★ BRIEFS ★

Syntactic Networks as an Endophenotype of Developmental Language Disorders: An Evo-Devo Approach to Clinical Linguistics

Lluís Barceló-Coblijn, Antonio Benítez-Burraco & Aritz Irurtzun

Usually, developmental language disorders are defined either symptomatically (based on a constellation of linguistic deficits appearing recurrently within a population) or etiologically (on the basis of a common underlying deficit), or both. On paper, each of these clinical categories is expected to be distinguished from other close entities at several levels of analysis (phenotypic, cognitive, neurobiological, genetic, etc.). Nonetheless, this is not typically the case: Comorbidity, variability, and heterogeneity are in fact a common outcome of the clinical practice. Ultimately, different disorders may share the same underlying deficit (e.g., phonological dysfunction in dyslexia and SLI); conversely, different deficits may give rise to the same disorder (e.g., both visual problems and phonological deficits may contribute to dyslexia) (Benítez-Burraco 2013).

If we want to achieve a better—and earlier—diagnosis of these conditions, we should improve the tools we employ at present. A promising approach is one relying on the endophenotypes of disorders. Endophenotypes may be defined as cognitive, neuroanatomical, neurophysiological, endocrine, or biochemical quantifiable components of the space between genes and diseases (Gould & Gottesman 2006). Endophenotypes refer to more specific (and more physiological) aspects of the body function, therefore they allow us to gain a more accurate diagnosis of its dysfunction (Gottesman & Gould 2003). Here we would like to advance a putative endophenotype of language disorders that combines four factors: (1) linguistic analysis (syntactic computation), (2) information management (communicative strategies), (3) recent evo-devo insights in the nature of phenotypic variation, and (4) network approaches to emergent properties of complex systems (surely, language it is; Deacon 2005).

To begin with, we would like to note that, although the set of pathological conditions already described by clinical linguists is ample, it is not unlimited either. In other words, variation is constrained or canalized, even in pathological states. At the same time, we observe that language is both sensitive to damage (e.g., some aspects of language processing are perturbed in nearly all disorders, like the proper use of inflectional cues in verbal and nominal morphology) and resistant to perturbation (e.g., a nearly functional language faculty may emerge at the term of growth in spite of severe underlying deficits).

Preparation of this work was supported in part by funds from the Spanish Ministry of Economy and Competitiveness (grant numbers FFI2013-44007-P (LB), FFI-2013-43823-P (ABB & AI), and FFI2014-61888-EXP (ABB)), the Basque Government (grant number IT769-13 (AI)), and the European Commission (grant number EC FP7/SSH-2013-1 AThEME 613465 (AI)).

ISSN 1450–3417
Biolinguistics 9: 043–049, 2015
http://www.biolinguistics.eu

Put simply, language is both plastic and robust, whereas language development is significantly canalized (Benítez-Burraco & Boeckx 2014). In evolutionary-developmental (evo-devo) approaches, the limited set of phenotypes that result from the interplay of the different factors regulating development are usually referred to as points within the morpho-space or adaptive landscape (McGhee 2006). Consequently, language disorders may well be characterized as possible—although dysfunctional—phenotypes within the whole landscape of language development potential. The real problem is that these phenotypes of the language faculty are still characterized in terms of the clinical categories we regard unsatisfactory (e.g., dyslexia, SLI, and the like). This may be optimized if we move downwards and consider instead some of their endophenotype(s). Because of their more biological nature, endophenotypes may reflect in a more reliable way how the impaired brain grows and how a more or less functional language capability instantiates in the pathological mind. More importantly, we expect (some of) them to be the axis delimiting the adaptive landscape of language development in the species (either normal or pathological). Nonetheless, not many confident endophenotypes of language disorders have been proposed up to now (see some exceptions in Neuhoff et al. 2012 or Peter et al. 2012). We believe that an evo-devo approach to disorders may further help to narrow and optimize the set of endophenotypes that are currently available.

