The Principle of Least Effort and diachronic lexical change

"The magnitude of words tends...to stand in an inverse...relationship to the number of occurrences." This is the Principle of Least Effort, or PLE (Zipf, 1935). This inverse relationship between word length and frequency has since been observed in a range of natural languages (Sigurd et al., 2004; Piantadosi et al., 2011; Ferrer-i Cancho and Hernández-Fernández, 2013), and possibly also animal communication systems (Ferrer-i Cancho et al., 2013) and computer programming languages (Ellis and Hitchcock, 1986).

A closely related principle, Uniform Information Density (UID), states that word length is inversely proportional to its probability *in context*, rather than overall probability in a corpus (Levy and Jaeger, 2007); i.e., word length is proportional to its information content, where information content is measured as the −log probability-in-context of a word. Like the PLE, UID is observed in a range of languages (Piantadosi et al., 2011). Moreover, there is evidence it applies beyond orthographic word length, modulating the use of English complementisers (Jaeger, 2010), morphological inflection (Fedzechkina et al., 2012) and derivation (Mahowald et al., 2013), and phonetic reduction (Gahl and Garnsey, 2004).

Given the spectrum of communication systems that exhibit the PLE and UID, we ask: what causes these systems to align utterance length and probability in these ways? It is clear these principles optimise communicative efficiency. Specifically, PLE ensures that the maximum information is transmitted in the minimum utterance length, while avoiding ambiguity. UID ensures that the maximum information is communicated through a noisy channel. Thus, the explanation may be that languages evolve under a pressure for communicative efficiency; as the average predictability of a word changes over time, speakers adjust its length accordingly, with a bias for aligning shorter words with more predictable meanings and vice-versa (Zipf, 1935; Mahowald et al., 2013). However, an alternative explanation (hypothesis 2) states that PLE and UID may be default states that have nothing to do with optimisation by speakers, since even randomly generated 'languages' display these relationships (Moscoso Del Prado Martin, 2013; Ferrer-i Cancho and Moscoso del Prado Martin, 2012).

If the mechanisms explaining PLE and UID can be traced to a speaker bias for efficiency, as hypothesis 1 states, then we should observe the predicted diachronic effect. However, prior work is based solely on synchronic data. We conduct the first investigation of this hypothesis using a diachronic dataset, the Google Books N-grams corpus. Using synonymous or near-synonymous word pairs differing in length, e.g. derivationally-related clipped pairs such as 'info' and 'information', we find that, as a meaning becomes more frequent over time in the corpus, the relative frequency of the shorter form tends to increase, and vice-versa. This is predicted by hypothesis 1, but unexplained by hypothesis 2. Our results therefore suggest that speaker biases towards communicative efficiency do play a role in explaining the PLE.

References

Ellis, S. R. and Hitchcock, R. J. (1986). The emergence of zipf's law: Spontaneous encoding optimization by users of a command language. *Systems, Man and Cybernetics, IEEE Transactions on*, 16(3):423–427.

Fedzechkina, M., Jaeger, T. F., and Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, 109(44):17897–17902.

Ferrer-i Cancho, R. and Hernández-Fernández, A. (2013). The failure of the law of brevity in two new world primates. statistical caveats. *Glottotheory International Journal of Theoretical Linguistics*, 4(1):45–55.

Ferrer-i Cancho, R., Hernández-Fernández, A., Lusseau, D., Agoramoorthy, G., Hsu, M. J., and Semple, S. (2013). Compression as a universal principle of animal behavior. *Cognitive Science*, pages n/a–n/a.

Ferrer-i Cancho, R. and Moscoso del Prado Martin, P. E. (2012). Information content versus word length in random typing. *JSTAT*.

Gahl, S. and Garnsey, S. M. (2004). Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation. *Language*, page 748775.

Jaeger, T. F. (2010). Redundancy and reduction: Speakers manage syntactic information density. *Cognitive Psychology*, 61(1):23–62.

Levy, R. and Jaeger, T. F. (2007). Speakers optimize information density through syntactic reduction. *Advances in neural information processing systems*, 19.

Mahowald, K., Fedorenko, E., Piantadosi, S. T., and Gibson, E. (2013). Info/information theory: Speakers choose shorter words in predictive contexts. *Cognition*, 126(2):313–318.

Moscoso Del Prado Martin, F. (2013). The missing baselines in arguments for the optimal efficiency of languages.

Piantadosi, S. T., Tily, H., and Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, 108(9):3526–3529.

Sigurd, B., Eeg-Olofsson, M., and Van Weijer, J. (2004). Word length, sentence length and frequency-Zipf revisited. *Studia Linguistica*, 58(1):3752.

Zipf, G. K. (1935). *The psycho-biology of language*, volume ix. Houghton, Mifflin, Oxford, England.