

# Intrasentential Resolution of Japanese Zero Pronouns in a Machine Translation System using Semantic and Pragmatic Constraints

Hiromi NAKAIWA, Satoru IKEHARA  
NTT **Communication Science Laboratories**  
1-2356 Take, Yokosuka-shi, Kanagawa-ken, JAPAN 238-03  
{nakaiwa,ikehara}@nttkb.ntt.jp

## Abstract

This paper proposes a method to resolve intrasentential references of Japanese zero pronouns suitable for application in widely applicable and practical machine translation systems. It is highly effective because the volume of knowledge that must be prepared beforehand is not so large and its accuracy is good. This method focuses on semantic and pragmatic constraints such as conjunctions, verbal semantic attributes and modal expressions to determine intrasentential antecedents of Japanese zero pronouns. According to a window test, intrasentential antecedents could be resolved correctly for 98% of the zero pronouns examined using rules consistent with those for intersentential and extrasentential resolution.

## 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.

Several methods have been proposed with regard to this problem (Kameyama;1986) (Walker et al.;1990) (Yoshimoto;1988) (Dousaka;1994). When considering applications for a practical machine translation system for which we can not limit the translation target area, it is not possible to apply these methods directly because their precision of the resolution is low as they only use limited information and because the volume of knowledge that must be prepared beforehand is so large.

The zero pronouns that must be resolved by a machine translation system can be classified into 3 types; (a) 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). Regarding type (b), Nakaiwa et al. (1992) proposed a method to determine the intersentential antecedents using verbal semantic attributes; focus is placed on the dynamic characteristics of verbs and the relationships between verbs. The rules vided in this method are independent of the field of the source text. Therefore, anaphora resolution may be conducted with a relatively small volume of knowledge making the proposed method very suitable for machine translation systems. Furthermore, for type (c), Nakaiwa et

al.(1995) proposed a method to determine the extrasentential referents of Japanese zero pronouns using semantic constraints such as verbal semantic attributes and pragmatic constraints such as modal expressions and types of conjunctions.

In this paper, we propose a widely applicable method to determine the intrasentential antecedents of Japanese zero pronouns (type (a)) that must be determined in a Japanese to English machine translation system using semantic constraints such as verbal semantic attributes and pragmatic constraints such as of conjunction type and modal expressions.

## 2 Zero Pronouns as viewed from Machine Translation

Zero pronouns are very common in Japanese discourse, but the kinds of zero pronouns that actually require resolution vary according to the purpose for which the analysis results are to be used. In the case of machine translation of text, depending on the translation languages, some zero pronouns may not require resolution. This paper considers the task of extracting zero pronouns in a Japanese to English text machine translation system. We first examine the three basic factors important in implementing such a system (Nakaiwa et al.; 1992).

### 2.1 The difference in conception between Japanese and English expressions

Referential analysis of zero pronouns in a Japanese sentence is not necessary if it can be rewritten into an easy to translate Japanese sentence with no zero pronoun (Japanese-Japanese rewriting) (Shirai et al.; 1993) before translating it into natural English.

- (1)  $[\emptyset-ga]^1$        $[\emptyset-o]$       *tsuki-400-dai*      *seisan-suru*  
 $\emptyset$ -SUBJ       $\emptyset$ -OBJ      400-unit-per-month      produce  
They produce 400 units of it per month
- (2)                      *Gessan-ha*<sup>2</sup>                      *400-dai*      *da*  
→      monthly-production-TOP/SUBJ      400 units      is  
The monthly production is 400 units

### 2.2 The difference in case frame patterns between Japanese and English

If an English translation exists where the case with the zero pronoun in Japanese is not translated, then there is no need to resolve the zero pronouns.

- (3)  $[\emptyset-ga]$       *X(FACILITY)-de*      *X(ANIMALS)-o*      *kau*<sup>3</sup>      →      This sentence has no subject  
 $\emptyset$ -SUBJ      X-at                      Y-OBJ      keep      but could be translated into  
    $\emptyset$  is keeping Y at X                                      the expression, “X raise Y”

### 2.3 Restrictions by Voice

Elements which are zero pronouns in Japanese may not need to be resolved, if the voice is changed. For example,

- a) A sentence originally in the passive voice can limit the zero pronouns whose referential elements must be identified by converting the English expression to a passive voice.

