A textual analysis of subjects in English and Japanese: from the viewpoint of information structure

Takumi Ihara

Liberal Arts Center, Nagano University of Health and Medicine

ABSTRACT: Our analysis of subjects in Japanese and English starts from the assumption that there are four corresponding textual interpretations: theme, contrast, neutral description, and exhausting listing, as proposed by Kuno in 1973. First, we clarify the definitions of the various conceptions in information structure. Second, we characterize those four corresponding textural interpretations one by one from the point of view of information structure. As a result of those characterizations, it becomes clear that in both English and Japanese, the subject of thematic interpretation must carry known item and, at the same time, old information. On the other hand, subjects from the other three interpretations may be considered either known or unknown item, but also bear new information. Lastly, we conclude that subjects in English and Japanese coincide in their textual features, though they may differ in their syntactic ones.

Key words: Comparative Linguistics, Subjects in English and Japanese, Textual Grammar

1. Introduction

In order for an utterance to be considered satisfactory, it must, first of all, be grammatical. Beyond that, it must be deemed appropriate, since sentences must be consistent with the stream of discourse. In that sense, we may say that an utterance comes into existence in accordance with its context.

When a message is presented in sequence, the initial element is of great significance in the following two respects: first, as the communicative point of transfer of the preceding context, and secondly, as the communicative point of departure for the rest of the sentence.

The initial element of a sentence may be called its "theme." In a declarative sentence, it is common that what is situated in the initial position of a sentence is, both in English and Japanese, considered to be the subject.

In this sense, a sentence's subject, whether in English or Japanese, plays an important role in textual as well as syntactic function.

Ihara[1] examined subjects in English and Japanese, first from a syntactic and then from a semantic viewpoint. He attempted to clarify syntactic and textual features of the two types of subjects, defining explicitly the various conceptions often used ambiguously in explanations of syntactic and textual analysis.

As a result, the following two postulations became evident. First, subjects in English and Japanese are quite different from each other in their syntactic features. Secondly, according to the stream of discourse, they have four textual interpretations each other: theme, contrast, neutral description, and exhaustive listing. In addition, these four textual interpretations in both subjects completely correspond to each other in meaning and function. That is to say, they have four corresponding textual interpretations.

The latter findings allow us to assume that English and Japanese subjects may have the same textual features. However, judging from the fact that these interpretations are produced in the stream of discourse, we may assume that they have something to do with information structure, which has a great influence on the stream of discourse.

Hence our task in this paper is to characterize these four interpretations from the point of view of information structure. By so doing, we hope to be able to clarify the textual features of subjects in English and Japanese more accurately.

2. Conceptions in Information Structure

Before we address these four interpretations from the point of view of information structure, it is necessary to make explicit the definitions of the

various conceptions in information structure. This is advisable because these conceptions, as things stand now, are used so ambiguously as to lead to possible misunderstanding or confusion in characterization.

The conceptions often used in the explanation of information structure are "old information" and "new information." It seems that the former has been used to indicate what has already been given, and the latter, what has not yet been given. That is why the term "old information" is sometimes referred to as "given information." These rough definitions, however, seem to have led to misunderstanding of information structure. The problem lies in what has already been given or has not yet been given. Murata[2] assumes that old information means that which the speaker supposes the hearer can infer from the preceding context, and that new information means that which the speaker supposes the hearer cannot infer from the preceding context. That is to say, Murata considers that whether information is old or new expresses whether it can or cannot be inferred from the preceding context. Acceptance of his position assumes that what can be inferred from the preceding context is equivalent to the presupposition of a statement, and what cannot be inferred from the preceding context is that which is indefinite in the presupposition. Hence, we may assume that old information indicates a presupposition which has already been given in the preceding context, and that new information indicates a focus or solution which has not been given in the presupposition until the response is given. This is illustrated in the following:

(1)　Whom did Alexander hit?
　　　Alexander hit MARY.
　　　old information new information
　　　　　　　　　　(focus)
　　　Alexander hit someone
　　　　　　presupposition

We will henceforth employ "old information" instead of "presupposition," and "new information" instead of "focus," whether it is marked or unmarked.

