distance between the demonstrative and the candidate referent.

##### Rule when the referent is a verb phrase

###### Candidate enumerating rule 2

<sup>2</sup>The order is based on order applying rules.

Table 3: Points given in the case of demonstrative pronouns

Sim.	0	1	2	3	4	5	6	Exact
Points	0	0	-10	-10	-10	-10	-10	-10

Sim. = Similarity level

When a pronoun is “*kore/sore/are*” or a demonstrative adjective,

{( The previous sentence (or the verb phrase which is a conditional form containing a conjunctive particle such as “*ga* (but)”, “*daga* (but)”, and “*keredo* (but)” if the verb phrase is in the same sentence), 15)}

The following is an example of a pronoun referring to the verb phrase in the previous sentence.

*tengu-wa maenoban-noyouuni utattari odottari shihajimeta.*  
 (tengu) (the previous night) (sing) (dance) (begin to do)  
 (Tengus began singing and dancing just as they had done the previous night.)

*ojiisan-wa sore-wo mite, kon'nahuuni utai-hajimeta.*  
 (the old man) (it) (see) (as follows) (begin to sing)  
 (When the old man saw this, he began to sing as follows.)

(1)

In these sentences, a demonstrative pronoun “*sore* (it)” refers to the event “*tengutachi-ga utattari odottari shihajimemashita* (tengus began singing and dancing just as they had done the previous night.)”<sup>3</sup>.

##### Rule using the feature that demonstrative pronouns usually do not refer to people

###### Candidate judging rule 1

When a pronoun is a demonstrative pronoun and a candidate referent has a semantic marker HUM (human), it is given  $-10$ . We used the Noun Semantic Marker Dictionary (Watanabe et al. 92) as a semantic marker dictionary<sup>4</sup>.

###### Candidate judging rule 2

When a pronoun is a demonstrative pronoun, a candidate referent is given the points in Table 3 by using the highest semantic similarity between the candidate referent and the codes {5200003010 5201002060 5202001020 5202006115 5241002150 5244002100} in “Bunrui Goi Hyou (BGH)” (NLRI 64)<sup>5</sup> which signify human beings.

<sup>3</sup>A tengu is a kind of monster.

<sup>4</sup>This dictionary includes semantic categories shown in Table 4.

<sup>5</sup>In BGH, each word has a number called a *category number*. In an electrical version of BGH, each word has a 10-digit

Table 4: Modification of category number of “*bunrui goi hyou*”

Semantic marker	Original code	Modified code
ANI(animal)	156	511
HUM(human)	12[0-4]	52[0-4]
ORG(organization)	12[5-8]	53[5-8]
PLA(plant)	155	611
PAR(part of living thing)	157	621
NAT(natural)	152	631
PRO(products)	14[0-9]	64[0-9]
LOC(location)	117,125,126	651,652,653
PHE(phenomenon)	150,151	711,712
ACT(action)	13[3-8]	81[3-8]
MEN(mental)	130	821
CHA(character)	11[2-58],158	83[2-58],839
REL(relation)	111	841
LIN(linguistic products)	131,132	851,852
Others	110	861
TIM(time)	116	a11
QUA(quantity)	119	b11

“125” and “126” are given two category numbers.

When we calculate the semantic similarity, we use the modified code table in Table 4. The reason for this modification is that some codes in BGH (NLRI 64) are not suitable for semantic constraints.

These rules use the feature that a demonstrative pronoun rarely refers to people. This reduces the number of candidates of the referent. For example, we find “*sore* (it)” in the following sentences refers to “*konpyuuta* (computer)”, because “*sore* (it)” can only refer to only a thing which is not human and the only noun which is near “*sore* (it)” and which is not human is “*konpyuuta* (computer)”.

*taroo-wa saishin-no konpyuuta-wo kaimashita.*  
 (Taroo) (new) (computer) (buy)  
 (Taroo bought a new computer.)

*jon-ni sassoku sore-wo misemashita.*  
 (John) (at once) (it) (show)  
 ([He] showed it at once to John.)

