Does frequency of occurrence make relative clause processing easier in Japanese?

Atsushi Sato (Hiroshima University), Barış Kahraman (Hiroshima University), & Hiromu Sakai (Hiroshima University)

Reali & Christiansen (2007) conducted a large-scale corpus analysis of English relative clauses and found that frequency of exposure influenced the processing difficulty of relative clauses (RC). The corpus frequency of subject and object relative clauses (SRC and ORC, respectively) was reflected in the result of self-paced reading experiments. Since processing asymmetry of SRCs and ORCs had been reported in a number of different languages, their findings raise an interesting question about the universality of frequency-based account for RC processing. We examined the usage of Japanese SRCs and ORCs in a written Japanese corpus and conducted a self-paced reading experiment. Our results showed that corpus frequency failed to account for existing empirical results on Japanese RCs.

Due to the absence of a large-scale parsed corpus in Japanese, an automated morpheme analysis was conducted on 3 million words from a written Japanese corpus, KOTONOHA (developed by the National Institute for Japanese Language) using Mecab 0.98 (developed by Taku Kubo). NPs that contained modifying clauses were selected by ChaKi Ver.2.1.16 (developed by NAIST). Then, we manually classified them into SRC, ORC, and other structures. The results showed that the number of ORC was marginally larger than that of SRC in Japanese (SRC: 1546; ORC: 1641; $x^2(1)=2.832, p=.091$). If the distributions of SRCs and ORCs were the major factor for the processing asymmetry, it would not explain the results of previous studies, which showed that SRCs are easier to process than ORCs in Japanese (e.g. Miyamoto & Nakamura, 2003). Therefore, we considered animacy of RC head nouns, which plays an important role in RC processing (Mak et al., 2002) and reanalyzed the corpus data. The results showed that frequency of SRCs with animate head nouns was significantly higher than that of ORCs with animate head nouns (SRC: 1019; ORC: 137; $x^2(1)=672.945, p<.01$). The result of corpus counting modified by animacy thus enables us to account for the processing asymmetry between SRCs and ORCs. As for RCs with inanimate head nouns, ORCs with inanimate head nouns occurred more frequent than SRCs with inanimate nouns (SRC: 865; ORC: 1471; $x^2(1)=157.207, p<.01$). The frequency account predicts that ORCs should be read faster than SRCs. Thus, we conducted a self-paced reading experiment using RCs with inanimate head nouns as shown in (1), in order to verify whether the frequency account modified by animacy can account for the processing asymmetry between SRCs and ORCs. However, the results revealed that SRCs were still read faster than ORCs at the relative head region even when animacy of RC head was manipulated in favor of ORCs ($F_1(1,18)=6.241, p<.05; F_2(1,21)=3.520, p=.075$).

The present study shows that the simple frequency in corpora is not reflected in the reading times of relative clauses in Japanese even when animacy is taken into consideration. Taken together, corpus frequency alone cannot explain the result of existing empirical research on Japanese relative clauses. To investigate how experience reflected in corpora plays a role in sentence processing and to understand how the human processor uses experience, further research is needed on a wide variety of languages in the world.

(1a) SRC sigikai-no iinkai-o kouseisi-ta senmonbukai-ni-wa the city council-GEN committee-ACC form-PAST section-at-TOP “the section that formed the city council committee …”

(1b) ORC sigikai-no iinkai-ga kouseisi-ta senmonbukai-ni-wa the city council-GEN committee-NOM form-PAST section-at-TOP “the section that the city council committee formed …”