

INTEGRATION OF STRUCTURAL PROBABILITY IN SPEECH PRODUCTION: EVIDENCE FROM JAPANESE RELATIVE CLAUSES

Amber B. Camp¹ and Nozomi Tanaka²

¹University of Hawai'i and ²Indiana University
acamp@hawaii.edu, tanakan@indiana.edu

ABSTRACT

The predictability of words (e.g., word frequency, subcategorization bias) is previously known to influence acoustic output, in which more predictable words are produced with a shorter duration. This study explores whether structural probability is also reflected in incremental speech production, replicating a previous study by Kurumada (2011). We used Japanese relative clauses to address this question, as these relative clauses are prenominal, and the embedded subject can be marked with the nominative marker *-ga* or the genitive marker *-no*; the nominative marker is also used in the independent clause, but the genitive marker is not, increasing the probability of an upcoming head noun. Results from nine native speakers of Japanese did not show any significant difference between the duration of head nouns between when they were produced following the genitive marker and when they were produced following the nominative marker.

Keywords: word duration, relative clauses, *ga-no* conversion, syntactic probabilities

1. INTRODUCTION

Acoustic properties of incremental speech production, such as word duration, are known to reflect probabilities of words, sentence structure, and contexts chosen by the speakers. For example, words with high frequency are produced with shorter syllable duration [1] and highly frequent function words are more likely to be longer in less predictable contexts [3]. The source of this phenomenon has been attributed to articulatory practice and ease of lexical retrieval, as well as speaker-controlled variation based on the accommodation of speakers' needs, as reviewed in [6].

Further studies demonstrated that syntactic probabilities also affect the duration of words. Gahl & Garnsey [6] showed that contextual probabilities based on verb subcategorization have impact on word duration, among other types of pronunciation variation. They compared verbs that are more likely to take a direct object than a sentential complement, such as *confirm*, and verbs that are more likely to take

a sentential complement than a direct object, such as *believe*.

- (1) a. The CIA director *confirmed* the rumor once it had spread widely.
b. The CIA director *confirmed* the rumor should have been stopped sooner.
- (2) a. The job applicant *believed* the interviewer when she discussed things with her.
b. The job applicant *believed* the interviewer had been dishonest with her.

They asked participants to read sentences like (1) and (2) aloud and found that the verbs were lengthened in the less probable contexts; for example, when a verb that is more likely to take a direct object is followed by a sentential complement, as in (1b). Similarly, direct objects were longer following a verb that is more likely to take a sentential complement, as in (2a).

In a more recent study, Tily et al. [14] used the dative alternation, (3), to examine whether word duration is influenced by structural probabilities.

- (3) a. They sent us two tickets.
b. They sent two tickets to us.

In English, different factors, such as givenness, animacy, length, and plurality of the recipient and theme influence which dative alternant is more likely to be used. Tily et al. [14] analyzed spontaneous speech and found that the probability of the speakers' choice of dative patterns is a predictor of word duration.

However, in both of these studies, the probability is estimated based on a local lexical cue. In [6], the verb serves as the cue to the complement that immediately follows. In [14], the probability was estimated based on two neighboring arguments.

For these reasons, [10] used the Japanese *Ga-No* conversion to investigate whether contextual probabilities can be estimated by a more global cue.

Ga-No conversion refers to a phenomenon in Japanese relative clauses, in which the embedded subject can be marked with the nominative marker *-ga* or the genitive marker *-no* [7, 8], as shown in the following examples taken from Watanabe [15].

- (4) a. [*kinoo John-ga katta*] *hon*
 yesterday John-NOM bought book
 ‘the book that John bought yesterday’
- b. [*kinoo John-no katta*] *hon*
 yesterday John-GEN bought book
 ‘the book that John bought yesterday’

While this alternation is allowed for the subject inside a relative clause, (4b), this is not the case in the independent clause, (5b).

- (5) a. *Kinoo John-ga hon-o katta*
 yesterday John-NOM book-ACC bought
 ‘John bought a book yesterday.’
- b. **Kinoo John-no hon-o katta*
 yesterday John-GEN book-ACC bought
 ‘John bought a book yesterday.’

As seen in (4), Japanese relative clauses are prenominal and use no relative pronoun. Another notable property of Japanese grammar is the frequent omission of NPs. This means that when speakers hear (4a), it is not initially obvious whether they are hearing an independent clause, like in (5a), or a relative clause, like in (4a). On the other hand, when speakers hear (4b), the genitive marker serves as an unambiguous cue for the relative clause, increasing the probability of a head noun. Kurumada [10] used a sentence reading production task to investigate whether the head nouns were produced with a shorter duration when following the genitive marker than when following the nominative marker. Her findings supported this hypothesis, indicating that such a non-local cue is reflected in the speakers’ acoustic output.

