Speakers sacrifice some (of the) precision in conveyed meaning to accommodate robust communication

Judith Degen and T. Florian Jaeger
University of Rochester

The process of encoding an intended meaning into a linguistic utterance is well-known to be affected by production pressures (Bock & Warren, 1985; Branigan et al, 2008; Ferreira & Dell, 2000). For example, there is mounting evidence that speakers prefer to distribute information uniformly across the linguistic signal, which maximizes information transmission (Uniform Information Density (UID), Aylett & Turk 2004; Levy & Jaeger 2007) and ensures robust communication. Support for UID comes from apparently meaning-equivalent alternating forms: speakers prefer the form with more linguistic signal when the encoded meaning is contextually unexpected (e.g. different word pronunciations, object drop, auxiliary-contraction, optional *that*-mentioning, Bell et al. 2003; Frank & Jaeger 2008; Resnik 1996; Jaeger 2006). However, it is an open question whether and to what extent precision in encoding an intended meaning trades off with the preference to distribute information uniformly when two linguistic forms are similar, but arguably not meaning-equivalent.

We asked whether the choice between two seemingly non-meaning-equivalent forms can be affected by UID. Specifically, we investigated the choice of simple versus partitive "some", (1a, b) respectively:

(1a) Alex ate some cashews.
(1b) Alex ate some of the cashews.

These two variants are generally assumed to differ in meaning, though there is no consensus as to the specific difference. One proposed restriction on the partitive is that the embedded NP should be definite or specific (e.g. Jackendoff 1977, Ladusaw 1982). Further, the partitive is more likely to give rise to a scalar implicature (Degen et al 2009). However, the two forms also arguably overlap in meaning. We assess this shared information, I(SOME), as the logarithm-transformed summed contextual probability of simple and partitive "some" (cf. Shannon 1948). If speakers are willing to tolerate deviation from subtler aspects of the intended meaning in order to safely communicate the core meaning SOME, UID predicts that speakers prefer partitive over simple "some", the higher I(SOME), i.e. the more unpredictable SOME is in context.

We extracted 1362 cases of "some"-NPs from spontaneous speech data. We used mixed logit models to assess the effects of discourse givenness of the NP head noun (which served as a proxy for the hypothesized meaning difference) and of three independent estimates of the information of SOME: I(SOME) conditioned on (a) the previous word ("ate" above), (b) the NP head ("cashews" above), and (c) the NP's grammatical function (direct object above). For all three predictability measures, speakers preferred the longer partitive form, the higher the information of SOME. Meaning differences also affected form choice: given referents were more likely to be used in the partitive. These effects are additive in that they hold independently of each other and while simultaneously controlling for other measures known to affect speakers' preferences, such as the

animacy, frequency, and bigram predictability of the head noun, as well as random effects of speakers.

This result is surprising given that simple and partitive "some" are assumed to at least partially differ in the message they encode. The data suggest that even when two forms do not encode the same (but a similar enough) message, speakers may sacrifice precision in meaning in order to robustly communicate a core meaning. This would have far-reaching consequences for theories of meaning, form choice, pragmatics, and processing.

References

Aylett, M.P. and A. Turk (2004). The smooth signal redundancy hypothesis: A functional explanation for relationships between redundancy, prosodic prominence, and duration in spontaneous speech. *Language and Speech, 47(1),* 31 - 56.

Bell, A., D. Jurafsky, E. Fosler-Lussier, C. Girand, M. Gregory, and D. Gildea (2003). Effects of disfluencies, predictability, and utterance position on word form variation in English conversation. *Journal of the Acoustical Society of America, 113(2),* 1001 – 1024.

Bock, J. K. and R. K. Warren (1985). Conceptual accessibility and syntactic structure in sentence formulation. *Cognition, 21(1),* 47 – 67.

Branigan, H.P., Pickering, M.J, & Tanaka, M. (2008). Contributions of animacy to grammatical function assignment and word order during production. *Lingua, 118,* 172-189.

Degen, J., P. A. Reeder, K. Carbary, and M. K. Tanenhaus (2009). Using a novel experimental paradigm to investigate the processing of scalar implicatures. *Experimental Pragmatics 2009*, Lyon, April 23 – 35.

Ferreira, V. S. and G. S. Dell (2000). The effect of ambiguity and lexical availability on syntactic and lexical production. *Cognitive Psychology, 40,* 296 – 340.

Frank, A. and T. F. Jaeger (2008). Speaking rationally: Uniform information density as an optimal strategy for language production. In *The 30th annual meeting of the Cognitive Science Society (CogSci08)* (pp. 939 - 944), Washington, DC.

Jackendoff, R. (1977). X-bar Syntax: *A Study of Phrase Structure*. Cambridge, MA: MIT Press.

Jaeger, T. F. (2006). Redundancy and syntactic reduction in spontaneous speech. Ph.D. thesis, Stanford University, Stanford, CA.

Ladusaw, W. (1982). Semantic constraints on the English partitive constructions. In D. Flickinger, M. Macken, & N. Wiegand (eds.), *Proceedings of WCCFL 1* (p. 231 - 242).

Levy, R. and T. F. Jaeger (2007). Speakers optimize information density through syntactic reduction. In B. Schlökopf, J. Platt, and T. Hoffman (Eds.). *Advances in neural information processing systems (NIPS)* (Vol. 19, pp. 849 - 856). Cambridge, MA: MIT Press.

Resnik, P. (1996). Selectional constraints: An information-theoretic model and its computational realization. *Cognition, 61,* 127 – 159.

Shannon, C. (1948). A mathematical theory of communications. *Bell Systems Technical Journal, 27(4),* 623 – 656.