(2)

#### Rule with feature that “*koko*” and “*soko*” often refer to locations

##### Candidate judging rule 3

When a pronoun is “*koko* (here) / *soko* (there) / *asoko* (over there)” and a candidate referent has a semantic marker LOC (location), the candidate referent is given 10 points.

##### Candidate judging rule 4

When a pronoun is “*koko/soko/asoko*”, a candidate referent is given the points in Table 5 based on the semantic similarity between the candidate referent and the

category number. This 10-digit category number indicates seven levels of an is-a hierarchy. The top five levels are expressed by the first five digits of a category number. The sixth level is expressed by the following two digits of a category number. The last level is expressed by the last three digits of a category number.

Table 5: Points given demonstrative pronouns which refer to places

Sim.	0	1	2	3	4	5	6	Exact
Points	-10	-5	0	5	10	10	10	10

codes {6563006010 6559005020 9113301090 9113302010 6471001030 6314020130} which signify locations in BGH (NLRI 64).

“*soko* (there)” commonly refers to location. For example, “*soko*” in the following sentences refers to “*baiten* (shop)” which signifies location.

*koora-wo kaini baiten-ni hairimashita.*  
 (cola) (buy) (shop) (enter)  
 (Taroo entered a shop to buy a cola.)

*jiroo-wa soko-de guuzen dekuwashimashita.*  
 (Jiroo) (there) (by chance) (meet)  
 (Jiroo met Taroo there by chance.)

(3)

#### Rule when “*kokode*” or “*sokode*” is used as a conjunction

##### Candidate enumerating rule 3

When a pronoun is “*kokode*” or “*sokode*”,  
 {(the pronoun is used as a conjunction, 11)}

This rule is for when “*kokode* (here or then)” or “*sokode* (there or then)” is used as a conjunction. If a word that signifies location is not found near “*kokode*” or “*sokode*”, the candidate listed by this rule has the highest score, and “*kokode*” or “*sokode*” is judged to be a conjunction. By using this rule, “*sokode*” in the following sentences is judged to be a conjunction.

*ojiisan-wa tengu-ga kowakunakunatte-imashita.*  
 (old man) (tengu) (lose all fear of)  
 (The old man lost all fear of the tengus.)

*sokode ojiisan-wa kakureteita ana-kara detekimashita.*  
 (so) (old man) (be hiding) (hole) (leave)  
 (So, he left the hole where he had been hiding.)

(4)

This rule is necessary when the system translates “*sokode*” into English, judges whether it is used as a demonstrative or as a conjunction, and translates it into “there” or “then.”

#### Rule when an anaphor does not have its antecedent

##### Candidate enumerating rule 5

When a pronoun is a demonstrative pronoun, a demonstrative adverb, or a demonstrative adjective,  
 {(Introduce an individual, 10)}

This rule is used when there is no referent of a pronoun in the sentences. This rule makes the system introduce a certain individual.

### 3.2 Rule for Demonstrative Adjectives

Demonstrative pronouns such as “*kono* (this)”, “*sono* (the)”, “*ano* (that)”, “*kon’na* (like this)”, and “*son’na* (like it)” are classified into two reference categories: *gentei*-reference and *daikou*-reference.

In a *Gentei*-reference although a demonstrative adjective does not refer to an entity by itself, the phrase of “demonstrative adjective + noun phrase” refers to the antecedent. For example “*kono ojiisan* (this old man)” in the following sentences:

*ojiisan-wa tengutachi-no-maeni deteitte odori-hajimemashita*  
(old man) (before the tengus) (appear) (begin to dance)  
(He appeared before the tengus, and began to dance.)

*keredomo kono ojiisan-wa uta-mo odori-mo hetakuso-deshita*  
(but) (this old man) (sing) (dance) (poor)  
(But the old man was a poor singer, and his dancing was no better.)

(5)

In this example, although the demonstrative “*kono* (this)” does not refer to “*ojiisan* (old man)” in the first sentence, the noun phrase “*kono ojiisan* (this old man)” refers to “*ojiisan* (old man)” in the first sentence.

*Daikou*-reference is a demonstrative adjective that refers to an entity. In this case, we can analyze “*sono* (the)” as well as “*sore-no* (of it)”. In the following sentences, “*sono*” refers to “*tengu*” (tengus). It is an example of *daikou*-reference.

