Emergence of Zipf's Law in the Evolution of Communication

Bernat   Corominas-Murtra
Jordi Fortuny Andreu
Ricard V. Solé

SANTA FE INSTITUTE

Emergence of Zipf's Law in the Evolution of Communication

Bernat Corominas-Murtra[1], Jordi Fortuny Andreu[2] and Ricard V. Solé[1,3,4]

[1] *ICREA-Complex Systems Lab, Universitat Pompeu Fabra -Parc de Recerca Biomèdica de Barcelona (PRBB). Dr Aiguader 88, 08003 Barcelona, Spain*
[2] *Centre de Lingüística Teòrica (CLT), Facultat de Lletres, Edifici B, Universitat Autònoma de Barcelona, 08193 Bellaterra (Barcelona), Spain*
[3] *Santa Fe Institute, 1399 Hyde Park Road, New Mexico 87501, USA*
[4] *Institut de Biologia Evolutiva. CSIC-UPF. Passeig Marítim de la Barceloneta, 37-49, 08003 Barcelona, Spain.*

Zipf's law seems to be ubiquitous in human languages and appears to be a universal property of complex communicating systems. Following the early proposal made by Zipf concerning the presence of a tension between the efforts of speaker and hearer in a communication system, we introduce evolution by means of a variational approach to the problem based on Kullback's Minimum Discrimination of Information Principle. Therefore, using a formalism fully embedded in the framework of information theory, we demonstrate that Zipf's law is the only expected outcome of an evolving, communicative system under a rigorous definition of the communicative tension described by Zipf.

Keywords: Zipf's law, scaling, evolution of codes, Minimum Discrimination of Information Principle

I. INTRODUCTION

Zipf's law is one of the most common power laws found in nature and society [1–6]. Although it was early observed in the distribution of money income [7] and city sizes, [1], it was popularized by the linguist George Kingsley Zipf, who observed that it accounts for the frequency of words within written texts [2, 3]. Specifically, if we rank all the occurrences of words in a text from the most common to the least, Zipf's law states that the probability $q(s_m)$ that in a random trial we find the m-th most common word ($i = 1, ..., n$) falls off as

$$q(s_m) = \frac{1}{Z} m^{-\gamma},$$

where,

$$Z = \sum_{j \leq n} j^{-\gamma},$$

with $\gamma \approx 1$. The ubiquity of this scaling behavior suggested several mechanisms to account for the emergence of this distribution, among many others, see [4, 8–12].

Within the context of human language, G. K. Zipf early conjectured that this scaling law is the outcome of a tension between two *forces* acting in a communication system [3]. Following Zipf's proposal, speakers and hearers need to simultaneously minimize their efforts, under what he called *Principle of Least Effort*. Under this view, a conflict would be present while trying to simultaneously minimize the efforts of both elements. The speaker's economy would favour a reduction of the size of the vocabulary to a single word whereas the hearer's economy would do just the opposite, increasing the size of a vocabulary to a point where there would be a different word for each meaning. The resulting vocabulary would emerge out of this unification-diversification conflict [3]. Although both numerical and theoretical studies have explored this idea [10, 11, 13] no truly analytic proof
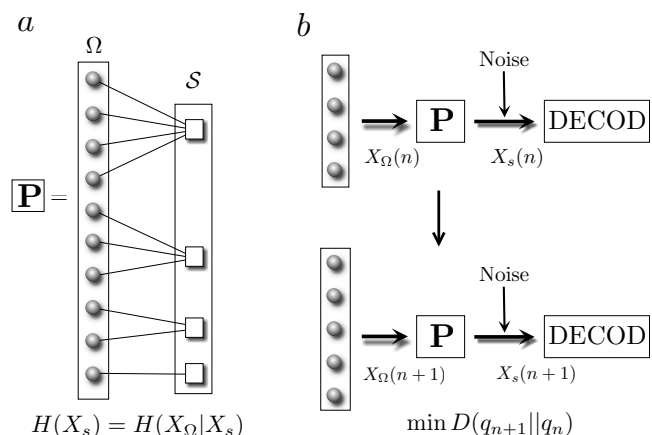

