Language Networks: their structure, function and evolution

Ricard V. Solé,[1,2] Bernat Corominas Murtra,[1] Sergi Valverde,[1] and Luc Steels[3,4]
[1]*ICREA-Complex Systems Lab, Universitat Pompeu Fabra (GRIB), Dr Aiguader 80, 08003 Barcelona, Spain*
[2]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe NM 87501, USA*
[3]*Sony CSL Paris, 6 Rue Amyot, 75005 Paris*
[4]*Vrije Universiteit Brussel, Pleinlaan 2, 1050 Brussels*

Several important recent advances in various sciences (particularly biology and physics) are based on complex network analysis, which provides tools for characterising statistical properties of networks and explaining how they may arise. This article examines the relevance of this trend for the study of human languages. We review some early efforts to build up language networks, characterise their properties, and show in which direction models are being developed to explain them. These insights are relevant, both for studying fundamental unsolved puzzles in cognitive science, in particular the origins and evolution of language, but also for recent data-driven statistical approaches to natural language.

Keywords: Language, evolution, neural networks, complex networks, syntax

I. INTRODUCTION

The origins and evolution of language and the relations to neural development are still largely unknown, despite the increased speculation and theorising going on at the moment (1; 2). Language does not leave fossils (at least not directly), and so solid grounds to support a well-defined theory of the evolution of the language faculty are largely missing. And yet human language is clearly one of the greatest transitions in evolution (3) and maybe the trait that makes us most essentially different from other organisms on our planet (1).

All languages share some universal tendencies at different levels of organization: the phoneme inventories, the syntactic and semantic categories and structures, as well as the conceptualisations being expressed. At the same time there are also very deep differences between languages, and often universal trends are implicational: They are about the co-occurrence of features and not the features themselves (4). For example, if a language has an inflected auxiliary preceding the verb, then it typically has prepositions. There are also universal statistical trends in human languages, such as Zipf's law (5), which is about the frequency with which common words appear in texts.

One of the most fundamental questions for a science of language is the origins of these universal trends. As with any other evolved system, there are many causal factors: There is first of all the nature of human embodiment and brain architecture. For example, the universal distinction between vowels and consonants is related to the structure of the human articulatory system, which can use vocal chords to produce vowels and strictures of the oral cavity for consonants (6). It has been argued that the brain features a kind of genetically determined language organ which strongly constrains the kinds of syntactic and semantic structures that we might expect in human languages (7; 8) and that language is subject to conditions imposed by other cognitive subsystems (8) or computational principles (9; 10), as well as memory limitations (particularly during language acquisition (11)).

A second causal factor is the nature of the tasks for which language is used, specifically communication. It is obvious that if language is to be adequate as a tool in communication, users will try to optimise communicative success and expressive power while minimising the cognitive and physical effort that they need to engage in. Various theories of language evolution explicitly take this to be the driving force in explaining the origins of grammatical features, such as expression of predicate-argument structure or perspective (12; 13). Other theories focus on the issue of transmission and argue that language is shaped largely so as to be able to be learnable by the next generation (14), or be genetically inheritable (15).

A third causal factor shaping universal trends comes from the family relations among languages (16). All Indo-European language have similar features, probably because they derived from a common ancestor, and they have continued to influence each other due to language contact. Many of the similarities we see among languages may therefore be a matter of historical contingency without any further deeper explanation.

In this paper we look at a fourth possible causal factor which is of a quite different nature. Over the past decade, it has become clear that complex dynamical systems exhibit a number of universal patterns both in their structure and in their evolution (17–19). Recently, important advances in graph theory, and specifically the theory of complex networks, have given a number of ways for studying the statistical properties of networks and for formulating general laws that all complex networks abide by, independently of the nature of the elements or their interactions (see box 1) (20; 21). Thus the study of ecological webs (22), software maps (23), genomes (24), brain networks (25; 26) or Internet architectures (27) all reveal common traits with characteristic efficiency and fragility (28). In this context, two main features seem

