A Criterion for Choosing Between Nominative Case Marker GA, Topic Marker WA, and Zero Pronouns in Japanese

A Criterion for choosing between Nominative Case Marker *GA*, Topic Marker *WA*, and Zero Pronouns in Japanese

Satoshi Imamura

Einar Andreas Helgason

Masatoshi Koizumi

1. Introduction

Japanese is a language that has a lot of options to mark the informational status of noun phrases. In particular, it has long been noted that the nominative case marker *ga* is used for new information and the topic marker *wa* is employed for given information (see e.g. Mikami, 1963; Kuno, 1972, 1973b; Ono, 1973). Moreover, it has been said that a zero pronoun is preferred when its referent is recoverable from context (Kuno, 1978) and is the unmarked option to express topic continuity in discourse (Hinds, 1978, 1982). Kuno (1972: 277) cites (1) in order to illustrate the usage of *ga* and *wa*. He points out that only the *wa*-marked subject *sono-gōtō* "the robber" is acceptable in (1b) because it has already been mentioned in (1a). If it were attached to *ga*, it would be unacceptable because *ga* marks new information, but *sono-gōtō* "the robber" is given information.

(1) a. gōtō-ga   boku-no-ie-ni   hait-ta
         robber-NOM   I-GEN-house-into   enter-PAST
         'A robber broke into my house.'
     b. sono-gōtō   *ga/wa   boku-ni-pisutoru-o   tsukitsukete
         the-robber   NOM/TOP   I-to-gun-ACC   point
         kane-o   da-se-to   it-ta
         money-ACC   give-IMP-QT   say-PAST
         'The robber, pointing a pistol at me, said "give me money."'

However, the given-new distinction is not sufficient to explain the differences among *wa*, *ga*, and zero pronouns because both *wa*-marked referents and zero pronouns are given information. What kinds of criteria, then, are necessary to distinguish them? In order to answer this question, we scrutinized the usage of *wa*, *ga*, and zero pronouns more deeply and objectively, using a quantitative approach. On the basis of quantitative data, we will propose that the usage of *ga*, *wa*, and zero pronouns largely depends on 'activeness', a term introduced by Chafe (1987, 1994). Specifically, we conclude that the nominative case marker *ga* marks inactive information, the topic marker *wa* is appropriate for semi-active information, and zero pronouns are suitable for active information. Based on this proposal, we hypothesize a constraint about the ordering of *ga*, *wa*, and zero pronouns: the more repeatedly a referent is mentioned, the more active it becomes, and an activated element cannot be referred to with a less activated form until it loses its activeness with time.

This paper is organized in the following way. Section 2 presents an overview of previous studies regarding *ga*, *wa*, and zero pronouns, Hinds and Hinds (1979) and Hinds (1983), and activeness mainly by Chafe (1987). Section 3 presents our discourse analysis about Referential distance based on written corpus data to see if Hinds (1983)' claims are valid. Section 4 provides a discourse analysis based on activeness by Chafe (1987, 1994) and discusses the relationship between *ga*, *wa*, and zero pronouns from the viewpoint of activeness. Section 5 is devoted to conclusions.

2. Previous Studies
2.1. *GA*, *WA*, and Zero Pronouns

As mentioned above, it has been said that the nominative case marker *ga* marks new information e.g. unpredictable information or what is not shared between the speaker and the hearer, and *wa* marks given information e.g. predictable information or what is known to both the speaker and the hearer. However, Kuno (1972: 270) points out that a *wa*-marked element is not necessarily anaphoric (i.e. previously mentioned) when it has a contrastive meaning. In other words, contrastive *wa* can mark both given information and new information. In fact, Miyagawa (1987) observed that thematic *wa* cannot follow a wh-phrase as in (2a) but contrastive *wa* can be attached to a wh-phrase as in (2b). Note that wh-phrases generally require new information and are not anaphoric because they have no specific referents. Thus, wh-phrases cannot be accompanied with thematic *wa*, which usually requires an anaphoric antecedent. However, there is no such constraint for contrastive *wa*.