Our study replicates [10], but is further modified by inserting an adverbial phrase to increase the distance between *-ga/-no* and the head noun. We also implemented additional changes to improve the experimental design, such as increasing the filler-to-critical items ratio to 2:1 and presenting items in pseudo-randomized order. Following [10], we conducted a sentence reading task, in which the participants were asked to read a set of experimental sentences aloud. In the subsequent sections, we will describe the methodology then present the results from our experiment.

2. METHODS

2.1. Participants

Participants included nine adult native speakers of Japanese (female $n = 6$) between the ages of 22 and

37. All the participants resided in the United States at the time of the study.

2.2. Materials

The stimuli included 24 critical sentences, prepared in two versions: one with the nominative subject, as in (6a), and the other with the genitive subject, (6b):

- (6) a. [*otoosan-ga senshuu katta kutsu*]-*ga*
 father-NOM last.week bought shoes -NOM
suguni kowareteshimatta.
 immediately broke
- b. [*otoosan-no senshuu katta kutsu*]-*ga*
 father-GEN last.week bought shoes -NOM
suguni kowareteshimatta.
 immediately broke
 ‘The shoes that the father bought broke immediately.’

The critical sentences were Latin-squared into two lists so that the same item did not appear in both conditions in either list. Each list also contained 48 filler sentences that did not include the target phenomena. The order of the items within the lists was pseudo-randomized so that no more than two critical sentences appeared consecutively.

2.3. Procedure

Sentences appeared one-by-one on a computer screen. Participants were instructed to read the sentence silently first, then read it aloud at their own pace while being recorded. They were instructed to count two seconds before proceeding to the next item in order to avoid speeding up.

The production task started with 20 practice items so that participants could become comfortable with the task. Participants then read through one of the lists. After they completed reading the list, they were asked to repeat the task so that we have a back-up in case there is equipment failure. In the second reading, however, they were assigned the other list. They were told that the items might appear in a different order but were not informed that the nominative marker and the genitive marker were switched between the two lists. In the end, all participants saw and read both lists. The ordering of the two lists was counter-balanced between participants.

Following the production task, each participant filled out a language background survey. Participants were compensated for participation.

3. RESULTS

The boundaries of each critical sentence and target word were carefully hand-coded in Praat [4]. For the target words, boundaries were placed where formant transitions and waveform patterns were clear.

We took two durational measures: (1) target word/head noun and (2) full sentence. In addition, we also counted the number of morae, in both the head noun and the full sentence, in order to calculate the speech rate by item.

We excluded three trials from three participants due to experiment error and another trial from another participant for a disfluency in the head noun region. Trials with disfluencies in other parts of the sentence, such as a pause or stutter, were not excluded, but false starts (when a participant starts the sentence, stops, and re-starts) were not included; in these cases, the sentence duration was simply measured from the onset of the fluent production of the sentence. Disfluencies that did not lead to exclusions were very few and only appeared in 2.1% of all trials.

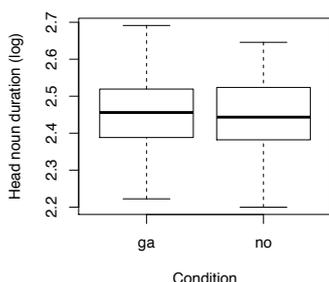
Head noun and sentence durations were measured in milliseconds and then log-transformed for analysis.

Speech rate was calculated by dividing the number of morae in the sentence minus the morae in the head noun by the duration of the sentence excluding the head noun duration. This was done so that the speech rate captured the global speech rate for that trial, but did not take into account the critical measure, the duration of target words between the two conditions. [5] The speech rate was then log-transformed for further analysis.

Table 1: Mean durations of head nouns (in milliseconds and log-transformed) in each condition

Condition	Head Noun (ms)	Head Noun (log)
-ga	289.173 ms	2.452
-no	285.280 ms	2.446

Figure 1: Log-transformed duration of head nouns in each condition.



The results were fit to a linear mixed-effects model with the log-transformed head noun duration as the outcome variable, condition and centered log-transformed speech rate as fixed effects, and items, participants, and trial order as random factors. The maximal random effects structure permitted by the design of the experiment was used. [2] Analysis was conducted in R [12] using the lmerTest package [11].

The results did not indicate a significant effect of condition ($p = .51$). Removing terms did not change the pattern of significance. Excluding trials in which the duration of target words was above or below two standard deviations (3.7% of all trials) also did not affect the pattern of significance, nor did replacing target word duration values in those trials with the mean.

