

Inter-textual Hapax Legomenon Growth Model

Luo Weihua
Dalian Maritime University

This empirical study examines the inter-textual distribution of hapax legomena and their growth patterns from the angle of EFL (English as a Foreign Language) teaching. The result shows that hapax legomena have a normal distribution across texts of equal size, and that they have a linear relationship with the size of vocabulary. Within a collection of texts, the number of hapax legomena increases as the number of texts increases. Their growth curve rises rapidly initially but the rise gradually slows down. However, as the cumulative number of word tokens reaches 1,000,000, the growth curve is still on the rise, at which point hapax legomena still account for around 38% of the total vocabulary. The inter-textual hapax legomenon growth pattern can be described with the power model, which provides a reasonably good fit to the observed hapax legomenon growth curves. The theoretical vocabulary growth rate formula based on hapax legomena ($P=V(1,N)/N$) does not correspond to the empirical data and therefore needs adjustment for inter-textual vocabulary growth.

Key words: inter-textual hapax legomena growth, power model