Antecedent Retrieval with Repeated Name Anaphors in Japanese: Topic and Subject

Shinichi Shoji

Organization for the Development of Higher Education and Regional Human Resources, Mie University, Tsu, Japan

Abstract—The present study tested whether repeated-name anaphors in Japanese elicit different effects in retrieving antecedents, depending on the anaphors being either a topic anaphor appended with the topic marker *wa* or a non-topic anaphor with the nominative marker *ga*. Early studies have shown that pronoun anaphors facilitate faster retrieval of their antecedents relative to repeated name anaphors, when the anaphors are grammatical subjects that refer to salient antecedents. In Japanese, however, grammatical subjects can be further classified into topic-subjects marked by *wa* and non-topic subjects marked by *ga*. Therefore, different antecedent-retrieval patterns may be possible between topic-*wa* and topic-*ga* even when they both were repeated-name anaphors. In addition, the present study investigated this issue with native English speakers who were learners of Japanese. Because their first language, English, does not overtly mark an entity as topic or non-topic, it was predicted that they might be relatively insensitive to anaphors' topic-hood and may not show different effects between topic-*wa* and non-topic-*ga*. A self-paced reading experiment showed that native Japanese speakers retrieved antecedents faster for repeated-name topic-*wa* anaphors than for non-topic-*ga*. On the other hand, native English speakers showed only marginally faster retrieval of antecedents for topic anaphors.

Index Terms-antecedent, anaphor, topic, Japanese

I. INTRODUCTION

In resolving referential relationship between anaphors and antecedents, an anaphor functions as a cue to retrieve the antecedent. The reference resolution has been examined by several studies in the field of psycholinguistics. Gernbacher's (1989) experiment is one of the studies that examined the different forms of anaphors, which are unequally efficient for retrieving antecedents. Her word-by-word timed-reading experiment showed that repeated name anaphors elicited faster antecedent-retrieval than pronouns, in sentences such as the ones below.

(1) a. Bill_i handed John some tickets to a concert but he_i took the tickets back immediately.

b. Bill_i handed John some tickets to a concert but Bill_i took the tickets back immediately.

(Gernsbacher, 1989, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4467536/)

Gernsbacher utilized probe recognition task, using antecedents as probe words, i.e., *Bill* in (1). The participants were presented with the probe words and asked to rapidly verify whether the probe words appeared in the sentences that they were reading. The results showed that, when the probe words were shown to the participants after they read anaphors, they responded faster when the anaphors were repeated names than when they were pronouns. Gernsbacher explained that the information that repeated name anaphors contain fully matches with that of antecedents, and thus repeated name anaphors facilitate faster reference resolution than pronoun anaphors.

However, Gordon, Grosz and Gilliom (1993) pointed out a methodological problem in Gernsbacher's experiment. In her experiment, because probe words (i.e., antecedents) were the same as the repeated name anaphors, what was retrieved from the probe words were possibly the anaphors, not the antecedents. Thus, Gernsbacher's results might have not reflected the speed of antecedent-retrieval. Gordon et al.'s self-paced sentence-by-sentence reading experiment elicited opposite results, which showed a faster reference resolution with pronoun anaphors than with repeated name anaphors when their antecedents were salient entities, e.g., grammatical subjects or first-mentioned entities in the sentences. This relative advantage of pronoun anaphors was observed in sentences such as the followings.

(2) a. Bruno_i chased Tommy all the way home from school one day. He_i watched Tommy hide behind a big tree and start to cry.

b. $Bruno_i$ chased Tommy all the way home from school one day. $Bruno_i$ watched Tommy hide behind a big tree and start to cry.

(Gordon et al., 1993, p. 318, with modification)

Sentences with a pronoun anaphor like the one in (2a) were processed faster than those with a repeated name anaphor like (2b).¹ The relative faster processing time of sentences with pronoun anaphors indicates the faster retrieval of antecedents. According to Gordon and Hendrick (1998), the different processing time occurs because, unlike pronouns,

¹ The difference of the processing times is not merely a reflection of the difference in phonological or orthographic lengths between pronouns and repeated names. If the outcomes were a reflection of the length-difference, then the processing-time difference should be uniform regardless of the salience of antecedent. Actually, when antecedents are not salient, there is no difference in their processing times.

readers of a repeated name anaphor do not initially interpret it as an anaphor. They first build an independent representation of the repeated name, and only then they realize that it refers to an antecedent, resulting in its slower retrieval. This advantage of pronoun anaphors over repeated name anaphors for antecedent-retrieval when antecedents are salient seems to be held as a general consensus among linguists (Arnold, 1998).