FIG. 1: A growing communication system. In (a) possible meaning-signal associations made by the coder module **P** in which eq. (2) holds is depicted. In (b) we summarize the evolution rules of our communicative system. Suppose that symmetry between coder and decoder -i.e., eq. (2)- holds for the step n (above). At each step (below) a new element is added to the set Ω and eq. (2) holds again for this new configuration. Furthermore, the new configuration is constrained by the $MDIP$, which introduces a path dependency in the evolutionary process.

of unicity has been provided under realistic, information-theoretic constraints. We will refer to the proposals made in [10, 11, 13] as *static* for they consider a fixed size of the code. As we shall see, we need another ingredient, pointed out -in a different context- in [14], namely the active role played by the evolutionary path followed by the code. As it occurs with other systems growing out of equilibrium, such as scale-free networks [15] we will consider the evolution of the communicative exchange under system's growth.

Here the evolutionary component is variationally introduced by minimizing the divergence between code configurations belonging to successive time steps. This minimal

change follows the so-called *Minimum Discrimination Information Principle* (henceforth $MDIP$), a general variational principle considered analogous to the Maximum Entropy Principle [16], from which statistical mechanics can be properly formalized [17, 18]. The $MDIP$ states that, under changes in the constraints of the system, the most expected probability distribution is the one minimizing the Kullback-Leibler divergence (also referred as to *Kullback-Leibler entropy* or *relative entropy*) from the original one [17]. Such a variational principle constrains the changes of the internal configurations of an statistical ensemble when the external conditions change in the same way that internal configurations of an statistical ensemble change when we introduce moment constraints in a Jaynesian formalism. In our context, this information theoretic functional assumes the role of a Lagrangian whose minimization along the process defines the possible ensemble configurations one can observe at certain point of an evolutionary path. Using the $MDIP$ and recent results from two of the authors on the convergence properties of the entropy of growing stochastic objects and its relation to the emergence of Zipf's law [12], we provide a proof of unicity for the emergence of Zipf's law in evolving codes.

The remainder of the paper is structured as follows: In section II we rigorously define the communicative tension intuitively defined by Zipf and explicitly characterize the evolutionary process in terms of the mathematical statement of such a tension. In section III we apply the $MDIP$ as the guiding, variational principle which accounts for the possible evolutionary paths of the code. Finally, we demonstrate that the consequences of the application of both the communicative tension and the $MDIP$ account for the emergence of Zipf's law as the unique possible solution of the evolving code. In section IV we discuss the implications of our results.

II. THE EVOLUTION OF THE COMMUNICATIVE SYSTEM

In this section we mathematically define 1) the communicative tension described by Zipf and 2) the evolution or growing of a given code subject to such a tension. We furthermore define the range of application of our formalism. As we shall see in section III, the proposal made in this section defines a framework whose key piece to work with is eq. (6).

A. The explicit description of the communicative conflict

The first task is to properly define the communicative tension between the coder and the decoder and how this tension is solved. Following the standard nomenclature used in studies of the evolution of communicating, autonomous agents [19–21], in our system there are two agents: the coder agent, $\mathbf{P}$, encoding information from a set of external events, Ω, and the decoder or external observer, which infers the behavior of Ω through the code provided by the coder agent $\mathbf{P}$. In this way,

$$\Omega = \{m_1, ..., m_n\}$$

is the set of external events acting as the input alphabet, and

$$\mathcal{S} = \{s_1, ..., s_n\}$$

is the set of signals or output alphabet. The coder module $\mathbf{P}$ -fig. (1a)- is fully described by a matrix $\mathbb{P}(X_s|X_\Omega)$, where X_Ω is a random variable taking values on the set Ω following the probability measure p; being $p(m_k)$ the probability to have symbol m_k as the input in a given computation. Complementarily, X_s is a random variable taking values on $\mathcal{S}$ and following the probability distribution q which, for a given $s_i \in \mathcal{S}$, reads:

$$q(s_i) = \sum_{k \leq n} p(m_k)\mathbb{P}(s_i|m_k), \qquad (1)$$

i.e., the probability to obtain s_i as the output of a codification. We assume that

$$(\forall m_i \in \Omega) \sum_{i \leq n} \mathbb{P}(s_i|m_k) = 1.$$

For the decoder agent inferring the input set from the output set with least effort, the best scenario is a one-to-one mapping between Ω and $\mathcal{S}$. In this case, $\mathbf{P}$ generates an unambiguous code, and no supplementary amount of information to successfully reconstruct X_Ω is required. However, from the coding device perspective, this coding has a high cost. In order to characterize this conflict, let us properly formalize the above intuitive statement: The decoder agent wants to reconstruct X_Ω through the intermediation of the coding performed by $\mathbf{P}$. Therefore, the amount of *bits* needed by the decoder of X_s to unambiguously reconstruct X_Ω is

