

Difficulty in Assessing Knowledge of Ambiguous Sentences

Tokiko Okuma

1. Introduction

In empirical studies of second language (L2) acquisition, various methods have been developed to assess people's knowledge of languages. Multiple-choice tasks are one of them and have been commonly used to investigate interpretations of sentences by native speakers as well as L2 speakers. This paper reports on puzzling results obtained in a series of multiple-choice tasks investigating interpretations of Japanese numeral classifiers by native Japanese speakers and discusses whether the tasks correctly tapped native Japanese speakers' knowledge. Following White, Bruhn-Garavito, Kawasaki, Pater, and Prévost (1997), I will argue that multiple-choice tasks could reflect preference rather than linguistic competence in interpreting ambiguous sentences; therefore, researchers should be careful in drawing conclusions based solely on them.

The rest of this paper is organized as follows: Section 2 presents examples of multiple-choice tasks, including that by Hirakawa (1990), and discusses their potential methodological limitations, as suggested by White et al. (1997). Section 3 turns to another example of multiple-choice tasks: that by Okuma (in press). Section 4 presents the follow-up experiments, which show contradictory results. Section 5 discusses the implications of the studies presented in the previous sections. It is argued that Okuma's puzzling results and her follow ups are attributable to a task effect and the lack of a valid interpretation of some numeral classifiers. Finally, section 6 concludes the paper, warning of the danger of drawing conclusions solely based on a multiple-choice task.

2. A multiple-choice task in second language acquisition

A multiple-choice task is an experimental method in which participants choose appropriate answers from a set of options when interpreting sentences. This task is extensively used to assess people's knowledge of first and second languages¹. The following parts of this section provide examples of multiple-choice tasks used in L2 studies, including that by Hirakawa (1990). Note that this section only reports the results of native speakers, not L2 speakers, in those studies because the present paper discusses whether this task correctly reflects people's knowledge of their mother tongue rather than their L2.

2.1. Hirakawa (1990)

Hirakawa (1990) investigated the acquisition of Binding Principle A by native Japanese speakers studying English as an L2. Binding Principle A suggests that reflexives need to be bound by their antecedents in language-specific domains. For example, both English and Japanese observe Binding Principle A but have different binding domains. English reflexives (e.g., *himself* and *herself*) need to be bound by their antecedents within the same clause (i.e., local binding). In contrast, the Japanese reflexive *zibun*, meaning “self,” can take antecedents in a different clause (i.e., long-distance binding). In other words, the binding domain of English reflexives is a subset of the Japanese counterparts. In order to see whether native Japanese speakers can acquire the binding domain of English reflexives, Hirakawa conducted a multiple-choice task, as shown in (1), to compare L2 speakers with native English speakers². In the task, participants read English sentences containing reflexives and chose the antecedents from five options, from (a) to (e). In example (1), native English speakers were expected to choose (b) *Bill* as the antecedent of *himself*. The results conformed to this expectation, as the native English participants (n = 20) chose (b)

99% of the time. Thus, the multiple-choice task by Hirakawa successfully assessed native English speakers' knowledge of unambiguous sentences in English, as seen in (1).

(1) (Two-clause sentence)

John said that Bill hit himself.

Q. Who does *himself* refer to?

- A. a. John
b. Bill
c. Either John or Bill
d. Someone else
e. Don't know

On the other hand, however, the multiple-choice task was not necessarily successful in assessing the knowledge of ambiguous sentences, as seen in (2). In sentence (2), the reflexive *himself* ambiguously takes either the subject (*Bob*) or the object (*Paul*) as its antecedent. Therefore, native English speakers are expected to choose (c) *either Bob or Paul* for the antecedent of *himself*. Nevertheless, the native English participants in Hirakawa's work chose (a) *Bob* 67% of the time and (b) *Paul* 21% of the time. They chose the target response, (c) *either Bob or Paul*, only 12% of the time. In other words, the participants mostly overlooked logically possible object antecedents and only chose subject antecedents, which they probably preferred. Thus, it seems that a multiple-choice task can reflect people's preferences rather than their knowledge (i.e., linguistic competence) of ambiguous sentences. Note that this failure does not adversely affect Hirakawa's conclusion because the main focus of her study was L2 speakers' interpretation of bi-clausal sentences like (1), not native English speakers' interpretation of mono-clausal sentences like (2). Nevertheless, the unexpected response by native English speakers in Hirakawa's study suggests some

difficulty inherent to assessing people's knowledge of ambiguous sentences. Similar examples of native speakers unexpectedly overlooking logically possible interpretations of ambiguous sentences in multiple-choice tasks were also reported by Thomas (1989).

