A System of Verbal Semantic Attributes Focused on the Syntactic Correspondence between Japanese and English

Hiromi NAKAIWA, Akio YOKOO and Satoru IKEHARA

NTT Communication Science Laboratories
1-2356 Take, Yokosuka-shi, Kanagawa-ken, 238-03 Japan
E-mail: nakaiwa {yokoo, ikehara} @nttkb.ntt.jp

Abstract

This paper proposes a system of 97 verbal semantic attributes for Japanese verbs which considers both dynamic characteristics and the relationship of verbs to cases. These attribute values are used to disambiguate the meanings of all Japanese and English pattern pairs in a Japanese to English transfer pattern dictionary consisting of 15,000 pairs of Japanese valence patterns and equivalent English syntactic structures.

1. Introduction

Various machine translation systems have approached the stage of being put to practical use. However, the quality of the finished translation is not satisfactory in any of these systems. This is due to difficulties in limiting linguistic phenomena that are handled by machine translation systems. In particular, the analysis of linguistic expressions such as ellipsis and anaphoric references, which require contextual analysis, is imperfect. To introduce constraints brought about by context requires an enormous volume of knowledge of word meanings that can be used to determine the semantic relationships between one sentence and another.

To avoid an explosion in the volume of knowledge, a technique is proposed for classifying word meanings and determining the relationships between words or between sentences using the typical attribute values of each word. Particularly in the case of context processing, the verbal semantic attributes that become the key factors in analyzing the flow of sentences constitute important knowledge.

Various efforts have been made in researching verb classification. Muraki (1985) suggested a method for grouping Japanese verbs using their word meanings and their syntactic

features. Tomiura et al. (1986) proposed a method for representing the meaning of verbs divided into fundamental meanings and reasoning rules. Ogino et al. (1989; EDR 1990) proposed a method for verb classification based on relations between verbs and co-occurring elements. Various efforts have been made to classify English verbs. For example, Levin (1993) proposed a method for the classification of 3000 English verbs that uses the relationship between syntactic behavior and shared meaning.

The research about verb classification still tends to be limited solely to classification of the semantics of verbs per se. It does not take into account the relationship between word meanings and their usage within sentences and is not aimed at natural language processing. Thus, the full benefits that could be achieved in the analysis of tracking semantic relationships between sentences and eliminating the polysemy of verbs have not been realized.

This paper focusses on the relationship between word meanings of verbs and their usage, and seeks to classify the semantic attributes of verbs. These semantic attributes are used in defining the method of use of each verb in Japanese to English transfer pattern dictionaries. They furnish the key to tracing the semantic relationships of verbs that are used in a text.

2. Semantic Structure of Verbal Patterns

This chapter examines the relationship between the usage of verbs[1] and the semantic structure of verbs. In machine translation systems, it is well known that the translation pattern pairs of source

---

[1] ALT-J/E's pattern dictionaries include both verbs and adjectives. Japanese adjectives are the equivalent of English 'be Adjective': for example "A-*ga utsukushii*" => "A is beautiful". 'verbs' will be used to refer to both verbs and adjectives from now on.

language and target language sentence are effective in determining the meaning of verbs.

Our machine translation system, ALT-J/E, uses two types of Japanese to English transfer pattern dictionaries based on verbs: the semantic valence pattern transfer dictionary and the idiomatic expression transfer dictionary (Fig. 1). These dictionaries consist of pairs made of

[The semantic valence pattern transfer dictionary]
(1) N1(SUBJECTS)-*ga* N2(FOOD)-*wo* *taberu*
                                    eat
    => **N1 eat N2**

(2) N1(*)-*ga* *yomigaeru*
                revive
    => **N1 revive**

[The idiomatic expression transfer dictionary]
(1) N1(SUBJECTS)-*ha* *se-ga* *takai*
                      back   high
    => **N1 is tall.**

Fig. 1 Japanese to English Transfer Pattern
Dictionaries
(The semantic constraints are shown in parenthesis,
* indicates there is no semantic constraint.)

Japanese unit sentence patterns derived from Japanese verbs[2] with semantic constraints to their case elements and English patterns which correspond to the Japanese expressions. For example, pattern(1) in Fig. 1 shows how, if the Japanese verb is "*taberu*" and the noun phrase with a "*ga*" particle, which shows a subject, has the semantic attribute SUBJECTS and the noun phrase with a "*wo*" particle, which shows a direct object, has the semantic attribute FOOD then the verb should be translated as "eat". The noun phrase with the "*ga*" particle is translated as the English subject. The noun phrase with the "*wo*" particle is translated as the English direct object. Here, we examine the relationship between the usage of verbs and the semantic structure of verbs using verbal patterns that have been entered into the Japanese to English transfer pattern dictionaries of ALT-J/E.