$$H(X_\Omega, X_s).$$

From the codification process, the decoder receives $H(X_s)$ bits, and thus, the remaining uncertainty it must face will be

$$H(X_\Omega, X_s) - H(X_s) = H(X_\Omega|X_s).$$

The tension between the coder and the decoder is solved by imposing a symmetric balance between its associated *efforts* -see fig. (1a)-, i.e.: The coder sends as many bits as the additional bits the observer needs to perfectly reconstruct X_Ω:

$$H(X_s) = H(X_\Omega|X_s). \qquad (2)$$

The above *ansatz* is the mathematical formulation of the symmetric balance between the efforts of the coder and

the decoder. We will refer to this equation as the *symmetry condition* and, as pointed out in [11], it mathematically describes how the communicative tension is solved by using a cooperative strategy between the coder and the decoder agents. It is worth to note that different equations sharing the same spirit were formerly proposed, within the framework of the so-called *code-length* game [10]. Using the fact that

$$I(X_\Omega : X_s) = I(X_s : X_\Omega),$$

we reach a general relation between the informative richness of the input variable X_Ω and the informative richness of the messages sent by the coder, constrained by eq. (2):

$$\frac{1}{2}H(X_\Omega) \leq H(X_s) \leq H(X_\Omega). \tag{3}$$

The first relation becomes equality only in the case of **P** performing a deterministic codification process. The second relation becomes equality when the coding device performs completely random associations. It is clear that eqs. (2) and (3) alone cannot explain the emergence of Zipf's law since one could tune the parameters of, say, an exponential distribution to reach the desired relation between entropies. Therefore we need to introduce another ingredient to obtain Zipf's law as the unique possible solution of our problem.

B. Evolution

The unicity in the solution is provided by the evolution, which is now explicitly introduced -see fig (1b). Let us suppose that our communicative success grows over time, thereby increasing the number of input symbols that **P** can encode. Formally, this implies that the cardinality of the set Ω defined above increases. We introduce this feature by defining a sequence of Ω's $\Omega(1), ..., \Omega(k), ...$ satisfying an inclusive ordering, i.e.,

$$\Omega(1) \subset \Omega(2) \subset ... \subset \Omega(k), ...,$$

which is introduced, without any loss of generality, assuming that

$$\begin{aligned}
\Omega(1) &= \{m_1\}, \\
\Omega(2) &= \{m_1, m_2\}, \\
... \quad &\quad ... \\
\Omega(n-1) &= \{m_1, ..., m_{n-1}\}, \\
\Omega(n) &= \{m_1, ..., m_{n-1}, m_n\}.
\end{aligned}$$

At time step n, **P** will be able to process the n symbols of $\Omega(n)$. The elements $m_1, ..., m_i, ...$ are members of some infinite, countable set $\tilde{\Omega}$, i.e., $(\forall i)(\Omega(i) \subset \tilde{\Omega})$. $\tilde{\Omega}$ can be understood, using a thermodynamical metaphor, as a *reservoir of information*. Following the characterization, we say that for every set $\Omega(i)$ there is a random variable $X_\Omega(i)$, taking values in $\Omega(i)$ following the ordered

probability distribution p_i. Furthermore, we assume that $\exists! \mu \in (0, 1)$ such that $(\forall \epsilon > 0)(\exists N) : (\forall n > N)$,

$$\left| \frac{H(X_\Omega(n))}{\log n} - \mu \right| < \epsilon. \tag{4}$$

i.e., that the entropy of the input set is unbounded when its size increases, which implies that the potential input set $\tilde{\Omega}$ acts as an *infinite* reservoir of information. The behavior of output set at the stage n is described by a random variable $X_s(n)$, which follows the ordered probability distribution q_n, as defined in eq. (1), taking values on $\mathcal{S}(n) = \{s_1, ..., s_n\}$. We observe that $\mathcal{S}(n) \subseteq \mathcal{S}(n+1)$, defining a sequence $\mathcal{S}(1), ..., \mathcal{S}(k), ...$ also ordered by inclusion. At every time step, the consequences of the symmetry condition -see eq. (2)- depicted in eq. (3) are satisfied, which implies that the sequence

