

RETHINKING ZIPF'S FREQUENCY-MEANING RELATIONSHIP: IMPLICATIONS FOR THE EVOLUTION OF WORD MEANING

BODO WINTER

*Cognitive and Information Sciences, University of California, Merced
Merced, U.S.A.
bodo@bodowinter.com*

DAVID H. ARDELL

*Quantitative Systems Biology, University of California, Merced
Merced, U.S.A.
dardell@ucmerced.edu*

In 1945, George Zipf discovered that more frequent words have more meanings listed in dictionaries than less frequent words. For example, a look at Google Ngram (Michel et al., 2011) reveals that the verb “follow” has a relative frequency of ~0.01%, 100 times more frequent than the verb “adorn” (0.0001%). On macmillandictionary.com, “follow” has 14 senses, “adorn” only 1.

Numerous proposals within cognitive and functional linguistics argue that words acquire new meanings through contextual re-interpretation (e.g., Evans & Wilkins, 2000; Traugott & Dasher, 2001). If a word occurs often enough in a novel context, meaning imbued by the context may become conventionalized. We argue that Zipf's frequency-meaning relationship is in fact reflective of this fundamental mechanism by which semantic systems evolve over time.

Word frequency is positively correlated with contextual diversity (Adelman et al., 2006), a measure of the number of different contexts that a word occurs in. In this paper, we use contextual diversity measures from Google Ngram (the number of different Ngrams a word occurs in) as well as from SUBTLEX-US (the number of different movies a word occurs in, Adelman et al., 2006). We take dictionary data from the MacMillan Online Dictionary as well as from WordNet (Miller, 1995; Fellbaum, 1998).

We perform a series of statistical analyses (negative binomial regressions) of the frequency-meaning relationship, showing that if contextual diversity is

controlled for, word frequency is, in fact, not positively correlated with the number of dictionary senses. Zipf's frequency-meaning relationship is driven by contextual diversity, consistent with cognitive/functional accounts of the evolution of word meaning.

This general result holds for a number of different ways of operationalizing contextual diversity and sense counts. Moreover, previous findings on the distribution of dictionary meanings can be shown to be consistent with our approach: Verbs have more dictionary senses than nouns (Fellbaum, 1990), which our data suggests is due to verbs being more contextually diverse. We also show that the more morphemes a word has (adding semantic specificity), the weaker the relationship between contextual diversity and senses: Words with many morphemes (e.g., "antidisestablishmentarianism") are less prone to acquiring new senses in novel contexts, presumably because they are too specific to be reinterpreted in novel contexts. Finally, we show that contextual diversity from 200 years ago (Google Ngram) predicts present dictionary senses.

Our results suggest a re-interpretation of Zipf's frequency-meaning relationship, and they suggest avenues for novel computational models of evolutionary semantics. We outline implications for models of the evolution of vocabularies (Smith, 2004), as well as for models of Zipfian distributions.

References

- Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, 17(9), 814-823.
- Evans, N., & Wilkins, D. (2000). In the mind's ear: The semantic extensions of perception verbs in Australian languages. *Language*, 76, 546-592.
- Fellbaum, C. (1990). English verbs as a semantic net. *International Journal of Lexicography*, 3(4), 278-301.
- Fellbaum, C. (1998). *WordNet: An electronic lexical database*. Cambridge: MIT Press.
- Michel, J. B., Shen, Y. K., Aiden, A. P., Veres, A., Gray, M. K., Pickett, J. P., Hoiberg, D., Clancy, D., Norvig, P., Orwant, J., Pinker, S., & Nowak, M. A., Aiden, E. L. (2011). Quantitative analysis of culture using millions of digitized books. *Science*, 331, 176-182.
- Miller, G. A. (1995). WordNet: a lexical database for English. *Communications of the ACM*, 38, 39-41.
- Smith, K. (2004). The evolution of vocabulary. *Journal of Theoretical Biology*, 228(1), 127-142.
- Traugott, E.C., & Dasher, R. B. (2001). *Regularity in semantic change*. Cambridge: Cambridge University Press.
- Zipf, G. K. (1945). The meaning-frequency relationship of words. *The Journal of General Psychology*, 33, 251-256.