Zero Pronoun Resolution in a Japanese to English Machine Translation System by using Verbal Semantic Attributes.

Hiromi Nakaiwa and Satoru Ikehara
NTT Network Information Systems Laboratories
1-2356 Take Yokosuka-Shi Kanagawa 238-03 Japan

Abstract

A method of anaphoral resolution of zero pronouns in Japanese language texts using the verbal semantic attributes is suggested. This method focuses attention on the semantic attributes of verbs and examines the context from the relationship between the semantic attributes of verbs governing zero pronouns and the semantic attributes of verbs governing their referents. The semantic attributes of verbs are created using 2 different viewpoints: dynamic characteristics of verbs and the relationship of verbs to cases. By using this method, it is shown that, in the case of translating newspaper articles, the major portion (93%) of anaphoral resolution of zero pronouns necessary for machine translation can be achieved by using only linguistic knowledge.

Factors to be given special attention when incorporating this method into a machine translation system are examined, together with suggested conditions for the detection of zero pronouns and methods for their conversion. This study considers four factors that are important when implementing this method in a Japanese to English machine translation system: the difference in conception between Japanese and English expressions, the difference in case frame patterns between Japanese and English, restrictions by voice and restriction by translation structure. Implementation of the proposed method with due consideration of these points leads to a viable method for anaphoral resolution of zero pronouns in a practical machine translation system.

1 Introduction

In all natural languages, components that can be easily deduced by the reader are frequently omitted from expressions in texts. In Japanese in particular, the subject and object are often omitted. These phenomena cause problems in machine translation because components not overtly indicated in the source language (i.e. Japanese) become mandatory elements in the target language (i.e. English). Thus, in Japanese to English translation systems it becomes necessary to identify corresponding case elements omitted from the Japanese original (these are referred to as "zero pronouns") to be translated into English expressions. Therefore, the technique of zero pronoun resolution is an extremely important function.

Several methods have been proposed with regard to this problem. Grotz et al. proposed the method of resolving definite noun phrases by using a centering algorithm. Kameyama expanded this concept by introducing property sharing constraints and applied it to zero pronoun resolution in Japanese. This method relies on the types of postpositional particle and whether there are any empathy-loaded verbs to exercise control over priority rankings for the focus of discourse segments.

Yoshimoto suggested a method that uses topics from a dialogue. This method has focused attention on the characteristic of the Japanese language where the case for the sentence is determined by the type of postpositional particle (e.g. "*ha*" (pronounced "wa"), "*ga*", "*wo*" and "*ni*" indicate the theme, subject, direct object and indirect object respectively). The method uses case elements accompanied by the postpositional particle "*ha*" and case elements become the theme or subject matter through expressions governed by a special sentence structure pattern.

Kuno classified zero pronouns into two categories (pseudo-zero, real-zero) and suggested separate resolution methods for each category. This method handles pseudo-zero pronouns (omitted by across-the-board discourse deletion) and real-zero pronouns (topicalized noun phrase or a noun phrase existing in a dialogue scene which can become a referent, somewhat resembling personal pronouns in the English language) separately from the point of the referent detection method.

The foregoing methods of anaphoral resolution can be divided into two major groupings. One uses comparatively superficial information such as the types of postpositional particles or the existence / non-existence of interjections. The other introduces the concepts of plans and scripts. When considering application to machine translation, the former leads to problems in the precision of resolutions because it is restricted to using specified information. The latter needs common knowledge and world models and to develop a translation system handling texts over a broad field, the volume of knowledge to be prepared beforehand is so large that this method can be regarded as impossible to realize.

Thus in this paper, attention has been focused on verbal semantic attributes. We propose a method of resolving zero

pronouns common in Japanese discourse. The method uses the dynamic characteristics of verbs and the relationship between verbs. The rules needed by this method are independent of the fields of the source text. Therefore, anaphora resolution may be conducted with a relatively small volume of knowledge, so the proposed method is very suitable for machine translation.

2 Zero Pronouns as viewed from Machine Translation

Zero pronouns are very common in Japanese discourse, but the number of zero pronouns that actually require resolution varies according to the purpose for which analysis results are to be used. For example, the case of a question and answer system involving a task such as replying to questions from a user who has just read a sentence. The questions, which can come from several points of view, must be anticipated, and practically all of the zero pronouns in the sentence will require resolution. In contrast, in the case of machine translation of text, depending on the translation languages, zero pronouns requiring resolution tend to limited. This paper considers the task of extracting zero pronouns in a Japanese to English text machine translation system. We first examine the four basic factors important in implementing such a system.