In our opinion, one useful endophenotype of this sort may be the 'syntactic fingerprints' characterizing the child's ability to combine words at different stages of development and, specifically, the kind of networks resulting from the measurement of the combinations of syntactic items (words or morphemes) in real samples of speech (this ultimately reflecting the syntactic links among words within utterances). Because we expect these 'fingerprints' to confidently reflect how the typically developed faculty of language unfolds within the child's mind, we have hypothesized them to be language-independent. We further expect that different clinical conditions are characterized by different 'syntactic fingerprints' throughout development, as a result of different language faculties being implemented in the child's mind. In turn, this plausibly results from different brain architectures emerging from different molecular backgrounds (e.g. gene mutations, changes in protein homeostasis, and the like). Overall, we expect that our syntactic networks fulfill the set of properties that endophenotypes have to meet (see Gottesman & Gould 2003, Gould & Gottesman 2006). Although we are still testing many of the details of our hypothesis, some promising results have been achieved.

For starters, we developed a new analytical tool for measuring the syntactic complexity of the utterances produced by speakers in real conversations. Our tool follows the basic lines of dependency grammar (Hudson 1990), representing the direction of dependency relations as well as the nature of the dependencies themselves. Thus, we can label each syntactic item by its category (noun, verb, etc.) and capture the dependencies between pairs of syntactic items (say, between a noun like *dog* and a determiner like *the*)—whether it is a head–complement relationship, like in the phrase 'the dog', where *the* is the head (cf. Abney 1987, Longobardi 2000, and others), or a modification relationship, as in constructions like 'walk quickly', where the adjunct *quickly* is a modifier of the event of walking

(cf. Pietroski 2005). Importantly, our technique allows us to treat each syntactic item separately (in a morpheme-by-morpheme fashion), which is essential for the analysis of agglutinative and polysynthetic languages. Once the syntactic analysis is done, the information is sent to the network program, which encapsulates words or morphemes into nodes and creates edges between nodes from the syntactic links between them (the program also imports the kind of syntactic relationship, e.g., subject, complement, etc.).

