When language breaks into pieces. A conflict between communication through isolated signals and language.

Ramon Ferrer i Cancho[1,*]

(1) Dip. di Fisica. Università "La Sapienza". Piazzale A. Moro 5. 00185 Roma, Italy.

Running tittle: When language breaks into pieces.

Abstract

Here we study a communication model where signals associate to stimuli. The model assumes that signals follow Zipf's law and the exponent of the law depends on a balance between maximizing the information transfer and saving cost the cost of communication. We study the effect of tuning that balance on the structure of signal-stimulus associations. The model starts from two recent results. First, the exponent grows as the weight of information transfer increases. Second, a rudimentary form of language is obtained when the network of signal-stimulus associations is almost connected. Here we show the existence of a sudden destruction of language once a critical balance is crossed. The model shows that maximizing the information transfer through isolated signals and language are in conflict. The model proposes a strong reason for not finding large exponents in complex communication systems: language is in danger. Besides, the findings suggests that human words may need to be ambiguous to keep language alive. Interestingly, the model predicts that large exponents should be associated to decreased synaptic density. It is not surprising that the largest exponents correspond to schizophrenic patients since, according to the spirit of Feinberg's hypothesis, decreased synaptic density leads to schizophrenia. Our findings suggests that the exponent of Zipf's law is intimately related to language and that it could be used to detect anomalous structure and organization of the brain.

Keywords: Zipf's law, communication, human language, syntax, symbolic reference, schizophrenia

(*) Contact information.
Ramon Ferrer i Cancho. Dip. di Fisica, Università "La Sapienza". Piazzale A. Moro 5. 00185 Roma, Italy.
e-mail: ramon@pil.phys.uniroma1.it
Phone: +39 064455705 / +39 064452045 (ext. 46-50)
FAX: +39 064463158

1

1 Introduction

The XX century witnessed the birth and development of information theory [1, 2], a theoretical framework devoted to the study of communication systems. Recently, various new models have been introduced to explain the organization of word frequencies in human language using an information theory approach [3, 4, 5]. Word frequencies in human language obey a universal regularity, the so-called Zipf's law [6]. If $P(f)$ is proportion of words whose frequency is f in a text, we obtain

$$P(f) \sim f^{-\beta} \tag{1}$$

where we typically have $\beta \approx 2$. Eq. 1 is a way of defining Zipf's law. The models mentioned above assume a system where signals from a set S communicate about stimuli from a set R. Signals are equivalent to words and stimuli are the basic ingredients of word meaning. For instance, the word 'dog' is associated to visual stimuli (e.g. the shape of a dog), auditive stimuli (e.g. barking),... All these stimuli are elicited by the word 'dog' [7]. Stimuli are sometimes called objects or events in the origins of language literature (e.g [8, 9]). Those models assume a set of n signals $S = \{s_1, ..., s_i, ..., s_n\}$ and a set of m stimuli $R = \{r_1, ..., r_j, ..., r_m\}$. Signals link to stimuli and connections are defined by an $n \times m$ binary matrix $A = \{a_{ij}\}$ where $a_{ij} = 1$ if s_i and r_j are linked and $a_{ij} = 0$ otherwise.

According to Shannon's standard theory [1], the goal of communication through isolated signals is maximizing $I(S, R)$, the information transfer between S and R. One of the most important contributions of the models above is that Zipf's law with non-extremal exponents can not be explained by maximizing $I(S, R)$ alone, which would lead to $\beta \to \infty$. Zipf's law with exponents close to the typical values are obtained when $I(S, R)$ is maximized with a further constraint. $H(S)$, the entropy of signals has been shown to be, as far as we know, the best candidate for that constraint [3, 10, 5]. It is known in psycholinguistics that the availability a word is positively correlated with its frequency. The higher the frequency of a word, the higher its availability. That is the so-called word frequency effect [11]. That frequency dependent availability concerns both the speaker and the hearer of a conversation. Imagine we have n words (or signals). When all words are equally likely, that is, when all words have frequency $1/n$, all words are taking the smallest frequency possible. In that case, $H(S) = \log n$, where $\log n$ is the maximum value of $H(S)$ [2]. In contrast, when a word has frequency one (which implies that the remaining words have frequency zero), $H(S) = 0$, which is the minimum value of $H(S)$ [2]. $H(S)$ is a measure of the cost of communication. The higher the value of $H(S)$ the higher the cost (and the lower the word availability). Notice that computers do not have the same information access and retrieval constraints of human brains. In general, information is accessed at a very high speed and frequency effects, when present, are not so heavy as those imposed by the human brain. One can, in general neglect $H(S)$ in many computer problems but not in brain word access and retrieval.

If we restrict ourselves to Shannon's classic information theory, the goal of

2

a communication system is maximizing the function

$$\Omega_0 = I(S, R) \tag{2}$$

If we take into consideration the cost of communication, we may write

$$\Omega = \lambda I(S, R) - (1 - \lambda)H(S) \tag{3}$$

as the function that a natural communication system should maximize [3, 10, 5]. λ is a parameter controlling the balance between maximizing the information transfer and minimizing the cost of communication. We assume $\lambda \in [0, 1]$. We have $\Omega_0 = \Omega$ when $\lambda = 1$. Ω_0 is suitable for computer or robotic problems where $H(S)$ can be neglected and Ω (with $\lambda < 1/2$ [10, 5]) is specially suitable for brain based communication systems. Ω seems, a priori, a better choice than Ω_0 for natural communication systems.