<sup>1</sup> When case elements in Japanese become zero pronouns, postpositional particle is not expressed. So, the particle in parenthesis only indicates the syntactic function of zero pronouns and is not expressed.

<sup>2</sup> The postpositional particle “*ha*” which indicates the theme is pronounced as “*wa*” and sometimes transliterated as “*wa*”

<sup>3</sup> The case word in parenthesis indicates the semantic constraints for the case. If symbol “\*” is in parenthesis, it indicates that the case has no semantic constraints by the verb.

- b) Some sentences containing verbs which take the passive voice in Japanese become active in English.

(4)	<i>A(*)-ga-</i>	<i>B (DOCUMENT)-ni</i>	<i>keisai-saseru</i>		This sentence is the passive expression
	A-OBJ	B-in	publish-PASSIVE	→	of “ $\emptyset$ -SUBJ publishes A in B”
		A is published in B			but, it is possible to say “A appears in B”

In a Japanese to English machine translation system, it is important to classify zero pronouns with due consideration of the factors outlined above.

### 3 Appearance of Zero Pronouns in Japanese Texts

To grasp the distribution of zero pronouns and their antecedents that appear in the same sentence, we examine which zero pronouns must be resolved and where their antecedents appear in a test set designed to evaluate the performance of Japanese to English machine translation systems (Ikehara et al.; 1994) considering the analysis presented in section 2. The results of the examination of zero pronouns and their referential elements in the functional test set (3718 sentences) are shown in Table 1. There were a total of 512 zero pronouns. According to this study of the functional test set, 27% (139 out of 512 instances) of the antecedents were in the same sentence

In the 139 instances, zero pronouns that were the subject with the postpositional particle, *ga*, and that are anaphoric to the theme with the postpositional particle, *ha*, are the most common: 102 instances. Of these, 8 were forward-looking anaphora where the zero pronoun comes before the location of their antecedents. These expressions can not be resolved by the anaphora resolution method based on the types of postpositional particles, so to resolve such kinds of zero pronouns, resolution based on the types of conjunctions and so on is needed.

There were 10 cases where zero pronouns whose syntactic function is the same as the syntactic function of their antecedent (for example, the subject becomes a zero pronoun and its antecedent is also a subject in the same complex sentence<sup>4</sup>). We believe that these zero pronouns can be resolved using constraints of the types of conjunctions that control the coverage of shareable case elements.

Furthermore, for zero pronouns whose antecedents are in the same sentence, zero pronouns in embedded or quoted sentences amounted to 9 instances. Zero pronouns whose antecedents are in embedded or quoted sentence amounted to 4 instances. To resolve such kinds of zero pronouns correctly, we need to determine the semantic relationship between embedded sentences or quoted sentences and the main clause in the same sentence using the information such as the types of verbal semantic attributes, modal expressions and nouns modified by the embedded sentences.

## 4 Intrasentential resolution of Japanese Zero Pronouns

Based on the results shown in section 3, we developed a method to resolve Japanese zero pronouns whose antecedents exist in the same sentences.

### 4.1 Intrasentential resolution using syntactic constraints based on postpositional particles

To resolve Japanese zero pronouns whose antecedents appear in the same complex sentence, we can use centering algorithms that rank the center of case elements in the sentence based on

<sup>4</sup> ‘complex sentence’ will be used to refer to both complex and compound sentences in this paper.

Table 1: Distribution of Zero pronouns and their referential elements

Location of zero pronoun	Location of 'referential elements'											Total number of cases	
	Intrasentential						Extrasentential						
	<i>ha</i>	<i>ga</i>		<i>o</i>	<i>ni</i>	misc	Psve	I we	you	HU- MAN	it		misc
	quoted	embedded											
<i>ha</i>	1	0	0	0	0	0	5	0	0	0	2	0	8
<i>ga</i>	102	2	6	2	0	1	151	69	28	23	50	3	444
embedded	3	0	1	0	0	0	15	0	0	2	0	0	21
quoted	4	0	0	0	0	0	0	0	0	0	0	0	4
<i>o</i>	3	0	0	0	3	1	0	0	0	0	11	0	18
embedded	1	0	0	0	0	0	0	0	0	0	0	0	1
<i>ni</i>	1	0	0	0	0	0	2	2	5	0	0	2	12
misc	0	0	0	0	1	0	0	1	1	0	1	0	4
Total	139						373					512	

(Source of Sample Sentences: Test Sentences to Evaluate Japanese to English MT System (3718 sentences).  
Of the test sentences, 463 contained zero pronoun(s) )

the types of postpositional particles and the existence or non-existence of the empathy loaded verbs, and that determine the inheritance of topic between unit sentences. This method can be applied to intrasentential resolution by applying the algorithm to each unit sentence in a complex sentence. Consider the following sentence.

