# THE EVOLUTION OF ZIPF'S LAW OF ABBREVIATION

JASMEEN KANWAL, KENNY SMITH, JENNIFER CULBERSTON, SIMON KIRBY

*Centre for Language Evolution*
*University of Edinburgh*
*s1252181@sms.ed.ac.uk*

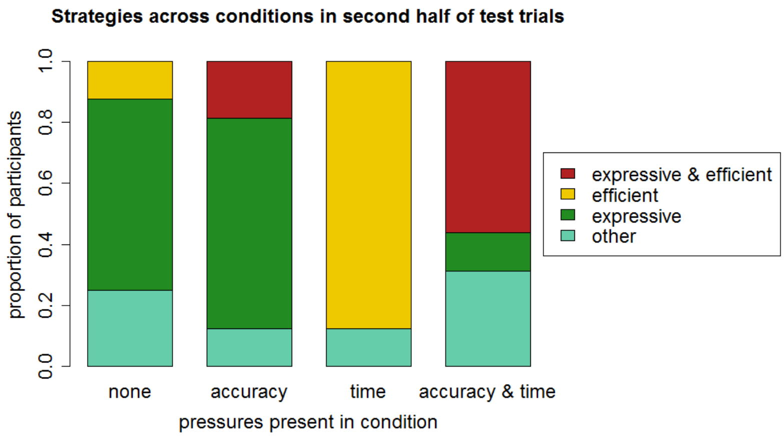
As Zipf observed in 1935, human languages appear to exhibit an inverse relationship between word length and word frequency; the higher the frequency of a word, the shorter it tends to be. Since then, this inverse relationship (Zipf's Law of Abbreviation, or ZLA) has been observed in a wide range of languages (see, e.g., Sigurd et al., 2004; Piantadosi et al., 2011; Ferrer-i-Cancho & Hernández-Fernández, 2013).

What causes languages to align length and frequency in this way? Zipf hypothesised that speakers operate in accordance with a Principle of Least Effort (PLE). Combined with a pressure to communicate successfully, the pressure to minimise effort for speakers–here, by minimising the length of an utterance–would lead optimally to the shortest forms being mapped to the most frequent meanings, leaving longer forms to describe rarer referents. This hypothesis (H1) relies on the assumption that speakers are sensitive to frequency differences, and can optimise form-meaning mappings accordingly. However, an alternative hypothesis (H2) states that ZLA can be explained purely by invoking cognition-external statistical facts about randomly generated systems (Moscoso del Prado, 2013; Ferrer-i-Cancho & Moscoso del Prado, 2012).

If ZLA is a result of the PLE, as H1 suggests, then we should be able to observe speakers actively optimising form-meaning mappings during communication. We test this by using an artificial language task manipulating the frequency of meanings. We predict, following H1, that speakers will spontaneously map forms to meaning in a way that optimises efficiency.

We trained participants to learn a long name and a 'clipped' name for each of two objects. The clipped name was the same for both objects, while the long names were distinct. Crucially, one object appeared more frequently than the other. In the critical condition, pairs of participants played a communication game in which they took turns transmitting the name of the object they saw to their partner, who then had to guess which of the two objects the first partner was seeing. Participants could choose to transmit either the long or short name to their partner. The longer names took a longer amount of time to transmit, so that greater length was associated with greater ‘effort’. Pairs were rewarded for completing the task in the quickest time (introducing a pressure to minimise effort), while maximising the number of correct guesses (pressure to communicate successfully). Three control conditions were included for a full 2x2 manipulation of time and communicative pressures.

In the condition with a time pressure but no communicative pressure (N=8), participants mapped both objects to the ambiguous short form (‘efficient strategy’). In the condition with only a communicative pressure (N=8 pairs), most participants retained the unique long forms for each object (‘expressive strategy’). Crucially, in the condition with pressures to communicate both accurately and quickly (N=8 pairs), by the end of test trials, most participants converged on the optimal strategy wherein the most frequent object was mapped to the ambiguous short name, and the infrequent object to its unique long name, making the ‘language’ both efficient and expressive.




Because the optimal strategy was significantly more likely to occur in the critical condition than in any of the controls, we conclude that speakers are sensitive to frequency differences, and when subject to pressures to communicate accurately and efficiently, they actively map forms to meanings in a way that optimises both these factors. These results are proof of concept that the PLE is in fact a plausible explanation for why the inverse relationship between word length and frequency is so widespread in human languages and communication systems; it results from a gradual accumulation over time of speakers’ optimising behaviour.

## References


Ferrer-i-Cancho, R., & 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., & Moscoso del Prado, F. (2012). Information content versus word length in random typing. *JSTAT*.

Moscoso del Prado, F. (2013). The missing baselines in arguments for the optimal efficiency of languages. In *Proceedings of the 35th annual conference of the cognitive science society* (p. 1032-1037).

Piantadosi, S. T., Tily, H., & Gibson, E. (2011). Word lengths are optimized for efficient communication. *PNAS*, *108*(9), 3526–3529.

Sigurd, B., Eeg-Olofsson, M., & 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* (Vol. ix). Oxford, England: Houghton, Mifflin.