Scaling laws in language evolution

Ricard V. Solé[1 2]

[1] ICREA-Complex Systems Lab, Universitat Pompeu Fabra(GRIB), Dr Aiguader 80, 08003 Barcelona, Spain ricard.sole@upf.edu

[2] Santa Fe Institute, 1399 Hyde Park road, 8765 Santa Fe, New Mexico USA sole@santafe.edu

Human language stands as one of the most important leaps in evolution (Bickerton, 1990; Deacon, 1997; Maynard Smith and Szathmáry, 1995). It is one of its most recent inventions: it might have appeared as recently as 50.000 years ago. Our society emerges, to a large extent, from the cultural evolution allowed by our simbolic minds. Words constitute the substrate of our communication system and the combinatorial nature of language (with a virtually infinite universe of sentences) allows to describe and eventually manipulate our world. By means of a fully developed communication system, human societies have been able to store astronomic amounts of information far beyond the limits imposed by purely biological constraints. As individuals sharing our knowledge and the cumulative experience of past generations, we are able to forecast the future and adapt in ways that only cultural evolution can permit.

The faculty of language makes us different from any other species (Hauser et al., 2002). The differences between animal communication and human language are fundamental, both in their structure and function. Although evolutionary precursors exist, it is remarkable to see that there seems to be no intermediate stage between them (Ujhelyi, 1996). Such a shift might result from a number of causes (Wray, 2002) from rare events (making human language a rather unique, unlikely phenomenon) to so called macromutations. But alternative scenarios stem from a sudden transition (Ferrer and Solé, 2003).

Unfortunately, language does not fossilize and to a large extent we might forever ignore the tempo and mode of language evolution. But we certainly know that languages are not static and are strongly influenced by social constraints. Economic and cultural forces modify the fate of human communication systems and they can trigger their disappearance or favour their predominance[1].

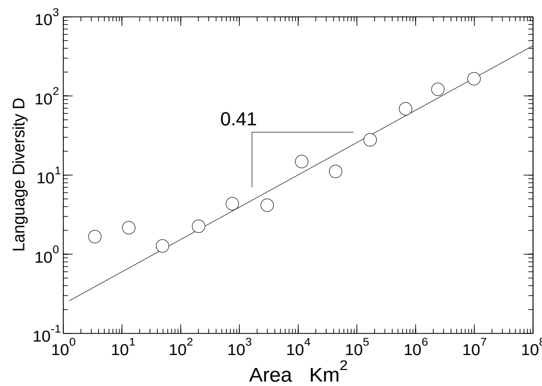

Fig. 1. Scaling law in the distribution of language diversity D as a function of area. The best fit to the power law $D \sim A^z$ is shown. Redrawn from Gomes et al., 1999.

As information spreads and breaks geographical barriers, so do dominant languages expand their reach. The direct effect is the generalized decline of rare languages. As urban centers attract more and more people and small communities disappear, native languages are wiped out. Starting from the late XXth century, with the development and widespread exchange of information through the planet, the trend has been accelerating. Linguistic change is a multiscale process, and known to develop at different time scales within different social groups. But as it occurs with other social and economic complex systems, language itself displays well defined trends suggesting the presence of universal patterns of organization. Such patterns are particularly obvious when looking at the scaling laws observed at multiple levels. In this chapter we present several scaling phenomena arising within human language at different scales.

1 Language biodiversity

One of the universal laws of ecological organization is the so called species-area relation (Rosenzweig, 1995). It establishes that the diversity of species D (measured as the number of different species) in a given area A follows a

[1] New languages can rapidly emerge provided that appropriate conditions are met. As an example, a full sign language emerged among deaf children in Nicaragua when young children deprived of exposure to any language invented a new one, unrelated to spanish nor the so called american sign language. After a few decades, it became fully developped.

power law

$$D \sim A^z \tag{1}$$

where the exponent z varies from $z = 0.1$ to $z = 0.45$ typically. Interestingly, languages seem to follow similar trends. They exhibit an enormous diversity, strongly tied to geographical constraints. As it happens to occur with species distributions, languages and their evolution are shaped by the presence of physical barriers, population sizes and contingencies of many kinds. And as it happens with biological entities, languages emerge but also get extinct. In this context, differences are also clear: speciation in ecosystems can take place without necessarily implying physical barriers, whereas languages seem to need some type of population isolation to evolve differences. A second difference involves the way extinction occurs. Species get extinct once the last of its members is gone. Languages get extinct too once they are not used anymore, even if its native speakers are still alive.

