DO LINGUISTIC LAWS EMERGE FROM VOICE?

Antoni Hernández Fernández
antonio.hernandez@upc.edu

Complexity and Quantitative Linguistics Lab, Laboratory for Relational Algorithmics, Complexity and Learning (LARCA), Departament de Ciències de la Computació / Institut de Ciències de l’Educació
OUR GROUP

Complexity and Quantitative Linguistics Lab,
Laboratory for Relational Algorithmics, Complexity and Learning (LARCA),
Departament de Ciències de la Computació / Institut de Ciències de l’Educació

https://recerca.upc.edu/larca/en
https://www.cs.upc.edu/~cqllab/
COLLABORATORS IN THIS WORK

Iván González-Torre
(The Hand-Walker Artist)

Bartolo Luque

Lucas Lacasa

Jordi Luque
REFERENCES


This conference:


CONTENTS

• Introduction: Theoretical framework
• Scaling laws
• Materials & Methods
• Results
• Discussion and open questions
THE QUESTION(S)

How to explain Linguistic Laws theoretically?

• Least Effort Ferrero (1894); Zipf (1949)
• Compression Principle and/or other principles from Information Theory (maximization of mutual information, minimization of entropy… )…

Ferrer i Cancho et al (2013)
Dębowski (2015, 2018 in prep)
Ferrer i Cancho (2018)

Evidence from voice or from texts? Are both sources “equal”?

Empirical evidence of robust linguistic laws holding in written texts across different human languages has been reported many times (Baayen, 2001; Altmann & Gerlach, 2016), and it has been shown that these laws are not observed in random texts (Ferrer-i-Cancho & Elvevag, 2010).

Text is interesting but…
…is a product of our TECHNOLOGY (Scripture).

…inferences of statistical patterns of language in acoustics are biased by the arbitrary segmentation of the signal (language dependent), and virtually precludes the possibility of making (not-biased) comparative studies between human voice and other animal communication systems.


Studies with oral corpus are much less abundant, and they imply:

- a transcription of the acoustical waves into words (case of human speech)
- or some ill-defined analog of words (animal communication)

... as the main segments to analyze statistically.

This problem leads researchers to manually segment acoustic signals guided by their expertise and prevents to explore signals of unknown origin (Doyle et al, 2011).

It involves two major problems in communication studies:

(i) The **impossibility of performing fully objective comparative studies** between human and non-human signals.

(ii) A rather arbitrary definition of the units of study guided by **orthographic conventions** already **produces non-negligible epistemological problems at the core of Linguistics** (Bunge, 1984; Köhler, 2005).


THE QUESTION(S)

Do linguistic laws emerge from voice?

• What is the origin of the linguistic laws that we know (Zipf’s law, brevity law…)?

• Have they a physiological (physical) origin?

• Could they be Scaling Laws/SOC?

• Can we find computable patterns at levels lower than the phoneme?
  (TECHNICAL APPLICATIONS, SPEECH TECHNOLOGIES)
Speech synthesis software fail to be “natural” so engineers introduce “residuals” in synthesis algorithms (small pieces of real human voice).

Sure, because human voice evidences nonlinearities at fine grained level (deviations from source filter theory which is linear and assumes voice is a combination of Gaussians).
Scaling Law (SL) is a functional relationship between two quantities, independent of the initial size of those quantities: one quantity varies as a power of another (POWER LAW).

Self-Organized Criticality (SOC) is a property of dynamical systems that have a critical point as an attractor. (Bak et al, 1987) SOC is a phenomenon observed in complex systems of multiple interacting components, that produce power-law distributed avalanche sizes. (Hoffman & Payton, 2018)

SL in human voice?

