

Analyzing Japanese Double-Subject Construction having an Adjective Predicate

Masahiro Oku

NTT Information and Communication Systems Laboratories
1-2356 Take, Yokosuka-shi, Kanagawa, 238-03 Japan
oku@isl.ntt.jp

Abstract

This paper describes a method for analyzing Japanese double-subject construction having an adjective predicate based on the valency structure. A simple sentence usually has only one subjective case in most languages. However, many Japanese adjectives (and some verbs) can dominate two surface subjective cases within a simple sentence. Such sentence structure is called the double-subject construction. This paper classifies the Japanese double-subject construction into four types and describes problems arising when analyzing these types using ordinary Japanese construction approaches. This paper proposes a method for analyzing a Japanese double-subject construction having an adjective predicate in order to overcome the problems described. By applying this method to Japanese sentence analysis in Japanese-to-English machine translation systems, translation accuracy can be improved because this method can analyze correctly the double-subject construction.

1 Introduction

We have been developing a Japanese-to-English machine translation system called ALT-J/E (Ikehara et al., 1987). In ALT-J/E, Japanese sentence analysis is performed based on the valency structure for the predicate of the input sentence. ALT-J/E can translate several special Japanese constructions such as predicate idioms and function verb expressions (Oku, 1992). The Japanese double-subject construction was not handled well by the original ALT-J/E. This paper describes a method for analyzing a Japanese double-subject construction having an adjective predicate based on the valency structure.

Simple sentences normally have only one subjective case in most languages. However, it is possible for many Japanese adjectives (and some verbs) to dominate two surface subjective cases within a simple sentence. Such sentence structure is called the double-subject construction. JEIDA (the Japan Electronic Industry Development Association) has issued evaluation

standards for machine translation systems (JEIDA, 1995). The main aim of these standards is to establish objective test sets for machine translation system evaluation. Therefore, they include almost of all linguistic phenomena that the systems have to process. In the standards, the double-subject construction is listed as one of the constructions that are difficult to process successfully.

Many Japanese philologists have studied the double-subject construction because it is necessary to reveal the syntactic or semantic structure of this construction if we are to establish a syntactic or semantic theory for the Japanese language (Ishigami, 1977) (Kuno, 1973) (Mikami, 1960). However, there are a few studies for processing this construction from the point of view of computational linguistics or engineering (Murata, 1990). Therefore, the aim of this paper is to discuss a method for analyzing the Japanese double-subject construction having an adjective predicate from the point of view of engineering.

2 Definition of key terms

Key terms used in this paper are defined as follows:

• case-marking particle

In English, each case is marked by its relative position in terms of the predicate or by a preposition. In contrast to this, Japanese language marks each case by a certain sort of postpositional particle located next to noun phrases such as "ga" or "wo". Such particles are called "case-marking particles".

• adverbial particle

"Adverbial particles" give a case an additional function (topicalization etc.) by their attachment to the case-marking particle. Some adverbial particles such as "wa" or "mo" often stand-in for case-marking particles and give the case an additional function.

• sentence having an adjective predicate

In the Japanese language, adjectives function as predicates in sentences as do verbs. Therefore, in this paper, the sentence in which an adjective acts as a predicate is called a "sentence having an adjective predicate".

• double-subject construction

Usually, a simple sentence has only one subjective case. However, many Japanese adjectives (and some

verbs) can dominate two surface subjective cases within a simple sentence. Such sentence structure is called the "double-subject construction". This construction includes both adverbial particle "wa" and subjective case-marking particle "ga". Therefore, the Japanese double-subject construction is also referred to as the "wa"- "ga" construction.

• **valency structure**

The sentence structure can be considered as a combination of a predicate and its modifiers. Such sentence structure is called the "valency structure". The valency structure represents what surface cases the modifiers dominated by a given predicate correspond to.

• **valency pattern**

"Valency patterns" are structure patterns that formulate possible valency structures for predicates. A valency pattern is defined for each usage for each predicate. Figure 1 shows an example of a valency pattern for a verb "shoukai suru (introduce)". For verbs and adjectives, valency patterns formulated as shown in figure 1 are collected beforehand in the dictionary called the valency pattern dictionary.

Japanese valency pattern for "紹介する"

N1 (subjective)	SR: [(agent)] JR: が、から
N2 (objective1)	SR: [(agent)] JR: を
N3 (objective2)	SR: [(agent)] JR: に

Remarks:
 Ni: Label for a valency element
 SR: Semantic restriction on a noun
 JR: Restriction on case-marking particles

Figure 1. An example of a valency pattern.