(2)   a.   *dare-wa   ki-ta-no?
              who-TOP   come-PAST-Q
         b.   dare-wa   ki-te,   dare-wa   ko-nakat-tano?
              who-TOP   come-GER   who-TOP come-do not-PAST-Q

(Miyagawa 1987: 186)

Zero pronouns, which are unexpressed arguments of the verb, also belong to given information. Hinds (1978, 1982) argues that zero pronouns are chosen as the unmarked form of topic continuity. In other words, zero pronouns seem to be an option to express that the speaker continues to talk about the same topic in the following discourse. Kuno (1978: 8) observed that Japanese prefers to drop constituents when they are recoverable from the discourse context. Furthermore, Kuno (1978: 15-6) generalizes that information can only be omitted if all older or less important information is also dropped. (3a) presupposes *you went to Kanda yesterday* and the focus is *how* and *whether you went to Kanda on foot or not*. It is *on foot* that is newer or more important information than any other elements in (3a). In this context, (3c) is unacceptable because older or less important information *went* is left in but newer or more important information *on foot* is dropped. This example clearly shows that the usage of ellipsis depends heavily on pragmatic factors.

(3)   a.   Ø   kinō   Kanda-made   aruite   it-ta-ndesu-ka?
           (you)   yesterday   Kanda-to   on foot   go-PAST-AUX-Q
           'Did you go to Kanda on foot yesterday?'
      b.   Hai,   Ø   Ø   aruite-it-ta-ndesu
           yes   (I)   (there)   on foot-go-PAST-AUX
           'Yes, I went there on foot.'
      c.   *Hai,   Ø   Ø   ~~aruite~~-it-ta-ndesu
           yes   (I)   (there)   (on foot)-go-PAST-AUX
           (Same intended meaning as (3b))

(Kuno 1978: 35)

In sum, *ga* marks new information, and *wa* and zero pronouns are used for given information. However, contrastive *wa* is an exception to this rule because it can be employed for both given and new information. As for zero pronouns, they express the speaker's intention to talk about the same topic in the conversation and rely heavily on discourse context.

2.2. Hinds and Hinds (1979) and Hinds (1983)

Regarding the ordering of *ga*, *wa*, and zero pronouns, Hinds and Hinds (1979) argues that there are three steps in the identification of referents. On first mention, the nominative case marker *ga*, the dative marker *ni*, or the accusative marker *o* are used to introduce new referents. On second mention, they are marked with the topic marker *wa*. By doing this, the speaker indicates to the hearer what he or she is going to talk about. After that, they are referred to by zero pronouns repeatedly to indicate that the same referent remains the topic of the discourse. Therefore, the most natural information flow seems to be *ga*, *ni*, or *o* to *wa* to a zero pronoun.

Hinds (1983) calculated the Referential Distance (RD) of *ga*, *wa*, and zero pronouns based on Givón (1983) in order to measure the relative topicality of them. RD assesses the gap between a referent in the current clause and its antecedent. If there is no antecedent in the previous clauses, RD is assigned 20 because otherwise it will be infinite without limitation[1]. Hence, RD is expressed by the number of clauses from 1 to 20. Let us illustrate this concept with figure 1. Suppose that each square represents a clause and black squares include the same referent. In order to measure the RD of the referent in the black square furthest to the right (target clause), you need to go back to the middle black square. Since there are two clauses between the middle and the right, RD for the referent in the right black square is 3. Although the same referent is once mentioned in the black square furthest to the left, this has nothing to do with the RD of the referent in the black square furthest to the right. This is because RD is the value of the distance between the target referent and its nearest antecedent.
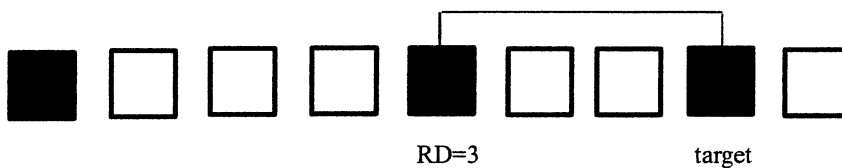

RD=3   target

Figure 1. Image illustrating Referential Distance

As a result of Hinds' analysis, *ga* demonstrated the least amount of topic continuity because RD for it was the largest, being consistent with its role as an indicator of new information. Zero pronouns were the most continuing option because RD for them was the smallest of all, showing their role as a continuation of the same topic in discourse. *Wa* was an intermediate category between *ga* and zero pronouns due to the intermediate value of RD.