*mata karasu-no-youna kao-wo-shita tengu-mo imashita*  
(also) (like crows) (with face) (tengu) (exist)  
(There were also some tengus with faces like those of crows.)

*sono kuchi-wa torino-kuchibashi-noyouni togatte-imashita*  
(their mouths) (like the beaks of birds) (be pointed)  
(Their mouths were pointed like the beaks of birds.)

(6)

Rules for *gentei*-reference and *daikou*-reference are as follows:

#### Candidate enumerating rule 7

When a pronoun is “demonstrative adjective + noun  $\alpha$ ,”  
{ (the noun phrase containing a noun  $\alpha$ , 45)  
(the topic which is a subordinate of noun  $\alpha$  and which has weight  $W$  and distance  $D$ ,  $W - D + 30$ )  
(the focus which is a subordinate of noun  $\alpha$  and which has weight  $W$  and distance  $D$ ,  $W - D + 30$ )}

The relationships between a super-ordinate word and a subordinate word are detected by judging the last word in the definition of the word  $\alpha$  in EDR Japanese word dictionary (EDR 95a) to be the super-ordinate of the word  $\alpha$ .

Because of this rule, when a pronoun is “demonstrative adjective + noun phrase  $\alpha$ ” and there is the same noun phrase  $\alpha$  near it, it is judged to be “*gentei*-reference” and is selected as a candidate of the referent. When there is a subordinate of a noun phrase  $\alpha$  near it, it is also selected as a candidate of the referent. These rules give higher points to a candidate referent than other rules do. The following is an example of the “demonstrative adjective + noun phrase  $\alpha$ ” referring to

Table 6: Points given to so-series demonstrative adjective

Sim.	0	1	2	3	4	5	6	Exact
Points	-10	-2	-1	0	1	2	3	4

Table 7: Examples of the form “the mouth of Noun X”

Examples of Noun X
<i>hukuro</i> (sack), <i>ruporaitā</i> (documentary writer) <i>iin</i> (member), <i>akachan</i> (baby), <i>kare</i> (he)

the subordinate of noun phrase  $\alpha$ .

*ojiisan-wa toonoiteiku tsuru-no sugata-wo miokurimashita.*  
(old man) (recede) (crane) (figure) (watch)  
(The old man watched the receding figure of the crane.)

*“ano tori-wo tasukete yokatta” to iimashita.*  
(that bird) (save) (glad) (say)  
(“I’m glad I saved that bird,” said the old man to himself.)

(7)

In this example, the underlined “*ano tori* (that bird)” refers to a subordinate “*tsuru* (crane)” in the previous sentence.

#### Rules for *daikou*-reference of so-series demonstrative adjective

##### Candidate judging rule 5

When a pronoun is a *so*-series demonstrative adjective, the system consults examples of the form “noun X *no* noun Y” whose noun Y is modified by the pronoun, and gives a candidate referent the points in Table 6 according to the similarity between the candidate referent and noun X in “Bunrui Goi Hyou” (NLR1 64). The Japanese Co-occurrence Dictionary (EDR 95c) is used as a source of examples of “X *no* Y”.

This rule is for checking the semantic constraint (For a *daikou*-reference, candidates of the referent are selected by Candidate enumerating rule 1 in Section 3.1.).

We explain how to use the rule in the underlined “*sono* (the)” in the sentences (6). First, the system gathers examples of the form “Noun X *no kuchi* (mouth of Noun X)”. Table 7 shows some examples of “Noun X *no kuchi* (mouth of Noun X)” in the Japanese Co-occurrence Dictionary (EDR 95c). Next, the system checks the semantic similarity between candidate referents and Noun X, and judges the candidate referent having a higher similarity to be a better candidate referent. In this example, “*tengu*” is semantically similar to Noun X in that they are both living things. Finally, the system selects “*tengu*” as the proper referent.

