A BEHAVIORAL CHARACTERIZATION OF TOPICALIZATION: WHAT IS TOPICALIZATION?-- FROM A COMPUTER SYSTEM POINT OF VIEW[1)]

H. Takahasi,[2)] O. Fujimura, and H. Kameda

The use of the grammatical particle wa in marking a noun phrase generally indicates presupposition of the marked noun phrase. Further, it is quite often the case that a noun phrase with wa constitutes the so-called "topic term" of a sentence. The notions of presupposition and topic in this connection, however, are, formally, not quite clear. In this paper we shall attempt to outline the typical semantic characteristics of wa by formulating a first-order approximation of performance processes in a man-machine communication situation.

Let us assume a machine information processing system as a model of the speaker/listener in a conversation in Japanese. We shall further assume that the memory system of the "speaker" or "listener" is divided into two portions, a primary memory and a secondary memory: the primary memory stores the incoming information temporarily whereas the secondary memory stores information permanently, analyzed and reorganized for different ways of retrieval. During the course of conversation, let us assume, the content of incoming and outgoing sentences is stored in the primary memory in linguistic forms, but not directly in surface forms (see infra). Some interpreting processes will take place subsequently operating on these sentences, and the results of these operations will be absorbed into the secondary memory in some"digested" forms of semantic structures under the relevant semantic entries represented by some kind of linguistic forms. Here we will have to limit our immediate concern to the level of operation in the primary memory.

First, as an example for ease of conceptualization, let us consider the following sequence of statements in air terminal announcements. The repeated use of the particle wa in this example may represent a typical case of topicalization of a presupposed NP.

(1-1) Tadaima kara honzicu no syuppacubin o
(now) (from) (today) ('s) (departing flight)(object case particle)

osirase si masu.
(inform)(do) (cordial ending).

(1-2) Syuppacubin wa JAL-531, PAN-315, NW-153 desu.
(be)

(1-3) JAL-531 wa Honolulu-keeyu San-Francisco-yuki de, 22:30-hacu
(via) (for)(be) (take-off)

desu.

(1-4) PAN-315 wa Alaska-keeyu New-York-yuki de, 23:00-hacu desu.

(1-5) NW-153 mo Alaska-keeyu New York-yuki de, kore wa 23:15-
(this)

hacu desu.

In the discourse above, the use of wa in (1-2) is based on the preceding sentence (1-1) which introduces the pertinent noun phrase syuppacubin. Thus if (1-1) were skipped, the sentence (1-2) would suggest an omission of some preceding statement that would be more or less equivalent to (1-1); or in other words there must be some environmental situation strong enough to justify such an omission. In a machine model of the primary memory it is clear that we will have to exclude all such implicit environmental information from our consideration (except for a highly limited universally defined set of notions, see infra). At this point in the discussion, let us assume that at the beginning of a discourse the content of the primary memory is blank for the speaker/listener.[3] Thus a first sentence such as (1-1) containing no wa-NP introduces in the primary memory several noun phrases. The new terms thus "registered" in the primary memory can be readily identified and referred to directly, without implicational rules, when the same terms appear in the succeeding sentences. These are the items that are qualified to be cited as "topic" (or "theme") in subsequent statements.

Any sentence stored in the primary memory, let us assume, is represented as a structural relation between its constituent substantive

items. We may here assume that this structure is something similar to the deep structure in Chomsky (1965). It is clear now, however, that a surface sentence structure can carry more information than its underlying deep structure (Kuroda 1965; Chomsky 1970). Even though it is still not quite clear what kinds of syntactic elements supply such surface-proper information, it seems appropriate to say that the use of the Japanese particle wa represents such a case (Kuroda ibid). The purpose of the present paper is to try to represent the actual performative function of this syntactic device in terms of a simplified but concrete performance model and to discuss the functional effects in relation to the inherent and general semantic characteristics of the particle wa.