- (5) *kare-ha hooteishiki-o toi-te, [ø-ga] kotae-o dashi-ta*  
 he-TOP equation-OBJ solve-AND ø-SUBJ answer-OBJ calculated  
 He solved the equation and calculated the answer

This sentence can be divided in to two parts, *kare-ha hooteishiki-o toku* 'He solved the equation' and *[ø-ga] kotae-o dashi-ta* 'ø-SUBJ calculated the answer'. The centering algorithm determines the antecedent of the zero pronoun which is the subject of verb *dasu* 'calculate' as *kare* 'he' which is topicalized by the postpositional particle *ha* in the same sentence.

The centering algorithm is very simple and is easy to realize but has insufficient resolution accuracy. For example, the following sentence can not be resolved because this algorithm was designed to resolve only anaphora whose antecedents appear before them in the text.

- (6) *[ø-ga] nawa-o kake-te, kodomotachi-ha ason-da*  
 ø-SUBJ rope-OBJ hung-AND children-TOP played  
 Those children hung a rope and played

In this sentence, the zero pronoun's antecedent is *kodomotachi* 'children' which is the topicalized subject of the subsequent verb *ason-da* 'played'. Because the zero pronouns exist in the unit sentence in the front part of the complex sentence and precede their antecedent, the centering algorithm can not resolve them.

These kinds of zero pronouns can be resolved using the methods which determine the antecedents of zero pronouns which exist in the same sentence depending on the types of conjunctions, verbal semantic attributes and modal expressions. For example, even in sentence (6), the meaning can be estimated as 'children's manner of playing' and the antecedent of zero pronoun can be estimated as 'children' because the conjunction *te* has the meaning that the former unit sentence indicates the manner of the latter unit sentence and the verb governing the zero pronouns indicates action.

## 4.2 Intrasentential resolution using semantic and pragmatic constraints

We analyzed the results shown in section 3 and found the types of conjunctions, verbal semantic attributes and the modal expressions that are useful in determining the intrasentential antecedents of Japanese zero pronouns. In this section, we examine three kinds of semantic and pragmatic constraints: conjunctions, verbal semantic attributes and modal expressions.

### 4.2.1 Constraints based on conjunctions

Conjunctions are expected to be powerful constraints in the determination of intrasentential antecedents of Japanese zero pronouns. This constraint is based on the sharing of cases depending on the types of conjunctions. Minami (1974) and Takubo (1987) proposed the characteristic of sharing coverage of cases depending on Japanese conjunction type. For example Minami divided Japanese conjunctions into three kinds: A, B and C. The complex sentence including A type Japanese conjunctions, such as *tsutsu* 'while' and *nagara* 'while', shares *ha*-case TOPIC and *ga*-case SUBJECT. In the case of B type Japanese conjunctions, such as *nod* 'because' and *tara* 'if', *ha*-case is shared but not the *ga*-case. In the case of C type Japanese conjunctions, such as *keredo* 'but' and *kedo* 'but', neither the *ha*-case nor the *ga*-case are necessarily shared. According to this classification, if a *ga*-case in one unit sentence of a complex sentence joined by an A type Japanese conjunction was to become a zero pronoun, the antecedent of the zero pronoun is a *ga*-case in the other unit sentence in the same sentence. These characteristics of Japanese conjunctions can be used to determine the antecedents of zero pronouns. Some Japanese conjunctions, such as *shi* 'and' and *te* 'and', are ambiguous in terms of their type. For these conjunctions, the meaning of the conjunction must be analyzed using semantic constraints and the type of conjunction must be determined (see 4.2.2(b)) to resolve zero pronouns with intrasentential antecedents.

### 4.2.2 Constraints based on verbal semantic attributes

Constraints based on verbal semantic attributes can be classified into the following two types.