Fig. 2 shows an example of entries in the Japanese to English transfer pattern dictionary which indicate the patterns of the

Japanese verb "*tsutsumu*". This verb has three patterns.

[Japanese Verb : *tsutsumu* "wrap"]
(1) N1(SUBJECTS)-*ga*
    N2(CONCRETE OBJECTS or PEOPLE)-*wo*
    N3(CLOTHES or PAPERS)-*de*
    tsutsumu
    => **N1 wrap N2 in/with N3**
    Verbal Semantic Attribute: N1's bodily action

(2) N1(FIRE, ATMOSPHERE or AIR)-*ga*
    N2(CONCRETE OBJECTS, CULTURE or PLACES)-*wo*
    tsutsumu
    => **N1 envelop N2**
    Verbal Semantic Attribute: N1 changes N2's attributes

(3) N1(FOG)-*ga*
    N2(CONCRETE OBJECTS or PLACES)-*wo*
    tsutsumu
    => **N1 veil N2**
    Verbal Semantic Attribute: Natural Phenomena

Fig. 2 Example of a Japanese Verb with
multiple patterns

The first example shows a pattern pair indicating that the equivalent of the Japanese expression "N1 (SUBJECTS) *ga* N2 (CONCRETE OBJECTS or PEOPLE) *wo* N3(CLOTHES or PAPERS) *de tsutsumu*" is the English expression "N1 wrap N2 in/with N3". When the Japanese verb "*tsutsumu*" was used with these cases, this sentence gives the impression that N1 really does the wrapping action. So, in this case, this pattern has the verb meaning "N1 conducts bodily action.".

The second example shows a pattern pair indicating that the equivalent expression of the Japanese expression "N1(FIRE, ATMOSPHERE or AIR) *ga* N2(CONCRETE OBJECTS, CULTURE or PLACES) *wo tsutsumu*" is the English expression "N1 envelop N2". This sentence gives the impression that the state of N2 which isn't usually enveloped by N1 changes to the enveloped state. So, even though the same Japanese verb "*tsutsumu*" was used with these cases, in this case, the pattern has a verb meaning of "N1 changes N2's attributes.".

The third example shows a pattern pair indicating that the equivalent of the Japanese expression "N1(FOG) *ga* N2(CONCRETE OBJECTS or PLACES) *wo tsutsumu*" is the English expression "N1 veil N2". In this case, this sentence gives the impression that a natural

---

[2]In the idiomatic expression transfer dictionaries, these are the core sector of idiomatic expressions such as "*Abura wo uru*" literally, "to sell oil", but idiomatically, "to idle away time".

phenomenon, 'fog', has occurred. So, this pattern has the meaning "Natural Phenomena have arisen".

As shown in these examples, maintaining expressions in pairs which indicate both the common meaning and their usage between the Japanese and English, enables us to eliminate many conceptual ambiguities and makes it possible to give detailed and accurate attribute values to the Japanese verb "*tsutsumu*".

As in the case of the Japanese verb "*tsutsumu*", one verb normally has several kinds of conceptual structures. But one verbal pattern which indicates common word meanings and their use between the Japanese and English (which differ so vastly in syntactic structure) corresponds to one conceptual structure. So, it is possible to eliminate the conceptual ambiguity of verbs by selecting verbal patterns in syntactic semantic analysis. In Japanese to English machine translation, we estimate there are tens of thousands of verbal patterns which need to be defined. If the usage of these patterns can be expressed by a small number of verbal semantic attributes, it is possible to track the semantic relationships of verbs easily. When giving verbal semantic attributes to a pair of individual Japanese and English patterns, it is possible to refer to the meaning of verbs not only in Japanese but also in English.

3. System of Verbal Semantic Attributes
3.1 Classification Standards for Verbal Semantic Attributes

Regarding the classification of verbs for use in machine translation, Nishida et al. (1980) proposed a system of verbal classification. This system of classification was introduced to resolve syntactic and semantic ambiguities of English in English to Japanese machine translation. To this system, they added the semantic attributes of verbs to the patterns of English verbs proposed by Hornby (1975) and determined the case structures depending on the combination of these two kinds of information. This system of verbal semantic attributes was introduced on the condition that the features of syntactic structures are expressed by Hornby's patterns of English verbs. So, this system of

classification focused only on word meaning. Therefore this system can not be applied as such to the classification of Japanese verbs because Hornby's patterns can't be applied directly to Japanese verbs. No one has yet to propose exhaustive patterns like Hornby's for Japanese verbs.