(2) (One-clause sentence)

Bob talked to Paul about himself.

Q. Who does *himself* refer to?

- A. a. Bob
b. Paul
c. Either Bob or Paul
d. Someone else
e. Don't know

2.2. White et al. (1997)

In order to cover the potential methodological drawbacks of multiple-choice tasks that were found in Hirakawa's and Thomas's studies, White et al. (1997) adopted truth-value judgement tasks to investigate interpretations of the same linguistic property, Binding Principle A, in English. In the task, pairs consisting of a written context and a test sentence, like (3), were presented to the participants. All test sentences were grammatical, and the participants were instructed to judge whether the test sentence matched the context. For example, in (3), the context provides the interpretation in which the reflexive *herself* refers to *Susan*. If the participants chose *True*, it followed that they knew that English reflexives can take object antecedents. Conversely, if the participants chose *False*, it indicated that they had misunderstood that English reflexives cannot take object antecedents. Thus, in a truth-value judgment task, contexts determine interpretations of ambiguous sentences and participants

answer whether each interpretation is expressed by the test sentence or not. White et al. intended to prevent the participants from demonstrating their preference for a particular option in the task by not making them choose one from a list of options.

- (3) (Context) Susan wanted a job in a hospital. A nurse interviewed Susan for the job. The nurse asked Susan about her experience, her education, and whether she got on well with people.

(Test sentence) The nurse asked Susan about herself. True / False

The results in White et al.'s study yielded a significantly higher proportion of object antecedents for reflexives than previous studies, including Hirakawa's and Thomas's works. Seventeen out of the total 19 native English participants (89%) consistently accepted the object antecedents in mono-clausal sentences like the test sentence in (3). Given that it is widely assumed that English reflexives allow object antecedents, White et al. concluded that truth-value judgement tasks with written contexts provide more accurate assessments of native speakers' knowledge of reflexives than multiple-choice tasks. It should be noted that, interestingly, not all types of truth-value judgement task are successful in assessing linguistic competence. White et al. also conducted another type of truth-value judgement task in which contexts were provided by pictures and found that only 21% of the native English participants consistently accepted object antecedents for reflexives. In other words, the pictures in the truth-value judgement task failed to control preference, just like in multiple-choice tasks. This result highlights the difficulty of tapping people's knowledge in empirical studies.

3. Okuma (in press)

3.1. Linguistic property

Another study that showed puzzling results on interpreting ambiguous sentences by native speakers is Okuma (in press). She investigated interpretations of Japanese numerals followed by classifiers (henceforth, NCs). Japanese NCs occur in different positions in a sentence. In (4a), for example, the NC *futa-ri*, meaning “two people,” appears before the host noun *gakusei*, or “student(s),” which the NC modifies. In contrast, the NC in (4b) appears right after the host noun, and the NC in (4c) is separated from the host noun—namely, the NC is floating (Nakanishi, 2008).

(4)

a. Prenominal NCs

Yamada sensei-wa [futa-ri-no gakusei]-o sagashi-teiru.

Yamada professor-Top [two-CL-Gen student-Acc look for

‘Professor Yamada is looking for two (specific/non-specific) students.’

b. Postnominal NCs

Yamada sensei-wa [gakusei futa-ri]-o sagashi-teiru.

Yamada professor-Top [student two-CL]-Acc look for

‘Professor Yamada is looking for two (specific/*non-specific) students.’

c. Floating NCs

Yamada sensei-wa [gakusei-o] futa-ri sagashi-teiru.

Yamada professor-Top student-Acc two-CL look for

‘Professor Yamada is looking for two (*specific/non-specific) students.’