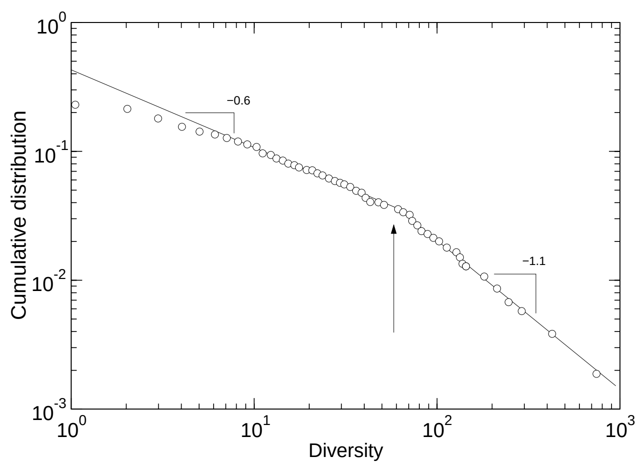

Fig. 2. Scaling laws in language diversity. Here we plot the cumulative distribution of languages using the number of countries with a language diversity greater than D. Redrawn from Gomes et al., 1999.

Studies of geographical patterns of language distribution reveal complex phenomena at multiple scales. As an example, it was shown that they also display a diversity-area scaling law, with $z = 0.41 \pm 0.03$ (Gomes et al., 1999). In figure 1 we show the results of this analysis for a complilation listing more than 6700 languages spoken in 228 countries. The power law fit is very good and spans over almost six decades (with a deviation for areas smaller than $30Km^2$) (Gomes et al., 1999). Similar results are obtained by using population size N instead of areas. In this case, it was shown that the new power law reads:

$$D \sim N^\nu \tag{2}$$

with $\nu = 0.50 \pm 0.04$.

The species-area relation has been explained in a number of ways through models of population dynamics on two-dimensional domains. Beyond their differences, these models share the presence of stochastic dynamics involving multiplicative processes. In ecology, such type of processes are characterized by positive and negative demographical responses proportional to the current populations involved: a larger population will be more likely to increase, but also more likely to suffer the attack of a given parasite (and thus experience a rapid decline).

A different measure of language diversity involves the language richness among different countries (figure 2). If $\mathcal{N}(D)$ is the frequency of countries with D diferent languages each, we can plot the cumulative distribution $\mathcal{N}_>(D)$ defined as:

$$\mathcal{N}_>(D) = \int_D^\infty \mathcal{N}(D) dD \tag{3}$$

The resulting plot is rather illustrating: the distribution follows a two-regime scaling behavior, i. e.

$$\mathcal{N}_>(D) \sim D^{-\beta} \tag{4}$$

with $\beta = 0.6$ for $6 < D < 60$ and $\beta = 1.1$ for $60 < D < 700$. What is revealed from this plot? The first domain has an associated power law with a small exponent (here $\mathcal{N}(D) \sim D^{-1.6}$): many countries have a small language diversity. But once we cross a given threshold $D \approx 60$ the decay becomes faster. One possible interpretation is that countries having a very large diversity will have harder times to preserve their unity under the social differentiation associated to ethnic diversity (Gomes et al., 1999).

The population-level view of languages provides the top of a hierarchy of levels down to the words forming them and their interactions. Are there scaling laws at those lower levels? How are they inter-related? The answer to these questions can be obtained by looking at language as a complex adaptive system (CAS), where words emerge, evolve and interact leading to emergent patterns which we identify as the characteristic traits defining language.

2 Language epidemics

Words constitute the basic meaningful units of language architecture. One of the challenges of current evolutionary theories is understanding how words originate, change and spread within and between populations, eventually being fixed or extinct. Moreover, meaningful communication beyond non-syntactic patterns requires the emergence of a set of rules able to easily exploit the underlying combinatorial power of word-word interactions. A very first approximation to word dynamics in populations should give account for the spread of words as a consequence of learning processes. Such model should be able to establish the conditions favouring word fixation.

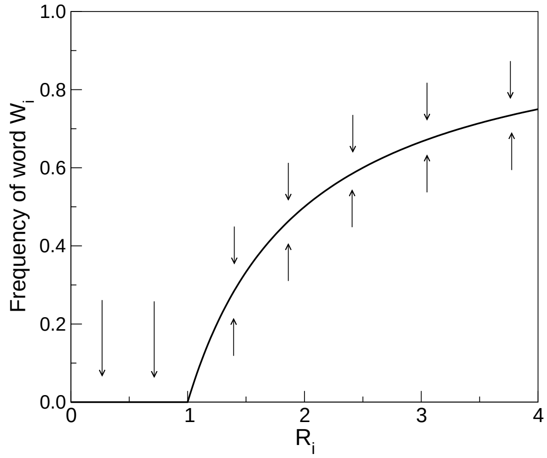

Fig. 3. Bifurcations in word learning dynamics: using a simple model of epidemic spreading of words, two different regimes are present. If the rate of word learning exceeds one (i. e. $R_i > 1$) a stable fraction of the population will use it. If not, then a well-defined threshold is found (a phase transition) leading to word extinction.