2.1 The difference in conception between Japanese and English expressions

When extracting zero pronouns in machine translation, whether the zero pronouns require resolution analysis or not needs to be decided. For example, in the sentence.

(1)*X-sha ha 2-gatsu-1-nichi, ha-dodhisuku-shouchi wo*
CompanyX TOP February 1 hard disc device OBJ
hatsubai-suru.
place on sale
"Company X will put on sale the hard disc device from February 1."

øsubj øobj tsuki-400-dai seisan-suru.
400 units per month produce
"They produce 400 units of it per month."

The second sentence has a structure that is centered around the verb "*seisan-suru*(produce)" and the subject and object have become zero pronouns. But to translate the sentence into natural English, there is a need to rewrite it into a predicate noun sentence ("*da*" sentence, so called because of the original Japanese "*Gessan wa 400 dai da*") to read

(2) *Gessan ha 400-dai da.*
Monthly production TOP/SUBJ 400 units is
"Monthly production is 400 units".

To translate the expression in this form, referential analysis of the zero pronouns of the subject and object of the verb "produce" is no longer necessary. When translating this type of expression, the syntactic/semantic structure of the sentence to be translated is first converted into an English type structure in the source language (This is makes the Japanese-Japanese conversion) in an analysis phase. Selection of only zero pronouns whose referent needs to resolved becomes possible.

2.2 The difference in case frame patterns between Japanese and English

There are verbs, the case elements of which are mandatory in Japanese but optional when translated into English. For example, an expression such as,

(3) *X* (facility) *de* *Y* (animals) *wo kau.*
X at Y OBJ keep
" At Y(facility), X(animals) are being kept."

in which there is no subject in Japanese, it would be possible to translate this by using the expression," X raise Y". In cases such as this, it would be useful to prepare case patterns to be used for syntactic analysis for each and every translation of English verb form and designate the English case structure when analyzing the Japanese. Elements which do not become mandatory cases in English will then not be mandatory cases in Japanese either. Thus deciding which zero pronouns must be analyzed can be done accurately.

2.3 Restrictions by Voice

Elements which have become zero pronouns in Japanese will, if the voice can be changed to give natural English, not need to be resolved. For example,

- A sentence originally in the passive voice

In this case, converting the English expression to passive voice will limit the zero pronouns for which the referent must be identified.

- Sentences containing verbs which take the passive voice in Japanese become active in English. For example, the expression,

(4)*A ga B* (document) *ni keisai-sareru.*
A OBJ B in publish-PASSIVE
" A is published in B."

is the passive expression of "øsubj publishes A in B" in which the subject has become a zero pronoun. In English, however, even though there is no subject in Japanese, it is possible to translate this to the expression "A appears in B". In cases such as this, case frame patterns must be prepared by modifying the English language to be used in syntactic analysis. When analyzing the Japanese, it is possible to limit the number of zero pronouns which must be resolved by limiting mandatory case patterns to those instances that are accompanied by passive aspects which are mandatory cases in the English case pattern.

2.4 Restriction by translation structure

In the expression,

(5) *X-sha ha haadodhisuku-souchi wo hatsubai-suru.*
Company X TOP hard disc device OBJ place on sale
"X Company will place on sale the hard disc device."

øsub sofuto wo OS ni Kumikomu-koto de
software OBJ OS into incorporate-EMBEDDED by
setsuzoku-daisuu wo fuyasi-ta
number of units to be connected OBJ increase-PAST
"They increased the number of units to be connected by incorporating the software into the OS."

the verbs "incorporate" and "increase" have turned the subject into a zero pronoun. The sentence with "*Kumikomu-koto*(incorporate-EMBEDDED)" is structured as an "embedded sentence" modifying the action "*koto*". Translated into English, the portion "*koto de*" becomes the methodical case "by incorporating software into the OS" and assumes a gerund phrase expression. That is the embedded sentence in Japanese becomes a prepositional phrase accompanied by a gerund phrase. Because different sentence structures are generated between Japanese and English, zero pronouns need to be extracted by converting the Japanese original to an English like syntactic/semantic structure.

