

# Information Structure and the Comprehension of Non-Canonical Word Orders in Japanese

Edson T. Miyamoto, Junna Yoshida, Ryosuke Kohita, Shiori Seki, Anubhuti Chauhan, Tayebbeh Norouzi, Bronwyn G. Wilson, and Muchamad Zinuri

Keywords: word order, scrambling, information structure, reading times, Japanese

## 1. Introduction

In Japanese, non-canonical word orders are known to be harder to understand compared to canonical orders (Miyamoto 2006a, for a summary). However, difficulty with the non-canonical OVS (object-verb-subject) order in Finnish has been shown to decrease when the scrambled constituent is mentioned in prior context (Kaiser & Trueswell 2004), and since then comparable results have been reported for Japanese (Koizumi 2010; but see Ferreira & Yoshita 2003, for production data).

We report a reading-time experiment replicating Koizumi's results with a key modification. Participants read the sentences for content and were not required to judge whether each sentence was natural or not.

## 2. Previous results: Koizumi (2010)

Koizumi (2010) reported a self-paced reading experiment in Japanese with pairs of sentences as items. The first sentence provided a simple context introducing either of two names. The second sentence used the two names in a transitive construction (e.g., “Kuroki welcomed Kaneda”). Two factors were manipulated: word order (whether the order in the second sentence was the canonical SOV or the scrambled OSV) and information structure (whether the order of the names in the second sentence was old-new or new-old, depending on whether each name had been mentioned in the first sentence or not). The 48 test items were distributed into four lists according to a Latin Square design. To each list, 84 distractor filler items were added. Some of the fillers were grammatically or semantically anomalous. Fifty-six native Japanese speakers read the sentences and judged whether they were acceptable or not.

Reading-time results for the second NP in the second sentence were as follows. There was a main effect of word order such that the OSV order was slower than the SOV order. There was also a main effect of information structure as the new-old conditions were slower than the old-new conditions. More importantly, there was an interaction as the information-structure effect (i.e., new-old slower than old-new) was larger between the

OSV conditions than between the SOV conditions.

Because readers are unlikely to pay attention to anomalous sentences in everyday use, there is an interest in verifying whether the result would also hold when readers are not explicitly instructed to look for anomalies. Moreover, previous results suggest that word-order difficulty interacts with the frequency of the verb in the sentence (Miyamoto 2006b), therefore it is also of interest to determine whether the frequency of the names used would interact with the difficulty associated with the OSV order. A reading-time experiment was conducted to examine these questions.

### 3. Experiment

#### 3.1. Method

##### 3.1.1. Participants

A total of 48 native Japanese speakers, students at the University of Tsukuba, were paid to participate in the experiment. There were two versions of the experiment (see the materials section for details) with 29 participants in version 1A and 19 participants in version 1B.

##### 3.1.2. Materials

Each item contained two sentences as in the following example.

###### 1. Item in the SOV/Old-New condition

###### a. Sentence 1: context

Saikin	wadai-no	senshu-no	hitori-wa	Sato	da-souda.
recently	talked-about-Gen	player-Gen	one-Top	Sato	is-Aux

“One of the recently talked about players is Sato.”

###### b. Sentence 2: crucial sentence

Sato-ga	Suzuki-o	sidosite-iru	rasii.
Sato-Nom	Suzuki-Acc	advising-be	Aux

“Sato seems to be advising Suzuki.”

The experimental design closely followed Koizumi (2010). The first sentence provided a context for the second sentence, which was the point of interest. There were four conditions in a 2 by 2 design (see Table 1 for schematic representations and Appendix A for the four versions of the item in Japanese characters).

The factor *order* refers to the word order in the second sentence: SOV or OSV. The factor *information status* refers to the status of the two nouns in the second sentence (column *N1-N2* in Table 1) depending on whether they were mentioned in the first sentence or not (old or new). Therefore, (1) is an example of the SOV/Old-New condition. Reading times to each corresponding region in the sentences were compared. Thirty-two sets of items were created. Each set contained four versions following the

scheme in Table 1.

