ADDRESSEES USE ZIPF'S LAW AS A CUE FOR SEMANTICS

FREEK VAN DE VELDE & DIRK PIJPOPS
Dept. of linguistics, University of Leuven,
Leuven, Belgium
freek.vandevelde@kuleuven.be, dirk.pijpops@kuleuven.be

1. Zipf's size-meaning correlation

In a number of publications George Kingsley Zipf (1932, 1935, 1949) set out to describe several quantitative tendencies that hold universally in language. Apart from the famous direct inverse correlation between the relative frequency of a word and its rank in a frequency list, Zipf also noted (i) an inverse correlation between the relative frequency of a word and its phonetic substance, and (ii) a correlation between the relative frequency of a word and its level of polysemy. Combining (i) and (ii), an inverse correlation can be deduced between the phonetic substance of a word and its polysemy: shorter words tend to have more meanings (Pustet 2004). We will call this Zipf's size-meaning tendency. Zipf himself saw one over-arching principle behind the phonetic size of words: the principle of least effort. Speakers are under a constant pressure to reduce articulatory effort: frequent, and semantically less specific words can be shortened by speakers. It is often assumed that this benefit for speakers is diametrically opposed to the addressee's interest (Langacker 1977): addressees want clear articulatory distinctiveness, to overcome noise in the signal. In this paper, however, we argue that addressees can benefit from Zipf's size-meaning tendency as well. If they are able to segment the speech signal in words, they are able to work on the correlation between the phonetic size and the semantics (see also Piantadosi et al. 2011). In our study, polysemy was measured using a proxy: cross-linguistically stable asymmetries of marked-unmarked related semantic pairs. The inverse correlation between polysemy and markedness, is grounded in recent work by Winter et al. (2013), who adduce evidence from corpus data and psycholinguistic experiments. The relation between phonetic size and semantic complexity is corroborated by recent work by Lewis & Frank (subm.).

2. Research Design and Findings

We set up an experiment in which 370 native-Dutch speakers (after filtering of problematic responses) were presented with 9 pairs of visual stimuli (drawings), based on Urban's (2011) asymmetrical word-pairs (e.g. SUN – MOON) and Berlin & Kay's color hierarchy (e.g. GREEN – GREY). Together with the visual pairs 2x9 verbal targets were presented, in the form of fake words that differed in phonetic size. Subjects were asked to match the verbal targets to the visual stimuli, under the pretext that the verbal targets were actually attested in children's speech. We only selected Urban pairs of visual stimuli which did *not* conform to the expected size-meaning correlation in Dutch. This 'hyper-conservative' setup was needed to avoid interference from the known language. We put phonotactic constraints on the verbal targets and controlled for phonetic similarity to actual words in Dutch. Visual stimuli were pretested to make sure they were interpreted as intended. Additionally, a number of fillers, both of unrelated word pairs and with equally sized verbal targets were used to mask the actual research goal. Mixed-effect logistic regression was used to detect the impact of factors such as semantic domain (nature, cultural artefacts, color), and the difference in phonetic size on whether subjects coupled the visual stimuli to the verbal targets in conformity with Zipf's size-meaning tendency.

Even with the 'hyper-conservative' research set-up, and with visual stimuli of which language users are often unaware that they show semantic asymmetries, we found that test subjects indeed act on Zipf's size-meaning tendency: they significantly coupled the semantically unmarked visual Urban stimulus to the shorter verbal targets ($p < 0.0001$). We controlled for whether test subjects indicated in the debriefing that they consciously used a Zipf-like heuristic.

The regression analysis showed that the effects, though relatively small, are stronger if the difference in phonetic size in the verbal targets increases, and are more apparent in cultural artefacts (e.g. $\text{CAR}_{\text{unmarked}} - \text{TRAIN}_{\text{marked}}$), nature terms (e.g. $\text{SUN}_{\text{unmarked}} - \text{MOON}_{\text{marked}}$) and color terms (e.g. $\text{GREEN}_{\text{unmarked}} - \text{GREY}_{\text{marked}}$) than in body terminology (e.g. $\text{HEART}_{\text{unmarked}} - \text{BELLY}_{\text{marked}}$).

Our research results support the idea that Zipf's size-meaning tendency is not only under evolutionary selection by speakers, in their attempt to minimize articulatory effort, but benefits addressees as well, who can use this tendency as a cue: through their life-time experience with language, they know that in general, shorter words have more unmarked meanings, and they apply this implicit knowledge when they are confronted with a new language when other cues are absent.

Acknowledgements

We gratefully acknowledge the input of the students who collaborated on this project: Mathias De Backer, Wendy Feytons, Alek Keersmaekers, Veerle Monserez, Thomas Van Hoey, and Marie Vanwingh.

References

Langacker, R.W. (1977). Syntactic reanalysis. In Ch.N. Li (Ed.), *Mechanisms of syntactic change* (pp. 57-139). Austin: University of Texas Press.

Lewis, M.L. & M.C. Frank (submitted). The length of words reflects their conceptual complexity.

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

Pustet, R. (2004). Zipf and his heirs. *Language Sciences*, 26, 1-25.

Urban, M. (2011). Asymmetries in Overt Marking and Directionality in Semantic Change. *Journal of Historical Linguistics*, 1, 3-47.

Winter, B., G. Thompson & M. Urban. (2013). Cognitive factors motivating the evolution of word meanings: Evidence from corpora, behavioral data and encyclopedic network structure. In: E. A. Cartmill, S. Roberts, H. Lyn, & H. Cornish (Eds.), *10th International Conference on the Evolution of Language* (pp. 353-360). New Jersey: World Scientific.

Zipf, G.K. (1932). *Selected Studies of the Principle of Relative Frequency in Language*. Harvard: Harvard University Press.

Zipf, G.K. (1935). *The Psycho-biology of Language. An Introduction to Dynamic Philology*. Boston: Houghton Mifflin Company.

Zipf, G.K. (1949). *Human behavior and the principle of least effort*. Cambridge, (Mass.): Addison-Wesley.