EVOLUTION TOWARDS AN OPTIMAL MANAGEMENT OF LINGUISTIC INFORMATION

LLUÍS BARCELÓ-COBLIJN

Department of Psychology, University of the Balearic Islands,
Palma, Spain
lluis.barcelo@uib.cat

Network science shows great applicability, also in cognitive science related areas. Network-based approaches to the brain are able to extract a pattern, called the connectome. Typically, humans develop a small-world (SM) brain pattern, which seems to be an optimal network in terms of information management. However, in atypical development the pattern changes (or deviates). In the case of network-based approaches to language development, syntactic networks show that children develop their language capability till reaching a stage characterized by a SM network. It is suggested that during evolution human brains evolved in order to develop an optimal brain able to deal with a huge quantity of linguistic information. In order to better understand the typical path, information about atypical language development should be recovered. To date, there is still no information about atypical linguistic networks. In the present work three different linguistic disorders (down syndrome, hearing impairment and language specific impairment) are put to the test. This allows comparisons between different biological conditions that affect the global patterns of linguistic phenotypes.

1. Introduction

Network Science (Newman, 2010) has shown a great applicability, also in cognitive science (Baronchelli et al., 2013). In the last years, a new impulse from the interaction of network science and linguistic theory has emerged, so that new information – invisible to the unaided eye – about the speaker's linguistic capability is now available. Networks provide global information, which is only perceptible once many pieces of a system are connected; therefore, to focus directly on words or on individual sentences is not enough to extract this particular kind of information. In other words, the kind of information that is sought is related to the complexity of the system as a whole. After several network-based studies of language ontogeny – focused on the syntactic capability to combine words –, the study of atypical language development has also been addressed (Barceló-Coblijn et al. 2015), putting forth a research agenda which envisages syntactic networks as endophenotypes. The ultimate goal will be the configuration of a phenotypic morpho-space in which

each linguistic phenotype – either typical or atypical – can be placed Barceló-Coblijn & Gomila, 2014). Three sets of syntactic networks – including Down syndrome (DS), Specific language impairment (SLI) and Hearing Impairment (HI) – will be compared to a control group and to longitudinal studies of language acquisition. These analyses represent the first steps towards the morpho-space of linguistic phenotypes.

2. Complexity

Neil Fraser Johnson (2007) defines complexity science as the "study of the phenomena which emerge from a collection of interacting objects". To adopt this general definition implies that complexity could be applied to language studies too, and so complex systems theory offers a promising new approach to language. Networks can represent any system in which there are elements interacting to each other. A crucial and basic aspect of Network science is that two networks could contain the same number of nodes, but a different number of edges. Hence, these two networks would show a completely different behavior. By analyzing them, one would see radical different scores in the network measures: a network with k nodes could be *regular*, if all nodes have exactly the same number of connections; however, it could be *scale-free*, if there appear a small number of highly connected nodes (also known as *hubs*). A *scale-free* network of k nodes can develop the characteristic of *small-world*, if it has a high *cluster coefficient* (indicating how many neighbours of a given node are also neighbours of each other) and a low *path length* (how many connections must be traversed from a randomly chosen node to another randomly selected node). Thus, the n of nodes is of course important, but even more important is the particular connectivity a network displays (Watts & Strogatz, 1998).

3. Network Science and Linguistic studies

Applications of network approaches to language studies are easy to figure out: nodes can be phonemes, or words, or morphemes, etc, while edges would be interactions between those elements. Phonology has received quite a bit f attention (e.g., Vitevitch 2008): the goal was to capture particular aspects of the mental lexicon, where *Neighbourhood density* is an important concept: a word is considered a "phonological neighbour" if one single phonological change transforms that word into the target word. Thus, the attested similarities between languages as different as Spanish, Mandarin, Hawaiian and Basque strongly suggested commonly shared psycholinguistic mechanisms used in the architecture of the mental lexicon (Arbesman, et al. 2010). Its application is also being expanded to clinical studies (Vitevitch & Castro, 2015). Interestingly, the application of Network science to language is reaching other linguistic areas:

