Zero Copula in Japanese: Learners' Use and Attitude

Satoshi Nambu
Monash University

1   Introduction

As in the case of copula absence in African American Vernacular English (AAVE) (Labov 1969, Cukor-Avila 1999, Bender 2000, and many others), also known as zero copula, a copula in Japanese can be optionally absent, as shown in (1) and (2) (Abe, 2001, 2015; Tanomura 2008, Shioda 2010).[1]

(1)   *Copula omission*
        A: Kore(-wa)   nani(-des-u-ka)?
             this-TOP[2]   what-COP-PRS-Q
             'What's this?'
        B: Sore-wa   sakura-no   hana(-des-u).
             that-TOP   cherry-GEN   flower(-COP-PRS)
             'That's a cherry flower.'

(2)   *Zero copula*
        Sore-wa   hitsuyou(-da)-to   omo-u.
        it-TOP   necessary-COP.PRS-COMP   think-PRS
        '(I) think that it is necessary.'

The copula absence in a main clause in (1) is considered as a case of copula omission due to casualness of the speech (Yamaguchi 2007). On the other hand, in the case of the copula absence in a subordinate clause with the complementizer *to* in (2), Nambu (2019) provides empirical evidence that it is not directly linked to casualness of speech, and claims that it has a unique historical origin implemented in grammar, labeling it as 'zero copula'. Nambu (2019) regards overt and zero copula as a case of linguistic variation and examined the nature of zero copula from variationist sociolinguistic and historical perspectives, drawing on data from a questionnaire survey and a corpus search.[3] While the results of his questionnaire survey confirmed that the overt copula is regarded as the correct form in comparison to zero copula, his historical corpus data suggests that zero copula was the default form in the certain linguistic environments until the use of the overt copula started increasing in the 17th century. In order to explain the discrepancy between the norm for contemporary Japanese and the historical data, Nambu (2019) proposed a hypothesis that the change involves two stages as follows;

(3)   **Language change hypothesis on zero copula (Nambu 2019)**
        Stage 1: *Change from below*
        The change began as the emergence or more frequent use of an overt copula in the relevant linguistic environments, which is due to a change in grammar (triggered by another linguistic change).

        Stage 2: *Change from above*
        While the use of an overt copula has been increasing over time, the overt copula in those linguistic environments has gained its status in the norm at the expense of zero copula; thus, the new norm has now been pushing the ongoing usage shift further as a case of "change from above (the level of consciousness)".

---

[1] The copula *da* (or *desu*/*dearu*) follows a noun or appears as part of an adjectival noun (Tsujimura 2014).
[2] The abbreviations in the glosses are defined as follows: COMP = complementizer, COP = copula, GEN = genitive, NOM = nominative, POL = polite form, PRS = present, PST = past, Q = question particle, TOP = topic.
[3] See also Tanomura (2008) and Abe (2015) for discussions about zero copula using a corpus data.

In short, the change in the use of zero copula lasting for centuries until the present time stems from a change in grammar triggered by another linguistic change, thus considered as a case of *change from below* in variationist sociolinguistics (Labov 1990)[4]; it is then followed by a change in the norm for Japanese language that became a driving force for a further change, i.e., a case of *change from above*.

Based on the findings reported in Nambu (2019), this article investigates Japanese learners' use and attitude of zero copula in Japanese, drawing upon data from a corpus search and a questionnaire survey. From a perspective of second language learning, the aim of this preliminary study is (i) to find any difference in the use of zero copula between native Japanese speakers and learners of Japanese and (ii) to identify how much the norm for Japanese language is shared with learners; the findings from those then lead to a discussion about how learners acquire linguistic variants that are currently undergoing a change in progress. Although the data in this study is not quantitatively sufficient to lead to a firm conclusion, the general trend can still be discussed with implications for further research.