We expanded our system based on the discussions in section 2, using the following two factors.

•**Dynamic Characteristics of verbs**
    Classification based on a verb's meaning
    and its effects on the discourse:
This classification is based on the types of action that can be understood to have occurred when a verb is expressed and what situations have been brought about.

    Ex. "*motsu*"(to have)   -- Possession
        "*kaihatsusuru*"(to develop) -- Production

The verb "*motsu*" indicates that there is an act of possession within the context. In contrast, the verb "*kaihatsusuru*" indicates that there is something being produced within the context.

•**Relationship of Verbs to Cases**
    Classification based on the role which the
    cases play with the verbs that govern them:
This classification is based on the roles played by the case elements governed by the verb expressed.

    Ex."*kanseisuru*":SUBJ be completed
            ->SUBJ be produced
    "*kaihatsusuru*":SUBJ develop OBJ
            ->SUBJ produce OBJ

"*kanseisuru*" and "*kaihatsusuru*" are both verbs which indicate acts of production. But whereas "*kanseisuru*" indicates that the SUBJ is being produced, "*kaihatsusuru*" indicates that the SUBJ produces the OBJ.

3.2 Semantic Attribute System considering the Semantic Relationship between Verbs

We created a system of verbal semantic attributes as explained above. The semantic attribute values were determined using the usage patterns of typical Japanese verbs. First we classified verbs focussing on their dynamic characteristics. Next, we classified each group again focussing on the relationships of verbs to

their cases. The top levels of the created system of verbal semantic attributes are shown in Fig. 3. The left side of this figure lists classifications as based on the dynamic characteristics of the verbs (their meanings). The right side lists the classifications based on the relationship of verbs with their cases (their usage). On the basis of these classification criteria, 97 verbal semantic attributes have been established.

4. Result of Application for the Semantic Descriptions of Verbal Patterns

We evaluated the coverage of the verbal semantic attributes shown in chapter 3 by examining the verbal semantic attributes for each Japanese to English pair (about 15,000 pairs) in the Japanese to English transfer pattern dictionaries[3].

Fig.4 shows how many transfer patterns were created for each verb in the semantic valence pattern transfer dictionary and the idiomatic expression transfer dictionary. This figure shows the results that were counted for each different verb. The percentage of patterns that came from verbs with more than one pattern was 73.4%. In these verbs that have multiple patterns, the percentage that had different kinds of verbal semantic attributes added to the patterns were 70.1%. This result shows that it is possible to classify semantic attributes for each verb by adding verbal semantic attributes to Japanese and English transfer pairings.