e.g., based on word co-occurrence, Beckage, Smith & Hills (2011) analyzed the semantic network growth in typically developed (from now on TD) and late talkers. Thus, their results support the thesis that *small-world* connectivity is somehow linked individual children's lexical development. Moreover, it has also been shown that language corpora let see a complex system behavior, and could be represented into a graph of word interactions (Ferrer-i-Cancho & Solé 2001). Thus, a novel like *Moby Dick* shows precisely a small-world network pattern of connectivity (Solé, et al. 2010). A different technique – this time combining syntactic theory and network science – shows that syntactically annotated corpora can have a network representation (Corominas-Murtra, et al. 2009). This work showed that it is possible to apply a network-based technique to the longitudinal stages during language development – as a way to go beyond the analysis of isolated sentences or utterances. Corominas-Murtra and collaborators' work represented in fact the birth of a particular technique of linguistic analysis, which is expanded here.

4. Syntactic networks

Following the line of research sketched out in the previous section, a new technique of linguistic analysis has emerged, which includes the syntactic analysis of each sentence in a corpus, phrase or utterance – following the Dependency Grammar (Hudson, 1990). Analyses of linguistic corpora covering one year of children's life have shown that children, regardless the language they acquire, follow the very same developmental stages (Corominas-Murtra, et al. 2009; Barceló-Coblijn et al. 2012). These three stages can be represented by a type of network: first a *tree-like* network, then a *scale-free* network and finally a *small-world* network. The syntactic analysis of language in combination with graph representation, not only applies the formal procedure of networks, but these are networks confidently encoded by a human linguist. Thus, lexical categories are differentiated, as well as syntactic relationships (i.e. in *a > b,* element *"a" syntactically depends on* element *"b"*). This technique uncovers how speakers combine words: which are the most connected words, what kind of relations they establish to each other and how the system behaves. The integration of all manually annotated sentences from a spontaneous speech corpus into a graph typically lets see that the graph is made up of one or more networks (called *components*). Metaphorically speaking, it is like a picture of a city taken from a hot-air balloon in the air. While standing on the street one can see the houses of that street – and many details of each house –, nothing can be said about how streets connect each other. Instead, a macroscopic picture of the town lets see how streets connect to each other, which houses are centric and which peripheral to the structure. It is the very same object, though from a

different perspective. Thus, a syntactic network becomes a kind of routing map of the speaker's public language.
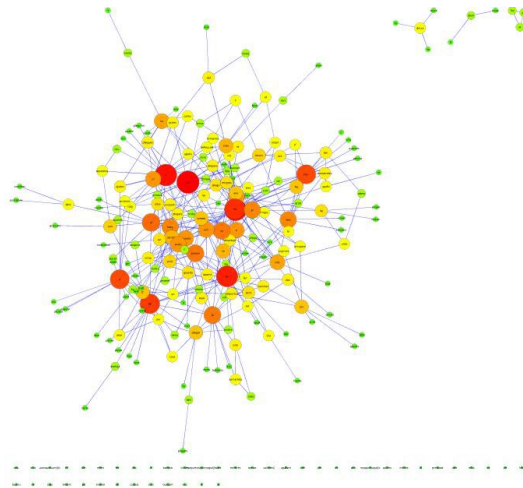
Figure 1. Graph from a sample of a Catalan speaker, age 2 years and 8 months.