We do not claim that Ω is the best function for natural communication systems but there are some results supporting its usefulness:

- Maximizing Ω, Zipf's law is obtained for a particular value of λ. If one replaces $H(S)$ in Eq. 3 by the the effective lexicon size, namely, the number of signals with at least one association with stimuli, Zipf's law is not obtained [3, 5]. Vocabulary size is an important factor for the cost of communication [12] but does not seem to be essential for Zipf's law. Zipf's law is still reproduced if $I(S, R)$ is replaced by $H(R|S)$ in the model in [3] but not in the model in [5].

- The exponent of Zipf's law in single author text satisfies $\beta \in [1.6, 2.42]$ [10]. Maximizing Ω in a system following Zipf's law (i.e. searching the value of β maximizing Ω) can explain the interval of variation of β in human language [10].

If one considers texts from a single author [13, 14] and does concentrate on words of a certain type (e.g. nouns) [15, 4], the extremes of the interval of variation of β correspond to schizophrenic patients [10]. The aim of the present paper is deepening our understanding of what may happen when β is large and in particular, what may be happing in schizophrenics with that β. We will show that that language breaks into pieces when the balance between maximizing $I(S, R)$ and minimizing $H(S)$ favours too much the former. More precisely, we will show that the network of signal-interactions becomes suddenly disconnected when λ takes a critical value in a communication system following Zipf's law.

2   The model

Maybe the simplest approach for reproducing Zipf's law for word frequencies is combining two assumptions. First,

$$P(k) \sim k^{-\beta}, \tag{4}$$

3

where $P(k)$ is the probability that a signal has k connections. Second, $p(s_i) \sim \mu_i$, where $p(s_i)$ is the probability of using s_i and

$$\mu_i = \sum_{j=1}^{m} a_{ij}. \tag{5}$$

Eq. 4 and $p(s_i) \sim \mu_i$ give Eq. 1. Various models recover Zipf's law when maximizing Ω without the constraint in Eq. 4 for a critical value of λ [3, 5].

Going further, we assume

$$p(s_i) = \frac{\mu_i}{M}, \tag{6}$$

where

$$M = \sum_{i=1}^{m} \mu_i \tag{7}$$

is the total amount of links. Assuming Eq. 6 has the virtue of simplicity and allowing one to explain the interval of variation of β in humans. Interestingly, Eq. 6 makes some important assumptions that need to be explicited. To that aim, let us start from a general assumption about $p(s_i, r_j)$, the joint probability of s_i and r_j, namely

$$p(s_i, r_j) = \frac{a_{ij} p(r_j)}{\omega_j}, \tag{8}$$

where $p(r_j)$ is the probability of the j-th stimulus and

$$\omega_j = \sum_{k=1}^{n} a_{kj} \tag{9}$$

is the number of links of that stimulus.

If we assume

$$p(r_i) = \frac{\omega_i}{M}, \tag{10}$$

and replace it in Eq. 10 Eq. 8 we obtain

$$p(s_i, r_j) = \frac{a_{ij}}{M}. \tag{11}$$

Replacing Eq. 11 into $p(s_i) = \sum_{j=1}^{m} p(s_i, r_j)$ we recover Eq. 6. The models in [10, 5] assume Eq. 10 (hence assume Eq. 6 and 11). In contrast, the model in [3] assumes that $p(r_j)$ is constant for each j and considers a particular case, i.e. $p(r_j) = 1/m$.

We define $H(R|S)$ as the conditional entropy of stimuli when signals are known. We may write Eq. 3 as

$$\Omega = -\lambda H(R|S) - (1 - \lambda)H(S) \tag{12}$$

as in the model in [3] when $p(r_j)$ is constant. That is not the case of the present article and related models [10, 5].

4

The assumption $p(r_j) \sim \omega_j$ means that the probability of each stimulus is dictated by the structure of signal-stimulus associations. In other words, the probability of perceiving $p(r_j)$ in the 'real world' is neglected. One may think that is a very radical assumption but in fact, human language is a communication tool allowing one to detach from the here and now. Nonetheless, displaced reference, our ability to talk about something that is distant in time or space, is a salient feature of human language [16, 17], Displaced reference is not uniquely human since bees have it [18]. Because of displaced reference, we can talk of 'dogs' even when there is no 'dog' in front of us. It seems wise to assume that talking about present stimulus is not the rule of human language and it seems that in some cases such as schizophrenia, the detachment from the here and know could be extreme. Various core aspects of schizophrenia such as false believes, hallucinations [19] and various cognitive impairments, including attention problems [20], suggest that interacting with the 'real world' is difficult. In fact, schizophrenics seem optimal candidates for $p(r_j) \sim \omega_j$. Schizophrenics speakers are a very special case in the results that will follow. We will return to them in the discussion.