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 Newspaper Articles

With due consideration of the conditions as presented in Chapter 2, we examine where troublesome zero pronouns and their referents appear in newspaper articles. Newspaper articles generally tend to use compressed forms of expressions. Thus, declinable words are frequently turned into nouns by compressing the declinable suffixes. Thus, more often than not, it is impossible to determine the zero pronoun's referent merely by relying on postpositional particle information, themes or the types of empathy-loaded verbs. For example,

(6) *NTT ha shingata-koukanki wo dounyuu-sita.*
NTT TOP new model switchboard OBJ introduce
"NTT will introduce a new model switchboard."

øsubj jiko-shindan-kinou wo tousai,
self checking function OBJ equip
"The new model switchboard is equip with self checking function and"
øsubj 200-shisutemu wo secchi-suru yotei-da.
200 systems OBJ install be-planning-to
"NTT is planning to install 200 systems."

In the first sentence, the subject is topicalized, but in the second sentence, the subject of the first portion of the sentence and the subject of the latter portion of the sentence are zero pronouns. Of the two zero pronouns, in the former case, the "*shingata-koukanki*"(new model switchboard), which is the object of the former sentence, and in the latter case, "*NTT*" , which is the subject of the former sentence become the referents. Thus, when there are elements which have been topicalized, and there are no other elements that can be topicalized, it cannot be taken for granted that topicalized elements will become the resolution elements for zero pronouns. Under such circumstances, there is a need for information other than whether the element has been topicalized or not, such as further semantic restrictions.

The lead paragraphs in 29 newspaper articles, totaling 102 sentences in all, were examined for zero pronouns and their referents, and the results are shown in Table 1. There were 88 cases of zero pronouns. According to this study, the case where elements topicalized by the postpositional particle "ha" in the first sentence became the referents of zero pronouns when being made the subject in the second sentence, was most common, with 45 instances (51%). Furthermore, zero pronouns having referents in the first sentence, totalled 76 instances (86%). With newspaper articles, the first sentence contains information that gives an outline of the entire article and thus the case element tends to become the referent. There were 67 instances (74%) of zero pronoun referents in the second and following sentences being used by the first sentence amounted to 67 instances(74%) which strongly suggests the importance of the first sentence.

Referent Appearnce location* / Zero Pronoun Appearnce location		1st sentence				2nd sentence and thereafter. Within Same Sentence				2nd sentence and thereafter. Not in the Same Sentence				Non in the Sent-ence	Sub Total [Cases]
		Ha	Ga	Wo	Etc.	Ha	Ga	Wo	Etc.	Ha	Ga	Wo	Etc.		
1st sent-ence	SUBJ	6	0	1	0	----	----	----	----	----	----	----	----	1	9
	OBJ	0	0	1	0	----	----	----	----	----	----	----	----	0	
	ETC	0	0	0	0	----	----	----	----	----	----	----	----	0	
2nd Sentence and after	SUBJ	45	4	12	1	7	0	0	1	0	0	0	0	3	79
	OBJ	0	0	6	0	0	0	0	0	0	0	0	0	0	
	ETC.	0	0	0	0	0	0	0	0	0	0	0	0	0	
Sub Total [Cases]		76				8				0				4	88

Table 1 Frequency of Appearance of Zero Pronouns and Their Referents
(Source of Sample Sentences: Nikkei Sangyo Newspaper, Information column,lead paragraphs during February, 1988. 29 articles (102 sentences) 2-8 sentences per article.
Of the newspaper articles tested, the number of sentences with zero pronoun(s) contained was 56 out of 102.)
* "Ha"(pronounced "Wa"),"Ga","Wo", which are postpositional particles in Japanese,respectively indicating the theme, subject, direct object.

Moreover, there were 12 instances (14%) where the referent was neither the theme nor the subject; the zero pronoun is the subject. From this, it can be observed that it would be inappropriate to rely solely on the technique of selecting the referent from case elements that have been topicalized or of determining the order of priorities for resolution elements from the type of postpositional particle. These 12 instances were studied further and found to contain verbs that included the referent. Such verbs were "*hatsubaisuru*" (sell), "*kaisetsusuru*" (establish), "*kaihatsusuru*" (develop) and other such words intended to introduce new object elements. Verbs for zero pronouns tend to be a noun predicate as in "*LAN da*" (That is LAN) -- [In English, it would correspond to the expression, "ø be <noun>"] or, to words such as "belong to" indicating attributes. To resolve this type of zero pronoun, it would appear essential that verb attributes be categorized and the zero pronoun referent be determined from the relationships of verbal semantic attributes.

