Message Planning and Lexical Encoding in Sentence Production
—A View from Japanese Existential Construction—

Hajime ONO† and Hiromu SAKAI‡

† Faculty of Foreign Languages, Kansai Gaidai University
Nakamiya Higashinocho 16-1, Hirakata-shi, Osaka, 573-1001 Japan
‡ Graduate School of Education, Hiroshima University
Kagamiyama 1-1-1, Higashihiroshima-shi, Hiroshima, 739-8524 Japan

E-mail: † onohajim@kansaigaidai.ac.jp, ‡ hsakai@hiroshima-u.ac.jp

Abstract Griffin & Bock (2000) monitored the eye-movement of speakers as they describe simple events and found that the speakers gaze a certain object before they decide what to say, suggesting that the result supports Wundt’s idea of incremental lexical encoding. Sentence formulation, however, involves other stages such as message planning including determination of speakers’ point of view, grammatical function assignment, or linearization of constituents (Bock & Levelt, 1994). To examine the time course of these multiple stages, we conducted an eye-tracking experiment in which the participants were required to describe a picture by an existential construction in Japanese where the elements are aligned as Location-Object word order. The result shows that speakers initially gaze the Object at the message planning stage, which was followed by the lexical encoding stage where speakers’ gaze is shifted to the Location.

Key words: production, eye-tracking, gaze, sentence, Japanese, existential construction, word order, linearization, message

1. Introduction

As the research progresses on the mechanism of language production, it becomes obvious that language production involves extremely complex processes (Fromkin, 1968; Levelt, 1989). The visual world paradigm, which uses an eye-tracking measurement, is getting quite popular to investigate what information the speakers integrate and exactly when they integrate various kinds of information in the course of producing an utterance (Tannenhaus, et al., 1995). Griffin & Bock (2000) found that the speakers gaze a certain object before they actually produce the utterance in their picture description task, and argued that the timing of the gaze reflects the process of sentence formulation, rather than the apprehension of the pictures.

Sentence formulation, however, obviously involves multiple stages, such as determination of the speaker’s point of view and focus structure, assignment of grammatical functions to the nouns, ordering of each elements in the sentence, and so on (Bock & Levelt, 1994). In Griffin & Bock (2000), the participants only produce regular transitive sentences that start with a subject noun phrase (e.g., The mouse is squirting the turtle with water). It is well known that the subject noun phrases are not only the first element in a transitive sentence in general, but also they tend to attract the speaker’s point of view, and to be salient in the given discourse. Then, it seems fair to say that measurement of the gaze timing of the subject noun phrase by itself does not give us a good grasp of which stage(s) in sentence formulation the gaze timing reflects on.

We conducted an eye-tracking experiment (the visual world paradigm) in which the participants were required to describe a scene by using an existential construction in Japanese where the elements are aligned as a non-canonical
"Location (Locative PP) / Locatum (Nominative NP) / Verb" order. We call that particular word order *non-canonical*, in the sense that a locative PP precedes a nominative NP. We contrasted the results against the gaze data obtained from the same picture but the participants utter another sentence which involves a canonical "Theme (Nominative NP) / Verb" order. The gaze-onset of the first noun phrase in two different word order conditions would give us information about different stages of speech production model.

2. Experiment

2.1 Participants, material, procedure

Students at Hiroshima University (N=20) participated the study. Data from 3 participants were eliminated due to the mechanical or other problems during the experiment; then analyses were made based on the remaining data from 17 participants.

24 pairs of simple line-drawn pictures were prepared as the target pictures for the study. Those pictures are initially selected from the collection (The International Picture Naming Projects / USCD), and further checked by a pilot naming task, so that five native speakers of Japanese labeled the picture with the same name. Among the 24 pairs of pictures, nine pairs were animate entities and other pairs were inanimate. Paring of the pictures were made (e.g., an elephant and fish) so that the one of them is in general heavier than the other. 48 pairs of filler pictures were also selected mostly from the same collection mentioned above.

Two task blocks (Simple and Complex) were prepared for the study. In the Simple Task block, the participants were told to describe what they see on the screen, in the sentence form of "[a heavy object] and [a light object] exist," using the coordinate structure where the heavy object must be uttered first. For instance, when the participants saw a picture of a bucket and a paper clip, they had to utter a sentence such as the one shown in (1), since in general, a bucket is heavier than a paper clip.

(1) baketu to kurippu-ga arimasu

bucket and clip-NOM exist

In the Complex Task block, on the other hand, the participants were asked to describe where the heavier entity locates on the screen. They were required to use the form of sentence as the one shown in (2), in which they have to mention the location of the heavier entity (kurippu-no yoko ‘next to a paper clip’) before they actually mention the locatum object itself (baketu-ga ‘bucket-NOM”).