5. Brain networks: the connectome

Network approaches have also dealt with brain structures, given the fact that the brain can also be considered a complex system. By means of combinations of neuroimaging techniques (e.g., diffusion tensor imaging) and network calculations, brain networks have become a real, useful tool in clinical studies too. According to Dehaene & Changeux (2011), higher order cognitive functions depend on a more integrative network topology. The structural configuration of the brain has indeed caught many researchers' attention, and several techniques have been developed in order to get more insight about it (Sporns, Tononi & Kötter, 2005). Following this point of view, van den Heuvel & Sporns (2011) could show that human brain networks have a kind of inner sub-structure only made of *hubs*: Up to 12 strongly connected bihemispheric hub regions. This structural characteristic has been dubbed as the *rich-club* effect. Collin & van den Heuvel (2013) review the connectome approach showing that there is a good deal of evidence pointing to the usefulness of network-based approaches to the study of both brain development and brain connectivity, which can be applied not only to ontogeny but also to ageing studies. Why could the macroscopic view be interesting for cognitive science? To be sure, it is a different perspective and hence the information quality will also be different than (and complementary to) the information from micro- or even mesoscopic approaches. Perhaps the most interesting thing of network-based techniques is

their capability to capture global dynamics. It seems that the *H. sapiens'* brain is structurally organized as a *small-world* network (henceforth SMN). This structure seems to be more suitable than others for dealing with huge quantities of information. As it has been mentioned before, a SMN has a high *clustering coefficient* and a low *path length*. Among other factors, the presence of *hubs* in the SMN structure has been related to those characteristics. Interestingly, Crossley and colleagues (2014) have observed that hubs of brain networks are generally central to clinical disorders like Alzheimer or schizophrenia. In their study, Crossley and colleagues could observe that "lesions were also disproportionately concentrated in rich-club regions rather than in the much larger number of peripheral regions". On another front, Ahmadlou et al. (2013) made an analysis of the functional brain connectivity in participants with Down syndrome (DS). The authors used a technique that combined the information from electroencephalogram frequency bands in a framework of Small-Worldness network. Results suggested that the connectivity of DS patients was disrupted in the whole brain in alpha and theta bands and that the global organization of the DS brain is closer to a *random* network than to a *SMN*.

Taking into account these data, it is worth thinking about the brain structure in pathological conditions, and how these conditions affect brain structure development. And then, how such an atypical brain development leads to an atypical linguistic profile. Are there similarities between the atypical brain networks and the atypical syntactic networks?

5.1. *Evolution towards an optimal structure?*

From the above mentioned studies on the connectome, the conclusion is that TD humans develop a brain structure reaching *small-worldness*. When particular biological factors intervene in the developmental path, the brain phenotype can develop atypically. It has repeatedly been observed that when a deviation from the typical developmental path takes place, the final connectome is also different. The above mentioned network-based studies on language ontogeny seem to agree with the connectome studies: TD children's ontogeny leads to a language capacity characterized by a *small-world* pattern, where for each word, two syntactic links are present. The syntactic network developed there can then be conceived as a reflection of the linguistic phenotype and therefore, for every person, a phenotype (a connectome) – and hence for every connectome a linguistic phenotype too. Thus, there are a (huge) limited number of brain phenotypes according to biological theory (in the lines of the Evo-Devo perspective), and therefore there will also be a (huge) limited number of linguistic phenotypes (Lorenzo & Balari, 2012).

This altogether suggests that there is in general a statistical tendency for brain developing towards the *small-world* pattern. The SM pattern, according to

physics, seems to be an optimal pattern for information processing. In the case of language development, current results also point to the same direction: TD speakers develop their capability for syntactically combining words following the same stages and reaching a characteristic SM pattern. It seems, thus, that something happened during the evolution of hominins that lead *H. sapiens* to develop so fast (during ontogeny) into that specific, linguistic, brain phenotype: A brain able to deal with an enormous amount of information, in an optimal way. However, in order to reinforce this hypothesis, it is necessary to check what happens in syntactic networks from atypically developing speakers.

