Syntactic probability influences duration

Claire Moore-Cantwell (University of Massachusetts, Amherst)
cmooreca@linguist.umass.edu

Duration; Phonetic reduction; Structural probabilities; Training

Spoken words reduce more the more predictable, or redundant, they are given their context (Aylett and Turk, 2004; Jaeger, 2010). Because of this, knowledge of predictability must have a place in the production system. This paper demonstrates, more particularly, that the production system must have access to information about the probability of syntactic structures which is not retrievable from the lexicon.

This study builds on work by Gahl and colleagues, which demonstrates that the predictability of the syntactic structure of a verb's complement given the verb's subcategorization bias affects the duration of words in that structure in a production task (Gahl and Garnsey, 2004) and in a spoken corpus (Tily et al., 2009). This work demonstrates that speakers use the probability of a structure given its context in the production process. This work does not address the question of where these probabilities are stored. Because they examine probabilities of complement structures which are conditioned on the identity of the verb, one option is that the probabilities of complement structures are affiliated with each verb's lexical entry.

In the current study, I aim to demonstrate that speakers know and use online probabilities over syntactic structures which are not conditioned by a particular lexical item, and cannot be affiliated with a particular lexical entry. I will separate the probability of a structure from its inherent complexity by manipulating the probabilities of structures within the experiment. The results of this show the need for some extra storage mechanism for probabilities over structures, which must be accessible to the production system.

For animate extracted constituents, subject extraction is much more common in corpora than object extraction (Roland et al., 2007). The present study measured the duration of the clefted constituent, not including pauses, in experimentally-induced pronounciations of subject- and object-extraction clefts in English.

Object extraction (OE) is more difficult to process than subject extraction (SE) (e.g. Gibson, 1998). In order to disentangle the probability of OE structures from their inherent difficulty, the probability of each structure was manipulated within the experiment. Two groups of participants each saw: First, two of each type of cleft; second, eight of either SE or OE clefts; and finally, another two of each type of cleft.

Before training, the clefted constituent was longer in OE clefts (mean 407ms) than in SE clefts (370ms, $t=2.4$, $p=.02$). After OE training, this difference was no longer present (OE: 385ms, SE: 397ms), but it was still present after SE training (OE: 448ms, SE: 388ms). A linear mixed effects regression model was fitted to the duration data from the two training conditions, with extraction type and training condition and their interaction as the fixed effects, and subject and item as random effects. A significant effect ($p=.04$) was found for the interaction of extraction type and training condition.

These results demonstrate that speakers track probabilities over different types of clefts, and thus know probabilities over syntactic structures not affiliated with any particular lexical entry, and use them in production.

(1)**Subject-extraction:** It was Edward who (t) scammed Melvin out of some money.

(2)**Object-extraction:** It was Edward who Melvin scammed (t) out of some money.

M. Aylett and A. Turk. The smooth signal redundancy hypothesis. Language and Speech, 47:31–56, 2004.

S. Gahl and S. M. Garnsey. Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronounciation variation. Language, 80(4):748–775, 2004.

E. Gibson. Linguistic complexity: locality of syntactic dependencies. Cognition, 68:1–76, 1998.

T. Florian Jaeger. Redundancy and reduction: Speakers manage syntactic information density.

Cognitive Psychology, 61(1):23–62, 2010.

D. Roland, F. Dick, and J. L. Elman. Frequency of basic english grammatical structures: A corpus analysis. Journal of Memory and Language, 57:348–379, 2007.

H. Tily, S. Gahl, I. Arnon, N. Snider, A. Kothari, and J. Bresnan. Syntactic probabilities affect pronounciation variation in spontaneous speech. Language and Cognition, pages 147–165, 2009.