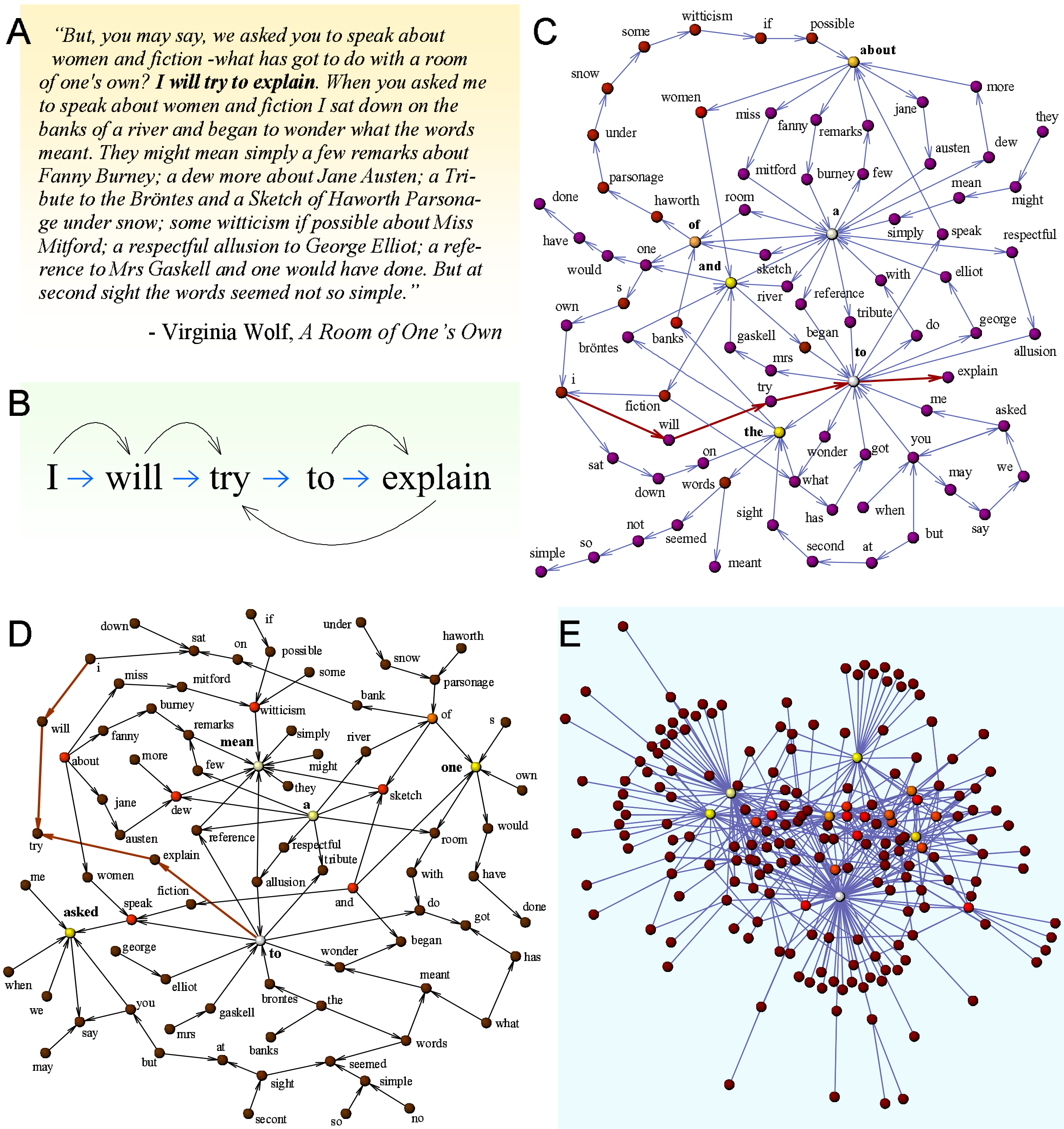

FIG. 1 How to build language networks. Starting from a given text (a) we can define different types of relationships among words, including precedence relations and syntactic relations. In (b) we show them using blue and black arrows, respectively. In figures (c) and (d) the corresponding co-occurrence and syntactic networks are shown. Paths on network (c) can be understood as the potential universe of sentences that can be constructed with the given lexicon. An example of such path is the sentence indicated in red. Nodes have been coloured according to the total (in- and out-) word degree, highlighting key nodes during navigation (The higher the degree the lighter its colour). In (d) we build the corresponding syntactic network, taking as a descriptive framework dependency syntax (50), assuming as criterion that arcs begin in complements and end in the nucleus of the phrase; taking the verb as the nucleus of well-formed sentences. The previous sentence appears now dissected into two different paths converging towards "try". An example of the global pattern found in a larger network is shown in (e) which is the coocurrence network of a fragment of *Moby Dick*. In this graph hubs we have *the, a, of, to.*

to be shared by most complex networks, both natural and artificial. The first is their *small world* structure. In spite of their large size and sparseness (i. e. a rather small number of links per unit) these webs are very well connected: it is very easy to reach a given element from another one through a small number of jumps (29). In social networks, it is said that there are just *six degrees of separation* (i.e. six jumps) between any two randomly chosen individuals in a country. The second is less obvious, but not less important: these webs are extremely heterogeneous: most elements are connected to one or two other elements and only a handful of them have a very large number of links. These *hubs* are the key components of web complexity. They support high efficiency of network traversal but are for the same reason their Achilles heel. Their loss or failure has very negative consequences for system performance, sometimes even promoting a system's collapse (30).

Language is clearly an example of a complex dynamical system (31; 32). It exhibits highly intricate network structures at all levels (phonetic, lexical, syntactic, semantic) and this structure is to some extent shaped and reshaped by millions of language users over long periods of time, as they adapt and change them to their needs as part of ongoing local interactions. The cognitive structures needed to produce, parse, and interpret language take the form of highly complex cognitive networks as well, maintained by each individual and aligned and coordinated as a side effect of local interactions (33). These cognitive structures are embodied in brain networks which exhibit themselves non-trivial topological patterns (25). All these types of networks have their own constraints and interact with the others generating a dynamical process that leads language to be as we find it in nature.

The hypothesis being explored in this article is that the universal laws governing the organization and evolution of networks is an important additional causal factor in shaping the nature and evolution of language (34). Commonalities seen in all languages might be an indicator of the presence of such common, inevitable generative mechanisms, independent on the exact path followed by evolution. If this is the case, the origins of language would be accessible to scientific analysis, in the sense that it would be possible to predict and explain some of the observed *statistical universals* of human languages in terms of instantiations of general laws. Current results within the new field of complex networks give support to this possibility.