6. Atypical development of language and networks

It has been observed since long time ago in biology that many morphological outcomes are possible, one for each biological phenotype. Biological variability can be represented by a set of possibilities that are the most probable within a population from the statistical point of view. There exist biological factors that provoke a deviation from the typical developmental path. Some deviations can be more abrupt than others (and hence in some cases the deviation of the linguistic phenotype would seem – from the external / observational viewpoint – more or less close to the typical linguistic phenotype of healthy speakers). We know that for brains, and for language too, thanks to decades of fruitful studies and discussions on linguistic disorders. However, there is still no information about atypical linguistic networks. In fact, if TD children develop a typical brain and a typical linguistic capability that both can be characterized by a *small-world* pattern, a simple prediction would be that any atypical connectome should correlate with an atypical linguistic network. This is something that has not been addressed until now.

6.1. *Procedure*

In this work, new information on networks is now presented, showing interesting aspects that uncover information about the linguistic capability under three different pathological conditions: Down syndrome, Hearing Impairment and Specific Language Impairment, in comparison to a control group. Each group has 20 Dutch speakers. The corpora were created by Bol & Kuiken (1990) and are available at CHILDEs data base. The syntactic analyses of the present work have been done by means of new software, *Netlang*, specifically created for this purpose (see http://neurolang.wordpress.com).
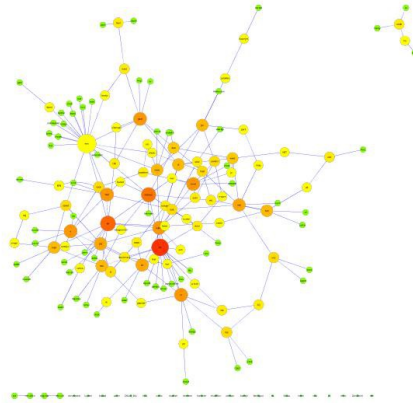
Figure 2. Graph from a sample of a speaker affected by SLI, age 5 years, 1 month and 2 days.

In contrast to previous works, the analyses made with this software has allowed encoding syntactic relationships in the edges: the relationship "subject" can be now seen "inside" the link between two nodes. An additional enhancement is that the software allows the inclusion of lexical categories too. So, the expression "house" can be categorized as "noun", and this information can be integrated into the network.

6.2. *Results and first conclusions*

Among other things, it will be shown that many of the atypical linguistic networks present a *non*-small-world pattern like in the TD children's (typically characterized by the 1:2 relation of words/syntactic links; that is to say, for each word, the child produces two syntactic links). Moreover, differences hubs connectivity – like in the atypical connectome. Thus, it seems that also in the case of atypical syntactic networks, hubs are specially affected, in comparison to TD syntactic networks.

There are a number of reasons to pursue the study of networks as endophenotypes, i.e. as biological markers that help to see the connection between genotype and phenotype (Barceló-Coblijn, et al. 2015). To develop such an enterprise, it is first necessary the development of new computational tools that reduce the number of problems due to typological differences between languages. Additionally, larger groups of speakers must be analyzed. Those seminal works included few speakers since these were longitudinal studies (up to 17 files for each speaker, covering between the second and third year of life) and hence focused on ontogeny. Cross-sectional studies can also be informative,

since they have a larger number of participants – and hence statistical calculations are more reliable (this does not mean that a longitudinal study are worse than a cross-sectional one, when in fact they complement each other).

The present work represents a step forward in the study of linguistic networks, since it begins to put to the test the hypothesis put forward here about brain evolution towards an optimal structure for the management of linguistic information. By including three different linguistic disorders and larger sets of speakers, it gives way to compare between different biological conditions that affect the global patterns of the linguistic phenotypes.

Acknowledgements

References

Ahmadlou, M., Gharib, M., Hemmati, S., Vameghi, R., & Sajedi, F. (2013). Disrupted small-world brain network in children with Down Syndrome. *Clinical Neurophysiology: Official Journal of the International Federation of Clinical Neurophysiology*, *124*(9), 1755-1764.

Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos*, *20*(03), 679-685.

Balari, S. & Guillermo L. (2012). *Computational Phenotypes: Towards an Evolutionary Developmental Biolinguistics*. Oxford: Oxford University Press.