**Table 1 Design of the experiment**

Conditions		Context			Sentence			
Order	N1 - N2	C1	C2	C3	S1	S2	S3	S4
SOV	Old-New	Recently	Sato	is.	Sato-	Suzuki-	advise	seems.
	New-Old	talked	Suzuki		Nom	Acc		
OSV	Old-New	about	Suzuki		Suzuki-	Sato-		
	New-Old	player	Sato		Acc	Nom		

The family names used were obtained from a list of 200 high-frequency names in a national telephone directory (Nihon Soft 1997). All the names were written with two kanji characters and their most common pronunciations were three-morae long. Because lexical frequency may interact with the difficulty in processing non-canonical orders, frequency was varied as follows. Sixteen items contained pairs of high-frequency names (between 59,993 and 481,980) and the remaining 16 items used pairs of low-frequency names (between 26,543 and 41,289). The frequency difference was reliable (mean raw frequencies: high 132,345.59, low 32,404.5;  $F(1,30)= 13.62$ ,  $p < .01$ ; log-transformed means: high 11.58, low 10.38,  $F(1,30)= 61.63$ ,  $p < .001$ ).

Sixty distractor filler items similar to the test sentences were created containing two sentences each. The second sentence included various types of constructions and word orders. Because of a formatting mistake, all filler items in version 1A of the experiment were such that the first NP in the second sentence was always old information. The fillers were corrected in version 1B so that information status was counter-balanced across the items.

### 3.1.3. Procedure

The self-paced reading experiment was conducted using Doug Rohde's Linger program (version 3) by presenting sentences one region at a time in a non-cumulative moving-window fashion (Just, Carpenter & Woolley 1982) using double-byte Japanese characters with the uniform-width font MS Mincho. Regions were initially shown masked with underscores and were revealed one at a time at each button press. The two sentences of each item were presented on the same screen on separate lines. After each item, a yes-no question was presented on a new screen and feedback was provided if the answer was incorrect.

The 32 sets of items were distributed into four lists according to a Latin Square design so that each list contained one version of each set. Each participant saw one list interspersed with 60 filler items in pseudo-random order so that test items were shown with at least one intervening filler item.

### 3.1.4. Analysis

Questions to six items were incorrectly phrased and were not included in the comprehension analyses but the corresponding reading times were all included in the reading-time analyses reported. For the remaining 26 items, reading times were included only if the corresponding question was answered correctly. For these 26 items, all participants' comprehension performance was above 80% (above 83% when fillers were also included), therefore participants were all likely to be reading carefully and none was eliminated from further analysis.

Mixed-effects analyses were conducted in R (R Core Team 2013). For response accuracy to the comprehension questions, logit mixed-effect models were used with the function *glmer* (*lme4* package; Bates, Maechler, Bolker & Walker 2014). For reading times, the function *lmer* (*lme4* package) was used. P-values reported for main effects and interactions were based on Wald chisquare tests (*Chisq*) calculated using the function *Anova* (*car* package version 2.0.20, Fox & Weisberg 2011). Tukey-adjusted pairwise comparisons are reported based on least square means (*lsmeans* package, Lenth 2014).

Information status (new-old or old-new), word order (SOV or OSV), experiment version (1A or 1B) and their interactions were included as fixed factors. Log-transformed frequencies of the proper names were included as a fixed factor as follows: for region S1, the log-frequency of the first noun and its interactions with the other factors; for region S2, the log-frequency of the second noun and its interactions with the other factors; for regions S3 and S4, the log-frequency of both nouns and all interaction terms. As random factors, intercepts for participants and items, as well as by-participants and by-items slopes for information status, word order and interaction terms were included.

Reading times are reported in milliseconds (ms). Because participants tend to speed up as an experiment progresses, reading times were residualized over trial number. Residual reading times beyond three standard deviations from the mixed-model predictions were removed (affecting less than 2.6% of the data) and the remaining data points were used to run the model reported (see Baayen 2008).

Similar trends to those reported were observed when all data points were included with residualized as well as raw reading times.

Compared to participants in version 1A of the experiment, participants in version 1B were faster to read the sentences but less accurate when answering the comprehension questions. Interaction between version of the experiment and the two factors of interest did not reach the 5% significance level, and therefore experiment version is not discussed further, since it is not of theoretical interest.

All means reported in the text and in the figures are over participants.

## 3.2. Results

### 3.2.1. Comprehension question accuracy

Comprehension accuracy for 26 items (6 items were not included in this analysis, see the analysis section) revealed no reliable differences ( $ps > .13$ ).