For the present article, we assume a communication system following Zipf's law by means of Eq. 6. The distribution of links per signal is given by $P = \{P(1), ..., P(k), ..., P(m)\}$ and the distribution of links per stimulus is given by $Q = \{Q(0), ..., Q(k), ..., Q(n)\}$, where $Q(k)$ is the probability that a stimulus has k links. We are assuming that $Q(k)$ is defined for $k = 0$ while $P(k)$ does not because we allow unlinked stimuli but do not allow unlinked signals. Here we take the simplest distribution for Q as in [10], that is

$$Q \sim binomial\left(\frac{\langle k \rangle_P}{m}, n\right),\tag{13}$$

where $\langle ... \rangle_P$ is the expectation operator over P. Thus, $\langle k \rangle_P$ is the mean signal degree. We may define the information theory measures that matter in the calculation of Ω assuming $p(r_j) \sim \omega_j$ (or $p(s_i) \sim \mu_i$) for any pair of P and Q. The calculation of Ω is straightforward once we know [10, 5]

$$H(S) = \log M - H(R|S)\tag{14}$$
$$H(R) = \log M - H(S|R)\tag{15}$$

where $M = n \langle k \rangle_P = m \langle k \rangle_Q$ and

$$H(R|S) = \frac{\langle k \log k \rangle_P}{\langle k \rangle_P}\tag{16}$$

$$H(S|R) = \frac{\langle k \log k \rangle_Q}{\langle k \rangle_Q}.\tag{17}$$

The present model integrates two recent results. The first result is that β^*, the value of β maximizing Ω grows with λ till $\lambda = \lambda^*$. Then, we have $\beta \to \infty$ for $\lambda > \lambda^*$ [10]. The behavior of β^* is illustrated in Fig. 1. It can be shown that

5

$\lambda^* < 1/2$ and a heuristic argument suggests the existence of a discontinuity at $\lambda = \lambda^*$ [10]. The idea is very simple. Eq. 3 can be written as

$$\Omega = (2\lambda - 1)\lambda H(S) - \lambda H(S|R) \tag{18}$$

knowing that $I(S, R) = H(S) - H(S|R)$ [2]. Eq. 18 indicates that maximizing Ω minimizes $H(S)$ if $\lambda < 1/2$ and maximizes $H(S)$ if $\lambda > 1/2$. $\beta \to \infty$ minimizes $H(S|R)$ and maximizes $H(S)$ (recall Eq. 17 and 14), so $\beta \to \infty$ is expected for $\lambda > 1/2$. Since maximizing Ω for $\lambda = 0$ gives a finite value of β^* (Fig. 2), β must diverge for $0 < \lambda \le 1/2$. Notice that maximizing Ω for $\lambda = 0$ is equivalent to minimizing $H(S)$, the signal entropy.

The second result is that a communication system gets a rudimentary form of language if the bipartite network of signal-stimulus associations is connected or almost connected [9]. Roughly speaking, connectedness is the possibility of starting from a signal (or a stimuli) and reaching the remaining signals and stimuli of the network crossing the links of the network. Fig. 3 A and B show, respectively, an almost connected and a disconnected bipartite networks. Almost connectedness means that a wide majority of vertices (e.g. 90%) lay in the largest connected component [9]. When exponents are close to the real ones, it has been shown that Zipf's law provides almost connectedness under a general set of conditions [9]. Connectedness is intimately related to two essential traits that researchers have identified as essential aspects of human language: syntax and symbolic reference [21]. Signal-stimulus associations allow one to define signal-signal associations. More importantly, the network of signal-stimulus association specifies allowed and forbidden signal-signal associations. Taking the example of words, we can explain why the syntactic combination of "drive cars" is a sensible combination in the sentence "John drives cars" and why it is not the combination "drives onions" in the sentence "John drives onions". The combination of 'drive' and 'car' in "John drives cars" exemplifies the relationship between a verb and its argument. As in [9], we adopt the convention that two signals (or two words) s_i and s_k can be combined syntactically if and only if they are linked to at least one common stimulus, that is, if $\xi > 0$ where

$$\xi_{ik} = \sum_j a_{ij} a_{jk} \tag{19}$$

The idea behind $\xi_{ij} > 0$ is that s_i and s_k must be semantically compatible. If $s_i =$'drive' and $s_k =$'car' we would have $\xi > 0$ and if $s_i =$'drive' and $s_k =$'onion' we would have $\xi = 0$ (Fig. 4). If two signals are linked to the same stimulus it does not mean that the signals are synonyms since stimulus here are not meanings but components of meaning. The meaning of 'drive' is linked among others, to the visual, tactile,... experiences of driving, the objects that can be driven,... whereas 'car' is associated to the visual shape of a car, the action of driving, ... The fact that 'drive' and 'car' share one or more stimuli does not mean that 'drive' and 'car' are synonyms. When the network of signal-object associations is connected, we have that for every signal there is at least another signal sharing stimuli. We could also define a network of signal-signal

6

associations defined by a binary $n \times n$ matrix $B = \{b_{ik}\}$ where $b_{ik} = 1$ if $\xi_{ik} > 0$ and $b_{ik} = 0$ otherwise. B is a rudimentary syntactic network where vertices are words and two words are linked if the can be combined syntactically [22]. The properties of real syntactic networks have been studied at the global [22] and sentence level [23]. The small-word phenomenon and heterogenous degree distribution have been reported at the global level. In a system following Eq. 4 with $\beta \approx 2$, the signal degree distribution in B has a power tail with the same exponent [9], which is consistent with the degree distribution of real syntactic networks [22].

