Glottometrics 11, 2005, 98-119

Hidden communication aspects in the exponent of Zipf's law

Ramon Ferrer i Cancho[1]

Abstract. This article focuses on communication systems following Zipf's law, in a study of the relationship between the properties of those communication systems and the exponent of the law. The properties of communication systems are described using quantitative measures of semantic vagueness and the cost of word use. The precision and the economy of a communication system is reduced to a function of the exponent of Zipf's law and the size of the communication system. Taking the exponent of the frequency spectrum, it is demonstrated that semantic precision grows with the exponent, whereas the cost of word use reaches a global minimum between 1.5 and 2, if the size of the communication system remains constant. The exponent of Zipf's law is shown to be a key aspect for knowing about the number of stimuli handled by a communication system, and determining which of two systems is less vague or less expensive. The ideal exponent of Zipf's law, it is therefore argued, should be very slightly above 2.

Keywords: Zipf's law, frequency spectrum, exponent, precision, economy

INTRODUCTION

Word frequencies in human language arrange themselves according to what is known as Zipf's law. If *P*(*f*) is the proportion of words whose frequency is *f* in a given sample (e.g. a text), we say that a sample follows Zipf's law (Zipf, 1932, 1935, 1949) if

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

where β is the exponent of the law. We assume that $\beta > 1$.

The previous equation appears as a straight line when *P*(*f*) is plotted on a logarithmic scale. Although different functions have been proposed for modelling *P*(*f*) (Chitashvili & Baayen, 1993; Tuldava 1996; Naranan & Basubrahmanyan, 1998), the basic trend described in simplified form by Eq. 1 appears to hold without exceptions in word frequencies. This article uses the functional form in Eq. 1 because its simplicity is extremely helpful for the analytical calculations discussed here.

Typically, $\beta \approx 2$ is found (Zipf, 1932, 1935, 1949) but significant deviations from that value have been reported in single author samples:

- $\beta > 2$ in fragmented discourse schizophrenia. This type of speech is characterized by multiple topics and the absence of a consistent subject. The lexicon of such a text may be varied and chaotic (Piotrowski *et al.* 1995, Piotrowska *et al.,* to appear). $\beta \in [2.11, 2.42]$ is found. Schizophrenic patients of this kind tend to be in the acute phase of the disease.

[1] Address for correspondence: Ramon Ferrer i Cancho, Dip. di Fisica, Università 'La Sapienza', Piazzale A. Moro 5, ROMA 00185, ITALY. E-mail: ramon@pil.phys.uniroma1.it

- Values suspiciously above the ideal $\beta = 2$ have been found in nouns from single author samples. More precisely, $\beta \in [2.15,2,32]$ (Balasubrahmanyan & Naranan, 1996).
- $1 < \beta < 2$ in advanced forms of schizophrenia (Whitehorn & Zipf, 1943; Zipf, 1949; Piotrowski *et al.*, 1995; Piotrowska *et al.*, to appear). Texts are filled mainly with words and word combinations related to the patient's obsessional topic. The variety of lexical units employed here is restricted and repetitions are many. $\beta = 1.66$ is reported in (Piotrowski *et al.* 1995; Piotrowska *et al.*, to appear).
- $\beta = 1.6$ in very young children (Brillouin, 1960; Piotrowski *et al.*, 1995). Older children conform to the typical $\beta \approx 2$ (Zipf, 1942).
- Exponents larger than $\beta \approx 2$ can be obtained as a result of deficient sampling from a text with the typical $\beta \approx 2$ (Piotrowski *et al.*, 1995; Piotrowska *et al.*, to appear).

Therefore, the exponents that are of interest here seem to be constrained to a very narrow domain, i.e. $\beta \in [1.66,2.42]$ (Ferrer i Cancho, 2005b). Whether Zipf's law can distinguish between acute and chronic schizophrenic patients is a matter of current research. The main message concerning schizophrenia here is that the disease shows exponents on both sides of the interval of variation in humans and that the value of the exponent may be related to the stage of the disease. Significant variations of β have also been found in multi-author samples (Piotrowski *et al*., 1994; Ferrer i Cancho, 2005d, Ferrer i Cancho & Solé, 2001; Montemurro, 2001; Montemurro & Zanette, 2002), particular word classes (Balasubrahmanyan & Naranan, 1996) and both (Ferrer i Cancho, 2005a).

The focus of the present paper is communicative aspects of single individuals. Significant deviations in multi-author texts will not be considered. The aim of the present paper is to show the connection between the exponents and various types of quantitative measures suggesting that the variation of the exponent may be due to tuning the vagueness and the cost of word use. Most of the measures of vagueness and the measure of cost of word use that are employed here are defined using Shannon's information theory (Ash, 1990). Support for the hypothesis of the strong association between Zipf's law and communication, comes from recent models where Zipf's law and/or the value of the exponent can be explained as the outcome of minimizing or constraining various standard information theory measures (Ferrer i Cancho, 2005a, 2005d; Ferrer i Cancho & Solé, 2003).

THE MODEL

We assume a general communication system mapping words to stimuli. We have a set of n words $S = \{s_1, ..., s_i, ..., s_n\}$ and a set of m stimuli $R = \{r_1, ..., r_j, ..., r_m\}$. We assume that words connect to stimuli to build their meaning. Word-stimuli associations are defined by a binary matrix $A = \{a_{ij}\}$ where $a_{ij} = 1$ if s_i and r_j are linked and $a_{ij} = 0$ otherwise. Let us consider in greater detail what is meant here by "stimuli". Various experiments have shown that words are associated with the activation of different brain areas (Pulvermüller, 2003). Generally speaking, nouns tend to activate visual areas. Verbs tend to activate motor areas if the corresponding action can be performed by the individual and visual areas otherwise. The activated areas are associated to different types of stimuli experienced with the word. Let us take one of the definitions of the Webster's Revised Unabridged Dictionary (1913)[2] for the word *write*: "to inscribe on any material by a suitable instrument". In our view, the verb *write* is associated to the motor stimuli of the action of writing and the visual (tactile, olfactory,...) stimuli of the instruments used for writing. The construction of a complex meaning would

[2] www.dict.org.

involve a structure combining diverse stimuli. From that point of view, a word in *S* does not refer to stimuli in *R*, but it is merely associated to them. We do not claim that words in *S* refer to stimuli in *R* via *A* although they may. We do not use the term reference because it is stronger than association. In our example, *write* can only refer to the motor stimuli of the action of writing. *write* cannot refer to the instrument used for writing, although it is associated with it. The action and the instrument are both stimuli involved in the construction of the complex meaning of the verb *write*. Defining "word meaning" is an open problem in various fields ranging from cognitive science to philosophy. In our view, complex meaning would emerge from the interaction between different stimuli. Referential associations may be a subset or a higher order structure of the associations defined by *A*. It makes sense to assume that the more stimuli a word is associated with, the higher the probability of using that word.

It is important to note that when we say that a word has no meaning we usually mean that it has no referential power. Nonetheless, if a word lacks referential power it does not imply that it has no associations with stimuli. Our framework is not inconsistent with the existence of words with no apparent meaning, such as prepositions, conjunctions or articles. Real words with no apparent meaning are the words with the highest frequencies. The five most frequent word in the British National Corpus[3], a large collection of text samples, are *the*, *of*, *and*, *to* and *a*. The framework here predicts that the most frequent words would have the largest number of connections with stimuli in *R*. Since those connections are merely associative (and *not* always referential) there is no inconsistency here. Furthermore, that high number of associations may underlie those words' lack of referential power or "meaning". The uncertainty associated with the interpretation of highly connected words is so large (Ferrer i Cancho, 2005c,e) that reference cannot be effectively attributed. Words with no meaning may have two different origins: words that have no links, and words having too many links. It makes sense to suppose that words with no meaning may have an excess of connections rather than a lack thereof, although those connections could be very weak given the high frequency of the words involved (Ferrer i Cancho & Reina, 2002).