It is important to note that Gordon, Grosz and Gilliom's (1993) study observed the slower processing of repeated name anaphors only when the anaphors were grammatical subjects (e.g., 'Bruno chased Tommy all the way home from school one day. *Bruno* watched Tommy hide behind a big tree and start to cry.'). In Japanese, however, grammatical subjects can be further classified into topic-subjects appended with the topic marker *wa* (e.g., *Bruno-wa ookina ki-no ushiro-ni kakureru Tommy-o mita.* 'Bruno watched Tommy hide behind a big tree.') and mere grammatical subjects appended with the nominative marker *ga* (e.g., *Bruno-ga ookina ki-no ushiro-ni kakureru Tommy-o mita.* 'Bruno watched Tommy hide behind a big tree.') and mere grammatical subjects appended with the nominative marker *ga* (e.g., *Bruno-ga ookina ki-no ushiro-ni kakureru Tommy-o mita.* 'Bruno watched Tommy hide behind a big tree.'). The present study reports on an experiment that examined whether either subject anaphors or topic-subject anaphors would show a comparative advantage for antecedent-retrieval when both anaphors are repeated names. Furthermore, while the experiment examined this issue with native Japanese-speaking participants, the same experiment was also conducted with native English speakers who were learners of Japanese as a second language. It tested whether native English speakers would react similarly to native Japanese speakers, while their first language, English, does not overtly differentiate topic-subjects and non-topic subjects.

II. ANAPHORS, TOPIC-WA AND SUBJECT-GA

In general, a topic noun phrase (NP) refers to an entity that the speaker and listener have already acknowledged (Chafe, 1987; Kuno, 1973). Prince (1978) argues that a speaker marks an entity as a topic when the listener already knows the entity. In response, the entity marked as a topic leads the listener to search for its referent in the preceding context, the ongoing situation, or the listener's long-term memory (Haviland & Clark, 1974). In short, a topic is essentially anaphoric, as Halliday (1967) argues: a topic is concerned with the relation of what is currently being said to what was said earlier in the discourse.

In Japanese, the topic marker wa is a morphological device that presents an overt NP as a topic. Because a topic-NP-wa is most frequently a grammatical subject (i.e., topic-subject) (Martin, 1975; Nishimura, 1989), it has been widely argued how different it is from a mere subject-NP with ga-marking (Kasuga, 1918; Matsushita, 1928, among others). Kuno (1973) maintains that, because topic-wa refers to already-known information, it can be anaphoric while subject-ga cannot, unless it exhaustively lists its antecedent. Noda (1996) argues that subject-ga could be anaphoric only when the overall sentence describes an unpredicted event or when the sentence with it begins a discontinuous context, at a change of scenes, etc. In other words, a subject-ga should not be an anaphor except in the above cases, and thus, readers of a repeated name marked by ga would not expect that it anaphorically refers to an antecedent, and they would quickly retrieve it.

Unlike Japanese, English does not utilize a morphological marking system. Thus, a topic-NP in English cannot be immediately understood as a topic, i.e., readers of a topic-NP without a topic marker would not immediately realize that the NP refers to an entity that was already presented in its preceding context. Recall that earlier studies, e.g., Gordon, Grosz and Gilliom (1993), found that pronouns are preferred to repeated names for retrieving antecedents because repeated names are not immediately interpreted as anaphors while pronouns are immediately interpreted as anaphors. Based on the above discussion that a topic is essentially an anaphor and that the morpheme *wa* marks a topic in Japanese, repeated-name-*wa* in Japanese seems close to English pronouns, while repeated names in English are likely equivalent to repeated-name-*ga* in Japanese.² If this argument is true, native Japanese speakers would retrieve antecedents faster when the antecedents are referred to by topic-subject anaphors with *wa* rather than by non-topic subject anaphor with *ga*. On the other hand, it is also possible that native English speakers who are learners of Japanese may not show this antecedent-retrieval difference if they interpret both repeated-name-*ga* and repeated-name-*wa* merely as repeated name anaphors as in English, which does not utilize morphological markings such as *wa* or *ga*.