A list as sketched in Fig. 1 will appear in the primary memory of the speaker/listener upon transmission of the first message in the discourse

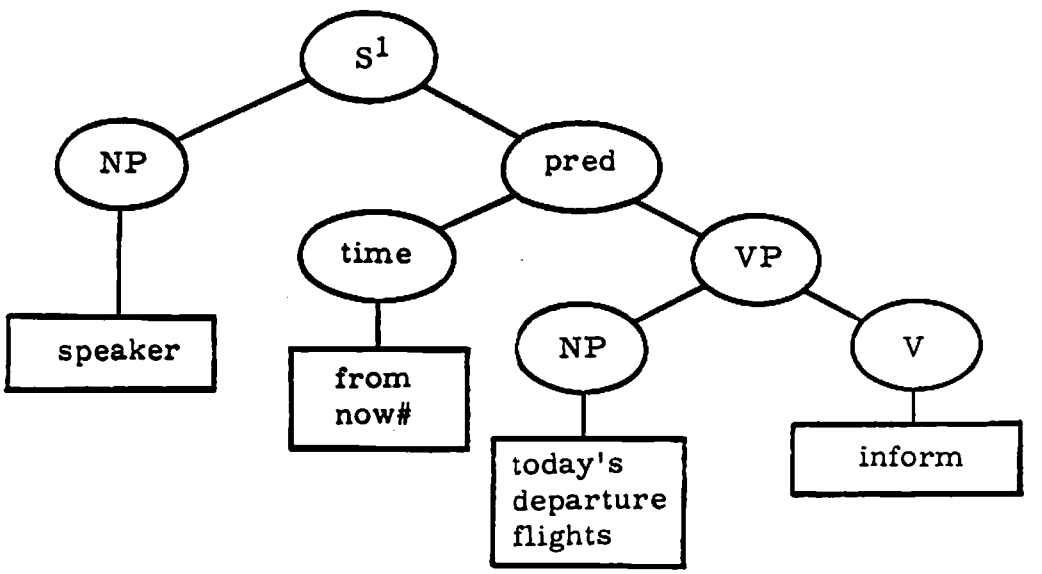

Fig. 1. Sentence (1-1) as stored in the primary memory when it is (to be) received by the listener. The mark # indicates a variable time record, which we will not discuss in this paper.

example above. Among the list items of sentences in general, some are given the status of "list head." A list head has a sentence pointer which points to an S-node. This occurs, perhaps among other cases, whenever in a discourse an NP appears at the head of a sentence and is marked by the particle wa. This situation may be illustrated as in Fig. 2, where a primary memory item is represented by a rectangle, a list head by a double rectangle, and a sentence pointer by an arrow attached to a rhombus enclosing a numeral in it.