A first approach to the semantic vagueness of the set of words could be the average number of links per word, that is, $<k>$. The number of links of a vertex (e.g. a word) is called "degree" in standard graph theory (Bollobás, 1998), so $<k>$ is the mean signal degree. The idea behind the relationship between $<k>$ and vagueness is very simple: the more links a word has, the higher the number of possible interpretations in the context where it appears. The higher the value of $<k>$, the lower the precision of the communication system. Hereafter we assume that 'precision' and 'vagueness' have opposite meaning. $H(R|\, S)$, that is, the average uncertainty (or entropy) associated with the interpretation of every stimulus once the corresponding word is known, is a more precise measure, from the information theory point of view. That measure is defined as

$$H(R \mid S) = \sum_{i=1}^{n} p(s_i) H(R \mid s_i)\,, \qquad (2)$$

where $H(R|s_i)$ is the uncertainty (or entropy) associated with the interpretation of s_i , and $p(s_i)$ is the probability of using s_i. $H(R|S)$ is the average uncertainty associated with the interpretation of the words in *S*. The higher the value of $H(R|S)$, the lower the precision of the communication system. Since $H(R|S)$ is mathematically a hard function to manipulate, a simpler version has been considered (Ferrer i Cancho, 2005a):

[3] www.natcorp.ox.ac.uk

$$G(R \mid S) = \frac{1}{n}\sum_{i=1}^{n} H(R \mid s_i)\,. \tag{3}$$

$G(R|S)$ is the amount of uncertainty per word associated with the interpretation of the words in S. $G(R|S)$ and $H(R|S)$ have similar properties. The upper and lower bounds are the same, i.e. $0 \le G(R \mid S), H(R \mid S) \le \log m$. $G(R|S)$ has the virtue of allowing Zipf's law (Eq. 1) to be derived using the maximum entropy principle (Ferrer i Cancho, 2005a).

A possible approach to the cost of word use is $H(S)$, the entropy of the set of words (Ferrer i Cancho, 2005a,d; Ferrer i Cancho & Solé, 2003). This is defined as

$$H(S) = -\sum_{i=1}^{n} p(s_i)\log p(s_i)\,. \tag{4}$$

Support for $H(S)$ as a measure of the cost of word use comes from two different sources. Firstly, it is known in psycholinguistics that the availability of a word in various linguistic tasks is correlated with the frequency of that word. The availability of a word obeys the so-called *word frequency effect*, i.e. the more frequent the word, the higher its availability (Akmajian *et al.*, 1995; Carroll, 1994). The best availability is achieved when a word has probability one, which means that the rest of the words have probability 0. In that case, $H(S) = 0$. The worst case is when all words are equally likely, that is when $p(s_i) = 1/n$ for each word. In that case, $H(S) = \log n$. That is, $H(S)$ is a good measure of cost of word use. Second, the use of $H(S)$ as a measure of cost is justified by models leading to Zipf's law when the information transfer is maximized while $H(S)$ is minimized (Ferrer i Cancho, 2005d; Ferrer i Cancho & Solé, 2003). Those models explain Zipf's law as the outcome of maximizing the communicative efficiency, but saving as much cost as possible. Interestingly, if those models replace $H(S)$ with the effective vocabulary size (i.e. the proportion of words with at least one link) as a measure of cost of word use, Zipf's law is not reproduced. Vocabulary size is an important ingredient for the cost of a communication system (Köhler, 1986, 1987) but it does not seem to be essential for Zipf's law.

We may assume that the $p(s_i)$, the probability of occurrence of word s_i, is proportional to k_i, the number of connections of s_i, that is

$$p(s_i) = \frac{k_i}{M}, \tag{5}$$

where

$$k_i = \sum_{j=1}^{m} a_{ij} \tag{6}$$

and

$$M = \sum_{i=1}^{n} k_i \tag{7}$$

(as in Ferrer i Cancho, 2005a,b,d). Eq. 5 contains the basic assumption that words are used according to the number of semantic associations they have. Eq. 5 states that a word is used with a probability proportional to the number of stimuli it is associated with. Eq. 5 is chosen for simplicity and its predictive power: it can explain the interval of variation of β in human language (Ferrer i Cancho, 2005b).

We may also assume that $P(k)$, the proportion of words with k links obeys

$$P(k) \sim k^{-\beta}. \quad (8)$$

Zipf's law (Eq. 1) is recovered from Eqs. 5 and 8 (Ferrer i Cancho, 2005a,b). We assume a fixed $P(k)$ or $P(f)$, given the surprising tendency of human language to arrange according to Zipf's law even in atypical cases. Although there is variation in β for (human) words, the basic trend described by Eq. 1 has essentially no exceptions, as far as we know. From Eq. 5 and

$$p(s_i) = \sum_{j=1}^{m} p(s_i, r_j) \quad (9)$$

it follows that the probability that s_i and r_j are associated by the communication system is

$$p(s_i, r_j) = \frac{a_{ij}}{M}. \quad (10)$$

We may write Eq. 5 as

$$p(s_i \mid k_i = k) = \frac{k}{M} = \frac{k}{n < k >}, \quad (11)$$

where <...> is the expectation operator over $P = \{P(1),...,P(k),...P(m)\}$ and $P(k)$ is the proportion of words having k connections.

Assuming Eq. 8, the uncertainty (or entropy) associated to the interpretation of s_i becomes $H(R|s_i) = \log k$ if $k_i = k$ (Ferrer i Cancho, 2005a). Thus, $H(R|S)$ in Eq. 2 becomes (Ferrer i Cancho, 2005b)

$$H(R \mid S) = \frac{< k \log k >}{< k >} \quad (12)$$

and $G(R|S)$ becomes (Ferrer i Cancho, 2005a)

$$G(R \mid S) = < \log k >. \quad (13)$$

Figs. 1-3 show that $<k>$, $G(R|S)$ and $H(R|S)$ are decreasing functions of β for different values of m. The three functions grow with m for a given value of β.

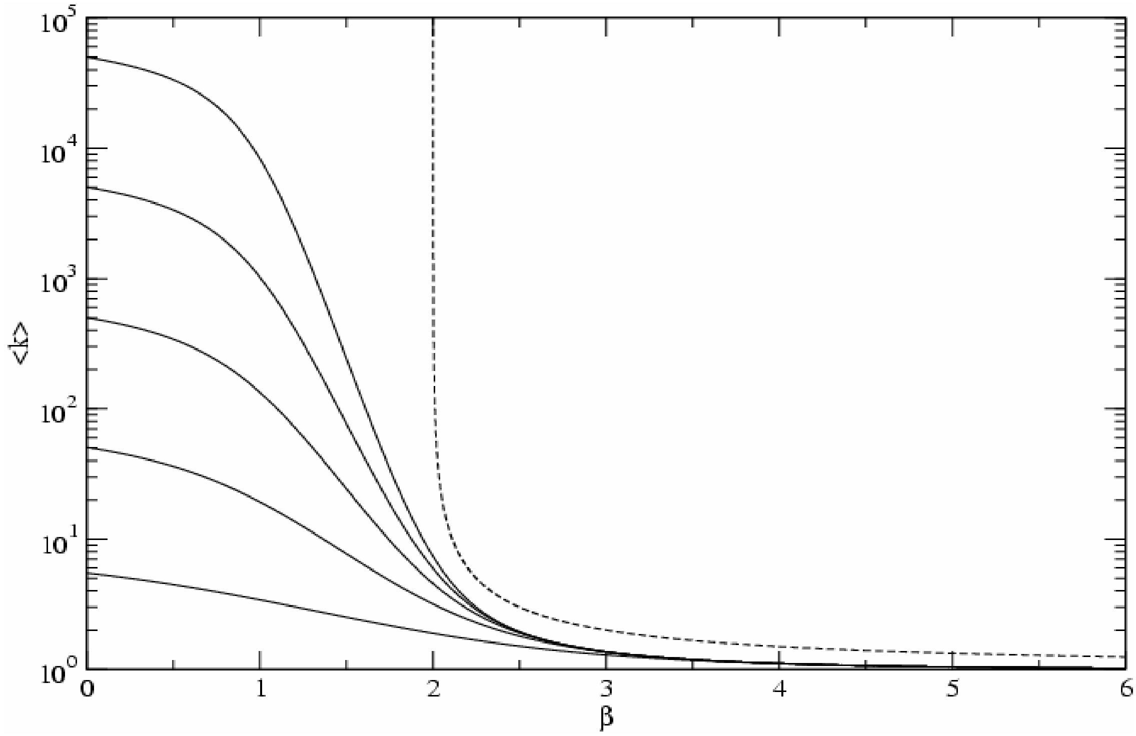

Fig 1. <*k*>, the mean word degree, versus β, the exponent of the frequency spectrum of Zipf's law. Series from the bottom to the top are for $m = 10$, $m = 10^2$, $m = 10^3$, $m = 10^4$ and $m = 10^5$ (solid lines). The approximated expected curve for $m \to \infty$ is also shown (dashed line).

