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

Ramon Ferrer i Cancho

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

Received 11 July 2005; received in revised form 30 November 2005; accepted 1 December 2005

Abstract
Here, we study a communication model where signals associate to stimuli. The model assumes that signals follow Zipf’s law and thus the exponent of the law depends on a balance between maximizing the information transfer and saving the cost of signal use. 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. 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, i.e. decreased synaptic density may lead to schizophrenia. Our findings suggest 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

1. Introduction
The XX century witnessed the birth and development of information theory (Shannon, 1948; Ash, 1965), 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 (Ferrer i Cancho and Sole, 2003; Ferrer i Cancho, 2005a,d). Word frequencies in human language obey a universal regularity, the so-called Zipf’s law (Zipf, 1972). If \( P(f) \) is the proportion of words whose frequency is \( f \) in a text, we obtain

\[
P(f) \sim f^{-\beta},
\]

where we typically have \( \beta \approx 2 \). Eq. (1) is a way of defining Zipf’s law. Zipf’s law is a regularity that appears in many contexts (Li, 2002; Newman, 2005). The ubiquity of Zipf’s law is the origin of many misunderstandings. First, the fact that Zipf’s law is everywhere (Li, 2002) does not imply that Zipf’s law is the only frequency distribution, not even the most common one. A recent compilation of Zipf’s law contains hundreds of distributions (Wimmer and Altman, 1999) among which one is the...
The recent information theory models mentioned at the beginning of the article 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), auditory stimuli (e.g. barking), etc. All these stimuli are elicited by the word ‘dog’ (Palermomüller, 2003). Stimuli are sometimes called objects or events in the origins of language literature (e.g. Nowak, 2000a; Ferrer i Cancho et al., 2005). Those models assume a set of \( n \) signals \( S = \{s_1, s_2, \ldots, s_n\} \) and a set of \( m \) stimuli \( R = \{r_1, r_2, \ldots, r_m\} \). 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 (Shannon, 1948), the goal of communication through isolated signals is maximizing \( H(S, R) \), 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 \( H(S, R) \) alone, which would lead to \( \beta \to \infty \). Zipf’s law with exponents close to the typical values are obtained when \( H(S, R) \) 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 (Ferrer i Cancho and Solé, 2003; Ferrer i Cancho, 2005c,d). 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 (Akmajian et al., 1995). 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 \( H(S) \) (Ash, 1965). In contrast, when a word has probability one (which implies that the remaining words have probability zero), \( H(S) = 0 \), which is the minimum value of \( H(S) \) (Ash, 1965). \( H(S) \) is a measure of the cost of communication, more precisely, of the cost of signal use. 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 the en-
 tropy of units in many computer or engineering problems but not in real brain word access and retrieval.

If we restrict ourselves to Shannon’s classic information theory, the goal of a communication system is to maximize the function

\[ I(S, R) = \log_2 \frac{1}{\sum_{k=1}^{M} p(k)} \]  

(2)

If we take into consideration the cost of signal use, we may write

\[ I_\Omega(S, R) = \lambda I(S, R) - (1 - \lambda) H(S) \]  

(3)

as the function that a natural communication system should maximize (Ferrer i Cancho and Solé, 2003; Ferrer i Cancho, 2005c,d). \( \lambda \) is a parameter controlling the balance between maximizing the information transfer and minimizing the cost of signal use. We assume \( \lambda \in [0, 1] \).

We have \( I_\Omega = I_\Omega(S) \) when \( \lambda = 0 \). \( I_\Omega \) is suitable for computer or robotic problems where \( H(S) \) can be neglected and \( I_\Omega \) (with \( \lambda < 1/2 \) (Ferrer i Cancho, 2005c,d)) is specifically suitable for brain based communication systems. \( I_\Omega \) seems, a priori, a better choice than \( I_\Omega \) for natural communication systems.

We do not claim that \( I_\Omega \) is the best function for natural communication systems but there are some results supporting its usefulness.