4 Classification of Verbal Semantic Attributes

As mentioned in the preceding chapter, the resolution of certain types of zero pronouns that could not be dealt with by conventional methods, may now be resolved by using semantic information. Therefore, in this chapter, the verbal semantic attributes will be categorized for the purpose of resolving zero pronouns using only linguistic knowledge (i.e. not world knowledge). The referent of zero pronouns will be determined by the relationship between attributes.

Japanese verbs will be categorized using the following 2 viewpoints.

Verb Categorization Standards

•Dynamic Characteristics of Verbs
Categorization based on the inherent concepts of verbs and the reaction brought about to discourse situation by the verbs
Ex. "*motsu*"(to have) --- Possession
"*kaihatsusuru*"(to develop) --- Production

•Relationship of Verbs to Cases
Ex."*kanseisuru*":SUBJ be completed->SUBJ be produced
"*kaihatsusuru*":SUBJ develop OBJ->SUBJ produce OBJ

The conceptual system of verbs as categorized by these standards is shown in Figure 1.

Next, we consider the relationship between verbs, by examining the information regarding the relationships within sentences containing zero pronouns and assess whether this information will be furnished anew to sentences containing the referent. The verbal semantic attribute (VSA) between verbs governing the referent and the verb governing the zero pronoun can be summarized in the form shown in Table 2. The use of this relationship will make it possible to make an assumption of verbal relationship and to determine the referential elements of zero pronouns based on the relationship of the two factors of verbal semantic attributes.

As mentioned in Chapter 3, the first sentence of the lead paragraph in a newspaper article often consists of a discourse structure that presents an outline of the contents of the entire article. Here, we shall refer to a unit sentence of this type as a "topicalized unit sentence", and based on its semantic attributes, the referents of zero pronouns in sentences that follow will be selected.

By relying on the categorization of verbal semantic attributes, and observing the rules for determining the referential elements of zero pronouns as described by its attribute value, we find that it is possible to describe multi-purpose anaphora resolution analysis rules which do not rely on the target domain of the analysis. Thus because, the information that is required for analysis is contained within the scope of linguistic knowledge, anaphora resolution of zero pronouns using this method can be applied to machine translation.

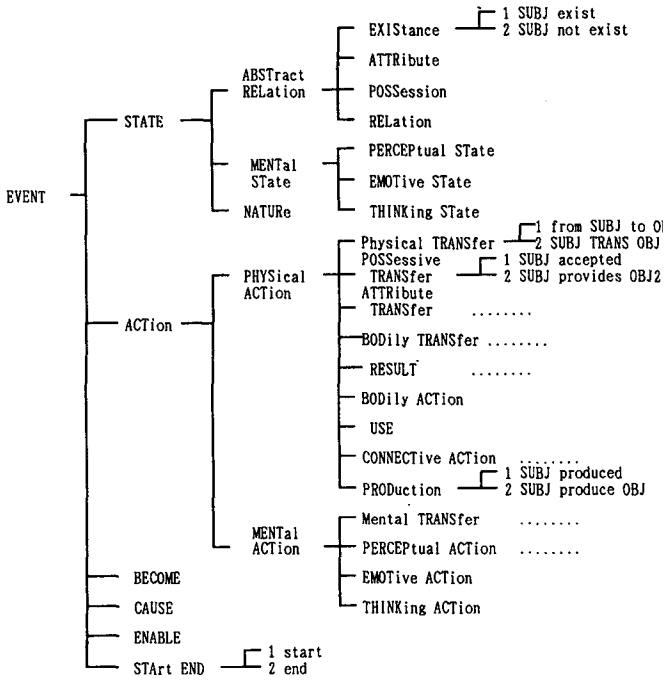

Figure 1 System of Verbal Semantic Attributes

Conditions for zeropronouns		Conditions for referents	Verbal Relationship	Assumed referents
VSA	case	VSA		
POSS	Subject	POSS-TRANS1 & START	Detailed explanation	Object
THINK-ACT	Subject	POSS-TRANS1 & START	Policy decision	Subject
.....	.....	.....	.....	.....

Table 2 Rules for Determining Resolution Elements by Verbal Semantic Categories

5 Format of Anaphoral Resolution

5.1 Algorithm

The structure of the system for resolution of zero pronouns using verbal semantic attributes is shown in Figure 2. The Japanese sentence to be analyzed has already undergone morphological analysis, syntactic/semantic analysis, and the results are input to context analysis. In context analysis, anaphora resolution of zero pronouns is conducted as follows.

(Step 1) --Detection of zero pronouns.
If they exist, examine whether there are referents within the same sentence.
If they exist, and resolution is concluded, proceed to Step 4

Resolution of referents within the same sentence relies on two types of methods.