II. LANGUAGE NETWORKS

If network structure is a potential key for understanding universal statistical trends then the first step is clearly to define more precisely what kinds of networks are involved. It turns out that there are several possibilities (Fig 1). First of all we can look at the network structure of the language elements themselves, and this at different levels: semantics and pragmatics, syntax, morphology, phonetics and phonology. Second, we can look at the language community and the social structures defined by their members. Social networks can help understanding how fast new conventions propagate or what language variation will be sustained (35). Moreover, the network organization of individual interactions has been shown to influence the emergence of a self-consistent language (36)

Box 1. Measuring network complexity

Several key statistical measures can be performed on a given language network as we illustrate here by focusing on the co-occurrence of words. If $\mathbf{W} = \{W_i\}$ is the set of words ($i = 1, ..., N$) and $\{W_i, W_j\}$ is a pair of words, a link can be defined based on a given choice of word-word relation. The number of links of a given word is called its *degree*. The total number of links is indicated by L, and the average degree $< k >$ is simply $< k >= 2L/N$.

Path length: it is defined as the average minimal distance between any pair of elements.

Clustering coefficient: is the probability that two vertices (e.g. words) that are neighbors of a given vertex are neighbors of each other. It measures the relative frequency of triangles in a graph.

Random graphs: they are obtained by linking nodes with some probability. The most simple example is given by the so called Erdös-Renyi (ER) graph for which any two nodes are connected with some given probability. For a random graph, we have very small clustering (with $C\ 1/N$) and $D \approx \log N / \log \langle k \rangle$.

Small world (SW) structure: this networks have a high clustering (and thus many triangles) but also a very short path length, with $D \sim \log N$ as in random graphs. A small world graph can be defined as a network such that $D \sim D_{rand}$ and $C \gg C_{rand}$.

Degree distributions: They are defined as the frequency $P(k)$ of having a word with k links. Most complex networks are characterized by highly heterogeneous distributions, following a power law (scale-free) shape $P(k) \sim k^{-\gamma}$, with $2 < \gamma < 3$.

Third we can study the interaction between language, meaning and the world, for example in terms of semiotic landscapes and the semiotic dynamics that occurs over these landscapes, both in the individual and the group (37; 38). Each viewpoint can be the basis of network analysis but in this article we only focus on the first type of relations as this has been already been explored the most extensively.

To further narrow our focus, we will take words as the fundamental interacting units, partly because this is very common in linguistic theorising but also because it is relatively straightforward to obtain sufficient corpus data to be statistically significant, and because several large-scale projects are under way for manual annotation of text based on lexical entries, such as Wordnet (39)

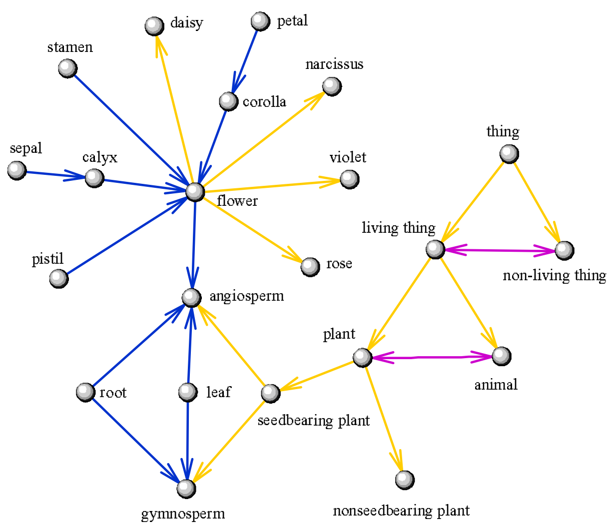

FIG. 2 Semantic webs can be defined in different ways. The figure shows a simple network of semantic relations among lexicalised concepts. Nodes are concepts and links semantic relations between concepts. Links are coloured to highlight the different nature of the relations. Yellow arcs define isa-relations (hypernymy) (*Flower* → *Rose* implies that *Flower* is a hypernym of *Rose*). Two concepts are related by a blue arc if there is a part-whole relation (*meronymy*) between them. Relations of binary opposition (*antonymy*) are bidirectional and coloured violet. Hypernymy defines a tree-structured network and other relations produce shortcuts that leads the network to exhibit a *small world* pattern *-see Box 1-*, making navigation through the network more easy and effective. (redrawn from (45))

and Framenet (40). We can then build for example the following kinds of networks:

Co-occurrence networks. (Fig.1b,e) Spoken language consists of linear strings of sounds and hence the first type of network that we can build is simply based on co-occurrence. Two words are linked if they appear together within at least one sentence. Such graphs can be undirected or directed. In undirected graphs the order of words in a sentence is considered irrelevant. It has been used for example in (41). A more realistic approach is to consider the order in which words co-occur and represent that in directed graphs (44; 46). Word order could partially reflect syntactic relations (47) and closeness between lexicalized concepts (48). Hence the study of such networks provides a glimpse of the generative potential of the underlying syntactic and semantic units. and we find hubs for words with low semantic content but important grammatical functions (such as articles, auxiliaries, prepositions, etc.). They are the key elements in sustaining an efficient traffic while building sentences. In figure 1C and example of shuch network is shown, with nodes colored proportionally to their degree. In these webs degree is directly related to frequency of appearance. Analysis of these networks in children, during language acquisition reveal that a trade-off is present in terms of a balance between lexicon size and flexibility: Children with smaller lexicon display a network with higher connectivity. Moreover, the ontogeny of these webs is well described by looking at the emergence of hubs. Content words such as *mama* or *juice* are important at early stages but fade out in later stages as function words (particularly *you, the, a*) emerge together with grammar (46).