Consider a population of individuals in which each subject is born ignoring all words. These words, however, can be acquired through learning. Words are likely to be learned in a non-independent manner, but as a first approximation let us assume that they are incorporated independently of each other (Nowak et al., 1999). If x_i indicates the fraction of the population knowing the word W_i, the population dynamics of such word read:

$$\frac{dx_i}{dt} = R_i x_i (1 - x_i) - x_i \qquad (5)$$

with $i = 1, ..., n$. The first term in the right-hand side of the previous equation introduces the way words are learned. The second deals with deaths of individuals at a fixed rate (here normalized to one). The way words are learned involve a nonlinear term where the interactions between those individuals knowing W_i (a fraction x_i) and those ignoring it (a fraction $1 - x_i$) are present. The parameter R_i introduces the rate at which learning takes place.

Two possible equilibrium points are allowed, obtained from $dx_i/dt = 0$. The first is $x_i^* = 0$ and the second

$$x_i^* = 1 - \frac{1}{R_i} \qquad (6)$$

The first corresponds to the extinction of W_i (or its inability to propagate) whereas the second involves a stable population knowing W_i. The largest the value of R_i, the highest the number of individuals using the word. We can see that for a word to be maintained in the population lexicon, we require the following inequality to be fulfilled:

$$R_i > 1 \tag{7}$$

This means that there is a threshold in the rate of word propagation to sustain a stable population. By displaying the stable population x^* against R_i (figure 3) we observe a well-defined phase transition phenomenon: a sharp change occurs at $R_i^c = 1$, the critical point separating the two possible phases. The subcritical phase $R_i < 1$ will inevitably lead to loss of the word.

The previous toy model of word dynamics within populations is an oversimplification. But it illustrates fairly well a key aspect of language dynamics: thresholds exist and play a role (Nowak and Krakauer, 1999). They remind us that, beyond the gradual nature of change that we perceive through our lives (mainly affecting the lexicon) sudden changes are also likely to occur. An important aspect not taken into account by the previous model is the process of word generation and modification. Words are originated within populations through different types of proceses. They become also incorporated by invasion from foreign languages. Once again, the processes of word invasion and origination recapitulate somehow the mechanisms of change in biological populations. Both ecosystems and languages reveal features indicating universal principles of organization. Within the former, the distribution of species abundances follows a common scaling behavior irrespective of their intrinsic differences. In languages too, in spite their overwhelming diversity, all share a common trait: they have evolved in order to express the needs of their users. As a consequence of such needs and fundamental principles of optimization in communication, universal patterns have also emerged. In the next section, the best known of them -the Zipf's law- will be explored.

3 Language structure: Zipf's law

Roughly speaking, the Zipf's law states that, by ordering the words of a long text (as a sample of a given language) by how often they are spoken, the second most frequent word is about half as frequent as the most frequent, the third most frequent is about a third as frequent as the most frequent, and so on. In mathematical terms, Zipf's law links i, the rank of a word (in a list of words decreasingly ordered by frequency) with $P(i)$, its frequency. The relation follows a power law in the form:

$$P(i) = p_1 i^{-\alpha} \tag{8}$$

where $\alpha \approx 1$ and p_1 is the probability of the most frequent word. The same law can also be presented as probability density function:

$$Q(j) \propto j^{-\beta} \tag{9}$$

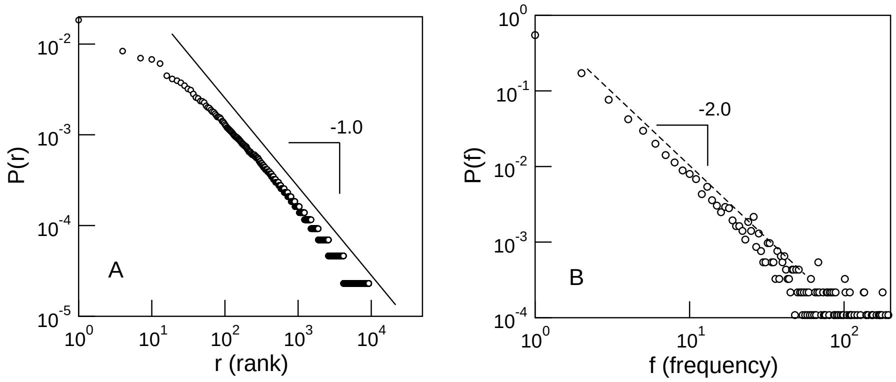

Fig. 4. Power laws in language: by ordering the most frequent words against their rank (a) we obtain the standard plot of the Zipf's law. An equivalent plot is obtained by using the frequency distribution (b) where the frequency of words appearing f times in a text is shown (redrawn from Ferrer and Solé, 2002). The first domain corresponds to the standard Zipf's law scaling. The second is associated to specific words.

where $Q(j)$ is the probability that a word is present j times in a text.