2 Learners' use of zero copula

2.1 *Corpus data*   To investigate how much learners of Japanese use zero copula, i.e., the copula absence in a subordinate clause followed by the complementizer *to*, I drew upon data from the *International corpus of Japanese as a second language* (I-JAS, Sakoda et al. 2016).[5] The I-JAS corpus consists of spoken and written language elicited from 560 learners of Japanese whose native languages are English, Chinese, Korean, Russian and others. The data in the corpus were elicited by multiple tasks such as interviews and storytelling. One of the important features of the corpus for the current study is that it includes scores on Japanese language tests, which makes it possible to examine a correlation between the Japanese language level and the use of zero copula. In addition, the corpus also includes data of 100 native Japanese speakers elicited in the same procedure, which can be used to compare the use by learners and native speakers. Using the online query available on the I-JAS corpus website[6], I extracted the use of overt copula (*da*, *dearu*) and zero copula followed by the complementizer *to* with verbs *omow-* 'think'.

Regarding errors by learners in the use of copula in general, omitting a copula in a linguistic environment where it must be overtly used is very common among learners, as exemplified below;

(4)   Sore-wa   hitsuyou*(-da)-kara,   kat-ta.
       it-TOP   necessary-COP.PRS-because   buy-PST
       '(I) bought it because it is necessary.'

This implies that it is difficult for learners to discern linguistic environments as to whether a copula can be absent in a particular environment. Therefore, being too cautious about omitting a copula (or using zero copula) may induce a case where learners use the overt copula more often than native speakers even in linguistic environments where zero copula is allowed, which may be viewed as a sort of *hypercorrection* (Labov 1966). That being said, since native speakers regard the overt copula as the correct form in comparison to zero copula, it is expected that more advanced level Japanese learners use overt copula more often than the others.

2.2 *Results*   Table 1 shows the use of overt/zero copula by learners and native speakers in the I-JAS corpus, indicating that learners use zero copula more frequently than native speakers. The result of the Pearson's chi-square test in Table 1 also supports the trend. For comparison, Table 2 is the data of native speakers from two corpora in Nambu (2019).

	Learners	Native speakers
Overt copula	1,055 (58.1%)	112 (88.1%)
Zero copula	762 (41.9%)	15 (11.9%)

X^2=43.65, *df*=1, *p*<.001

Table 1: Use of overt/zero copula by learners in I-JAS

The data of native speakers in Table 2 shows that the use of zero copula is much higher when the subordinate clause with the complementizer *to* is followed by the verb *omoware-* 'seem', which is a spontaneous form of

---

[4] Nambu (2019) argues that the diachronic change in the use of zero copula derives from a change in the function of *to* that follows zero copula.

[5] The data was obtained from the corpus in December 2018.

[6] https://chunagon.ninjal.ac.jp/static/ijas/about.html

omow- 'think', than with *omow-*. I will discuss this effect in Section 3. The data from the I-JAS corpus in Table 1 did not contain any examples with *omoware-*, and comparing the data in Table 1 with the data for *omow-* in Table 2 confirms that learners use zero copula more frequently than native speakers.

	Written Japanese (BCCWJ)		Spoken Japanese (CSJ)	
	omow-	*omoware-*	*omow-*	*omoware-*
Overt copula	20,721 (82.9%)	621 (21.8%)	2,434 (91.8%)	125 (61.3%)
Zero copula	4,281 (17.1%)	2,227 (78.2%)	218 (8.2%)	79 (38.7%)

X^2=5322, *df*=1, *p*<.001 $\quad\quad$ X^2=185.92, *df*=1, *p*<.001

Table 2: Use of overt/zero copula by native speakers in Nambu (2019)[7]

Table 3 shows frequencies of overt/zero copula by the type of preceding item, and the corresponding data of native speakers in Nambu (2019) is given in Table 4 for comparison specifically because the data of native speakers in Table 3 is sparse. Example sentences in the learners' data are given in (5).