#### Rules when non-*so*-series demonstrative has *daikou*-reference

##### Candidate judging rule 6

When a pronoun is a non-*so*-series demonstrative adjective, the system consults examples of the form “Noun X *no* (of) Noun Y (Y of X)” whose Noun Y is modified by the pronoun, and gives candidate referents the points in Table 8 according to the similarity between the candidate referent and noun X in “Bunrui Goi

Table 8: Points given in the case of non-*so*-series demonstrative adjective

Sim.	0	1	2	3	4	5	6	Exact
Points	-30	-30	-30	-30	-10	-5	-2	0

Table 9: Results of investigating whether “*kon’na* noun” (noun like this) refers to the previous or next sentences

Postpositional particle	previous sentence	next sentence
<i>wa</i> (topic)	9	0
<i>wa-nai</i>	5	0
<i>ni</i> (indirect object)	17	0
<i>ni-mo</i>	1	0
<i>ni-wa</i>	2	0
<i>de</i> (place)	15	0
<i>de-wa</i>	5	0
<i>no</i> (possessive)	9	0
<i>sura</i>	2	0
<i>ga</i> (subject)	27	22
<i>wo</i> (object)	43	26
<i>mo</i> (also)	2	4
<i>de-wa-nai</i>	0	1
Total	137	53

Hyou” (NLRI 64). Since a non-*so*-series demonstrative adjective rarely is a *daikou* reference (NLRI 81) (Yamamura et al. 92), the number of points is footnotesizeer than in the case of the *so*-series.

#### Rule when a pronoun refers to a verb phrase

Like a demonstrative pronoun, a demonstrative adjective can refer to the meaning of the verb phrase in the previous sentence. This case is resolved by Candidate enumerating rule 2 in Section 3.1.

#### Rule for “*kon’na* noun” (noun like this)

“*kon’na* noun” can also refer to the next sentences in addition to a noun phrase and the previous sentences.

*ojiisan-wa odorinagara kon’na uta-wo utaimashita.*  
 (old man) (dance) (song like this) (sing)  
 (As he danced, he sang the following song: )

“*tengu tengu hachi tengu.*  
 (tengu) (tengu) (eight tengu)  
 (“‘Tengu,’ ‘tengu,’ eight ‘tengus.’”)

(8)

In the above example, “*kon’na uta* (song like this)” refers to the next sentence “*tengu, tengu, hachi tengu.*”

But we cannot decide whether “*kon’na* noun” (noun like this) refers to the previous or next sentences only by the expression of “*kon’na* noun” (noun like this) itself. To make the decision, we gathered 317 sentences containing “*kon’na*” (like this) from about 60,000 sentences in Japanese essays and editorials, and counted the total frequency of cases in which “*kon’na*” refers to the previous and next sentences. The results are shown in Table 9. This table indicates that “*kon’na* noun” followed by other particles, specifically “*ga*” and “*wo*,” which are used when representing new information, very often refers to the previous sentence. Therefore, the system judges that the desired antecedent is the previous

sentence. When “*kon’na* noun” is followed by the particles “*ga*” or “*wo*,” the proper referent is determined by the expression in quotation marks (“”).

### 3.3 Rule for Demonstrative Adverbs

#### Rule when *so*-series demonstrative adverb refers to the previous sentences

##### Candidate enumerating rule 9

When an anaphor is a *so*-series demonstrative adverb such as “*sou* (so),”

{(the previous sentences, 30)}

The following is an example.

“*tengu tengu hachi tengu.*  
 (tengu) (tengu) (eight tengu)  
 (“‘Tengu,’ ‘tengu,’ eight ‘tengus.’”)

*sou utatta-nowa sokoni hachihiki-no tengu-ga itakara-desu.*  
 (sing so) (there) (eight) (tengu) (exist)  
 (He sang so because he counted eight of them there. )

(9)

“*sou* (so)” refers to the previous sentence “*tengu tengu hachi tengu.*”

#### Rule when *so*-series demonstrative adverb cataphorically Refers to the Verb Phrase in the Same Sentence

##### Candidate enumerating rule 10

When an anaphor is “*sou/soushite/sonoyouni*” and is in the subordinate clause which has a conjunctive particle such as “*ga*,” “*daga*,” and “*keredo*” or an adjective conjunction such as “*youni*,”

{(the main clause, 45)}

### 4 Heuristic Rule for Personal Pronouns

#### Candidate enumerating rule 1

When an anaphor is a first personal pronoun,  
 {(the first person (the speaker) in the context, 25)}

