Ferrer i Cancho, R. (2008). Network theory. In: The Cambridge encyclopedia of the language sciences, Colm Hogan, P. (ed.). Cambridge University Press, in press.

NETWORK THEORY

Network theory concerns itself with the study of elements, called “vertices” (e.g., words), and their connections, called “edges” or “links” (e.g., two words are connected if one word has been elicited by the other in a word association experiment; Fig. 1). This theory has many applications in language sciences and is the outcome of intersecting work of mathematicians and physicists, who usually call it “graph theory” (Bollobás 1998) or “complex network theory” (Newman 2003), respectively. One of the major contributions of physicists has been unravelling the statistical properties of real networks (Newman 2003), e.g., the World Wide Web or protein interaction networks. Firstly, physicists discovered that practically all real networks exhibited the small world phenomenon. The term small-world comes from the observation that everyone in the world can be reached through a short chain of social acquaintances although the number of people of the whole social network is huge. In the word association network partially shown in Fig. 1, *volcano* is reached from *ache* through a chain of at least four links, while only one links separates *fire* from *volcano*. Secondly, physicists found that many real networks had an heterogeneous degree distribution. Loosely speaking, this property means that there are vertices (e.g., words) with a disproportionately large number of connections, the so-called hubs). For instance, in the network partially shown in Fig. 1, the five words with the highest degrees are *food*, *money*, *water*, *car* and *good* (Steyvers and Tenenbaum 2005). Finally, another fundamental property of real networks is clustering, i.e. roughly speaking, that if two vertices are connected to the same vertex they are likely to be directly connected as well.

Network theory has contributed to the study of language in three ways: (a) characterizing the statistical properties of linguistic networks, such as networks of word association (Steyvers and Tenenbaum 2005), thesauri (Sigman and Cecchi 2002) and syntactic dependencies (Ferrer i Cancho *et al.* 2004) (b) modelling the properties of these networks (Steyvers and Tenenbaum 2005; Motter *et al*. 2002) and (c) proposing abstract models that provide a further understanding of the faculty of language (Ferrer i Cancho *et al.* 2005).

Although the systematic application of network theory to language is a young field (starting in the early twenty first century) within *QUANTITATIVE LINGUISTICS*, it can be concluded that the small-world phenomenon, high clustering and heterogeneous degree distribution are common properties of linguistic networks (Mehler 2007). Most models proposed are based on the preferential attachment principle proposed by Barabási and Albert (1999): vertices (e.g., words) with more connections are more likely to become more connected in the future than those with fewer connections (Steyvers and Tenenbaum 2001,2005, Dorogovtsev and Mendes 2001, Motter *et al*. 2002).

The challenges of the application of network theory are explaining the properties of these networks (most studies are merely descriptive), incorporating deeper statistical techniques, e.g., degree correlation analysis (Serrano *et al.* 2006) and extending the studies to more languages (most studies are in English). For these reasons, it is too early to argue that the heterogeneous degree distributions and other statistical patterns found constitute *LAWS OF LANGUAGE* in the sense of *ABSOLUTE UNIVERSALS*. When applied to syntactic networks, network theory has helped to explain the origins of the properties of the syntactic dependency structure of sentences, e.g., the exceptionality of syntactic dependency crossings (Ferrer i Cancho 2006) and provides new tracks for understanding the *SYNTAX*, *UNIVERSALS OF* at the large scale of syntactic organization (Ferrer i Cancho *et al.* 2004), above the traditional sentence level.

In their pioneering application of network theory, Steyvers and Tenenbaum (2001,2005) studied the large scale organization of various kinds of semantic networks (e.g. word association networks; Fig. 1) and proposed a simple model for explaining the small-worldness, high clustering and an heterogeneous degree distribution of semantic networks. Over time, new vertices (e.g. words) are added and attached to existing vertices using two principles: Barabási-Albert's preferential attachment (see above) and differentiation. Differentiation means that a new vertex tends to mimic the connectivity pattern of an existing vertex.

Network theory has shed new light on the *EVOLUTION OF LANGUAGE* by defining the necessary conditions for the existence of language (e.g., word ambiguity) and also suggesting the possibility that language could have appeared for free as a side-effect of communication principles (Ferrer i Cancho *et al*. 2005).

--Ramon Ferrer i Cancho

Works Cited and Suggestions for Further Reading

“Bibliography on Linguistic and Cognitive Networks”, at http://www.lsi.upc.edu/~rferrericancho/linguistic_and_cognitive_networks.html.

Bollobás, Béla. 1998. *Modern Graph Theory*. New York: Springer. A helpful introduction to graph theory.

Barabási, Albert László and Réka Albert. 1999. “Emergence of Scaling in Random Networks” *Science* 286.5439: 509-512.