- Maximizing \( I_\Omega \), Zipf’s law is obtained for a particular value of \( \lambda \). If one replaces \( H(S) \) in Eq. (3) by the effective lexicon size, namely, the number of signals with at least one association with stimuli, Zipf’s law is not obtained (Ferrer i Cancho and Solé, 2003; Ferrer i Cancho, 2005d). Vocabulary size is an important factor for the cost of word use (Köhler, 1987) but does not seem to be essential for Zipf’s law: We define \( H(R; S) \) as the conditional entropy of stimuli when signals are known. Zipf’s law is still reproduced if \( I_\Omega \) is replaced by \( -H(R; S) \) in the model in Ferrer i Cancho and Solé (2003), but not in the model in Ferrer i Cancho (2005d).
- The exponent of Zipf’s law in single author text satisfies \( \beta \in \{1.6, 2.4\} \) (Ferrer i Cancho, 2005c). Maximizing \( I_\Omega \) in a system following Zipf’s law (i.e. searching the value of \( \beta \) maximizing \( I_\Omega \)) can explain the interval of variation of \( \beta \) in human language (Ferrer i Cancho, 2005c).

If one considers texts from a single author (Ferrer i Cancho and Solé, 2003; Möttemüller, 2001) and does not concentrate on words of a certain type (e.g. nouns) (Balasubramanyan and Nairnian, 1996; Ferrer i Cancho, 2005a), the extremes of the interval of variation of \( \beta \) correspond to schizophrenic patients (Ferrer i Cancho, 2005c). The aim of the present paper is deepening our understanding of what may happen when \( \beta \) is large and, in particular, what may be happening in schizophrenics with that \( \beta \). We will show 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 \( \lambda \) takes a critical value in a communication system following Zipf’s law.

3. The model

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

\[ p(k) \sim k^{-\mu} \]  

(4)

where \( p(k) \) is the probability that a signal has \( k \) connections. Second, \( \mu_i \sim \mu_i^* \), where \( p(s_i) \) is the probability of using \( i \) and \( \mu_i \) the number of links of that stimulus.

Eq. (4) and \( \mu_i \) give Eq. (1). Various models recover Zipf’s law when maximizing \( I_\Omega \) without the constraint in Eq. (4) for a critical value of \( \lambda \) (Ferrer i Cancho and Solé, 2003; Ferrer i Cancho, 2005d).

Going further, we assume

\[ p(s_i) = \frac{\mu_i}{M} \]  

(6)

where

\[ M = \sum_{i=1}^{M} \mu_i \]  

(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 \( \beta \) in humans (Ferrer i Cancho, 2005c).

Interestingly, Eq. (6) makes some important assumptions that need to be made explicit. 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

\[ p(s_i, r_j) = \frac{\omega_{ij} p(r_j)}{\omega_j} \]  

(8)

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

\[ \omega_j = \sum_{k=1}^{n} \omega_{kj} \]  

(9)

is the number of links of that stimulus.
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 $P = \{(P(1), \ldots, P(k), \ldots, P(m))\}$ and the distribution of links per stimulus is given by $Q = \{Q(0), \ldots, Q(k), \ldots, Q(n)\}$, 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 Ferrer i Cancho (2005c), that is

$$Q \sim \text{binomial} \left( \frac{1}{m} \right),$$

where $(\ldots)_p$ 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 $\Omega$ assuming $p(r_j) \sim \omega_j$ (or $p(s_k) \sim \mu_j$) for any pair of $P$ and $Q$. The calculation of $\Omega$ is straightforward once we know (Ferrer i Cancho, 2005c,d)

$$H(S) = \log M - H(R|S)$$

(15)

$$H(R(S)) = \log M - H(S|R)$$

(16)

where $M = n(k_p) = m(k_{Q})$ and

$$H(R(S)) = \frac{(k \log k)_P}{\langle k \rangle_P}$$

(17)

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

(18)