Summing up, Hinds and Hinds (1979) pointed out the natural information flow from *ga* through *wa* to zero pronouns. Even though Hinds (1983) tried to validate this tendency and obtain objective data by calculating RD, his analysis is not sufficient because no statistical test was conducted for his data. In order to gain objectivity, it is necessary to confirm his hypothesis by conducting statistical tests. Moreover, even if his explanation is descriptively right, it is not explanatorily sufficient. In other words, he describes *how* but does not explain *why*.

2.3. Activation

Chafe (1987: 25) proposes three different states of activation: active, semi-active (or accessible), and inactive. An active concept is one that is currently lit up, a concept in a person's focus of

---

[1] The limitation of RD is rather arbitrary. For example, Givón (1994) proposed that it should be 3 and Cooreman (1992) suggested that it should be 15 because there was no example with RD higher than 15. However, we observed sentences with RD higher than 16, so we followed the criteria of Givón (1983).

consciousness at a particular moment. A semi-active concept is one that is in a person's peripheral consciousness, a concept of which a person has background awareness, but one that is not being directly focused on. Chafe (1994: 86) states that semi-active referents 'may be in the semi-active rather than the inactive state and thus be more accessible than new referents. It may be a referent that (a) was active at an earlier time in the discourse, (b) is directly associated with an idea that is or was active in the discourse, or (c) is associated with the nonlinguistic environment of the conversation and has for that reason been peripherally active but not directly focused on.' An inactive concept is one that is currently in a person's long-term memory, neither focally nor peripherally active. Generally speaking, active information refers to given information and inactive information is related to new information. Chafe's proposal is unique in that semi-active information is introduced as an intermediate type that exists between them.

Many studies have applied activation for their analyses. One example is Lee (2002), who tried to explain the mechanism for case-drop in Japanese. According to her, the nominative case marker *ga* cannot be dropped when a subject is inactive. However, *ga* can be dropped when a subject is semi-active, and both the subject and *ga* can be dropped when the subject is active. Her generalizations are demonstrated in (4).

(4)   When the information status of the subject NP-*ga* is:
        a. Inactive (new), no element can be deleted.
        b. Semi-active (accessible), *ga* deletion is possible.
        c. Predictable therefore Active (given), subject NP deletion is possible

(Lee 2002: 17)

Take Lee's examples to explain the validity of her generalizations. In (5a), the speaker knows that the listener's shoelace is untied and he is talking about it. Hence, *kutsu-no himo* 'shoelace' is part of the hearer's outfit and thus it is associated with the nonlinguistic environment of the conversation. This means that *kutsu-no himo* is semi-active and *ga* can be deleted. In (5b), the speaker sees a handkerchief that the hearer has dropped on the ground. Since *hankachi* 'handkerchief' is a part of the discourse, it is also semi-active information. Thus, the deletion of *ga* is natural again (see Lee, 2002 for more detail).

(5)   a.   Kutsu-no   himo-ga/Ø   hodokete-(i)ru-yo
              shoes   lace-nom   untied-is-s-f-p
              'Your shoelace is untied.'

b. Hankachi-ga/ Ø otimasita-yo.

  handkerchief-nom fell-s-f-p

  'You dropped your handkerchief.'

<div align="right">(Lee 2002: 23)</div>

To summarize, there are three activation states: inactive, semi-active, and active. Since activation consists of three concepts, it is more accurate for some analyses such as Lee (2002). We will also apply this concept for our analyses.

3. Discourse Analysis based on Referential Distance

Hinds (1983) observed that *ga*-marked elements had a larger RD than *wa*-marked elements, which in turn had a larger RD than zero pronouns in the Japanese folk tale *Momotaro*. However, his analysis is not objective in that no statistical test was conducted for his data. Hence, we will conduct a test to see whether his observation is on the right track or not. Moreover, we take theta-roles into consideration and control this factor by adding passive sentences to our analysis. By doing this, we can directly compare the RD of subjects to objects because both of them are patients.

