

Analyzing Embedded Noun Phrase Structures Derived from Japanese Double-Nominal-Case Construction

Masahiro Oku*

NTT Information and Communication Systems Laboratories

Embedded noun phrase structures are frequently used in the Japanese language because embedded sentences make modified noun phrases more expressive. In embedded noun phrase structures, a noun phrase modified by an embedded sentence is usually created by extraposing the case, such as the subjective case, of the sentence. In contrast, there is another type of embedded noun phrase structure. In this type, the noun phrase modifying a certain case of the embedded sentence is extraposed. As conventional methods for analyzing embedded noun phrase structures do not take the existence of the Japanese double-nominal-case construction into consideration, the latter type of embedded noun phrase is handled as an exception. This paper describes a method for analyzing the latter type of embedded noun phrase structure based on the valency structure in the Japanese-to-English translation system called ALT-J/E. We regard the latter type as the structure derived from the Japanese double-nominal-case construction in which the adverbial particle "wa" is a proxy for the pre-nominal particle "no". This point of view leads us to analyze both the former and the latter types of embedded noun phrase structures uniformly based on the valency structure of embedded sentences. Moreover, this paper describes an evaluation that clarifies the validity of the proposed method. The evaluation results show that about 86 % of the latter type of the embedded noun phrase structures are correctly analyzed by the method proposed here. This means the proposed method is valid for analyzing embedded noun phrase structures derived from Japanese double-nominal-case construction.

1. INTRODUCTION

This paper describes a method for analyzing embedded noun phrase structures that is derived from the Japanese double-nominal-case construction based on the valency structure used in ALT-J/E. ALT-J/E is a Japanese-to-English translation system (Ikehara et al. 87).

An embedded noun phrase structure is the structure in which a sentence (hereafter, called the embedded sentence) modifies a noun phrase (hereafter, called the modified noun phrase). This structure is equivalent to the English structure in which a noun phrase is modified by a relative clause. Embedded noun phrase structures are frequently used in the Japanese language because an embedded sentence makes the modified noun phrase more expressive. In embedded noun phrase structures, the noun phrase modified by an embedded sentence is usually derived from a certain case of a predicate for the sentence. For this type of embedded noun phrase structure, several analysis methods have been proposed (Tsujii 83; Yokoo & Hayashi 87). These methods analyze the modified noun phrase as the case extraposed from the embedded sentence.

In contrast, there is another type of embedded noun phrase structure in which the noun phrase modifying a certain case, especially the subjective case, of the embedded sentence is extraposed. Because, in this type, the modified noun phrase is not derived from the valency element for the Japanese verb or adjective of the embedded sentence, it is difficult to apply the analysis methods developed to handle the former type. This is the reason why an exceptional process is required to analyze the embedded noun phrase structure in Japanese syntactic analysis. If the modified noun in the latter type can be connected to any valency element in the embedded sentence, then we can avoid processing the latter type as an exception and analyze both types uniformly. For this purpose, we regard the latter type as the structure derived from the Japanese double-nominal-case construction in which the adverbial particle "wa" is a proxy for the pre-nominal case-marking particle "no". This point of view allows us to treat the modified noun in the latter type as a virtual valency element for the embedded sentence.

In the following, chapter 2 defines terms used in this paper, and chapter 3 summarizes the Japanese sentence analysis based on the valency structure in ALT-J/E. Chapter 4 proposes processing the embedded noun phrase structure derived from Japanese double-nominal-case construction, and chapter 5 describes an evaluation that confirms the validity of the method proposed here.

* NTT Information and Communication Systems Laboratories, 1-1 Hikari-no-oka, Yokosuka-shi, Kanagawa, 239 JAPAN, E-mail: oku@isl.ntt.co.jp

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 either 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.

- **valency structure**

Sentence structure can be considered as a combination of a predicate and its modifiers. Such a 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 the verb "shoukai-suru (introduce)". For verbs and adjectives, valency patterns formulated as shown in figure 1 are collected beforehand in the dictionary called a 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: Restrictions 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.