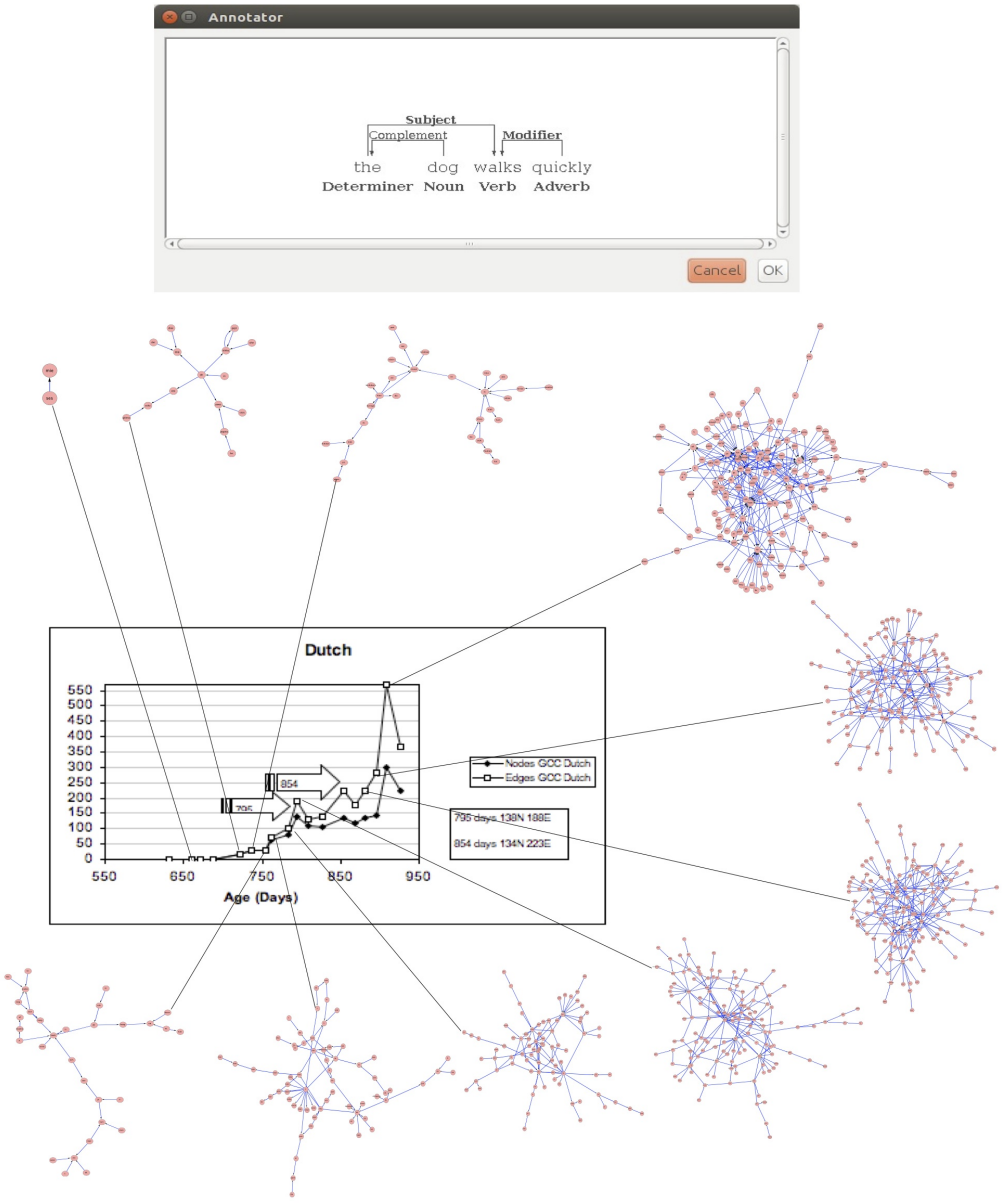

Figure 1: Language development as resulting from the network approach. On the top, the syntactic analysis is carried out. On the bottom, the development of the linguistic performance. In the pink networks, each word/morpheme is a node and each edge a syntactic relationship. This graphic belongs to the Dutch corpus Daan from CHILDES. In the graphic, white dots represent edges/ syntactic relationships, whereas black dots represent words/morphemes. White arrows point to the abrupt transition and change in the topology of the resulting networks.

We have shown that this approach confidently characterizes the development in the child of her ability to produce complex utterances with the combination of multiple words or morphemes. Interestingly, we found abrupt phase transitions in the syntactic complexity of the child's speech as she grows (from chain networks to scale-free networks to small-world networks) (Figure 1 above). We take this categorical difference in production to be a reflex of the different stages in the acquisition of the child's syntactic knowledge (Corominas-Murtra et al. 2009, Barceló-Coblijn et al. 2012). Importantly, these patterns are also informative about the words that are center-stage in the child's speech and ultimately, about communicative strategies. Importantly, this analytical tool is not language-dependent: We found similar network profiles at similar developmental stages when applied to the speech of typically developing (TD) children acquiring languages which are typologically diverse and belong to different phylogenetic groups (Germanic: Dutch, German, English; Romance: Catalan, Spanish, French, Italian; non-Indo-European: Basque) (Barceló-Coblijn et al. 2012, Barceló-Coblijn et al. submitted). For example, at 27–28 months of life, the ability to syntactically combine words achieved by TD children acquiring any of these languages can be regularly identified by a small-world network with a ratio of words/nodes vs. syntactic links/edges of 1:2 on average.