#### (a) Constraints based on the types of verbs

Intrasentential resolution of zero pronouns, in which zero pronouns or their antecedents exist in an embedded or quoted sentence, requires us to determine the semantic relationship between the expression in the embedded or quoted sentence and the expression in main clause in the same sentence. For this purpose, it is effective to use the semantic attributes of verbs. For example,

- (7) *minato-ku-ha*                      *shihon-sanka-suru-to*                      [ $\emptyset$ -*ga*]                      *it-ta*  
minato-ku-TOP                      capital-take-part-in-QUOTE                       $\emptyset$ -SUBJ                      announced  
Minato-ku announced they would take part in capital investment

In this sentence, the *ga*-case SUBJECT of verb *iu* 'announce' in the main clause becomes a zero pronoun and its antecedent is *ha*-case TOPIC in quoted sentence, *minato-ku-ha shihon-sanka-suru* 'Minato-ku take part in capital investment'. In this expression, verb *sanka-suru* 'take part in' in the quoted sentence has the semantic attribute '*ga*-case SUBJECT's attribute transfer' and verb *iu* 'announce' in the main clause has the semantic attribute '*ga*-case SUBJECT transfer mental information expressed in the quoted sentence' Depending on these kinds of pair of verbal semantic attributes, the meaning of this sentence, '*ha*-case TOPIC transfer mental information that the same *ha*-case TOPIC's attribute is changed', can be analyzed and its antecedent can be determined as *minato-ku*. In this manner, even if the antecedents of zero pronouns can not be determined by the types of conjunctions only, semantic relationships between verbs can be determined using semantic attributes of verbs and intrasentential resolution of zero pronouns can be realized.

#### (b) Constraints based on the types of verbs and conjunctions

The constraints based on the types of conjunctions shown in section 4.2.1 can be estimated extremely well. Unfortunately, some Japanese conjunctions are ambiguous in their type. For example, the conjunction *te* has three kinds of ambiguities: A type meaning, a sub clause indicates the manner of main clause; B type meaning, where it indicates temporal relation or causal relation; and C type meaning, where it indicates a coordinate expression. For example, the conjunction *te* in example (6) indicates the manner of children's play and so is A type.

- (8) *kare-ha seichooshi-te, [ø-ga] rippa-ni shinshi-ni nat-ta*  
 he-TOP grow-up-AND ø-SUBJ well gentleman-IND-OBJ became  
 He has grown and matured into a gentleman

In (8), however, the conjunction *te* indicates causal relation and so is B type. Accordingly, for ambiguous conjunctions, the types of conjunction must be identified from the meaning using cooccurrence of verbal semantic attributes and modal expressions in clauses connected by the conjunctions. For example, in the case of sentence (6), because the semantic attributes of verb *kakeru* ‘hung’ and verb *asobu* ‘play’ are the same ‘BODILY ACTION’ and because *ha*-case TOPIC is located after verb *kakeru* ‘hung’, the meaning of the conjunction *te* can be determined as a sub clause indicating the manner of the main clause. In the same way, in the case of sentence (8), because the semantic attributes of verb *seichoosuru* ‘grow up’ and verb *naru* ‘become’ are the same, ‘ATTRIBUTE TRANSFER’, we understand the meaning of the sentence as ‘attribute transfer causes attribute transfer’ and so determine the antecedent of the zero pronoun, *kare* ‘he’. As shown in these examples, verbal semantic attributes are effective for analyzing conjunctions which have ambiguities, and intrasentential antecedents can be determined. Furthermore, this analysis of conjunctions is effective for generating accurate translations of conjunctions.

To write these constraints based on types of verbs effectively, we used the 97 verbal semantic attributes (VSA) proposed by Nakaiwa et al. (1994). These attribute values were added to the 15,000 patterns in the Japanese to English transfer pattern dictionaries used in our Japanese to English machine translation system, ALT-J/E (Ikehara et al.; 1991). The meanings of Japanese verbs determined by the verbal semantic attributes can be effectively limited when verbs are viewed in terms of pattern pairings in Japanese and English.