Both scaling laws are related. Let us indicate by

$$m_n = TQ(n)$$

the number of words having population n, being T the total number of words. The rank is given by

$$R(n) = \int_n^\infty m_{n'} dn' \tag{10}$$

and the most frequent word has $R = 1$, the second most frequent word has $R = 2$, and so on, for decreasing values of n in the integral. The last equation establishes a general relation between the rank of an event in the sample and the probability distribution according to the event frequency. From

$$R \sim n^{-1/\alpha}$$

(obtained from 8) and 9 in equation 10 we obtain

$$n^{1-\beta} \simeq n^{-1/\alpha}$$

from where a formal relationship between exponents is derived:

$$\alpha = \frac{1}{\beta - 1} \tag{11}$$

$$\beta = \frac{1}{\alpha} + 1 \tag{12}$$

If $\alpha = 1$ (standard Zipf's distribution) then $\beta = 2$. In figure 4 the two alternative plots are shown. Here the data have been compiled from Melville's *Moby Dick* (Ferrer and Solé, 2001, 2002). The estimated exponents are consistent with our previous discussion.

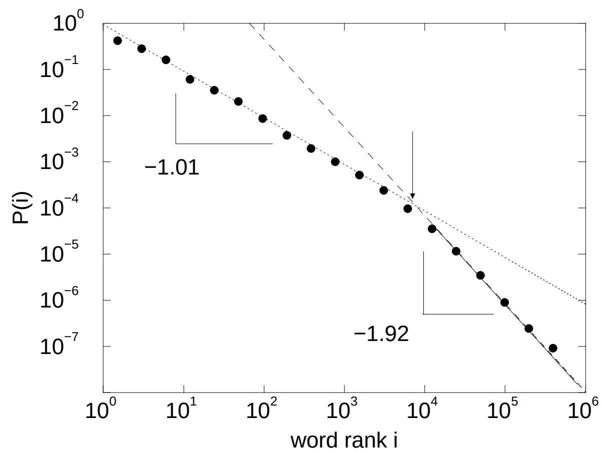

Fig. 5. Two regimes in the Zipf's law. Although standard analyses indicated that the Zipf's law involves only one regime, detailed analysis of large corpuses (such as the British National Corpus, used here) indicates that two different scaling laws are actually involved (redrawn from Ferrer and Solé, 2001).

Although early analyses of different corpuses (most based on single books) confirmed the presence of a single-scale distribution, a more careful exploration of the rank ordering plot (using much larger corpuses) revealed a very interesting pattern: the existence of two *different* exponents in the same rank ordering plot (Figure 5). Using the British National Corpus (BNC), which includes a total of $\approx 9 \times 10^7$ words and $N = 588.030$ *different* words, two exponents were found to be required to fully describe language structure (Ferrer and Solé, 2001). Here they are indicated as $\alpha_1 = \alpha \approx 1$ and $\alpha_2 \approx 2$, for ranks $i < N$ and $i \geq N$, respectively. Thus, the frequency of words is actually a double law, the initial Zipf's law followed by another, steeper scaling law:

$$P(i) = \begin{cases} p_1 i^{-\alpha_1} & \text{if } i < N \\ N^{\alpha_2} p_N i^{-\alpha_2} & \text{otherwise} \end{cases} \tag{13}$$

where p_N is a the probability of the n-th most frequent word (it can also be obtained from Eq. 8 and thus be $1/p_1 N_1^{\alpha} \approx p_n/N$).

The two scaling regimes are associated to two different sets: a *kernel lexicon* formed by $\approx N$ versatile, common words and a very large lexicon used for specific communication. The size of the kernel lexicon is likely to be related with cognitive constrains. In this context, it has been suggested that here the change in exponents is related to the average number of words reliably stored in a human brain. The set of words included within the standard Zipf's regime are those shared by most users, whereas words within the second scaling regime are very specific and obviously not shared by all speakers. The point of separation between both regimes seems to occur at $N \approx 5000 - 6000$.

The existence of a kernel lexicon indicates that there might be well-defined bounds to lexicon complexity associated to proper communication within a developped society sharing a common language. What is the smallest size required to obtain fluid communication among users? Pidgin languages provide examples of very small lexicons. Roughly speaking, a pidgin is a system of communication which has emerged within a given group of individuals not sharing a common language but who need to communicate. They may arise when speakers of different languages try to have a makeshift conversation. The final lexicon usually comes from one language but structure can come from the other. Estimates of the number of items of a pidgin vary from about $300 - 1500$ words, depending on the language and are thus far lower from the number of lexical items of a common speaker of an ordinary language (which is estimated to be $25,000 - 30,000$).

For the first domain, a minimal set of words is clearly required to perform fluently in communicating ideas and needs. But since pidgins already allow communication to take place, the increased richness of developped languages might reflect the stability and richness of the underlying society. The second scaling domain is tied to the statistical pattern of complexity of a diverse social structure with multiple and specialized communities using their own jargon. As it happened with whole languages, social forces also shape the patterns of diversity exhibited by word frequencies. However, an additional ingredient is here in place: cognition. In order to explain the Zipf's law within the first domain, we have to look into how information is transmitted among users in order to reach optimal information transfer. A simple principle formulated by Zipf might allow to explain it.