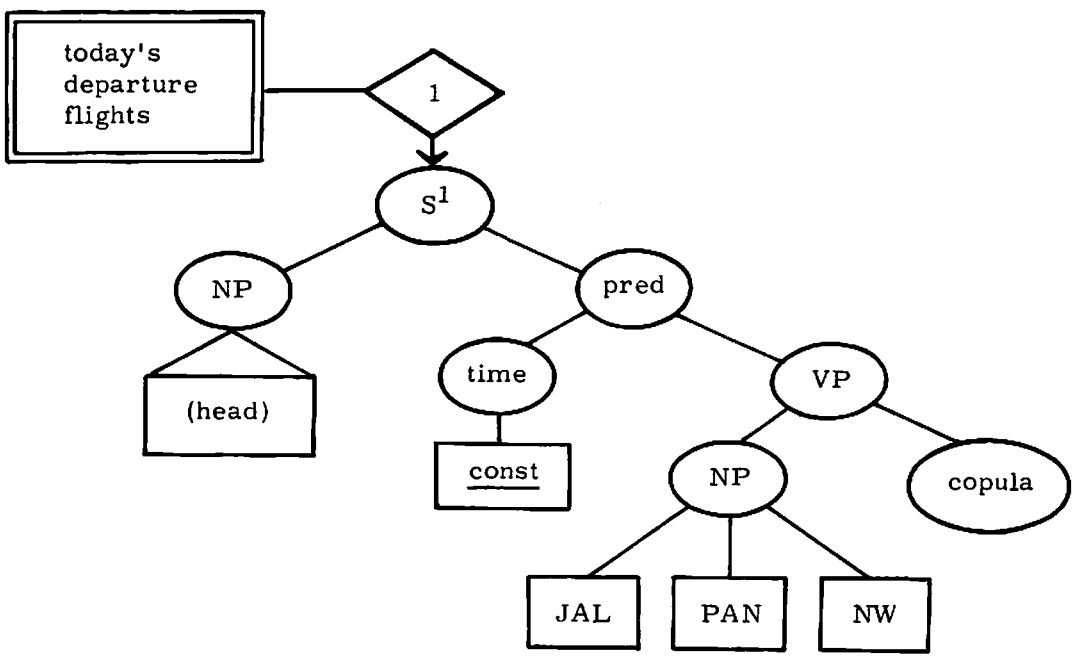

Fig. 2. The list structure after receiving sentence (1-2).

The marker wa tells the listener that the NP to which it is attached must exist in his primary memory (i. e. this item is "known") and therefore can be treated as a "presupposed" term, and that the content of the new sentence (in the form of deep structure) shall be attached to this item (through a sentence pointer).[4] Upon receipt of such a sentence, in

general, the machine (listener) will look for the pertinent term in the primary memory, first by searching through all list heads. If it finds the term, then another sentence pointer is added to it and this connects the new sentence to the list head; if not, as in the example above, it creates a new list head, and the sentence is listed under this head.

The list structure in Fig. 2 is essentially different from the one in Fig. 1 in that the former is attached to a list head whereas the latter is not. A semantic structure under a list head is ready for storage as part of the listener's knowledge (in his secondary memory), whereas one without a list head simply represents the logical structure of the received sentence as a direct memory (but in a form of deep structure) of the listener's experience.

As is well known, copula statements typically require the particle wa for their subjects. In the examples cited above, in particular, it is grammatically impossible to replace the particle wa by the subject-case particle ga in any of the sentences (1-2) through (1-5) in this discourse context. The particle ga can be used for the same sentences only when the NP is used in other discourse environments as a focus term (e. g. as a reply to a question which employs a wh-NP in place of the pertinent subject NP[5]). Thus in the usual cases of copula statements, unless under such strong contextual conditions, we have no possible memory structure in our model but the one depicted in Fig. 2.

If a sentence is used for an announcement of some actual event that is taking place, however, the particle ga may be used as in the following sentence where there is no presupposition of the predicate:

(2-1) Honolulu keeyu San Francisco-yuki JAL-531 ga (departure) syuppacu

itasi masu.
(is going to do- polite ending)

(2-2) Gotoozyoo no okyakusama wa geeto ni oisogi kudasai.
(on-boarding) (passenger) (gate) (onto) (hurry) (please)

In (2-1), the subject NP is not treated as a presupposed topic term; that is to say, there is no intention that the subject NP be treated as a list head,

containing the predicate "is going to take off" under it. Rather, the sentence will be left "unprocessed" in a form like Fig. 1, which may merely be kept as a record as it is.