### 4.2.3 Constraints based on modal expressions

Modal expressions in Japanese are the most powerful constraints in estimating extrasentential referents (Nakaiwa et al.; 1995). For example, in the case of zero pronouns in *ga*-cases SUBJECT, the referent becomes the writer if the sentence has the modal expressions, *-shitai* ‘</> want to -’ HOPE or *-shitehoshii* ‘ø want ø to -’ CAUSATIVE HOPE, and the referent becomes the reader if the sentence has the modal expressions, *-shitewa-ikenai* ‘ø must not -’ PROHIBIT or *-subekida* ‘ø should -’ OBLIGATION. This characteristic is also effective in the intrasentential resolution of zero pronouns. For example,

- (9) *kare-ga tenmongakusha-nanode [ø-ga] ano-hoshi-o shitteiru-daroo*  
 he-SUBJ astronomer-BECAUSE ø-SUBJ that-star-OBJ know-WOULD  
 Because he is an astronomer, he would know that star

In this expression, because the conjunction *nanode* ‘because’ is a B type conjunction, we can not say that *ga*-case SUBJECT usually shares in the same sentence. Thus, these zero pronouns can not be resolved only by the type of conjunction. However, because the verb *shiru* ‘know’ is with the modal expression *daroo* ‘would’ which indicates estimation, *ga*-case SUBJECT can be resolved as the element to be not referring to the writer so *kare* ‘he’ becomes its antecedent. As in this example, even in the case of intrasentential references, the antecedents can be determined by modal expressions.

## 4.3 Algorithm

In this section, we propose an algorithm for the intrasentential resolution of Japanese zero pronouns using the constraints proposed in section 4.2. This algorithm was realized in a Japanese to English machine translation system, so the only zero pronouns that must be resolved are those that become mandatory elements in English. To realize the previously proposed conditions in an algorithm, we must consider cases when these antecedents do not exist in the text, as well as when these antecedents exist in another sentences in the text, and we must design the algorithm to increase the overall accuracy of zero pronouns' resolution.

Anaphora resolution of zero pronouns is conducted as follows<sup>5</sup>.

- (Step 1) Detection of zero pronouns.
  - If they exist, examine the type of sentence that is being analyzed now.
  - If the sentence is a complex sentence, proceed to step 2.
  - If the sentence is a simple unit sentence, proceed to step 3.
- (Step 2) Resolution of zero pronouns in a complex sentence will be conducted in the following order.
  - 1) intrasentential resolution of Japanese zero pronouns using the types of conjunctions verbal semantic attributes and modal expressions (conditions for resolution were proposed in section 3.2.2(b) and 3.2.3).
  - 2) intrasentential resolution of Japanese zero pronouns using the types of conjunctions, (conditions for resolution were proposed in section 3.2.1).
  - If their antecedents can be found, finish the resolution process. Else, proceed to step 3.
- (Step 3) If the sentence currently being analyzed includes an embedded or quoted sentence, intrasentential resolution of Japanese zero pronouns using verbal semantic attributes will be conducted, (conditions for resolution were proposed in section 3.2.2(a)).
  - If their antecedents can be found, finish the resolution process. Else, proceed to step 4
- (Step 4) Examine whether there are antecedents within other sentences in the text.
  - (For example, anaphora resolution is performed using Nakaiwa's method (1992).)
  - If their antecedents can be found, finish the resolution process. Else, proceed to step 5.
- (Step 5) Extrasentential resolution of Japanese zero pronouns using verbal semantic attributes, modal expressions and the types of conjunctions are conducted(Nakaiwa et al.; 1995).
  - If their referents can be found, finish the resolution process. Else, proceed to step 6.
- (Step 6) If referential elements can not be found and translation in the passive voice can be done, translate to the passive voice, else, based on the semantic restrictions imposed on the zero pronoun by the verbs, and deductively generate anaphora elements.
  - Finish the resolution process.

## 5 Evaluation

### 5.1 Evaluation Methods

In this section, we show the results of an evaluation of the proposed method. The target of resolution is the zero pronouns whose antecedents are in the same sentences shown in table 1 (139 entries). The rules to resolve these 139 zero pronouns were created by examining these zero pronouns, that is. the evaluation was conducted using a window test. We conducted the following two types of tests to compare the accuracy and the simplicity of creation with the consistency for all the rules.

To evaluate the former factor, we examined the dependency between the types of conditions in resolution such as conjunctive expression, verbal semantic attributes and modal expression and the accuracy of the resolutions so achieved. We evaluated the accuracy according to the type of constraint used. To evaluate the latter factor, we examined the dependency between the complexities of the rules that were used in the resolution and the accuracy of the resolutions. This examination evaluated the accuracy for simple rules that are easy to create and that can become universal rules.

