The Physics of Language Learning

Based on work of K. Kosmidis, A. Kalampokis, J. M. Halley and P. Argyrakis.

A Problem

- Can Human Language be described using concepts of Physics ?
- Physics deals with collections of molecules, atoms or elementary particles.
- Unlike systems common in Biophysics, Geophysics and other interdisciplinary areas, there are NO molecules in Language!

Our Target

- What we hope is to use the framework, terminology and methods of Statistical Mechanics to describe Human Language.
- This is what is done in Econophysics (application of physics in Economics), where also there are no molecules.

Basic Statistical Mechanics

System at constant temperature *T*.

The probability that it is found at a given state *i* with energy E_i is proportional to the "Boltzmann factor."

$$p = \frac{1}{Z}\,\mathrm{Exp}\left(-\frac{E_i}{k_B T}\right)$$

Some Remarks

- Common sense indicates that some words are more useful than others. The word “food” is essential and no organized group of people will go very far without it in their vocabulary. The word “heterogeneous” is probably not so useful and I can imagine groups of people that will survive without knowing it.
- Speakers may have different ability or willingness to communicate.

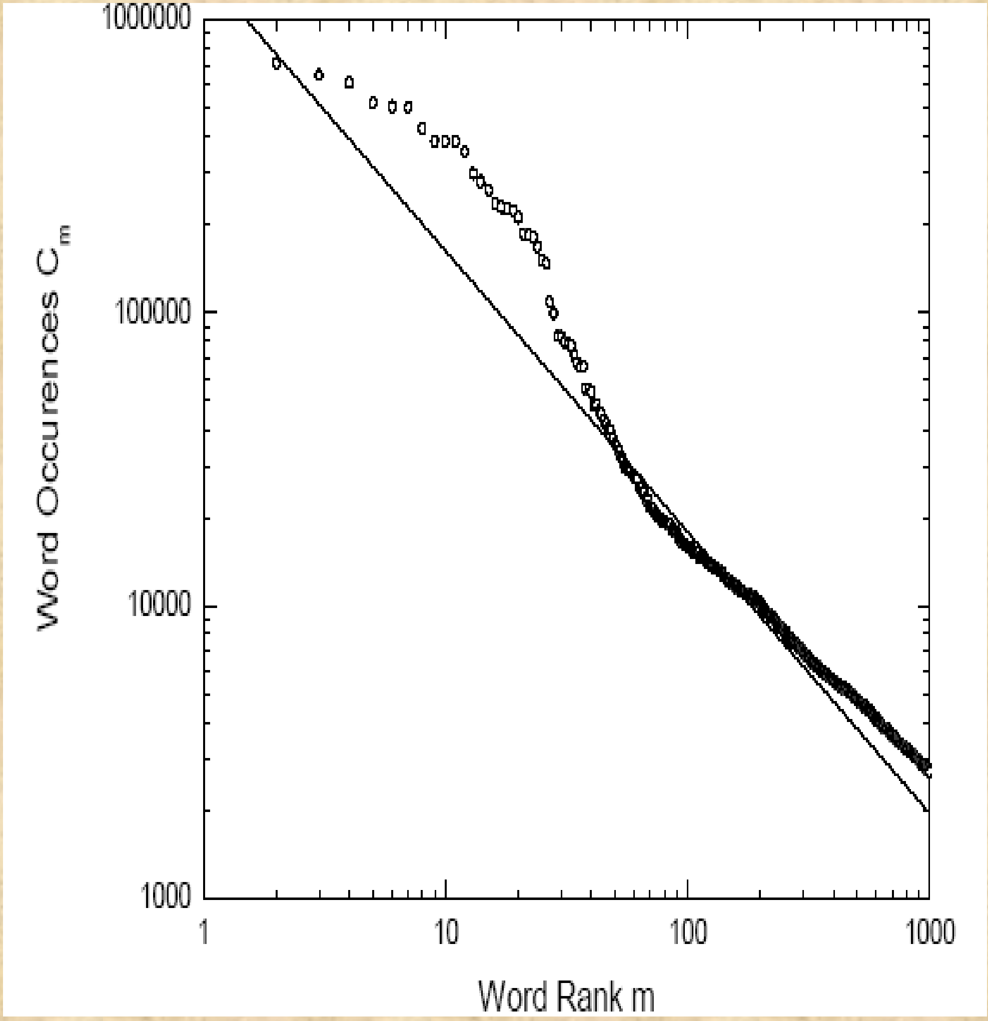
A Possibility

- Language of N words is a physical system that can be found in N states.
- Energy of state: $H(k)=\varepsilon \ln k$,

  k = Words usefulness.
- Temperature of the system T = communication ability (or willingness) of the speaker(s).