	Learners		Native speakers	
	Adjectival noun	Noun	Adjectival noun	Noun
Overt copula	398 (63.2%)	657 (55.3%)	20 (95.2%)	92 (86.8%)
Zero copula	232 (36.8%)	530 (44.7%)	1 (4.8%)	14 (13.2%)

X^2=10.03, *df*=1, *p*<.01 $\quad\quad$ *p*=.46 (Fisher's test)

Table 3: Use of overt/zero copula and preceding item by learners in I-JAS

	Written Japanese (BCCWJ)		Spoken Japanese (CSJ)	
	Adjectival noun	Noun	Adjectival noun	Noun
Overt copula	3,283 (82.9%)	18,059 (75.6%)	314 (92.4%)	2,245 (89.2%)
Zero copula	679 (17.1%)	5,829 (24.4%)	26 (7.6%)	271 (10.8%)

X^2=277.66, *df*=1, *p*<.001 $\quad\quad$ X^2=2.81 *df*=1, *p*=.09

Table 4: Use of overt/zero copula and preceding item by native speakers in Nambu (2019)

(5)   Examples of learners' use in I-JAS
        a. *Adjectival noun + overt copula*
        Jikan-ga   taisetsu-da-to   omoi-mas-u.
        time-NOM   important-COP.PRS-COMP   think-POL-PRS
        'I think that time is important.'

        b. *Noun + overt copula*
        Sore-wa   ii   aidea-da-to   omoi-mas-u.
        it-TOP   good   idea-COP.PRS-COMP   think-POL-PRS
        'I think that it is a good idea.'

        c. *Adjectival noun + zero copula*
        Jikan-wa   juuyou-to   omoi-mas-u.
        time-TOP   important-COMP   think-POL-PRS
        'I think that time is important.'

        d. *Noun + zero copula*
        Aji-wa   futsuu-to   omoi-mas-u.
        taste-TOP   normal-COMP   think-POL-PRS
        'I think that the taste is normal.'

The data of learners and native speakers in Table 3 and 4 indicate the same trend in that both cohorts use zero copula more frequently with a noun than with an adjectival noun, although the difference in the spoken data of native speakers was not statistically significant (*p*=.09). This result implies that, in addition to the grammar rules as to where the linguistic variants in question can be used, learners of Japanese can acquire the knowledge of frequency as to how often they should be used in a certain linguistic environment.

---

[7] BCCWJ is *Balanced Corpus of Contemporary Written Japanese* that is a corpus of written Japanese (Maekawa et al. 2014), and CSJ is *Corpus of Spontaneous Japanese* that is a corpus of spoken Japanese (Maekawa 2004).

Figure 1 illustrates the relationship between the rate of overt copula over zero copula and the Japanese proficiency level.
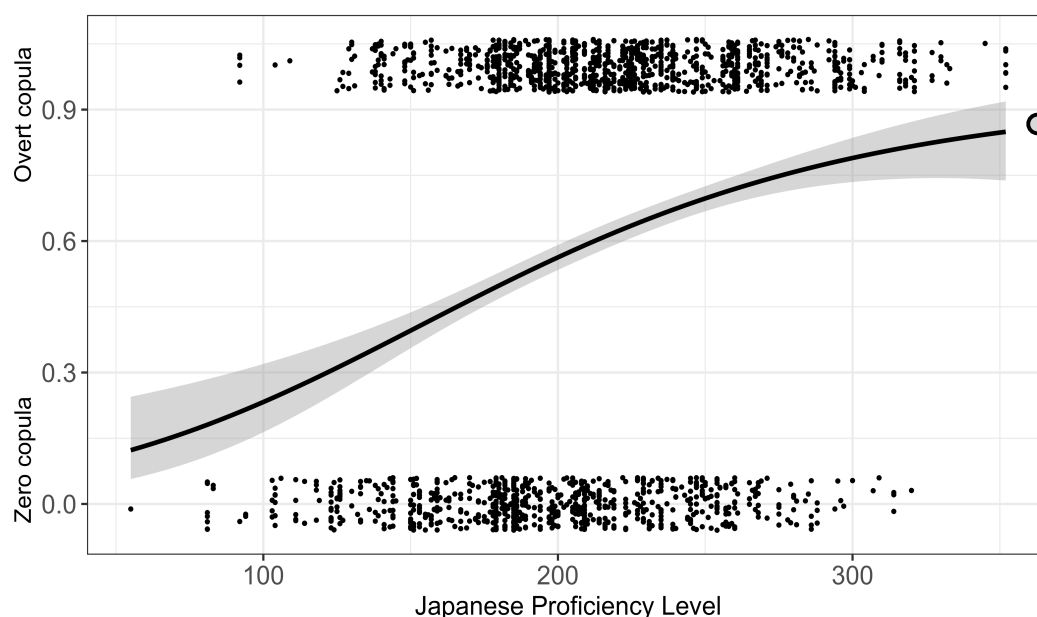