In Peirce's view, there are three ways in the which words and objects of the 'world' can associate: iconically (by similarity), indexical (by spatial or temporal cooccurrence) or symbolically (by convention) [24]. According to Deacon, an essential aspect of symbolic reference is that real words not only evoke stimuli (or meanings) but also other words [24]. Deacon tried to define symbolic reference but his proposal has been criticized due its lack of precision [25, 26]. Taking the idea of 'signals evoking other signals', Ferrer i Cancho and colleagues have defined symbolic reference as connectedness in the network of signal-stimulus associations [22]. The definition is not ambiguous and relies on standard concepts of graph theory [27]. When a network is connected, one may start from a certain signal and reach its first neighbours (stimulus) and from them one can get to the second neighbours (signals). One can continue from 2nd neighbours to 3rd, 4th, and so on till all the signals and stimulus in the network have been reached. We define L, the normalized size (in number of vertices) of the largest connected component, as

$$L = \frac{l}{n + m},$$ (20)

where l is the number of vertices in the largest connected component and $n + m$ is the total amount of vertices. L is a measure of the expressive power of the rudimentary language emerging from signal-object associations. If $L = 1$ then all signals can be combined in a grammatically correct discourse. If $L < 1$ then that is possible only for a fraction of signals. We will show that L is controlled by λ.

3   Results

For each value of λ,

- We obtained β^*, value of β maximizing Ω, exploring $\beta \in [0, 10]$ with a resolution $\epsilon = 0.1$.

- We calculated the mean value of L in random bipartite network where signal degree follows Eq. 4 with $\beta = \beta^*$. Links with stimuli are formed chosing stimuli at random (all stimuli are equally likely so Eq. 13 follows). Means were calculated over 1000 replicas.

Fig. 5 shows the evolution of a small network of signal-stimulus associations as λ grows. At a critical value of λ, the size of the largest connected component falls

7

abruptly. In general, L falls abruptly to a small value for $\lambda = \lambda^*$ (Fig. 6). λ^* is the point where β diverges and $I(S, R)$ and $H(S)$ reach their maximum value [10]. The steepness of the fall grows with n. Fig. 7 illustrates what happens to L, $I(S, R)$ and $H(S)$ at the same time.

4 Discussion

We have seen that a communication system maximizing Ω undergoes an abrupt transition to disconnectedness for $\lambda > \lambda^*$. We have seen that the transition is caused by a sudden transition from a finite value of β to $\beta \to \infty$, where the chance that a stimulus has two links vanishes as m grows. The disconnection of the network when $\beta \to$ is easy to understand. In general, a unipartite graph with N vertices and M edges cannot be connected if $M < M^*$ where $M^* = N - 1$ is the number of edges of a tree of N vertices [27]. Thus, a bipartite graph with $N = n + m$ vertices cannot be connected if $M < n + m - 1$. In other words, connectedness is not possible if $\langle k \rangle_P < (n + m + 1)/n$. When $\beta \to \infty$, we have $\langle k \rangle_P = 1$ and $\langle k \rangle_P < (n + m + 1)/n$ holds trivially provided $m > 1$. In sum, connectedness is impossible for $\beta \to \infty$ and $m > 1$.

In Section 2, we have reviewed a heuristic argument suggesting the transition from a highly connected phase to disconnectedness in our model is discontinuous. Discontinuous phase transitions are widespread in nature. For instance, the melting of ice into water or the transformation of boiling water into vapour are discontinuous in normal circumstances. In a communication context, the models in [3, 5] show a continuous phase transition between no communication and a perfect communication phase when Ω is minimized with no constraint on P. There, the presence of Zipf's law in the vicinities of the phase transition is the hallmark of a continuous phase transition. In contrast, the phase transition from disconnectedness to connectedness in a classic Erdös-Rényi graph [28, 29] is continuous [30, 31]. The hallmark of continuous phase transition in classic unipartite graphs is a power distribution of connected component sizes [30], which is related to a critical branching process [32] at the threshold for connectedness. Other examples of continuous phase transitions are the transition from resistivity to superconductivity (continuous in the absence of an external magnetic field) and the conversion of iron from paramagnetic to ferromagnetic form [33]. In a communication context, the model examined here shows not only apparently shows a discontinuous transition to disconnectedness but also to maximum information transfer and maximum cost for $\lambda = \lambda^*$ [5].

The divergence of β for $\lambda = \lambda^*$ is accompanied by a jump to maximum information transfer (Fig. 7). Increasing λ increases $I(S, R)$ but decreases the size of the largest connected component (the significance of the decrease depends on the size of the system). At the point where the $I(S, R)$ is maximum, L is minimum. In a communication system maximizing Ω, communication using isolated signals and language are in conflict. Human speakers may need to regulate λ in order to maximize information transfer but avoid reducing the size of the largest connected component too much. Interestingly, the regulation of the size

8

of the largest connected component can be done indirectly because increasing $I(S, R)$ also increases $H(S)$, the cost of communication. Word ambiguity may not be a mere defect but a requirement for connectedness and thus language. Our findings suggests a possible scenario for the origins of language. Reducing λ (giving more weight to minimizing $H(S)$) maximizes the chance of connectedness. The emergence of connectedness could be a side effect of saving the cost of communication.