#### Candidate enumerating rule 2

When an anaphor is a second personal pronoun,  
 {(the second person (the hearer) in the context, 25)}

A first or second personal pronoun is often presented in quotations, and can be resolved by estimating the first person (speaker) or the second person (hearer) in advance. The estimation of the first person and the second person is performed by regarding the *ga*-case (subjective) and *ni*-case (objective) components of the verb phrase representing the speaking action of the quotation as the first and second persons, respectively. The detection of the verb phrase representing the speaking action is performed as follows. If the quotation is followed by a speaking action verb phrase such as “*to itta* (was said),” the verb phrase is regarded as the verb phrase representing the speaking action. Otherwise, the last verb phrase in the previous sentence is regarded as the verb phrase representing the speaking action. For example, the second personal pronoun “*omae-san* (you)” in the following sentences refers to the second person “*ojiisan* (the old

<i>ojiisan-wa</i>	<i>jimen-ni</i>	<i>koshi-wo-oroshimashita.</i>
(old man)	(ground)	(sit down)
(The old man sat down on the ground.)		
<i>yagate</i>	<i>(ojiisan-wa)</i>	<i>nemutte-shimaimashita.</i>
(soon)	(old man)	(fall asleep)
(He soon fell asleep.)		
<b>Semantic Marker</b>		
HUM/ANI <i>ga</i> (agent)		<i>nemuru</i> (sleep)
<b>Example</b>		
<i>kare</i> (he)/ <i>inu</i> (dog) <i>ga</i> (agent)		<i>nemuru</i> (sleep)
Figure 3: How to check semantic constraint		

man)” in this quotation.

“*asu*, *mata* *mairimasuyo.*” *to*,  
 (tomorrow) (again) (come)  
 (“I’ll come again tomorrow,”)  
*ojiisan-wa* *yakusoku-shimashita.*  
 (old man) (promise)  
 (promised the old man.)  
 “*mochiron* *omae-san-wo* *utagauwakedewanainodaga.*”  
 (of course) (you) (don’t mean to doubt)  
 (“Of course, we don’t mean to doubt you.”)  
*tengu-ga* *ojiisan-ni* *iimashita.*  
 (tengu) (old man) (said)  
 (said one of the “tengu” to the old man.)

(10)

The second person in the quotation is estimated to be “*ojiisan*” because the *ni*-case component of the verb phrase “*iimashita* (said)” representing the speaking action of the quotation is “*ojiisan*”.

#### Candidate enumerating rule 3

When an anaphor is a third personal pronoun,  
 {(a first person, -10) (a second person, -10)}

## 5 Heuristic Rule for Zero Pronoun

### Rule proposing candidate referents of general zero pronoun

#### Candidate enumerating rule 1

When a zero pronoun is a *ga*-case component,  
 {(A topic which has weight  $W$  and distance  $D$ ,  $W - D * 2 + 1$ )  
 (A focus which has weight  $W$  and distance  $D$ ,  $W - D + 1$ )  
 (A subject of a clause coordinately connected to the clause containing the anaphor, 25)  
 (A subject of a clause subordinately connected to the clause containing the anaphor, 23)  
 (A subject of a main clause whose embedded clause contains the anaphor, 22)}

#### Candidate enumerating rule 2

When a zero pronoun is not a *ga*-case component,  
 {(A topic which has weight  $W$  and distance  $D$ ,  $W - D * 2 - 3$ )  
 (A focus which has weight  $W$  and distance  $D$ ,  $W - D * 2 + 1$ )}

Table 10: Points given by a verb-noun relationship

Sim.	0	1	2	3	4	5	6	Exact
Points	-10	-2	1	2	2.5	3	3.5	4

### Rule using semantic relation to verb phrase

#### Candidate judging rule 1

When a candidate referent of a case component (a zero pronoun) does not satisfy the semantic marker of the case component in the case frame, it is given -5.

#### Candidate judging rule 2

A candidate referent of a case component (a zero pronoun) is given the points in Table 10 by using the highest semantic similarity between the candidate referent and examples of the case component in the case frame.