$$\mathcal{H} = H(X_s(1)), H(X_s(2)), ..., H(X_s(k)), ...$$

also satisfies the convergence ansatz made over the sequence of normalized entropies of the input -see eq. (4). The only difference is the value of the limit, ν. The value of ν can be bounded by using eqs (3) and (4), thereby obtaining:

$$\frac{1}{2}\mu \leq \nu \leq \mu \tag{5}$$

Therefore, in this case, by virtue of eqs. (3), (4) and (5), the convergence condition for the normalized entropies of the sequence of random variables $X_s(1), ..., X_s(n), ...$ reads: $\exists! \nu \in (\frac{1}{2}\mu, \mu)$ such that $(\forall \epsilon > 0)(\exists N) : (\forall n > N)$:

$$\left| \frac{H(X_s(n))}{\log n} - \nu \right| < \epsilon. \tag{6}$$

The above equation depicts two crucial facts in the forthcoming derivations: If the potential informative richness of the input set is unbounded, so is the informative richness of the output set, under the constraints imposed by the symmetry condition -see eq. (2).

III. THE EMERGENCE OF ZIPF'S LAW UNDER THE *MDIP*

The *MDIP* is presented in this section as the variational principle guiding the evolution of the code. As we shall see at the end of this section, the consequences of its application result in a proof of unicity for the emergence of Zipf's law in evolving codes.

A. The MIDP and its consequences for the evolution of codes

The question is thus how the probability distribution q_n evolves along the growth process. Under the *MDIP* we face a variational problem which is stated as follows:

During the growth process, the most likely code at step $n + 1$ is the one minimizing the *distance* with respect to the code at step n, consistently with the $MDIP$. Furthermore, the evolution of the code must satisfy, along all the evolutionary steps, the symmetry condition depicted by eq. (2). The crucial contribution of the $MDIP$ is that it naturally introduces the footprints of the path dependence imposed by evolution. Following the thermodynamical metaphor, this variational principle acts, in our context, as a principle on energy minimization acting over the transitions of successive codes. Putting it formally, let

$$D(q_{n+1}||q_n) \equiv \sum_i q_{n+1} n(s_i) \log \frac{q_{n+1}(s_i)}{q_n(s_i)}$$

be the *Kullback-Leibler* Divergence of the distribution q_{n+1} with respect the distribution q_n [22]. Therefore, the $MDIP$ is achieved by minimizing the following functional [17]:

$$\mathcal{L}(q_{n+1}, \lambda) = D(q_{n+1}||q_n) + \lambda \left(\sum_{i \le n+1} q_{n+1}(s_i) - 1 \right).$$

We observe that this functional has a role equivalent to the one attributed to the Lagrangian function in a given continuous, differentiable system; therefore, the *trajectories* minimizing it will govern the evolution of the system. Furthermore, the symmetry condition on coding/decoding -eq. (2)- imposes that the solutions must lie in the region defined by eq. (6). The minimum of $\mathcal{L}$ is found when q_{n+1} satisfies:

$$q_{n+1}(s_i) = \begin{cases} \lambda q_n(s_i) & \text{iff } i \le n \\ 1 - \lambda & \text{iff } i = n+1, \end{cases} \qquad (7)$$

being λ the Lagrange multiplier, which is a positive, unique constant for all elements of the probability distribution q_{n+1}. We observe that, for $\lambda = 1$, $D(q_n||q_{n+1}) = 0$, but, in this case, $H(X_s(n)) = H(X_s(n+1))$, in contradiction to the assumption provided by eq. (6), according to which informative richness grows during the evolutionary process.

