



More than words - speakers are sensitive to the frequency of four-word phrases

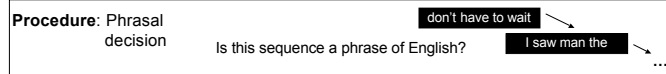
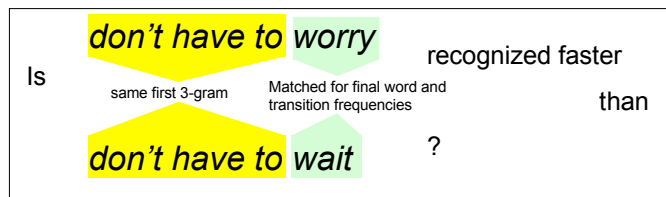
Inbal Arnon¹ (*inbalar@stanford.edu*) & Neal Snider² (*nsnider@bcs.rochester.edu*)

¹Department of Linguistics, Stanford University, ²Department of Brain & Cognitive Sciences, University of Rochester



Abstract

What are the building blocks of language? In a series of studies, we show that speakers are faster to recognize higher frequency compositional 4-word phrases even when the frequency of the substrings is controlled. This happens across the frequency range and is better modeled using a continuous measure of frequency and not a binary one (high/low). These findings suggest that (a) speakers store information about multi-word phrases, (b) not just very frequent ones or non-compositional ones, and (c) every time counts: the more often a sequence occurs, the faster it is recognized. We discuss implications for models of the lexicon.



Sample materials

	High	Low
High bin:	a lot of places 10.45 i don't know why 35.15	a lot of days 0.55 i don't know who 7.00
Mid bin:	it takes a lot 7.35	It takes a little 1.25
Low bin:	all over the country 9.55 i want to say 3.6	all over the house 0.85 i want to sit 0.2

Result summary

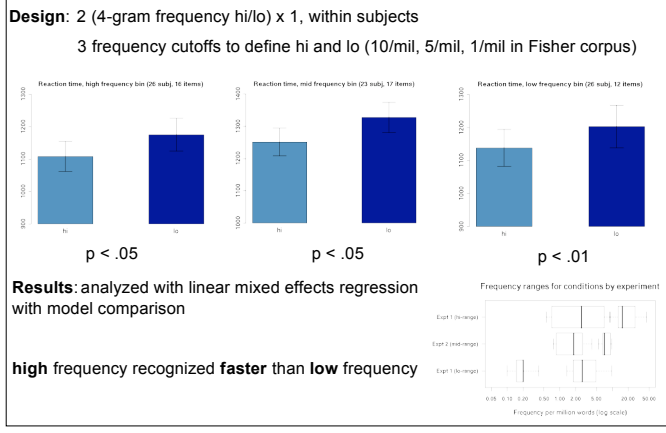
- More frequent compositional phrases processed faster than low frequency ones
- Across the frequency continuum
- Actual frequency is a better predictor than hi/lo bin

Introduction and Background

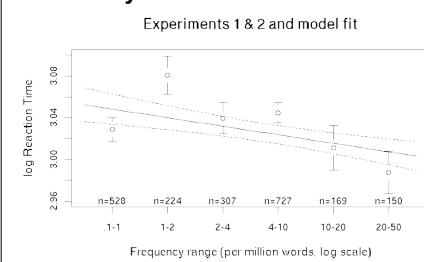
Do speakers represent compositional multi-word phrases?

- Contrast generative and usage-based views of language.
 - Generative: atomic elements and rules or constraints (Chomsky, 1981; Jackendoff, 2002)
 - Usage-based: Complementary representations at different grain-sizes (Bod, 1998; Bybee, 1998; Goldberg, 2006; Langacker, 1987)
- Add multi-word phrases to the units speakers are sensitive to
 - Expand the units whose frequency influences processing (words, constructions, etc.)
 - Call for lexical models that can capture co-occurrence patterns beyond the bigram
- How to assess representation?
 - Look for phrase-frequency effects when substring frequency is controlled for
 - Speakers have a memory trace of the complex unit (drawing on morphological literature on regular inflected forms (Taft 1979, Baayen, 2007))
- Previous findings
 - Few studies look beyond the bigram. Of those, most look only at very frequent phrases (Bybee & Schiebman, 1999), or don't control fully for substring frequency (Underwood et al., 2004)
- What we did
 - Two reaction time experiments, three frequency bins (different cutoff points between high and low)
 - Meta-analysis of reaction times

Experimental Results



Meta-analysis



Analysis: linear mixed effects regression with model comparison. **Dependent variable:** log(reaction time)

Controls: log(unigram3), log(unigram4), log(bigram1), and log(trigram1), block-order, number-of-letters, random effects of subject and item

Results for factors of interest:

- Log (frequency) is better fit than a binary measure.
 - log(4gram frequency) $p < .005$
 - frequency > 10/mil $p > .3$

Implications

- Speakers store information about compositional complex units
 - Doesn't entail that those units are accessed as a whole or that they are unanalyzed
- Processing latencies are affected by the frequency of fully realized multi-word phrases
- Need a language model to account for that
 - Theory of representation where speakers represent linguistic units at multiple grain-sizes (words, multi-word phrases, constructions)
 - An implementation that can capture and predict when such frequency effects arise (e.g. Bod, 2006)

Future work

- How detailed is their internal structure?
- The relation of the smaller and larger units?
- How do different grain-sized units interact in processing?

References

Arnon, I., & Snider, N. (2015). More than words: Speakers are sensitive to the frequency of four-word phrases. *Journal of Experimental Psychology: Applied*, 21(1), 1-15.

Bod, R. (1998). *Usage-based models of language*. Cambridge, MA: MIT Press.

Bybee, J., & Schiebman, R. (1999). *Morphology: A comprehensive guide*. London: Routledge.

Chomsky, N. (1981). *Lectures on government and binding*. Cambridge, MA: MIT Press.

Goldberg, A. (2006). *Usage-based models of language*. Cambridge, MA: MIT Press.

Jackendoff, R. (2002). *Foundations of language: Computational principles of natural language*. Cambridge, MA: MIT Press.

Langacker, R. (1987). *Concept, structure, and language: A cognitive grammar*. Cambridge, MA: MIT Press.

Underwood, J. L., & Snider, N. (2004). *Frequency effects in word recognition*. In J. L. Underwood & N. Snider (Eds.), *Frequency effects in word recognition* (pp. 1-15). Cambridge, MA: MIT Press.

Taft, J. (1979). *Morphological processing in reading: A case study*. In J. R. Hayes (Ed.), *Reading: A cognitive process* (pp. 15-32). Hillsdale, NJ: Erlbaum.

Underwood, J. L., & Snider, N. (2004). *Frequency effects in word recognition*. In J. L. Underwood & N. Snider (Eds.), *Frequency effects in word recognition* (pp. 1-15). Cambridge, MA: MIT Press.