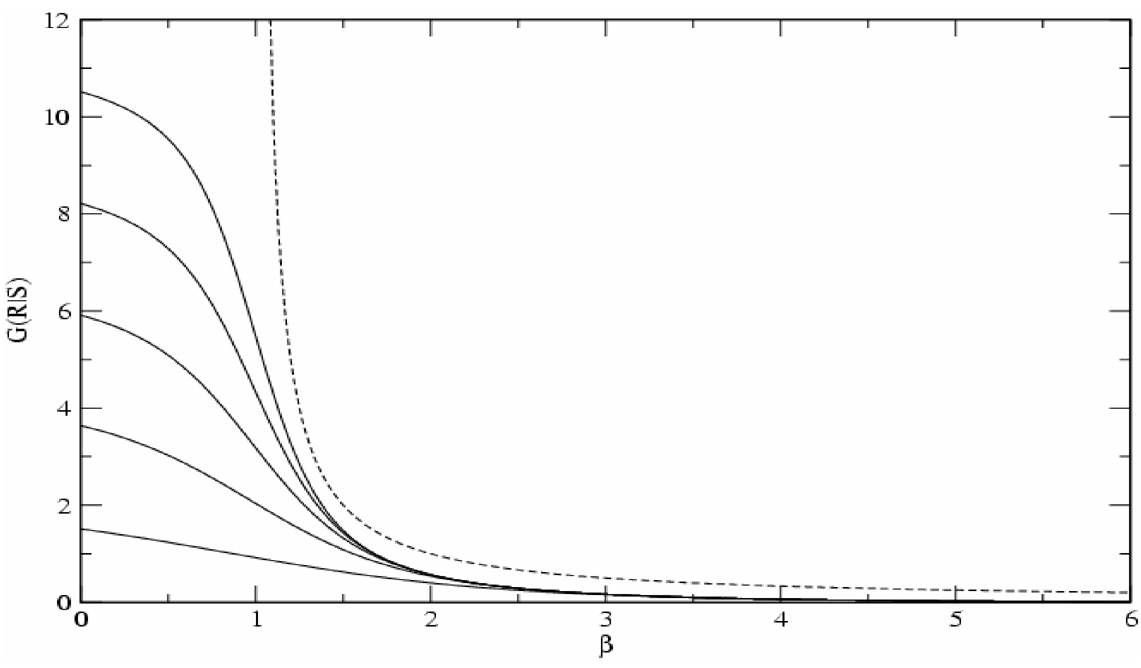

Fig 2. $G(R|S)$, the uncertainty per word associated with the interpretation of every word, versus β, the exponent of the frequency spectrum of Zipf's law. Series from the bottom to the top are for $m = 10$, $m = 10^2$, $m = 10^3$, $m = 10^4$ and $m = 10^5$ (solid lines). The approximated expected curve for $m \to \infty$ is also shown (dashed line). Natural logarithms were used.

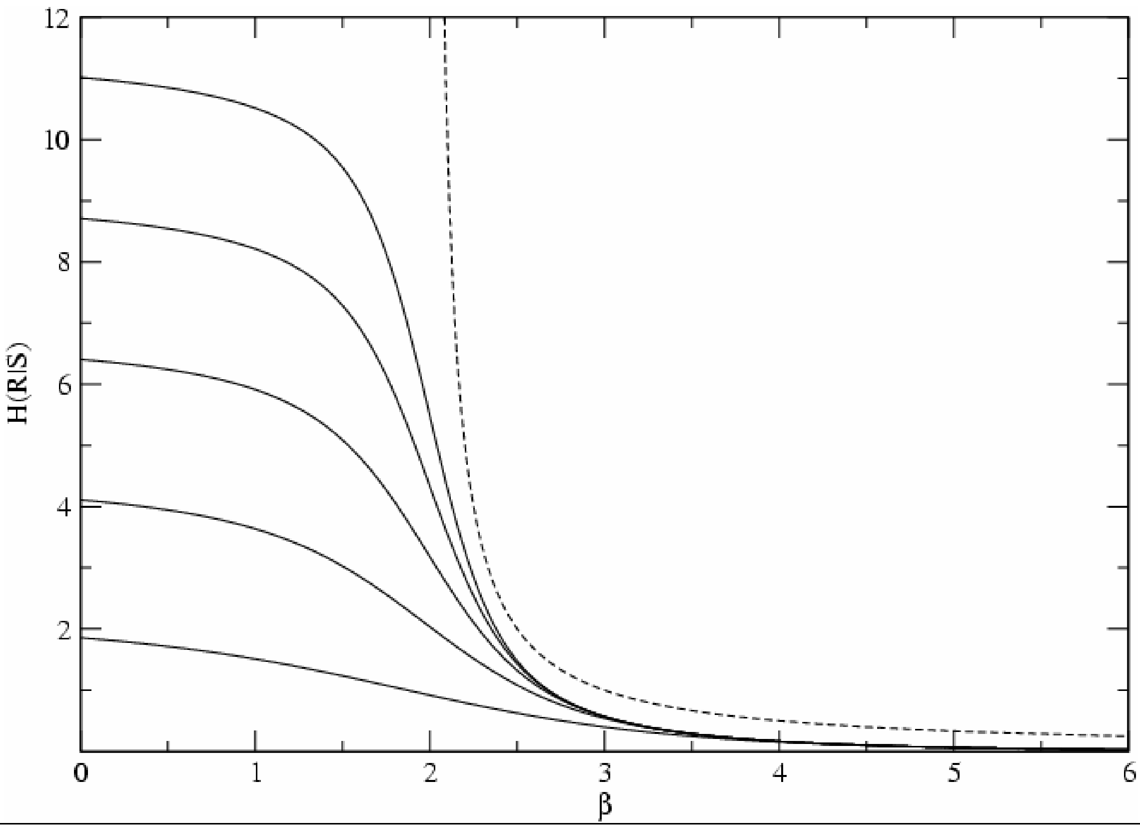

Fig 3. $H(R \mid S)$, the average uncertainty associated with the interpretation of every word, versus β, the exponent of the frequency spectrum of Zipf's law. Series from the bottom to the top are for $m = 10$, $m = 10^2$, $m = 10^3$, $m = 10^4$ and $m = 10^5$ (solid lines). The approximated expected curve for $m \to \infty$ is also shown (dashed line). Natural logarithms were used.

Here we will define vagueness as the opposite of precision. $<k>$, $G(R|S)$ and $H(R|S)$ are inverse measures of precision and direct measures of vagueness.

As for cost of word use, substituting Eq. 11 into Eq. 4 we get,

$$H(S) = \log(n < k >) - \frac{< k \log k >}{< k >}, \tag{14}$$

where $M = n<k>$ is the total amount of links. Knowing Eq. 12, Eq. 14 can be written as

$$H(S) = \log(n < k >) - H(R \mid S). \tag{15}$$

Fig. 4 shows $H(S)$ for $n = 10^3$ and different values of m. $H(S)$ decreases as m grows for a fixed value of β whereas the vagueness measures behave inversely. $H(S)$ has a minimum at $\beta = \beta^*$, a critical value of β, such that $1 < \beta^* < 2$ for the values of m that we used here. Notice that although the exact value of $H(S)$ depends on n, β^* depends only on m (recall Eq. 15). Fig. 5 shows β^* versus m.