A theory of word frequencies needs answering different questions:

1. Why do words arrange themselves according to Zipf's law (Eq. 1)?

2. Why do humans choose some particular values of β?

3. Why is there variation in β?

4. What are the limits of that variation?

5. What is the link between Zipf's law and human language?

Many answers have been proposed for Questions 1-2 [5]. As far as we know, Question 1 and 2, have only been answered assuming that words are used according to their meaning in [3, 4, 5]. Choosing values of β near 2 could be an optimal solution for a conflict between maximizing the information transfer and saving the cost of communication [3, 5]. Questions 3 and 4 have begun to be addressed in [10, 4]. The idea is that the lower bound and the upper bounds of β are obtained when maximizing Ω for $\lambda = 0$ and $\lambda = \lambda^*$, respectively. The present article sheds new light on Questions 3, 4 and 5. As for 3, variation in β my be due to the chance of connectedness. As for 4, it has been argued that the variation of β is constrained by the fact that maximizing Ω for $\lambda \in [0, 1]$ gives a narrow interval of exponents [10]. It has been argued that the interval of variation of β excludes $\beta \to \infty$ because the maximum cost, i.e. $H(S) = \log n$, is paid in that case. The argument has some drawbacks. $H(S) = \log n$ is a slow growing function of n. In practice, significant difference in $\log n$ for two different systems can only be obtained if the respective values of n differ in at least one order of magnitude. In order to explain why $\beta \to \infty$ is not found, one has to argue that speakers, in general, are very sensitive to the variation of $\log n$, which we do not know. Instead, one may propose a stronger argument: $\beta \to \infty$ is not found because the chance of connectedness is 0 for $m > 1$ (as seen above). That is a compelling reason for not finding large β in human language. We do not mean that large β is impossible to attain in humans, but it would be surprising to find it in a system combining words through semantic constraints. In sum, the present article puts another step forward in the construction of a theory of word frequencies. As for 5, our work suggests that the exponent of Zipf's law is an important factor for the presence or absence of language.

The largest values of β than have been found up to now in single author text samples correspond to schizophrenic patients in the acute phase of the illness [10, 35]. One of the most salient features of schizophrenia is 'disorder of thought' [36]. Disorder of thought may be described as disturbances in the

structure, organization and coherence of thought that are reflected in reduced intelligibility and increased disorganization of speech that is difficult, if not impossible, for the listener to comprehend [37]. Our model makes two relevant predictions for the case of schizophrenics. First, the chance of being on the edge of an abrupt transition grows with the value of β, so schizophrenics with large exponents may be threatened by an apparently discontinuous phase transition where language breaks into pieces. Second, if n is small, the decrease in the size of the largest connected component with λ (and therefore β) is significant (recall Fig. 6). The larger the value of β, the smaller the size of the largest connected component. Both predictions are apparently consistent with the appearance of thought disorder in schizophrenia. It is hard to imagine how a schizophrenic can construct a coherent discourse if the size of the largest connected component has dramatically decreased.

The network of signal-stimulus associations is an emergent structure of the neural substrate. Integrating stimuli of various kinds with words implies connecting distant neural tissues. In order to have an example of mind, visual and temporal stimuli tend to be related to occipital and temporal areas of the human brain [7]. It is reasonable to think that the density of synapsis has an influence on the largest connected component of the network of signal-stimulus associations. Thus, β, specially for small n, can be seen as an indicator of the size of the largest connected component, which would be in turn an indicator of the density of the neural substrate. The link density of the network of signal-stimulus associations is $\delta = \langle k \rangle_P / m$. It can be easily seen that $\langle k \rangle_P$ decreases with β (see Appendix). For large m and $\beta > 2$ we have (see Appendix)

$$\langle k \rangle \approx \frac{1 - \beta}{2 - \beta}. \tag{21}$$

If our hypothetical correspondence between β and synaptic density (or size in words of the largest connected component) was correct, one would expect that the smallest synaptic density would be for the largest values of β, which corresponds to schizophrenic patients in the acute phase [35]. Interestingly, it has been speculated that excessive synaptic pruning occurs in schizophrenia, which may lead to psychosis when it reaches a threshold [19, 38, 39]. Our work is consistent with the spirit of Feinberg's hypothesis, relating the onset of schizophrenia to a critical decrease in synaptic density [40]. We do not mean that a critically low synaptic density is the only possible cause of schizophrenia and that reduced synaptic density must always originate through the exact mechanisms that Feinberg proposed. Instead, we claim it is not surprising that large exponents belong to schizophrenic patients since those exponents predict a decreased synaptic density, which is an important factor that may lead to schizophrenia [41, 19]. Our work suggests that the exponent of Zipf's law could be used to detect synaptic density alterations and more importantly, disconnected brain areas.

The model presented here indicates a track for other related cases. While schizophrenics with large exponents seem to face the problem of the destruction of connectedness, children seem to face an inverse problem, i.e. the development

10

of connectedness. The relatively short time elapsed from the single-word to multiple-word utterances (of the order of several months [34]), suggests that the emergence of syntactic communication in children could be a phase transition to connectedness in the network of word syntactic interactions. According to our model of a rudimentary form of language, that transition would be an epiphenomenon of a transition to connectedness in the network of signal-stimuli associations. Whether the presumable phase transition would be continuous or not would depend of the presence or not of a special signature: scaling in the distribution of connected component sizes in the network of word syntactic interactions. We know that the network of syntactic interactions of adults is (almost) connected [22] but the signature above may be found in children at a critical time. To sum up, our findings open new research prospects and support that Zipf's law is an essential aspect of human language.