These two rules are for checking the semantic constraint between the candidate referent and the verb phrase which has the candidate referent in its case component. Candidate judging rule 1 checks semantic constraints by using semantic markers. Candidate judging rule 2 checks semantic constraints by using examples. Figure 3 explains how to check semantic constraints in the example sentences.

In the method using semantic markers, a candidate referent is the proper referent if one of the semantic markers belonging to the candidate referent is equal or subordinate to the semantic marker of the case component. For example, with respect to the zero pronoun in Figure 3, since the *ga*-case component in the verb “*nemuru* (sleep)” has the semantic markers HUM (human being) and ANI (animal) and since “*ojiisan* (old man)” has the semantic marker HUM, the proper referent is judged to be “*ojiisan*.”

In the example-based method, the validity of a candidate referent is decided by the semantic similarity between the candidate referent and the examples of the case component in the verb case frame. The higher the semantic similarity is, the greater the validity is. For example, with respect to a zero pronoun in Figure 3, since the examples of the *ga*-case are “*kare* (he)” and “*inu* (dog),” and since “*ojiisan* (old man)” is semantically similar to “*kare* (he)”, the proper referent is “*ojiisan* (old man).”

These rules, which use semantic relationships to verbs, are also used in the estimation of the referent of demonstratives and personal pronouns.

### Rule using the feature that it is difficult for a noun phrase to be filled in multiple case components of the same verb

#### Candidate enumerating rule 4

When there is “Noun X” in another case component of the verb which has the analyzed case component (the analyzed zero pronoun), {(Noun X, -20)}

### Rule using empathy

This rule is based on empathy theory (Kameyama 86). When an anaphor is a *ga*-case zero pronoun whose verb is followed by an auxiliary verb such as “*kureru*” or “*ku-dasaru*,” the *ni*-case zero pronoun is analyzed first, and

*doru souba-wa kitai-kara* 130-yen-dai-ni joushoushita.  
 (dollar) (the expectations) (130 yen) (surge)  
 (The dollar has since rebounded to about 130 yen because of the expectations.)  
*kono doru-daka-wa oushuu-tono kankei-wo gikushaku-saseteiru.*  
 (the dollar's surge) (Europe) (relation) (strain)  
 (The dollar's surge is straining relations with Europe.)

Rule	Score of each candidate (points)				
	the previous sentence	new individual	130 yen (130 yen)	<i>kitai</i> (expectations)	<i>dorusouba</i> (dollar)
Candidate enumerating rule 2	15				
Candidate enumerating rule 5		10			
Candidate enumerating rule 1			17	15	15
Candidate judging rule 6			-30	-30	-30
Total score	15	10	-13	-15	-15

Figure 4: Example of resolving demonstrative “*kono* (this)”

Table 11: Results

Text	demonstrative	personal pronoun	zero pronoun	total score
Training	87% (41/47)	100% (9/9)	86% (177/205)	87% (227/261)
Test	86% (42/49)	82% (9/11)	76% (159/208)	78% (210/268)

The points given in each rule are manually adjusted by using the training sentences.

Training sentences {example sentences (43 sentences), a folk tale “*kobutori jiisan*” (Nakao 85) (93 sentences), an essay in “*tenseijingo*” (26 sentences), an editorial (26 sentences), an article in “Scientific American (in Japanese)” (16 sentences)}

Test sentences {a folk tale “*tsuru no ongaeshi*” (Nakao 85) (91 sentences), two essays in “*tenseijingo*” (50 sentences), an editorial (30 sentences), articles in “Scientific American (in Japanese)” (13 sentences)}

it is filled with the noun phrase that has high empathy such as the topic, and a *ga*-case zero pronoun is filled with another noun phrase.

## 6 Experiment and Discussion

### 6.1 Experiment

Before pronoun resolution, sentences were transformed into a case structure by a case structure analyzer (Kurohashi & Nagao 94). The errors made by the structure analyzer were corrected by hand. We used IPAL dictionary (IPAL 87) as a verb case frame dictionary. We put together the case frames of the verb phrases which were not contained in this dictionary by consulting a large amount of linguistic data.