The present model integrates two recent results. The first result is that $\beta^*$, the value of $\beta$ maximizing $\Omega$, grows with $\lambda$, till $\lambda = \lambda^*$. Beyond ($\lambda > \lambda^*$), we have $\beta \rightarrow \infty$.
Fig. 2. \( \beta^* \), the value of \( \beta \) minimizing \( H(S) \) vs. \( m \). \( H(S) \) is the signal entropy and \( m \) is the number of stimuli. (Ferrer i Cancho, 2005c). The behavior of \( \beta^* \) is illustrated in Fig. 1. It can be shown that \( \lambda^* < 1/2 \) and a heuristic argument suggests the existence of a discontinuity at \( \lambda = \lambda^* \) Ferrer i Cancho (2005c). The idea is very simple. Eq. (3) can be written as

\[
\Omega = (2\lambda - 1)\lambda H(S) - \lambda H(S|R) \quad (19)
\]

knowing that \( I(S, R) = H(S) - H(S|R) \) (Ash, 1965). Eq. (19) indicates that maximizing \( \Omega \) minimizes \( H(S) \) if \( \lambda < 1/2 \) and maximizes \( H(S) \) if \( \lambda > 1/2 \). Notice that maximizing \( \Omega \) for \( \lambda = 0 \) gives a finite value of \( \beta^* \) (Fig. 2). \( \beta \) must diverge for \( 0 < \lambda < 1/2 \). Since maximizing \( \Omega \) 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 (Ferrer i Cancho et al., 2005). 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. 3A and B shown, 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 (Ferrer i Cancho et al., 2005). 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 (Ferrer i Cancho et al., 2005). Connectedness is intimately related to two essential traits that researchers have identified as essential aspects of human language: syntax and symbolic reference (Knight et al., 2000). 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 Ferrer i Cancho et al. (2005), 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

\[
\xi_{ik} = \sum_j a_{ij}a_{jk}. \quad (20)
\]
According to Deacon, an essential aspect of symbolic reference is that real words do not only evoke stimuli (or meanings) but also other words (Deacon, 1997). Deacon tried to define symbolic reference but his proposal has been criticized due its lack of precision (Hurford, 1998; Hudson, 1999). Taking the idea of ‘signals evoking other signals’, Ferrer i Cancho et al. (2004) have defined symbolic reference as connectedness in the network of signal–stimulus associations. The definition is not ambiguous and relies on standard concepts of graph theory (Bollobás, 1998). 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 may 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

$$L = \frac{1}{n + m}$$

where $n$ 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 $\lambda$.

4. Results

For each value of $\lambda$,
Fig. 6. \( L \), the normalized size of the largest connected component vs. \( \lambda \), the parameter controlling the balance between \( I(S, R) \) and \( H(S) \) in \( \Omega \). 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) \( m = 10^5 \).

We obtained \( \beta^* \), value of \( \beta \) maximizing \( \Omega \), 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 choosing stimuli at random (all stimuli are equally likely so Eq. (14) follows). Means were calculated over 1000 replicas.

Fig. 5 shows the evolution of a small network of signal–stimulus associations as \( \lambda \) grows. At a critical value of \( \lambda \), the size of the largest connected component falls abruptly. In general, \( L \) falls abruptly to a small value for \( \lambda = \lambda^* \) (Fig. 6). \( \lambda^* \) is the point where \( \beta \) diverges and \( I(S, R) \) and \( H(S) \) reach their maximum value (Ferrer i Cancho, 2005c). The steepness of the fall grows with \( n \).

5. Discussion

We have seen that a communication system maximizing \( \Omega \) undergoes an abrupt transition to disconnectedness for \( \lambda > \lambda^* \). We have seen that the transition is

Fig. 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 vs. \( \lambda \), the parameter regulating the balance between maximizing \( I(S, R) \) and \( H(S) \) in \( \Omega \). A sudden change of behavior is found for \( \lambda \approx 0.37 \). \( n = m = 100 \) was used.
caused by a sudden jump from a finite value of \( \lambda \) to \( \lambda \to \infty \), where the chance that a stimulus has two links
vanishes as \( m \) grows. The disconnection of the network
when \( \lambda \to \infty \) 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 (Bollobás, 1998). 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 \( (k)_p < (n + m - 1)/n \). When \( \lambda \to \infty \), we
have \( (k)_p = 1 \) and \( (k)_p < (n + m - 1)/n \) holds trivially
provided \( m > 1 \). In sum, connectedness is impossible for
\( \lambda \to \infty \) and \( m > 1 \).

