Linguistic laws in oral communication

I. G Torre[1], B. Luque[1], L. Lacasa [2], J. Luque[3], and A. Hernández-Fernández[4],
[1]Departamento de Matemática Aplicada, ETSIAE, Universidad Politécnica de Madrid, Spain
[2]School of Mathematical Sciences, Queen Mary University of London, Mile End Road E14NS, London, UK
[3]Telefonica Research, Edificio Telefónica-Diagonal 00, Barcelona, Spain
[4]Complexity and Quantitative Linguistics Lab, Institut de Ciències de l'Educació, Universitat Politècnica de Catalunya, Spain

During the last decades, a great effort has been made to describe and quantify statistical regularities in natural language in written corpus and those patterns, or linguistic laws, have been proposed to be ubiquity and universal. Some of these well-known statistical laws are [1]: Zipf's law (relation between the frequency of appearance of words), Brevity Law (tendency of more frequent elements in communication to be shorter), Heap's Law (vocabulary of a text grows allometrically with the text length), Menzerath-Altmann law (the longer a word, measured in number of syllables, the shorter the syllabic duration) [1][2].

In our previous work we proposed a method that allows to measure those patterns in acoustic signal without needing access to the language corpus underneath. We recovered some linguistic laws of human communication at timescales below the phoneme and another link between complexity and criticality in a biological system was found [2].

In this research we inted to unify both point of view studying linguistic laws in oral communication. We use automated forced alignment techniques for segmenting at word and phonemic level speech signal coming from spontaneus conversation in Spanish. Then, we analyze well-known physical magnitudes of oral communication such as time elapsed in seconds and energy released. Statistical laws are usually explained in terms of efficiency of communication that have been optimized during evolution so it is expected that this process has been developed in oral language and that the laws studied in written corpus emerge from it [3].

In Fig. 1 we show the Brevity Law in terms of duration and energy released instead of the usual way to use letters or syllables. It is shown that there is a tendency of more frequent words to be shorter and also less energetic. Heap's law is represented in Fig. 2 showing that exists a clear regularity in the appearance of new words according to the length of the text measured both in the number of total words (tradicional way) and in elapsed time (seconds). Finally we report Menzerath-Altmann's law (Fig. 3) at two levels: Outer panel shows the relationship between the size of phrases in terms of number of words and the size of words in terms of their mean duration; Inner panel represents the relation between the size of words in terms of number of phonemes and the size of phonemes in terms of their mean duration.

Although there are previous studies in which these scaling laws of linguistics are analyzed in written texts and in oral corpus, as far as we know, this is the first time that these linguistic laws are reported from spontaneous speech using both physical magnitudes (duration and energy) of acoustical communication.

[1] M. D. Esposti, E. G. Altmann, and F. Pachet, *Creativity and Universality in Language* (Springer, 2016).

[2] I. G. Torre, B. Luque, L. Lacasa, J. Luque, and A. Hernández-Fernández, Emergence of linguistic laws in human voice, Sci. Rep. **7**, 43862 (2017).

[3] R. Ferrer-i-Cancho, A. Hernández-Fernández, D. Lusseau, G. Agoramoorthy, M. J. Hsu, and S. Semple, Compression as a universal principle of animal behavior, Cogn. Sci. **37**, 1565-1578 (2013).

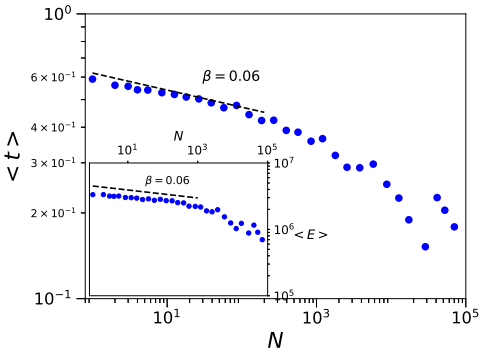

Fig. 1. Zipf's Law of Brevity in oral communication. It is shown mean duration of vocabulary (outer) and mean energy released (inner) depending on the frequency of appearance.

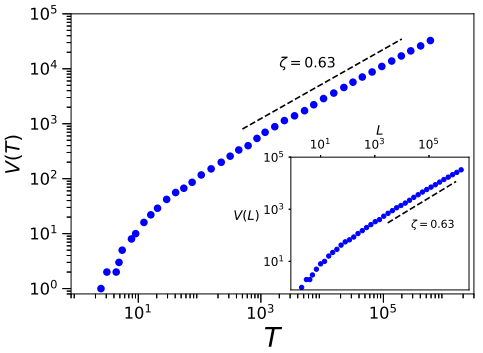

Fig. 2. Heap's law: Number of different words depending on time elapsed T (outer) and on the total words L that have been appeared (inner).

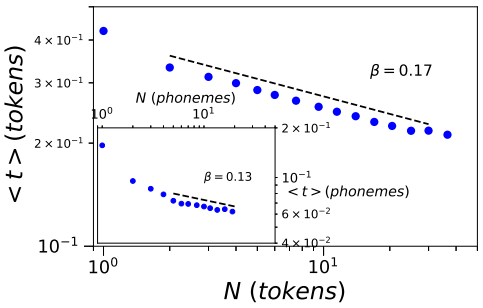

Fig. 3. Menzerath-Altmann's law in oral communication at using two different scales.