In order to test the above possibilities, a self-paced reading experiment was conducted with repeated-name-*ga* anaphors and repeated-name-*wa* anaphors, participated in by both native Japanese speakers and native English speakers learning Japanese. While Gordon, Grosz and Gilliom's (1993) experiment in English showed the relative disadvantage of repeated name anaphors, the present study further examines the advantage/disadvantage of repeated-name topic-*wa* anaphors would elicit faster antecedent-retrieval than with the repeated-name non-topic subject-*ga*. However, native English speakers might not, or more weakly, show this advantage with topic-*wa* anaphors because their first language does not morphologically differentiate topics and non-topics.

III. EXPERIMENT

² Japanese null pronouns should be the most equivalent to English overt pronouns (Kameyama, 1985), and Japanese overt pronouns do not have their counterpart in English (Hashimoto, 1948; Kuroda, 1965; Shibatani, 1990).

A. Participants

24 native Japanese speakers and 11 native English speakers recruited at the University of South Carolina participated in the experiment. The native Japanese speakers were all from an ESL (English as the second language) school that belongs to the university. The native English speakers were all undergraduate students, who have taken four semesters of Japanese at the university.

B. Items

A self-paced sentence-by-sentence reading experiment was conducted, following Gordon, Grosz and Gilliom (1993). In the experiment, participants read two-sentence discourses, which consisted of the first sentence including an antecedent and the second sentence including an anaphor. The antecedents were all salient, namely, sentence-initial grammatical subjects. The anaphors were all repeated names, which were either topic-subject-*wa* or non-topic subject-*ga*. Participants answered a yes-no comprehension question after they read the discourses. Example item is shown below. Native Japanese speakers saw only the Japanese texts (underlined parts) in the actual experiment. (3)

First sentence:	太郎が	図書館に	<u>行った。</u>
	Taro-ga	toshokan-ni	itta.
	Taro-NOM	library-DIR	went
	'Taro went to a library	,	
Second sentence:	太郎は/が	英語を	勉強した。
	Taro-wa/ga	eego-o	benkyooshita.
	Taro-TOP/NOM	English-ACC	studied
	'Taro studied English.'	,	
Question(問題):	太郎は	図書館に	行きました
	Taro-wa	toshokan-ni	ikimashita
	Taro-TOP	library-DIR	went
	'Did Taro go to a libra	ry?'	

For native English-speaking participants, in order to ensure that they understand each Japanese word, English translations were presented after each content word, but not for functional words (i.e., markers). Also, comprehension questions were presented only in English. An example is shown below. The underlined parts show what the native English speakers saw in the actual experiment.

か。 ka. Q

(4)

First sentence:	太郎(Taro)が	図書館(library)に	行った(went)。
	Taro(Taro)-ga	toshokan(library)-ni	itta(went).
	Taro-NOM	library-DIR	went
	'Taro went to a library	.,	
Second sentence:	太郎(Taro)は/が	英語(English)を	<u>勉強した(studied)。</u>
	Taro(Taro)-wa/ga	eego(English)-o	benkyooshita(studied)
	Taro-TOP/NOM	English-ACC	studied
	'Taro studied English.	,	
Question:	'Did Taro go to a libra	ry?'	

30 different experimental items such as the one above were prepared: half of them included repeated-name topic-subject-*wa* anaphors in the second sentences, and the other half included repeated-name non-topic subject-*ga*. The 30 experimental items were mixed with 40 distractor items. Thus there were 70 items given to the participants in total. The items were presented in random order.