4. DISCUSSION

The goal of the study was to find whether a cue to syntactic probabilities is integrated when it is not available locally, and whether this is reflected in the acoustic output. Our experiment did not find a significant difference in the duration of head nouns that follow the nominative marker and the genitive marker, and thus failed to replicate the results from Kurumada [10]. There are a few possible explanations for our results.

First, according to [7, 8], the acceptability of having an intervening element between the genitive-marked subject and the verb varies depending on speaker and depending on the nature of the intervening element. It is unclear whether time adverbs are also a type of intervening element that may cause the sentence to be ungrammatical. However, even if this did lead to reduced acceptability, it would not contribute to the increase of the probability of an upcoming head noun. A separate experiment without the addition of time adverbs should be run to confirm that this issue is indeed what nullified the results.

Second, it is possible that the time adverb was not presented in its canonical position. The canonical position of time adverbs in Japanese is currently debated. According to [9], which tested the canonical position of time adverbs experimentally, there was no difference between time adverbs appearing sentence-initially and those appearing after the subject. While their results suggest that both positions are equally preferred by participants, again, if the time adverb was perceived by the participants to be in a non-canonical position, it would not contribute to the shortening of the head noun duration. It may be more desirable to use an intervening element whose position is less equivocal.

Lastly, we noticed more disfluencies and false-starts in our filler sentences than in critical sentences. It is possible that our fillers included structures that were less probable, reducing the structural probability effects in our critical sentences in comparison.

5. CONCLUSION

In the current study, we investigated whether a global cue for structural probability is reflected in acoustic properties of output, specifically in word duration. This question was tested with Japanese relative clauses, whose embedded subjects can be marked with either the nominative marker or the genitive marker. The former is also used in an independent clause, but the latter is not, serving as a cue for an upcoming head noun, possibly shortening its duration. Our sentence reading task failed to find any effect of subject marking on head noun duration, failing to replicate the results in [10]. We identified a possible issue in the nature of the intervening element. In future research, this issue should be resolved by either removing the adverb or replacing it with another type of intervener.

Other possible future directions include using spontaneous speech or elicited productions to address the differences between read-aloud speech and speech that is more natural [13].

6. REFERENCES

- [1] Aylett, M., Turk, A. 2004. The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Lang. Speech*. 47, 31–56.
- [2] Barr, D. J., Levy, R., Scheepers, C., Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *J Mem. Lang.* 68, 255–278.
- [3] Bell, A., Jurafsky, D., Fosler-Lussier, E., Girand, C., Gregory, M., Gildea, D. 2003. Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *J. Acoust. Soc. Am.* 113, 1001–1024.
- [4] Boersma, P., Weenink, D. 2018. Praat: Doing phonetics by computer (Version 6.0.43) [Computer program]. www.praat.org
- [5] Campbell, W. N. 1988. Speech-rate variation and the prediction of duration. *Proc. 12th. COLING* Budapest (Vol. 1), 93–95.
- [6] Gahl, S., Garnsey, S. M. 2004. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*. 80, 748–775.
- [7] Harada, S. I. 1971. Ga-No conversion and idiolectal variations in Japanese. *Gengo Kenkyu*. 60, 25–38.
- [8] Harada, S. I. 1976. Ga-No conversion revisited: A reply to Shibatani. *Gengo Kenkyu*. 70, 23–38.
- [9] Koizumi, M., Tamaoka, K. (2006). Bun kaiseki jikken ni yoru nihongo fukushi rui no kihon gojun no hantei [The canonical positions of adjuncts in the processing of Japanese sentences]. *Cognitive Studies*. 13, 392–403.
- [10] Kurumada, C. 2011. Syntactic context affects probability estimation: Evidence from Japanese relative clauses. *85th LSA* Pittsburgh.
- [11] Kuznetsova, A., Brockhoff, P. B., Christensen, R. H. B. 2017. lmerTest package: Tests in linear mixed effects models. *J. Stat. Softw.* 82, 1–26.
- [12] R Core Team. R: A language and environment for statistical computing. <https://www.R-project.org/>
- [13] Schafer, A. J., Speer, S. R., Warren, P. 2005. Prosodic influences on the production and comprehension of syntactic ambiguity in a game-based conversation task. In: Trueswell, J. C., Tanenhaus, M. K. (eds), *Approaches to studying world-situated language use*. Cambridge, MA: MIT Press, 209–225.
- [14] Tily, H., Gahl, S., Arnon, I., Snider, N., Kothari, A., Bresnan, J. 2009. Syntactic probabilities affect pronunciation variation in spontaneous speech. *Lang. Cogn.* 1–2, 147–165.
- [15] Watanabe, A. 1996. Nominative-genitive conversion and agreement in Japanese: A cross-linguistic perspective. *J. East. Asian. Linguis.* 5, 373–410.