The three types of NC are also known for their interpretive differences, as shown in Table 1. Huang and Ochi (2014) suggested that prenominal NCs have either specific or non-specific interpretations. For example, the sentence in (4a) can be used when Professor Yamada is looking for two specific students in his class, Mori-kun and

Takeda-san. Similarly, (4a) can be used when Professor Yamada is looking for any two volunteer students to help with his project. Thus, (4a) is appropriate regardless of the specificity of the noun that the NC modifies. In contrast, postnominal NCs have only a specific interpretation, while floating NCs have only a non-specific interpretation.

Table 1. Interpretive differences among numerals followed by classifiers (NCs)

	Specific	Non-specific
Prenominal NCs	✓	✓
Postnominal NCs	✓	✗
Floating NCs	✗	✓

3.2. Methodology

Okuma focused on two types of NC, prenominal and floating NCs, which are shown in grey in Table 1, and investigated their interpretations by native Japanese speakers with no prior knowledge of linguistics. She conducted a multiple-choice task with 32 native Japanese speakers who were freshmen at a university in Japan and were not majoring in linguistics. The task consisted of combinations of a paragraph that provided either a specific or non-specific context and a test sentence, as shown in (5) and (6). All parts of the task, including paragraphs, were presented in Japanese to the participants, and (5) and (6) show the English translations. The participants completed each sentence by choosing either (i) the prenominal NC or (ii) the floating NC so that the sentence matched the meaning of the paragraph. They were instructed to choose both NCs when applicable. When neither option was appropriate, they chose (iii) “Both options are inappropriate.” When they could not judge, they chose (iv) “I don’t know.”

In the example (5), the paragraph provides a specific context, in which Professor

Yamada is looking for two students, Mori-kun and Takeda-san. The specific context is compatible with prenominal NCs, but not floating NCs. Therefore, native Japanese speakers were predicted to choose (i), not (ii).

(5) Stimuli Example (specific context, English translation)

Professor Yamada has an appointment to meet two students, Mori-kun and Takeda-san, at 1 pm today. It has already passed 1 pm, but the two students nevertheless haven't shown up, so Professor Yamada has gone to the student office to find them.

Test sentence: (✓ represents the target response)

Yamada sensei-wa [✓(i) futa-ri-no gakusei-o / (ii) gakusei-o futa-ri]

Yamada professor-Top [two-CL-Gen student-Acc / student-Acc two-CL]

sagashi-teiru.

look for

'Professor Yamada is looking for two students.'

(iii) Both options are inappropriate

(iv) I don't know

The paragraph in (6) provides another example of the stimuli. In contrast to (5), (6) provides a non-specific context, in which Mr. Hara is looking for any two students to help with his shop. The non-specific context is compatible with either prenominal or floating NCs. Therefore, native Japanese speakers were predicted to choose both options, (i) and (ii). The task consisted of six tokens of each stimuli type, (5) and (6), and six fillers. They were randomized and presented to the participants. Consequently, each participant judged 18 sentences in total.

(6) Stimuli Example (non-specific context, English translation)

Mr. Hara is the owner of a coffee shop. Recently, a university moved to his city; so, his shop has had more customers and he has become very busy. Therefore, he decided to look for students to help out in his shop and put an advertisement on the door of his shop. He wants to hire two students.

Test sentence: Same as (5)

Hara san-wa [✓ (i) futa-ri-no gakusei-o / ✓ (ii) gakusei-o futa-ri]

Hara Mr.-Top [two-CL-Gen student-Acc / student-Acc two-CL]

sagashi-teiru.

look for

‘Mr. Hara is looking for two students.’

(iii) Both options are inappropriate

(iv) I don’t know

3.3. Results

Figure 1 shows the group means of the frequency of choosing prenominal and floating NCs out of the total six tokens in the specific and non-specific contexts by the native Japanese speakers. Regarding the prenominal NCs, the native Japanese speakers chose the specific interpretation more often than the non-specific one (5.47 vs. 1.59, $t(31) = 16.1$, $p < 0.01$). In contrast, regarding the floating NCs, they chose the non-specific interpretation more often than the specific one (4.97 vs. 0.91, $t(31) = 16.5$, $p < 0.01$). These results show that native Japanese speakers make a clear distinction between prenominal NCs and floating NCs.