From the point of view of order, old information usually precedes new information as theme does rheme, because to comment on what can be inferred from the preceding context is, communicatively speaking, more natural than to say something about what cannot be inferred from the preceding one. This seems to be a fundamental rule in the stream of discourse. Consequently, we may assume that the subject of old information is "unmarked" because of following the rule, and, on the contrary, that the subject of new information is "marked" because it violates the rule, and, as the penalty for the violation,

receives intonation nucleus and contrastive sense. This explains in another respect why the theme (the subject) to which focus is shifted from its neutral position is "marked" as is discussed in Ihara[1].

From the point of view of information value, new information is more valuable than old information, since it is a solution to be given to the hearer. Hence, it is impossible to omit new information, whereas it is possible to omit old information, considering it no longer worth saying.

However, here is a problem. Consider the following:
(2)　Who is taller, John or Tom?
　　　JOHN is taller than Tom.

In (2), JOHN cannot be inferred from the preceding context until the response is given, and therefore we can take it as new information. However, we must note that it has already been mentioned in the preceding context. That is to say, although it is new information, it has already been given in the preceding context in another sense. In order to resolve the contradiction properly, we must regard the conception of what can or cannot be inferred from the preceding context as quite different from that of what has or has not been mentioned in the preceding context. We will name the former "known item," and the latter "unknown item" according to Yasui[3]. (As for these items, Kuno[4] employs the terms "anaphoric" and "nonanaphoric" as nearly equivalent to our terms "known item" and "unknown item" respectively. However, in our framework "known item" contains even generic noun phrase as is shown below though his term "anaphoric" does not always contain.)

In English it is usual that, as an indicator of known items, the anaphoric the or anaphoric determiners such as this, that, its, his, her, and theirs, are put before nouns, or, more often, the personal pronouns of subjective and objective cases are used. On the contrary, indefinite articles are put before nouns as indicators of unknown items. In Japanese, the indicators of known and unknown items do not come into existence so strictly as in English because there are no articles in Japanese, and therefore it may be said that the system of determiners in Japanese is not established so perfectly as it is in English. However, some anaphoric determiners such as *kono*, *sono*, *ano*, and *onaji* sometimes occur before nouns or some personal pronouns as indicators of known items. With respect to indicators of unknown item in Japanese, some indefinite determiners such as *aru* and *saru* seem to occur infrequently before nouns, or some appositives, such as *toiuhito* and *toiutokoro* after proper nouns.

Now, it should be noted that there is a case where

a noun phrase that has not been mentioned in the preceding context becomes a "known" item. Look at the following example:

(3)   The river which flows through London is the Thames[5].

We can paraphrase (3) as follows:

(4)   A river flows through London. The river is the Thames.

As will be easily understood, the cataphoric the in the river of (3) is equivalent to the anaphoric the in the river of (4). In this sense we can regard the river in (3) as a known item. That is to say, the cataphoric antecedent followed by the restrictive relative clause may be taken to be a known item. Going further still, we may assume that "cataphoricity" is a variant of "anaphoricity."

A similar phenomenon can also be seen in Japanese:

(5)   *Kare no tateta ie wa goka da*.
       'The house which he built is gorgeous.'

Ie in (5) may be taken to be known item because (5) can be paraphrased as (6):

(6)   *Kare wa ie wo tateta. Sono ie wa goka da*.
       'He built a house. The house is gorgeous.'

We defined the above "mentioned in the preceding context" (anaphoric) as "known" except in special cases such as in (3) and (5). But suppose that a teacher comes into the classroom, looks at the sentences on the blackboard that ought to have been erased, and asks, "What about the blackboard?" Although the blackboard in this case comes into existence for the first time in the discourse and therefore has not been mentioned in the preceding context, it is a referent which the hearers (the students) can recognize easily as soon as they hear it. In this sense we may also regard the blackboard in this case as known. If so, we may assume that all the words or phrases denoting the referents that the hearer can recognize easily from the situation are known items. In such cases, it is usual in English that the exophoric the or deictic determiners such as this, that, and its are put before nouns as indicators of a known item. Similarly, in Japanese, deictic determiners, such as kono, sono, and ano, are usually put before nouns.