In an imperative sentence like (2-2), the wa-NP is certainly presupposed, i. e. the listener is supposed to "know" what "on-boarding passengers" means-in particular whether it applies to him or not. But in our model, there will be found in the primary memory no such NP registered by means of a preceding sentence.[6] In order to treat this kind of case properly, we will have to assume, modifying our first assumption stated above, that a certain set of such items is presupposed as a sort of convention for a given conversational situation. There are also certain terms of this kind that are universally available regardless of the particular conversational situation, like the pronouns I and you, the adverbials now, here, etc. In addition, in certain kinds of discourse, we would need to include in the domain of presupposition some more general terms, perhaps even all the entry items of a sort of encyclopaedia (consider e. g. the sentence "Cuki wa cikyuu no eesee de aru-----The moon is a satellite of the earth, " in a textbook of high school physics). While these encyclopaedic entries with their descriptions are presupposed and therefore can appear with wa as "generic" terms (Kuno 1970), they are obviously beyond the capacity of our "primary memory, " and it is natural to assume that they are stored in the secondary memory. In these cases, let us assume, the search is first made through the list heads, and then all list items in the primary memory, and finally, at the expense of searching time, the entries in the secondary memory are looked up.

An imperative sentence such as (2-2) raises another problem about the function of wa. The wa-NP in this kind of situation is used for calling the attention of applicable addressees.[7] In fact, an appropriate English translation of this sentence may be: "Passengers going aboard, please hurry up to the departure gate. " In cases where the addressee is uniquely predetermined, the same wa may be used for calling the addressee's attention to a certain object within the scene. For example, an instruction given to a student sitting in an oral examination might be:

(3) Hon wa cukue no ue ni oki nasai.
(book) (table) (on) (put) (imperative)

(As for the book put it on the table.)

The book is certainly presupposed situationally, and it may be legitimate to assume that the instruction is temporarily stored in the primary memory under the list head "the book," even though the listener is urged by the "imperative sentence" to fetch its predicate and "do" it right away.

Kuno (1970) has argued that there are two distinct uses of wa, viz. "theme" and "contrast." The uses of wa in (1-2), (1-3), and (1-4) may be considered as examples of the theme in his sense. In the following examples, however, it does not seem quite clear which mode of use of the particle wa is intended. As a continuation of the announcements (1-1) etc., we may have:

(1-a+1) JAL-531 wa teezi ni syuppacu no yotee desu.
(on time) (schedule)

(JAL-531 is scheduled to depart on time.)

(1-a+2) NW-315 mo teezi ni syuppacu no yotee desu.

(1-a+3) PAN-153 wa 1-zikan okure-te syuppacu no yotee desu.
(hour) (delay)

It is clear to a native speaker of Japanese that the particle wa in the sentences above has its inherent meaning regardless of whether it is used as "contrast" or "theme." In terms of our model, the pertinent NP's exist in the primary memory at this stage of the discourse, and any predicate following a wa-NP will be placed under the proper list head.

It has been pointed out by some investigators that a sentence with a topic wa-NP at its head may not necessarily be derived simply by extraposing a constituent NP in the deep structure of the sentence. As a typical example, it is natural to say in Japanese,

(4) Sakana wa sake ga oisii.
(fish) (salmon) (good-tasting)

(As for fish the salmon is good-tasting.)

Regardless of how one derives this surface structure from its underlying structure (whatever that may be) we may assert that the head in our primary memory file structure will again be fish, and the description for fish (a generic term) is that the salmon is good, or better, the good kind is salmon. Here, too, we can sense some "presupposition" that our point of interest is selecting an item with a particular feature (i. e. good-tasting) among others in a given set specified by the "topic" term (fish), and "contrasting" this against other candidate topics (i. e. meat, fruit, etc.). The sense of contrast which is one of the inherent semantic features of the form wa, conforms with and is specialized by the context and, quite often by a particular syntactic construction like "A is a but B is not a. "

In connection with the use of the particle wa, we may also discuss briefly the other particle mo, which already has appeared in some of the sentences of our discourse examples, since a comparison of the two particles seems to point to some clearly delimited performative characteristics of wa.[8] Suppose we replace e. g. (1-a+2) by the following sentence using wa instead of mo:

(1-a+2') NW-315 wa teezi ni syuppacu no yotee desu.