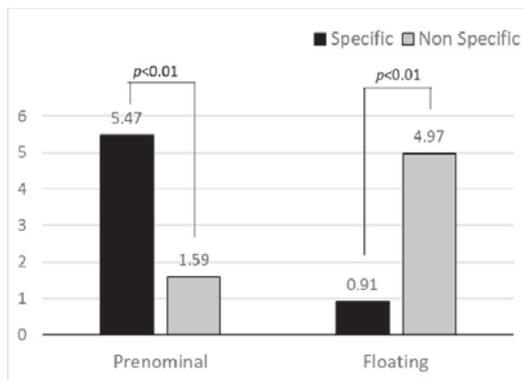


Figure 1. Acceptance of each interpretation of prenominal and floating NCs

Table 2. Interpretive differences between NCs

	Specific	Non-specific
Prenominal NCs	✓	✓
Floating NCs	✗	✓

However, one unexpected result was that the native Japanese speakers did not fully accept the non-specific interpretation of prenominal NCs. In fact, they chose the combination of the prenominal NCs and non-specific contexts only 1.59 out of the total 6 times (26.5%). This does not conform to the interpretation by Huang and Ochi (2014), which is presented in Table 2. If Huang and Ochi are right, the acceptance of prenominal NCs with non-specific interpretations should be above 3 out of 6 (i.e., 50%). However, it is not clear whether the low acceptance of prenominal NCs with non-specific interpretations truly reflects native Japanese speakers' knowledge. This is because, to the best of my knowledge, few studies have presented empirical data on the interpretation of NCs by native Japanese speakers. If the low acceptance of non-specific interpretations of prenominal NCs does not reflect the participants'

linguistic knowledge, it would be attributable to a task effect. In other words, the multiple-choice task by Okuma made the participants tend to choose only one option that they preferred, even though they were instructed to choose both options when applicable. This is parallel to Hirakawa, although Hirakawa and Okuma investigated different linguistic properties (i.e., reflexives and NCs) in different languages (i.e., English and Japanese).

4. Follow-up experiments

As we saw in the previous section, Okuma obtained a puzzling result: the native Japanese speakers unexpectedly did not fully accept the non-specific interpretation of prenominal NCs, and whether this truly reflects people's knowledge (i.e., linguistic competence) or is a task effect (i.e., performance) was left unexplained. To clarify this, two follow-up experiments were conducted for the present paper, extending Okuma's work. The aim was to examine whether native Japanese speakers make a specific/non-specific distinction between (i) prenominal NCs and postnominal NCs and between (ii) postnominal NCs and floating NCs, which Okuma did not investigate. The next section will report the results.

4.1. Methodology

Thirty-six native Japanese speakers participated in the experiments. They were all freshmen at a university in Japan and were not majoring in linguistics. None of them had participated in Okuma (in press). They were randomly divided into 2 groups, with a group consisting of 20 speakers taking part in follow-up experiment 1, and the other consisting of 16 speakers taking part in follow-up experiment 2.

The tasks were the same as those in Okuma's study, except different combinations of NCs were presented in the test sentences. As shown in Table 3, follow-up

experiment 1 contained postnominal and floating NCs in the test sentences. Similarly, follow-up experiment 2 contained prenominal and postnominal NCs in the test sentences. In both experiments, the participants read the context and completed the following test sentence so that the sentence matched the meaning of the context.

Table 3. Properties investigated in Okuma and the follow-ups

Study (sample size)	Combinations investigated (appropriate context)
Okuma (in press) (n = 32)	prenominal (either) vs. floating (non-specific)
Follow-up experiment 1 (n = 20)	postnominal (specific) vs. floating (non-specific)
Follow-up experiment 2 (n = 16)	prenominal (either) vs. postnominal (specific)

A stimuli example from follow-up experiment 1 is given in (7). The paragraph in (7) is the same as that presented in (5), which provides a specific context. In order to match the meaning of the following test sentence to the specific context, the test sentence needs to have the postnominal NC, not the floating NC. Therefore, native Japanese speakers were expected to choose option (i) *gakusei futa-ri-o*, “student two-CL-Acc” (i.e., the postnominal NC), not the floating option (ii) *gakusei-o futa-ri*, “student-Acc two-CL” (i.e., the floating NC). Each stimuli type consisted of six tokens.