Unlike Gordon et al.'s study, reading times of the second sentences would not be a valid variable to assess the speed of resolving referential relationship (i.e., antecedent-retrieval) because the items in the present experiment for native English speakers were mixtures of Japanese texts and English translations of content words. Instead, accuracy rates of the comprehension questions and the response times for the comprehension questions were used as dependent variables in this study. The comprehension questions for the present study asked about the first sentence with antecedents, as shown in (3) and (4). They were similar to Gernsbacher's (1989) comprehension questions in her experiment, which asked about the first clause containing antecedents. For example, for her item shown in (1b) 'Bill_i handed John some tickets to a concert but Bill_i took the tickets back immediately', the comprehension question was 'Who handed someone some tickets?'. According to Gernsbacher, this type of question asking about the clauses with antecedents ensured that participants correctly identified the antecedents of repeated name anaphors. The author of the present study considers that the response times for the questions would reflect how fast participants retrieved antecedents. In addition, the accuracy rates for this kind of comprehension questions would reflect how successfully participants retrieved

antecedents. Thus, the accuracy rates and question-response times were used as the dependent variables to be analyzed in this study.³ Also, when analyzing the question-response times, the data from those with wrong answers were removed prior to the analysis, as incorrect answers may indicate that the participants did not accurately identified antecedents. Only the data of question-response times that reflected how long it took the participants to retrieve accurate antecedents was analyzed.

C. Procedure

The discourse items in the experiment were presented using E-Prime. Participants read the two-sentence discourses sentence-by-sentence, in a self-paced reading fashion, following Gordon, Grosz and Gilliom (1993). The experiment was carried out with each participant viewing the sentences on a computer. The participants first received the welcome message and instruction on the computer screen and proceeded to the practice block by hitting the space key. The practice block provided four practice questions to familiarize the participants with the task. After the participants finished the practice questions, they received the end-of-practice message, and they proceeded to the actual experiment by hitting the space key. In the practice block and actual experiment, the first sentence of each experimental discourse appeared after the fixation mark, "+". After participants read each discourse, a yes-no comprehension question was given, which could be answered by hitting the keys, "1 (yes)" or "2 (no)". After the comprehension question, the fixation "+" appeared, which was followed by the first sentence of the next discourse. The experimental and distractor items were given in random order. One session lasted approximately 20 minutes.

IV. RESULTS

A. Native Japanese Speakers

Native Japanese speakers' results of accuracy rates and response times for comprehension questions are summarized in the table below.

	INDEE I.				
C	CCURACY AND RESPONSE TIMES FOR COMPREHENSION QUESTIONS (NATIVE JAPANESE SPEAKERS)				
	Conditions	Accuracy % (SD) (SE)	Response time ms (SD) (SE)		
	Topic-subject-wa	90.00 (.30) (.016)	3756.28 (1731.38) (96.19)		
	Non-topic subject-ga	90.56 (.29) (.015)	4090.89 (2510.65) (139.05)		

TABLEI А

One-way ANOVA tests were performed for the accuracy rates and response times for the comprehension questions, comparing those between two conditions: repeated-name topic-subject-wa anaphors and repeated-name non-topic subject-ga anaphors. Participants' accuracy rates did not show a significant difference between the conditions [F(1, 718)]= .063, p = .802]. However, the question-response time data showed a significant difference between the conditions (at a significant level of p < .05 or lower). Analyzing the data after removing the responses with wrong answers, affecting 9.72% of the entire data, the result showed that comprehension questions for topic-wa items were responded to faster than those for non-topic-ga items [F(1, 648) = 3.908, p = .048]. This result indicates that participants retrieved the antecedents of repeated-name topic-subject-wa anaphors faster than those of repeated-name non-topic-ga. As predicted, participants were likely to have immediately interpreted the repeated-name-wa as an anaphor and quickly realized its antecedent, resulting in faster response times to comprehension questions that asked about antecedents. In contrast, repeated-name-ga was not initially realized as an anaphor, resulting in relative slower retrieval of antecedents, which was reflected in slower question-response times.⁴

B. Native English Speakers

The results of native English speakers' comprehension accuracies and question-response times are summarized in the table below.