Appendix

We assume k is a random discrete variable whose probability is

$$P(k) = ck^{-\beta} \tag{22}$$

where β is a constant and

$$c = \frac{1}{\sum_{k=1}^{m} k^{-\beta}} \tag{23}$$

is a normalization term. $\langle k \rangle$, the mean value of k is

$$\langle k \rangle = c \sum_{k=1}^{m} k^{1-\beta}. \tag{24}$$

We can approximate $\langle k \rangle$ replacing summations by integrals and write

$$\langle k \rangle \approx \frac{\int_1^m k^{1-\beta} dk}{\int_1^m k^{-\beta} dk}. \tag{25}$$

Solving the integrals we obtain

$$\langle k \rangle \approx \frac{(1-\beta)(m^{2-\beta}-1)}{(2-\beta)(m^{1-\beta}-1)}. \tag{26}$$

For $m \to \infty$ and $\beta > 2$, we get

$$\langle k \rangle \approx \frac{1-\beta}{2-\beta}. \tag{27}$$

Acknowledgments

11

714

References

[1] C. E. Shannon, Bell Systems Technical Journal **27**, 379 (1948).

[2] R. B. Ash, *Information Theory* (John Wiley & Sons, New York, 1965).

[3] R. Ferrer i Cancho and R. V. Solé, Proc. Natl. Acad. Sci. USA **100**, 788 (2003).

[4] R. Ferrer i Cancho, Physica A **345**, 275 (2004), doi:10.1016/j.physa.2004.06.158.

[5] R. Ferrer i Cancho, Submitted to Eur. Phys. J. B (2005).

[6] G. K. Zipf, *Human behaviour and the principle of least effort. An introduction to human ecology* (Hafner reprint, New York, 1972), 1st edition: Cambridge, MA: Addison-Wesley, 1949.

[7] F. Pulvermüller, *The neuroscience of language. On brain circuits of words and serial order* (Cambridge University Press, Cambridge, 2003).

[8] M. A. Nowak, Phil. Trans. R. Soc. Lond. B **355**, 1615 (2000).

[9] R. Ferrer i Cancho, O. Riordan, and B. Bollobás, Proc. R. Soc. Lond. Series B **272**, 561 (2005).

[10] R. Ferrer i Cancho, Eur. Phys. J. B **44**, 249 (2005).

[11] A. Akmajian, R. A. Demers, A. K. Farmer, and R. M. Harnish, *Linguistics. An Introduction to Language and Communication* (MIT Press, 1995).

[12] R. Köhler, Theor. Linguist. **14**(2-3), 241 (1987).

[13] R. Ferrer i Cancho and R. V. Solé, J. Quantitative Linguistics **8**(3), 165 (2001).

[14] M. A. Montemurro, Physica A **300**, 567 (2001).

[15] V. K. Balasubrahmanyan and S. Naranan, J. Quantitative Linguistics **3**(3), 177 (1996).

[16] N. Chomsky, Paper presented at the Univ. de Brasilia, Nov. 26 (1996).

[17] C. F. Hockett, *A course in modern linguistics* (McMillan, New York, 1958).

12

[18] K. von Frisch, Scientific American **207**, 79 (1962).

[19] K. T. Mueser and S. R. McGurk, The Lancet **363**, 2063 (2004).

[20] B. Elvevåg and T. E. Goldberg, Critical Reviews in Neurobiology **14**, 1 (2000).

[21] C. Knight, M. Studdert-Kennedy, and J. R. Hurford, in *The evolutionary emergence of language* (Cambridge University Press, Cambridge, 2000), pp. 303–322.

[22] R. Ferrer i Cancho, R. V. Solé, and R. Köhler, Physical Review E **69**, 051915 (2004).

[23] R. Ferrer i Cancho, Physical Review E **70**, 056135 (2004).

[24] T. W. Deacon, *The Symbolic Species: the Co-evolution of Language and the Brain* (W. W. Norton & Company, New York, 1997).

[25] J. Hurford, The Times Literary Supplement (October 1998).

[26] R. Hudson, Journal of Pragmatics (August 1999).

[27] B. Bollobás, *Modern graph theory*, Graduate Texts in Mathematics (Springer, New York, 1998).

[28] P. Erdös and A. Rényi, Publ. Math. Inst. Hung. Acad. Sci. **5**, 17 (1960).

[29] B. Bollobás, *Random graphs*, Cambridge Studies in Advanced Mathematics, vol. 73 (Cambridge University Press, 2001), 2nd Edition.

[30] M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. E **64**, 026118 (2001).

[31] V. E. Stepanov, Theory Probab. Appl. **15**, 187 (1970).

[32] T. Harris, *The theory of branching processes* (Prentice-Hall, N.J, 1963).

[33] J. Binney, N. Dowrick, A. Fisher, and M. Newman, *The theory of critical phenomena. An introduction to the renormalization group* (Oxford University Press, New York, 1992).

[34] C. J., H. Davis, and M. Macken, in A. Lock and C. R. Peters, eds., *Handbook of human symbolic evolution* (Blackwell, Oxford, 1999), pp. 686–746.

[35] R. G. Piotrowski, V. E. Pashkovskii, and V. R. Piotrowski, Automatic Documentation and Mathematical Linguistics **28**(5), 28 (1995), first published in Nauchno-Tekhnicheskaya Informatisiya, Seriya 2, Vol. 28, No. 11. pp. 21-25, 1994.

[36] B. Elvevåg and T. Goldberg, CNS Spectrums **2** (1997).

13

[37] E. Breuler, *Dementia Praecox or the group of schizophrenia* (International University Press, New York, 1911/1950).

[38] G. M. Innocenti, F. Ansermet, and J. Parnas, Mol. Psychiatry **8**, 261 (2003).