ZIPF LAW

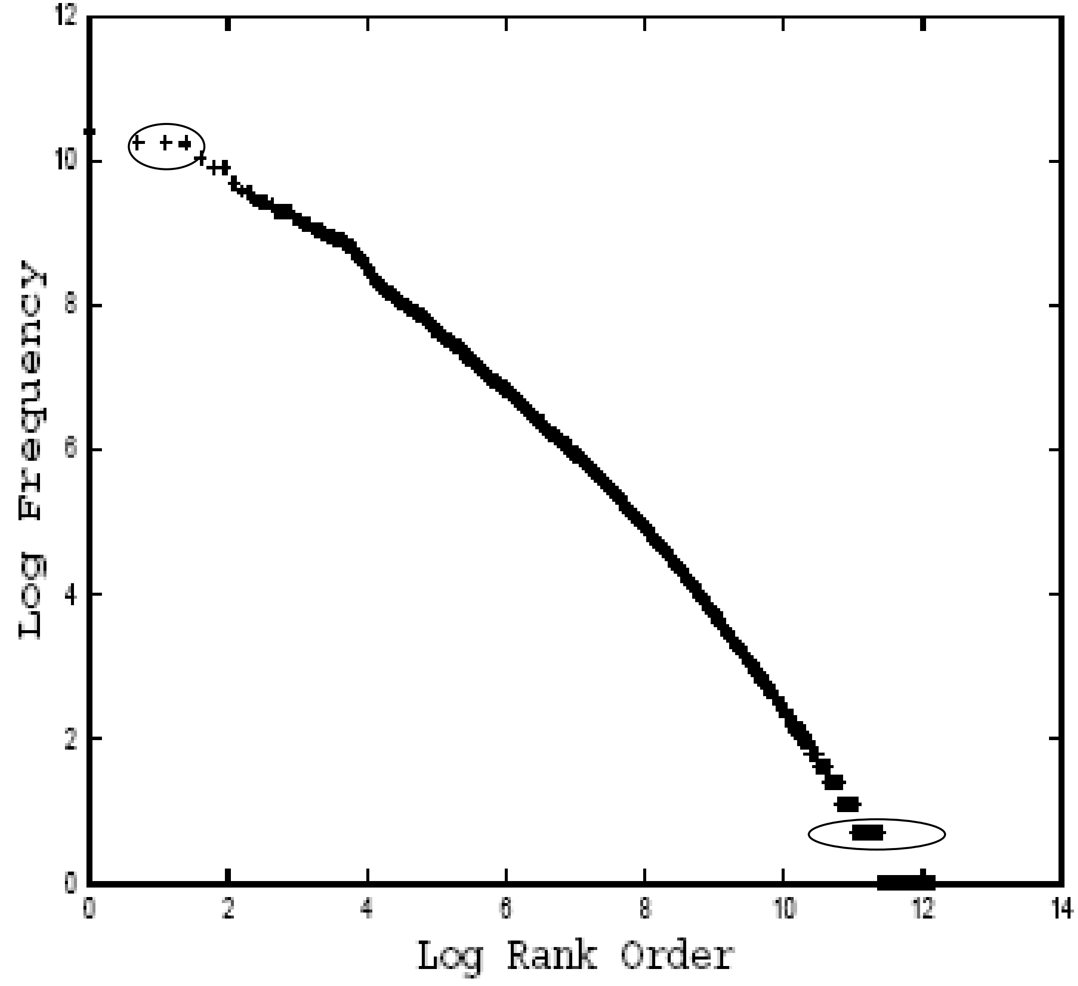
- Empirical law
- Frequency –Rank plot is a power law
- Believed to be universal
- Figure for Greek (Modern)



Predictions

- Probability to find a word with usefulness k
- Similar, but not identical, to Zipf law.

$$p = \frac{1}{Z} k^{\left(\frac{-\varepsilon}{k_B T}\right)}$$



Log Frequency
Log Rank Order
0
2
4
6
8
10
12
14

Universality of Exponent in Zipf law

- Absolute value of the Zipf exponent.
- High values => low temperature => low communication ability.
- Most modern languages are similar in ability to convey messages.
- This is experimentally verified.

$$n = \frac{\varepsilon}{k_B T}$$

Partition Function

- How to calculate the partition function?