In Section 3, we have reviewed a heuristic argument
suggesting the transition from a highly connected phase
to disconnectedness in our model is discontinuous. Dis-
continuous phase transitions are widespread in nature.
For instance, the melting of ice into water or the trans-
formation of boiling water into vapour are discontinuous
in normal circumstances. In a communication context,
the models in Ferrer i Cancho and Solé (2003). Ferrer
i Cancho (2005d) shows a continuous phase transition
between no communication and a perfect communica-
tion phase when \( \Omega \) is minimized with no constraint on
\( P \). There, the presence of Zipf’s law in the victims of
an abrupt change is the hallmark of a continuous phase
transition. In contrast, the phase transition from discon-
ectedness to connectedness in a classic Erdős–Rényi
graph (Erdős and Rényi, 1960; Bollobás, 2001) is con-
tinuous (Newman et al., 2001; Stepanov, 1970). The hall-
mark of continuous phase transition in classic unipartite
graphs is a power distribution of connected component
sizes (Newman et al., 2001), which is related to a criti-
ical branching process (Harris, 1963) at the threshold
for connectedness. Other examples of continuous phase
transitions are the transition from resistivity to super-
conductivity (continuous in the absence of an external
magnetic field) and the conversion of iron from param-
agnetic to ferromagnetic form (Binney et al., 1992). In
a communication context, the model examined here not
only apparently shows a discontinuous transition to dis-
connectedness but also to maximum information transfer
and maximum cost for \( \lambda = \lambda^* \) (Ferrer i Cancho, 2005d).

The divergence of \( \beta \) for \( \lambda = \lambda^* \) is accompanied by
a jump to maximum information transfer (Fig. 7). In-
creasing \( \lambda \) increases \( I(S, R) \) but decreases the size of the
largest connected component (the significance of the de-
crease depends on the size of the system). At the point
where the \( I(S, R) \) is maximum, \( k = \lambda \) is minimum. In a com-
munication system maximizing \( \Omega \), communication us-
ing isolated signals and language are in conflict. Human
speakers may need to regulate \( \lambda \) in order to maximize
information transfer but avoid reducing the size of the
largest connected component too much. Interestingly,
the regulation of the size of the largest connected com-
ponent can be done indirectly because increasing \( I(S, R) \)
also increases \( H(S) \), the cost of signal use. Word ambi-
guity 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
\( \lambda \) (giving more weight to minimizing \( H(S) \)) maximizes
the chance of connectedness. The emergence of connect-
edness could be a side effect of saving the cost of signal
use.

A theory of word frequencies needs answering differ-
ent questions:

1. Why do words arrange themselves according to
Zipf’s law (Eq. (1))?  
2. Why do humans choose some particular values of \( \beta \)?
3. Why is there variation in \( \beta \)?
4. What are the limits of that variation?
5. What is the link between Zipf’s law and human lan-
guage?

Many answers have been proposed for Questions 1-2
(Ferrer i Cancho, 2005d). As far as we know, Question 1
and 2 have only been answered assuming that words are
used according to their meaning in Ferrer i Cancho and
Solé (2003), Ferrer i Cancho (2005a). Choosing values
of \( \beta \) near 2 could be an optimal solution for a conflict
between maximizing the information transfer and saving
the cost of word use (Ferrer i Cancho and Solé, 2003;
Ferrer i Cancho, 2005d). Questions 3 and 4 have begun
be addressed in Ferrer i Cancho (2005c,a). The idea
is that the lower bound and the upper bounds of \( \beta \) are
obtained when maximizing \( \Omega \) for \( \lambda = 0 \) and \( \lambda = \lambda^* \), re-
spectively. The present article sheds new light on Ques-
tions 3, 4 and 5. As for Question 3, variation in \( \beta \) may be
due to the chance of connectedness. As for Question 4, it
has been argued that the variation of \( \beta \) is constrained by
the fact that maximizing \( \Omega \) for \( \lambda \in [0, 1] \) gives a narrow
interval of exponents (Ferrer i Cancho, 2005c). It has
been argued that the interval of variation of \( \beta \) 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 prac-
tice, significant differences in \( \log n \) between two differ-
ent 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
That is a compelling reason for not finding large $\beta$ in human language. We do not mean that large $\beta$ is impossible to attain in humans, but it would be surprising to find it in a system combining words through semantic constraints. As for Question 5, our work suggests that the exponent of Zipf’s law is an important factor for the presence or absence of language. In sum, the present article puts another step forward in the construction of a theory of word frequencies.