• **valency element**

In the valency structure, each relation between a predicate and its modifier is called a "valency element". As shown in figure 1, valency elements are described using both the semantic restriction on the noun including the modifier and the restriction on the case-marking particles including the modifier.

• **sentence analysis (or analyzing a sentence)**

"Sentence analysis" is the process that reveals the valency structure of the input sentence. ALT-J/E performs sentence analysis by binding the modifiers for the predicate in the input sentence to the valency elements in the valency pattern for the predicate in the valency pattern dictionary.

• **semantic category**

A "semantic category" is a class for dividing nouns into concepts according to their meaning. For example, both "man" and "woman" belong to the same semantic category [(human)]. In the below, the words surrounded by [()] are semantic categories.

3 Japanese Sentence Analysis based on the valency structure

Figure 2 outlines the translation process in ALT-J/E (Hayashi, 1987). Morphological analysis segments the input Japanese sentence into its component words such as predicates and nouns. Dependency analysis determines the dependency structure to indicate the association between words. From the association between a predicate and its modifiers in the dependency structure, sentence analysis tries to determine the valency structure, i.e., it determines, for the valency pattern for the predicate, which valency element each modifier corresponds to. This valency structure is then converted into the equivalent English sentence structure. Finally, the output English sentence is generated from the structure.

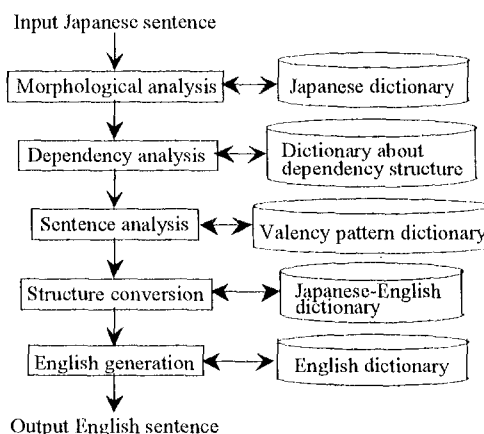


Figure 2. Outline of translation process in ALT-J/E.

The following describes how ordinary sentences are analyzed by ALT-J/E using the valency structure for the predicate of the input sentence.

Sentence analysis is the process that converts the dependency structure into the valency structure by referring to valency patterns. All valency patterns for each usage with each predicate are prepared beforehand and held in the valency pattern dictionary. As shown in figure 1, for each predicate, both the semantic restriction on a noun including its modifier and the restrictions on case-marking particles including its modifier are described for each valency element. When the modifier in the input sentence satisfies both restrictions for a certain valency element in the valency pattern for the predicate, the modifier is bound to that valency element. The valency structure is determined by binding all predicate modifiers to valency elements. An adverbial particle can correspond to more than one case-marking particle, for example, "wa" is a possible proxy for "ga", "wo", "ni" and so on. Therefore, sentence analysis first tries to bind modifiers that have case-marking particles, each of which represents which case is unambiguously marked by particle spelling like "ga"