Barceló-Coblijn, L., Benítez-Burraco, A., & Irurtzun, A. (2015). Syntactic Networks as an Endophenotype of Developmental Language Disorders: An Evo-Devo Approach to Clinical Linguistics. *Biolinguistics*, *9:* 43-49.

Barceló-Coblijn, L. & Gomila, A. (2014) Syntactic devlopment in phenotypic space. In Cartmill, Roberts, Lyn & Cornish (eds.) *The Evolution of Language. Proceedings of the 10th International Conference (EVOLANG 10)*, pp. 33-40. Singapur, World Scientific.

Barceló-Coblijn, L., Corominas-Murtra, B., & Gomila, A. (2012). Syntactic trees and small-world networks: syntactic development as a dynamical process. *Adaptive Behavior*, *20*(6), 427-442.

Baronchelli A., Ferrer-i-Cancho, R., Pastor-Satorras, R., Chater, N. & Christiansen, M. Networks in Cognitive Science, *Trends in Cognitve Science*, 17(7): 348-360.

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PloS One*, *6*(5), e19348.

Bol, G. W. & Kuiken, F. (1990). Grammatical analysis of developmental language disorders: Astudy of the morphosyntax of children with specific

language disorders, with hearing Impairment and with Down's syndrome, *Clinical Linguistics and Phonetics*, 4, 77–86.

Collin, G., & van den Heuvel, M. P. (2013). The ontogeny of the human connectome: development and dynamic changes of brain connectivity across the life span. *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry*, *19*(6), 616-628.

Corominas-Murtra, B., Valverde, S., & Solé, R. V, (2009). The Ontogeny Of Scale-Free Syntax Networks: Phase Transitions In Early Language Acquisition. *Advances in Complex Systems (ACS)*, *12*(03), 371-392.

Crossley, N. A., Mechelli, A., Scott, J., Carletti, F., Fox, P. T., McGuire, P., & Bullmore, E. T. (2014). The hubs of the human connectome are generally implicated in the anatomy of brain disorders. *Brain: A Journal of Neurology*, *137*(Pt 8), 2382-2395.

Dehaene, S., & Changeux, J.-P. (2011). Experimental and theoretical approaches to conscious processing. *Neuron*, *70*(2), 200-227.

Ferrer-i-Cancho, R., & Solé, R. V. (2001). The small world of human language. *Proceedings. Biological Sciences / The Royal Society*, *268*(1482), 2261-5.

Goñi, J., Avena-Koenigsberger, A., Velez de Mendizabal, N., van den Heuvel, M. P., Betzel, R. F., & Sporns, O. (2013). Exploring the Morphospace of Communication Efficiency in Complex Networks. *PLoS ONE*, *8*(3), e58070.

Hudson, R. A. (1990). *English Word Grammar*. Blackwell Pub.

Johnson, N. F. (2007). Two's company, three is complexity: A simple guide to the science of all sciences. Oxford: Oneworld.

Newman, M. (2010). *Networks: An Introduction*. OUP Oxford.

Solé, R. V., Corominas‑Murtra, B., Valverde, S., & Steels, L. (2010). Language networks: Their structure, function, and evolution. *Complexity*, *15*(6), 20-26.

Sporns, O., Tononi, G., & Kötter, R. (2005). The human connectome: A structural description of the human brain. *PLoS Computational Biology*, *1*(4), e42.

van den Heuvel, M. P., & Sporns, O. (2011). Rich-club organization of the human connectome. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *31*(44), 15775-15786.

Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech, Language, and Hearing Research: JSLHR*, *51*(2), 408-422.

Vitevitch, M. S., & Castro, N. (2015). Using network science in the language sciences and clinic. *International Journal of Speech-Language Pathology*, *17*(1), 13-25.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of «small-world» networks. *Nature*, *393*(6684), 440-442.