

Maike Vorholt

Can frequency predict length?

A crosslinguistic investigation of Zipf's law for European adpositions

The aim of this crosslinguistic investigation is to shed some light on European adpositions. The proportion of adpositions in press corpora is compared for the sample languages. Additionally, a relationship between length and frequency as stated by Zipf's law of abbreviation (Zipf 1935) will be examined. This assumption has been tested for several languages and text types (Sigurd et al. 2004; Strauss et al. 2007; Piantadosi et al. 2011). In this presentation we will see if Zipf's law applies to the word class of adpositions as well.

Languages were chosen from different language families which led to a convenience sample of 16 languages, including Maltese. The comparison reveals that adposition inventories are quite divergent in size and composition. Moreover, it will be shown that adposition proportions vary significantly across corpora. Linear regression analysis is used to see if frequency can predict orthographic length. The results show that length can be predicted by frequency for the majority of languages in the sample.

Since comparative research in the area of adposition frequencies has been very limited so far, this study can lead to new input concerning adpositions in a crosslinguistic perspective.

Literature

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 de Weijer, J. (2004) Word length, sentence length and frequency - Zipf revisited. *Studia Linguistica* 58(1), 37–52.

Strauss, U.; Grzybek, P. & Altmann, G. (2007) Word length and word frequency. In P. Grzybek, ed. *Contributions to the science of text and language*. Dordrecht: Springer, 277–294.

Zipf, G.K. (1935) *The psycho-biology of language. An introduction to dynamic philology*. Boston: Houghton Mifflin Company.