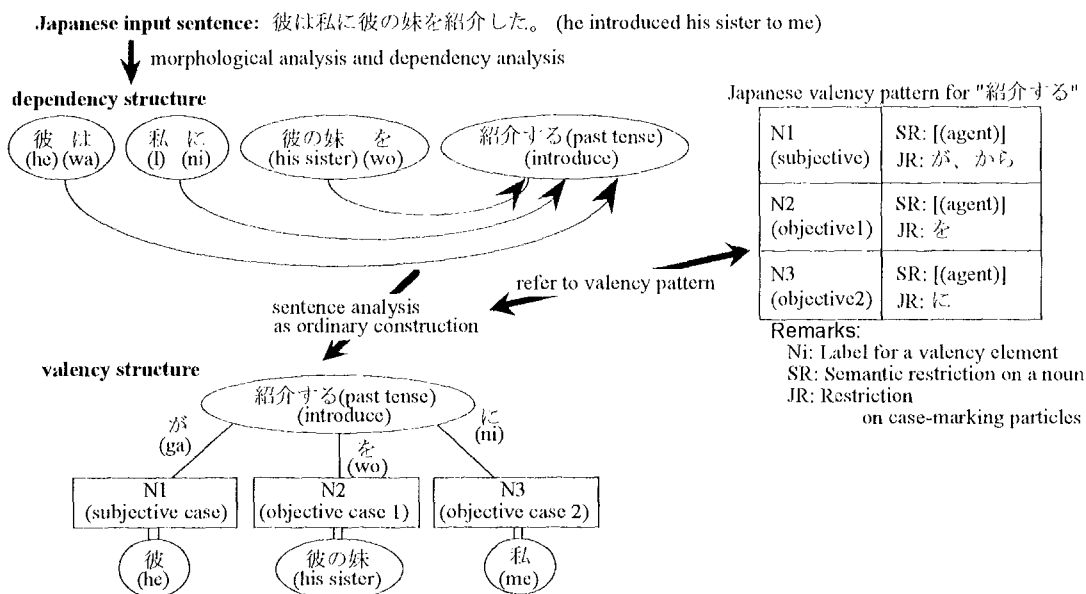


Figure 3. An example of the Japanese sentence analysis.

marks subjective case, to valency elements in the valency pattern for the predicate. The analysis then tries to bind modifiers that have an adverbial particle to the non-bound valency elements.

Figure 3 shows an example of this type of Japanese sentence analysis for the sentence "kare wa watashi ni kare no imouto wo shoukai shita (he introduced his sister to me)". First, the analysis tries to bind modifiers with case-marking particles "wo" and "ni" to the valency elements in the valency pattern for the predicate "shoukai suru (introduce)" which is obtained from the valency pattern dictionary. As modifiers satisfy both restrictions on the noun meaning and case-marking particles, they are bound to N2 (objective case 1) and N3 (objective case 2), respectively. The analysis then tries to bind the modifier with adverbial particle "wa" to a non-bound valency element. So far, N2 and N3 in the valency pattern have already been bound. Therefore, as shown in figure 3, the only non-bound valency element is N1 which is a subjective case. The adverbial particle "wa" can stand-in for case-marking particle "ga", which is the non-bound valency element, and the noun "kare (he)" satisfies the semantic restriction on the subjective case N1 i.e. [(agent)]. Therefore, the modifier with "wa" is bound to the subjective case N1. Finally, the valency structure as shown in figure 3 is obtained.

4 Processing the Japanese double-subject construction

Many Japanese adjective predicates dominate two subjective cases and so form the double-subject construction. The double-subject construction having an adjective predicate actually has several variants, so

no one approach can be used to analyze it. Accordingly, this section classifies the four types and the characteristics of each type are described.

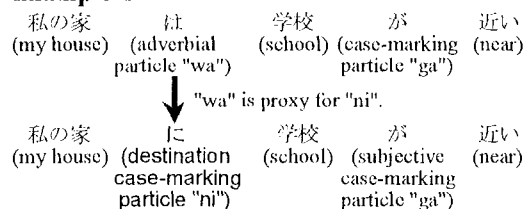
4.1 Classification of the Japanese double-subject construction

The Japanese double-subject construction (also called the "wa"- "ga" construction) can be classified into the following four types based on a previous study (Ishigami, 1977):

• type-1

In this variant, adverbial particle "wa" is a proxy for a case-marking particle such as "ni" other than subjective case-marking "ga". Example-1 shows "wa" as a proxy for case-marking particle "ni" in the sentence "watashi no ie wa gakkou ga chikai (the school is near my house)".

Example-1



• type-2

In this variant, adverbial particle "wa" is a proxy for case-marking particle "no" representing a noun modifier (pre-nominal). Example-2 shows "wa" as a proxy for pre-nominal case-marking particle "no" in the sentence "zou wa hana ga nagai (elephants have long trunks)". In this variant, the case of "wa" modifying a predicate must be analyzed as "no" modifying the