- **semantic category**

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

- **double-nominal-case 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-nominal-case construction. This construction includes both the adverbial particle "wa" and subjective case-marking particle "ga". Therefore, the Japanese double-nominal-case construction is also referred to as the "wa"- "ga" construction.

- **embedded noun phrase structure**

An embedded noun phrase structure is the structure in which a sentence called the embedded sentence modifies a noun phrase called the modified noun phrase, such as "kare-ga watashi-ni shoukai-shita kare-no-imouto (his sister whom he introduced to me)". In this example, the modified noun phrase "kare-no-imouto (his sister)" is modified by the embedded sentence "kare-ga watashi-ni shoukai-shita (he introduced to me)".

- **analyzing embedded noun phrase structure**

Analyzing the embedded noun phrase structure is one process in sentence analysis and is the process that reveals the relation between the embedded sentence and the modified noun phrase. In the previous example, the embedded noun phrase structure is analyzed as that the objective case "kare-no-imouto (his sister)" is extraposed from the original sentence "kare-ga watashi-ni kare-no-imouto-wo shoukai-shita (he introduced his sister to me)".

3. JAPANESE SENTENCE ANALYSIS BASED ON THE VALENCY STRUCTURE

3.1 OUTLINE OF JAPANESE SENTENCE ANALYSIS 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 (Hayashi 87).

Sentence analysis 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 (hereafter, called SR) on a head noun including its modifier and the restrictions on case-marking particles including its modifier (hereafter, called JR) 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 unambiguously marks a case such as "ga" which 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 unbound valency elements.

3.2 OUTLINE OF ANALYZING THE JAPANESE DOUBLE-NOMINAL-CASE CONSTRUCTION IN ALT-J/E

The Japanese double-nominal-case construction has already been studied (Mikami 60; Kuno 73; Ishigami 77; Murata 90). From the point of view of engineering, we have proposed a method for analyzing this construction having an adjective predicate (Oku 96). In that paper, we classified the construction into four types. One of them is involved in the proposed method for analyzing the embedded noun phrase structure. This type is the Japanese double-nominal-case construction in which adverbial particle "wa" is a proxy for the pre-nominal case-marking particle "no". This section describes this variant of the Japanese double-nominal-case construction in detail.

Figure 2 shows the analysis of this variant of the Japanese double-nominal-case construction. The predicate "nagai (long)" in the sentence "zou-wa hana-ga nagai (elephants have long trunks)" has only one valency element