3.1. Method

3.1.1. Corpus Data

In order to collect *ga*, *wa*, and zero pronoun examples, we employed *Aozora bunko*, which is a database of a collection of novels written in modern Japanese.

3.1.2. Materials

We assembled 12 types of sentences based on three factors[2]. The first factor is voice: (6a, b, c, d, e, f, g) are active voice and (6h, i, j, k, l) are passive voice. The second factor is word order: only (6f, g) are scrambled word order sentences and other examples are all canonical word order. The third factor is how patients are marked: they are marked with accusative case *o* (6a, b, c, f, g), nominative case *ga* (6h, j), topic marker *wa* (6i, k), and zero pronoun (6d, e, l). Note that all the target references correspond to direct object in active voice and they are underlined in (6).

(6) a. NP-ga/NOM <u>NP-o/ACC</u> V-PAST

    b. NP-wa/TOP <u>NP-o/ACC</u> V-PAST

    c. Ø   <u>NP-o/ACC</u> V-PAST

    d. NP-ga/NOM <u>Ø(zero)</u> V-PAST

---

[2] The reason why we choose them is that one aim of our study was to investigate the discourse function of scrambling and passives. However, this paper focuses on the particles.

e.	NP-wa/TOP	Ø(zero)	V-PAST
f.	NP-o/ACC	NP-ga/NOM	V-PAST
g.	NP-o/ACC	NP-wa/TOP	V-PAST
h.	NP-ga/NOM	NP-ni/DAT	V-PASS-PAST
i.	NP-wa/TOP	NP-ni/DAT	V-PASS-PAST
j.	NP-ga/NOM	NP-niyotte	V-PASS-PAST
k.	NP-wa/TOP	NP-niyotte	V-PASS-PAST
l.	Ø(zero)	NP-ni/DAT	V-PASS-PAST

3.1.3. Procedure

The RDs of target referents were divided into four categories (*o*, *ga*, *wa*, and zero pronoun) based on direct realization that does not take into account bridging relations between the preceding clause and the current clause (Grosz, Joshi, and Weinstein, 1995). In (7), for instance, *mado* 'window' is bridged with *heya* 'room' because *mado* is relevant to *heya*. However, *heya* is not a direct antecedent of *mado*, and hence the RD of *Mado* in (7b) is not 1.

(7) a . Kinō   pro   sono-heya-e   hait-ta.
         yesterday (I)   that-room-LOC   go.into-PAST
         'I went into that room yesterday.'
     b. Mado-ga   warete-i-ta.
         window-NOM   broken-be-PAST
         'The window was broken.'

The merit of direct realization is that this approach omits ambiguous examples. Since the scope of indirect realization is not clear, it is more desirable to deal with only clear examples.

3.1.4. Data Analysis

A series of t-tests were conducted to examine whether there were significant differences in RD between the accusative case marker *o*, the nominative case marker *ga*, the topic marker *wa*, and zero pronouns.

3.2. Results

The summary of the RDs and the number for each condition are shown in table 1.

Table 1. RD and Number for Target Referents

Types	RD		Number
	M	SD	
NP-*o*	10.14	8.84	2021
NP-*ga*	15.04	7.98	23
NP-*wa*	5.66	6.90	35
Ø	1.15	0.42	93

There are significant differences between *o* vs. *ga* (t(2042)=2.649, p<.01), *o* vs. *wa* (t(2054)=2.980, p<. 005), *o* vs. Ø (t(2112)=9.795, p<. 001), *ga* vs. *wa* (t(56)=4.763, p<. 001), *ga* vs. Ø (t(114)=16.922, p<. 001), and *wa* vs. Ø (t(126)=6.311, p<. 001). In other words, significant differences were observed between all conditions.