Dorogovtsev, Sergey and José Fernando Mendes. 2001. “Language as an Evolving Word Web.” *Proceedings of The Royal Society of London Series B*, Biological Sciences, 268.1485: 2603-2606.

Ferrer i Cancho, Ramon, Ricard V. Solé and Reinhard Köhler. 2004. ”Patterns in Syntactic Dependency Networks.” *Physical Review E* 69: 051915.

Ferrer i Cancho, Ramon, Oliver Riordan and Béla Bollobás. 2005. “The Consequences of Zipf's law for Syntax and Symbolic Reference.” *Proceedings of the Royal Society of London Series B* 272.1562: 561-565.

Ferrer i Cancho, Ramon. 2006. “Why do Syntactic Links not Cross?” *Europhysics Letters* 76.6: 1228-1234.

Mehler, Alexander 2007. Large Text Networks as an Object of Corpus Linguistics Studies. In *Corpus linguistics. An international handbook of the science of language and society*, ed. Anke Lüdeling and Merja Kytö. Berlin/New York: de Gruyter.

Motter, Adilson E., Alessandro P. S. de Moura, Ying-Cheng Lai and Partha Dasgupta. 2002. “Topology of the Conceptual Network of Language.” *Physical Review E* 65.6: 065102.

Newman, Mark. 2003. “The Structure and Function of Complex Networks.” *SIAM Review* 45.2: 167-256. A helpful introduction to complex network theory.

Serrano, Mari Ángeles, Marian Boguñá, Romualdo Pastor-Satorras Alessandro Vespignani. 2006. “Correlations in complex networks.” In *Structure and Dynamics of Complex Networks, from Information Technology to Finance and Natural Science* ed. Guido Caldarelli and Alessandro Vespignani. World Scientific, in press.

Sigman, Mariano and Guillermo A. Cecchi. 2002. “Global Organization of the Wordnet lexicon.” *Proceedings of the National Academy of Sciences USA* 99.3: 1742-1747.

Steyvers, Mark and Joshua B. Tenenbaum 2005. “The Large-scale Structure of Semantic Networks: Statistical Analyses and a Model for Semantic Growth.” *Cognitive Science* 29.1: 41-78. The first version available is from 2001: http://arxiv.org/ftp/cond-mat/papers/0110/0110012.pdf.

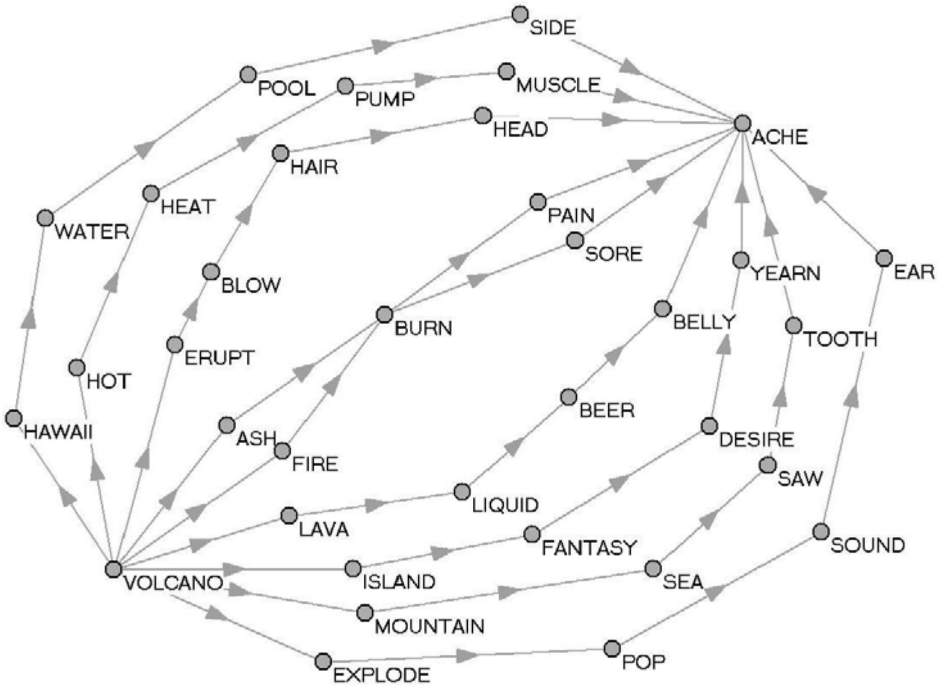

Fig. 1. A subset of a word association network appearing in Steyvers and Tenebaum (2005). Links go from the stimulus to the response word. Reproduced by permission of Cognitive Science Society, Inc.