1) Anaphoral resolution of zero pronouns based on the type of conjunction

2) Anaphoral resolution based on verbal semantic attributes

The first method uses constraints where anaphoral elements determine the syntactic structure depending on the type of postpositional particle and of conjunctions. A portion of the rules for determining anaphoral elements depending on the type of conjunctions is shown in Table3. The second method is when, within the same sentence, anaphoral elements cannot be determined based on conjunctions (for example, when three or more types of unit sentences exist within the same sentence), anaphoral resolution is then conducted using VSA.

(Step 2)--When they do not exist within the same sentence,referent candidates are selected from among the case elements of topicalized unit sentences that are retained within the contextual information stage sector. The standard for selection will be based on the relationship between VSA of verbs governed by zero pronouns and VSA of topicalized unit sentences and on the rules for designating verbs given in Table 2.
When constraints by verbs are satisfied, anaphoral relationships become valid and proceed to Step 4.

(Step 3)--When the referent cannot be detected, handle as "processing impossible".
Based on the semantic restrictions imposed on the zero pronoun by the verbs, conjecture anaphoral elements.

(Step 4)--From the knowledge base for sentence structure control, use the rules for extraction of topicalized unit sentences determined by relying on the sentence structure of target field of analysis[1] to select the topicalized unit sentence and have the context information retaining sector retain the sentence.
Proceed to the next sentence.

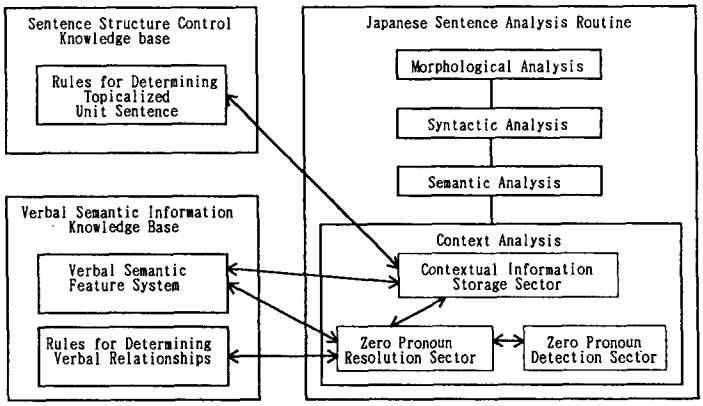

Figure 2 Structure of This System

ExampleofConnecting Words of Zero Pronouns	ConstainttotheCaseMarker	Connection with Referents*
kara"(because),*"shi"*(and) ,*"ba"*(if..then..)	*"ha"*(TOP/SUBJ)	sub sent. ->main sent.
"tame"(so tnat)	*"ha"*(TOP/SUBJ)	sub sent.<-->main sent.
"mama"(wile)	*"ha"*(TOP/SUBJ), *"ga"*(SUBJ)	sub sent. -> main sent.
"tari"(and),*"te"*(after)	*"ha"*(TOP/SUBJ), *"ga"*(SUBJ)	sub sent.<-->main sent.
"to"(when)	*"ha"*(TOP/SUBJ), *"wo"*(OBJ)	sub sent. -> main sent.
"tsutsu"(while),** *"nagara"*(while)**	*"ha"*(TOP/SUBJ), *"ga"*(SUBJ), *"wo"*(OBJ)	sub sent.<-->main sent.

Table 3 Constraints to Zero Pronouns and their referent with Connecting Words

* The arrows go from the sentence which include referents to the sentence including the zero pronouns capable of correspondence.

** In the case of *"tsutsu"* and *"nagara"*, the *"wo"* case will become the target of referents only when its connection is "CONTRARY-AFFIRMATIVE"(This type of connection is translated as "although" in our system)

5.2 Examples

Using the example sentence (6) and using the technique mentioned here, an example of zero pronoun resolution is given in (7).