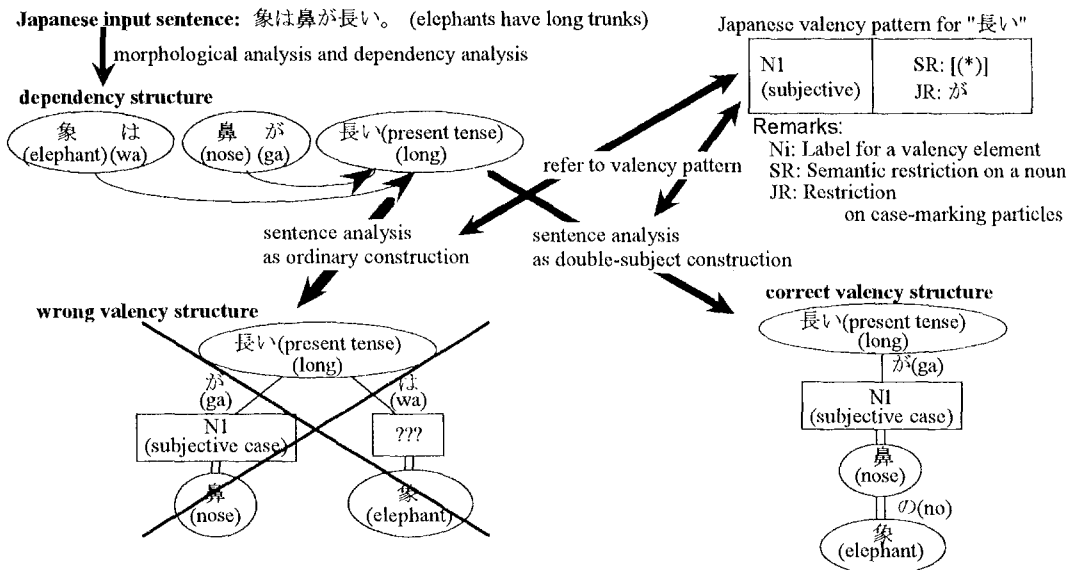
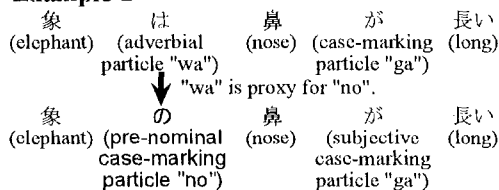


Figure 4. An example of analyzing double-subject construction (type-2).

case with subjective case-marking particle "ga". Therefore, this analysis involves re-formation of the valency structure.

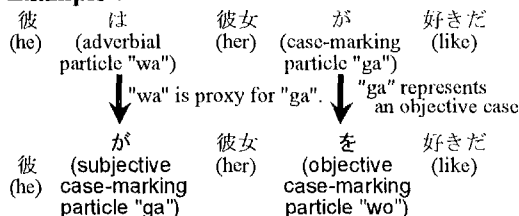
Example-2



• type-3

In this variant, the case with case-marking particle "ga" sometimes represents an objective case. Although an objective case is usually marked by case-marking particle "wo", some adjective predicates have an objective case marked by case-marking particle "ga". Example-3 shows that "kanojo (she)" with "ga" is an objective case and "kare (he)" with "wa" is a subjective case in the sentence "kare wa kanojo ga sukida (he likes her)". As case-marking particle "ga" normally indicates the subjective case, binding "ga" to the subjective case leads to incorrect analysis if only surface spelling is considered.

Example-3

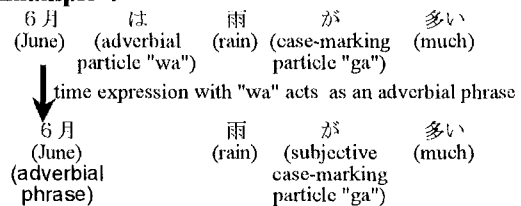


• type-4

In this variant, the case with adverbial particle "wa" acts as an adverbial phrase representing time and is actually a special form of type-1. Representing time is

optional for most predicates. Moreover, Japanese time expressions are often translated into English adverbial phrases. Therefore, type-4 is separated from type-1 in this classification from viewpoint of engineering. Example-4 shows that time expression "6-gatsu wa (in June)" acts as an adverbial phrase in the sentence "6-gatsu wa ame ga ooi (it has much rain in June)".

Example-4



4.2 Problems in processing the Japanese double-subject construction

Type-1 and type-4 cases can be analyzed using the processing flow described in section 3 because adverbial particle "wa" simply acts as a case-marking particle. However, the following problems arise when processing type-2 and type-3 cases in the normal way.

• Problem with type-2 cases

Figure 4 shows the analysis of the type-2 example, "zou wa hana ga nagai (elephants have long trunks)". The predicate "nagai (long)" has only one valency element N1 with "ga". According to ordinary sentence analysis, the modifier "hana ga" is bound to the valency element N1, which means that the other modifier "zou wa" is left unbound. That is, sentence analysis cannot be completed as shown in the left bottom of figure 4. This complicates the accurate translation of this modifier into English.