Figure 1: Rate of overt copula over zero copula and Japanese proficiency level in I-JAS
(The large dot at the right edge represents the rate by native Japanese speakers.)

The solid line and shaded area represent a regression line and a 95% confidence interval, respectively. The regression line indicates a strong correlation between the rate and the Japanese level; more advanced level learners use the overt copula more often than the others. The large dot at the right edge in Figure 1 represents the rate by native speakers, which suggests that the shift in the rate by the learners targets the rate by native speakers, i.e., learners of Japanese can acquire the knowledge of frequency regarding the use of linguistic variants by native speakers.

3 Learners' attitude

3.1 *Questionnaire survey* Using a questionnaire survey, Nambu (2019) investigated attitudes of native speakers and revealed that the overt copula is generally regarded as the correct form in comparison to zero copula. More noteworthy is the fact that, contrary to the general trend, they preferred zero copula to the overt copula when a predicate that takes the subordinate clause is *omoware-* 'seem' (the spontaneous form of *omow-* 'think'), which is congruent to the corpus data given in Table 2 in Section 2.2. On the ground that the spontaneous form *omoware-* is considered as more formal in style compared to *omow-*, the result represents that zero copula is preferred to be used when a given context is formal. Since this is opposite to the link between the copula omission in main clauses and casualness of speech discussed in Section 1, it supports the hypothesis that zero copula is not a case of copula omission. Based on the observation of a longitudinal linguistic change from zero copula to the overt copula, Nambu (2019) explained that zero copula is the conservative variant and thus preferred to be used with the formal style *omoware-*.

This section investigates how much learners of Japanese share the norm for Japanese language, employing a questionnaire survey on correctness and formality regarding the use of zero/overt copula. With respect to the methodology, this survey is to elicit the explicit attitude of learners toward zero copula and thus considered as a direct approach, in which participants are explicitly asked about the linguistic forms to elicit their evaluations, rather than an indirect approach, in which their attitude to the forms is implicitly elicited through an elaborate design as a form of experiment such as the matched-guise technique (Campbell-Kibler 2008, Garrett 2007, 2010). Therefore, in this survey, participants were explicitly asked to judge zero/overt copula in sentences using a five-point Likert scale. The participants were 25 learners of Japanese recruited at Monash University including two heritage speakers of Japanese who are also studying Japanese at the university. Their Japanese levels vary but all of them were in the upper two of four levels or above provided at Monash University.

Target sentences for the survey for native speakers' attitude in Nambu (2019) were used for the survey in this study. The following linguistic conditions were included in the set of target sentences:

(6) a. Predicate: *omow-* 'think', *omoware-* 'seem' (the spontaneous form of *omow-*)
     b. Preceding item that the copula is attached to: noun, adjectival noun
     c. Subject marker in a subordinate clause: the nominative case particle *ga*, the accusative case particle *o*[8]

Each of the three conditions with two levels (e.g., noun, adjectival noun) in (6) was provided with and without a copula, yielding 12 total target sentences. Since it is a direct approach, the participants were given a pair of two sentences with an overt/zero copula successively so that they could directly compare the variants to mark their evaluations; thus, the elicited values are relative rather than absolute. Example sentences used in the questionnaire are given in (7).