Syntactic networks. (Figure 1c,d) A second type of network focuses on syntax (42). They can be built up based on constituent structures, where units form higher level structures which in turn then behave as units in other structures. Constituent structures are usually described as the product of fundamental operations like *merge* (49) or *unify* (2). Dependency grammar (50) is a useful linguistic framework to build up syntactic networks because it retains the words themselves as fundamental nodes in a graph (Figure 1c). Many of these networks could be extracted automatically using techniques from data-driven parsing (51). Here hubs are functional words but we can see that the in- and out- degrees are different from the previous web based on precendence relations.

Semantic networks. (Fig.2) Semantic networks can be built starting from individual words that lexicalise concepts and by then mapping out basic semantic relations such as isa-relations, part-whole or binary opposition. They can potentially be built up automatically from corpus data (43; 48; 52–54). The topology of these networks reveals a highly efficient organization where hubs are polysemous words, which have a profound impact on the overall structure. It has been suggested that the the hubs organize the semantic web in a categorical representation and might explain the ubiquity of polysemy across languages (43). In this context, although it has been argued that polysemy can be some type of historical accident (which languages should avoid) the analysis of these webs rather suggests that they are a necessary component of all languages. Additionally, as discussed in (48), the scale-free topology of semantic webs places some constraints on how these webs (and the previous ones) can be implemented in neural hardware. The high clustering found in these webs favours search by association, while the short paths separating two arbitrary items makes search very fast (52).

Many of these networks have now been analysed and they exhibit non trivial patterns of organization *-see box 1-*: Coocurrence, syntactic, and semantic networks exhibit scaling in their degree distributions and they display *Small World* effects with high clustering coefficients and short path lengths between any given pair of units (41–43; 48), which remarkably are the same universal features as found in most of the natural phenomena studied by complex network analysis so far (24; 27; 28).

	Co-occurrence networks	Syntactic Networks	Semantic Networks
Size	$N \sim 10^3 - 10^6$	$N \sim 10^3 - 10^4$	$N \sim 10^4 - 10^5$
Average degree	$< k >\sim 4 - 8$	$< k >\sim 5 - 10$	$< k >\sim 2 - 4$
Link relation	Precedence	Syntactic dependence	Word-word association
Functional meaning	Sentence production	Grammar architecture	Psycholinguistic web
Path length	$d \sim 3 - 4$	$d \sim 3.5$	$d \sim 3 - 7$
Clustering	$C/C_{rand} \sim 10^3$	$C/C_{rand} \sim 10^3$	$C/C_{rand} \sim 10^2$
Link distribution	SF, $\gamma \sim 2.2 - 2.4$	SF, $\gamma \sim 2.2$	SF, $\gamma \sim 3$
Hubs	Words with low semantic content	functional words	Polysemous words
Effect of hub deletion	Loss of optimal navigation	Articulation loss	Loss of conceptual plasticity

TABLE I Statistical universals in language networks. The data shown are based on different studies involving different corpuses and different languages (41–44; 48). The small world character of the webs appears reflected in the short path lengths and the high clustering. Here we compare the value of C with the one expected from a purely random graph C_{rand}. The resulting fraction C/C_{rand} indicates that the observed clustering is orders of magnitude larger than expected from random.

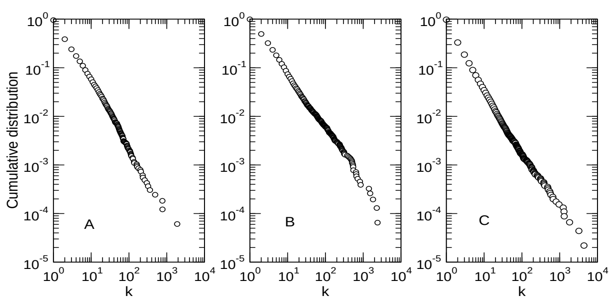

FIG. 3 Scaling laws in language webs. Here three different corpuses have been used and the co-occurrence networks have been analysed for: (a) basque, (b) english and (c) russian. Specifically, we computed the degree distribution $P(k)$ (box 1) and in order to smooth the fluctuations the cumulative distribution has been used, being defined as $P_>(k) = \sum_{j>k} P(j)$. Each corpus has 10^4 lines of text. Although some differences exist (44), their global patterns are rather similar, thus suggesting common principles of organization.

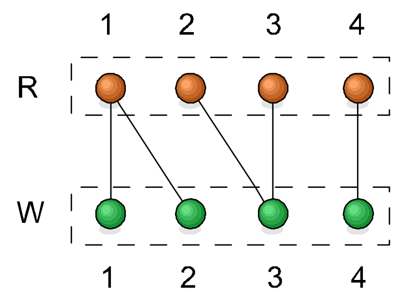

FIG. 4 Defining a bipartite graph of word-meaning association. A set of words (signals) W is linked to a set R of objects of reference. Here some words are connected to a single object (W_1, W_2, W_4) and other are polysemous (W_3). Some objects are linked to more than one word: W_1 and W_2 are thus synonyms.

This is highly relevant for two reasons. First it shows that language networks indeed exhibit certain significant statistical properties, confirming the discovery of a new type of universals in human languages similar to the universals found in (statistical) physics. Second, it strengthens the case that these statistical properties are a consequence of universal laws governing all types of evolving networks, independent of the specific cognitive and social processes that generate them. In the next section, we explore this topic further.