Next we counted the number of verbal semantic attribute values given for each pattern. Fig. 5 shows how many verbal semantic attributes
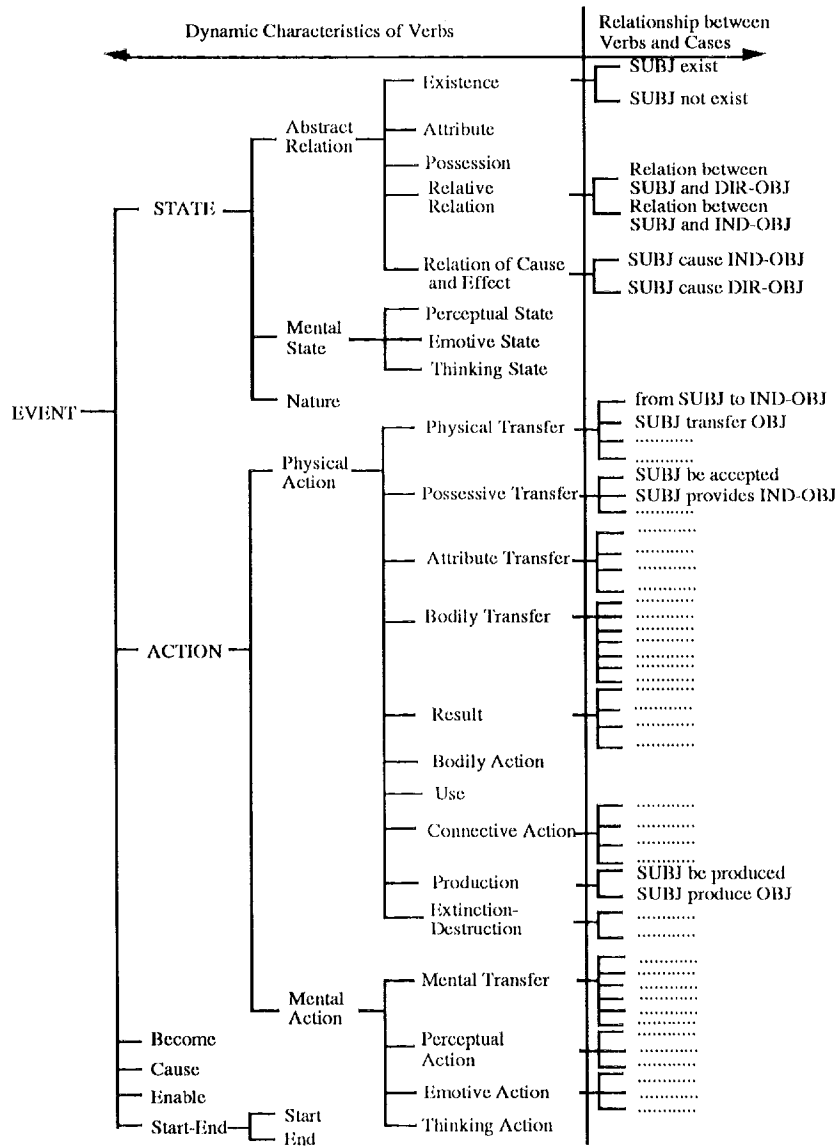
Fig. 3 System of Verbal Semantic Attributes

[3]Attribute values from a general noun attribute system classified into some 2,800 types have been provided as semantic constraints to the case elements of these patterns (Ikehara et al. 1991) enabling accurate selections of syntactic structures.
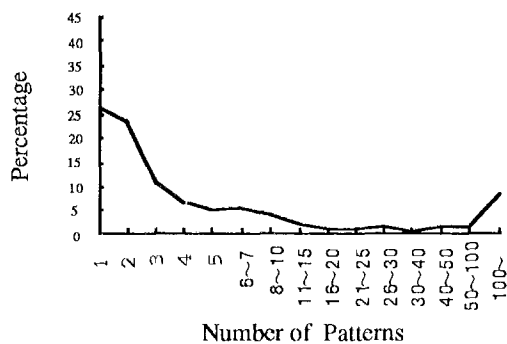
Fig. 4 Ratio of the number of patterns to each verb

were used by how many patterns. About 90% of patterns can be described by just one attribute value. This result shows that by giving the verbal semantic attributes proposed in this paper to each pattern in ALT-J/E, even in instances where multiple meanings may exist for a given Japanese verb, meanings can be selectively limited when verbs are viewed in terms of pattern pairings. The verbal semantic attributes which were given in each pattern have the potential to become an important key to tracking semantic relationships between sentences as is shown in chapter 5.

Fig.6 shows the most frequent ten verbal semantic attributes for all the patterns. In these verbal semantic attributes, the patterns that ATTRIBUTE was added can almost all be described by only one attribute value (26.4% out of 27%). By contrast, the many patterns
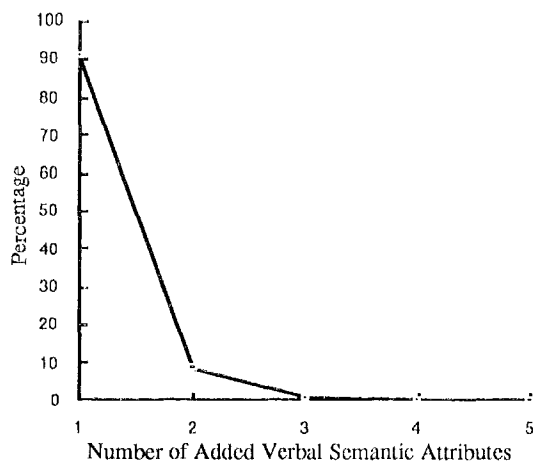

Fig. 5 Ratio of the number of added verbal semantic attributes to each pattern

No.1 :ATTRIBUTE ,   Coverage: 27.0%
    Number of added VSA: 1:26.4%, 2 or more:0.6%
    Ex. N1(SUBJECTS)-ga   chikarazuyoi
        =>   N1 be   reliable