Furthermore, consider the following examples:

(7)   Who said so?
       JOHN did.

JOHN in (7) neither is referred to in the preceding context nor can be recognized from the situation. But we can suppose that in this case he is a person who has already been "known" to both of the speaker and the hearer. In other words, he is a person who has already been recorded in the registry of the present discourse. If his entry in the registry had not been accomplished, the speaker first would have to establish it and then

talk about John. In this sense we may also regard JOHN in (7) as a known item. Similarly, nouns of unique reference in this universe of discourse, such as the sun, the moon, my wife, my children, may be taken to be known items because they seem to be in the permanent registry.

Talking of the permanent registry, generic noun phrases seem to be in the permanent registry of discourse and do not have to be reentered into the temporary registry for each discourse because they refer to classes such as men (human beings in general), Americans (Americans in general, all Americans, any American), and the linguist (linguists in general, all linguists, any linguist) and not to some arbitrary members of the classes[6]. In the sense that they are permanently stored in the registry of discourse, generic noun phrases are also known items.

We will sum up by saying that a known item is what the hearer has already known or can easily recognize from the situation, in other words, a word or phrase of definiteness, or a generic noun phrase.

Although it is normal that a known item bears old information and an unknown item provides new information, it sometimes happens that a known item bears new information as was seen in (2). However, it by no means happens that an unknown item bears old information. Consequently it can be said that old information must always be a known item while new information can be either a known item or an unknown item. Therefore, it is possible that there are three combinations of conceptions in information structure as in the following:

(8)   Whom does John love?
       He loves MARY.
       ⌐known item
       └old information

(9)   Who is the tallest among John, Tom, and Mary?
       JOHN is.
       ⌐known item
       └new information

(10) Did anyone come to visit while I was away?
       A WOMAN I HAVE NEVER SEEN did.
       ⌐unknown item
       └new information

Now that it has been made clear that there are three combinations of conceptions in information structure, our next task is to characterize the subjects of four textual interpretations one by one from the point of view of information structure.

3. Characterization

Since it is recognized as a result of Ihara[1] that subjects in English and Japanese have four textual interpretations which correspond to each other, we can deal with both sets of subjects simultaneously in the

process of characterization.

Firstly, the subject of thematic interpretation must be a known item because theme, being "what is discussed," is required in and of itself to be what the hearer has already known or can recognize easily from the situation. Talking about what the hearer has not known yet or cannot recognize from the situation would be communicatively senseless. Besides, it must be old information. When we see "what is discussed" in the stream of discourse, we find that it is the communicative point of transfer of the preceding context. Needless to say, communication proceeds most smoothly when the communicative point of transfer of the preceding context is what the hearer can infer from the preceding one. Let us look at the following example:

(11) Whom did Alexander hit?

　　　Alexander (He) hit MARY.

　　　(Alexander wa) MARY wo nagutta.
　　　⌐ known item
　　　└ old information

Since the subjects in the response of (11) are referred to in the preceding context and it can be inferred from the preceding context that it was Alexander who did the hitting, we can regard them as given items and at the same time old information. We can also affirm this from the fact that in actual utterances the personal pronoun he is substituted for Alexander in the response of (11) in English and that *Alexander wa* in the response of (11) is usually omitted in Japanese.

Secondly, the subject of contrastive interpretation may be either a known or an unknown item, but it must be new information.

(12) Did Alexander and Tom hit Mary?

　　　ALEXANDER hit Mary, but TOM didn't.

　　　ALEXANDER wa Mary wo nagutta ga, *TOM*
　　　⌐ known item　　　　　　　　　⌐ known item
　　　└ new information　　　　　　　└ new information

　　　wa nagurana katta.

