

Donnerstag  
09.03.2017  
12:15 – 12:45  
B3 2, 0.03

---

## Mathematical barriers to channel coding-based explanations of variation in language production

---

Eric Meinhardt  
University of California, San Diego  
[emeinhardt@ucsd.edu](mailto:emeinhardt@ucsd.edu)

In the pursuit of computational-level analyses of language production, three influential theories (Genzel & Charniak, 2002; Aylett & Turk, 2004; Jaeger, 2010) have offered information theoretic explanations of variation. The common eponymous idea of all three theories is roughly that, from timestep to timestep the amount of information transmitted in natural language signal sequences should be approximately *constant*, *ceteris paribus*. They claim this property of language is implied by a rationality assumption about human communication and the *noisy channel coding theorem* ('NCCT') of Shannon (1948). I demonstrate that this property is not, in fact, a consequence of the NCCT through two lines of reasoning.

First, these theories have not appreciated ways in which natural language represents a channel with significantly different mathematical properties from Shannon's, or that these differences require much more sophisticated and possibly novel proofs. Second, Shannon's results and the component mathematical objects do not (presently) have clear relevance to incremental online choices in a language-like encoding. For example, *entropy rates* are asymptotic in the length of the signal sequence, whereas natural language features relatively short signals. As well, in Shannon's channel, there are no choice points for senders – each message maps to one signal. Finally, channel capacity is a *maximum over all possible distributions on the source*. This is of clear interest to an engineer without knowledge of the future source distribution, but it is not immediately obvious why or how which alternative distributions over word or sound sequences are a relevant part of an upper bound on the performance of a speaker in a particular situation.

**References:** • Aylett, M. & Turk, A. (2004). The smooth signal redundancy hypothesis. *Language and Speech*, 47 (Pt 1), 31–56. • Genzel, D. & Charniak, E. (2002). Entropy Rate Constancy in Text. *Proceedings of the 40th Annual Meeting of the ACL*, 199–206. • Jaeger, T.F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61 (1), 23–62. • Shannon, C. (1948). A Mathematical Theory of Communication. *Bell System Technical Journal*, 27 (3), 379–423.