This is obviously as grammatical as (1-a+1) or (1-a+3) as an independent sentence, but if the sequence (1-a 1), (1-a+2'), and (1-a+3) were used for an air terminal announcement, it would sound either erroneous or inadequately mechanical. In view of the file structure in our primary memory model, (1-a+2') would be perfectly all right in leading to a similar deep structure listing as (1-a+1) or (1-a+3) do, and their resultant listing seems to be satisfactory as the stored information in the primary memory. Why is (1-a+2') unacceptable in Japanese? We may interpret the effect of the use of mo in this case as follows. In comparing (1-a+2) with (1-a+1) it is clear that the predicate as a sister constituent of the dummy NP (head) in

Fig. 3 is identical for these sentences. Thus the sentence pointer for these two cases giving two list heads will contain the same identification number, saving the creation of another predicate structure in the file upon receipt of (1-a+2).[9] The point is, however, this is not just a matter of economy.

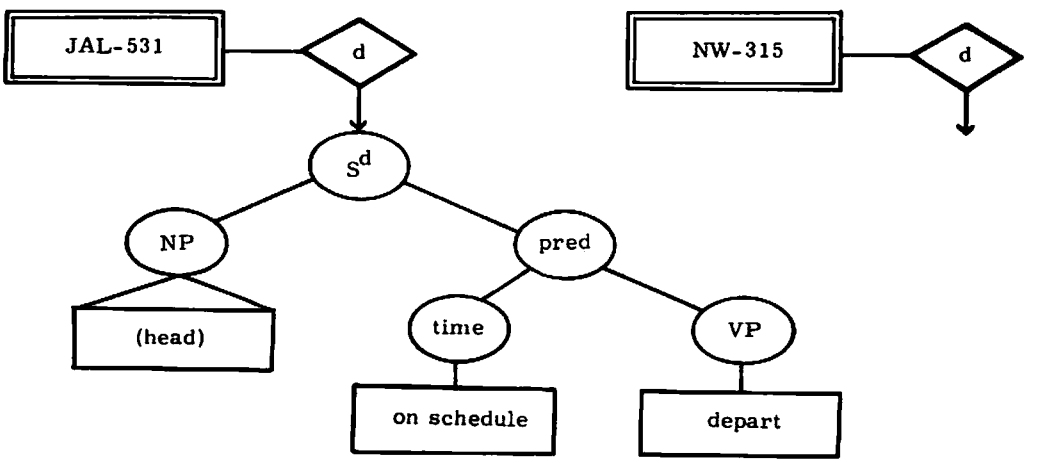

Fig. 3. The list structure for (1-a+1) and (1-a+2).

Human beings are destined to make mistakes. Typical errors are, for example, the misuse of negation, the wrong selection among semantically (or phonologically) similar items with the same syntactic features, etc. Sentence (1-a+2') could well be taken as an error of this sort; the predicate in this sentence sounds like an inadvertent copying of the preceding sentence, unless the speaker's intention of this duplication is explicitly marked, according to an apparently universal linguistic convention. The particle mo in (1-a+2) serves exactly as this redundancy marker, specifying explicitly that the listener must have in his primary memory the same structure

except for an NP with this article. The NP itself also exists as a primary memory item, as in the case of wa-NP's. Our primary memory model simply exploits the existence of the pertinent marker in a straightforward manner, viz, in the form of cross reference to the same predicate through the identical sentences pointers attached to the different heads. It is in these senses that in (1-a+2) the predicate and the subject NP are "presupposed." The focus of (1-a+2) is the syntactic combination of the subject and the predicate, rather than the introduction of any of the constituents themselves. The use of wa in (1-a+2'), in contrast, requires that some "focus" term(s), which contrasts in some way with the preceding predicate, be contained in a new predicate.[10] Similarly, if in (1-5) the particle mo were replaced by wa, the listener would assume that there was an error in the phrase "Alaska-keeyu New York-yuki de" either in (1-4) or in (1-5). This would be particularly true even if the second predicate concerning the departure time were identical for (1-4) and (1-5).