(2) kurippu-no yoko-ni baketu-ga arimasu

clip-GEN side-at bucket-NOM exist

In the experiment, in all of the target pairs of the pictures, objects were placed side by side, while in some of the filler pictures, items are aligned vertically. The participants were told not to use the words “hidari (left)” or “migi (right)” and encouraged to use the word “yoko (side)” so that they do not get confused from left to right. Each participant performed both tasks, and the order between the two task blocks was counter-balanced. Using 24 pairs of pictures, four lists were created, by counter-balancing (a) where each object appears (either left or right on the screen) and (b) in which task block the pair appears (either Simple or Complex task block).

Stimuli were presented by using E-Prime (Psychology Software Tools, Inc.). An experimental trial begins by pressing a button in the Serial Response Box (see below). A crosshair appears and remains for 500 ms when the participant presses a button, and the pictures show up and remained on screen for 3500 ms. Participants were instructed to describe the scene as soon as possible, using the appropriate sentence form in a given task blocks. Task sessions were preceded by six practice trials, and the experimenter made sure that the participants understood the whole procedure. The response latency of each trial was recorded by the Voice-Key that is part of the Serial Response Box, working with E-Prime. Also, the utterances were recorded by an IC-Recorder so that the experimenter can check whether the participants followed the procedure and made correct utterances.

Eye gaze pattern during the experiment was recorded by
a Tobii 1750 Eye Tracker (Tobii Technology: Stockholm, Sweden). Participants’ gaze data were recorded while the target pictures were shown on the screen. Recording was done with the sampling rate of 50 Hz; in general, 50 data points were recorded per second. Before the experiment begins, the participants received some introduction of the eye-tracking system, and the calibration was done with five points on the screen. Gaze data recording was done, using the calibration information from each participant. Gaze data was recorded and classified as (a) looking at the picture on the left, (b) looking at the picture on the right, (c) looking neither. A small break was available between the tasks, and in general, each task block took about 6 min.

### 2.2 Analysis and Results

First, we will report data other than the gaze measurement. The mean accuracy (i.e., the rate of the trials in which each participant follows the instruction accurately on the target trials) from each participant was 88.2% (SD = 10.1). Major errors came from the trials in which the participant mentioned the heavy object first in the Complex Task block or the non-heavy object first in the Simple Task block (and therefore s/he did not follow the instruction of the task), or the trials in which the participant could not name the object correctly, or s/he could not start naming while the picture was on screen, etc.).

Second, the latency data were submitted to ANOVA, taking the following three factors: Task (Simple or Complex), Location (the heavy object is on the left or right side of the screen), and Task Block (first or second). The results showed that there was a main effect of Task Block, showing that the latency from the task performed in the first half was shorter than that in the second half ($F_1=4.84, p<.05; F_2=6.14, p<.03$). There was no interaction among the three factors. Table 1 shows the mean latencies, SD, and SE classified by the Task Block. Note that the mean latency from the task that has been performed in the first half (whichever the task was) was about 100 ms shorter than that from the task that has been performed in the second half.

Table 1: The Mean Latencies from Each Block

<table>
<thead>
<tr>
<th>Task Block</th>
<th>Mean Latency</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>First half</td>
<td>1616.5 ms</td>
<td>455</td>
<td>34</td>
</tr>
<tr>
<td>Second half</td>
<td>1713.8 ms</td>
<td>463</td>
<td>34</td>
</tr>
</tbody>
</table>

Table 2 shows the mean latencies for the each task condition. Statistically not significant though, the mean latency from the Complex Task condition was slightly longer than that from Simple Task.

Table 2: The Mean Latencies for Each Task

<table>
<thead>
<tr>
<th>Task</th>
<th>Mean Latency</th>
<th>SD</th>
<th>SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Complex</td>
<td>1689.5 ms</td>
<td>455</td>
<td>34</td>
</tr>
<tr>
<td>Simple</td>
<td>1642.5 ms</td>
<td>467</td>
<td>35</td>
</tr>
</tbody>
</table>

Next, we will report the gaze data. As illustrated above, two line drawn pictures were presented side by side in the target trials, and one of the objects was in general heavier than the other. We first calculated the gaze rate (whether the participants look the heavy or non-heavy object) at a given time in each task condition. Figure 1 illustrates the gaze data during the Simple Task, while Figure 2 during the Complex Task.