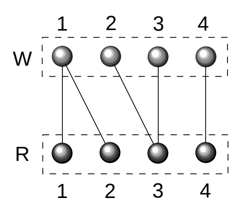

Fig. 6. Word-meaning associations are defined through a bipartite graph. Here words and objects of reference are indicated as lighter and darker balls, respectively. Links (associations) define a matrix of relations among them. We can see polysemous words (w_1) having several meanings and also synonymous words (w_2, w_3).

4 Least effort and power laws in language

The Zipf's law is common to all known languages. How is this regularity interpreted? Some authors concluded that it is just some irrelevant feature common to random texts (Li, 1992; Miller 1957). But careful inspection of this idea revealed that it is fundamentally wrong: many statistical features exhibited by real language are not shared by their random counterparts (Ferrer and Solé, 2002). An old conjecture concerning the origins of Zipf's law was actually made by Zipf himself, and is known as the *principle of least effort*. The essential idea in our context is that two efforts are implicit in communicating a message. The first involves the effort of the speaker, who wants to describe the world at the minimum cost (by means of a reduced lexicon). The second deals with the hearer's effort, who wants to be able to understand the message with the smallest level of ambiguity (by means of a diverse lexicon). More precisely, Zipf's law would emerge from a conflict between speaker's unification and hearer's diversification forces acting upon vocabularies. Unification reduces the repertoire of polysemous words, whereas diversification leads to large vocabularies, tipically formed by unambiguous words. Such a conflict, as it happens with many disordered systems analysed in statistical physics, can lead to phase transitions between two qualitatively different types of behavior (Solé et al., 1996; Solé and Goodwin, 2001). The least effort principle has been made explicit by formally measuring the two efforts using measures borrowed from information theory (Ash, 1965).

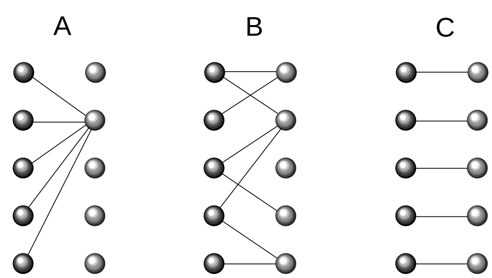

Fig. 7. Schematic distribution of links in the word-object universe for different efforts of the hearer and the speaker. Words and objects are represented as light and dark balls, respectively. In (A) we have one word being used to name all objects, whereas in (C) a one-to-one mapping of word-object associations is present. An intermediate situation is shown in (B) where both specific and ambiguous words are being used.

The principle of least effort can be properly mapped into a mathematical model based on information theoretic quantities. Let us consider an external world defined as a finite set of m objects of reference (i. e. meanings):

$$\mathcal{R} = \{r_1, ..., r_i, ..., r_m\} \tag{14}$$

and a set of n words (signals) used to label them:

$$\mathcal{W} = \{w_1, ..., w_i, ..., w_n\} \tag{15}$$

If a word w_i is used to name a given object of reference r_j, then a link will be established between both. Let us call $A = (a_{ij})$ the matrix connecting them, where $1 \leq i \leq n$ and $1 \leq j \leq m$. Here $a_{ij} = 1$ if the word a_i is used to refer to r_j and zero otherwise. A graph is then obtained, including the two types of elements (words and meanings) and their links. This is known as a *bipartite* graph. An example of such graph is shown in figure 6.

The matrix A is often called the association or *lexical matrix*. For the previous example, it would read:

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{16}$$

The lexical matrix is a binary matrix and allows to compute a number of relevant quantities. The first is the number of links (degree) of each word. For a given word $w_i \in W$ this is given by

$$k_w^i = \sum_{j=1}^{n} a_{ij} \tag{17}$$

and provides a measure of the ambiguity (polysemy) associated to w_i. Conversely, the degree k_r^i of a given object/meaning $r_i \in R$ will be given by

$$k_r^i = \sum_{j=1}^{n} a_{ji} \tag{18}$$

and measures the number of synonyms associated to r_i.

The lexical matrix also allows to calculate a set of probabilities from which the efforts can be derived. Assuming for simplicity that $n = m$, the conditional probability $P(w_i|r_j)$ of using word $w_i \in W$ given an object of reference $r_j \in R$ is

$$P(w_i|r_j) = \frac{a_{ji}}{k_r^i} \tag{19}$$

The so called joint probability $P(w_i, r_j)$ of having an association $\{w_i, r_j\}$ will be defined (from Bayes' rule) by:

$$P(w_i, r_j) = P(r_j)P(w_i|r_j) \tag{20}$$

where we will take $P(r_j) = 1/n$ for all objects in R. The frequency of a given word is thus simply

$$P(w_i) = \sum_{j=1}^{n} P(w_i, r_j) \tag{21}$$

By using these sets of probabilities, we can properly weight the efforts of the speaker Ω_s and the hearer Ω_h.

In order to illustrate the spectrum of possible trade-offs between speaker and hearer efforts, consider the three situations depicted in figure 7. The two extremes correspond to (a) least (largest) effort for the speaker (hearer) and (c) least (largest) effort for the hearer (speaker), respectively. The intermediate situation (b) corresponds to a compromise between both. For each bipartite graph we can define average quantities based on information theory and use them as measures of the efforts involved.