**Table 2 Means and standard errors for the comprehension question of 26 items**

Word order	N1-N2	Mean Percentage Correct (SE)
SOV	Old-New	96.31 (1.12)
SOV	New-Old	96.79 (1.02)
OSV	Old-New	93.99 (1.39)
OSV	New-Old	91.73 (1.56)

### 3.2.2. Reading times

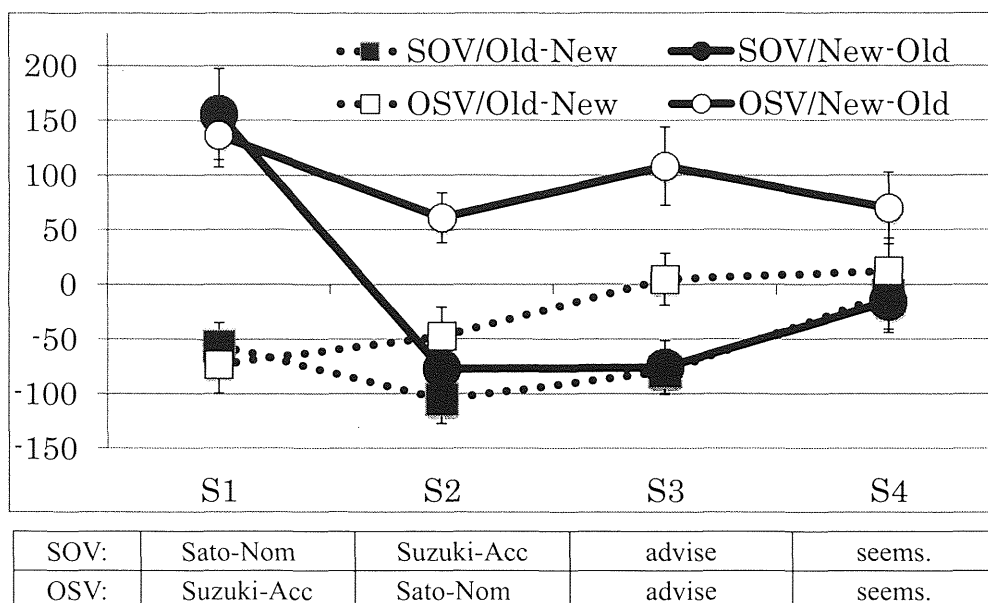
See Figure 1 for the mean residual reading times to each region of the second sentence (see Appendix B for the reading times to all regions). Because the frequency of the names used interacted with the version of the experiment, results including this factor are not reported below (analyses not including frequency as a factor yielded similar trends, see Appendix C).

In region S1 (the first NP in the second sentence), there was a main effect of information structure as old information (the NP was mentioned in the first sentence) was read more quickly than new information (the NP was not mentioned in the first sentence;  $Chisq = 39.277$ ,  $p < .0001$ ). The result is unsurprising as repeated words tend to speed up processing (e.g., by facilitating lexical access). Word order and interaction were not significant ( $ps > .72$ ).

In region S2 (the second NP), there were main effects of word order (OSV slower than SOV,  $Chisq = 12.728$ ,  $p < .001$ ) and of information structure (new-old slower than old-new,  $Chisq = 11.351$ ,  $p < .001$ ). Their interaction was also reliable ( $Chisq = 4.417$ ,  $p = .036$ ) as the word-order effect was larger in the new-old conditions ( $p < .001$ ) than in the old-new conditions ( $p = .0997$ ).

The trends in region S2 persisted in region S3 (the transitive verb). There were a main effect of word order (OSV slower than SOV,  $Chisq = 15.21$ ,  $p < .001$ ) and a marginal main effect of information structure (new-old slower than old-new,  $Chisq = 2.91$ ,  $p = .088$ ). The interaction was reliable ( $Chisq = 7.79$ ,  $p = .005$ ) as the word-order effect was larger in the new-old conditions ( $p < .001$ ) than in the old-new conditions ( $p = .0597$ ). However, the interaction was not consistent across experiment versions and should be interpreted with caution (3-way interaction:  $Chisq = 8.56$ ,  $p = .003$ ).

In region S4 (the auxiliary verb), there were no reliable differences ( $ps > .13$ ).



**Figure 1** Mean residual reading times and standard errors per region (in ms) for the crucial sentence regions

### 3.3. Discussion

Overall the results replicated the reading times reported by Koizumi (2010) even though fewer items and fewer participants were included (48 participants and 32 items, as opposed to Koizumi’s 56 participants and 48 items). More crucially, the result was obtained from participants who were reading sentences for content, and were not required to judge whether each sentence was acceptable or not (as was required in Koizumi’s experiment to make participants pay close attention to the constructions).