TINTI

	TABLE II.					
ACCURACY AND RESPONSE TIMES FOR COMPREHENSION QUESTIONS (NATIVE ENGLISH SPEAKERS)						
	Conditions	Accuracy % (SD) (SE)	Response time ms (SD) (SE)			
	Topic-subject-wa	89.70 (.30) (.024)	2842.34 (1300.14) (106.87)			
	Non-topic subject-ga	93.94 (.24) (.019)	3225.81 (2086.28) (167.57)			

One-way ANOVA compared the accuracy rates and question-response times between the conditions. Similar to native Japanese speakers, the results of the accuracy rates did not show a significant difference between repeated-name-wa anaphors and repeated-name-ga anaphors [F(1, 328), p = .161]. On the other hand, the result of question-response times, excluding the data with wrong answers (affecting 8.18% of the data), showed a marginally significant difference

³ Gernsbacher's experiment would have possibly shown faster question-response times for pronoun anaphors than for repeated name anaphors, which indicates the advantage of pronouns over repeated names, unlike her results from probe recognition task.

Although reading times for the second sentences with anaphors were not a reliable variable for native English speakers' results, the author analyzed the reading-time data from native Japanese speakers, excluding the data with wrong answers for comprehension questions. The result showed no significant difference [F(1, 648) = .004, p = .952].

between the topic anaphors with wa and non-topic anaphors with ga [F(1, 301), p = .057]. This result is compatible with the prediction: native English speakers' speed of antecedent-retrieval did not significantly differ between the conditions with repeated-name topic anaphors and repeated-name non-topic anaphors, possibly due to the influence of their first language, which does not overtly differentiate between the two anaphors. However, their result was close to that for native Japanese speakers. The participating English speakers' sensitivity to wa and ga was only slightly weaker than native Japanese speakers.

V. DISCUSSION

The result of native Japanese speakers showed a significant difference in antecedent retrieval between two types of repeated-name anaphors. When asked about antecedents, they responded faster when the repeated-name anaphors were topic-*wa* than when they were non-topic-*ga*. The result may indicate that the topic marker *wa* signaled that the repeated name was a topic that referred to an antecedent. Thus, the participants immediately retrieved the antecedents after they saw the anaphor, and both the anaphor and antecedent were activated as the same entity in their working memory while they were reading the discourse. Accordingly, the participants quickly responded to comprehension questions when the questions asked about the antecedents. In contrast, the participants might have not initially realized the anaphor-antecedent relationship when the anaphors were non-topic-*ga*. Possibly, they shifted their attention away from antecedents. They might have been reminded of the antecedents only when they read comprehension questions. In sum, the test results from native Japanese speakers indicate that, unlike in English, there is a difference in retrieving antecedents of repeated name anaphors in Japanese, depending on whether the anaphors are repeated-name topics or repeated-name non-topics.

The results for native English speakers indicate that they may have been not sensitive to the markers, *wa* and *ga*, enough to elicit a significant difference in using them as cues to retrieve antecedents. They might have often interpreted both repeated-name topic-subject-*wa* and repeated-name non-topic-subject-*ga* as mere repeated name subjects. However, the fact that they showed the same trend as native Japanese speakers, i.e., marginally faster question-response times for topic anaphors than non-topic anaphors, may indicate that they are in an ongoing process of linguistically acquiring the difference between *wa* and *ga*. As mentioned, the participants were Japanese learners who have completed four semesters of Japanese language classes in a university. This implies further predictions: more advanced-level Japanese learners (e.g., after completing six semesters of Japanese classes) might show a significant difference between *repeated-name-ga*, while more elementary-level Japanese learners (e.g., after completing that native English speakers learning Japanese gradually acquire the anaphoric that native English speakers learning Japanese gradually acquire the anaphoric function of the topic marker *wa* and the sensitivity to the morphological markings in Japanese.

VI. LIMITATION

A limitation of this study is that the experimental items were different between conditions. The same items with only a difference between *wa* and *ga* would have more accurately assessed the possible different reference resolutions with topic anaphors and non-topic anaphors. A future experiment with a modification for this limitation should be conducted, and as also mentioned above, an experiment with different levels of Japanese learners would explore learners' acquisition of Japanese more comprehensively.