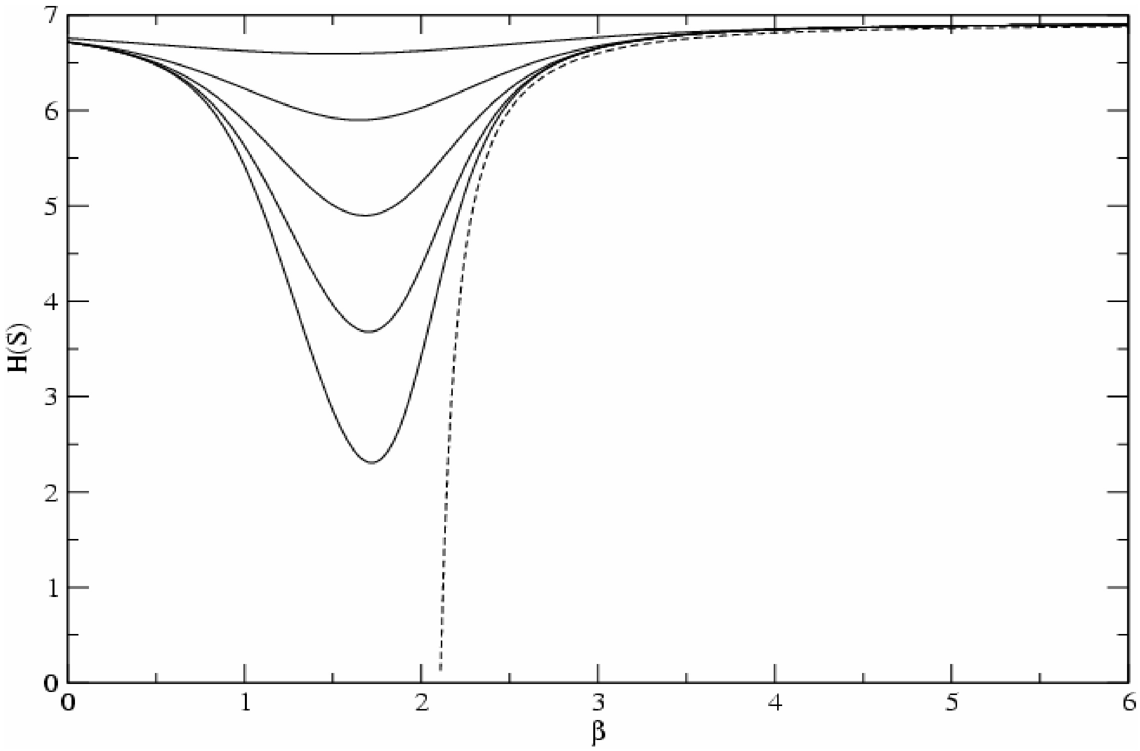

Fig. 4. $H(S)$ versus β where $H(S)$ is the entropy of the set of words S and β is the exponent of the frequency spectrum of Zipf's law. Series from top to the bottom are for $m = 10$, $m = 10^2$, $m = 10^3$, $m = 10^4$ and $m = 10^5$ (solid lines). $n = 10^3$ is used in all cases, although the point where the minimum $H(S)$ is reached is independent of n. The approximated expected curve for $m \to \infty$ is also shown (dashed line). Natural logarithms were used.

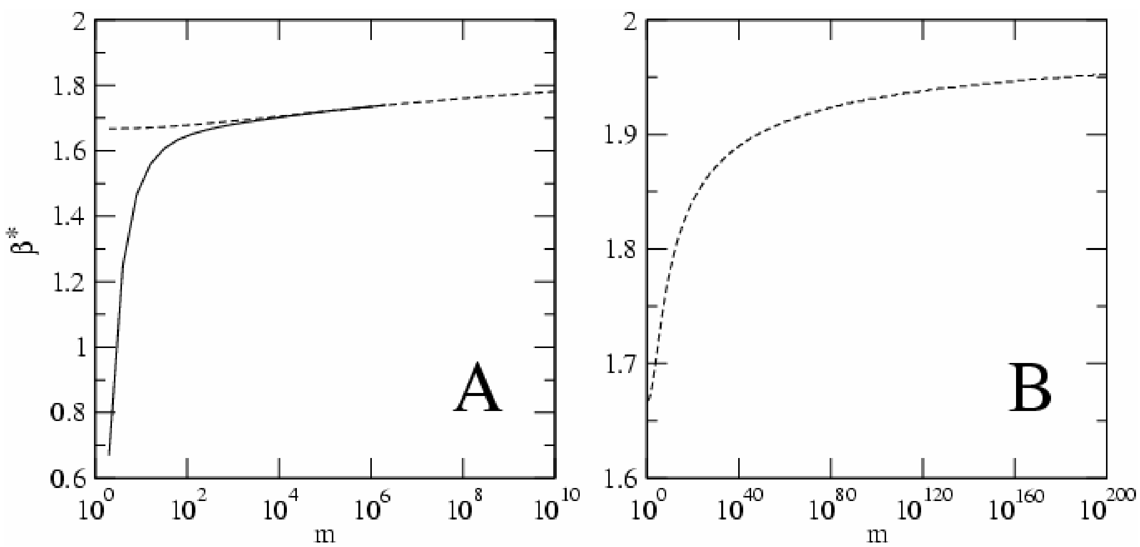

Fig. 5. β^*, the value of β minimizing $H(S)$, versus m. β is the exponent of the frequency spectrum representation of Zipf's law, $H(S)$ is the entropy of the set of words S and m is the number of stimuli. A. β^* versus m calculated without integrals using Eq. 15 (solid line) and with integrals using Table 1 (dashed line) B. β^* versus m calculated with integrals till very large values of m.

RESULTS

We can obtain formulae for the measures of vagueness and cost using approximation with integrals (see Appendix A). Results are summarized in Table 1. The measures of vagueness are functions of β and m whereas $H(S)$ is a function of β, m and n. When $m \to \infty$, we can obtain simple mathematical expressions in particular domains of β (Table 2).

Table 1
Summary of the relationship between Zipf's law and communication measures

Function	Information theory	Approximation	
$<k>$	-	$\beta \neq 1$ and $\beta \neq 2$	$\frac{(1-\beta)(m^{2-\beta}-1)}{(2-\beta)(m^{1-\beta}-1)}$
		$\beta = 1$	$\frac{m-1}{\log m}$.
		$\beta = 2$	$\frac{\log m}{1-\frac{1}{m}}$
$<\log k>$	$G(R\|S)$	$\beta \neq 1$	$\frac{1}{m^{1-\beta}-1}\left[m^{1-\beta}\left(\log m - \frac{1}{1-\beta}\right)+\frac{1}{1-\beta}\right]$
		$\beta = 1$	$\frac{\log m}{2}$
$\frac{<k \log k>}{<k>}$	$H(R\|S)$	$\beta \neq 1$ and $\beta \neq 2$	$\frac{1}{m^{2-\beta}-1}\left[m^{2-\beta}\left(\log m - \frac{1}{2-\beta}\right)+\frac{1}{2-\beta}\right]$
		$\beta = 1$	$\frac{m(\log m - 1)+1}{m-1}$
		$\beta = 2$	$\frac{\log m}{2}$

$<k>$ is the mean word degree, $G(R|S)$ is the uncertainty per word associated with the interpretation of every word and $H(R|S)$ is the average uncertainty associated with the interpretation of every word. $<\ldots>$ is the expectation operator over k. m is the number of stimuli. β is the exponent of the power spectrum of Zipf's law.

$<k>$ when $\beta > 2$ and $m \to \infty$ is shown as dashed line in Fig. 1. $G(R|S)$ when $\beta > 1$ and $m \to \infty$ is shown as dashed line in Fig. 2. $H(R|S)$ when $\beta > 2$ and $m \to \infty$ is shown as a dashed line in Fig. 3. $H(S)$ when $\beta > 2$ and $m \to \infty$ is shown as a dashed line in Fig. 4.

When $\beta > 1$ and $m \to \infty$, we have $G(R|S) = 1/(\beta - 1)$. Zipf's law can be alternatively defined as $P(i) \sim i^{-\alpha}$, where $P(i)$ is the frequency of the i-th most frequent word in a piece of text and $\alpha = 1/(\beta - 1)$ (Chitashvili & Baayen, 1993). Notice that $G(R|S) = 1/(\beta - 1)$ tells us that $\alpha = G(R|S)$. The value of β where $\beta = \alpha$ (and thus $\beta = G(R|S)$) can be calculated solving $\beta = 1/(\beta - 1)$, which has two solutions

$$\beta_1 = \frac{1+\sqrt{5}}{2} \approx 1.61. \qquad (16)$$

and

$$\beta_2 = \frac{1-\sqrt{5}}{2} \approx -0.61. \qquad (17)$$

The first solution (Eq. 16) is the only valid one here since $\beta > 1$ is assumed in $\alpha = 1/(\beta - 1)$ (see also Ferrer i Cancho & Solé, 2001). There are two points of interest with regard to $\beta = \alpha \approx 1.61$. Firstly, $\approx$ 1.61 is close to the exponent found in certain children and schizophrenics. Only in those cases, β is also a direct measure for $G(R|S)$. Secondly, T. Hernández noticed that $(1+\sqrt{5})/2$ (Eq. 16) is the golden ratio, the value to which the fraction of two consecutive numbers of the Fibonacci series converges (Dunlap, 1997). The golden ratio has been the topic of many speculations about its role in nature and our sense of aesthetics (Ghyka, 1927). Future work should be devoted to investigating the origins of the appearance that striking coincidence.

Table 2
Summary of the relationship between the exponents of Zipf's law and various communication measures when $m \to \infty$.

Communication measures	$m \to \infty$			
	β		α	
	Approximation	*Condition*	*Approximation*	*Condition*
$<k>$	$\frac{\beta-1}{\beta-2}$	$\beta > 2$	$\frac{1}{1-\alpha}$	$\alpha < 1$
$G(R\|S)$	$\frac{1}{\beta-1}$	$\beta > 1$	α	$\alpha > 0$
$H(R\|S)$	$\frac{1}{\beta-2}$	$\beta > 2$	$\frac{\alpha}{1-\alpha}$	$\alpha < 1$
$H(S)$	$\log\frac{n(1-\beta)}{2-\beta} - \frac{1}{\beta-2}$	$\beta > 2$	$\log\frac{n}{1-\alpha} - \frac{\alpha}{1-\alpha}$	$\alpha < 1$

$<k>$ is the mean word degree, $G(R|S)$ is the uncertainty per word associated with the interpretation of every word and $H(R|S)$ is the average uncertainty associated with the interpretation of every word. β and α are, respectively, the exponents of the power spectrum and the frequency versus rank representation of Zipf's law. Recall $\alpha = 1/(\beta - 1)$ (Chitashvili & Baayen, 1993).