3.3. Discussion

The results have demonstrated statistically that Hinds (1983)' conclusion is valid. The nominative case marker *ga* has a significantly larger RD than the accusative case marker *o*. Moreover, *o* has a significantly larger RD than the topic marker *wa*, which in turn has a significantly larger RD than zero pronouns. Considering that *o* is the unmarked option for patient, *ga*-marked NPs are newer information than default accusative objects. This supports the idea that *ga* plays a role as an indicator of new information. Since zero pronouns are drastically lower than other options in RD, they tend to take over the same topic from the immediately preceding sentence. As Hinds observed, *wa* showed an intermediate value of RD between *ga* and zero pronouns. However, these results do not explain the *ga* to *wa* to zero pronouns ordering. We will explore this question in the next section.

4. Discourse Analysis based on Activeness

In order to explain Hinds and Hinds (1979), Hinds (1983), and the results of our analysis based on RD, we will propose a hypothesis based on activation by Chafe (1987, 1994). Specifically, we suppose that *ga* is inactive, *wa* is semi-active, and zero pronouns are active (see (8)). In addition, we hypothesize a one-way traffic information flow from less active to more or equally active (see (9)).

(8)　　Hypotheses 1

　　　　a. the nominative case marker *ga* is inactive

　　　　b. the topic marker *wa* is semi-active

　　　　c. zero pronouns are active

(9)   Hypothesis 2

The identification of NPs changes from less active to more (or equally) active. Therefore, inverse flow of identification from more active to less active violates this constraint.

If our hypotheses are on the right track, the possible patterns are the following: from *ga* (inactive) to *ga* (inactive), *ga* (inactive) to *wa*(semi-active), *ga*(inactive) to zero pronoun(active), *wa*(semi-active) to *wa*(semi-active), *wa*(semi-active) to zero pronoun(active), and zero pronoun(active) to zero pronoun(active). To put it the other way around, the antecedents of *ga* are nothing or *ga*, those of *wa* are nothing, *ga*, or *wa*, and those of zero pronouns are nothing, *ga*, *wa*, or a zero pronoun. We will investigate our predictions about how to mark NPs, which are summarized in (10).

(10)   Predictions about the possible antecedents of each condition:

a. *ga* (inactive): nothing or *ga*

b. *wa* (semi-active): nothing, *ga*, or *wa*

c. zero pronoun (active): nothing, *ga*, *wa*, or zero pronoun

4.1. Method

4.1.1. Corpus Data

Aozora bunko was employed again to assemble target referents.

4.1.2. Materials

We analyzed the immediately preceding antecedent of target references (*ga*-marked NPs, *wa*-marked NPs, and zero pronouns respectively).

4.1.3. Procedure

The forms of antecedents are divided into four categories: *ga*, *wa*, zero pronouns, and others. If there is no antecedent, they are categorized into "no antecedent". Therefore, we set five categories. Let us take an actual example from the corpus data. In (11), the antecedent of *Mito-wa* 'Mito' is the underlined full NP *Mito-kisha-ga*. Thus, this example is categorized into NP-TOP condition with an antecedent NP-NOM.

(11)

Ōkina nibui-oto-ga	okot-ta.	Suwat-to,	**Mito-kisha-ga**	yoko-o-miru-to,
think-sound-NOM	make-PAST	Oh-PTCL	Mito-reporter-NOM	side-to-look-and
Dorego-kisha-ga		yuka-ni	buttaoretei-ta.	"Ah,	ya-rare-ta?"
Dorego-reporter-NOM		floor-on	lie collapse-PAST.	Oh	knock out-PASS-PAST

Hortens-mo   sore-ni   kigatsui-ta.   Soshite   futari-no-kisha-wa

Horens-also   this-DAT   notice-PAST   then   two-GEN-reporter-TOP

Dorego-no-soba-ni   hizaotsui-ta.   Dorego-wa   chikaku-ga-nakat-ta.

Dorego-GEN-beside-LOC   kneel-PAST   Dorego-TOP   conscious-NOM-NEG-PAST

Mito-wa   ichijirushii-fuan-ni   torawa-re-ta.

Mito-TOP   extreme-anxiety-DAT   take-PASS-PAST

'A thick sound was made. Reporter Mito, surprised, looks to the side and Reporter Dorego lies collapsed on the floor. "Oh, he was knocked out?" said Hortens as he noticed. The two reporters then kneeled beside Dorego. Dorego was unconscious. Mito was taken by extreme anxiety.'