III. NETWORK GROWTH AND EVOLUTION

What could the structural principles for language networks be? Language networks are clearly shaped at two different scales (at least). The first is linked to language acquisition by individuals, and the second to its emergence within human populations. The ontogeny of language is strongly tied to underlying cognitive potentials in the developing child. But it is also a product of the exposure of individuals to normal language use (without formal training). On the other hand, the child (or adult) is not isolated from other speakers. The emergence and constant change in language needs to be understood in terms of a collective phenomenon (32) with continuous alignment of speakers and hearer to each other's language conventions and ways of viewing the world (33). Although the two previous views seem in conflict, they are actually complementary. It depends on what level of observation is considered. Without a social context in which a given language is sustained through invention, learning and consensus, no complex language develops. But unless a minimal neural substrate is present, it is not possible for language users to participate in this process.

The next question is: What is the nature of the shaping that affects the statistical properties of network growth? This question is still wide open and some successful approaches based on simple rules of network growth have been presented (48; 55). An interesting approach, based on constraints operating on communication and coding, can be defined in terms of two forces: one that pushes towards communicative success, and another one that pushes towards least effort, a principle already used by Zipf (5) and Simon (56). The first step is to define idealised simpler models of communicative interactions in the form of *language games* (57), possibly instantiated in embodied agents (58). In the simplest language game (usually called the Naming Game) a mapping between signals (words) and objects of reference is maintained by each agent (Box 2). This map defines a bipartite network (Fig.4). The most successful communication arises obviously when the individual lexicons take the form of the same one-to-one mapping between names and objects. Many simulations and theoretical investigations have now shown that this state can be rapidly reached when agents attempt to optimize their mutual understanding and adjust weights between signals and objects (37).

Box 2. Word-meaning association networks for Naming Games

Let us consider an external world defined as a finite set of m objects of reference (i. e. meanings):

$$\mathbf{R} = \{R_1, ..., R_i, ..., R_m\} \tag{1}$$

and the previous set words (signals) now used to label them $\mathbf{W} = \{W_i\}$. If a word w_i is used to name a given object of reference R_j, then a link will be established between both. Let us call $A = (a_{ij})$ the matrix connecting them, where $1 \leq i \leq n$ and $1 \leq j \leq m$. Here $a_{ij} = 1$ if the word a_i is used to refer to r_j and zero otherwise. A graph is then obtained, the so called *lexical matrix*, including the two types of elements (words and meanings) and their links. This is known as a *bipartite* graph. An example of such graph is shown in figure 3 where A reads:

$$A = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{2}$$

This matrix enables the computation of all relevant quantities concerning graph structure.

Next, we not only consider communicative success but also the cost of communication between hearer and speaker in order to connect lexical matrices as studied in language games with the kind of statistical regularities made visible by language network analysis. As noted by George Zipf the conflict between speaker and hearer might be explained by a lexical tradeoff which he named *the principle of least effort* (5). This principle can be made explicit using the lexical matrix, when the relative efforts of the hearer and the speaker are properly defined ((13). If we indicate by E_s and E_h the speaker and hearer's efforts, respectively, it is possible to assign a weight to their relative contribution by means of a linear function:

$$\Omega(\lambda) = \lambda E_s + (1 - \lambda) E_h \tag{3}$$

where $0 < \lambda < 1$ is a parameter. Effort is defined in terms of information theoretic measures (59). For E_s, the number of different words in the lexicon can be used (the signal entropy), whereas for E_h, the ambiguity for the hearer. The minimal effort for the speaker is obtained when a single word refers to many objects (Fig.4a) but this is maximal effort for the hearer (most ambiguous lexical network). The opposite case is indicated in Fig.4c. It provides minimal effort for the hearer, because the speaker is using one word for each object, but that means maximal effort for the speaker.

By tuning λ, we can move from one case to the other. What happens in between? Interestingly, a sharp change occurs at some intermediate λ_c value, where we observe a rapid shift from no-communication to one-to-one networks. This is a phase transition similar to the ones found in physics. At this phase transition (Fig.4b) we recover several quantitative traits of human language. The first is Zipf's law: if we map the number of links of a word to its frequency (as it occurs in real language) the previous scaling relation is recovered at close to the critical λ_c point. Such emergence of scaling close to critical points shares a number of characteristics with scenarios observed in physical systems under conflicting forces (19). An additional finding is that the bipartite graph with a word frequency distribution following Zipf's law can also impose strong constraints on syntactic networks (60; 61). It is worth mentioning that the presence of a gap has immediate consequences for understanding the origins of the unique character of human language compared to other species: a complexity gap in the patterning of word interactions would be required should be overcome in order to achieve symbolic reference.

IV. DISCUSSION

This article argued that there are statistical universals in language networks, which are similar to the features found in other 'scale free' networks arising in physics, biology and the social sciences. This observation is very exciting from two points of view. First, it points to new types of universal features of language, which do not focus on properties of the elements in language inventories as the traditional study of universals (e.g. phoneme inventories or word order patterns in sentences) but rather on statistical properties. Second, the pervasive nature of these network features suggests that language must be subject to the same sort of self-organization dynamics than other natural and social systems and so it makes sense to investigate whether the general laws governing