We can combine Zipf's law (Eq. 1) and the results in Table 2 in order to emphasize the relationship between communication and Zipf's law. When $\beta > 2$,

$$P(f) \sim f^{-\frac{1-2<k>}{1-<k>}} \qquad (18)$$

and

$$P(f) \sim f^{-\frac{1}{H(R|S)}-2} \tag{19}$$

When $\beta > 1$,

$$P(f) \sim f^{-\frac{1}{G(R|S)}-1} \tag{20}$$

A rewritten version of Zipf's law in terms of quantitative communication measures is summarized in Table 3 for both the frequency spectrum and the frequency versus rank representation. We have seen that $G(R|S)$ is the exponent of Zipf's law in the frequency versus rank representation. β also tells us about $G(R|S)$ but $\beta = G(R|S)$ holds only when $\beta \approx 1.61$ (recall Eq. 16).

Table 3
Explicit relationship between Zipf's law and various communication aspects

Representation	Variable		
	$<k>$	$G(R\|S)$	$H(R\|S)$
$P(f)$	$\sim f^{-\frac{1-2<k>}{1-<k>}}$	$\sim f^{-\frac{1}{G(R\|S)}-1}$	$\sim f^{-\frac{1}{H(R\|S)}-2}$
$P(i)$	$\sim i^{-1+\frac{1}{<k>}}$	$\sim i^{-G(R\|S)}$	$\sim i^{-\frac{H(R\|S)}{1+H(R\|S)}}$

$P(i)$ is the frequency of the i-th most frequent word in a sample (e.g. a text). $P(f)$ is the proportion of words in a sample with frequency f. $<k>$ is the mean word degree, $G(R|S)$ is the uncertainty per word associated to the interpretation of every word and $H(R|S)$ is the average uncertainty associated to the interpretation of every word. β and α are, respectively, the exponents of the power spectrum and the frequency versus rank representation of Zipf's law.

DISCUSSION

We have seen that β is an indicator of the degree of semantic precision of a communication system. Given the same values of m, the higher the value of β, the higher the precision. $<k>$, $G(R|S)$ and $H(R|S)$ are decreasing functions of β (Figs. 1-3). In contrast, $H(S)$ has a global minimum between 1.5 and 2 for sufficiently large m (Fig. 4). $G(R|S)$ is a measure of vagueness that does not diverge for $1<\beta < 2$. That is not surprising since $G(R|S)$ does not weight $H(R|s_i)$ by the probability of s_i. From the information theory point of view, $H(R|S)$ is the reference measure of semantic vagueness. We will leave $<k>$ and $G(R|S)$ as alternative simpler measures which correlate with $H(R|S)$ for certain values of β and m. For instance, notice that $G(R|S)$ has no counterpart when $1 < \beta < 2$ and $m \to \infty$ (since $<k>$ and $H(R|S)$ are not defined in that case).

Both $G(R|S)$ and $H(R|S)$ measure the semantic precision from an information theory approach. $G(R|S)$ and $H(R|S)$ can be defined as functions of a single parameter, β, for m large (recall Table 2). $\beta > 2$ is required for $H(R|S)$ while only $\beta > 1$ is required for $H(R|S)$. Thus, $H(R|S)$ cannot deal with the exponent of some children and schizophrenics having $\beta < 2$ if m is actually large. Briefly, $G(R|S)$ covers all the range of variations of β found in human language while $H(R|S)$ does not (provided m is large, of course).

There are strong constraints on the communication systems following Zipf's law (with our assumptions) that may exist when $m \to \infty$. First, $P(k)$ is a probability function only when $\beta > 1$, so systems with $\beta \leq 1$ are impossible at the thermodynamic limit. Second, it is easy to show that $H(R|S)$ and $<k>$ diverge when $m \to \infty$ and $1 < \beta < 2$. In other words, systems with finite vagueness are impossible if $1 < \beta < 2$. From the information theory point of view, if $H(R|S)$ is infinite, then communication is not possible due the infinite uncertainty associated with decoding a single word on average. Besides, we have seen that $H(R|S)$ takes finite values, regardless of how large m is, when $\beta > 2$. In a communication system with $\beta < 2$, m must be finite and not too large, otherwise vagueness is infinite, which contradicts the notion that our system is communicating. We can apply this to real problems. There are many cases where speakers are clearly communicating with $\beta < 2$: military combat texts with $\beta = 1.7$ and children with $\beta = 1.6$ (Piotrowski *et al.*, 1995). Some caution must be taken with schizophrenics with $\beta < 2$, where the assumption of communication may fail. There are reasons for thinking that the assumption is actually satisfied. If we assume that schizophrenics with $\beta < 2$ are communicating, then it follows that m should be small (it should be actually the smallest, since those schizophrenics take the smallest β among real speakers and the vagueness measures grow fast as β decreases). A dramatically reduced set of stimuli that can be perceived and thus can be conveyed using words (i.e. a dramatically low value of m) might explain the obsessive pattern found in that kind of schizophrenics. The same could be happening to children, whose perception of the world is under construction. We may synthesize the essence of the previous argument in a rule.

RULE 1. Suppose a communication system with exponent $\beta < 2$. Then m must be finite and not too large (otherwise the ambiguity would be too large). The chance of a large value of m decreases with β.

The predicted decrease in m in schizophrenics with $\beta < 2$ suggest that the apparent normality of category content and structure (Elvevåg *et al.*, 2005; Elvevåg *et al.*, 2002) may need to be revised.

The various results presented in this article allow us to face the following problem. What can be said about the communicative accuracy or the cost of a communication system when only the signal (e.g. word) frequency distribution is known? Candidates for this kind of analysis are atypical human speakers and the utterances of non-human species (McCowan *et al.*, 1999; McCowan *et al.*, 2002). If Zipf's law is found, the slope in log-log scale of the frequency spectrum is key to finding the answer. We will propose a series of lemmata that are helpful in determining which of two systems is more precise or economical (see Appendix B for outlines of proofs):

LEMMA 1. Suppose we have two communication systems A and B, with exponents β_A (or α_A) and β_B (or α_B) with $\beta_A, \beta_B > 1$, and the number of stimuli is m_A and m_B, respectively. If $\beta_A < \beta_B$ (or $\alpha_A > \alpha_B$) and $m_A \geq m_B$ then B is a strictly more precise communication system than A.

LEMMA 2. If we have two communication systems A and B, with exponents β_A (or α_A) and β_B (or α_B) with $\beta_A, \beta_B > 1$, their number of stimuli is m_A and m_B, and their lexicon size is n_A and n_B, respectively. We assume m_A, m_B, n_A and n_B are finite. If $\beta^* \leq \beta_A < \beta_B$ (or $\alpha^* > \alpha_A > \alpha_B$ with $\alpha^* = 1/(\beta^*-1)$) and $m_B \geq m_A$ and $n_A \geq n_B$ then it follows that A is a more economical communication system than B.

An apparently serious drawback to applying Lemma 2 is that we do not know, in general, if $\beta^* \leq \beta_A$. Interestingly, there are reasons for thinking that communication systems with $\beta^* \leq \beta_A$ are unlikely. The cost of word use and the vagueness increase at the same time as β decreases when $\beta < \beta^*$. It is hard to imagine how a communication system would tolerate decreasing the quality of communication and simultaneously expending more energy to communicate. Vagueness and cost of word use are in conflict for β above β^* (the former decreases with β and the latter grows with β), so it is reasonable to suppose that particular communication systems choose to favour one over the other. But it seems unlikely that a communication system would evolve against *both* factors below β^*. Since $\beta^* \leq \beta_A$ is apparently unlikely, we may propose a modified version of Lemma 2 that is likely to be true in natural communication systems:

> LEMMA 3. If we have two communication systems A and B, with exponents β_A (or α_A) and β_B (or α_B) with $\beta_A, \beta_B > 1$, their number of stimuli is m_A and m_B, and their lexicon size is n_A and n_B, respectively. We assume m_A, m_B, n_A and n_B are finite. If $\beta_A < \beta_B$ (or $\alpha_A > \alpha_B$) and $m_B \geq m_A$ and $n_A \geq n_B$ then it is likely that A is a more economical communication system than B.