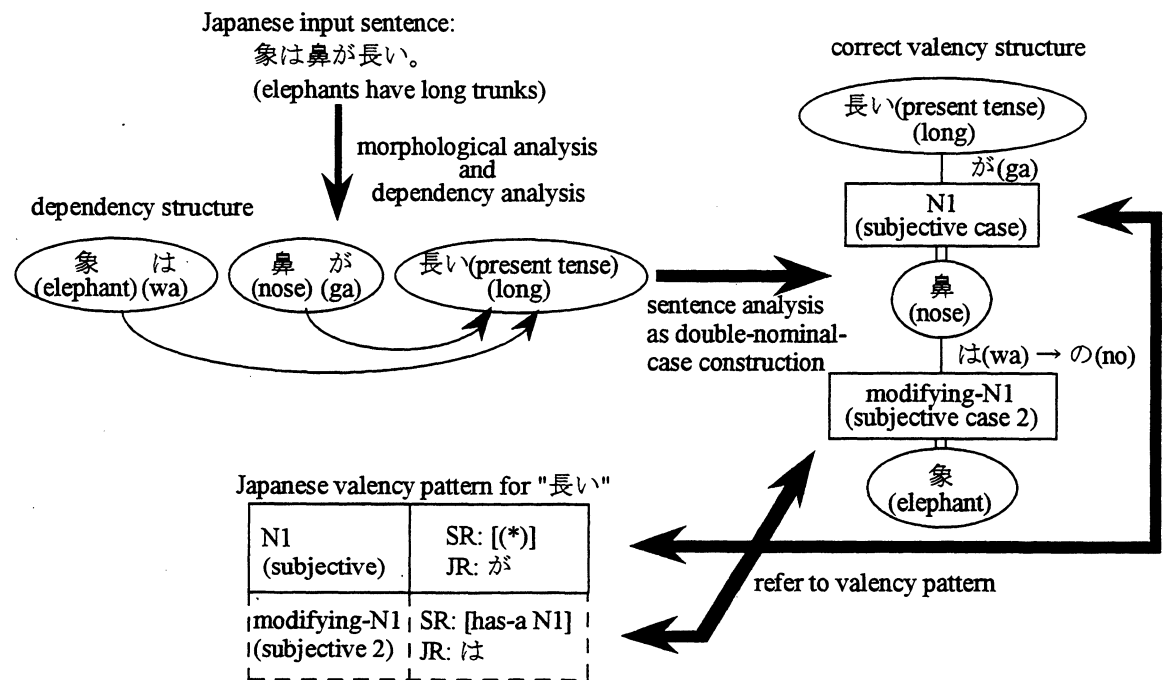


Figure 2: An example of analyzing double-nominal-case construction.

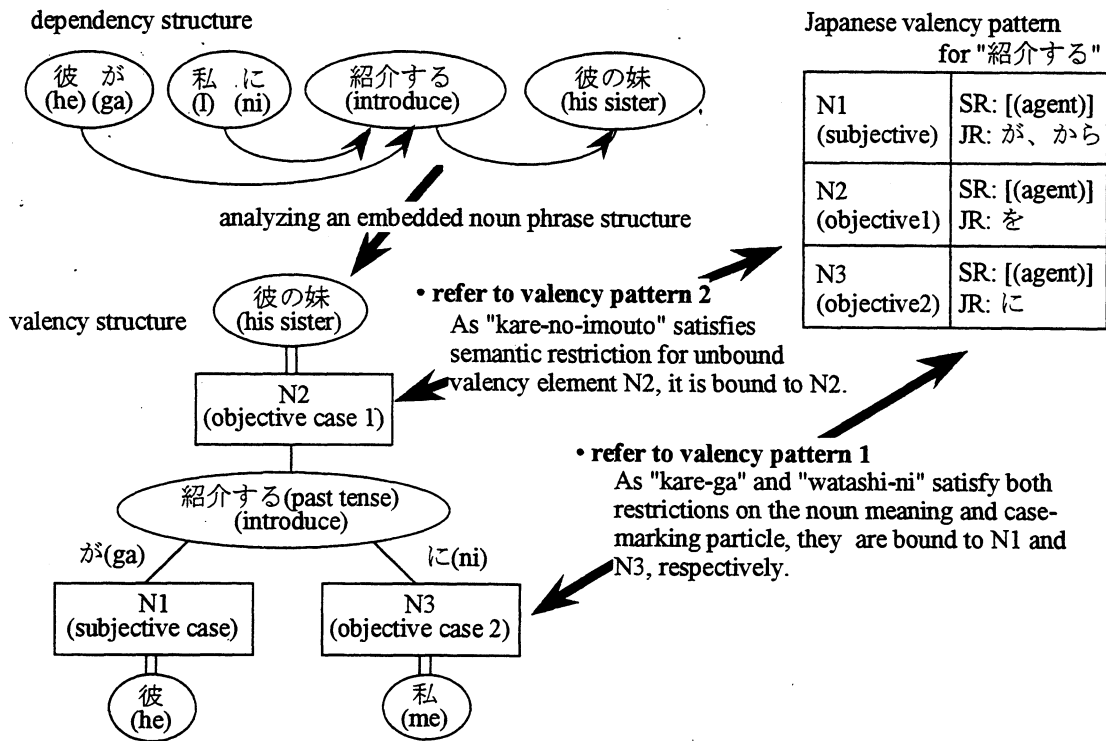


Figure 3: An example of analyzing embedded noun phrase structure (type-1).