[39] M. S. Keshavan, S. Anderson, and J. Pettegrew, J. Psychiatric Res. **28**, 239 (1994).

[40] I. Feinberg, J. Psychiatric Res. **17**, 319 (1982).

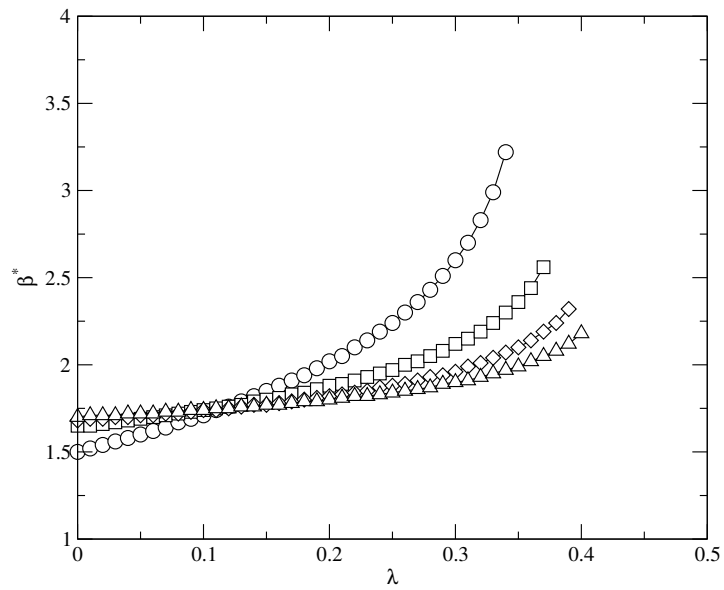[41] P. J. Harrison, Current Opinion in Neurobiology **7**, 285 (1997).

14

Figure 1: β^*, the value of β maximizing Ω for $n = m = 10$ (circles), $n = m = 10^2$ (squares), $n = m = 10^3$ (diamonds) and $n = m = 10^4$ (triangles). β is the exponent of Zipf's law, Ω is the energy function that communication maximizes, n is the number of signals and m is the number of stimuli. λ tunes the balance between information transfer and cost of communication. When $\lambda = 0$ communication is totally balanced towards saving the cost of communication whereas it is totally balanced towards information transfer when $\lambda = 0$.