We used two kinds of evaluation factors, recall and precision, to indicate the accuracy of resolution. Recall is the ratio of the number of zero pronouns that can be successfully resolved to the 139 zero pronouns that need to be resolved. Precision is the ratio of the number of zero pronouns that can be resolved accurately to the total number of matched by the intrasentential

---

<sup>5</sup> In each step in the algorithm, when the referential element within or without the text is determined, the system checks not only the conditions that are written in the 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.

resolution rules. In this evaluation, the conditions for intrasentential resolution were applied to 175 zero pronouns that could be resolved with the constraints proposed by Nakaiwa et al. (1995) (*ga*-case ← ‘I’ or ‘we’, ‘you’, HUMAN, ‘it’ and *ni*-case(indirect object) ← ‘you’; 5 kinds, 175 entries) and 173 zero pronouns that do not need to be resolved if we convert them to the passive voice in English, in addition to the 139 zero pronouns that need to be determined for intrasentential antecedents.

Because the sentences in Table 1 include various kinds of natural Japanese expressions, the rules used in this evaluation are expected to be general and universal.

## 5.2 Resolution Accuracy for Conditions of Resolution

To examine the resolution accuracy for conditions of resolution, we examined the recall and precision of the method proposed in this paper with 4 kinds of conditions: using conditions of conjunctions only, using conditions of conjunctions and verbal semantic attributes, using conditions of conjunctions and modal expressions and using conditions of conjunctions, verbal semantic attributes and modal expressions. We also examined the resolution accuracy when the centering algorithms proposed in section 4.1 were applied to these zero pronouns to compare resolution accuracy objectively.

Table 2 shows the results of the resolution according to the types of rules. As shown in this table, 136 out of 139 zero pronouns were resolved with the rules proposed in section 4.2. The accuracy is higher than the accuracy of the centering algorithm in both recall and precision. By introducing verbal semantic attributes to conjunctions, the accuracy of resolution exceeded the accuracy of the centering algorithm in both recall and precision. Hence, we can say that the verbal semantic attributes are effective for determining the intrasentential antecedents of zero pronouns. Also, the accuracy of recall, when only the conditions of conjunctions were used, was higher than with the centering algorithm. When realizing the resolution process of zero pronouns in machine translation systems, the resolution errors that degrade recall need post-editing to complete uncompleted referential elements. The resolution errors that degrade precision need post-editing to change completed elements to the correct referential elements when the system makes a mistake. Considering both the efficiency of the post-editing processes and the deterioration of translation quality, the decline of precision has more impact than the decline of recall. Thus our method is appropriate for realization in a machine translation system.

Table 2: Resolution Accuracy for Resolution Condition

Resolution Condition		Accuracy	
		Recall	Precision
Our Method	Conjunctions	71% (98)	96% (98/102)
	Conjunctions + VSA	88% (123)	98% (123/125)
	Conjunctions + Modal Exp.	73% (101)	97% (101/104)
	Conjunctions + VSA + Modal Exp.	98% (136)	100% (136/136)
Centering Algorithm		74% (103)	89% (103/116)

(In this table, the numerator in the parenthesis in precision indicates the number of accurately resolved zero pronouns, and the denominator in the parenthesis in precision indicates the number of zero pronouns matched by intrasentential resolution rules.)

## 5.3 Resolution Accuracy against Rule Complexity

We examine recall and precision of the method proposed in this paper against rule complexity. Rule complexity C is evaluated using the following formula based on the constraints used.

$$C = \text{num. of conjunction const.} + \text{num. of VSA const.} + \text{num. of modal const.}$$

According to this formula, the complexity of a rule that has a constraint for conjunctions



and for VSA in the main clause and for modal and VSA in sub clause, becomes 4 (= 1 (conjunction) + 2(VSA) + 1 (modal)).

Table 3 shows the accuracy of the resolution as a function of rule complexity. 70 kinds of rules were used in the extrasentential resolution of 139 zero pronouns as shown in table 3. Rules of complexity of 3 or less yielded a recall of 88% and precision of 99%. Increasing the complexity of 4 or less, increased the recall to 94%: and precision to 100%. This result shows that the use of constraints based on conjunctions, VSA, and modal expressions can achieve high accuracy with relatively simple rules.

