Study on So-called "Particle Ellipsis"

Naoko Maruyama, Minako Hashimoto, and Wakako Kuwahata
Tokyo Woman's Christian University
Suginami-ku, Tokyo 167, JAPAN
e-mail: maruyama@twcu.ac.jp

1. Introduction

1.1 Case Marking with No Particle

One of the language phenomena that are specific to spoken languages is so called "particle ellipsis." An example is Shimbun (newspaper) yo-mu (read)?, in which the case particle wo is dropped. There have been attempts of recovering the elided particle in such sentences (Shimbun wo yo-mu in the above example). However, recovering the particle that makes the recovered sentence most natural does not necessarily mean that we have identified the case of the constituent (e.g. wo and ga). For example, Hasami (a pair of scissors) a-ru (be)? can be recovered as either Hasami ga a-ru or Hasami wa aru, but the meanings of these sentences are different from the original sentence. In this example, we consider that the case marking for the constituent Hasami is intentionally done by omitting the case particle. We call this type of case marking case-marking-with-particle-ellipsis (hereafter called CMWPE), and treat it as if it is another case particle such as ga and wo.

1.2 Other Case Markings

The same discussion can also be applied when other particles (kakari-particle and fuku-particle) are used instead of case particles. In Watashi (I) wa yo-mu, for instance, the particle wa indicates that the case of Watashi is ga, while in Hon (book) wa yo-mu, the same particle wa indicates that the case of hon is wo. Similarly, Watashi in Watashi dake yo-mu is case ga, and Hon in Hon dake yo-mu is case wo. We need a single framework that treats these two distinctive types of language phenomena in the same way.

1.3 Written and Spoken Languages

In ancient written Japanese, the case particles wo and ga were not used in general. Consequently, there are many situations in contemporary written Japanese in which these particles are not used, especially where the space is limited such as newspapers headlines and teletext. Since the usage of particle ellipsis in written language is somewhat different from that in spoken language, these two should be studied separately. Also particle ellipsis occurs in company with adverbial modifiers and conjunction.

1.4 Purpose of our study

From these observations, we have conducted our study on particle ellipsis with the following three objectives:

* Identify of the case indicated by CMWPE,

* Study various particle ellipsises in the overall grammar structure, and

* Reveal the difference between written and spoken languages.

2. Identification of case

2.1 Identification of Case for CMWPE

As argued above, we need to identify the case that is implied by CMWPE, not to recover the particle itself. In this paper, we define case as that is used in valency patterns, such as wo and ga (in this paper, we use italics for surface words such as ga and wo, and nonitalics for cases). Valency cases are "deeper" than corresponding surface particles but not as deep as the deep case defined in Case Grammar. One exception is the treatment of case ni. Usually a valency pattern does not distinguish the usage of case ni although there are several different ways of using case ni. Whenever possible, we tried to subcategorize the usage of case ni.

We used the database reported in "Research Report on Intelligent Speech Processing, Heisei 3" as the source of data (86 dialogs, 8,000 sentences), and investigated the case implied by a CMWPE. The case were determined based on the valency pattern of the predicate. The valency pattern used here is taken from "IPAL Verb Dictionary." If the predicate is not listed in this dictionary, we used "Nihongo Kihon Doushi Youhou Jiten (Dictionary of Japanese Basic Verb Usage)" and the reports issued by "Keiryo Keikaku Kenkyujo." The results showed the following important insights about the use of CMWPE:

* Most CMWPEs are used for cases ga, wo, and ni.

* When a constituent of a CMWPE modifies its immediately following verb, its case function is most likely wo. This seems natural since particle ellipsis is generally used for the case that is required by the predicate most, and case wo is the highest priority case of the most verbs.

* Other than wo, cases that are described near the end in the valency pattern seem preferred. These cases are considered more strongly tied to the verb.

* Case ni has several different meanings. The most frequent one is a target point, and an aim of a movement verb (e.g., Tsuri (fish) ni i-ku (go) (go fishing)) and special kinds of object (e.g. Shiken (exam) ni uka-ru (pass)) follow.

* If the position of a constituent with CMWPE is far apart from the predicate verb, the order is reversed and ga is stronger than the others. Probably this is accounted for by the fact that a constituent has a more topical nature if it modifies a distant verb.

* Sometimes it is difficult to identify the case even if it can be considered as CMWPE.

These results give rise to insights for case identification for other case markings as described in the following subsection.

2.2 Identification of Case for Other Case Markings

CMWPE is not the only situation where no case particle is present for marking the case. The following is a partial list of other related open problems that can be tackled in the similar way in a broader sense.