No.2 :BODILY ACTION ,   Coverage: 12.7%
    Number of added VSA: 1:9.9%, 2 or more:2.8%
    Ex. N1(HUMAN)-ga   odoru
        =>   N1 dance

No.3 :ATTRIBUTE TRANSFER(Subj's attribute) ,
                                Coverage: 9.4%
    Number of added VSA: 1:8.1%, 2 or more:1.3%
    Ex. N1(CONCRETE OBJECTS)-ga
        N14(FIRE, HEAT or LIGHT)-de tokeru
        =>   N1 be melted   by   N14

No.4 :THINKING ACTION ,   Coverage: 8.9%
    Number of added VSA: 1:7.5%, 2 or more:1.4%
    Ex. N1(HUMAN)-ga N2(CULTURE)-wo
    fukushuusuru
        =>   N1 review N2

No.5 :ATTRIBUTE TRANSFER
    (Subj changes Dir-Obj's attribute) ,Coverage: 7.9%
    Number of added VSA: 1:5.8%, 2 or more:2.1%
    Ex. N1(SUBJECTS)-ga
        N2(PRODUCTS or CULTURE) wo
        moyasu
        =>   N1 burn N2

No.6 :EMOTIVE ACTION(Subj acts) , Coverage: 7.7%
    Number of added VSA: 1:6.6%, 2 or more:1.1%
    Ex. N1(SUBJECTS)-ga N2(DEATH)-wo
        kanashimu
        =>   N1 mourn N2

No.7 :MENTAL TRANSFER
    (Subj transfers Dir-Obj to Ind-Obj), Coverage: 4.9%
    Number of added VSA: 1:4.5%, 2 or more:0.4%
    Ex. N1(SUBJECTS)-ga
        N2(LITERATURE)-wo
        N3(PUBLICATION or BOOK)-ni
        kankousuru
        => N1 publish N2   in N3

No.8 :EMOTIVE STATE ,   Coverage: 2.1%
    Number of added VSA: 1:1.8%, 2 or more:0.3%
    Ex. N1(SUBJECTS)-ha N2(ABSTRACT)-ga
        kuyashii
        =>   N1 regret N2

No.9 :RELATIVE RELATION
    (between Subj and Ind-Obj) ,   Coverage: 1.8%
    Number of added VSA: 1:1.3%, 2 or more:0.5%
    Ex. N1(HUMAN)-ga N2(CULTURE)-wo
        N3(HUMAN)-ni   shijisuru
        => N1 study N2   under N3

No.10:POSSESSIVE TRANSFER
    (Subj provides Ind-Obj) ,   Coverage: 1.6%
    Number of added VSA: 1:1.4%, 2 or more:0.2%
    Ex. N1(SUBJECTS)-ga N3(SUBJECTS)-ni
        zouwaisuru
        =>   N1 bribe N3

Fig. 6 Coverage of the top 10 verbal semantic attributes

676

described by BODILY ACTION or ATTRIBUTE TRANSFER was added can't be described by one attribute value (2.8% out of 12.7% and 2.1% out of 7.9 %, respectively). These 2 kinds of attribute values indicate the SUBJECT's Physical Action, and it tends to be difficult to resolve the semantic ambiguities for these patterns.

As shown in Fig.6, a few verbal semantic attributes cover a large proportion of patterns. For example, the sum of the coverage of the most frequent attribute value, ATTRIBUTE, and the second most frequent attribute value, BODILY ACTION, cover 39.7 % of all patterns. For these attributes, even if there are several patterns for a given verb, sometimes the same attribute value was given to all the patterns. So the system of verbal semantic attributes is not sufficient to resolve the semantic ambiguities. For such attributes, we need more detail. We are planning to subdivide these attribute values in the future.

5. Applications for Context Processing

In this chapter, we show examples of applications in context processing.

5.1 Analysis of Anaphoric Reference of Japanese Zero Pronouns

Using verbal semantic attributes to analyze anaphoric referents of zero pronouns appearing in Japanese texts is one application that has been considered (Nakaiwa et al. 1992). This technique pays attention to verbal semantic attributes and the relationship between the semantic attributes of the verbs which govern zero pronouns and the semantic attributes of the verbs which govern case element candidates which may be anaphorically referred to. The contexts are carefully examined to determine anaphoric reference elements.