The overall pattern in the Simple Task condition (in which the participants had to mention the heavy object first) revealed that, at 200 ms after the onset of the picture presentation, the gaze rate for the non-heavy object (dark line) was larger than that for the heavy object (gray line), though the difference was very small. But, soon the gaze rate...
for the heavy object increased, and the pattern remains the same for about 1000 ms. Then, the pattern got reversed. The gaze rate for the non-heavy object was larger than that for the heavy object. The pattern continued at least until 3000 ms after the picture onset. As for the overall pattern in the Complex Task condition (in which the participants had to name the non-heavy object first), as seen in Figure 2, not only what the participants looked at, but also the overall pattern was a lot different, compared with that in the Simple Task condition. At 200 ms after the picture onset, there was a major increase of the gaze rate for the heavy object for a short period of time. Then around 800 ms after the picture onset, the gaze rate for the non-heavy object exhibited a sharp increase. Then finally, after 1400 ms after the picture onset the participants continued to look at the heavy object.

Here, we report the statistical results for the gaze patterns. Data were submitted to ANOVA with repeated measure, taking Task (Simple vs. Complex) and Gaze (Heavy or Non-Heavy) as within-subject factors. Only significant or marginally significant results are reported (Table 3). The statistical results were calculated based on three different time frames: (a) 0 to 1000 ms after the picture onset, (b) 0 to 600 ms, and (c) 0 to 400 ms. In either time frame, there was a significant or marginally significant main effect of Gaze, which seems to be due to the large gaze rates for the heavy object in general, compared with that for the non-heavy object.

Next, we report the interaction between Task and Gaze. Interactions between the two factors were observed in (a) 0 to 1000 ms after the picture onset and (c) 0 to 400 ms after the picture onset, but clearly the two interactions were qualitatively different. First, the interaction observed in 0 to 1000 ms after the picture onset revealed that there was a simple main effect of Gaze in the Simple Task, due to the large gaze rate for the heavy object ($F_1=31.02, p<.01; F_2=14.98, p<.01$). On the other hand, there was no simple main effect of Gaze in the Complex Task ($F_1<1$). Now, as for the interaction observed during 0 to 400 ms after the picture onset, there was a significant simple main effect of Gaze in the Complex Task due to the larger gaze rate for the heavy object ($F_1=5.35, p<.04; F_2=6.26, p<.03$). On the other hand, there was no simple main effect in the Simple Task ($F_1<1$).

### Table 3: Statistical Results of the Gaze Rates

<table>
<thead>
<tr>
<th>Time Frame (ms)</th>
<th>Task</th>
<th>Gaze</th>
<th>Task x Gaze</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$F_1$, $F_2$, $p$</td>
<td>$F_1$, $F_2$, $p$</td>
<td>$F_1$, $F_2$, $p$</td>
</tr>
<tr>
<td>&lt; 1000</td>
<td>13.32 &lt; 0.01 *</td>
<td>16.38 &lt; 0.01 *</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.92 &lt; 0.07 #</td>
<td>18.90 &lt; 0.01 *</td>
<td></td>
</tr>
<tr>
<td>&lt; 600</td>
<td>6.01 &lt; 0.03 *</td>
<td>4.75 &lt; 0.05 *</td>
<td></td>
</tr>
<tr>
<td>&lt; 400</td>
<td>3.57 &lt; 0.08 #</td>
<td>3.80 &lt; 0.07 #</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.94 &lt; 0.06 #</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

3. Discussion and Summary

Recall that we have observed that the response latency for the task, measured by the Voice Key Device, was shorter in the first half of the experiment than the one in the second half of the experiment. Usually, it is likely that the participants response to the trials faster as the experiment progresses, since they got used to the procedure, but in our current experiment, such facilitation was not observed. Also, this seems to have nothing to do with which task condition, either Simple or Complex, the participants engaged with. However, we need to be careful to draw any strong conclusion from this result. First, in our experiment, we did
not eliminate the trials in which the participants uttered disfluencies (i.e., “etto,” “eee,” etc). Therefore, it is not quite clear whether the “recorded” response latencies actually and accurately correspond well to the very onset of the first noun phrase the participants uttered in the experiment. Obviously, we would like to overcome such technical limitations, but that was left for future studies. Furthermore, the participants were asked to mention the heavy object first in the Simple Task condition and the non-heavy object first in the Complex Task condition; counter-balance between the task conditions was not perfect. We obviously plan to run an experiment in which the participants were asked to mention either the heavy object or the non-heavy object first, depending on the task condition they are in. That way, we can obtain a much finer measurement of the response latencies. Therefore, at this moment, it is quite hard to speculate whether and to what extent the response latency reflects any specific stage(s) in the production processes, in fully meaningful ways.