- It is just a sum of the Boltzmann factor over all energy states.

PARTITION FUNCTION

$$Z \approx \int_{1}^{N/b} b\, k^{-\varepsilon / k_B T}\, dk = k_B T \left[N \left(\frac{b}{N}\right)^{\varepsilon / k_B T} - b \right] / (k_B T - \varepsilon).$$

- b= average number of words having the same usefulness (i.e energy) (degeneration)

Vocabulary size of Children

- We will study the vocabulary growth as a warming process. The temperature in this case is a measure of communication ability. We suppose that this ability increases proportionally with time. So we set T=γ t

$$T = \gamma\, t$$

- We assume that a word is known if the probability to appear is greater or equal to one over 10000.
- We solve the probability equation and find a (rather complex) equation for the vocabulary size vs Time

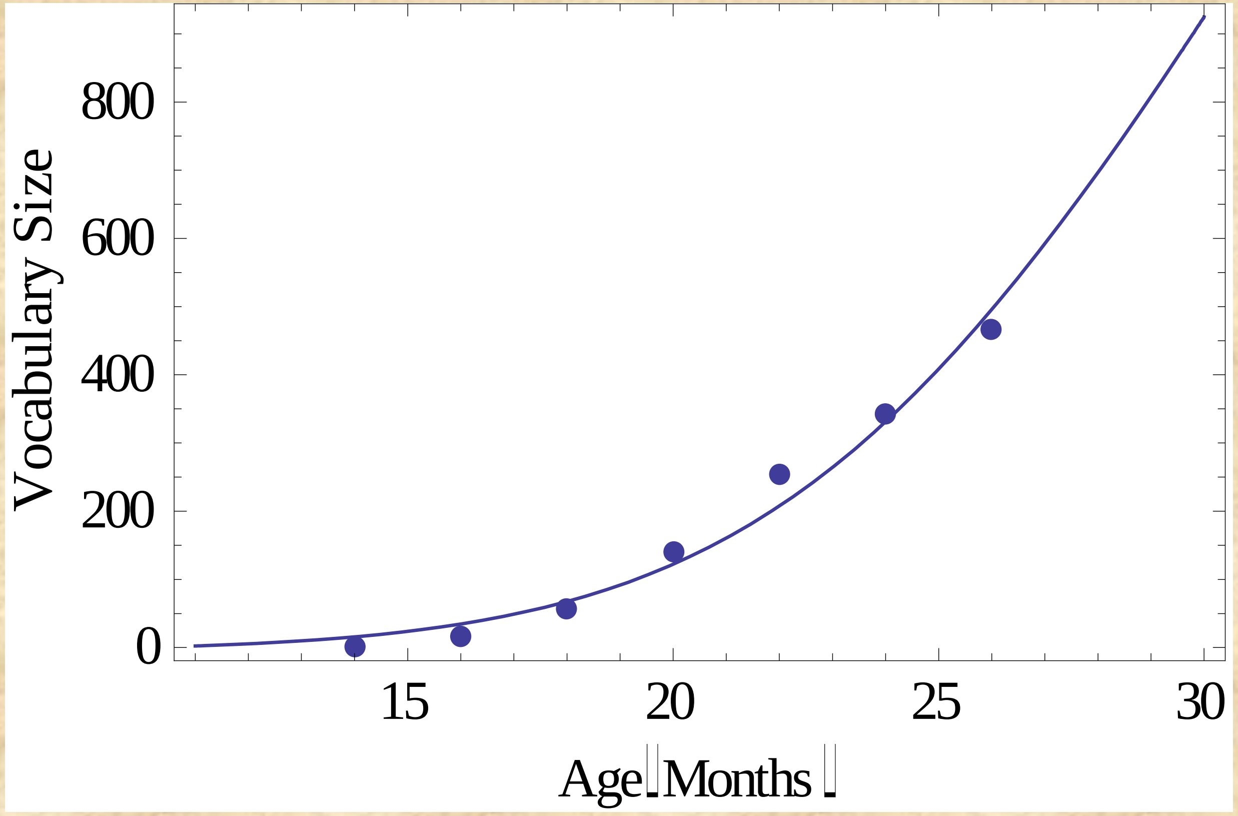
$$\frac{1}{Z}\, k_{\max}^{-\varepsilon / k_B T} = p_c$$

