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

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). 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 the intended message trades off with the preference to distribute information uniformly when two linguistic forms are similar, but arguably not meaning-equivalent.

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

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

These two variants are generally assumed to differ in meaning, although they also arguably overlap in meaning. We assess this shared information, I(SOME), as the logarithm-transformed summed contextual probability of bare 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 bare "some", the higher I(SOME)..

We extracted 1362 cases of "some"-NPs from spontaneous speech data. We used linear mixed models to assess the effects of three independent estimates of the information of SOME: I(SOME) conditioned on (a) the previous word ("ate" above), (b) the NP head ("chard" above), and (c) the NP's grammatical function (direct object above). For all three measures, speakers preferred the longer partitive form, the higher the information of SOME. 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, givenness, frequency, and bigram predictability of the head noun, as well as random effects of speakers.

While supporting UID, the result is surprising, given that bare and partitive "some" are assumed to at least partially differ in the message they encode. The data are thus consistent with two possible scenarios. Either the bare and partitive form are (in general) meaning-equivalent, or we have provided evidence that even when two forms do not encode the same (but a similar enough) message, speakers may sacrifice precision in meaning for increased processing efficiency. This would have far-reaching consequences for theories of meaning, form choice, pragmatics, and processing. We discuss these findings in light of recent Bayesian and game-theoretic approaches to pragmatics and expression choice (Frank et al, 2009, Franke, 2009).

481 words

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.

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.

Frank, M. C., N. D. Goodman, P. Lai, and J. B. Tenenbaum (2009). Informative communication in word production and word learning. In *Proceedings of the 31st Annual Meeting of the Cognitive Science Society*.

Franke, M. (2009). Signal to act: Game Theory in Pragmatics. Ph.D. thesis, Universiteit van Amsterdam.

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

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.