### Information-Theoretic Approaches to Language

**Length and Informativity**

**General hypothesis:** Word lengths reflect a pressure for efficient communication. Less informative words should be shorter and more reduced (Aylett & Turk, 2006). More informative words should be longer and more distinctive.

**Informativity is Surprisal:** $-\log p(w)$

**Example:** The fundamental problem of communication is that of reproducing at one point either exactly or approximately a message selected at another point. Frequently these messages have meaning that they refer to, or are correlated according to some context.

(Shannon, 1948)

In an efficient code, you can minimize your expected code length by making shorter the words shorter and informative words longer (Shannon, 1948, 1951).

---

**Average surprisal in context predicts word length**

**Efficiency and word length:** Word lengths are determined by the average amount of information conveyed by a word in context. Information content (average surprisal in context) predicts word length better than frequency of usage (Bell, 1936, 1949; Plantadosi, 2011).

---

**Surprisal in context predicts word duration**

**Efficiency and word duration:** Controlling for length, words that are more predictable in a certain context are spoken faster (Bell et al., 2009). Their vowels are also more reduced (Aylett & Turk, 2006). Also, speakers are more likely to use abbreviated forms (Mahowald et al., 2012).

---

### Efficient Communication Forward and Backward

Richard Futrell (MIT), Kyle Mahowald (MIT), Steve Plantadosi (University of Rochester), and Edward Gibson (MIT)

Contact: futrell@mit.edu

---

**Length Correlates with Backward Surprisal**

### Crosslinguistic Corpus Study

Using the paradigm of Plantadosi et al. (2011), we correlate word length with average surprisal as measured in the Google Web N-Gram corpora. We include more diverse languages than the previous study, and we measure the effect of average backward surprisal.

---

**Word Shortening Does Too**

For words that have short and long variants, such as chimp/chimpanzee, the shorter variant has lower average surprisal—measured with both forward and backward surprisal. This indicates that speakers prefer shorter forms when a word is predictable given following material.

---

### Conclusions

The relevance of preceding context for surprisal seems intuitively obvious: it is the order in which linguistic units are sent and received. However, when viewed in the context of the whole processes of production and comprehension, it is less clear that forward context is what matters.

In producing an utterance, a speaker has knowledge of what she is going to talk about, so the decision to produce each word is not conditioned solely on the words previously produced. Similarly, a comprehend is trying to understand a whole utterance, if one part of the utterance doesn’t make sense when it initially hears it, that is fine if following parts of the utterance make it sense in retrospect. We hope the demonstrated effect of backward surprisal motivates a move away from models that use preceding context only.

---

### References


Bell, S., & Turk, D. (2009). Word lengths are determined by the average amount of information conveyed by a word in context. Information content (average surprisal in context) predicts word lengths better than frequency of usage. Journal of Memory and Language, 60(3), 308–326.

Bell, S., & Turk, D. (2010). The relevance of preceding context for surprisal seems intuitively obvious: it is the order in which linguistic units are sent and received. However, when viewed in the context of the whole processes of production and comprehension, it is less clear that forward context is what matters. In producing an utterance, a speaker has knowledge of what she is going to talk about, so the decision to produce each word is not conditioned solely on the words previously produced. Similarly, a comprehend is trying to understand a whole utterance, if one part of the utterance doesn’t make sense when it initially hears it, that is fine if following parts of the utterance make it sense in retrospect. We hope the demonstrated effect of backward surprisal motivates a move away from models that use preceding context only.