We have seen in Eq. 15 that the cost of word use is a function of vagueness when the latter is measured using $H(R|S)$. When $\beta < \beta^*$, cost (of signal use) and vagueness decrease with β. In contrast, cost grows with β while vagueness decreases with β when $\beta > \beta^*$. Vagueness and cost are in conflict when $\beta > \beta^*$, which, as argued above, is likely to hold in natural communication systems.

Lemma 3 can be safely used if β_A is sufficiently large. Fig. 5 suggests that it is rather unlikely that a value of β very near to 2 minimizes the cost of word use. When m is about 10^{80}, a rough estimate of the number of atoms in the universe (Gribbin, 1986), we obtain $\beta^* \approx 1.923$ (recall here and later that our calculations are based on approximations using integrals). When m is about the number of neurons in the brain, about 10^{11} (Damasio, 1999), we get $\beta^* \approx 1.789$. If the number of neurons in the brain is taken, then β cannot minimize $H(S)$ if $\beta > 1.789$. In practice, if $\beta > 1.923$, that means that $\beta > \beta^*$. In a less compelling fashion, if $\beta > 1.789$, that means that $\beta > \beta^*$ is likely to be true. So, we do not need to worry about β^* if β is sufficiently large. In sum, imagine a communication system with exponent β. If $\beta > 1.923$ then $\beta > \beta^*$ is very likely and if $\beta > 1.789$ then $\beta > \beta^*$ is likely.

Now, let us try to apply the lemmata above to real problems. First, we may ask whether the correlation between β and semantic precision is consistent with the variation of β found in the real cases summarized in the introduction section, with regard to an ideal normal language with $\beta_A = 2$. Maybe nouns are the only unquestionable case of communication that is *a priori* more precise than normal language. There seems to be a certain consensus in philosophy and linguistics about the semantic rigidity of many nouns (Kripke, 1980; Mcbeth, 1995; Devitt & Sterelny, 1999). Let us define A as ideal normal language, and B as nouns. We have $\beta_A = 2$ and $\beta_B \in [2.15,2,32]$ (Balasubrahmanyan & Naranan, 1996). Thus, $\beta_A < \beta_B$ for the largest values of β_B. We will focus on that case. Since nouns are associated with a (probably strict) subset of all possible stimuli, that is, $m_A \geq m_B$, it follows from Lemma 1 that nouns are more precise that the entire set of words on average.

With the theory presented here and the support of the previous test, we can move to increasingly complicated cases. Imagine we take schizophrenics with $\beta < 2$ as A and ideal normal language as B. Lemma 1 can not be applied because we may have $m_A < m_B$ as discussed above. A similar problem is poised by children. Imagine we take ideal normal language as A and schizophrenics with $\beta < 2$ as B. We do not know if $m_A \geq m_B$ holds so we cannot safely use Lemma 1. We must be conservative because the decrease in m that is predicted for schizophrenics with $\beta < 2$ could also happen in schizophrenics with $\beta > 2$.

Second, we may try to shed light on the cost of word use in real cases against ideal normal speakers. Let us define A as an ideal normal language and B as nouns. Since nouns are a strict subset of all words, we have $n_A > n_B$. We have seen above that $\beta_A < \beta_B$ and $m_A \geq m_B$. The latter means we cannot use Lemma 2. As for schizophrenic patients, we assume that lexicon size, n, is the same as in normal speakers, as it may be inferred that the lexicon is intact in schizophrenia (Goldberg *et al.*, 2000; Elvevåg *et al.*; 2001; Allen *et al.* 1993). Let us take A as schizophrenics with $\beta < 2$ and B as ideal normal language. We have $\beta_A < \beta_B$ and $m_B \geq m_A$, as we have deduced above. From Lemma 3, we discover that schizophrenics with $\beta < 2$ are more economical speakers. Let us take A as ideal normal language and B as schizophrenics with $\beta > 2$. We do not know if $m_B \geq m_A$. Again, recall that schizophrenics with $\beta > 2$ may have an anomalously low m_B, as schizophrenics with $\beta < 2$. If we take A as children with $\beta = 1.6$ and B as ideal normal speech, Lemma 3 cannot be applied because $n_A \geq n_B$ is not warranted. n_A, vocabulary size, could be one or more orders of magnitude smaller than that of normal adults (Johnson *et al.*, 1999). Since we know that older children eventually converge to $\beta = 2$ (Zipf, 1942), children with $\beta = 1.6$ must be sufficiently young. Knowing that vocabulary grows with age (Johnson *et al.*, 1999), children with $\beta = 1.6$ should have a significantly smaller vocabulary than adults. Although we do not know the exact value of their set of stimuli, and the size of that set depends on the level of their brain development, Lemma 3 cannot be safely used. Nonetheless, the expected significantly small vocabularies may reduce the value of $H(S)$ below that of normal adults.

Let us summarize all the inferences we have made till now:

- Nouns are more precise than mean words in ideal normal adults.
- Schizophrenics with $\beta < 2$ are likely to have a reduced value of m and a more economical communication system with regard to normal adult speakers.

We have assumed that $\beta = 2$ is the ideal exponent of normal adults. Are there reasons for thinking that ideal exponent should be very near 2? Imagine a communication system trying to transmit information about the largest set of stimuli possible. The latter would mean $m \to \infty$. In that case, what is the most economical communication system following Zipf's law? $m \to \infty$ imposes that $\beta > 2$ so that communication has finite vagueness. Since we have seen that the cost of communication grows with β when $\beta > 2$, communication should not go far above $\beta = 2$. Therefore, the ideal communication system minimizing the cost but avoiding infinite vagueness should have $\beta = 2+\varepsilon$, where ε is a small positive quantity (e.g. $\varepsilon = 0.01$ so $\beta = 2.01$). The effect of minimizing β when $\beta > 2$ admits a complementary view. We have considered the negative dimension of $H(R|S)$: the higher the value of $H(R|S)$, the higher the vagueness of words. We could make a positive complementary argument from the point of view of semantic versatility: the higher the value of $H(R|S)$, the higher the semantic versatility of words. It is important to avoid $\beta \leq 2$ to elude infinite vagueness, but it is important to remain near $\beta = 2$, since $H(R|S)$ decreases with β. The arguments above may shed light on the expected exponent of Zipf's law in ideal conditions.

In summary, this article has approached the relationship between the exponent of Zipf's law and various properties of communication systems. Given two communication systems, it is possible to infer which of them is the more economical or vague. It is clear that we need additional information, such as m or n, as well the exponent of Zipf's law, in order to know more about the features of a communication system. Interestingly, we have seen that the amount of extra information that is needed is reduced, and available in some cases. These findings indicates that the little information provided by real communication systems can be squeezed to increase our knowledge about them.

ACKNOWLEDGMENTS

We thank Vito D.P. Servedio and Toni Hernández for helpful comments. We are grateful to Brita Elvevåg for advice about schizophrenia. This work was supported by the ECAgents project, 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

Here, we shall obtain analytical approximations for $<k>$, $G(R|S)$ (Eq. 13), $H(R|S)$ (Eq. 12) and $H(S)$ (Eq. 15), assuming Eq. 5. We will approximate summations using integrals (Cormen *et al.*, 1990). When a summation can be expressed as

$$\sum_{k=k_{\min}}^{k_{\max}} f(k), \quad (21)$$

where $f(k)$ is a monotonically increasing function, we can approximate it by integrals (Cormen, 1990) holding

$$\int_{k_{\min}-1}^{k_{\max}} f(k)dk \leq \sum_{k=k_{\min}}^{k_{\max}} f(k) \leq \int_{k_{\min}}^{k_{\max}+1} f(k)dk . \quad (22)$$

When $f(k)$ is a monotonically decreasing function, then

$$\int_{k_{\min}}^{k_{\max}+1} f(k)dk \leq \sum_{k=k_{\min}}^{k_{\max}} f(k) \leq \int_{k_{\min}+1}^{k_{\max}} f(k)dk . \quad (23)$$

Here, we will use the approximation

$$\sum_{k=k_{\min}}^{k_{\max}} f(k) \approx \int_{k_{\min}}^{k_{\max}} f(k)dk , \quad (24)$$

which is used often in physics (e.g. Cohen & Havlin, 2002; Newman, 2005).

Before providing approximations for $<k>$, $G(R|S)$, $H(R|S)$ and $H(S)$, we need to introduce a function

$$F_x(\gamma, m) = \sum_{k=1}^{m} k^{-\gamma} \log^x k \approx \int_1^m k^{-\gamma} \log^x k dk , \quad (25)$$

which will be used recurrently later on. Interestingly, $F_0(\gamma,m) = H_m^{\gamma}$, where H_m^{γ} is the harmonic number of order γ.