Table 3: Resolution Accuracy against Rule Complexity

Resolution Condition			Complexity	Number of Rules	Accuracy	
Conjunctions	VSA	Modal Exp.			Recall	Precision
1	0	0	1	38	71% (98)	96% (98)
1	0	1	2	40(+2)	72%(+1%) (100(+2))	97% (100/103)
1	2	0	3	57(+17)	87%(+15%) (121(+21))	99% (121/122)
1	0	2	3	58(+1)	88%(+1%) (122(+1))	99% (122/123)
1	2	1	4	65(+7)	94%(+6%) (130(+8))	100% (130/130)
2	2	0	4	66(+1)	94%(+1%) (131(+1))	100% (131/131)
2	3	0	5	68(+2)	96%(+2%) (134(+3))	100% (134/134)
2	2	2	6	70(+2)	98%(+2%) (136(+2))	100% (136/136)

(In this table, the numerator in the parenthesis in precision indicates the number of accurately resolved zero pronouns, and the denominator in the parenthesis in precision indicates the number of zero pronouns matched by intrasentential resolution rules.)

## 6 Conclusion

This paper has proposed a powerful method for the intrasentential resolution of Japanese zero pronouns which is suitable for implementation in widely used machine translation systems. It was found that the antecedents that were in the same sentence resolved 98% of the zero pronouns in the window test with the introduction of rules based on three kinds of constraints based on conjunctions, verbal semantic attributes, and modal expressions with consistent rules for intersentential and extrasentential resolution. In the future we will examine the universality of the rules that were created in this paper by applying them to other texts.

## References

- [Dousaka 94] K. Dousaka (1994) Identifying the Referents of Japanese Zero-Pronouns based on Pragmatic Condition Interpretation. *Transactions of Information Processing Society of Japan*, Vol.35 No.10, pp.768-778 (in Japanese).
- [Ikehara et al. 91] S. Ikehara, M. Miyazaki and A. Yokoo (1991) Semantic Analysis Dictionary for Machine Translation. *Technical Reports of SIG on Natural Language Processing, Information Processing Society of Japan*, Vol.84-13 (in Japanese).
- [Ikehara et al. 94] S. Ikehara, S. Shirai and K. Ogura (1994) Criteria for Evaluating the Linguistic Quality of Japanese to English Machine Translation. *Journal of Japanese Society for Artificial Intelligence*, Vol.9 No.5, pp.569-579.
- [Kameyama 86] M. Kameyama (1986) A property-sharing constraint in centering. *Proc. of the 24th Annual Meeting of the Association for Computational Linguistics*. pp.200-206.
- [Kuno 78] S. Kuno (1978) *Danwa no Bunpoo*, Taishukan Publ. Co., Tokyo (in Japanese).
- [Minami 74] F. Minami (1974) *Gendai Nihon-go no Kouzou*, Taishukan Publ. Co., Tokyo (in Japanese).

- [Nakaiwa et al. 92] H. Nakaiwa and S. Ikehara (1992) Zero Pronoun Resolution in a Japanese to English Machine Translation System by using Verbal Semantic Attributes. *Proc. of ANLP'92*, pp.201-208.
- [Nakaiwa et al. 94] H. Nakaiwa, A. Yokoo and S. Ikehara (1994) A System of Verbal Semantic Attributes Focused on the Syntactic Correspondence between Japanese and English. *Proc. of COLING'94*, pp.672-678.
- [Nakaiwa et al. 95] H. Nakaiwa, S. Shirai, S. Ikehara and T. Kawaoka (1995) Extrasentential Resolution of Japanese Zero Pronouns using Semantic and Pragmatic Constraints. *Proc. of AAAI 1995 Spring Symposium Series, Empirical Methods in Discourse Interpretation and Generation* (to appear).
- [Shirai et al. 93] S. Shirai, S. Ikehara and T. Kawaoka (1993) Effects of Automatic Rewriting of Source Language within a Japanese to English MT System. *Proc. of TMI'93*, pp.226-239.
- [Takubo 87] Y. Takubo (1987) Tougo Kouzou to Bunmyaku Jyohou, *Nihon-go gaku*, Vol.6 No.5, pp.37-47 (in Japanese).
- [Walker et al. 90] M. Walker, M. Iida and S. Cote. (1990) Centering in Japanese Discourse. *Proc. of COLING'90*.
- [Yoshimoto 88] K. Yoshimoto (1988) Identifying zero pronouns in Japanese dialogue. *Proc. of COLING'88*, pp.779-784.