N1 (subjective case) with case-marking particle "ga". According to ordinary sentence analysis, described in section 3.1, the modifier "hana-ga" is bound to the valency element N1, which means that the other modifier "zou-wa" is left unbound. In order to analyze the unbound modifier with adverbial particle "wa", let us consider a virtual valency element, modifying-N1 (subjective case 2), shown by the dotted square in figure 2. The virtual valency element has "wa" as JR and has "has-a N1" or "is-a N1" relation as SR. "zou-wa" is bound to this virtual valency element because "zou (elephant)" has "hana (nose)" bound to N1, i.e., "zou (elephant)" satisfies SR, "has-a N1", and "zou-wa" satisfies JR, "wa". Consequently, "zou-wa", associated with the predicate "nagai (long)", is analyzed as the pre-nominal case associated with the modifier "hana-ga". Generally, in processing this variant, adverbial particle "wa" is taken as a proxy for the 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, the analysis involves changing the valency structure (see the right side of figure 2).

3.3 OUTLINE OF CONVENTIONAL ANALYSIS FOR EMBEDDED NOUN PHRASE STRUCTURES

In ALT-J/E, the analysis of embedded noun phrase structures reveals the relation between the modified noun phrase and the embedded sentence by referring to valency patterns (Yokoo & Hayashi 87).

Figure 3 shows an example of embedded noun phrase structure analysis for "kare-ga watashi-ni shoukai-shita kare-no-imouto (his sister whom he introduced to me)". First, the analysis tries to bind modifiers with case-marking particles "ga" 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 the modifiers satisfy both SR and JR, they are bound to N1 (subjective case) and N3 (objective case 2), respectively. The analysis then tries to bind the noun phrase modified by the embedded sentence to an unbound valency element. So far, N1 and N3 in the valency pattern have already been bound. Therefore, as shown in figure 3, the only unbound valency element is N2 which is objective case 1. As the modified noun has no case-marking particle to indicate its association with the predicate of the embedded sentence, SR is only checked. The noun phrase "kare-no-imouto (his sister)" satisfies SR on the objective case 1, N2, i.e. [(agent)]. Therefore, the modified noun phrase "kare-no-imouto (his sister)" is bound to objective case 1, N2. The result is the valency structure shown in figure 3.

4. PROCESSING THE EMBEDDED NOUN PHRASE STRUCTURES

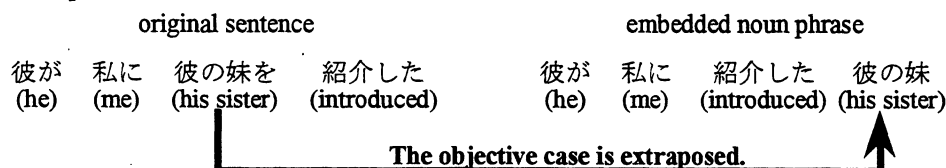
4.1 CLASSIFICATION OF EMBEDDED NOUN PHRASE STRUCTURES

This paper classifies embedded noun phrase structures into the following three types from the viewpoint of the relation between a modified noun phrase and an embedded sentence based on previous studies (Tsujii 83; Yokoo & Hayashi 87):

• type-1

In this variant, the modified noun phrase is analyzed as an extraposition of one case which is dominated by the predicate of the embedded sentence. Example-1 shows that the embedded noun phrase structure "kare-ga watashi-ni shoukai-shita kare-no-imouto (his sister whom he introduced to me)" is constructed by the extraposition of objective case 1 "kare-no-imouto (his sister)" from the original sentence "kare-ga watashi-ni kare-no-imouto-wo shoukai-shita (he introduced his sister to me)".

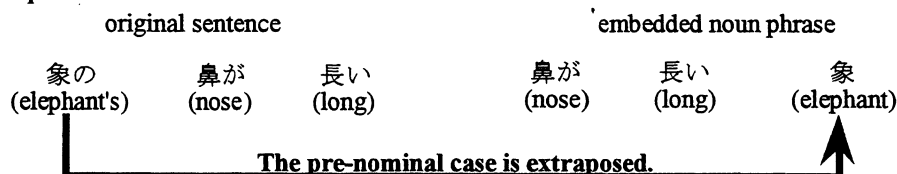
Example-1



• type-2

In this variant, in contrast to type-1, the modified noun phrase is analyzed as a pre-nominal case associated with one of the cases, especially subjective case, of the embedded sentence. Example-2 shows that the embedded noun phrase "hana-ga nagai zou (elephants whose trunks are long or elephants which have long trunks)" is constructed by extraposing the pre-nominal case "zou (elephant)" from the original sentence "zou-no hana-ga nagai (elephants have long trunks)".