(7) *NTT ha shingata-koukanki wo dounyuu-sita.*
NTT TOP new model switchboard OBJ introduce
"NTT will introduce a new model switchboard."
Topicalized Unit Sentence:
(introduce (VSA (POSS-TRANS2 & START))
(SUBJ "NTT")(OBJ "new model switchboard"))

[1] In the case of newspaper articles, the first sentence in the article becomes the topicalized unit sentence. When the first sentence consists of a number of unit sentences, set an order of priority for the topicalized unit sentence depending on the type of conjunction used. Specifically, in the case of compound sentences, rules such as the main sentence taking precedence will be applied

øsubj jiko-shindan-kinou wo tousai,
self checking function OBJ equip
"The new model switchboard is equipped with a self checking function and"
(equip (VSA (POSS))
(SUBJ øSUBJ) (OBJ "self checking function"))
øSUBJ= "new model switchboard"

øsubj 200-shisutemu wo secchi-suru yotei-da.
200 systems OBJ install be-planning-to
"NTT is planning to install 200 systems."
(be-planing-to (VSA (THINK-ACT))
(SUBJ øSUBJ) (OBJ ....))
øSUBJ = "NTT"

Topicalized Unit Sentence:
(introduce (VSA (POSS-TRANS2 & START))
(SUBJ "NTT")(OBJ "new model switchboard"))

The results of analyzing the first sentence are used to extract the topicalized unit sentence. In example (7), the first sentence is structured from the unit sentence and the result of analysis is stored in the context information storage sector as the topicalized unit sentence. Next, from the analysis results of the second sentence, it can be understood that the subjects of "*tousaisuru* (is outfitted with or equipped with)" and "*yoteida* (is planning to)" have been converted to zero pronouns. Since there are no referents within the same sentence, the case element within the topicalized unit sentence becomes the referent candidate. The VSA of "*tousaisuru*" and "*yoteida*" are respectively, "POSS", "THINK-ACT", and the VSA of topicalized unit sentence verb are "POSS-TRANS2" and "START". Thus, according to the rules given in Table 2, "Detailed explanation" and "Policy decision" are established as the verbal semantic relationships and the object and subject of the topicalized unit sentence respectively, and become the referents.

6 Implementation in a Machine Translation System

The following is an outline of the processing undertaken by the Japanese to English machine translation system, ALT-J/E (See Figure 3). First, a morphological analysis of the input Japanese sentence is conducted, followed by a dependency analysis of elements in the sentence. Unit sentences[2] are extracted based on results of the relationships between verbs, and from these a simple unit sentence[3] is extracted. Subjective expression information such as modality , tense and aspect is extracted from the simple unit sentence to yield the objective simple unit sentence. This objective simple unit sentence, as shown in Figure 4, is collated with two types of pattern dictionaries having predicates as index words (the idiomatic expression transfer dictionary and the semantic valentz pattern transfer dictionary). When there is no appropriate pattern, a general pattern transfer rule is applied. This determines the syntactic and semantic structure pattern that is used in Japanese to English conversion. In the cases of (3) and (4) in Chapter 2,

[2]a unit sentence is a part of the sentence in which the tree structure is centered around one predicate in the sentence; there are occasions when embedded sentences are included in a unit sentence.

[3]a simple unit sentence is one where a unit sentence has been parsed to the level where it has only one predicate. .
(Ex.(in English)
"This is the only paper that contains the news" <- unit sentence
"This is the only paper" , "the only paper contains the news"
<- simple unit sentences)

(1) Morphological analysis:
Separation of words, determination of words part of speech
(2) Dependency analysis:
-Determination of relations between sentence elements
(3) J-J conversion:
-Conversion of expressions within Japanese
(4) Simple sentence extraction:
-Determining the scope of influence of all predicates from dependency analysis results
(5) Simple sentence analysis:
(5.1) Predicate analysis:
-Extraction of modality and other elements and conversion to an ordinary sentence
(5.2) Gerund phrase analysis:
-Determination of semantic structure of gerund phrases and compound words
(6) Embedded sentence analysis:
-Determination of the semantic structure of embedded sentences
(7) Ordinary sentence conversion to English:
-Conversion of objective expression by means of pattern dictionary
(8) Connection analysis:
-Determination of relations between declinable words
(9) Optimal result selection:
-The best(semantically and syntactically most plausible) interpretation is selected
(10) Zero anaphora resolution:
-Resolution of zero anaphora by use of contextual information
(11) Resolved element conversion:
-Determination of the conversion method for resolved zero anaphora
(12) Unit sentence generation:
(12.1) Basic structure generation:
-Determination of the structure of the entire English sentence
(12.2) Adverbial phrase generation:
-Determination of adverbial phrase translation from modality, tense, verb and other elements
(12.3) Noun phrase generation:
-Conversion of phrase and compound word structures and embedding of embedded sentences
(13) Connecting structure generation:
-connection of the unit sentences according to connection attributes and the presence or absence of a subject
(14) Modality tense structure generation:
-Insertion of auxiliary verbs and infinitives, transformation of word model / syntactic structure
(15) English sentence coordination:
-Contraction, setting of determiner