Now we want to find the asymptotic behavior of q_n, $n \to \infty$ under the above justified conditions (6) and (7). The key feature we derive from the path dependency in the evolution imposed by the $MDIP$ is that the following quotient

$$(\forall k + j \le n) \quad f(k, k+j) = \frac{q_n(s_{k+j})}{q_n(s_k)} \qquad (8)$$

does not depend on n. Therefore, along the evolutionary process, as soon as

$$q_n(s_k), q_n(s_{k+j}) > 0,$$
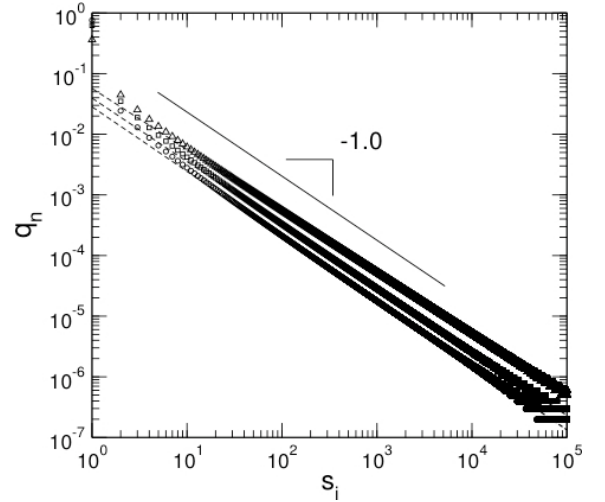
$f(k, k+j)$ remains invariant.

FIG. 2: Numerical simulation of the final distribution q_n, ($n = 10^4$) obtained by constraining the growth process with i)the consequences of the symmetry of coding/decoding -see eq. (2)- provided by eq. (6) and ii) the application of the $MDIP$ at every step of the growth process. Different convergence values are studied: a) $\nu = 0.2$, b) $\nu = 0.3$ and c) $\nu = 0.5$. The dashed lines are the best fit interpolations, which give estimated exponents $\gamma = 1.06, 1.04$ and 1.01, respectively (all with correlation coefficients $r < -0.99$).

B. The Emergence of Zipf's Law

The asymptotic behavior of quotient f and, thus, the tail of q_n, is strongly constrained by the entropy restriction provided by eq. (6) [12].

The first observation is that it can be shown that the convergence properties of the Riemann ζ-function on $\mathbb{R}^+$ [23] strongly constrain the convergence properties of the entropies of a given probability distribution [12]. Indeed, we find that, if $(\forall \delta > 0, n > m)(\exists N)$ such that:

$$(\forall m > N) \quad f(m, m+1) < \left(\frac{m}{m+1} \right)^{1+\delta},$$

then $(\exists C < \infty \in \mathbb{R}^+)$ such that $(\forall n)(H(X_s(n)) < C)$, which contradicts the assumptions of the problem, depicted by eq. (6). Indeed, primarily, one can observe that the above statement implies that q_n is *dominated* by a power-law having exponent $1 + \delta$, i.e. that q_n decays faster than q'_n, defined as:

$$q'_n(s_i) = \frac{i^{-(1+\delta)}}{Z_{1+\delta}},$$

where $Z_{1+\delta}$ is the normalization constant. Now, we write the explicit form of the entropy of $X'_s(n) \sim q'_n$ -to be written as $H(X'_s(n))$- when $n \to \infty$:

$$\lim_{n \to \infty} H(X'_s(n)) = \frac{1+\delta}{\zeta(1+\delta)} \sum_{i=1}^{\infty} \frac{\log i}{i^{1+\delta}} + \log(\zeta(1+\delta)).$$

We observe that all the elements of the above equation are finite constants, since

$$\sum_{i=1}^{\infty} \frac{\log i}{i^{1+\delta}} < \infty.$$

Thus, having q'_n as defined above,

$$\lim_{n\to\infty} H(X'_s(n)) < \infty.$$

Therefore, during the growth process, due to the constraint imposed by eq. (6),

$$f(m, m+1) > \left(\frac{m}{m+1}\right)^{(1+\delta)}, \qquad (9)$$

with δ arbitrarily small, provided that n can increase unboundedly. Otherwise,

$$\frac{H(X_s(n))}{\log n} \to 0,$$

in contradiction to the assumption that $\nu > 0$ as depicted in eq. (6).

