Length-based phrase-ordering tendencies as a product of word-order flexibility

Hiroko Yamashita¹ and Tadahisa Kondo²
¹Rochester Institute of Technology, ²NTT Communication Science Laboratories
contact: hiroko.yamashita@rit.edu

Studies report that in both head-initial and head-final languages, speakers tend to produce sentences that place salient words earlier in a sentence [1][2]. In producing sentences with a long phrase, however, the two types of languages diverge: speakers of head-final languages with flexible word-order such as Japanese show the Long-before-short (LbS) tendency, a tendency to position a long phrase before a short phrase [3][4]. The frequency of scrambling a long phrase ahead of a short phrase varies across sentence structures in Japanese [4][5], suggesting that length-based phrase-ordering may be guided by multiple principles. Using the Corpus of Spontaneous Japanese, we investigated whether (i) LbS is motivated by a lighter production load, and (ii) speakers make incremental production or advance planning for (a) forthcoming phrase(s) in producing structures of different degrees of linguistic constraints.

One of the functions of fillers is to signal efforts to formulate utterances [6]. Using fillers such as ee, ano ‘uh, um’ as a measure of production/planning load, we examined global/local production loads and how the local production load interacts with the length of forthcoming and preceding phrases. High Canonical Structure (HCS), the Transitive structure shown in (1), and Low Canonical Structure (LCS), the Locative structure shown in (2), were tested. The length of two underlined phrases and the occurrences of fillers before and within those phrases were analyzed. When the underlined words receive modification and gain in length, the word-order within the pre-nominal modifier clause is mostly flexible in both structures. However in HCS the subject-object is 95% in natural speech and in LCS the order of locative phrase before object and reverse order is approximately equal [4].

(1) High Canonical Structure: Transitive Structure
John-ga  pizza-o  tabeta.  “John ate some pizza”
John-nom pizza-acc ate

(2) Low Canonical Structure: Locative Structure
(John-wa/ga)  teeburu-ni pizza-o oita.  “(John) put the pizza on the table.”
(John-top/nom) table-loc pizza-acc put

The results revealed striking differences between the two structures. In LCS, there were twice more LbS sentences than Short-before-long (SbL). In contrast, the number of LbS and that of SbL structures in HCS were about the same. As for global production load, the Logistic regression analysis revealed that the probability of fillers occurring before and within the two phrases was no difference between SbL and LbS in LCS ($p>.05$), whereas in HCS, fillers in SbL occurred more frequently than those in LbS ($p<.05$). As for advance planning, LCS was incremental in the sense that the length of forthcoming phrases did not influence the occurrences of fillers. In contrast, HCS showed advanced planning because before the first phrase in HCS the filler frequency was influenced by the length of second phrase, suggesting that speakers were aware of the forthcoming long object phrase as they spoke the canonically-motivated subject.

The LbS tendency was observed for a structure that allows flexible word-order within
a clause and within a modifier, and it was unrelated to a production load. In producing a structure with a linguistic constraint (canonicality), LbS tendency was not observed; speakers produced SbL with increased production loads and they showed evidence of advance planning. Overall the results indicate that even within a language speakers’ production processes are adjusted according to how strongly linguistic constraints require the speakers to observe a rigid word-order, and LbS tendency emerges only when speakers have a high level of flexibility in arranging words within a sentence and a modifier. Implications for producing languages with high positional requirements such as English are discussed.

References