The inconsistent results related to the frequency of the names used needs further study. Perhaps the frequency range was too restricted as the names in this experiment were chosen from among 200 high-frequency Japanese names in national telephone directories (Nihon Soft 1997).

## 4. General Discussion

The results were largely compatible with previous results (Koizumi 2010). In regions S2 and S3 in particular, the OSV order was reliably slower than the SOV order when the NPs were presented in the new-old order, while only marginally so when presented in the old-new order. In other words, the OSV order was particularly hard when the scrambled object was new information and the subject was old information.

Most previous experiments on scrambling presented sentences in isolation without embedding them in contexts, therefore both NPs were new information (see Miyamoto 2006a, for a summary). Taken together with the present results, the trend seems to be

that scrambled orders are hard to understand when the scrambled NP is new information. But there are at least two possible interpretations for such an outcome.

One possibility is that there is always a processing cost for the OSV order and the cost is compounded when new information is scrambled. A second possibility is that scrambling is only costly when new information is scrambled. The distinction is important because it points to two different ways of understanding the nature of scrambled orders. The second possibility suggests a tight relationship between scrambling and information structure such that scrambling is only natural when old information is scrambled and perhaps only used in order to avoid a new-old ordering of constituents. For example, if the subject is new information and the object is old information, producing a sentence in the OSV order (i.e., old-new) guarantees that old information comes first and that might be the only kind of situation where scrambling would be warranted.

One piece of evidence against this second alternative is that even though the old-new OSV order is relatively fast compared to the new-old OSV order, it is nevertheless slower than the old-new SOV order. For example, Koizumi (2010) reports a main effect of word order at the verb regardless of information structure (see also the marginal difference between OSV and SOV, when both are old-new, in regions S2 and S3 of the present experiment; also, Kaiser & Trueswell 2004, for comparable results for Finnish). If this is correct, it suggests that scrambling is not directly conditioned on information structure. Although the cost of scrambled orders can be modulated by information structure, it is still observed in the optimal scenario from the information-structure perspective (i.e., in the old-new order). This conclusion also seems to fit results in production of ditransitive constructions in Japanese where old-new objects were produced more often than new-old objects regardless of whether the resulting word order was the canonical dative-accusative or the scrambled accusative-dative (Ferreira & Yoshita 2003). If scrambling was motivated by information structure, the scrambled order should have been produced relatively more often to avoid the new-old order. In sum, the cost in processing scrambled orders is unlikely to be exclusively due to information structure (see Miyamoto 2006a, for a possible explanation based on the working memory cost of maintaining the dependency between scrambled constituent and its canonical position).

## **Appendix A**

What follows is the item set in Table 1 using Japanese characters (spaces indicate the segmentation used in the self-paced presentation).

a. SOV / Old-New:

最近話題の選手の一人は 佐藤 だそうだ。 佐藤が 鈴木を 指導している らしい。

b. SOV / New-Old:

最近話題の選手の一人は 鈴木 だそうだ。 佐藤が 鈴木を 指導している らしい。

c. OSV / Old-New:

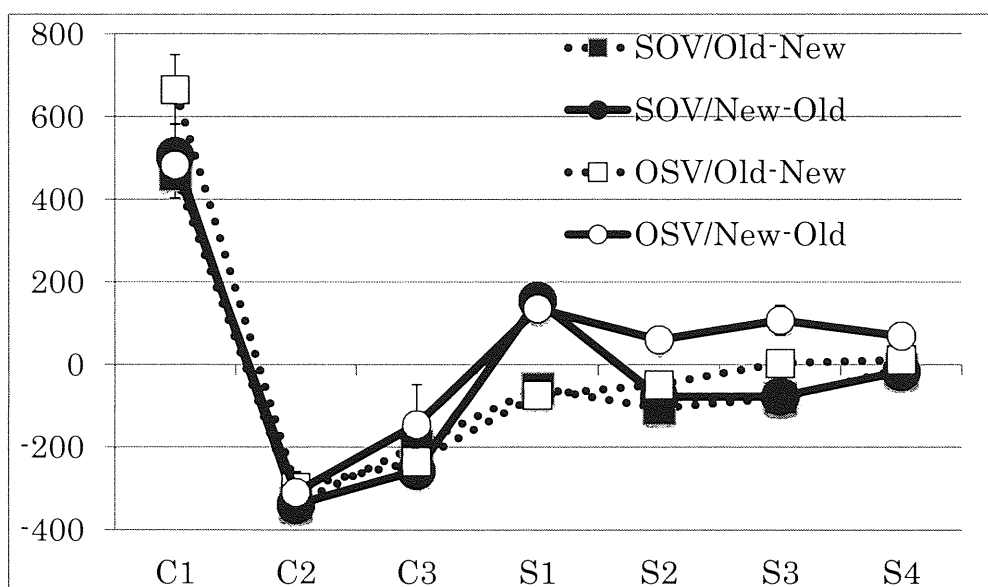
最近話題の選手の一人は 鈴木 だそうだ。 鈴木を 佐藤が 指導している らしい。

d. OSV / New-Old:

最近話題の選手の一人は 佐藤 だそうだ。 鈴木を 佐藤が 指導している らしい。

## Appendix B

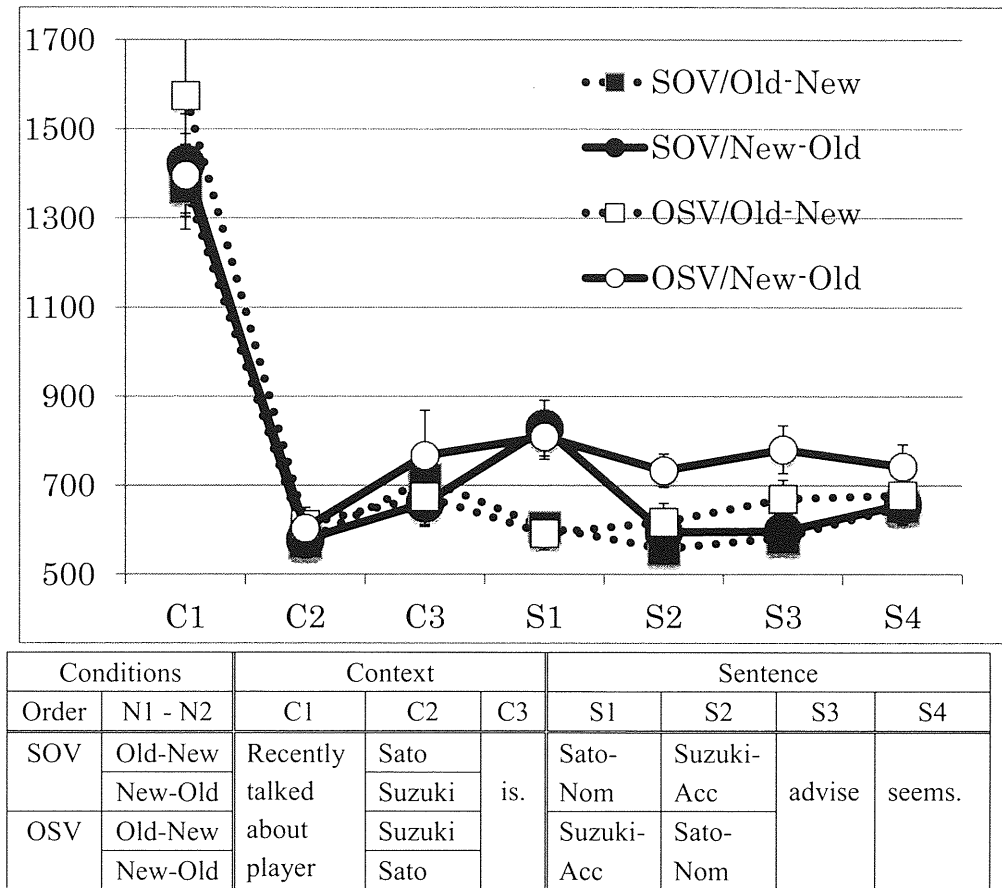
Mean residual reading times to all regions (context and crucial sentence) are shown in Figure 2. Mean raw reading times are shown in Figure 3.



Conditions		Context			Sentence			
Order	N1 - N2	C1	C2	C3	S1	S2	S3	S4
SOV	Old-New	Recently talked	Sato	is.	Sato-Nom	Suzuki-Acc	advise	seems.
	New-Old		Suzuki		Sato-Nom	Suzuki-Acc		
OSV	Old-New	about player	Suzuki		Suzuki-Acc	Sato-Nom		
	New-Old		Sato		Suzuki-Acc	Sato-Nom		

Figure 2 Mean residual reading times and standard errors per region (in ms)





**Figure 3 Mean raw reading times and standard errors per region (in ms)**

## Appendix C

For the following analyses, frequency was not included as a factor. Overall the trends were similar to those reported in the main text.