This method has been realized in the machine translation system ALT-J/E. The enhanced ALT-J/E was assessed by processing common Japanese newspaper articles. It was found that 95% of the Japanese zero pronouns requiring anaphoral resolution in the 102 sentences from 30 newspaper articles' lead paragraphs had their referents determined correctly using rules tuned for the 102 sentences(window test). In the case of a blind test, the rate of success in anaphora resolution in

which the zero pronoun referent exists within the sentence in another 98 sentences from newspaper articles was about 83% using the rules. To demonstrate the effectiveness of this method, we evaluated the performance of the method proposed by Walker et.al. (1990) using the 98 sentences. Its rate of success in anaphora resolution where the zero pronoun referents existed within the sentence was about 74%. This result shows that our method is more effective than Walker's method, and that the rules used in our method determine universal relationships between verbs. If a few rules appropriate for the 98 sentences are added, the rate increases to 95%. This result shows that the load imposed by rule customization is low.

Even in the case of sentences in machine translation systems for which target areas cannot be constrained, this method allows the construction of rules independent of the translation target areas by means of verbal semantic attribute pairings. Using the verbal semantic attributes, anaphoric reference resolution of zero pronouns can be conducted with a limited volume of knowledge.

5.2 Supplementation of Elements Outside Sentences against Elliptical Case Elements

Verbal semantic attributes can be used with elliptical case elements in Japanese texts to supplement case elements whose referents do not appear within the texts. To analyze such elliptical phenomena, it is possible to use case elements' semantic constraint conditions to estimate supplementary elements. Semantic information used to estimate supplementing elements is a constraint on cases for selecting the transfer pattern. With this method, therefore, the majority of the constraints involve abstract semantic information, frequently posing difficulties in pinpointing elements to be supplemented. For example, if in Fig. 1(2), "N1(*)-ga yomigaeru(revive)", N1 were to be omitted, the case element N1 has no semantic constraint, and supplementary elements to the case can't be determined. In this case, it is effective to complete the case element corresponding to SUBJECT using the verbal semantic attributes of the pattern, "N1's bodily state is transferred". Thus if a method presuming supplementary elements of elliptical

case elements corresponding to the verbal semantic attributes is used, the deduction of more accurate supplementary elements would be possible.

5.3 Application for Other Context Processings

The verbal semantic attributes can be applied to other context processing problems. Estimating the relationship between verbs by pairing of the verbal semantic attributes, analysis of the tenses relationship of events as indicated by certain sentences and events indicated by another, together with sentence abridgment can be considered.

6. Conclusion

This paper has proposed a system of 97 verbal semantic attributes for Japanese verbs which considers dynamic characteristics and the relationship of verbs to cases. These attribute values were used to disambiguate the meanings of all Japanese and English pattern pairs in a Japanese to English transfer pattern dictionary consisting of 15,000 pairs of Japanese valence patterns and equivalent English syntactic structures. As a result of examining the verbal semantic attributes for each pattern of Japanese to English paring, 90% of patterns can be described by only one attribute values. This result shows that the meanings of Japanese verbs determined by the verbal semantic attributes can be effectively limited when verbs are viewed in terms of pattern parings. Further attentions to details and tightening of standards together with extensive application of this system are now being worked on.

References

EDR (1990) *Concept Dictionary*, TR-027.

Hornby, A. S. (1975) *Guide to patterns and usage in English*, 2nd edition, London, Oxford University Press.

Ikehara, S., M. Miyazaki and A. Yokoo (1991) Semantic Analysis Dictionary for Machine Translation. *Information Processing Society of Japan, Natural Language Processing, Vol.84-13* (in Japanese).

Levin, B. (1993) *English Verb Classes and Alternations*, The University of Chicago Press.

Muraki, S. (1985) Jyutsugo-so niyoru dousi no bunrui (Classification of Verbs by Predicates). *Information Processing Society of Japan, Natural Language Processing, Vol.48-5* (in Japanese).

Nakaiwa, H 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.

Nishida, F. and S. Takamatsu (1980) English-Japanese Translation through Case-Structure Conversion. *Proc. of COLING '80*, pp. 447-454.

Ogino, T. et al. (1989) Verb Classification Based on Semantic Relation of Co-occurring Elements. *Information Processing Society of Japan, Natural Language Processing, Vol.71-2* (in Japanese).

Tomiura, Y. and S. Yoshida. (1986) A Research of the Polysemy and Description of Verbs. *Information Processing Society of Japan, Natural Language Processing, Vol.55-2* (in Japanese).

Walker, M., M. Iida and S. Cote (1990) Centering in Japanese Discourse. *Proc. of COLING'90*.