Example-2



• type-3

In this variant, the relation between a modified noun and an embedded sentence is appositive such as "kare-ga chikoku-shita riyuu (a/the reason why he was late)". Only a few dozen nouns including indefinite nouns such as "mono", "koto" can act as modified nouns in type-3 construction.

4.2 PROBLEMS IN PROCESSING TYPE-2

Figure 4 shows the analysis of the type-2 example, "hana-ga nagai zou (elephants which have long trunks)". The predicate "nagai (long)" has only one valency element N1 with "ga". According to the conventional analysis described in section 3.3, the modifier "hana-ga (nose)" is bound to the valency element N1, which means that there is no unbound valency element. Therefore, "zou (elephant)" cannot be bound to the valency element for the predicate "nagai (long)". Conventional analysis thus processes type-2 as an exception. That is, type-1 and type-2 cannot be analyzed uniformly.

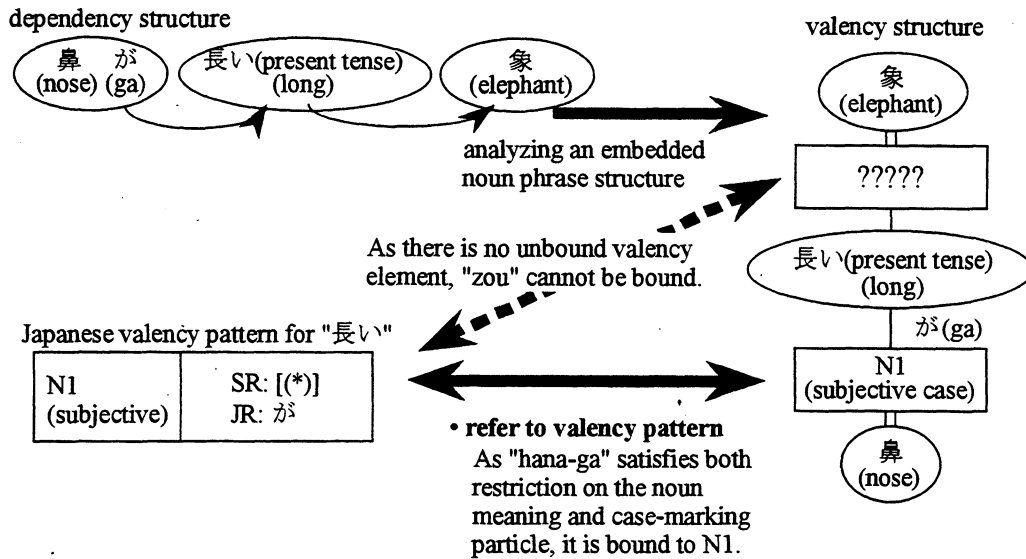


Figure 4: Analysis of type-2 example.

4.3 PROPOSED METHOD FOR ANALYZING EMBEDDED NOUN PHRASE STRUCTURES

In order to analyze both type-1 and type-2 uniformly, this section proposes a method for analyzing the type-2 embedded noun phrase structure.

Let us regard the type-2 embedded noun phrase structure as the phrase derived from the Japanese double-nominal-case construction in which adverbial particle "wa" is a proxy for the pre-nominal case-marking particle "no". As shown in section 3.2, this allows a virtual valency element, modifying-N1 to be considered in the valency structure. This virtual valency element has "wa" as JR and has "is-a N1" or "has-a N1" relation as SR. As the modified noun phrase has no particle that would indicate its association with the predicate of the embedded sentence, JR cannot be checked. Therefore, if the modified noun phrase satisfies SR, it can be bound to the virtual valency element modifying-N1.