15
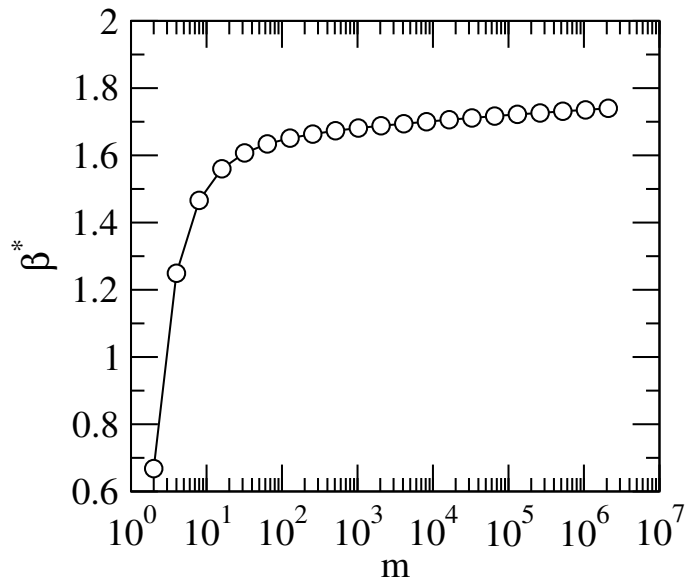
Figure 2: β^*, the value of β minimizing $H(S)$ versus m. $H(S)$ is the signal entropy and m is the number of stimuli.
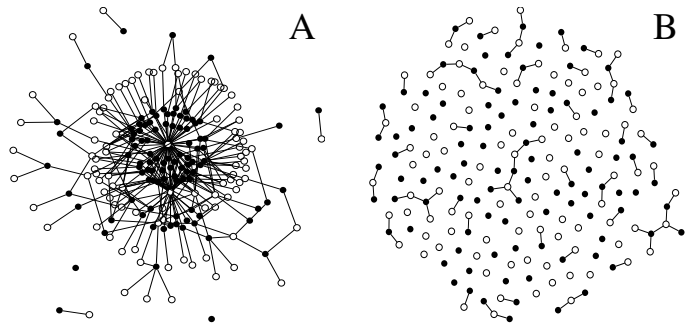
Figure 3: Two bipartite networks. White and black are used for each vertex partition. A. An almost connected networks. B. A disconnected network.
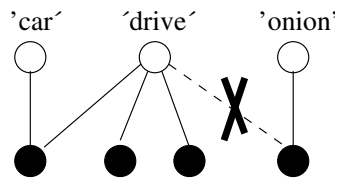
Figure 4: A possible implementation of the constrains of the verb 'drive' with possible arguments 'car' and 'onion'. White circles are words and black circles are stimuli. 'car' is an allowed argument of the verb 'drive' and therefore there is a link between 'drive0 and a stimulus associated to 'car'. 'onion' is not a valid argument of 'drive', so no stimulus linked to 'drive' is linked to 'onion'. ξ_{ik}, the number of shared stimulus by the pair (s_i, s_k), is 1 for ('drive','onion') and 0 for ('drive','onion').
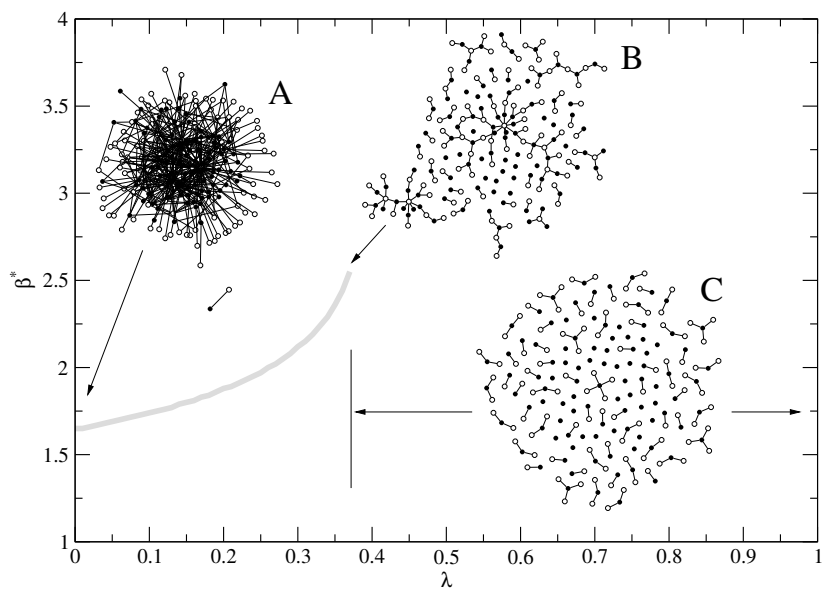
Figure 5: The evolution of β^* versus λ (gray curve) and the structure of network of signal-stimulus associations with $n = m = 100$. White and black circles indicate, respectively, signals and stimuli. The curve for β^* ends at the point of divergence at $\lambda = \lambda^* \approx 0.37$. A, B and C are examples of the kind of the topologies found for $\lambda = 0$, $\lambda = \lambda^*$ and $\lambda > \lambda^*$.
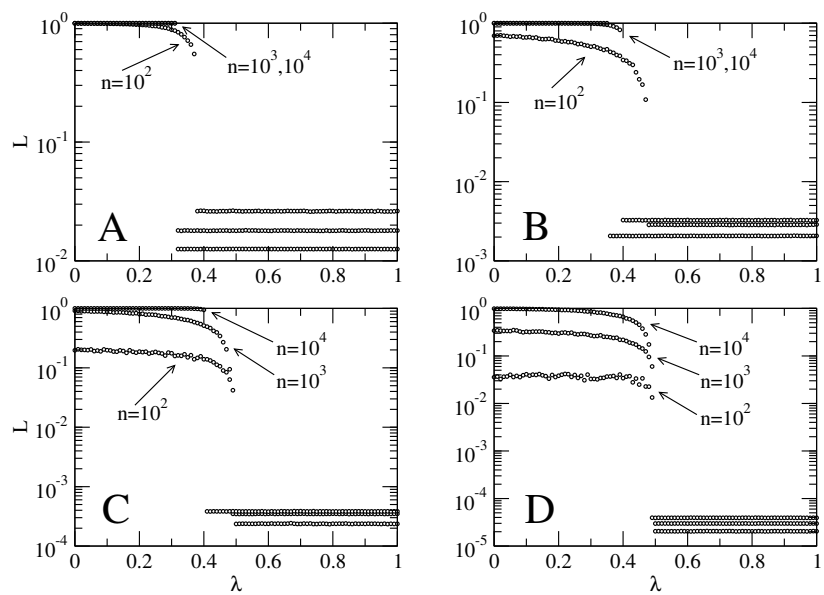
19

Figure 6: L, the normalized size of the largest connected component versus λ, the parameter controlling the balance between $I(S, R)$ and $H(S)$ in Ω. The connected component size is measured in vertices. n is the number of signals and m is the number of objects. A. $m = 10^2$. B $m = 10^3$. C. $m = 10^4$. D. $m = 10^5$.
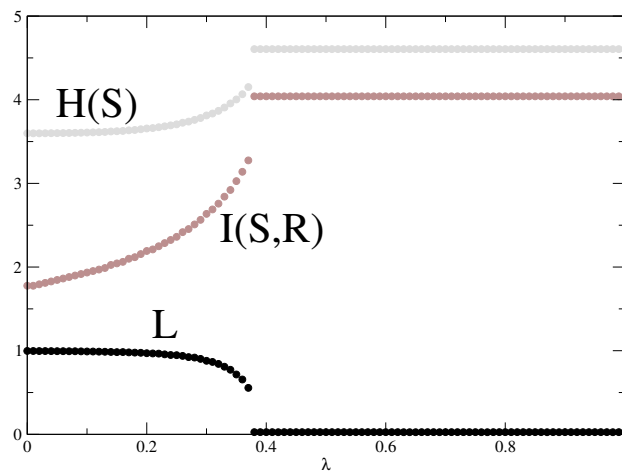
20

723

Figure 7: An example of the behavior of L (black), the normalized size in vertices of the largest connected component, $I(S, R)$ (dark gray), the information transfer and $H(S)$ (light gray), the signal entropy, versus λ, the parameter regulating the balance between maximizing $I(S, R)$ and $H(S)$ in Ω. A sudden change of behavior is found for $\lambda \approx 0.37$. $n = m = 100$ was used.

21