(7) a. Sobo-wa   sono konbini-ga   benri-**da**-to   omot-ta.
         grandmother-TOP   that convenience.store-NOM   convenient-COP.PRS-COMP   think-PST
         '(My) grandmother thought that the convenience store is convenient.'

     b. Sobo-wa   sono konbini-ga   benri-to   omot-ta.
         grandmother-TOP   that convenience.store-NOM   convenient-COMP   think-PST
         '(My) grandmother thought that the convenience store is convenient.'

3.2 *Results* Since the target sentences in this survey were identical to the ones in the survey for native speakers in Nambu (2019), the data of learners in this study is compared with the data of native speakers. Regarding correctness, Figure 2 illustrates mean ratings of overt/zero copula by predicate type (*omow-* or *omoware-*). There is a similar trend between learners and native speakers in that both graphs in Figure 2 show that the rate of zero copula was higher with the verb *omoware-* than with *omow-*. This conforms to the trend observed in the corpus data in Nambu (2019) in Section 2.2.

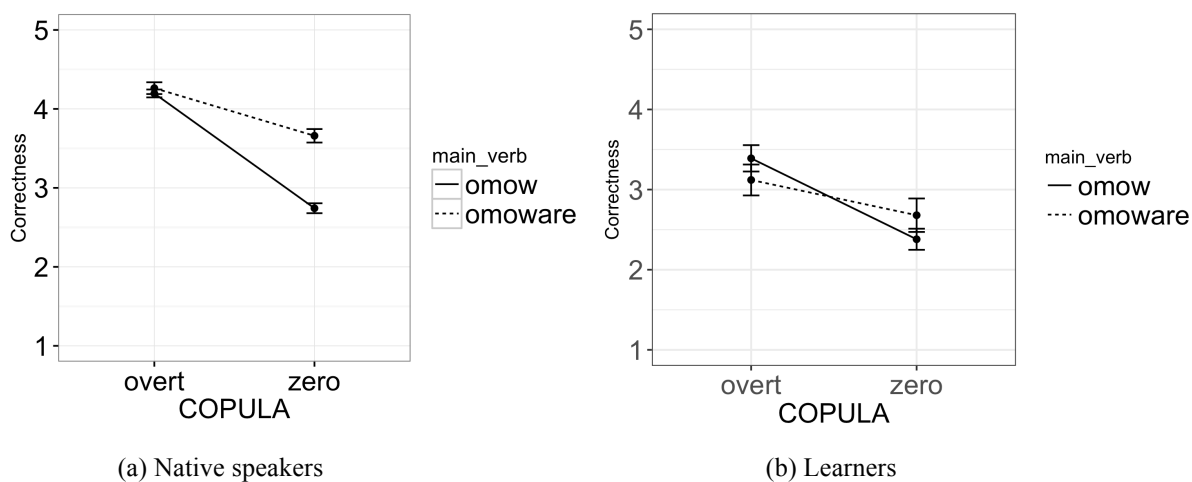(a) Native speakers   (b) Learners

Figure 2: Mean ratings of correctness by predicate type

To confirm the significance of the trend, a linear mixed-effects model was constructed, incorporating the fixed effects of COPULA (overt, zero), VERB (*omow-*, *omoware-*), and their interaction, as well as random intercepts and slopes for participants. The analysis was conducted in the R environment (R Core Team 2016), using the *lmerTest* package for R (Kuznetsova et al. 2017). The results indicate that the data of native speakers confirmed the improvement in the ratings of zero copula with *omoware-* in contrast to the *omow-* condition (the interaction of the factors COPULA and VERB: native speakers, $t=8.15$, $p<.001$), but the data of learners was only marginally significant ($t=1.76$, $p=.08$).