Figure 3 Process Outline of Japanese-English Machine Translation System, ALT-J/E

[Example of Idiomatic Expressions]
(1) Example of idiomatic phrase pattern
X(Subject) *ha se ga takai* => X be tall
X TOP back SUB high

(2) Example of functional verb combination
X (subject) *ha Y*(subject) *no hinan wo abiru*
X TOP Y by criticism OBJ be-subjected-to
" X (subject) is subjected to criticism by Y"
(-> X is criticized by Y) | Conversion within
(-> Y criticizes X + passive) |Japanese language
(=> Y claim X (+passive) |Application of Japanese to
| English conversion pattern
=> X be claimed by Y. | Transformation of English

[Example of Semantic Combined Value Pattern]
X (subject) *ga Y* (cultural, human activity) *wo anki-suru.*
X SUBJ Y OBJ memorize
=> X learn Y by heart.
"X(subject) memorizes Y (cultural, human activity)."

X (facility) *de Y* (animals) *wo kau*. => X raise Y
X at Y OBJ be-kept
"Y (animals) are kept at X (facility). "

X (subject) *ga Y* (food) *wo taberu*. => X eat Y
X (subject) SUBJ Y OBJ eats
"X (subject) eats Y (food) ."
Ex. Y =<niwatori> => Y = chicken
(1) bird ..... hen
(2) food ... chicken

Figure 4 Example of Application of Japanese-English Conversion Pattern Dictionary

they are not identified during processing as cases of zero pronouns. If numerous interpretations remain at this point, a single and final interpretation is decided on, based on the results of interpretation of the pattern at the objective simple unit sentence level. Also, as seen in (1) and (7) of Chapter 2, when there is a wide difference between the structures in Japanese and English, converting the Japanese structure resulting from analysis to a structure as close as possible to the English expression can make it possible to avoid referential analysis; only the zero pronouns that are used in the English translation need to be treated. If, after the foregoing analysis, zero pronouns still remain, anaphora resolution using the context is conducted as shown in Chapter 5. At this stage, the sentence pattern used in generating the unit sentence is established and all that remains is to use this to generate the backbone expression in English, adding other relevant information such as modality, tense and conjunction. In doing so, care should be taken to avoid the situation where extracting zero pronouns after correspondence analysis results in verbose English. In this case elliptical pronouns and definite articles should be used.

7. Evaluation

The 102 sentences from 29 newspaper articles' lead paragraphs, as introduced in Chapter 3, were used as target sentences; the results of processing zero pronouns, appearances, and rate of resolution in analysis, are shown in Table 4. The rate of success in anaphoral resolution by this method including zero pronouns outside the scope of target processing (referent not appearing within the text) was about

Referent appearnce location / Zero Pronouns appearnce location		1st sentence				2nd sentence and thereafter Within same sentence				2nd sentence and thereafter. Not in the same sentence				None in the Sentence	Sub Total Cases
		Ha	Ga	Wo	Etc.	Ha	Ga	Wo	Etc.	Ha	Ga	Wo	Etc.		
1st Sentence	SUBJ	6 / 6	0	0 / 1	0	--	--	--	--	--	--	--	--	0 / 1	7 / 9 [78%]
	OBJ	0	0	1 / 1	0	--	--	--	--	--	--	--	--	0	
	ETC.	0	0	0	0	--	--	--	--	--	--	--	--	0	
2nd Sentence and after	SUBJ	45 / 45	4 / 4	12 / 12	0 / 1	7 / 7	0	0	1 / 1	0	0	0	0	0 / 3	75 / 79 [95%]
	OBJ	0	0	6 / 6	0	0	0	0	0	0	0	0	0	0	
	ETC.	0	0	0	0	0	0	0	0	0	0	0	0	0	
Sub Total [Cases]		74 / 76 [97%]				8 / 8 [100%]				0				0 / 4 [0%]	82 / 88 [93%]

Table 4 The Frequency of Successful Resolution of Zero Pronouns by This Method
* With the fractions in the above table, the denominator denotes the number of cases of zero pronouns occurrence, and the numerator the number of cases of zero pronouns succeeding in resolution.

93%. The rate exclusive of the zero pronouns outside the scope of target processing was as high as 98%.