The equivalence of power laws with a particular scaling exponent can have a deeper origin in the dynamical processes that generate the power-law relation.


https://commons.wikimedia.org/wiki/File:Major_levels_of_linguistic_structure.svg
MATERIALS

Dataset 1: KALAKA2
• TV broadcast speech dataset
• 4 hours per language
• 6 languages (Basque, Catalan, Galician, Spanish, Portuguese and English)
• Different conditions (planned & spontaneous speech, different environments, excluding telephonic channel)
• CD quality (16 bit / 44.1 kHz / stereo) Roland Edirol R-09 ultralight digital audio recorder
• Signals downsampled at 16kHz, left & right channel averaged via SoX and stored in WAV

Dataset 2: NIST Language Recognition Evaluation 1996
• conversations drawn mainly from LDC Friendcall corpus
• 2—4 hours per language
• 11 languages (English, Arabic, French, Mandarin, German, Hindi, Japanese, Spanish, Korean, Tamil, Vietnamese).
• several speakers from several conversations but speaking the same language
• signals correspond to one side of a 4-wire telephonic conversation
• standard 8 bit 8kHz mu-law digital telephone data
• samples converted into 2byte PCM digital format
SET A THRESHOLD $\theta$ DEFINED AS THE % OF DATA THAT ARE LARGER THAN $\theta$.

DEFINING ENERGY RELEASES VIA THRESHOLDING

METHOD

Defining Energy Releases via thresholding

METHOD

• During speech, the energy is unevenly released and power-law distributed. (Gutenberg–Richter law)
• ‘Earthquakes in speech’ show temporal correlations and are power-law distributed.
• The process responsible for this complex phenomenon is not cognitive, but it resides in the physiological mechanisms (alveolar) of speech production.
RESULTS

Binned histogram of integrated avalanches $P(E)$ – Gutenberg-Richter-like law


we fix threshold $\theta=80\%$
constant & small to remove background noise
RESULTS

P(E) – Gutenberg-Richter-like law

Results are independent of the threshold (invariant under rescaling)

\[ E \to E\langle E \rangle / \langle E^2 \rangle, \quad P_\theta(E) \to P_\theta(E) \langle E^2 \rangle^2 / \langle E \rangle^3. \]

Varying \( \theta \) between 90% and 50% allows to set a threshold that ranges several orders of magnitude in energy.
RESULTS

HEAPS - HERDAN LAW

Sublinear growth of the number of different elements $V$ in a text with text size $L$

$V \sim L^{\alpha}$, $\alpha < 1$

Log-log plot of the Heaps' law for the Portuguese sample (KALAKA) and several thresholds. In the inner panel we show how different tokens ($V$) increases sublinearly with the size of the series ($L$), where the slope can be estimated properly for about three decades.

González-Torre et al (2017)
RESULTS

ZIPF’s LAW

Number of different “words” (vocabulary) which occur exactly n times decays as $\mathcal{N}(n) \sim n^{-\zeta}$ (or) number of times the word with rank r occur decays as $n(r) \sim r^{-z}$ $z = \frac{1}{\zeta - 1}$
RESULTS

ZIPF’S BREVITY LAW

Tendency of more frequent words to be shorter or smaller (Zipf 1935).

Log-log plot in the case of English (KALAKA), for several thresholds. In the upper panel we plot the histogram $M(t)$ that describes the relative frequency of a type of mean duration $t$. 

González-Torre et al (2017)
RESULTS

SUMMARY

<table>
<thead>
<tr>
<th>Exponent</th>
<th>$\phi$</th>
<th>$\zeta$</th>
<th>$\alpha$</th>
<th>$\beta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Basque</td>
<td>1.13 ± 0.04</td>
<td>1.77 ± 0.14</td>
<td>0.90 ± 0.03</td>
<td>3.1 ± 0.3</td>
</tr>
<tr>
<td>Catalan</td>
<td>1.17 ± 0.05</td>
<td>1.89 ± 0.14</td>
<td>0.92 ± 0.03</td>
<td>2.8 ± 0.4</td>
</tr>
<tr>
<td>English</td>
<td>1.16 ± 0.05</td>
<td>1.85 ± 0.14</td>
<td>0.91 ± 0.01</td>
<td>2.9 ± 0.3</td>
</tr>
<tr>
<td>Galician</td>
<td>1.18 ± 0.04</td>
<td>1.80 ± 0.14</td>
<td>0.89 ± 0.03</td>
<td>2.9 ± 0.4</td>
</tr>
<tr>
<td>Portuguese</td>
<td>1.16 ± 0.05</td>
<td>1.77 ± 0.14</td>
<td>0.91 ± 0.01</td>
<td>3.0 ± 0.3</td>
</tr>
<tr>
<td>Spanish</td>
<td>1.15 ± 0.04</td>
<td>1.79 ± 0.14</td>
<td>0.91 ± 0.03</td>
<td>2.8 ± 0.4</td>
</tr>
</tbody>
</table>