When $\gamma \neq 1$, we have

$$F_0(\gamma,m) \approx \int_1^m k^{-\gamma} dk = \frac{m^{1-\gamma}-1}{1-\gamma} \tag{26}$$

and

$$F_1(\gamma,m) \approx \int_1^m k^{-\gamma} \log k dk = \frac{1}{1-\gamma}\left[m^{1-\gamma}\left(\log m - \frac{1}{1-\gamma}\right) + \frac{1}{1-\gamma}\right]. \tag{27}$$

When $\gamma = 1$, we have

$$F_0(\gamma,m) \approx \log m \tag{28}$$

and

$$F_1(\gamma,m) \approx \frac{\log^2 m}{2}. \tag{29}$$

Table 4
Summary of definitions of different functions and their relationships

Function	Information theory	Definition
$F_x(\gamma,m)$	-	$\sum_{k=1}^{m} k^{-\gamma} \log^x k$
c	-	$1/F_0(\beta,m)$
$< k >$	-	$F_0(\beta\text{-}1,m)/F_0(\beta,m)$
$< \log k >$	$G(R\|S)$	$F_1(\beta,m)/F_0(\beta,m)$
$\frac{< k \log k >}{< k >}$	$H(R\|S)$	$F_1(\beta\text{-}1,m)/F_0(\beta\text{-}1,m)$

Table 4 summarizes the relationship between the auxiliary function $F_x(\gamma,m)$ and the functions of vagueness.

Thus, we may write,

$$< k > = c\sum_{k=1}^{m} k^{1-\beta} = cF_0(\beta-1,m), \tag{30}$$

where c is the normalization constant of Eq. 8, defined as

$$c = \frac{1}{\sum_{k=1}^{m} k^{-\beta}} = \frac{1}{F_0(\beta,m)}. \tag{31}$$

Eqs. 30 and 31 together give

$$< k > = \frac{F_0(\beta - 1, m)}{F_0(\beta, m)}. \quad (32)$$

If $\beta \neq 1$, substituting Eq. 26 on Eq. 31 gives (Cohen & Havlin, 2002)

$$c \approx \frac{1-\beta}{m^{1-\beta} - 1}. \quad (33)$$

If $\beta = 1$, substituting Eq. 28 on Eq. 31 gives

$$c \approx \frac{1}{\log m}. \quad (34)$$

When $\beta \neq 1$ and $\beta \neq 2$, substituting $F_0(\beta - 1,m)$ with Eq. 26 with $\gamma = \beta - 1$ and $F_0(\beta,m)$ by Eq. 26 with $\gamma = \beta$ into Eq. 32 we obtain

$$< k > \approx \frac{(1-\beta)(m^{2-\beta} - 1)}{(2-\beta)(m^{1-\beta} - 1)}. \quad (35)$$

When $\beta = 1$, substituting $F_0(\beta - 1,m)$ with Eq. 26 (with $\gamma = \beta - 1 = 0$) and $F_0(\beta,m)$ by Eq. 28 (since $\gamma = \beta = 1$) into Eq. 32 we obtain

$$< k > \approx \frac{m-1}{\log m}. \quad (36)$$

When $\beta = 2$, substituting $F_0(\beta - 1,m)$ with Eq. 28 (since $\gamma = \beta - 1 = 1$) and $F_0(\beta,m)$ by Eq. 26 (with $\gamma = \beta = 2$) into Eq. 32 we obtain

$$< k > \approx \frac{\log m}{1 - \frac{1}{m}}. \quad (37)$$

When $\beta > 2$ and $m \to \infty$, Eq. 35 becomes

$$< k > \approx \frac{\beta - 1}{\beta - 2}. \quad (38)$$

The previous equation is shown as a dashed line in Fig. 1.

$G(R|S)$ can be written as

$$G(R \mid S) = < \log k > = c \sum_1^m k^{-\beta} \log k = \frac{F_1(\beta, m)}{F_0(\beta, m)}. \quad (39)$$

When $\beta \neq 1$, substituting $F_1(\beta,m)$ by Eq. 27 and $F_0(\beta,m)$ by Eq. 26 (both with $\gamma = \beta$) into Eq. 39 we get

$$G(R\,|\,S) \approx \frac{1}{m^{1-\beta}-1}\left[m^{1-\beta}\left(\log m - \frac{1}{1-\beta}\right) + \frac{1}{1-\beta}\right]. \tag{40}$$

When $\beta = 1$, substituting $F_1(\beta,m)$ by Eq. 29 and $F_0(\beta,m)$ by Eq. 28 (since $\gamma = \beta = 1$ in both cases) into Eq. 39 we get

$$G(R\,|\,S) \approx \frac{\log m}{2}. \tag{41}$$

When $\beta > 1$ and $m \to \infty$, Eq. 40 becomes

$$G(R\,|\,S) \approx \frac{1}{\beta - 1}. \tag{42}$$

(as in Ferrer i Cancho, 2005a). The previous equation is shown as a dashed line in Fig. 2.
As for $H(R|S)$, the numerator in Eq. 12 can be expressed as

$$< k \log k > = c\sum_{1}^{m} k^{1-\beta}\log k = cF_1(\beta - 1, m). \tag{43}$$

Substituting Eqs. 30 and 43 into Eq. 12 we obtain

$$H(R\,|\,S) = \frac{F_1(\beta-1,m)}{F_0(\beta-1,m)}. \tag{44}$$

If $\beta \neq 1$ and $\beta \neq 2$, substituting $F_1(\beta - 1,m)$ with Eq. 27 and $F_0(\beta - 1,m)$ with Eq. 26 (with $\gamma = \beta - 1$ in both cases) into Eq. 44 we get

$$H(R\,|\,S) \approx \frac{1}{m^{2-\beta}-1}\left[m^{2-\beta}\left(\log m - \frac{1}{2-\beta}\right) + \frac{1}{2-\beta}\right]. \tag{45}$$

If $\beta = 1$, substituting $F_1(\beta\text{-}1,m)$ with Eq. 27 and $F_0(\beta\text{-}1,m)$ with Eq. 26 (both with $\gamma = \beta - 1 = 0$) into Eq. 44 we get

$$H(R\,|\,S) \approx \frac{m(\log m - 1) + 1}{m - 1}. \tag{46}$$

If $\beta = 2$, substituting $F_1(\beta - 1,m)$ with Eq. 29 and $F_0(\beta - 1,m)$ with Eq. 28 (since $\gamma = \beta - 1 = 1$ in both cases) into Eq. 44 we get

$$H(R\,|\,S) \approx \frac{\log m}{2}. \tag{47}$$

Eq. 45 with β > 2 and $m \to \infty$ becomes

$$H(R \mid S) \approx \frac{1}{\beta - 2}. \tag{48}$$

The previous equation is shown as a dashed line in Fig. 3.

When $\beta > 2$ and $m \to \infty$, substituting Eqs. 38 and 48 into Eq. 15 we obtain

$$H(S) = \log\left(\frac{n(\beta - 1)}{\beta - 2}\right) - \frac{1}{\beta - 2}. \tag{49}$$

The previous equation is shown as a dashed line in Fig. 4. Since Eq. 49 is an approximation, and $<k>$ and $H(R|S)$ diverge for $\beta = 2$, it is convenient to keep $\beta >> 2$.

APPENDIX B

Here we give an outline of proof for Lemma 1 and 2.

LEMMA 1. Suppose we have two communication systems *A* and *B*, with exponents β_A (or α_A) and β_B (or α_B) with $\beta_A, \beta_B > 1$, and the number of stimuli is m_A and m_B, respectively. If $\beta_A < \beta_B$ (or $\alpha_A > \alpha_B$) and $m_A \geq m_B$ then *B* is a strictly more precise communication system than *A*.

Proof: The proof is based on $H(R|S)$, the reference measure for word vagueness. Assuming $\beta_A, \beta_B > 1$ we warrant that $P(k)$ is a probability distribution even when $m \to \infty$. In general, there are only four situations:

1) m_A and m_B are finite. It is easy to see from the approximate equations in Table 1 Appendix A (recall also Fig. 3) that $H(R|S)$ is a monotonically decreasing function of β (when $\beta > 0$) when m_A and m_B are finite. Given a particular β, the larger the value of m, the larger the value of the measure.

2) m_A is finite and m_B is not. That contradicts $m_A \geq m_B$.

3) m_A is infinite and m_B is not. That contradicts the notion that A is a communication system if $\beta_A \leq 2$. $\beta_A > 2$ must be satisfied and thus we can proceed as in 1).