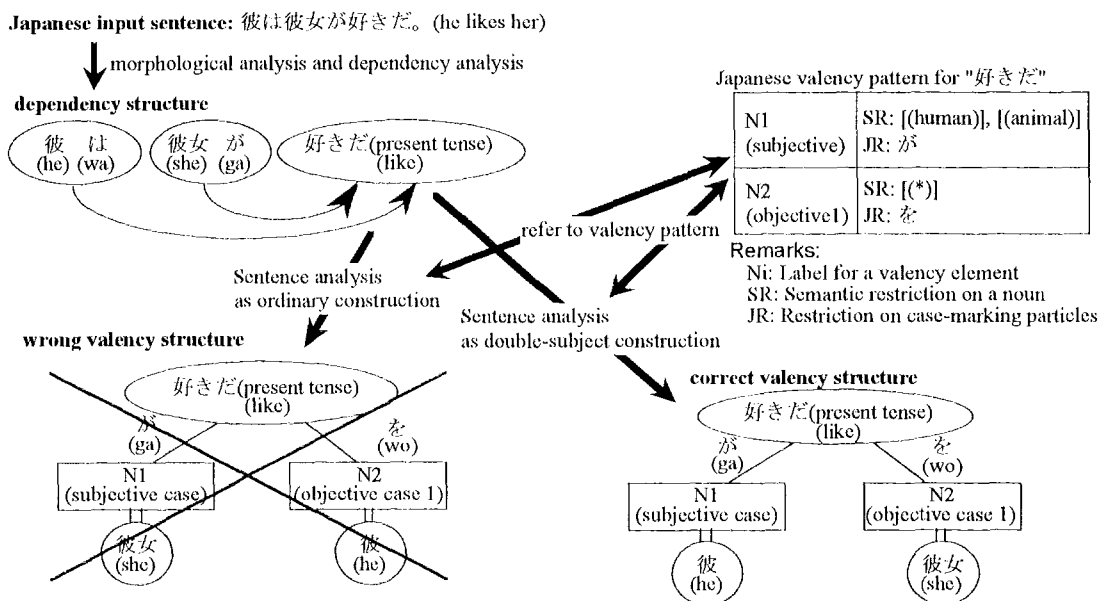


Figure 5. An example of analyzing double-subject construction (type-3).

• Problem with type-3 cases

Figure 5 shows the analysis of sentence analysis of the type-3 example, "kare wa kanojo ga sukida (he likes her)". If modifiers with a case-marking particle are preferentially processed as described in section 3, then the modifier with "ga" binds to subjective case N1 which is wrong in this example. This leads us to the wrong interpretation as shown in the left bottom of figure 5.

5 Proposed method for analyzing Japanese double-subject construction

In order to overcome the problems described in the previous section, this section proposes a method for analyzing a Japanese double-subject construction having an adjective predicate. The method has three processing phases. The first determines whether the input sentence has double-subject construction or not. The second determines which of the four variants the sentence is. The last processes the sentence according to its type.

Figure 6 shows the processing flow. The input sentence has already undergone morphological analysis and dependency analysis, i.e., it has been already determined which nouns modify the adjective predicate with what sort of postpositional particle expressions. Each processing phase is described below in detail with reference to figure 6.

• Judgment of double-subject construction

If the input sentence contains two modifiers that have adverbial particle "wa" and case-marking particle "ga", it is determined as a double-subject construction. Sentences other than double-subject construction, as

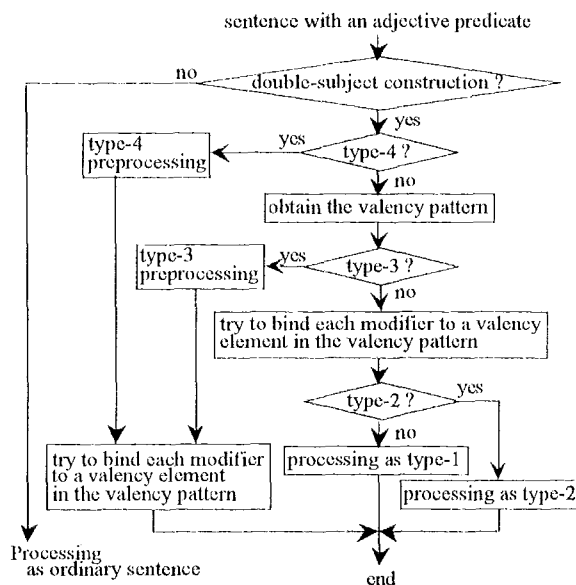


Figure 6. Processing flow for analyzing a Japanese double-subject construction.

well as sentences with verb predicates, are processed normally as described in section 3.

