

*Klaas Willems*

## **Zipf's Law**

empirical law according to which the frequency of occurrence of a word in a corpus of naturally occurring utterances is inversely proportional to its rank in the list of all the words of the text.

## ***Zipf'sches Gesetz***

*empirisches Gesetz, nach dem die Häufigkeit, mit der ein Wort in einem Korpus natürlichsprachlicher Äußerungen auftritt, umgekehrt proportional ist zu seinem Rang in der Liste aller im Text vorkommenden Wörter.*

What is today referred to as *Zipf's law* was termed the 'rank-size' law by ZIPF himself (ZIPF 1932; 1935; 1949). The corresponding formula is:  $f \cdot r_f = \text{constant}$  (ZIPF 1929), i.e., the frequency of occurrence of a word  $f$  multiplied by the rank of this word  $r$  in naturally occurring utterances is (roughly) constant. Thus a high position on a statistical scale of occurrence correlates with a low rank in the list of all the words of a text, in such a way that the most frequent word occurs roughly twice as often as the second most frequent one, which in turn occurs twice as often as the fourth most frequent one, etc.

ZIPF's law bears the stamp of ZIPF's lifelong dedication to the study of language from a naturalistic point of view, i.e. the belief that language can be studied as a self-regulating structure which evolves independently from other social and cultural factors, with words conceived as tools used to convey meanings in order to achieve certain objectives. By striving for optimal communication, languages balance between the *Force of Unification* and the *Force of Diversification*: on the one hand, speakers tend to minimize articulatory effort by shortening the length of utterances and words, with potential ambiguity as a result; on the other hand, diversification helps hearers to interpret the meaning of an utterance unambiguously. Fulfilling both requirements leads to the *Principle of Least Effort* (ZIPF 1949), the social sciences' analogue of the *Principle of*

*Least Action* in dynamics. Several linguistic schools continue to use ZIPF's idea of language as a self-regulating system in their research (see especially the *Journal of Quantitative Linguistics*). It is not entirely clear whether ZIPF's law reflects a property of languages or is a consequence of the properties of the formula itself, as every randomly generated text inevitably seems to produce ZIPF's law (LI 1992). Moreover, some scholars have argued that ZIPF's law is linguistically shallow (CHOMSKY/MILLER 1963). Others doubt that frequency can serve as a cognitive explanation of linguistic structure. In current linguistic research, reference to ZIPF's law commonly concerns the effects of frequency of use causing specific sound changes (compare, e.g., the initial /t/ in *two* with /tw/ in *twice* and *twenty*), syntagms or words becoming shorter over time (e.g., OE *þa hwile þe* > *while*; OE *feowertyne niht* 'fourteen nights' > *fortnight*; OE *hopian* > *hope*) or fusion with other words (e.g., Latin *cantare habemus* 'we have to sing/we will sing' > French *chanterons* 'we will sing'), as in processes of grammaticalization. The pragmatic interpretation of ZIPF's law is then that frequent forms are more predictable than infrequent ones and therefore tend to be shortened due to reasons of economy. However, these processes are far from being automatic or general, and not all frequent words are gradually shortened in the course of time. Whether frequency effects can also help to explain other linguistic phenomena that are usually accounted for by markedness and/or iconicity is still a matter of controversy.

## Literature

- LI, W. [1992] Random texts exhibit Zipf's-law-like word frequency distribution. In: TInfoTheo 38/6: 1842-1845
- MILLER, G.A./ CHOMSKY, N. [1963] Finitary models of language users. In: LUCE, R.D./ BUSH, R.R./ GALANTER, E. [eds.] Handbook of Mathematical Psychology. Vol. 2. New York [etc.]: 419-492
- ZIPF, G.K. [1929] Relative frequency as a determinant of phonetic change. In: HStPh 40: 1-95
- ZIPF, G.K. [1932] Selected studies of the principle of relative frequency in language. Cambridge, MA
- ZIPF, G.K. [1935] The psycho-biology of language. An introduction to dynamic philology. Boston
- ZIPF, G.K. [1949] Human behavior and the principle of least effort. Cambridge [Reprint: New York [etc.] 1965]