4) m_A and m_B are infinite. That contradicts the notion that A and B are communication systems if $\beta_A \leq 2$ and/or $\beta_B \leq 2$. $\beta_A, \beta_B > 2$ must be satisfied and thus we can proceed as in 1).

LEMMA 2. If we have two communication systems *A* and *B*, with exponents β_A (or α_A) and β_B (or α_B) with $\beta_A, \beta_B > 0$, their number of stimuli is m_A and m_B, and their lexicon size is n_A and n_B, respectively. We assume m_A, m_B, n_A and n_B are finite. If $\beta^* \leq \beta_A < \beta_B$ (or $\alpha^* > \alpha_A > \alpha_B$ with $\alpha^* = 1/(\beta^* - 1)$) and $m_B \geq m_A$ and $n_A \geq n_B$ then it follows that *A* is a more economical communication system than *B*.

Proof: $\alpha^* = 1/(\beta^* - 1)$ comes from the equivalence between, α, the exponent of the frequency versus rank representation and β, the exponent of the frequency spectrum (Chitashvili & Baayen, 1993). It is easy to show from the approximate equations in Table 1 (recall Figs. 4-5) that $H(S)$, the measure of cost, is a monotonically increasing function of β when $\beta > \beta^*$ and that given a particular β, the larger the value of m, the lower the cost, and that, the larger the value of n, the larger the cost.

REFERENCES

Allen, H. A., Liddle, P.F. & Frith, C. D. (1993). Negative features, retrieval processes and verbal fluency in schizophrenia. *British Journal of Psychiatry 163, 769-775*.

Akmajian, A., Harnish, R. M., Demers, R. A. & Farmer, A. K. (1995). *Linguistics. An introduction to language and communication*. Cambridge, MA: MIT Press.

Ash, R. B. (1990). *Information Theory.* New York: Dover Publications, Inc.

Balasubrahmanyan, V. K. & Naranan, N. (1996). Quantitative linguistics and complex systems studies. *Journal of Quantitative Linguistics 3, 177-228.*

Bollobás, B. (1998). Modern graph theory. New York: Springer.

Brillouin, L. (1960). *Science and theory of information* (Russian translation). Moscow: Gosudarstvennoe Izdatel´stvo Fiz.-Mat. Literatury.

Carroll, D. W. (1994). Psychology of language. Chapter 9, Conversational interaction. pp. 242-248. Pacific Grove, California: Brooks/Cole Publishing Company.

Chitashvili, R. J. & Baayen, R. H. (1993). Word frequency distributions. In: L. Hřebíček, G. Altmann (eds.), *Quantitative text analysis: 54-135.* Trier: Wissenschaftlicher Verlag Trier.

Cohen, R. & Havlin, S. (2002). Scale-free networks are ultrasmall. *Physical Review Letters 90, 057801.*

Cormen, T. H., Leisserson, C. E. & Rivest, R. L. (1990). *Introduction to algorithms*. Cambridge: MIT Press.

Damasio, A. (1999). *The Scientific American Book of the Brain.* New York: Scientific American. See http://hypertextbook.com/facts/2002/AniciaNdabahaliye2.shtml.

Devitt, M. & Sterelny, K. (1999). *Language and reality: an introduction to the philosophy of language.* Cambridge, MA: MIT Press.

Dunlap, R. A. (1997). *The golden ratio and Fibonacci series*. Singapore: World Scientific Publishing.

Elvevåg, B., Weinstock, M., Akil, M., Kleinman, J.E. & Goldberg, T. E. (2001) A comparison of verbal fluency tasks in schizophrenic patients and normal controls. *Schizophrenia Research 51, 119-126.*

Elvevåg, B., Fisher, J. E., Gurd, J.M. & Goldberg, T. E. (2002). Semantic clustering in verbal fluency: schizophrenic patients versus control participants. *Psychological Medicine 32, 909-917.*

Elvevåg, B., Storms G., Heit, E. & Goldberg, T. (2005). Category content and structure in schizophrenia: an evaluation using the instantiation principle. *Neuropsychology 19, 371-380.*

Ferrer i Cancho, R. (2005a). Decoding least effort and scaling in signal frequency distributions. *Physica A* 345, 275-284.

Ferrer i Cancho, R. (2005b). The variation of Zipf's law in human language. *European Physical Journal B 44, 249-257.*

Ferrer i Cancho, R. (2005c). The consequences of Zipf's law for syntax and symbolic reference. *Proceedings of the Royal Society of London B 272, 561-565.*

Ferrer i Cancho, R. (2005d). Core and peripheral lexicon from word length optimization. Submitted to the *Journal of Quantitative Linguistics.*

Ferrer i Cancho, R. (2005e). Zipf's law from a communicative phase transition. Submitted to *European Physical Journal B*.

Ferrer i Cancho, R. and Reina, F. (2002). Quantifying the semantic contribution of particles. *Journal of Quantitative Linguistics, 9, 35-47.*

Ferrer i Cancho, R. & Solé, R. V. (2001). Two regimes in the frequency of words and the origins of complex lexicons: Zipf's law revisited. *J. Quantitative Linguistics, 8, 165-173*. First appeared as *Santa Fe Institute Working Paper 00-12-068*.

Ferrer i Cancho & Solé, R. V. (2003). Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Science USA 100, 788-791.*

Ghyka, C. M. (1927). *Esthétique des proportions dans la nature et dans les arts*. Paris: Gallimard.

Gribbin, J. (1986). *In search of the big bang: Quantum physics and cosmology*. New York: Bantam Books. See http://www.sunspot.noao.edu/sunspot/pr/answerbook/universe.html.

Goldberg, T. E., Dodge, M., Aloia, M., Egan, M. F. & Weinberger, D. R. (2000). Effects of neuroleptic medications on speech disorganization in schizophrenia: biasing associative networks towards meaning. *Psychological Medicine 30, 1123-1130.*

Johnson, C., Davis. H. & Macken, M. (1999). Symbols and structure in language-acquisition. In: Lock, A. & Peters, C.R. (eds.), *Handbook of Human Symbolic Evolution: 686-746*. Oxford: Blackwell.

Kripke, S. A. (1980). Referring to artifacts. *Phylosophical Review 89, 109-114.*

Köhler, R. (1986). *Zur linguistischen Synergetik: Struktur und Dynamik der Lex*ik. Bochum: Brockmeyer.

Köhler, R. (1987). System theoretical linguistics. *Theoretical Linguistics 14, 241-257.*

Macbeth, D. (1995). Names, natural kind terms and rigid designation. *Philosophical Studies 79, 259-281.*

McCowan, B., Hanser, S. F. & Doyle, L. R. (1999). Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Animal Behavior 57, 409-419.*

McCowan, B., Doyle, L. R. & Hanser, S. F. (2002). Using information theory to asses the diversity, complexity and development of communicative repertoires. *Journal of Comparative Psychology 116, 166-172.*

Montemurro, M. (2001). Beyond the Zipf-Mandelbrot law in quantitative linguistics. *Physica A 300, 567-578.*

Montemurro, M. & Zanette, D. (2002). Frequency-rank distribution in large text samples: phenomenology and models. *Glottometrics 4, 87-98.*

Naranan, S. & Balasubrahmanyan, V. K. (1998). Models for power law relations in linguistics and information science. *Journal of Quantitative Linguistics 5, 35-61.*

Newman, M. E. J. (2005). Power laws, Pareto distributions and Zipf's law. *Contemporary Physics*, in press. *cont-mat/0412004.*

Piotrowska, X., Pashkovska, W., & R. Piotrowski, R. (to appear). Pathological text and its statistical parameters.

Piotrowski, R.G., Pashkovskii, V.E., Piotrowski, V.R. (1995). Psychiatric linguistics and automatic text processing. In: *Automatic Documentation and Mathematical Linguistics, 28(5), 28-35.* [First published in Naučno-Techničeskaja Informacija, Serija 2, Vol. 28, No. 11. pp. 21-25, 1994].

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

Tuldava, J. (1996). The frequency spectrum of text and vocabulary. *Journal of Quantitative Linguistics 3, 38-50.*

Whitehorn, J.C. & Zipf, G.K. (1943). Schizophrenic language. *Archive of neurology and psychiatry 49, 831-851.*

Zipf, G. K. (1932). Selected studies of the principle of relative frequency in language. Cambridge, MA: Harvard University Press.

Zipf, G. K. (1935). *The psycho-biology of language: an introduction to dynamic philology*. Boston, MA: Houghton-Mifflin.

Zipf, G. K. (1942). Children's speech. *Science 96, 344-345*.

Zipf, G. K. (1949). *Human behavior and the principle of least effort.* Reading: Addison-Wesley.