Vocabulary Size vs Time

$$V_s(t) = b(k_{\max} - 1) = b[-1 + (\frac{A}{p_c})^{\gamma k_B t/\varepsilon}],$$

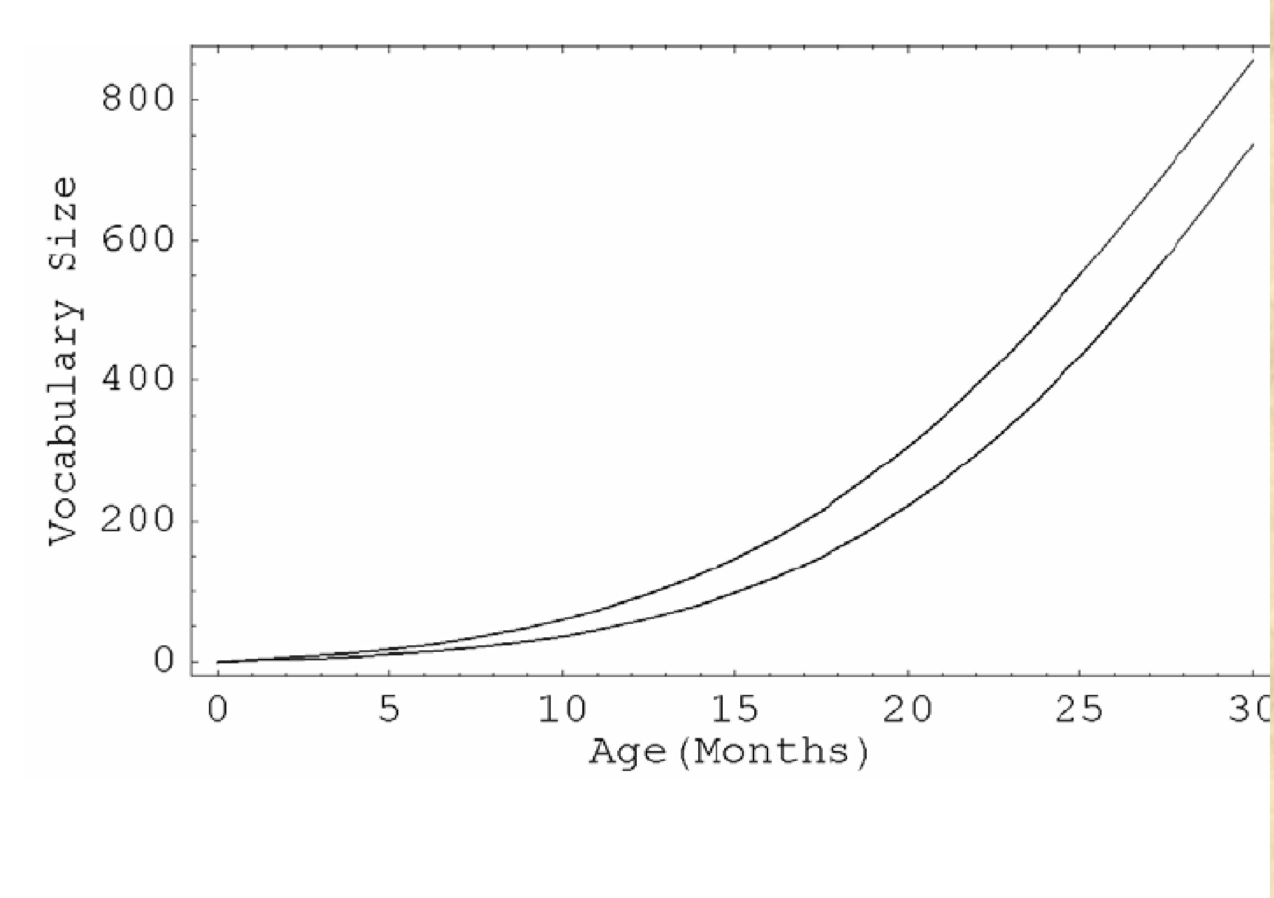
$$A = \frac{-1}{b - N(\frac{b}{N})^{\varepsilon/(\gamma k_B t)}} + \frac{\varepsilon}{(b - N(\frac{b}{N})^{\varepsilon/(\gamma k_B t)})\gamma k_B t}$$

Comparison



Only one adjustable parameter.