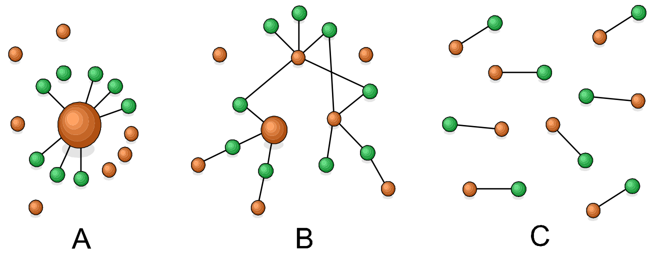

FIG. 5 Bipartite networks of word-meaning interactions in a least effort model of language networks. Here the size of the words is related to their use. As the effort of the speaker E_s increases (from left to right) more words are required to refer to objects. In (a) a low E_s implies a few words being used (with a high effort for the hearer) whereas in (c) a one-to-one signal-object association is obtained, with high E_s. At some intermediate, critical point (b) a transition occurs between both extremes, in which the frequency of word use follows Zipf's law (see text) and polysemy is present but limited to a small number of words. For clarity, isolated or rare words are drawn with a fixed size, although they should have a very small (or zero) radius.

complex dynamical systems apply to language as well and what aspects of language they can explain. We briefly discussed an example of such an effort for the most simplest form of language, namely names for objects.

The study of language networks and the identification of their universal statistical properties provides an tentative integrative picture. The previous networks are not isolated: In figure 6 we summarize some of the key relations, in which language networks define a given scale within a community of interacting individuals. The changes in lexicon and grammar through time are tied to social changes. A common language is displayed by any community of speakers, but it is far from a stable entity. Semantic and syntactic relations are at the basis of sentence production and they result from evolutionary and social constraints. Although we have a limited picture of possible evolutionary paths, the presence of language universals and the use of mathematical models can help elucidate them. On the other hand, the ontogeny of these networks offers an additional information that can also help distinguishing the components affecting the emergence of human language and the role played by genetic versus cultural influences.

The explanation of universal statistical network features is even more in its infancy. There are several reasons for this. First of all the explanation of these features generally requires that we understand the forces that are active in the building of the networks. This means that we must adopt an evolutionary point of view in the study of language, which contrasts with the dominating structuralist trend in the 20th century focusing only on the synchronic description of language. Second we must develop more complex models of the cognitive processes that underly the creation, maintenance, and transmission of language networks. The example discussed earlier argued that there are two basic forces at work: maximizing communicative success and expressive power while minimizing effort required, but the translation of these forces into concrete cognitive models and their effect for the explanation of the more complex aspects of language, in particular grammar, is to a large extend open. Third, in many biological systems, the general laws governing complex adaptive systems not only act as limiting forces, but also as forces *leading to* the emergence of complex and ordered structures (18). This fascinating hypothesis still remains to be examined seriously for the case of language evolution.

Although we believe that the application of network analysis to the study of language has an enormous potential, it is also worthwhile to point out the limits. First of all, there are other types of universal trends in language which cannot be explained by this approach, either because they are not related to statistical features of language networks or because they are due to the other causal factors discussed in the introduction: human embodiment, the nature of the tasks for which language is used, or historical contingencies and family relations between languages. Second, as Herbert Simon already argued (56) p.440, the ubiquity of certain statistical distributions, such as Zipf's law, and the many mechanisms that can generate them means that they do not completely capture the fine-grained uniqueness of language.

Clearly, the study of language dynamics and evolution needs to be a highly multidisciplinary effort (62). We have proposed a framework of study that helps to understand the global dynamics of language and brings language closer to many other complex systems found in nature, but much remains to be done and many disciplines within cognitive science will have to contribute.

Box 4. Questions for future research

How do language networks grow through language acquisition?

Are there statistical differences among networks for different languages?

Can a typology of languages be constructed where the genealogical relations are reflected in network features?

How do general principles discovered in statistical physics play a role in the topology of language networks?

Can artificial communities of agents develop languages with scale-free network structures?

How are language networks modified through aging and brain damage?

Is there a link between cortical maps involved in language and observed language networks?

Acknowledgments

The authors would like to thank the members of the Complex Systems Lab for useful discussions. RVS and

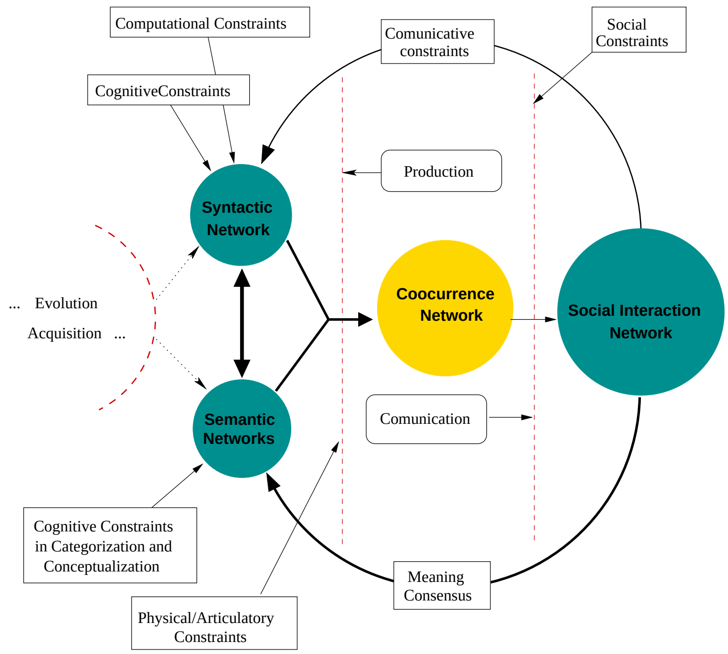

FIG. 6 The network of language networks. The three classes of language networks are indicated here, together with their links. They are all embedded within embodied systems belonging themselves to a social network. The graph shown here is thus a feedback system in which networks of word interactions and sentence production shape, and are shaped by, the underlying system of communicating agents. Different constraints operate at different levels. Additionally, a full understanding of how syntactic and semantic networks emerge requires an exploration of their ontogeny and phylogeny.