Since the subjects in the responses of (12) are referred to in the preceding context, we can take them to be known items. However, it cannot be inferred from the preceding context who hit Mary. That is to say, it is not clear until the response is given whether it was both Alexander and Tom, just Alexander, or just Tom who hit Mary. Consequently, we can regard the subjects as new information. Concerning this consideration, Chafe[7] proposes that the Japanese particle wa reflects old information and the particle ga new information. But this is against fact. Although thematic wa reflects old information, contrastive wa reflects new information as is shown in (12).

With regard to cases in which the subject of contrastive interpretation is an unknown item, look at the following:

(13) Was the party nice?

　　　SOMBER FOLKS joined it, but AMUSING ONES didn't.

　　　INKINA HITO wa kita ga, OMOSIROI HITO
　　　⌐ unknown item　　　　⌐ unknown item
　　　└ new information　　　└ new information

　　　wa konakatta.

Since the subjects in the responses of (13) are not mentioned in the preceding context, nor can they be recognized from the situation, we can consider them to be unknown items. Here we would add that, as subjects of contrastive interpretation, known items tend to occur more frequently than unknown ones. This is because the subject of contrastive interpretation bears the nature of theme in itself as is clear from the fact that it is shown by the same particle wa as the subject of thematic interpretation in Japanese.

Thirdly, the subject of neutral-description interpretation may be either a known or an unknown item, but again must be new information. Let us look at the following example:

(14) What happened next?

　　　Alexander hit Mary.

　　　Alexander ga Mary wo nagutta.
　　　⌐ known item
　　　└ new information

Although the subjects in the responses of (14) are not mentioned in the preceding context, we can suppose that in this case Alexander is a person the hearer knows and therefore that his entry in the registry is established. Consequently, we can regard them as known items. The neutral-description sentence consists of new information only, because the scope of marked focus extends over the whole statement as is referred to in Ihara[1]. Consequently, the subject of the neutral-description interpretation, one of the elements that constitutes a neutral-description sentence, is also new information.

With regard to cases in which the subject of neutral-description interpretation is an unknown item, look at the following example:

(15) What happened next?

　　　A girl came into the room.

　　　Aru onnanoko ga heya ni haittekitannda.
　　　⌐ unknown item
　　　└ new information

In both the English and the Japanese sentences, the subjects in the responses of (15) are not mentioned in the preceding context, nor are they recorded in the registry of the present discourse, nor can they be recognized from the situation, so that we can consider them to be unknown items.

Lastly, the subject of exhaustive-listing interpretation

may be either a known or an unknown item, but must also be new information. Look at the following example:

(16) Who hit Mary?
    ALEXANDER hit Mary.
    ALEXANDER ga Mary wo nagutta.
    ⌐known item
    └new information

Although the subjects in the responses of (16) are not mentioned in the preceding context, we can suppose that in this case Alexander's entry in the registry is established just as it was in (14). In addition, since they are the answer to the question who in (16), they represent information which cannot be inferred from the preceding context. Therefore, we can consider them to be known items and new information.

Concerning cases wherein the subject of the exhaustive-listing interpretation is an unknown item, look at the following example:

(17) Who came here?
    A MR. SMITH came here.
    SMITH SAN TOIUHITO ga kokoni kita.
    ⌐unknown item
    └new information

The subjects in the responses of (17) as in (15) are not referred to in the preceding context, nor are they recorded in the registry of the present discourse, nor can they be recognized from the situation. So, we can consider them unknown items. It goes without saying that in (17) the indefinite article a in English and appositive *toiuhito* in Japanese are indicators of unknown items. Let us add here that the subject of exhaustive-listing interpretation can be the focus in the cleft sentence:

(18) It was ALEXANDER who hit Mary.