Figure 5 shows the analysis of type-2 example, "hana-ga nagai zou (elephants which have long trunks)". First, in the same way as type-1 analysis, the modifier "hana-ga" for the predicate "nagai (long)" of the embedded sentence is bound to the subjective valency element N1. After this binding, as shown in figure 5, there is no

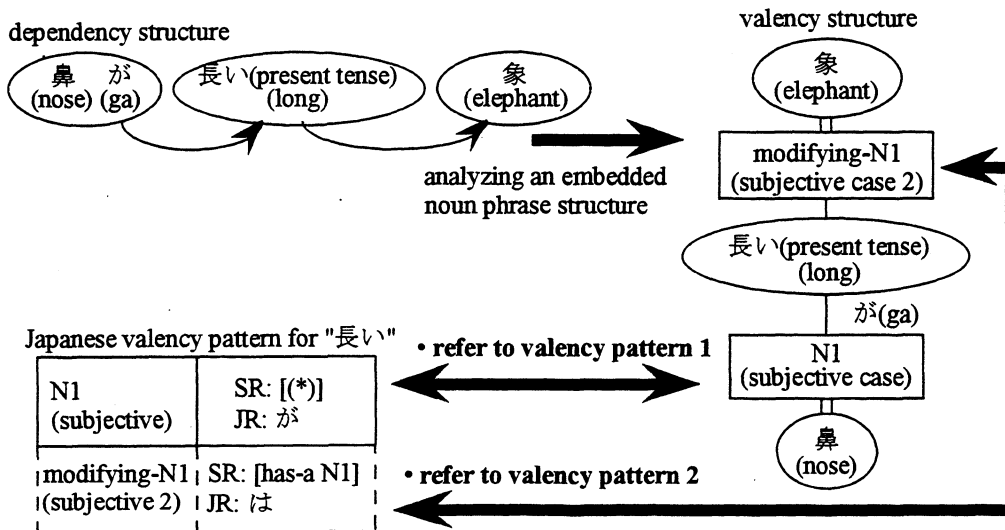


Figure 5: An example of analyzing noun phrase structure (type-2).