BCM thank Murray Gell-Mann, David Krakauer, Eric D. Smith, Constantino Tsallis for useful discussions on language structure and evolution. This work has been supported by grants FIS2004-0542, IST-FET ECAGENTS project of the European Community founded under EU R&D contract 011940, IST-FET DELIS project under EU R&D contract 001907, by the Santa Fe Institute and the Sony Computer Science Laboratory.

References

[1] Hauser, M.D., Chomsky, N. and Fitch, W.T. (2002). The faculty of language: what is it, who has it,and how did it evolve? Science 298, 1569-1579.

[2] Jackendoff, R. (1999) Possible stages in the evolution of the language capacity. Trends Cogn. Sci. 3(7), 272–279.

[3] Maynard-Smith, J. and Szathmàry, E. (1995) *The major transitions in evolution.* Oxford U. Press, Oxford.

[4] Greenberg J.H. (1963) *Universals of Language.* MIT Press. Cambridge, Ma.

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

[6] Peter Ladefoged & Ian Maddieson (1996). The sounds of the world's languages. Blackwell. UK

[7] Pinker, S. and P. Bloom (1990) Natural language and natural selection. Behav. Brain Sci. 13, 707-784.

[8] Chomsky, N. (2002) Interview on minimalism. In: *On Nature and Language*, Belletti, A. and Rizzi, Luigi (Editors) Cambridge University Press, New York.

[9] Chomsky, N. 2000. Minimalist inquiries: the framework. In Roger Martin, David Michaels, and Juan Uriagereka, eds., *Step by step: essays on minimalist syntax in honor of Howard Lasnik*, pp. 89-155. MIT Press. Cambridge, Mass.

[10] Piatelli-Palmarini, M. and Uriagereka, J. (2004) The immune syntax: the evolution of the language virus. In: Lyle Jenkins, ed, *Variation and Universals in Biolinguistics*, pp. 341–377. Elsevier.

[11] Newport, E. L. (1990). Maturational constraints on language learning. Cogn. Sci. 14: 11-28.

[12] Steels, L. (2005) The emergence and evolution of linguistic structure: from lexical to grammatical communication systems. Connection Sci. 17(3-4), 213–230.

[13] Ferrer i Cancho, R. and Solé, R. V. (2003) Least effort and the origins of scaling in human language. Procs. Natl. Acad. Sci. USA 100, 788-791

[14] Kirby, S. and Hurford, J. (1997) Learning, culture and evolution in the origin of linguistic constraints. In P. Husbands and I. Harvey, editors, ECAL97, pages 493–502. MIT Press.

[15] Nowak, M. A., Plotkin, J. B. & Krakauer, D. C., 1999. The Evolutionary Language Game, J. Theor. Biol. 200, 147-162.

[16] Greenberg, Joseph H. (2002) *Indo-European and its Closest Relatives: the Eurasiatic Language Family* Volume I and II. Stanford University Press, Stanford.

[17] Prigogine, I. and Nicolis, G. (1977) *Self-Organization in Non-Equilibrium Systems: From Dissipative Structures to Order Through Fluctuation.* J. Wiley & Sons, New York

[18] Kaufman, S. (1993) *The Origins of Order: Self-Organization and Selection in Evolution* Oxford Univer-

sity Press, Oxford.

[19] Solé, R.V. and Goodwin, B.C. (2001) *Signs of Life: how complexity pervades biology.* Basic Books, Perseus, New York.

[20] Albert, R. and Barabási, A-L. (2001) Statistical mechanics of complex networks. Rev. Mod. Phys. 74, 47-97.

[21] Dorogovtsev, S.N. and Mendes, J.F.F. (2003)*Evolution of Networks: from biological nets to the Internet and WWW* Oxford U. Press, Oxford

[22] Montoya, J. M. and Solé, R. V. (2002) Small world patterns in food webs. J. Theor. Biol. 214, 405-412 .

[23] Valverde, S., Ferrer-Cancho, R., and Solé, R. V. (2002) Scale Free Networks from Optimal Design, Europhys. Lett. 60, 512-517.

[24] Solé, R. V. and Pastor-Satorras, R. (2002) Complex networks in genomics and proteomics. In: *Handbook of Graphs and Networks*, S. Bornholdt and H.G. Schuster, eds. pp. 147-169. John Wiley-VCH. Berlin.

[25] Eguiluz, V., Cecchi, G., Chialvo, D. R., Baliki, M. and Apkarian, A. V. (2005) Scale-free brain functional networks. Phys. Rev. Lett. 92, 018102.

[26] Sporns O, Chialvo DR, Kaiser M, and Hilgetag CC. (2004) Organization, Development and Function of Complex Brain Networks. Trends Cogn. Sci. 8(9), 418-425.