Explicit mention of presupposition by special markers as we observed above is seen also in a wide variety of other adverbials in Japanese, and obviously in other languages, too. The words even and also in English, as treated by Kuroda in connection with his "attachment transformation" (1965, 1969, 1970), can be considered as such. In Japanese, the uses of sae (or made), on the one hand, and gurai, on the other, may be noted as another set of typical cases. Let us consider the following sentence examples:

(5-a) John sae (mo) { itta yo. (Even John went, I'm telling you!)
ikanakatta yo. (Even John didn't go, I'm telling you!)

(5-b) John gurai (wa) iku yo. (At least John will go, I'm sure!)

There is a scale of probability of John's going (or not going), either in the past or future. In the case of sae, the presupposed probability is low for John (not) to go there, whereas it is presupposed at the same time that many others with higher probabilities of (not) doing so actually did (not) do so. In the case of gurai, on the contrary, it is presupposed that there is a high probability of John's doing it, while the probability is not high for most

other people. The use of these syntactic devices in Japanese do not pertain to the logical structure of the predicate "John went" or "John didn't go," which is of immediate concern for the primary memory discussed here. Rather, it creates redundancy and is used in the "checking" procedure when the statement is input to the memory, and the attached information pertains to other related and similar sentence structures. In this case, unlike the mo-NP, it is probably not appropriate to assume that the referred structures are available in the primary memory.

It may also be mentioned that the particle pair wa and mo can have a semantic function somewhat different from what we have observed above, and in a sense rather similar to the adverbials in (5-a, b). For example,

(6-a) Kono BALL wa 100-yen wa suru yo. (This ball costs at least 100-yen, I'm sure.)

(6-b) Kono BALL ga 100-yen mo
- sita yo. (Just this ball cost as much as 100-yen; it surprised me.)
- suru no ka. (Just this ball costs as much as 100-yen; it surprises me.)

The second adverbial particle wa and mo in these examples seem to have inherent semantic values which are closely related to their meanings discussed above. In (6-a), 100 yen is understood to be within the expected price range, and taking this value as a presupposed value established between the speaker and the listener, the speaker emphasizes that he expects this much as the minimal price, and that the true price is probably higher. In (6-b), the presupposition is that the ball cost(s) something no doubt, but the speaker emphasizes that the actual price was (is) higher than what he had expected. In other words, not only is the presupposition proved true, but also even a high price of 100-yen is proved true.

In the two sentences above, the second NP apparently does not serve as a topic, i.e., the list head status is not given to it in our model, whether it is with wa or mo. It may be noted also that in (6-a), the second wa-NP may almost be considered as a case of "contrast" in the sense of Kuno, but

not quite. Thus it does not necessarily imply something like

(6-a') 100-yen wa suru keredo 200-yen wa sinai.

(It costs at least 100-yen, but not as much as 200-yen.)

but perhaps more typically implies:

(6-a'') 100-yen wa suru yo, zissai ikura ka (wa) siranai keredo.

(It costs at least 100-yen, though I don't know how much it actually is.)

We may derive both (6-a') and (6-a'') from an "abstract contrast, "

$$\text{(100-yen suru) (yes-CONTRAST-no)} \begin{cases} \text{200-yen suru.} \\ \text{ikura ka sitteru.} \\ \text{(how much)(I know)} \end{cases}$$

and derive (6-a) from a similar incomplete contrast with the second term unspecified. The particle wa always contains at least this sense of incomplete contrast. Even in its typical use in "theme, " as in Kuno (ibid.), one can point out a similar implicit contrast, whether it is strong or weak. This inherent semantic value of wa is quite consistent with (but not necessarly accompanied by) the semantic function of topicalization. The letter, if it is intended, must be always implemented by the NP-transposition to the head of the sentence. The wa-NP at the head (surface subject position) of a sentence, therefore, is always a list head specified as such by the speaker, whereas a ga-NP at the same position is not to be treated as a list head.