VII. CONCLUSION

This study exhibited that two types of repeated-name anaphors in Japanese showed different effects in antecedent-retrieval. Topic anaphors with *wa* facilitated a faster resolution of antecedents than non-topic anaphors with *ga* because of the function of the topic marker *wa* that overtly informed readers that the repeated names are anaphors and thus there should be the antecedents. Based on the results of this study, the author suggests that more experiments with languages that utilize morphological markings should be conducted. Exploration of the processing of anaphors appended with markers would allow us to more comprehensively understand reference resolution across different languages.

REFERENCES

- [1] Arnold, J. E. (1998). Reference form and discourse patterns. Ph.D. dissertation, Stanford University.
- [2] Chafe, W. (1987). Cognitive constraints on information flow. In R. Tomlin (ed.), *Coherence and grounding*. Amserdam: John Benjamins, 21-52.
- Gernsbacher, M. A. (1989). Mechanisms that improve referential access. Cognition 32.2, 99-156. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4467536/ (accessed 3/1/2017).
- [4] Gordon, P. C., B. J. Grosz & L. A. Gilliom. (1993). Pronouns, names, and the centering of attention in discourse. *Cognitive Science* 17, 311-347.

- [5] Gordon, P. C. & R. Hendrick. (1998). The representation and processing of coreference in discourse. *Cognitive Science* 22, 389-424.
- [6] Halliday, M. A. K. (1967). Notes on transitivity and theme in English. Journal of Linguistics 19, 377-417.
- [7] Hashimoto, S. (1948). Kokugoho kenkyu [Study of national language]. Tokyo, Japan. Iwanami Shoten.
- [8] Haviland, S. E., & H. H. Clark. (1974). What's new? Acquiring new information as a process in comprehension. *Journal of Verbal Learning and Verbal Behavior* 13, 512-521.
- [9] Kameyama, M. (1985). Zero anaphora: The case of Japanese. Ph.D. dissertation, Stanford University.
- [10] Kasuga, M. (1918). Jinjou shogaku kokugo dokuhon no gohou kenkyu [Linguistic study on elementary school textbook]. Tokyo, Japan. Shubunkan.
- [11] Kuno, S. (1973). Nihon bunpou kenkyuu [Study of Japanese grammar]. Tokyo, Japan: Taishukan Shoten.
- [12] Kuroda, S-Y. (1965). Generative grammatical studies in the Japanese language. Ph.D. dissertation, Massachusetts Institute of Technology.
- [13] Martin, S. (1975). A reference grammar of Japanese. New Haven, CT: Yale University Press.
- [14] Matsushita, D. (1928). Kaisen hyojun Nihongo bunpou [Re-selected standard Japanese grammar]. Tokyo, Japan: Kigensha.
- [15] Nishimura, M. (1989). The topic-comment construction in Japanese-English code-switching. World Englishes 8, 365–377.
- [16] Noda, H. (1996). Wa to ga [Wa and ga]. Tokyo, Japan: Kuroshio Publishers.
- [17] Prince, E. F. (1978). On the function of existential presupposition in discourse. In D. Farkas, W. Jacobson & C. Tudrys (eds.) Papers from the fourteenth regional meeting of the Chicago Linguistic Society. Chicago, IL: Chicago Linguistics Society, 362-376.
- [18] Shibatani, M. (1990). The Languages of Japan. Cambridge: Cambridge University Press.

Shinichi Shoji was born in Japan in 1975. He received a PhD in linguistics focusing on reference resolution in Japanese from the University of South Carolina, Columbia, SC, USA, in 2016, with his dissertation *The Repeated Name Penalty and the Overt Pronoun Penalty in Japanese*. Currently he is an Assistant Professor at Organization for the Development of Higher Education and Regional Human Resources, Mie University, Japan. He was previously an Instructor of Japanese at Winthrop University, Rock Hill, SC, USA, University of North Carolina at Charlotte, Charlotte, NC, USA, University of South Carolina, Columbia, SC, USA, and Clemson University, Clemson, SC, USA. Dr. Shoji is a recipient of The Hanako Ito Chaplin Memorial Award for Excellence in Japanese Language Teaching.