[27] Albert, R.; Jeong, H. and Barabási, A. (1999) Diameter of the World Wide Web. Nature 401, 130

[28] Solé, R, V.; Ferrer i Cancho, R.; Montoya, J. M. and Valverde, S. (2002) Selection, Tinkering and Emergence in Complex Networks. Complexity 8, 20-33

[29] Watts, D. J. and Strogatz, S. H. (1998) Collective dynamics of 'small-world' networks. Nature 393, 440-442

[30] Albert, A., Hawoong, J and Barabási, A.L. (2000) Error and attack tolerance of complex networks. Nature 406, 378 - 382

[31] Briscoe, E. J. (1998) Language as a Complex Adaptive System: Coevolution of Language and of the Language Acquisition Device. In: H. van Halteren et al., editor, *Proceedings of Eighth Computational Linguistics in the Netherlands Conference*.

[32] Steels, L. (2000) Language as a Complex Adaptive System. In: Schoenauer, M., editor, *Proceedings of PPSN VI, Lecture Notes in Computer Science*. pp. 17-26. Springer-Verlag. Berlin.

[33] Pickering, M. J. and S. Garrod (2004). "Toward a mechanistic psychology of dialogue." Behav. Brain Sci. 27, 169-226.

[34] Ferrer i Cancho, R. (2004) *Language: universals, principles and origins.* PhD Thesis. Universitat Politecnica de Catalunya. Barcelona, Spain.

[35] Labov, W. (2000) *Principles of Linguistic change.* Volume II: Social Factors. Oxford: Blackwell.

[36] Corominas, Bernat and Solé Ricard (2005) Network Topology and self-consistency in Language Games. Journal of Theoretical Biology *in press.*

[37] Baronchelli, A., Felici, M., Caglioti, E., Loreto, V., and Steels, L. (2005) Sharp Transition towards Shared Vocabularies in Multi-Agent Systems. http://arxiv.org/pdf/physics/0509075

[38] Steels, L. and Kaplan, F. (1999) Collective learning and semiotic dynamics. In: Floreano, D. and Nicoud, J-D and Mondada, F., editors, *Proceedings of ECAL99*, pp. 679–688. Springer-Verlag, Berlin.

[39] Fellbaum, C. (1987) *Wordnet: An electronic lexical database.* Cambridge, Ma: MIT Press.

[40] Baker, C. F., C.J. Fillmore, and J.B. Lowe (1998) The Berkeley Framenet Project. In: *Proceedings of the 17th international conference on Computational linguistics* - Volume 1, pp. 86 - 90. Association for Computational Linguistics. Morristown, NJ, USA.

[41] Ferrer i Cancho, R. and Solé, R. V. (2001) The Small-World of Human Language. Proc. R. Soc. Lond. Series B 268, 2261-2266.

[42] Ferrer i Cancho, R.; Koehler, R. and Solé, R. V. (2004) Patterns in syntactic dependency networks. Phys. Rev. E 69, 32767

[43] Sigman, M. and Cecchi, G.A. (2002) Global organization of the Wordnet lexicon. Procs. Natl. Acad. Sci. USA 99(3), 1742–1747.

[44] Solé, R.V., Corominas and B. Valverde, S.(2005) Global Organization of Scale-Free Language Networks. SFI Working Paper 05-12.

[45] Miller, G. A. (1996) *The Science of words.* Sci. Am. Library. Freeman and Co. New York.

[46] Jinyun, K. (2004) *Selforganization and language evolution: system, population and individual.* Phd Thesis. City University of Hong Kong.

[47] Kayne, R. (1994) *The Antisymmetry of Syntax.* Cambridge, Mass: MIT Press.

[48] Steyvers, M. and Tenenbaum, J. B. (2005) The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. Cogn. Sci. 29(1), 41-78.

[49] Chomsky, N. (1995). *The Minimalist Program* Cambridge, Mass: MIT. Press.

[50] Melçuck, I. (1989). *Dependency grammar: Theory and practice.* New York, University of New York.

[51] Bod, R., R. Scha and K. Sima'an (eds.), 2003. *Data-Oriented Parsing.* University of Chicago Press, Chicago.

[52] Motter, A. e. et al. (2002) Topology of the conceptual network of language. Phys. Rev. E65, 065102.

[53] Kinouchi, O., Martinez, A. S., Lima, G. F., Lourenço, G. M.and Risau-Gusman, S. (2002) Deterministic walks in random networks: an application to thesaurus graphs. Physica A 315, 665-676.

[54] Holanda, A. J., Torres Pisa, I., Kinouchi, O., Souto Martinez, A. and Seron Ruiz, E. (2004) Thesaurus as a complex network. Physica A 344, 530-536.

[55] Dorogovtsev, S. N. and Mendes J. F. F. (2001) Language as an evolving word web. Proc. Royal Soc. London B 268, 2603-2606.

[56] Simon, H. (1955) On a Class of Skew Distribution Functions, Biometrika 42, 425–440

[57] Steels, L. (1995) A Self-organizing Spatial Vocabulary. Artificial Life 2, 319-332.

[58] Steels, L. (2003) Evolving Grounded Communication for Robots. Trends Cogn. Sci. 7,308–312.

[59] Ash, J. (1965) *Information Theory* John Wiley & Sons, New York

[60] Ferrer i Cancho, R., Bollobás, R, & Riordan, O. (2005). The consequences of Zipf's law for syntax and symbolic reference. Proc. R. Soc. Lond. Series B 272, 561-565.

[61] Solé, R. V. (2005) Syntax for free? Nature 434, 289 (2005)

[62] Christiansen, M. H. and Kirby, S. (2003) Language Evolution: Consensus and Controversies. Trends Cogn. Sci., 7,300–307.