A second appearance of wa in a sentence, on the other hand, generally cannot be handled by this model, as we have seen above. Thus the list head treatment may be legitimately considered as a machine process proposed for handling the "topic" of a sentence, even though this term seems a little misleading, because what it does is merely a specification of the entry under which the statement is to be temporarily stored, a sort of activation of the term in the primary memory for facilitating the subsequent informa-

tion processing, not particularly calling the listener's attention to a central issue of the discourse extending beyond the domain of the sentence.[11]

Notes

1) A paper submitted for publication in the Proceedings of the International Meeting on Computational Linguistics, Debrecen, Hungary, September 1971. It is a substantially extended version of a paper presented by H. Takahasi at the Survey Seminar in Computational Linguistics, US-Japan Scientific Cooperation Program, Hawaii, 1967.

2) Information Science Laboratory, Faculty of Science, University of Tokyo, Hongo, Tokyo.

3) We here postulate that the speaker assumes a similar structure in the listener's primary memory as the one he forms in his own.

4) A similar function is observed in the use of the definite article the in English. It may be noted in this connection that a "known" NP is not necessarily given the particle wa. Known NP's without wa tend to be modified by demonstrative pronouns like sono. Thus we can say,

(7-1) Watasi wa kinoo gakkoo ni hon o wasure masita.
(I) (yesterday) (school)(in)(book) (leave) (cordial past)

(7-2) Hon wa kyoositu ni ari masita.
(classroom) (exist)

In (7-2) hon appears with wa and without sono. We can also have (7-2') in place of (7-2) employing sono but without wa, when we do not topicalize this NP hon.

(7-2') Sono hon ga kyoosicu ni ari masita.
(that)

However, we cannot replace (7-2) by the following sentence, which is perfectly appropriate in some other contexts:

(7-2") Hon ga kyoosicu ni ari masita.

If we do use (7-2") in place of (7-2'), then the NP hon means another book, or more exactly "a book" which at that point of the conversation does not refer to the book in question and the sentence has to be supplemented by some kind of explanatory statement such as "But it was an entirely different book."

5) In this kind of use of the particle ga, the semantic structure of the sentence (or even its underlying syntactic form) may be a tree with the presupposed predicate as the subject NP with wa and with the NP in question as the focus NP in the form of copula complement.

6) Kuno (1970) assumes that there are only two cases of NP's that can be a theme (or topic in our terms), viz. either "anaphoric" or "generic." Since (2-2) does not appear to represent either of these cases, we would have to assume that (2-2) represents a case of contrast in accordance with the assertion that a wa-NP always represents either theme or contrast. This interpretation does not seem appropriate, however, See the next paragraph in the text.

7) It may be mentioned, however, that this use of wa in calling people's attention to an announcement is substantially different from the use of vocatives which are not accompanied by wa. In paging a person in a similar environment, therefore, the particle wa will never be used with the name of the passenger.

8) There is another case of mo which may be paired with ga rather than wa (e. g. in response to a wh-question, such a mo-sentence may follow some ga-sentence(s) in a row.) We will not discuss this problem in this paper.

9) It may be noted that a sentence can contain two independent predicates, and possibly only one of them is under the effectual domain of mo. In the case of (1-5), this separation of two predicates for the same list head, and

consequently the creation of two sentence pointers, is explicitly stated by the use of pronoun kore as the subject for the second predicate. In the actual use of mo in daily conversation, cases may be observed with less explicit structure in this respect. Thus, for example, a mother may talk about her children's absence from home and say,

(8-1) Ue no ko wa gakko e itte masu.
(older) (child) (school) (to) (be gone)

(8-2) Sita no ko mo yoocien e itte masu.
(younger) (child) (kindergarten)

This kind of partial identity as a qualification for the use of mo may be treated in our model by analyzing each of the sentences into two constituent predicates, one of which is identical for the two sentences. The partial identity, however, may be only semantic, and there may be no actual repetition of any identical noun or verb in the two sentences.