Let us first consider Ω_s. The effort of the speaker is clearly related to the diversity of words being used. Diversity is thus naturally measured by means of Shannon's entropy (Ash, 1965):

$$H(\mathcal{W}) = -\sum_{i=1}^{n} p(w_i) \log p(w_i) \tag{22}$$

This quantity is maximal when all elements have the same probability, i. e. for $P(w_i) = 1/n$ for all $w_i \in W$. In such case, $H(\mathcal{W}) = \log(n)$. It corresponds to a maximal repertoire of words being used and thus maximal effort for the speaker. The minimal value of entropy is zero, and is reached when a single word is being used (i. e. if $P(w_i) = 1$ and $P(w_{j \neq i} = 0)$. Such situation corresponds to minimum effort for the speaker, since it is using a single (or few) words to name all objects.

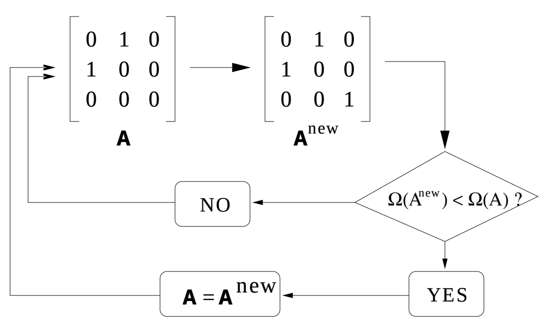

Fig. 8. Evolving language graphs. Here the basic flow diagram for the algorithm used here is presented. Each iteration involves change of one link (adding or removing it). If the effort Ω decreases as a consequence, the new interaction matrix is accepted. Iterative application of the algorithm leads to a stationary bipartite graph of word-object interactions (redrawn from Ferrer and Solé, 2003).

Defining hearer's effort Ω_h is more difficult. Here we need to weight how effective is the hearer's understanding when a given word is emitted by the speaker. More precisely, what is the likelihood that the hearer properly identifies the object r_j after receiving the word w_i? Such event has a probability given by $P(r_j|w_i)$ and thus the average effort will be measured by the diversity associated to this set of probabilities. This is given by the conditional entropy associated to $w_i \in W$:

$$H(\mathcal{R}|w_i) = -\sum_{j=1}^{n} p(r_j|w_i) \log p(r_j|w_i) \tag{23}$$

where $p(r_j|w_i) = p(r_j, w_i)/p(w_i)$. The effort for the hearer is defined as the average noise, that is

$$H(\mathcal{R}|\mathcal{W}) = \sum_{i=1}^{n} p(w_i) H(\mathcal{R}, w_i) \tag{24}$$

The total effort is defined in terms of a linear combination of the effort for the hearer and the effort for the speaker. More precisely,

$$\Omega(\lambda) = \lambda H^*(\mathcal{R}|\mathcal{W}) + (1 - \lambda) H^*(\mathcal{W}) \tag{25}$$

where the entropies have been normalized so that $0 \leq H^*(\mathcal{W}), H^*(\mathcal{R}, \mathcal{W}) \leq 1$. The energy function depends on a single parameter $0 \leq \lambda \leq 1$ which weights the contribution of each term. The question now is what type of distribution of words is obtained if we look for the word-meaning association matrix minimizing the total effort $\Omega(\lambda)$. If $\lambda = 0$ ($\lambda = 1$) then we only minimize the effort of the speaker (hearer). The interesting question is what type of pattern should be expected at intermediate values of λ when different compromises between the two efforts are at work.

In order to explore the previous question we need to scan the universe of different lexical matrices A, since all measures are based on the word-association wiring diagram. A simple algorithm was devised to this purpose (Ferrer and Solé, 2003). The basic rules are summarized in figure 8. They involve: (a) make a change in one or a few elements of a_{ij}, i.e. remove of add a few links; (b) compute the total effort $\Omega(\lambda)$ for the new lexical matrix; (c) if the new changes reduce the effort, accept the new matrix. Otherwise, keep the original. These steps are repeated iteratively until an equilibrium network of word-meaning interactions is achieved. In figure 9 the final efforts for each of these λ values are plotted in the left column. A well-defined, sharp change is observable close to a critical value $\lambda_c \approx 0.4$. Below this value, one or a few words are used to label all objects (and thus no communication is possible). Beyond the threshold, a one-to one mapping is observed. The conflict between both needs is solved at the transition point, where both are minimized. The right column of figure 6 displays the three tipical word-rank