Furthermore, we observe that, if $(\forall \delta > 0, \ n > m)(\exists N)$ such that:

$$(\forall m > N) \ f(m, m+1) > \left(\frac{m}{m+1}\right)^{(1-\delta)} \qquad (10)$$

then,

$$\lim_{n\to\infty} \frac{H(X_s(n))}{\log n} = 1,$$

again in contradiction to eq. (6), except in the extreme, pathological case where $\nu = 1$, when the coding process is completely noisy. To see how we reach this latter point we observe that statement (10) implies that q_n is *not dominated* by a power-law probability distribution q'_n having exponent $1 - \delta$, namely:

$$q'_n(s_i) = \frac{i^{-(1+\delta)}}{Z_{1-\delta}},$$

where $Z_{1-\delta}$ is the normalization constant. Putting explicitly the expression of the entropy for the random variable $X'_s(n)$, one obtains:

$$\begin{aligned}
\lim_{n\to\infty} \frac{H(X_s(n))}{\log n} &= \lim_{n\to\infty} \left(\frac{\delta(1-\delta)}{n^\delta \log n} \sum_{k\le n} \frac{\log k}{k^{1-\delta}} + \delta \right) \\
&= \lim_{n\to\infty} \frac{1-\delta}{\log n} \left(\log n - \frac{1}{\delta} \right) + \delta \\
&= 1,
\end{aligned}$$

which is the desired result. Accordingly, since from eq. (6) ν is generally different from 1,

$$f(m, m+1) < \left(\frac{m}{m+1}\right)^{(1-\delta)}. \qquad (11)$$

Thus, combining eq. (9) and (11), we have shown that the asymptotic solution is bounded by the following chain of inequalities:

$$\left(\frac{m}{m+1}\right)^{(1+\delta)} < f(m, m+1) < \left(\frac{m}{m+1}\right)^{(1-\delta)}.$$

The crucial step is that it can be shown that, if $n \to \infty$, we can set

$$\delta \to 0.$$

(The mathematical technicalities of this result can be found in [12].) This implies, in turn, that, for $n \gg 1$:

$$f(m, m+1) \approx \frac{m}{m+1}$$

and, from the definition of f provided in eq. (8), we conclude that

$$q_n(s_m) \propto \frac{1}{m},$$

thereby leading us to Zipf's law as the unique asymptotic solution.

In fig. (2) we numerically explored the behavior of the rank probability distribution of signals belonging to a growing code under the assumption of symmetry in coding/decoding provided by eqs. (2) and (6), and the $MDIP$ whose consequences in the evolution of q_n are depicted in eq. (7). The outcome perfectly fits with the mathematical derivations, showing very well-defined power-laws with exponents close to 1 although the convergence values ν diverge from 0.2 to 0.5. This numerical validation shows that the predicted asymptotic effects - i.e., the convergence of q_n to Zipf's law- are perfectly appreciated even in finite simulations where 10^5 signals are at work.

We end this section with a remark on the boundary conditions needed for the emergence of Zipf's law. In the section II B, we imposed that the potential information richness of the source must be unbounded. Such a condition is mathematically stated by (4). We observe that, more than an assumption, equation (4) is a boundary condition under which a growing code can (assymptotically) exhibit Zipf's law[26]. In this way, since $H(X_s(n))$ has a linear relation with $H(X_\Omega(n))$, the divergence of the latter implies the divergence of the former. And it is a required condition, since the entropy of a system exhibiting a power law with an exponent equal to 1 diverges with n. Otherwise, exponents are higher, or other probability distributions can emerge.

IV. DISCUSSION

The results provided in our study define a general rationale for the emergence of Zipf's law in the abundance of signals of evolving communication systems. The variational approach taken here as a formal picture of the least

effort hypothesis has two ingredients. First, starting from Zipf's conjecture, we reach a static symmetry equation to solve the communicative tension between coder and decoder. This is consistent with previous work, but reveals itself insufficient to derive Zipf's law as the unique solution, for it is easy to check that *static* equations of the kind of eq. (2) and (3) have infinite arbitrary solutions, even in the asymptotic regime, due to the possible parametrizations of the solutions. Secondly -and crucially- we consider that the code evolves through time, and that, consistently, there is a path dependence in its evolution, mathematically stated by imposing a variational principle, the $MDIP$, between successive states of the code. It is only by imposing evolution (and thus, path dependence) that we reach Zipf's law as the only asymptotic solution. Therefore, the origin of the power-law with exponent $\gamma = -1$ derives from three complementary, very general conditions:

- The unbounded informative potential of the code,

- the loss of information resulting from the symmetry condition, depicted in eq. (2), and

- evolution, and its associated path dependency, variationally imposed by the application of the $MDIP$ over successive states of the evolution of the system.