* Use of kakari-particle (e.g., wa) or fuku-particle (e.g., dake). Example: Watashi (I) wa yo-mu (read). Watashi should be identified as the case ga of the verb yo-mu.

* Relationship between a relative clause and the modified noun. Example: Hon (book) wo yo-mu (read) heya (room). Heya should be identified as the case de of the verb yo-mu.

* Noun-noun modification with case particle no. Example: Kakubakudan (nuclear bomb) no bakuhatsu (explosion). Kakubakudan should be identified as the case ga of the verb bakuhatsu-suru.

* Compound noun. Similar to No. 3 above.

These different situations pose different constraints on what cases are used. Generally speaking, they have broader usage than CMWPE. We believe that the method based on valency pattern that we used in identifying the case for CMWPE can also be applied for these situations.

3. Other Functions of Particle Ellipsis

In addition to the function as an implicit case marker, particle ellipsis has several other functions. Topicalization is one of them and has been studied for a long time. Since particle wa is normally used for topicalization, it is meaningful to compare the difference in their usage. They are similar in the following sense.

* A constituent in the beginning of a sentence has more topical nature.

* An already-known noun has more topical nature.

* Sometimes it is difficult to identify the case.

On the other hand, particle ellipsis is more neutral than particles wa and ga. Sometimes particles wa and ga imply exclusion. For example, when hearing the sentence Watashi (I) wa sono (the) hon (book) wo yo-mi-masu (read), the hearer naturally assumes that it is unlikely that others read the book. This is because of the exclusive nature of particle wa. Particle ellipsis does not have this exclusive nature, and therefore, we prefer Watashi sono hon wo yo-mi-masu if we do not want to give an impression that the only reader of the book is Watashi. If our standpoint were that topic should be exclusive by its own nature, probably we should argue that what particle ellipsis signifies is not a topicalization. Toridashi (extraction) as defined by Hasegawa would be a better word in that case.

Originally wa is considered to combine larger constituents and ga is used to combine smaller constituent. Particle ellipsis can be used for both cases. If we put wa after ryuhyou in Kyonen (last year) wa ryuhyou (drift ice) ko-na-katta-n-da-yo-ne (did not come, did it ?), it means contrast. If we put ga instead, it means that there was no such event as coming of drift ice. To state the fact of not coming as for drift ice, particle ellipsis is the most appropriate in this case. Smaller combinations are related to forming a compound noun as well. Examples are: "kyoumi (interest) a-ru (be)," "jikan (time) kaka-ru (consume)," and "onaka (stomach) su-ku (hungry)." Larger combinations follow a call. "Anta (you) naite-n-no-ne (are crying, aren't you ?)" is an example.

In summary, the roles of wa, wo, and particle ellipsis are described as follows:

wa: topicalization, contrast

ga (wo, ni): case marking, enumeration (i.e., no other things share the same statement)

particle ellipsis: simple omission of case particle, toridashi

We argue that topicalization indicated by the particle wa is different from toridashi in their nature. The function of toridashi is to extract and set the topic at the time of utterance, while topicalization by wa identify the topic that is already extracted and set.

Factors that make a particle ellipsis function as toridashi include (1) structural position of the constituent in the sentence, (2) characteristics of the noun (unknown/known, person, pronoun, etc.), (3) characteristics of the predicate (e.g., types of intransitive verbs), (4) characteristics of the sentence (interrogative, instructive, soliciting), (5) pragmatic factors (focus, antecedent, honorific), and (6) phonetic features (pause, speed, accent).

There are other usage of particle ellipsis than case relations. As for ones that follows a noun, adverbial modifiers and conjunctions are important ones. Adverbial modifiers describe situation, degree, quantity, time , and so on. Conjunctions include apposition and exemplification.

4. Written Language and Spoken Language

Particle ellipsis is used in written and spoken languages differently.

One of the phenomenon of particle ellipsis that is specific to written languages is called "temporary one-word." Temporary one-words are traditionally used in newspaper headlines. More recently, they are used in teletext in order to put as much information as possible in a limited space.

On the other hand, there are particles unique to spoken languages. Toka, nanka, and nante are among them. We are trying to incorporate them in our grammar.

One of the characteristics of spoken languages is that they are fragmentary. In a spoken conversation, a lot of information is provided by the surrounding context, and only fragmentary information appears in uttered sentences. Particle ellipsis can be considered as one of the manifestation of this fact. For example, the sentence Kyou (today) no sora (sky), ao-i-ne (is blue, isn't it?) brings up the object itself directly, and then states that it is blue. In this case, kyou no sora is more independent. Since it is trying to avoid an analytic statement, it has some effect on the communication underway. It is called yawarage (softening).

Keywords: identification of case, topicalization, written language and spoken language