(7) Stimuli Example (specific context)

Professor Yamada has an appointment to meet two students, Mori-kun and Takeda-san, at 1 pm today. It has already passed 1 pm, but the two students nevertheless haven’t shown up, so Professor Yamada has gone to the student office to find them.

Test sentence:

Yamada sensei-wa [✓ (i) *gakusei futa-ri-o* / (ii) *gakusei-o futa-ri*]

Yamada professor-Top [student two-CL-Acc / student-Acc two-CL]

sagashi-teiru.

look for

‘Professor Yamada is looking for two students.’

(iii) Both options are inappropriate

(iv) I don’t know

4.2. Results

4.2.1. Follow-up 1

Follow-up experiment 1 investigated interpretations of postnominal and floating NCs, which are summarized in Table 4.

Table 4. Interpretive differences between NCs

	Specific	Non-specific
Postnominal NCs	✓	✗
Floating NCs	✗	✓

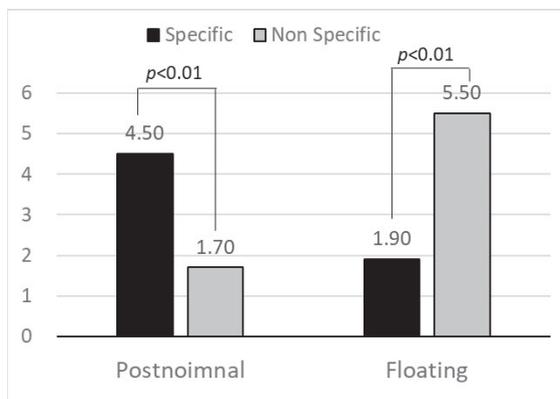


Figure 2. Acceptance of each interpretation of postnominal and floating NCs

Figure 2 shows the frequency the NCs were chosen in each context. Two findings were obtained. First, postnominal NCs allowed a specific interpretation more frequently than a non-specific interpretation and the difference was statistically significant (4.5 vs. 1.7, $t(19) = 6.98$, $p < 0.01$). Second, floating NCs allowed a non-specific interpretation more frequently than a specific interpretation and the difference between them was statistically significant (1.9 vs. 5.5, $t(19) = 10.73$, $p < 0.01$). These results are compatible with the interpretation in Table 5, which was suggested by Huang and Ochi. It follows that the task successfully reflected native speakers' linguistic competence regarding postnominal and floating NCs.

4.2.2. Follow-up 2

Follow-up experiment 2 investigated interpretations of prenominal NCs and floating NCs, which are summarized in Table 5. Figure 3 presents the responses by the native Japanese speakers.

Table 5. Interpretive differences between NCs

	Specific	Non-specific
Prenominal NCs	✓	✓
Postnominal NCs	✓	✗

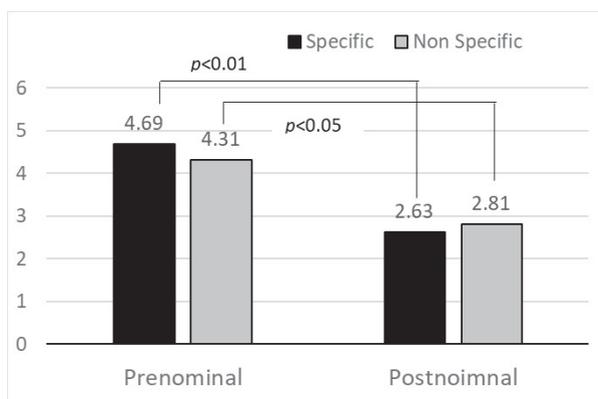


Figure 3. Acceptance of each interpretation of prenominal and postnominal NCs

Figure 3 shows two points. First, prenominal NCs allowed either specific or non-specific interpretations more than four times out of six. This is compatible with the interpretation in Table 5. Second, postnominal NCs unexpectedly allowed neither interpretation. The participants chose postnominal NCs only 2.63 and 2.81 times out of a total 6 times, numbers which were significantly smaller than the frequency of prenominal NCs in each context: 4.69 and 4.31 ($t(15) = 3.21, p < 0.01, t(15) = 2.16, p < 0.05$). These results on postnominal NCs are not compatible with Huang and Ochi's work. As shown in Table 4, postnominal NCs should have specific interpretations; consequently, their acceptance should be above the chance level of 3. Thus, the task was not successful in assessing native Japanese speakers' linguistic knowledge in

follow-up experiment 2, although it appeared to be successful in follow-up experiment 1. The next section discusses these puzzling results.