BILINGUAL CHILDREN



No evidence that bilingual children talk late

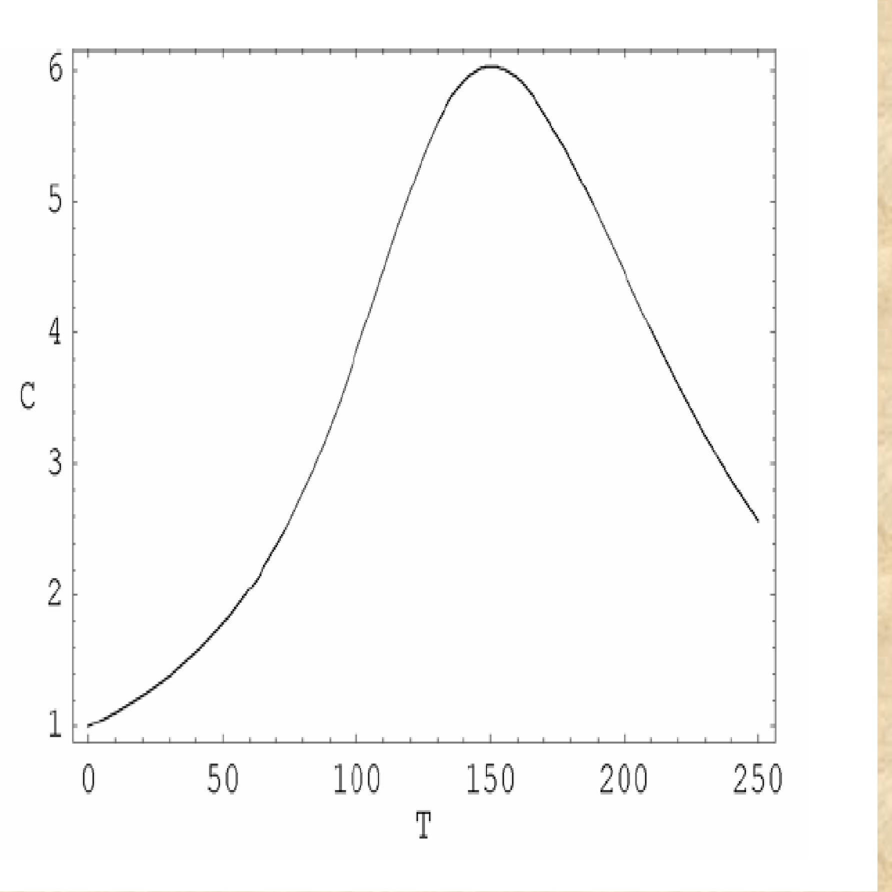
Syntactic Communication.

- *Z* remains finite for *N* -> infinity, only if k_B T < ε
- Low T= allowed to have a language with infinite number of words.
- High T = Overcome the difficulty through syntax.

Effort to Increase Communication Skills

- We may estimate effort needed to increase communication skills of a person or group by calculating

$$C = \frac{\partial E}{\partial T}$$



Schizophrenia

- Zipf exponent in schizophrenic patients is higher than normal. =>
- T for subjects with schizophrenia is lower than for “normal” subjects.
- Lower communication ability associated with mental illness.

Coding and non-coding DNA

- Mantegna et al (1994), partitioned the DNA sequence in sections of fixed length and these sections were considered as words.
- Non-Coding DNA has a Zipf exponent 2.2 times higher than coding DNA
- Low temperature for non-coding DNA.

OVERVIEW

- What the model does NOT answer?

1. There is no reason that the energy of a word is proportional to $\ln k$. Why not k or k^2 for example? We cannot conclude that Zipf law is a power law.
2. We cannot predict that the value of the exponent of Zipf law is close to one.

OVERVIEW

- What the model explains?

1. Universality of the Zipf Exponent
2. “Islands” on Zipf diagram
3. Vocabulary Size
4. Rate of learning for Bilingual Children.
5. Schizophrenia , Coding DNA.
6. Monte Carlo Simulations can be used for simulating more complicated situations.
7. Several quantities, like Entropy etc can be easily calculated.