An example of resolving the demonstrative “*kono* (this)” is shown in Figure 4, which shows that the referent of the noun phrase “*kono dorudaka* (this dollar’s surge)” was properly judged to be the previous sentence.

By Candidate enumerating rule 2 in Section 3, the system took a candidate “the previous sentence” and gave it 15 points. By Candidate enumerating rule 5 in Section 3, the system took a candidate “new individual” and gave it 10 points. By Candidate enumerating rule 1 in Section 3, the system took three candidates, “130 yen (130 yen)”, “*kitai* (expectations)”, and “*dorusouba* (dollar)”, and gave them 17, 15, and 15 points, respectively. The system applied Candidate judging rule 6 to them. This uses examples of “X no Y”. In this case, it used examples of “X no *dorudaka* (the dollar’s surge of X)”. The only example noun phrase X of this form “X no *dorudaka*” in the EDR occurrence dictionary was “*saikin* (recently)”. All three candidates, “130 yen (130

yen)”, “*kitai* (expectations)”, and “*dorusouba* (dollar)”, were low in similarity to “*saikin* (recently)” in “Bun Rui Goihyou”, and were given -30 points by Table 8. Two candidates, “the previous sentence” and “new individual”, so they are not noun phrases, and were not given points by Candidate judging rule 6. As a result, “the previous sentence” had the highest score and was judged to be the proper referent.

We show the results of our resolution of demonstratives, personal pronouns, and zero pronouns in Table 11. The detailed results for demonstratives are shown in Table 12. The precision rate of zero pronouns is in the case when the system knows whether the zero pronoun has a referent or not in advance.

### 6.2 Discussion

With respect to demonstratives, the precision rate was over 80% even in the test sentences. This indicates that the rules used in this system are effective. But since Japanese demonstratives are classified into many kinds, the precision may be increased by making more detailed rules. In this work we used the feature that “*kono* (this)” rarely functions as a *daikou*-reference. There were four cases analyzed correctly because of this rule.

With respect to personal pronouns, since only first and second personal pronouns appeared in the texts used in the experiment, almost all of the personal pronouns were resolved correctly by estimating the first and second persons in the quotation. The main reason for the errors in the personal pronoun resolution is that the *ni*-case zero pronoun was resolved incorrectly and the second person was estimated incorrectly.

Table 12: Detailed results for demonstrative

Text	demonstrative pronoun	demonstrative adjective	demonstrative adverb	total score
Training	83% (15/18)	86% (19/22)	100% (7/7)	87% (41/47)
Test	82% (14/17)	88% (23/26)	83% (5/6)	86% (42/49)

Table 13: Results of comparison between semantic marker and example-base

	Method 1	Method 2	Method 3	Method 4	Method 5
Demonstrative	87% (41/47)	83% (39/47)	87% (41/47)	83% (39/47)	79% (37/47)
	86% (42/49)	88% (43/49)	88% (43/49)	84% (41/49)	86% (42/49)
Personal pronoun	100% (9/9)	100% (9/9)	100% (9/9)	100% (9/9)	89% (8/9)
	82% (9/11)	64% (7/11)	82% (9/11)	55% (6/11)	64% (7/11)
Zero pronoun	86%(177/205)	83%(171/205)	86%(176/205)	82%(169/205)	66%(135/205)
	76%(159/208)	76%(158/208)	79%(164/208)	75%(155/208)	63%(131/208)

Method 1 : Using both semantic marker and example

Method 2 : Using semantic marker

Method 3 : Using example (using modified codes of *bunrui goi hyou*)

Method 4 : Using example (using original codes of *bunrui goi hyou*)

Method 5 : Using neither semantic marker nor example

There are several reasons for the errors of the zero pronoun resolution: there are errors in Japanese thesaurus "Bunrui goi hyou", Noun Semantic Marker Dictionary, and Case Frame Dictionary.

### 6.3 Comparison Experiment

As mentioned before, we use both the example rule and the semantic marker rule as judging rules. To check which rule is more effective, we made a comparison between the example method and the semantic marker method. The results are shown in Table 13. The upper and lower rows of this table show the accuracy rates for training and test sentences, respectively. The precision of the method using examples was equivalent or superior to that of the method using semantic markers, as shown in Table 13. This indicates that we can use examples as well as semantic markers. Since some codes in BGH are incorrect, we modified them. Since the precision using modified codes was higher than that using original codes, this indicates that the code modification is valid.