4.2. Results

Table 2. Distribution of antecedent forms of target NPs

Antecedent Forms	NP forms of Target References		
	NP-*ga*	NP-*wa*	Ø
Nothing	16	6	0
NP-*ga*	3	5	12
NP-*wa*	0	11	35
Ø	0	0	22
Others[3]	4	13	24

The distribution of antecedent forms of target references are summarized in table 2. Numbers indicate observed frequency. What we should note here is that there is no crucial counterexample to our predictions. All examples basically follow the patterns presented in (10). Firstly, in most cases the nominative case marker *ga* has no antecedent (16 examples) and some examples have a *ga*-marked antecedent (3 examples). Secondly, the antecedents of the topic marker *wa* are *ga* (5 examples), *wa* (11 examples), or nothing (6 examples). Note that there is no example with a zero pronoun as the immediate antecedent. This fact can be explained by the hypothesis that zero pronouns (active) are more active than *wa*-marked NPs (semi-active). Thirdly, the antecedents of zero pronouns are *ga* (12 examples), *wa* (35 examples), and zero pronouns (22 examples). What we should note here is that there are no examples of zero pronouns without an antecedent. This comes about as a natural result because readers cannot interpret zero pronouns without an explicitly mentioned antecedent.

---

[3] Others include *no, ni, de, mo*. and *o*. Since the purpose of this paper is to explore the relationships among *ga*, *wa*, and zero pronouns, others are out of the scope of our analysis.

4.3. Discussion

Our predictions have been clearly borne out because all actual examples followed them; supporting our hypotheses in (8) and (9). Therefore, we can conclude that the nominative case marker *ga* is inactive, the topic marker *wa* is semi-active, and zero pronouns are active. Moreover, the constraint that information flows from less active to more active is sufficient to explain the distribution of *ga*, *wa*, and zero pronouns. This also explains the ordering of *ga*, *wa*, and zero pronouns. First, *ga* cannot take *wa* or a zero pronoun as its antecedent because *ga* (inactive) is less active than both of them. Second, *wa* cannot take a zero pronoun as its antecedent because *wa* (semi-active) is less active than a zero pronoun (active). Third, a zero pronoun must have an antecedent which can be of any type. Taking Lee (2002) into consideration, it seems that *ga*-dropped NPs are as active as *wa*-marked NPs because both of them are semi-active referents. If this is true, it predicts that *ga*-dropped NPs cannot have zero pronouns as their antecedents but can take *ga*-marked NPs and *wa*-marked NPs as their antecedents. To put it the other way around, the identification of *ga*-dropped referents moves on to *wa*-marked NPs or zero pronouns. This is because *ga*-dropped NPs can take only semi-active or active referents. The predictions about the relationships between *ga*-marked NPs, *wa*-marked NPs, *ga*-dropped NPs, and zero pronouns are summarized in (20).

(20)   Predictions about the ordering of *ga*-marked NPs, *wa*-marked NPs, *ga*-dropped NPs, and zero pronouns:
   a. *ga*-marked NP → *ga*-marked NP, *wa*-marked NP, *ga*-dropped NP, or zero pronoun
   b. *wa*-marked NP → *wa*-marked NP, *ga*-dropped NP, or zero pronoun
   c. *ga*-dropped NP → *wa*-marked NP, *ga*-dropped NP, or zero pronoun
   d. zero pronoun   → zero pronoun

However, we did not observe any examples with a *ga*-dropped antecedent for wa-marked NPs or zero pronouns. Why did not we find *ga*-dropped antecedents in our data? One possibility is that our analysis depends largely on written Japanese, while it has been said that case drop is a common phenomenon in spoken Japanese (see e.g. Kuno, 1973a: 223). Therefore, it seems that our data are not appropriate for searching for *ga*-dropped NPs. Further studies are needed to confirm whether the predictions in (20) are correct or not by using spoken Japanese corpus data.