    In order to investigate whether the trend in the native speakers' data in Figure 2-a emerged from the link between the formality of the verb *omoware-* and zero copula as discussed in Section 3.1, I illustrate mean ratings of formality by the predicate type in Figure 3. As seen in Figure 3-a, the native speakers marked zero copula with *omoware-* as more formal in comparison to that with *omow-* (the interaction of the factors COPULA and VERB:

---

[8] (5c) is a grammatical phenomenon known as Exceptional Case Marking (ECM) in the generative paradigm, in which a subject in a subordinate clause is marked by the accusative marker *o* instead of the nominative marker *ga* (Kuno 1976, Chomsky 1981).

t=3.41, *p*<.001), which is congruent to the trend in Figure 2-a and signals the link between the formality of the verb *omoware-* and zero copula. On the other hand, the learners' data in Figure 3-b does not reveal an analogous trend but shows that the overt copula with *omoware-* was marked more formal across the conditions. To identify what caused this trend in the learners' data, I examined the data further in terms of their Japanese proficiency level but could not detect anything that is responsible for the trend. One potential factor to be considered is that their evaluation on overt/zero copula may be influenced by the formality of the verb *omoware-* itself, which masked the genuine effect; further study with substantial data is necessary to explore this possibility.

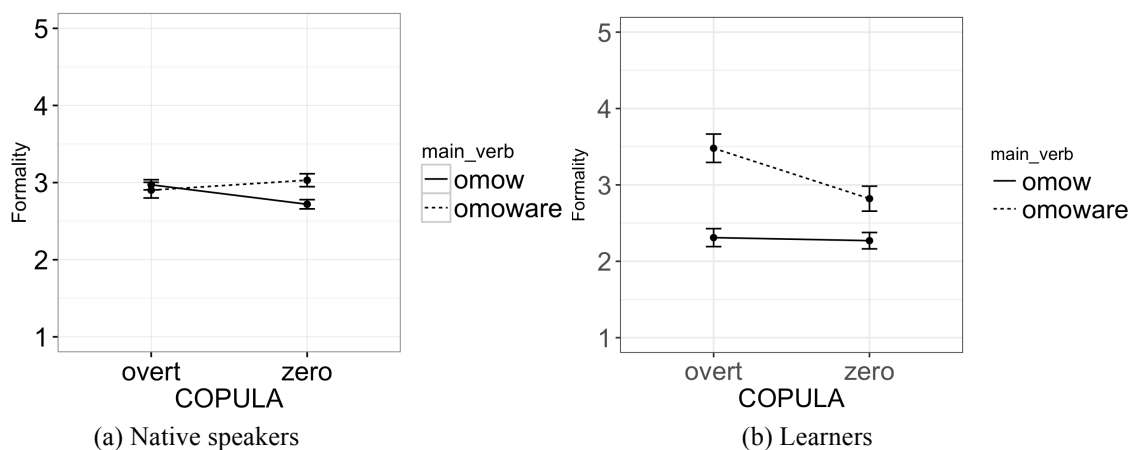(a) Native speakers   (b) Learners

Figure 3: Mean ratings of formality by predicate type

Figure 4 shows mean rating scores of correctness by type of preceding item, demonstrating that native speakers did not perceive any difference based on the type (the interaction of the factors COPULA and PREVIOUS_ITEM, *t*=1.21, *p*=.23) but learners rated the overt copula with a noun lower than with an adjectival noun (*t*=2.21, *p*<.05). Referring to the frequency in the corpus data in Table 3 and 4 in Section 2.2, the learners' evaluations in Figure 4-b reflect the distinction between the noun and adjectival noun in their use in the I-JAS corpus, whereas the native speakers' evaluations in Figure 4-a are not that straightforward since they do not conform to the trend in the written data but do so in the spoken data. Since the spoken data in the CSJ corpus consists of more recent language uses compared to the written data in the BCCWJ corpus that spans a wide range of published years, the fact that the trend in Figure 4-a as the current norm matches the trend in the spoken data is no surprise by assuming that the difference between the adjectival noun and noun may be diminishing in the course of the linguistic change.