• Type determination

First, type-4 is set if the modifier with adverbial particle "wa" represents a time expression or not. This is efficient because type-4 is the most specific type. This judgment is performed by checking whether the semantic category of the noun included in the modifier with "wa" is associated with [(time)] or not. Second, type-3 is set by the valency pattern for the predicate in the input. The predicate in a type-3 case has to cover

both subjective and objective cases in its usage. The valency pattern for each usage for each predicate is defined in the valency pattern dictionary. Therefore, this judgment is performed by checking whether the valency pattern for the predicate includes both a subjective case and an objective case. Finally, type-1 and type-2 are differentiated according to the result of binding between the modifiers in the input sentence and the valency elements in the valency pattern. Type-2 is set if the modifier with case-marking particle "ga" is bound to the subjective case and the modifier with adverbial particle "wa" is not. The remaining cases are judged as type-1. Here, modifiers with "wa" and "ga" are often bound in type-1. Notice that the binding process resolves what case-marking particle the adverbial particle "wa" stands-in for.

• **Determining the valency structure**

In the type-4 preprocessing step in figure 6, the modifier with "wa" (time reference) is considered as an adverbial phrase. Next, binding the other modifiers with the valency elements in the valency pattern for the predicate is attempted. This facilitates the translation of Japanese time expressions into English.

In the type-3 preprocessing step in figure 6, case-marking particle "ga" has to be converted into the case-marking particle "wo" before binding the former which represents an objective case. This conversion allows us to correctly bind each modifier to its appropriate valency element, because predicates in type-3 cases have both a subjective case with "ga" and an objective case with "wo". See the example in the right bottom of figure 5.

In type-2 processing, adverbial particle "wa" is a proxy for pre-nominal case-marking particle "no" and the modifier with adverbial particle "wa" must be analyzed as the phrase which modifies the subjective case with "ga". As a result, type-2 analysis involves changing the valency structure. In figure 4, the noun phrase "zou no hana" is formed from both "zou (elephant)" with "wa" and "hana (nose)" with "ga" by converting "wa" into "no", and "zou no hana" is bound to the subjective case for the predicate "nagai (long)". See the example in the right bottom of figure 4. Moreover, type-2 processing tries to determine the semantic relation between the noun with "ga" and the noun with "no" (originally "wa"). Determining the semantic relation helps us translate the Japanese double-subject construction into the appropriate English construction and expression. Although several semantic relations are known, at the present time ALT-J/E can resolve only two of them based on semantic categories: has-a relation and is-a relation.

As adverbial particle "wa" in type-1 cases is a proxy for a case-marking particle such as "ni", "de" and so on, type-1 cases can be processed in the way

described in section 3.

The above approach correctly determines the input sentence's valency structure which allows the machine translation system to produce more accurate output. As a result, the method proposed here improves the translation accuracy of ALT-J/E.

6 Conclusion

This paper has proposed a method for analyzing the Japanese double-subject construction that includes an adjective predicate. This paper has classified the construction into four types and described problems when applying the ordinary sentence analysis to the four types. The algorithm that overcomes these problems has been explained in detail.

This method has been applied to Japanese sentence analysis in ALT-J/E. Because this method can correctly analyze the double-subject construction, the method helps the translation of this construction into an appropriate English construction.

References:

- Yoshihiko Hayashi, "On the Japanese sentences analysis based on the valency structure", *Working group report on natural language processing, IPS Japan*, 62-6 (1987) (in Japanese).
- Satoru Ikehara, Masahiro Miyazaki, Satoshi Shirai and Yoshihiko Hayashi, "Speaker's recognition and multi-level-translating method based on it", *Trans. IPS Japan*, Vol.28, No.12, pp.1269-1279 (1987) (in Japanese).
- Teruo Ishigami, "Structure of Japanese double-subject construction with adjective predicates (nijuu-shukaku keiyoushi bun no kouzou)", *Essay on Japanese philology (nihongo-gaku shiron)*, Vol.3, pp.1-37 (1977) (in Japanese).
- JEIDA Evaluation standards on machine translation systems (JEIDA kikai hon'yaku system hyouka kijun), *JEIDA report* (1995) (in Japanese).
- Susumu Kuno, *The structure of the Japanese Language*, The MIT Press, 1973.
- Akira Mikami, *Elephants have long trunks (zou wa hana ga nagai)*, Kuroshio-shuppan, 1960 (in Japanese).
- Kenichi Murata, "On the formal description of the meanings of Japanese adjectives", *Proc. on the 40th annual meeting of IPS Japan*, 5F-3 (1990) (in Japanese).
- Masahiro Oku, "Methods for analyzing Japanese predicate idioms and function verb expressions", *Journal of Information Processing*, Vol.15, No.3, pp.425-433 (1992).