4. Conclusion

In the course of the textual analysis of subjects in English and Japanese, we started our discussion based on the position that they both have four corresponding textual interpretations such as theme, contrast, neutral description, and exhaustive listing. Then, after making explicit the definitions of the various conceptions in information structure, we characterized them from the point of view of information structure according to those four interpretations. As a result of the characterization, it has become clear that in both English and Japanese the subject of thematic interpretation must bear known item and at the same time old information, whereas the subjects in the other three interpretations may be either known or unknown items, but must bear new information. This is illustrated as below (Fig. 1):

Subject in Japanese		Subject in English	
	wa	Theme	● old information ● known item ● unmarked
		Contrast	● new information ● known or unknown item ● marked
	ga	Neutral description	● new information ● known or unknown item ● unmarked
		Exhaustive listing	● new information ● known or unknown item ● marked

Fig. 1. Textual features of subjects in English and Japanese.

From what has been stated thus far, we can conclude that subjects in English and Japanese coincide in their textual features though they may differ in their syntactic ones. This might allow us to foretell that various conceptions concerning information structure apply, not only to English and Japanese, but also to other languages.

We conclude by presenting two possible problems. One is about the assessment of degree of contrastive sense. In this paper any element bearing new information has been dealt with as having contrastive sense except those of neutral-description sentence. Therefore, in our framework, MARY in (19) is regarded as having contrastive sense:

(19) Who loves Tom?
    MARY loves Tom.

This is because insofar as MARY is uttered as an answer to a wh-question, it is implied that the speaker loves Mary, and not anyone else.

Compare (19) with (20):

(20) Do Mary and Jane love Tom?
    MARY loves Tom.

It is obvious that MARY in (20) is more intense in its contrastive sense than in (19). This means that there is a degree of contrastive sense. It then follows that we have dealt with new information without considering the degree of contrastive sense. For a more precise description, we would be required to clarify the relation of new information to its level of contrastive sense.

The other concern is in reference to the thematic sense of the subject in a sentence of neutral description. We have dealt with the interpretation of the subject in neutral-description sentences as being fixed or unchangeable, that is to say, neutral-descriptive interpretation only. In reality, Nakajima[8]

proposes that there should be no themes in the sentences of (21) even though they have their own subjects. In that sense, we might understand that Nakajima considers the subjects in the sentences of (21) to have no thematic interpretations at all:

(21) a. Bees swarm in the garden.

　　　b. Trout teem in this lake.

As is referred to in Ihara[1], an existential sentence and a presentational sentence have no theme because they consist of rheme only. In fact, if we take the sentences in (21) as variants of existential sentences, no problems will arise, since they are nearly equivalent to 'There are bees swarming in the garden' and 'There are trout teeming in this lake,' respectively.

However, even if the sentences in (21) have no theme, it cannot be denied that the subjects of the sentences in (21), Bees and Trout, are assumed to be, in a sense, "what are discussed" (themes). If so, the subject in the neutral-description sentence in (14) which is neither an existential nor presentational sentence, Alexander, might have a stronger thematic sense than Bees and Trout. Therefore, it is necessary to bear in mind that there might be some level of degree in the thematic sense of the subjects in neutral-description sentences.

It is not clear yet whether these two kinds of degrees may or may not influence the interpretations of contrast and neutral description, but if we should find such examples as bring about the difference in interpretations, our analysis would have to be a different one.

Referneces

1. Ihara T: Syntactic and Textual Features of Subjects in English and Japanese. 長野保健医療大学紀要, 2019; 3: 43–51.
2. 村田勇三郎：機能英文法. 大修館書店, 東京, 1982; p 185.
3. 安井　稔：新しい聞き手の文法. 大修館書店, 東京, 1978, p 183.
4. Kuno S: The Structure of the Japanese Language. The MIT Press, Cambridge, 1973; pp 39–49.
5. 中島文雄：英語の構造　上. 岩波書店, 東京, 1980; p 27.
6. Kuno S: The Structure of the Japanese Language. The MIT Press, Cambridge, 1973; p 41.
7. Chafe WL: Meaning and the Structure of Language. The University of Chicago Press, Chicago, 1970; p 233.
8. 中島文雄：英語の構造　下. 岩波書店, 東京, 1980; p 42.