(a) Native speakers   (b) Learners
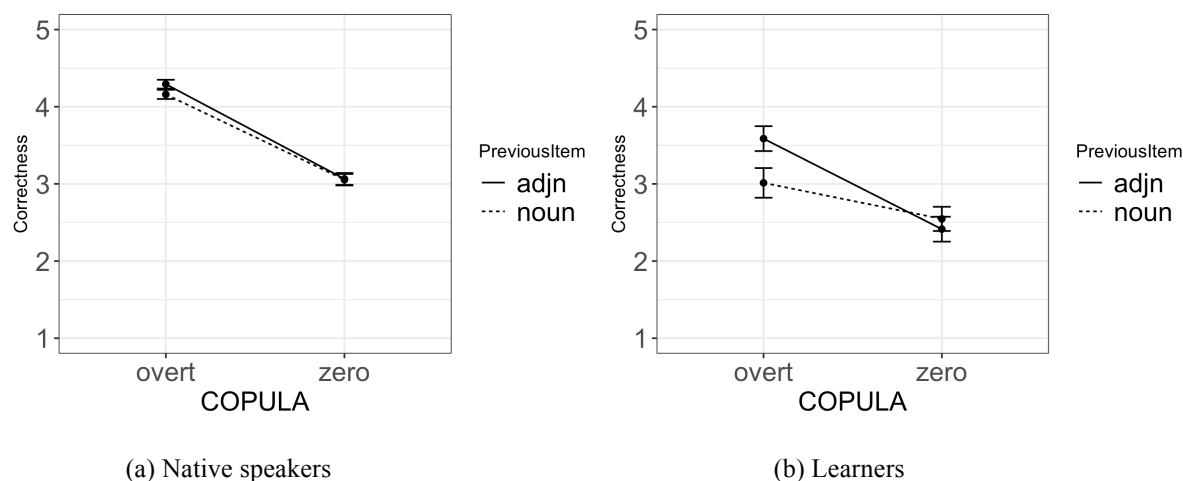
Figure 4: Mean ratings of correctness by type of previous item

To summarize, the data from the questionnaire survey demonstrates some sensitivity of learners to the linguistic conditions in terms of evaluations on zero/overt copula. As is the case with native speakers, the learners also generally consider the overt copula as the correct form in comparison to zero copula. In addition, they distinguish the evaluations by the predicates *omow-* and *omoware-*, but a reason for making the difference may be different from how native speakers do, as indicated by the difference between Figure 3-a and 3-b; while the

native speakers may have distinguished the evaluations by type of predicate in correctness (Figure 2-a) due to the formality difference (Figure 3-a), the learners may have distinguished them (Figure 2-b) simply based on the frequency difference in use. If this is the case, it suggests that the learners are not necessarily aware of what causes the difference between *omow-* and *omoware-* in terms of the use/evaluation on zero copula but they can still replicate the difference as the native speakers do.

4   Conclusion

Drawing on the data from a questionnaire survey and a corpus, this preliminary study examined how learners of Japanese perceive and use zero copula from a quantitative perspective. Regarding the use in the corpus data, the results suggest that they can acquire the difference in frequency of the zero copula use between linguistic conditions over the course of developing their language skills. The results of the questionnaire survey demonstrate a difference between the learners and the native speakers in evaluations on zero copula, which suggests that the learners may not be necessarily aware of the reason for a difference in use between linguistic conditions but can still replicate the difference in use. Further research is required to delve into the observed trends based on data substantial enough for statistical analyses such as examining data of evaluation by Japanese proficiency levels of learners.

References

Abe, Jiro. (2001) A o B da to omou to A o B to omou ["A o B da to omou" and "A o B to omou"]. *Nihongo to nihon bungaku* 33, 14-24. University of Tsukuba.