## 7 Summary

In this paper, we presented a method of estimating referents of demonstrative pronouns, personal pronouns, and zero pronouns in Japanese sentences using examples, surface expressions, topics and foci. Unlike conventional works, which use semantic markers for semantic constraints, we use examples for semantic constraints and showed in our experiments that examples are as useful as semantic markers. We also proposed many new methods for estimating referents of pronouns. For example, we used the form "X of Y" for estimating referents of demonstrative adjectives. In addition to our new methods, we used many conventional methods. As a result, experiments using these methods obtained a precision rate of 87% in estimating referents of demonstrative pronouns, personal pronouns, and zero pronouns for training sentences, and obtained a precision rate of 78% for test sentences.

## References

- Electronic Dictionary Research Institute, Ltd.: Electronic Dictionary, Japanese Word Dictionary, Version 1.5, (in Japanese), 1995.
- Electronic Dictionary Research Institute, Ltd.: Electronic Dictionary, Japanese Cooccurrence Dictionary, Version 1.5, (in Japanese), 1995.
- Hayashi, S.: Daimeishi-ga Sasumono Sono Sashi-kata, (in Japanese), Unyou I, Asakura Japanese New Lecture 5, Asakura Publisher, pp. 1-45, 1983.
- Information-technology Promotion Agency, Japan: IPA Lexicon of the Japanese Language for computers IPAL (Basic Verbs), (in Japanese), 1987.
- Kameyama, M.: A Property-sharing Constraint in Centering *Proc. of 24th Annual Meeting of ACL*, pp.200-206, 1986.
- Kinsui, B. and Takubo, Y.: Demonstrative, (in Japanese), Hit-suji Shobou, 1992.
- Kurohashi, S. and Nagao, M.: A Method of Case Structure Analysis for Japanese Sentences based on Examples in Case Frame Dictionary *IEICE Transactions on Information and Systems*, E77-D(2), pp.227-239, 1994.
- Matsumoto, Y., Kurohashi, S., Myoki, Y., and Nagao, M.: Japanese Morphological Analysis System JUMAN Manual version 1.0, (in Japanese), Nagao Lab., Kyoto University, 1992.
- Nakaiwa, H. and Ikehara, S.: Intrasentential Resolution of Japanese Zero Pronouns using Pragmatic and Semantic Constraints Viewed from Ellipsis and Inter-Event Relations (in Japanese), *IEICE-WGNLC 95-5*, pp.33-40, 1995.
- Nakao, K.: The Old Man with a Wen, (in Japanese), Eiyaku Nihon Mukashibanashi Series, Vol. 7, Nihon Eigo Kyouiku Kyoukai, 1985.
- The National Language Research Institute: Bunrui Goi Hyou, (in Japanese), Shuuei Publishing, 1964.
- System of "KO/SO/A": (in Japanese), The National Language Research Institute, 1981.
- Takada, S. and Doi, N.: Centering in Japanese: A Step Towards Better Interpretation of Pronouns and Zero-Pronouns, *Proc. of 15th COLING*, Vol.2, pp.1151-1156, 1994.
- Takahashi, T. et al.: Demonstrative, (in Japanese), Nihon-gogaku, vol. 9, Meiji Shoin, 1990.
- Walker, M., Iida, M., and Cote, S.: Japanese Discourse and the Process of Centering, *Journal of Computational Linguistics*, Vol.20, No.2, pp.193-232, 1994.
- Watanabe, Y., Kurohashi, S. and Nagao, M.: Construction of semantic dictionary by IPAL dictionary and a thesaurus, (in Japanese), *Proc. of 45th Convention of IPSJ*, pp.213-214, 1992.
- Yamamura, T., Ohnishi, N., and Sugie, N.: A Classification Scheme of Anaphora in Japanese Demonstrative Pronoun, (in Japanese), *IEICE Transactions on Information and Systems*, J75-D-II(2), pp.371-378, 1992.