Next we discuss the results of the gaze measurement. As mentioned in Introduction, the production model argued by Bock & Levelt (1994) specifies multiple levels in producing a sentence. For instance, it is claimed that there is a message planning stage in which the speaker decides the point of view or what to be focused in the utterance. Also, it is argued that there is a stage in which grammatical relations and thematic role assignment to noun phrases are determined. Finally, there is a stage in which the speaker decides the word order among the elements in the sentence. We use their model as we discuss the gaze data.

There is a common pattern found in both tasks: a major gaze switching from the heavy to non-heavy object in the Simple Task condition, and one from the non-heavy to heavy object in the Complex Task condition. Such patterns can be interpreted quite naturally that the major gaze shifts reflect the task requirement. In the Simple Task condition, the participants have to mention the heavy object first, and in the Complex Task condition, they have to mention the non-heavy object first. In other words, in the Simple Task condition, the participants gaze the heavy object for about 1000 ms, then, they turn to look at the non-heavy object. In the reverse way, in the Complex Task condition, the participants gaze the non-heavy object for about 1000 ms, then they switch to look at the heavy object to mention them. Following the model by Bock & Levelt (1994), this gaze pattern might reflect the process in which elements in the sentence got linearized or the process in which the linearized materials are converted to sounds to be pronounced. Also, considering the timing of this gaze switching, it could be speculated that the gaze switching occurs 300-400 ms before the participants actually start to utter anything, if we, somewhat seriously, believe that the overall reaction latencies for the task was about 1650-1700 ms after the picture onset. But, as we have discussed above, the sensitivity of the reaction latencies was not satisfactory, then it is hard to draw a strong conclusion for that point.

It is quite noteworthy to observe that the difference of the gaze patterns between the two task conditions was not simply due to the difference on the task requirement in terms of which object to be named first. In fact, the graph in Figure
1 cannot be produced just by switching the heavy and non-heavy objects in Figure 2. In order to highlight that point, we calculated the mean difference gaze rates, as shown in Figure 3.

We can compare how gaze patterns shifted in two different task conditions. In the Complex Task condition, within 50 ms after the picture onset, the gaze rate for the heavy object is much larger than that for the non-heavy object, suggesting that there is a strong motive to look at the heavy object. On the other hand, such a major difference was not found in the Simple Task condition. This observation was confirmed by the significant interaction found in the time frame 0 to 400 ms after the picture onset. Let us examine this finding by the model by Bock & Levelt (1994).

Let us consider how our finding can be fit into the speech production model by Bock & Levelt (1994). In order to perform the task requirement, the participants must first determine which object is heavier. That has a consequence for which object the participants have to mention first. In the Simple Task condition, the order between the objects was reflected the order in the subject noun phrase, as producing a sentence form in (3).

3. \[\text{[Heavy AND Non Heavy]-NOM exist}\]

However, those two objects were treated equally in terms of the properties regarding the message planning level, such as the speaker’s point of view, focus, as well as the thematic role they bear in the sentence. On the other hand, in the Complex Task condition, the decision of which object is heavy will create a major consequence regarding the grammatical property of the sentence of which object will become the subject noun phrase. The participants need to assign the subject role to the heavy object, and the locative role to the non-heavy object in the sentence form in (4).

4. \[\text{[Non Heavy]-SIDE-AT [Heavy]-NOM exist}\]

Also in the Complex Task condition, not the locative expression, but the subject noun phrase is focused, so that producing a sentence in the Complex Task condition requires rather complicated operations in the message planning stage or the stage in which the grammatical function is determined. It could be that the gaze pattern for the heavy object that has been observed in the early time frame can be considered as the cost increase due to the complications in those stages.

Summarizing, the results of the present study showed that eye-tracking measurement during the production of non-canonical sentence pattern, such as Japanese locative sentences, provide useful information for investigating multiple stages assumed by the current speech production model. Investigating whether the cost increase arises from the message planning stage or the grammatical function assignment stage is left for the future study.

4. References


Acknowledgment

This research was supported by a Grant-In-Aid for Scientific Research (B) “Neurocognitive basis for the processing of compound words in spoken language” (PI: Hiromu Sakai, #17320064) by the Japan Society for Promotion of Science, by a Grant-in-Aid for Scientific Research on Priority Areas (System study on higher-order brain functions) “Cortical functions for recursive linguistic computation: An investigation through syntactic and lexical priming” from the Ministry of Education, Culture, Sports, Science, and Technology of Japan (PI: Hiromu Sakai, #18020022), and by a Grant-In-Aid for Scientific Research (B) “An integrated research on eventuality expressions based on Chinese-Japanese contrastive studies” (PI: Shen Li, #19320064) by the Japan Society for Promotion of Science.