5. Conclusion

Using a statistical approach, we confirmed Hinds (1983)' observations that *ga* has a higher RD than *wa*, which in turn has a higher RD than zero pronouns. On the basis of Hinds and Hinds (1979) and Chafe (1987, 1994), we concluded that *ga* is inactive, *wa* is semi-active, and zero pronouns are active, and that identification of NPs changes from less active forms to more or equally active forms.

References

Chafe, W. (1987). Cognitive constraints and information flow. In R. Tomlin (ed.), *Coherence and grounding in discourse* (pp. 21-51). Amsterdam: John Benjamins.

Chafe, W. (1994). *Discourse, consciousness and time: The flow and displacement of conscious experience in speaking and writing.* Chicago: The University of Chicago Press.

Cooreman, A. (1992). The pragmatics of word order variation in Chamorro narrative text. In P. Doris (ed.), *Pragmatics of word order flexibility* (pp. 243-264). Oregon: John Benjamins.

Givón, T. (1983). Introduction. In T. Givón, (ed.), *Topic continuity in discourse* (pp. 5-41). Amsterdam: John Benjamins.

Givón, T. (1994). The pragmatics of de-transitive voice: Functional and typological aspects of inversion. In T. Givón (ed.), *Voice and inversion* (pp. 3-44). Amsterdam: John Benjamins.

Grosz, B., Joshi, A., K., & Weinstein, S. (1995). Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics*, 21, 203–26.

Hinds, J. (1978). Anaphora in Japanese conversation. In J. Hinds (ed.), *Anaphora in discourse.* Alberta, Canada: Linguistic Research, Inc.

Hinds, J. (1982). *Ellipsis in Japanese discourse.* Alberta, Canada: Linguistic Research, Inc.

Hinds, J. (1983). Topic continuity in Japanese. In T. Givón (ed.), *Topic continuity in discourse: A crosslinguistic quantitative study* (pp. 43-93). Amsterdam: John Benjamin.

Hinds, J. & Hinds, W. (1979). Where do we go from here? In D. Maisch, J. Hoard, & C. Sloat (eds.), *Proceedings of the eighth annual meeting of the western conference on linguistics.* Alberta, Canada: Linguistic Research, Inc.

Kuno, S. (1972). Functional sentence perspective: A case study from Japanese and English. *Linguistic Inquiry*, 3, 269-320.

Kuno, S. (1973a). *Nihon-Bunpō-Kenkyū* [A study of Japanese grammar]. Tokyo: Taishūkan.

Kuno, S. (1973b). *The Structure of the Japanese Language.* Cambridge, Mass: MIT Press.

Kuno, S. (1978). *Danwa-no Bunpō* [Grammar of Discourse]. Tokyo; Taishūkan.

Lee, K. (2002). Nominative case-marker deletion in spoken Japanese: An analysis from the perspective of information structure. *Journal of Pragmatics*, 34, 683-709.

Mikami, A. (1963). *Nihongo no ronri - wa to ga* [Logic of the Japanese Language]. Tokyo: Kuroshio Shuppan.

Miyagawa, S. (1987). 'WA and the Wh phrase', In J. Hinds et al. (eds), *Perspectives on topicalization: The case of Japanese wa* (pp. 185-217). Amsterdam: John Benjamins.

Ono, H. (1973). Japanese Grammar. Tokyo: The Hokuseido Press.

Satoshi Imamura, Graduate student, University of Oxford

Einar Andreas Helgason, Graduate student, Tohoku University

Masatoshi Koizumi, Associate professor, Tohoku University

日本語における主格の「ガ」、話題標識「ハ」、およびゼロ代名詞の選択基準について

今村怜

Einar Andreas HELGASON

小泉政利

主格の「ガ」、話題標識の「ハ」、およびゼロ代名詞の相互関係を談話機能の観点から計量的に分析した。その結果、「ガ」から「ハ」、「ハ」からゼロ代名詞という情報構造の流れを観察した。これは「ガ」が非活性、「ハ」が半活性、ゼロ代名詞が活性を示すと仮定することで説明がつくと主張する。非活性から活性への一方向的な流れと情報構造上の流れが一致し、助詞およびゼロ代名詞の選択が正しく予測できるとものと考えられる。