Examples of failure in anaphoral resolution are shown below. They fall into 2 types, those where world knowledge is necessary (a), and those where the referent appears in the sentence so that analysis is possible by converting the sentence structure in J-J conversion (b,c). In (b), however, a rule for anaphoral resolution that handles it as a different sentence within the same sentence is necessary. In (c), the sentence structure of the topicalized unit sentence needs to be changed to "*---ha ---sisutemu wo hanbaishi-hajimeru.*"(--- will begin selling the --- system) thus changing the case of "*--- sisutemu no*"(of the --- system).

•Examples of supplement processing failures

:(Total 6 cases)

(a) Those requiring worldwide knowledge (common sense)
---- 4 cases

e.g.

(9) *øsubj ofukon ni natte, ---*
the office computer IND-OBJ becoming
"(the mainstream product type)
becoming the office computer, ---"
(øsubj =the mainstream product type)

(10) *A-sha ga matome-ta*
Company A SUBJ gather-PAST
densen-toukei ..... niyoruto,
data wire and cable statistics according to
"According to data wire and cable statistics gathered by Company A, "
øsubj kouchou wo tsuzuke-teiru.
prosper OBJ continue to
"(the wire and cable industry) continues to prosper"
(*øsubj* =the wire and cable industry)

(b) The case element of "*wo*" case within the same sentence becomes the referent of "ga" case of zero pronouns residual B.----------1 case

e.g.

(11) *A-sha ha B-eigyousho wo shinsetsu,*
company A TOP Sales Office B OBJ open newly
" Company A will open its new sales office B and "
øsubj 2-gatsu-1-nichi kara .. eigyou wo hajimeru
February 1 from sales activities OJB begin
" (Sales Office B) begin sales activities from February 1."
(*øsubj* =Sales Office B)

(c) A noun modifying another noun by "*no*" turns it into a supplement candidate. 1 case

e.g.

(12) *--- ha --- sisutemu no hanbai wo hajimeru.*
TOP system of sales OBJ begin
"--- will begin sales of --- system"
øsubj ha --- no-mono
TOP belongs to
"(the --- system) belongs to ---"
(*øsubj* = the --- system)

8. Summary

This paper has suggested a powerful method for anaphoral resolution using VSA to deal with the zero pronouns appearing in Japanese texts. With previously suggested methods, it was difficult to realize pronominal resolution of zero pronouns in a practical translation system due to the huge volume of knowledge necessary (common sense and world knowledge). In contrast, the proposed method, which utilizes semantic attributes of categorized verbs, makes it unnecessary to describe rules unique to various fields. With a comparatively limited volume of knowledge, it is thus possible to anaphorically resolve zero pronouns. This method has been realized in the machine translation system ALT-J/E. ALT-J/E was assessed by processing common Japanese newspaper articles. It was found that 93% of the Japanese zero pronouns requiring anaphoral resolution had their referents determined correctly.

One possible application of this method in context processing would be to generate an abridged text based on a structural analysis of sentences in the entire article and categorization of contents of the articles focusing on the VSA of the first sentence in each text.

In this report, the target sentences were limited to newspaper article lead paragraphs and comparatively short sentences. In the future, studies need to be made on changes in topic and sentences with a complicated discourse structure.

References

Susumu Kuno. *Danwa no Bunpoo (Grammar of Discourse)*, Taishukan Publ. Co.,Tokyo, 1978.

Susumu Kuno. Identification of Zero-Pronominal Reference in Japanese. *In ATR Symposium on Basic Research for Telephone Interpretation*, 1989.

Barbara J.Grosz, Aravind K.Joshi, and Scott Weinstein.. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics*, 1983.

Megumi Kameyama "A property-sharing constraint in centering." In Proceedings.of the 24th Annual Meeting of the Association for Computational Linguistics, 1986.

Marilyn Walker, Masayo Iida, and Sharon Cote. Centering in Japanese Discourse." In *COLING'90*, 1990.

Kei Yoshimoto. "Identifying zero pronouns in japanese dialogue." In *COLING'88*, 1988.

Satoru Ikehara, Satoshi Shirai, Akio Yokoo, and Hiromi Nakaiwa. Toward an MT System without Pre-Editing - Effects of New Methods in ALT-J/E." In *Proceedings of MT Summit-III*, 1990.

Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai, and Akio Yokoo. "An approach to machine translation method based on constructive process theory. In *Review of ECL*, Vol.37, No.1, 1989

Hiromi Nakaiwa. Case element completion in Japanese texts. In *Proceedings of the 3rd Annual Conference of JSAI*, 1989.