The factors used were as follows. Information status (new-old or old-new), word order (SOV or OSV), experiment version (1A or 1B) and their interactions were included as fixed factors. As random factors, intercepts for participants and items, as well as by-participants and by-items slopes for information status, word order and interaction terms were included.

Results for the context regions were as follows. For region C1, there was a main effect of word order (OSV slower than SOV;  $\chi^2 = 6.07$ ,  $p = .0138$ ) but the effect is spurious as the same words had been read across the conditions at that point. Readers sometimes rest instead of reading the first region, leading to random differences in the beginning of the sentence. There were no reliable effects of information structure or interaction ( $ps > .29$ ).

There were no reliable differences in regions C2 or C3 (all  $ps > .10$ ).

Results for the second sentence were as follows. In region S1 (the first NP in the second sentence), there was a main effect of information structure as old information (the NP was mentioned in the first sentence) was read more quickly than new information (the NP was not mentioned in the first sentence;  $Chisq = 42.73$ ,  $p < .0001$ ). Word order and interaction were not significant ( $ps > .51$ ).

In region S2 (the second NP), there were main effects of word order (OSV slower than SOV,  $Chisq = 13.04$ ,  $p < .001$ ) and of information structure (old-new faster than new-old,  $Chisq = 11.84$ ,  $p < .001$ ). The interaction was also reliable ( $Chisq = 4.80$ ,  $p = .028$ ) as word order effect was reliable in the new-old conditions ( $p = .0001$ ) but not in the old-new conditions ( $p = .100$ ).

In region S3 (the transitive verb), there was a main effect of word order (OSV slower than SOV,  $Chisq = 15.01$ ,  $p < .001$ ). But there was no main effect of information structure ( $p > .14$ ). The interaction was reliable ( $Chisq = 8.63$ ,  $p = .003$ ) but the effect was not consistent across the two versions of the experiment (3-way interaction:  $Chisq = 7.16$ ,  $p = .007$ ).

In region S4 (the auxiliary verb), there was a marginal effect of word order (OSV slower than SOV,  $Chisq = 3.36$ ,  $p = .067$ ) but information structure and the interaction were not reliable ( $ps > .69$ ).

## 【References】

- Baayen, R. H. (2008) *Analyzing linguistic data: A practical introduction to statistics using R*. Cambridge University Press.
- Bates, Douglas, Maechler, Martin, Bolker, Ben, & Walker, Steven (2014) *lme4: Linear mixed-effects models using Eigen and S4*. R package version 1.1-7.  
<http://CRAN.R-project.org/package=lme4>
- Ferreira, Victor S., & Yoshita, Hiromi (2003) Given-new ordering effects on the production of scrambled sentences in Japanese. *Journal of Psycholinguistic Research*, 32, 669-692.
- Fox, John, & Weisberg, Sanford (2011) *An R Companion to Applied Regression*, 2nd ed. Thousand Oaks CA: Sage. URL: <http://socserv.socsci.mcmaster.ca/jfox/Books/Companion>
- Kaiser, Elsi, & Trueswell, John C. (2004) The role of discourse context in the processing of a flexible word-order language. *Cognition*, 94, 113-147.
- Koizumi, Masatoshi (2010) *Project report "A cognitive-neuroscientific study of the contextual effects on the comprehension of scrambled sentences"*, Grants-in-Aid for Scientific Research. <https://kaken.nii.ac.jp/pdf/2010/seika/jsps/11301/19320056seika.pdf> (In Japanese).
- Lenth, Russell V. (2014) *lsmeans: Least-squares means*. R package version 2.12.  
<http://CRAN.R-project.org/package=lsmeans>
- Miyamoto, Edson T. (2006a) Processing alternative word orders in Japanese. In M. Nakayama, R. Mazuka, & Y. Shirai (eds.), *Handbook of East Asian Psycholinguistics*, Vol. 2, Japanese

- (pp. 257-263). Cambridge, UK: Cambridge University Press.
- Miyamoto, Edson T. (2006b) Understanding sentences in Japanese bit by bit. *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society*, 13, 247-260.
- Nihon Soft (1997) As cited in Shirooka, Keiji <http://www.ipc.shizuoka.ac.jp/~jjksiro/nihonsei.html>  
Downloaded June 22, 2014. (In Japanese)
- R Core Team (2013) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org/>.