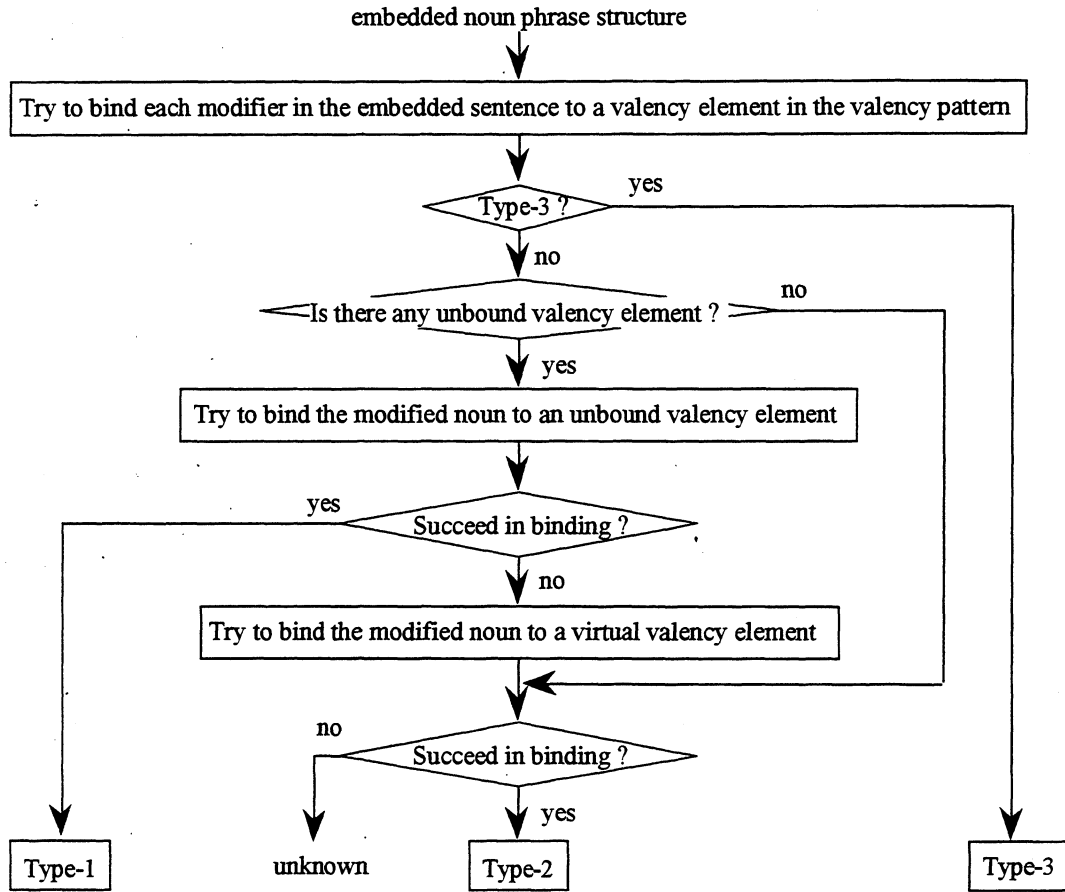


Figure 6: Processing flow for analyzing embedded noun phrase structure.

unbound valency element in the valency structure for the predicate "nagai (long)". Here, consider that the embedded noun phrase structure "hana-ga nagai zou" is derived from the Japanese double-nominal-case construction "zou-wa hana-ga nagai". Accordingly, we can consider the virtual valency element, modifying-N1, as shown by the dotted square in figure 5. As the relation between "zou (elephant)" and "hana (trunk)" is "has-a N1", the modified noun "zou (elephant)" satisfies the semantic restriction "has-a N1" for the virtual valency element modifying-N1. Therefore, the modified noun "zou (elephant)" is bound to the virtual valency element modifying-N1 and the valency structure shown in the right side of figure 5 is obtained. This process analyzes type-2 as the structure extraposing the valency element modifying-N1 in the same way as type-1.

Namely, this approach enables us to analyze type-1 and type-2 embedded noun phrase structures uniformly by binding the modified noun phrase to the valency element syntactically. Moreover, as shown by the semantic restriction "is-a N1" or "has-a N1" for the virtual valency element, the modified noun phrase bound to this element can be analyzed as the pre-nominal case associating the case bound to a real valency element (in figure 5, subjective case N1).

Figure 6 shows the processing flow for analyzing the three types of embedded noun phrase structure. First, an attempt is made to bind each modifier in the embedded sentence to a valency element in the valency pattern. Second, type-3 is separated because type-3 is the most specific. This separation is performed by checking whether the modified noun phrase is one of the noun phrases known to form type-3 or not. Here, the noun phrases for type-3, such as "koto", "mono", are pre-defined. Third, when there is any unbound valency element, an attempt is made to bind the modified noun phrase to an unbound valency element (type-1) and when there is no unbound valency element, an attempt is made to bind the modified noun phrase to a virtual valency element (type-2). Namely, the above approach enables us to analyze type-1 and type-2 structures uniformly in terms of binding a modified noun to a valency element in a valency structure.

5. EVALUATION

5.1 OUTLINE

In order to confirm the validity of the proposed method, we extracted embedded noun phrase structures, each containing one of the 153 most popular adjectives, as they appeared in newspaper articles over a 7 day period (about 15,000 sentences). We found 1,259 embedded noun phrase structures, and manually classified them into three types as described in section 4.1. Moreover, we checked whether they could be successfully analyzed or not by the flow shown in figure 6.

5.2 RESULTS

Table 1 shows the results of this evaluation. As shown in table 1, 77% (931+35 of 1,259) of extracted embedded noun phrase structures were type-1 and 96% (931 of 931+35) of type-1 were formed by extraposing the subjective case (see the column of manual classification in table 1).

The proposed method can successfully process type-3 and type-1, in which the subjective case is extraposed, to the same level as conventional methods. Moreover, the proposed method can analyze about 86% of type-2 forms that the conventional methods cannot handle so well. Namely, the results show that, by binding modified noun phrase in type-2 to the virtual valency element, the proposed method succeeds in processing uniformly both type-1 and type-2 embedded noun phrase structures. This fact confirms that the proposed method is valid for analyzing Japanese embedded noun phrase structures.

Table 1: Evaluation results.