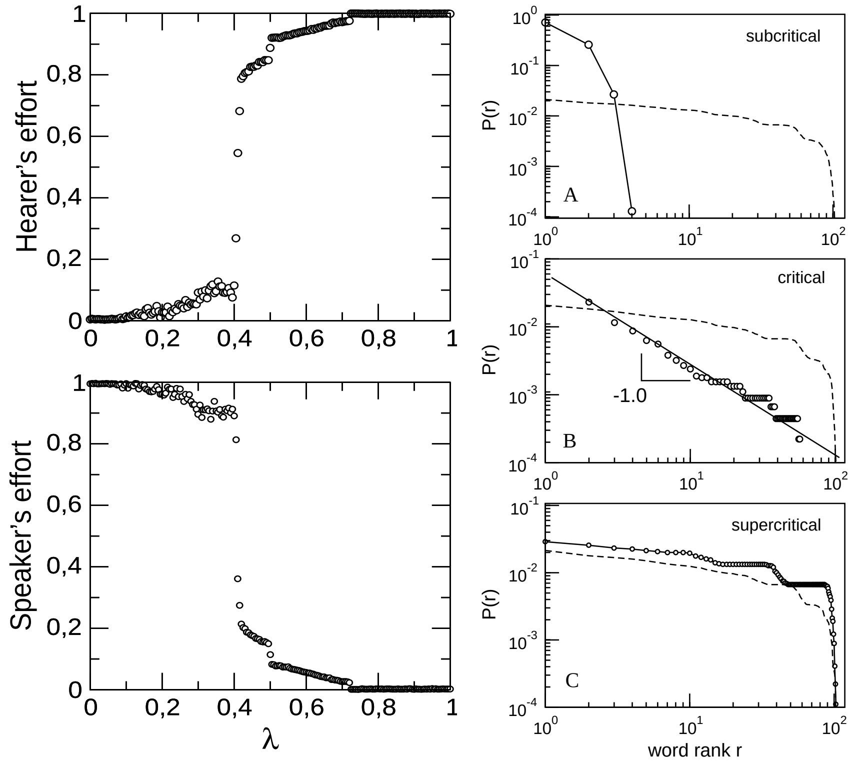

Fig. 9. Left: the efforts of the speaker and the hearer, plotted against λ. The two efforts experience a sharp change close to a critical value $\lambda_c \approx 0.4$. Rigth: The three types of rank distributions obtained for the least effort model (see text). These correspond (from top to down) to $\lambda = 0.3, 0.4$ and 0.5, respectively. Close to criticality, the Zipf's law is recovered.

distributions obtained for the three regimes: (a) subcritical ($\lambda < \lambda_c$), (b) critical ($\lambda = \lambda_c$) and (c) supercritical ($\lambda > \lambda_c$). At criticality, the Zipf's law is fully recovered.

5 Discussion and prospects

Scaling in language diversity and architecture are a consequence of both increasing returns intrinsic to social dynamics and constraints associated to communication needs. Human populations expand and shrink due to a number of socioeconomic forces. Words and languages, as an essential part of culture and social cohesion, change as well. The presence of universal patterns reveals the existence of convergent evolution. Appropriate communication results from a compromise between diversity and specificity and largely canalizes language architecture.

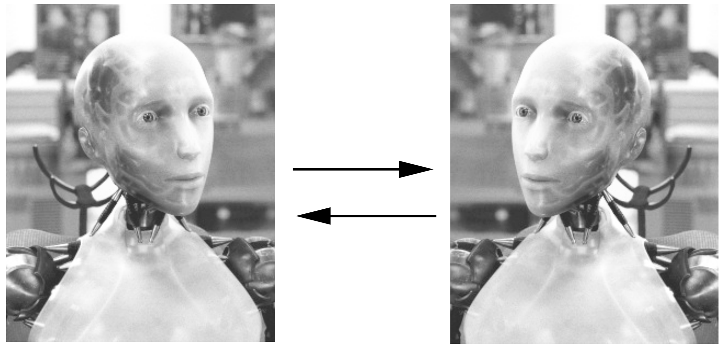

Fig. 10. Communication among artificial entities might provide insigth into the origins of language. Communicating robots have been shown to be able to develop primitive forms of lexical organization and grammar. Although far from inteligent agents, interacting artificial agents are able to generate self-organized lexicons and rudimentary forms of language.

Is it possible to really know the origins of language and how it shaped (and was shaped by) society? Language does not fossilize and thus we may never know how it emerged and how the first protolanguage sounded like.Our ancestors developped the capacity of language by mechanisms that we can only conjecture. But some alternative possibilities might be available to us. If universals are shown to be really robust and common to language architecture (Ferrer 2004), an answer to the previous question would be nevertheless available.

One possible approach to these questions is to analyse the patterns of communication emerging from interacting, artificial systems. Such an approximation has been proven successful within biology, and is known as *Artificial Life* (shortly Alife). Alife systems can be structurally far from their organic counterparts, but they often display very similar solutions to common problems. For example, evolving populations of programs competing for computer memory resources and incorporating mistakes when replicating can develop parasitism, sex or cooperation (Ray, 1991; Adami 1998). Such type of behaviors are easily recognized as essential traits of living systems. The observation of common traits strongly suggests convergent evolution at its fundamental level. In other words, if virtual creatures eventually behave as real ones, it might be the case that the spectrum of possible solutions displayed by complex systems is actually very narrow. Simple forms of language are actually known to emerge within populations of interacting, artificial agents. Such individuals have a simple cognitive architecture but the colective is nevertheless able to develop communication (Cangelosi and Parisi, 1998; Kirby, 2001). These developments define a whole area within Alife known as evolutionary linguistics (see Steels, 2003 and references therein).