TABLE I: Summary of scaling exponents associated to the energy release distribution ($\phi$), Zipf’s law ($\zeta$), Heaps’ law ($\alpha$) and Brevity law ($\beta$) for the six different languages. Power law fits are performed using maximum likelihood estimation (MLE) following Clauset [71] and goodness-of-fit test and confidence interval are based on Kolmogorov-Smirnov (KS) tests. In all cases, KS are greater than 0.99. Exponents associated to energy release are compatible with those found in rainfall [70]. Results are compatible with the hypothesis of language-independence.

DISCUSSION

• Human voice manifests the analog of classical linguistic laws found in written texts (Zipf’s law, Heaps’ law and the brevity law) in this level.

• These laws are invariant under changes of the energy threshold $\Theta$. As $\Theta$ is the only free parameter of the method, this invariance determines that the results are not afflicted by ambiguities associated to arbitrarily defining unit boundaries.

• Results are robust across a list of 16 different languages (indoeuropean and non-indoeuropean) and across timescales, energy threshold and conversational modes.
DISCUSSION

• Interpreting linguistic laws as **Scaling Laws** which emerged in communication systems actually opens the door for speculating on the existence of underlying **scale-invariant (physical) laws** operating underneath.

• The specific and complex alternation of air stops (silences) intertwined with voice production are at the core of the microscopic voice fluctuations (SOC?).

• First observation of scaling behavior with a clear exponent in the case of brevity law in speech. Our finding of a power law in brevity law differs from the case of **random typing** where a power law doesn’t conform.
DISCUSSION

• We are able to map an arbitrary acoustic signal into a sequence of types separated by silence events.

• Standard linguistic laws can then be directly explored in acoustic signals without needs to have an a priori knowledge neither of the signal code nor of the adequate segmentation process or the particular syntax of the language underlying the signal.

• This protocol can be used to make unbiased comparisons across different systems (comparative studies): Universal Segmentation Method.
OPEN QUESTIONS

• What are the values of the exponents indicating (in this level under the phoneme)?

• How can we connect these findings with information theory?

• Emergence of (“linguistic”) Scaling Laws already at the voice level: another hint of complexity? Is the system operating close to a critical point?

• Is there any evolutive gain?

• Relation with traditional linguistic laws (in upper levels)?
OPEN QUESTIONS

Is physiology the ultimate reason of the onset of complexity and SL (linguistic laws) in communication?

Is it necessary to study other “forgotten” physical magnitudes (Energy: Guttenberg-Richter…)?

Instead of introducing pieces of real speech (residuals): Is it better to model speech fluctuations at intraphoneme via simple SOC models?

Do linguistic laws emerge from voice?
THANK YOU FOR YOUR ATTENTION!

Dziękuję bardzo!

Iván González-Torre, Bartolomé Luque, Lucas Lacasa, Jordi Luque & Antoni Hernández-Fernández

Emergence of linguistic laws in human voice. Scientific Reports 7, 43862;
doi: 10.1038/srep43862 (2017).

This research was supported by the grant TIN2017-89244-R from MINECO (Ministerio de Economía, Industria y Competitividad; Spanish Government), and the recognition 2017SGR-856 (MACDA) from AGAUR (Generalitat de Catalunya).