Till now, we have studied the implications of large exponents in a theoretical model. We would like to provide a framework that can offer new insights in real cases. Schizophrenics speakers with large exponents will receive special attention. For that reason, it is important to review the facts that provide support for the soundness of the theory. With the theoretical framework used here, two types of successful predictions have been previously made: general predictions and specific predictions for schizophrenia. As for the general predictions, we have seen two predictions that are made by minimizing $\delta(\omega)$ in Section 1. That is not all. Nouns have a greater exponent that all parts-of-speech mixed together, $\beta \in [2, 1, 3]$, approximately (Ferrer i Cancho, 2005c).

There is a wide consensus in linguistics and philosophy about the greater semantic specificity of nouns (e.g. the concept of rigidity of nouns in Kripke’s work (Kripke, 1990)). The theoretical approach followed in this article allows one to predict a higher semantic precision for nouns because of their higher exponent (see Ferrer i Cancho, 2005b) for the details of the argument). As for the predictions specific to schizophrenia, we will concentrate on the speech of schizophrenics with large $\beta$, containing many words related to the patients topic of obsession (Piotrowski et al., 1995). It has been shown that if $m$ is kept constant and $\beta < 2$, then $H(R(S))$ (a measure of word ambiguity) grows as $\beta$ decreases (Ferrer i Cancho, 2005b). The theory predicts that $H(R(S))$ diverges when $\beta < 2$ and expressivity is maximized ($i.e. m$ is maximized). Thus, $\beta$ must be kept small to avoid having too ambiguous words, which explains the onset of obsession (Ferrer i Cancho, 2005b). Notice that the lowering of the exponent (if $n$ remains constant) translates into a greater repetition of words, but the latter does not imply that the speech is circumscribed very particular topic (in our simplified model, a narrow topic corresponds to a small $m$).

Our theoretical approach makes a strong prediction: obsession at the level of the topic of the discourse, not only more repetition at the surface level of words. In sum, we believe that there is a critical mass of successful predictions allowing one to move to cases where there is no available information for testing the predictions or predictions cannot be easily tested. Nonetheless, we hope that that what follows is not taken as the ultimate explanation or as an unthoughtful hypothesis, but rather as a suggestive research track.

The largest values of $\beta$ than have been found up to now in single author text samples correspond to schizophrenic patients in the acute phase of the illness (Ferrer i Cancho, 2005c; Piotrowski et al., 1995). One of the most salient features of schizophrenia is ‘disorder of thought’ (Elvén and Goldberg, 1997). 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 (Bleuler, 1911/1950). 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 $\beta$, 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 $\lambda$ (and therefore $\beta$) is significant (recall Fig. 6). The larger the value of $\beta$, 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 (Pulvermüller, 2003). 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, $\beta$, 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 $\delta = M/nm$. Knowing $M = n\langle k \rangle_p$, we may write $\delta = \langle k \rangle_p/m$. It can be easily seen that $\langle k \rangle_p$ decreases with $\beta$ (see Appendix A and Ferrer i Cancho, 2005b). For large $m$ and $\beta > 2$ we have (see Appendix A)