Artificial creatures are not just a window into language origins and universals. The near future will host the emergence of new communication forms among humans and robots. Advances in artificial intelligence and technology have made possible the development of embodied agents with the necessary degree of internal complexity to exhibit different types of emergent behavior. Robots can incorporate a high degree of behavioral plasticity, memory and interaction capabilities. Either under the presence of comunicating humans or other robots, they can actively respond to incoming information and develop new behavioral patterns. Communication among artificial creatures and humans is one of the fundamental issues of AI, but emergent communication among artificial beings is no less important. Our future society will experience considerable changes once robotic agents become incorporated to our daily life and start interacting with us. Perhaps new forms of language might finally emerge and start change our society in ways that we barely imagine right now.

Acknowledgments

The author thanks Ramon Ferrer Cancho for useful discussions. This work was supported by a grant BFM2001-2154, FET Open Project IST ECAGENTS and by the Santa Fe Institute.

6 References

1. Adami, C. 1998. *Introduction to Artificial Life.* New York:Springer.
2. Ash, Robert B. 1965. *Information Theory.* New York:Dover.
3. Bickerton, D. 1990. Language and Species. Chicago: Chicago U Press.
4. Cangelosi, Angelo. and Parisi, Domenico. 1998. Emergence of language in an evolving population of neural networks. Connection Science 10(2):83-97.
5. Deacon, T. W. 1997. The Symbolic Species: The Co-evolution of language and brain. New York: Norton and Co.
6. Ferrer Cancho, Ramon, and Solé, Ricard V. 2001. Two regimes in the frequency of words and th eorigins of complex lexicons: Zipf's law revisited. Journal of Quantitative Linguistics 8:165-174.
7. Ferrer Cancho, Ramon, and Solé, Ricard V. 2002. Zipf's law and random texts. Advances in Complex Systems 5(1):1-6.
8. Ferrer Cancho, Ramon, and Solé, Ricard V. 2003. Least effort and the origins of scaling in human language. Proceedings of the National Academy of Sciences USA 100: 788-791.
9. Ferrer Cancho, Ramon, 2004. *Language: Universals, Principles and Origins.* Ph D Dissertation. Universitat Politecnica de Catalunya.
10. Gomes, Marcelo A. F. et al. 1999. Scaling relations for diversity of languages. Physica A271: 489-495.

11. Hauser, Marc D., Chomsky, N., and Fitch, W. T. 2002. The faculty of language: what is it, who has it and how did it evolve? Science 298: 1569-1579.
12. Kirby, Simon. 2001. Spontaneous evolution of linguistic structure: an iterated learning model of the emergence of regularity and irregularity. IEEE Transactions on Evolutionary Computation 5(2): 102-110.
13. Li, Wentian. 1992, Random texts exhibit Zipf's-law-like word frequency distribution. IEEE Transactions on Information Theory, 38(6):1842-1845.
14. Maynard Smith, John and Szathmáry, Eörs. 1995. *The Major Transitions in Evolution*. Oxford: Oxford U. Press.
Miller, George A., 1957, Some effects of intermittent silence, American Journal of Psychology, 70:311-314.
15. Miller, George A., 1991. *The Science of Words*. Scientific American Library, New York: Freeman.
16. Nowak, Martin A., Krakauer, D. 1999. The evolution of language. Proceedings of the National Academy of Sciences USA 96: 8028-8033.
17. Nowak, Martin A., Plotkin, Joshua B., Jansen V. 2000. The evolution of syntactic communication. Nature 404: 495-498
18. Ray, Thomas S. 1991. In : Langton, C., C. Taylor, J., Farmer, D. and Rasmussen, S. [eds], *Artificial Life II*, pp. 371-408. Addison-Wesley: Redwood City, CA.
19. Rosenzweig, Michael L. 1995. *Species Diversity in Space and Time*. Cambridge: Cambridge U. Press.
20. Solé, Ricard V., Manrubia, Susanna C., Luque, Bartolo, Delgado, Jordi, and Bascompte, Jordi. 1996. Phase Transitions and Complex Systems. Complexity 4:13-19.
21. Solé, Ricard V., and Goodwin, Brian C. 2001. *Signs of Life: how complexity pervades biology*. New York: Basic Books.
22. Steels, Luc. 2003. Evolving grounded communication for robots. Trends in Cognitive Science 7(7): 308-312.
23. Ujhelyi, Maria. 1996. Is there any intermediate stage between animal communication and language? Journal of Theoretical Biology 180: 71-76.
24. Wray, Martin A., editor. 2002. *The transition to language*. Oxford: Oxford U. Press.