The relation between information-theoretic measures and reaction times of syntactic prediction during L2ers' sentence processing

Euhee Kim[a], Myung-Kwan Park[b] and Hye-Jin Seo[b]
[a]Shinhan University, [b]Dongguk University
parkmk@dgu.edu

This paper investigates how L2 speakers predict upcoming syntactic structure based on a newly received word during sentence processing. One way of predicting appropriate syntactic structure is *probabilistic inference*. For instance, readers are most likely to postulate the forthcoming syntactic structure predicted relying on their experiences with the language they use (DeLong, Urbach, & Kutas, 2005; Smith & Levy, 2013). Linzen and Jaeger (2016) employed subcategorization frame(s), which are the option(s) that a verb takes for their complement(s) to quantify the information-complexity metrics of a given word that reflects the processing difficulty during sentence processing. They used three kinds of the information-complexity metrics; surprisal, entropy, and entropy reduction. Surprisal is related to the unexpectedness of a given word after a string of preceding words; entropy is the uncertainty about upcoming syntactic structure; and entropy reduction is the fluctuation of the uncertainty from one word to the next. The latter two metrics in their L1 study were estimated from a probabilistic context-free grammar (PCFG) based on Penn Treebank corpus. Linzen and Jaeger's study demonstrates that the surprisal and the entropy reduction about the full prediction, which is the prediction of the whole syntactic structure of the sentence, has an impact on processing difficulty during sentence processing in English native speakers. In particular, English L1 speakers predict the syntactic structure for the upcoming context promptly after reading a verb.

The present study replicated the earlier study by Linzen and Jaeger (2016). However, it differs from Linzen and Jaeger's study in two respects. First, the experiment was conducted for Korean L2 learners of English. Second, the PCFG was estimated from a corpora of English textbooks published in Korea from 2001 to 2009 (consisting of 2,750,000 word tokens). It is expected that the PCFG from the language materials that L2 English learners are familiar with can provide a better account for their behavior in sentence processing.

All the experiments were run on Ibex Farm, which is the web-based experimental presentation platform (Drummond, 2013). Each sentence was presented in a word-by-word, self-paced moving window. The experimental stimuli consisted of 30 sentence pairs adapted from Linzen and Jaeger (2016). Each pair involved one type of sentences with the complementizer added after a main verb, *The men discovered that the island had been invaded by the enemy*; and the other without it, *The men discovered the island had been invaded by the enemy*. 35 undergraduates from Dongguk University in Seoul, Korea participated in this study (13 males, mean age = 25.2, range = 19~30). The data was analyzed by following the statistical methods in Linzen and Jaeger's study. The results show that a significant effect of Entropy on RTs in the ambiguous region was detected (β=-0.100, SE=0.040, t=2.454, p<.05): Higher entropy was correlated with longer RTs in the ambiguous region. In addition, a significant effect of Sentence Condition (ambiguous vs. unambiguous sentence) at the ambiguous (β=-0.254, SE=0.088, t=-2.883, p<.01) and the disambiguating regions (β=0.340, SE=0.160, t=2.125, p<.05) was found: RTs were longer in the ambiguous regions of the ambiguous than of the unambiguous sentences, but were longer in the disambiguating region of the unambiguous than of the ambiguous sentences, supporting the entropy reduction hypothesis. As a result, Korean L2 learners of English start to predict the upcoming syntactic structure in the post-verb, ambiguous and disambiguating regions when they are provided with more solid evidential sources such words and structures for syntactic (re)analysis.

Overall, Korean L2 learners of English differ from English native speakers in sentence processing. Still, the behavioral aspects of sentence processing by Korean L2 learners of English can be explained by the predictions that the entropy reduction and the surprisal hypotheses make, especially at the ambiguous and the disambiguating regions.

References

DeLong, K. A., Urbach, T. P., & Kutas, M. (2005). Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nature Neuroscience, 8(8), 1117–1121.

Drummond, A. (2013). Ibex farm. Retrieved February 12, 2017.

Linzen, T., & Jaeger, T. F. (2016). Uncertainty and expectation in sentence processing: Evidence from subcategorization distributions. Cognitive Science, 40(6), 1382-1411.

Smith, N. J., & Levy, R. (2013). The effect of word predictability on reading time is logarithmic. Cognition, 128(3), 302–319.

Appendix A. The interest region of target sentences

Words	The men	discovered	(that)	the island	had been invaded	by the enemy.
Region	Subject	Verb	That	Ambiguous	Disambiguating	Rest