There is another, very interesting point, intimately tied to a code exhibiting Zipf's law and, more specially, the consequences of the *symmetry condition*, the mathematical ansatz which abstractly encodes the Zipf's hypothesis of vocabulary balance: The presence of an inevitable ambiguity in the code. It is a common observation that natural languages are ambiguous, namely, that linguistic utterances or parts of linguistic utterances can be assigned more than one interpretation. If the principle of least effort is at work, and thus there is a cooperative strategy between the coder and the decoder, then the presence of a certain amount of ambiguity is expected, provided that the speaker tends to assign more than one meaning to certain signals. Therefore, ambiguity is a byproduct of efficient communication rather than a fingerprint of poor communicative design.

The presented framework is general, and rigorously demonstrates that Zipf's law is a natural outcome of a broad class of communication systems evolving under coding/decoding tensions. In other words, Zipf's law emerges in a system where the coder and decoder *coevolve* under a general problem of energy minimization. The range of application to real-world phenomena, however, must be contrasted with the validity of data, for it has been pointed out that many supposed power-law behaviors show deviations when the statistical analysis is performed accurately [24, 25]. It should be noted, however, that a deviation of the predicted behavior need not be necessarily attributed to a fail of the framework. One should take into account that other constraints, such as general memory limitations, can play a role in shaping the final distribution.

Acknowledgments

[1] F. Auerbach, Patermans Geograpische Mittelungen **59**, 74 (1913).

[2] G. K. Zipf, *The Psycho-biology of language* (Routledge (London), 1936).

[3] G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley (Reading MA), 1949).

[4] P. Bak, C. Tang, and K. Wiesenfeld, Phys. Rev. Lett. **59**, 381 (1987).

[5] X. Gabaix, The Quarterly Journal of Economics **114**, 739 (1999)

[6] M. E. J. Newman, Contemporary Physics **46** (2005).

[7] V. Pareto, *Cours de d'economie Politique* (Droz. Geneva, 1896).

[8] H. A. Simon, Biometrika **42**, 425 (1955).

[9] W. Li, Information Theory, IEEE Transactions on **38**, 1842 (1992).

[10] P. Harremoës and F. Topsœ, Entropy **3**, 191 (2001).

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

[12] B. Corominas-Murtra and R. V. Solé, Phys. Rev. E **82**, 011102 (2010).

[13] P. Vogt, in *Artificial Life IX Proceedings of the Ninth International Conference on the Simulation and Synthe-sis of Living Systems*, edited by I. J. Pollack, M. Bedau, P. Husbands, T. Ikegami, and R. A. Watson (MIT Press, 2004).

[14] T. Maillart, D. Sornette, S. Spaeth, and G. von Krogh, Phys. Rev. Lett. **101**, 218701 (2008).

[15] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks* (Oxford University Press. New York, 2003).

[16] E. T. Jaynes, Physical Review **106**, 620 (1957).

[17] S. Kullback, *Information Theory and Statistics* (John Wiley and Sons. New York, 1959).

[18] A. R. Plastino, H. G. Miller, and A. Plastino, Phys. Rev. E **56**, 3927 (1997).

[19] J. Hurford, Lingua **77**, 187 (1989).

[20] M. A. Nowak and D. Krakauer, Proc. Nat. Acad. Sci. USA **96**, 8028 (1999).

[21] N. L. Komarova and P. Niyogi, Art. Int. **154**, 1 (2004).

[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory* (John Wiley and Sons. New York, 1991).

[23] M. Abramowitz and S. I. (editors), *Handbook of mathematical functions*, vol. 55 of *NBS, Appl. Math. Ser.* (U.S. Government Printing office, Washington, D.C., 1965).

[24] I. Kanter and D. A. Kessler, Phys. Rev. Lett. **74**, 4559 (1995).

[25] D. Avnir, O. Biham, D. Lidar, and O. Malcai, Science **279**, 39 (1998).

[26] We notice that eq. (4) depicts a linear relation between $H(X_\Omega(n))$ and $\log n$; i.e.: $H(X_\Omega(n)) \sim \mu \log n$. There are strong reasons to believe that one could generalize this statement by saying that the only condition needed is that, in spite that

$$\lim_{n \to \infty} \frac{H(X_\Omega(n))}{\log n} = 0,$$

if $H(X_\Omega(n))$ is a monotonous, growing and unbounded function on n, then Zipf's law would emerge using similar arguments to the ones used in this paper. The lack of a rigorous demonstration for this latter point forces us to restrict our arguments to the region of application of eq. (4).