$$\langle k \rangle_p \approx \frac{1 - \beta}{2 - \beta}$$

If our hypothetical correspondence between $\beta$ and synaptic density (or size in words of the largest component of the network of signal–stimulus associations) is accurate, we would expect $\beta$ to be larger in the acute phase of the illness than the chronic phase.
nected component) was correct, one would expect that the smallest synaptic density would be for the largest values of \( \beta \), which corresponds to schizophrenic patients in the acute phase (Piotrowski et al., 1995). Interestingly, it has been speculated that excessive synaptic pruning occurs in schizophrenia, which may lead to psychosis when it reaches a threshold (Mueser and McGurk, 2004; Innocenti et al., 2003; Keshavan et al., 1994). See McGlashan and Hoffman (2000) for a review of recent evidence for reduced connectedness in schizophrenia. Our work is consistent with the spirit of Feinberg’s hypothesis, relating the onset of schizophrenia to a critical decrease in synaptic density (Feinberg, 1982). 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 (McGlashan and Hoffman, 2000; Harrison, 1997; Mueger and McGurk, 2004).

Our work suggests that the exponent of Zipf’s law could be used to detect synaptic density alterations and more importantly, brain area disconnections.

We intend to study schizophrenic language and communication from a specific specific framework. It is technically impossible that our approach accounts for the wide range of features of schizophrenia. Neural network models have accounted for important aspects of schizophrenia such as its unique symptoms, short- and long-term course, typical age of onset, neurodegenerative, mental deficits, limited neurodegenerative progression and sex differences (Hoffman and McClashan, 2001; McGlashan and Hoffman, 2000). Our model should be seen as an attempt to cover a very specific dimension of schizophrenia. As far as we know, no model before has faced the alterations in the exponent of Zipf’s law and the implications for language. The aim of the present article is not providing an ultimate explanation about what happens in schizophrenics with large exponents but putting forward a strong theoretical hypothesis that would need further research. Future work should be devoted to test the correlation suggested by the model between high exponents and brain alterations. Unfortunately, the brain alterations in the patients with high \( \beta \) examined by Piotrowski et al. (1995) are not available in their work. Our model is abstract enough to embrace not only schizophrenics speakers with large exponents but also other kinds of pathological speakers exhibiting large exponents. Among those, patients with Alzheimer’s disease are specially interesting because of their loss of synapses (Hamos et al., 1989), although there is no study of Zipf’s law on Alzheimer’s disease, as far as we know. The use of schizophrenics instead of other is due to the fact the schizophrenias is, as far as we know, the only brain alteration where Zipf’s law has been studied.

The model presented here suggests a track for understanding non-pathological 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 of connectedness. The relatively short time elapsed from the single-word to multiple-word utterances (of the order of several months (Johnson et al., 1999)), 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 (Ferrer i Cancho et al., 2004) 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, rather than an curious regularity, is an essential aspect of human language.

Acknowledgments

We are grateful to Vito Servedio for helpful comments. Discussions with Toni Hernández and Brita Elvevåg have had a great influence for the arguments of that article. We specially thank Brita Elvevåg for suggesting many improvements. This work was funded by the RC-Agents project (RFC), funded by the Future and Emerging Technologies program (IST-FET) of the European Commission under the EU RD contract IST-1940. The information provided is the sole responsibility of the authors and does not reflect the Commission’s opinion. The Commission is not responsible for any use that may be made of the data appearing in this publication.

Appendix A

We assume \( k \) is a random discrete variable whose probability is

\[
P(k) = c k^{-\beta}\]  

(23)
where $\beta$ is a constant and

$$\epsilon = \frac{1}{\sum_{k=1}^{\infty} k^{-\beta}} \quad (24)$$

is a normalization term. $(k)$, the mean value of $k$ is

$$\langle k \rangle = \frac{\sum_{k=1}^{\infty} k^{-\beta}}{\sum_{k=1}^{\infty} k^{-\beta}} \quad (25)$$

We can approximate $(k)$ replacing summations by integrals and write

$$\langle k \rangle \approx \int_{1}^{\infty} \frac{k^{-\beta} \, dk}{\int_{1}^{\infty} k^{-\beta} \, dk} \quad (26)$$

Solving the integrals, we obtain

$$\langle k \rangle \approx \left( 1 - \frac{\beta m^{-\beta}}{m^{-\beta} - 1} \right) \quad (27)$$

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

$$\langle k \rangle \approx \frac{1 - \beta}{2 - \beta} \quad (28)$$

References