5. Discussion

5.1. Why were these puzzling results obtained?

This section discusses whether the multiple-choice task used in Okuma (in press) and the follow-up experiments successfully assessed the linguistic competence of native Japanese speakers. Table 6 repeats the results presented in Figures 1, 2, and 3.

Table 6. Mean frequency of NCs (out of 6) in Okuma and the follow-up experiments (* represents contexts that are not appropriate for the NCs)

(paired NCs)	Prenominal NCs		Postnominal NCs		Floating NCs	
	specific	non-specific	specific	*non-specific	*specific	non-specific
(prenominal)			2.63	2.81	0.91	4.97
(postnominal)	4.69	4.31			1.90	5.50
(floating)	5.47	1.59	4.50	1.70		

In Table 6, the columns in bold lines represent the group means of the frequency of the three types of NC (i.e., prenominal, postnominal, and floating NCs) in each context (i.e., specific and non-specific). The leftmost column in the table presents NCs that were presented in a pair with the three types of NC in the test sentence.

In Table 6, we can see that among the three types of NC, only floating NCs had consistent interpretations throughout the three experiments. In other words, floating NCs consistently allowed non-specific interpretations (i.e., they were chosen 4.97 and

5.5 times out of 6) and disallowed specific interpretations (i.e., they were chosen only 0.91 and 1.90 times) irrespective of the paired NCs. Thus, interpretations of floating NCs are stable. By contrast, the other two types of NC unexpectedly did not have consistent interpretations. The grey cells in Table 6 present the two unexpected results reported in Section 4. First, prenominal NCs did not have non-specific interpretations when they were presented in pairs with floating NCs (i.e., they were chosen only 1.59 times) although they did have non-specific interpretations when presented in pairs with postnominal NCs (i.e., they were chosen 4.31 times). I suggest that these conflicting results are caused by the task. In Japanese, prenominal NCs are ambiguous and can have either specific or non-specific interpretations, as Huang and Ochi suggested. However, paired floating NCs are not ambiguous, as they only have non-specific interpretations. As a result, native Japanese speakers prefer non-specific interpretations of floating NCs to those of prenominal NCs, and they tend to overlook the less preferred interpretations. This is attributable to the multiple choice task; therefore, if we adopt a truth-value judgment task in which the interpretations are checked one by one, the non-specific interpretations of prenominal NCs would be accepted more often than in the present study, just like the object antecedents for reflexives mentioned by White et al.

The second puzzling result is that postnominal NCs did not have clear interpretations when they were paired with prenominal NCs (i.e., they were chosen 2.63 times in specific contexts and 2.81 times in non-specific contexts). This result is unexpected because the linguistics literature suggests that postnominal NCs have specific interpretations and not non-specific interpretations. Moreover, they did have clear interpretations, as suggested in the literature, when they were paired with floating NCs (i.e., the specific interpretation of postnominal NCs was chosen 4.50 times). I speculate that the low acceptance of postnominal NCs with specific and non-specific interpretations was caused by the low frequency of postnominal NCs. In fact, Downing

(1996) analyzed 226 samples of NCs taken from written and oral sources as well as her original questionnaires, and she found that postnominal NCs accounted for only 6% of the samples. This is much smaller than other uses of NCs, such as prenominal NCs (47%) and floating NCs (42%). I suggest that the low frequency of postnominal NCs may prevent people who are not linguists from making clear interpretations of prenominal NCs; consequently, the participants were indeterminate in the experiment. By contrast, they were determinate when prenominal NCs were paired with floating NCs. This is because floating NCs are commonly used and have unambiguous interpretations, and that helped the participants realize that floating NCs and postnominal NCs have complementary distributions in terms of their interpretations. Thus, I suggest that the unexpectedly low acceptance of specific and non-specific interpretations of postnominal NCs was caused by their low frequency rather than a task effect. Therefore, even in a truth-value judgement task, native Japanese non-linguists would be indeterminate in interpreting postnominal NCs.