At the level of completeness of our model, we should probably just not worry about these "elliptic" cases. However, if we were to deal with these problems, which obviously require a semantic analysis of sentences (of course, with reference to the "secondary memory"), we might simply leave an indication for the sentence pointer in the primary memory which indexically interrelates two apparently different predicates.

10) We may mention here that the notion of presupposition formulated as an item existing in the primary memory is not quite appropriate for characterizing the distinction between wa and mo. Even when a structure is presupposed in the sense above, it does not qualify as the predicate to be associated with the mo-NP, unless this predicate is identical with the immediately preceding one. Thus we cannot have,

(9-1) John wa otoko da.
(man)

(9-2) Mary wa onna da.
(woman)

(9-3) Bill mo otoko da.

(9-4) Ann mo onna da.

The sentences (9-3) and (9-4) must have wa rather than mo, though the same predicates have appeared before. The use of mo requires not presupposition but exactly "prediction" of the predicate to be combined with the pertinent NP.

11) We may mention also that a topic in the usual sense for a discourse can be treated by a non-topic term in our sense using the particle ga. For example, at the head of a story we may find a sentence like:

Mukasi Yosicune to yuu o-samurai ga i masita.
(Once upon a time)(proper noun) (by name of) (samurai) (there was)

In this sentence from our syntactic point of view, the pertinent ga-NP is actually the focus of the sentence.

References

N. Chomsky (1965): Aspects of the Theory of Syntax, M. I. T. Press, Cambridge.

N. Chomsky (1970): "Deep Structure, Surface Structure, and Semantic Interpretation" Studies in General and Oriental Linguistics (R. Jokabson and S. Kawamoto, eds.) TEC Co., Ltd., Tokyo, 1970, 52-91.

S. Kuno (1970): "Notes on Japanese Grammar," Sciences of Language No. 2, 27-89.

S. -Y. Kuroda (1965): "Generative Grammatical Studies in the Japanese Language" (M. I. T. Dissertation).

S. -Y. Kuroda (1969): "On Attachment Transformations," Modern Studies in English: Readings in Transformational Grammar (eds. S. A. Schane and D. A. Reibel) Prentice-Hall, Englewood Cliffs, 331-351.

S. -Y. Kuroda (1970): "Remarks on the Notion of Subject with Reference to Words like Also, Even, or Only, Illustrating Certain Manners in Which Formal Systems Are Employed as Auxiliary Devices in Linguistic Descriptions, Part 2, Annual Bulletin (Research Institute of Logopedics and Phoniatrics, University of Tokyo) No. 4, 127-152.

For particularly relevant classical works concerning the notions of topic (theme) and presupposition in connection with the use of wa, see:

M. Kasuga (1918): Jinjōshōgaku Kokugo Tokuhon no Gohō Kenkyū (A Study of Japanese Usage in Primary School Readers).

A. Matsumura (1932): "Shukaku Hyōgen ni okeru Joshi 'Ga' to 'Wa' no Mondai" ("The Problem of the Particles ga and wa in Subject Phrases"), Gendai Nihongo no Kenkyu (A Study of Modern Japanese) (ed. Kokugo Shinko Kai), Hakusuisha, Tokyo.

D. Matsushita (1930): Hyōjun Nihon Kōgo Hō (Standard Grammar of Colloquial Japanese), Chubunkan, Tokyo.

A. Mikami (1964): Zō wa Hana ga Nagai (As for the Elephant, its Nose is Long), Kuroshio Shuppan, Tokyo.

K. Sakuma (1940): Gendai Nihongohō no Kenkyū (A Study of the Usage of Modern Japanese), Kōseisha-Kōseikaku, Tokyo.