Type		manual classification (1)	analyzing by proposed method (2)	(2) / (1) [%]
Type-1	Subjective	931	931	98.6
	Other	35	21	
Type-2		42	36	85.7
Type-3		251	251	100
unknown			20	
Total		1,259	1,259	100

5.3 DISCUSSION

Of the 35 embedded noun phrase structures in which cases other than subjective case were extraposed, 14 were unsuccessfully analyzed (see the second line in table 1) and 6 of 42 type-2 examples were also unsuccessfully processed (see the third line in table 1). Both these examples were classified as "unknown" by the proposed method. The modified nouns classified as "unknown" should be analyzed as the extraposition of an optional case for the embedded sentence. The proposed method determines the valency structure by referring to the valency patterns in the valency dictionary. Therefore, in order to handle the extraposition of an optional case, it must be described in the valency pattern. However, since optional cases, such as time, place, can appear in the valency patterns for all verbs and adjectives, only those unique cases that characterize each verb or adjective are described as valency elements in the valency patterns. Accordingly, the proposed method cannot analyze them correctly.

For each of the embedded noun phrase structures classified as "unknown", it may be possible to determine which optional case is extraposed by investigating the semantic category of its modified noun; for example, the modified noun having the semantic category [(time)] is derived from the optional case showing time. To improve analysis accuracy, it is important to advance research on the analysis of optional cases and the flow shown in figure 6 should be modified according to the research results on optional cases.

6. CONCLUSION

This paper has described a method for analyzing embedded noun phrase structures derived from the Japanese double-nominal-case construction based on the valency structure in ALT-J/E. The proposed method regards type-2 as the structure derived from the Japanese double-nominal-case construction in which the adverbial particle "wa" is a proxy for the pre-nominal particle "no".

By introducing the concept of the virtual valency element, we can analyze modified noun phrases of both type-1 and type-2, as the extraposition of one case which is dominated by the predicate of the embedded sentence. This approach allows us to process type-1 and type-2 forms uniformly based on the valency structure for embedded sentences in Japanese sentence analysis. Moreover, for type-2, the proposed method can analyze the relationship between a noun phrase bound to subjective case element N1 and a modified noun phrase bound to a virtual valency element. The evaluation results show that about 86 % of the latter type of the embedded noun phrase structures are correctly analyzed by the method proposed here. This means the proposed method is valid for analyzing embedded noun phrase structures derived from Japanese double-nominal-case construction. Therefore, in a machine translation system, such as ALT-J/E, this method allows the type-2 structure to be correctly translated into English, i. e. translation accuracy is improved.

REFERENCES

- (Hayashi 87) Yoshihiko Hayashi, "On the Japanese sentences analysis base on the valency structure", Working group report on natural language processing, IPS Japan, 62-6 (1987) (in Japanese).
- (Ikehara et al. 87) 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).
- (Ishigami 77) Teruo Ishigami, "Structure of Japanese double-nominal-case 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).
- (Kuno 73) Susumu Kuno, The Structure of the Japanese Language, The MIT Press, 1973.
- (Mikami 60) Akira Mikami, "Elephants have long trunks (zou wa hana ga nagai)", Kuroshio-shuppan, 1960 (in Japanese).
- (Murata 90) Kenichi Murata, "On the formal description of the meanings of Japanese adjectives", In Proceedings of the 40th annual meeting of IPS Japan, 5F-3 (1990) (in Japanese).
- (Oku 92) Masahiro Oku, "Methods for analyzing Japanese predicate idioms and function verb expressions", Journal of Information Processing, Vol. 15, No. 3, pp. 425-433 (1992).
- (Oku 96) Masahiro Oku, "Analyzing Japanese Double-Subject Construction having an Adjective Predicate", In Proceedings of the 16th International Conference on Computational Linguistics (COLING-96), pp. 865-870 (1996).
- (Tsuji 83) Jun'ichi Tsujii, "Japanese Sentence Structure Analysis", Working group report on natural language processing, IPS Japan, 38-5 (1983) (in Japanese).
- (Yokoo & Hayashi 87) Akio Yokoo and Yoshihiko Hayashi, "Analysis of Japanese Embedded-Noun-Phrase Structures", In Proceedings of the 1st annual meeting of the Japanese Society of the Artificial Intelligence, 7-2, pp. 369-372 (1987) (in Japanese).