5.2. Are multiple-choice tasks successful?

As we saw in 5.1, a multiple-choice task can have two potential drawbacks. First, it can mask less-preferred interpretations. As found by Hirakawa and Okuma, people tend to overlook less-preferred options even when they are instructed to choose multiple options when applicable. In other words, not choosing some option does not mean the rejection of it. Therefore, believing just the results from a multiple-choice task could lead one to underestimate people's linguistic competence, as suggested by White et al. Second, non-linguists do not necessarily have clear interpretations of infrequent forms, such as postnominal NCs. As a result, their judgements on those forms can be unstable. This point is not specific to multiple-choice tasks but is probably applicable to other empirical methods as well. These two points highlight the

difficulty of tapping people's knowledge of languages in experiments. In order to assess people's linguistic competence without overestimating or underestimating it, researchers should be careful in drawing conclusions based on a single task, as White et al. pointed out.

6. Conclusion

In this paper, I discussed potential shortcomings of multiple-choice tasks as a tool to assess people's interpretation of ambiguous sentences. I presented L2 studies that showed puzzling results on sentence interpretation by native speakers, including works by Hirakawa and Okuma. In order to clarify the contradictory results obtained in the multiple-choice task used by Okuma, I conducted two follow-up experiments and showed that people's linguistic competence can be masked because of a task effect. As suggested by White et al., multiple-choice tasks could reflect preference rather than linguistic competence in interpreting ambiguous sentences. I also argued that interpretations of infrequent forms, such as postnominal NCs in Japanese, can be unstable because native Japanese non-linguists do not make clear interpretations. Based on these results, I concluded that researchers should be careful in drawing conclusions based solely on said tasks, which supports White et al.

Notes

1. See Shirahata, Wakabayashi, and Muranoi (2010) for an overview of data collection methods commonly used in L2 acquisition studies, although they do not specifically mention multiple-choice tasks, which are the focus of the present study.
2. Hirakawa also investigated another aspect of English reflexives, namely, the subject-orientation of antecedents. However, this aspect is not relevant to the present study and, accordingly, is not discussed.

Works Cited

- Downing, Pamela. *Numeral classifier systems: The case of Japanese*. Amsterdam/Philadelphia: John Benjamins, 1996.
- Huang, C.-T. James and Ochi, Masao. "Remarks on classifiers and nominal structure in East Asian." *Peaches and plums* (pp. 53–74). Ed. C.-T. James Huang and Feng-Hsi Liu. Taipei: Academia Sinica, 2014.
- Hirakawa, Makiko. "A study of the L2 acquisition of English reflexives." *Second Language Research* 6 (1), 60-85, 1990.
- Nakanishi, Kimiko. "The syntax and semantics of floating numeral quantifiers." *Oxford handbook of Japanese linguistics* (pp. 287–319). Ed. Shigeru Miyagawa and Mamoru Saito. Oxford: Oxford University Press, 2008
- Okuma, Tokiko. "L2 acquisition of the specificity of Japanese numeral quantifiers." *Proceedings of the 44th Annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, in press.
- Shirahata, Tomohiko, Wakabayashi, Shigenori, and Muranoi, Hitoshi. *Shousetsu dainigengoshuutoku kenkyuu* [A detailed explanation of second language acquisition studies]. Tokyo: Kenkyuusha, 2010.
- Thomas, Margaret. "The interpretation of English reflexive pronouns by non-native speakers." *Studies in Second Language Acquisition* 11(3), 281-301, 1989.
- White, Lydia, Bruhn-Garavito, Joyce, Kawasaki, Takako, Pater, Joe, and Prévost, Philippe. "The researcher gave the subject a test about himself: Problems of ambiguity and preference in the investigation of reflexive binding." *Language Learning* 47(1), 145-172, 1997.