Abe, Jiro. (2015) Inyoukunai niokeru kopyura no hishutsugen nitsuite: Da to omou to to omou [On the non-occurrence of a copula in a quotative phrase: "da to omou" and "to omou"]. In J. Abe, I. Iori, & T. Sato (eds.), *Bunpou Danwakenkyu to Nihongo Kyouiku no Setten* [Interface between grammar/discourse analysis and Japanese language education], 57-78. Tokyo: Kurosio Publishers.

Bender, Emily. (2000) *Syntactic Variation and Linguistic Competence: The Case of AAVE Copula Absence*. Doctoral dissertation, Stanford University.

Campbell-Kibler, Kathryn. (2008). I'll be the judge of that: Diversity in social perceptions of (ING). *Language in Society* 37(5), 637-659.

Chomsky, Noam. (1981) *Lectures on Government and Binding*. Dordrecht: Foris Publications.

Cukor-Avila, Patricia. (1999) Stativity and copula absence in AAVE: Grammatical constraints at the subcategorical level. *Journal of English Linguistics* 27(4), 341-355.

Garrett, Peter. (2007) Language attitudes. In C. Llamas, L. Mullany, & P. Stockwell (eds.), *The Routledge Companion to Sociolinguistics*, 116-121. New York: Routledge.

Garrett, Peter. (2010) *Attitudes to language*. Cambridge: Cambridge University Press.

Kuno, Susumu. (1976) Subject raising. In M. Shibatani (Ed.), *Syntax and semantics 5: Japanese generative grammar,* 17-49. New York: Academic Press.

Kuznetsova, Alexandra, Brockhoff, Per B., & Christensen, Rune H. B. (2017) lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software* 82(13), 1-26.

Labov, William. (1966) Hypercorrection by the Lower Middle Class as a Factor in Linguistic Change. In William Bright (ed.), *Sociolinguistics: Proceedings of the UCLA Sociolinguistics Conference* 1964, 84-113. The Hague: Mouton.

Labov, William. (1969) Contraction, deletion, and inherent variability of the English copula. *Language* 45(4), 715-762.

Labov, William. (1990) The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2, 205-254.

Maekawa, Kikuo. (2004) Design, compilation, and some preliminary analyses of the Corpus of Spontaneous Japanese. In K. Maekawa and K. Yoneyama (eds.), *Spontaneous speech: Data and analysis*. 87-108. Tokyo: The National Institute of Japanese Language.

Maekawa, Kikuo, Yamazaki, Makoto, Ogiso, Toshinobu, Maruyama, Takehiko, Ogura, Hideki, Kashino, Wakako, Koiso, Hanae, Yamaguchi, Masaya, Tanaka, Makiro, & Den, Yasuharu. (2014) Balanced corpus of contemporary written Japanese. *Language Resources and Evaluation* 48(2), 345-371.

Nambu, Satoshi. (2019) *A quantitative study on zero copula in Japanese.* Manuscript.

R Core Team. (2016) *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Sakoda, Kumiko, Madoka Konishi, Aiko Sasaki, Wakako Suga, and Yoko Hosoi. (2016) Tagengo no Nihongo Gakushuusha Oodan Koopasu [International Corpus of Japanese as a Second Language]. *NINJAL Project Review* 6(3), 93-110.

Shioda, Takehiro. (2010) Danuki Kotoba? ['Da' deletion?]. *Online article at NHK Broadcasting Culture Research Institute* (http://www.nhk.or.jp/bunken/summary/kotoba/term/140.html).

Tanomura, Tadaharu. (2008) Daikibona denshishiryou ni miru gendai nihongo no doutai [A corpus-based analysis of some time-related aspects of contemporary Japanese]. *Machikaneyama Ronso* 42, 55-77. Osaka University.

Tsujimura, Natsuko. (2014) *An Introduction to Japanese Linguistics*. Malden, MA: Wiley-Blackwell.

Yamaguchi, Toshiko. (2007) *Japanese Linguistics: An introduction*. London: Continuum.