More importantly, we also used this analytical approach to confidently characterize language growth in pathological conditions. Different developmental disorders entailing language deficits are known to display pretty variable patterns of linguistic behavior. For instance, whereas SLI or Down syndrome are typically associated with a sharp syntactic disability, the performance of children suffering from other conditions like Fragile X syndrome is closer to that of their TD peers (see, for instance, Martin et al. 2013). Similarly, other syndromes like Williams syndrome are characterized by fluent speech, which on the surface does not seem to display such a patent syntactic disorder (see Bartke & Siegmüller 2004 for discussion). Ultimately, the variability observed within pathological groups (in terms of language knowledge and use) is typically greater than the variability within the normal population. Overall, it is quite difficult to draw a distinctive linguistic profile of each disorder.

As we pointed out above, we expect that biologically-driven factors that affect typical development provoke a deviation from this regular pattern of network transition found in the TD population so that they are achieved differently or are never achieved. Our preliminary results (Barceló-Coblijn et al. submitted) confirm that the networks reflecting syntactic development in some pathological populations like Down syndrome differ from those observed in TD children in several aspects, including the kind of network (and hence several network parameters like the clustering coefficient or the path length), the lexical nature of hubs, and the ratio nodes/edges. Likewise, our first assessment of the Williams Syndrome discourse (Palmer 2014) is also indicative of an idiosyncratic pattern of language growth, which is characterized by the modular nature of the resulting networks, despite the appearance of a typical speech, as noted above.

Interestingly, it is the network technique that allowed us to capture and formalize the language deficits (and the deviant developmental pattern) characteristic of this group that may be otherwise difficult to identify or even to observe

(obviously, Williams syndrome can be confidently diagnosed cytogenetically, but this is not always the case with conditions that are defined symptomatically, like autism or mental retardation; moreover cytogenetic analyses are expensive and may not be available under certain socio-economic circumstances).

We wish to end by briefly discussing the main translational values of our approach. First, the tool we have developed enables one to extract valuable information from real speech samples, which we feel is a more reliable source of information about the child's language knowledge and use (in contrast, tools currently used for the diagnosis of language disorders usually involve batteries of normalized tasks that have to be passed in controlled environments that may affect the child's performance). Second, because networks are characterized by a number of precise mathematical properties (like the clustering coefficient, the average path length, etc.), we further expect that the observed patterns are easier to quantify and have clearer diagnostic and prognostic correlates. Third, because we focused on a linguistic dimension that appears quite early in the child's discourse (syntax), we expect that our tool (and the kind of endophentype we propose) also allows for an earlier diagnosis of disorders (e.g., dyslexia cannot be reliably diagnosed until the child starts reading, at age 4–6, depending on the educational system).

Last, we expect our approach to be also of interest for the biological analysis of language (aka biolinguistics). On the one hand, because we heavily relied on a network approach for our analysis of how syntax emerges in the child's language, we expect to be able to accurately characterize how the properties of a complex system like language emerges during the child's growth. On the other hand, given that graph theory has recently been employed to define a two-dimensional morpho-space for complex networks (Goñi et al. 2013), we expect to be able to contribute as well to define the morpho-space of the available language faculties in the species. This latter approach focuses on two measures that capture communication efficiency within the network (routing and diffusion) and has shown that it is connectivity that matters and not just the *n* of nodes comprising the network. Under this view, two language networks may contain the same *n* of words/nodes but have a rather different *n* of syntactic links/edges or a different edge distribution—and hence a different kind of structure. This approach should help us to confidently characterize the complex networks resulting from the analysis of language growth in pathological populations.

Overall we expect that the whole set of language disorders (and the language faculty of non-affected individuals) can be translated into a constellation of complex networks located in different points of the language morpho-space, each characterizing a specific developmental itinerary for language, either normal or pathological (Barceló-Coblijn & Gomila 2014). Incidentally, all this conforms evidence that atypical language faculties also have their own developmental paths, although they grow in rather different ways. Actually, the network technique allows us to capture and formalize the fact that brains with linguistic disorders are not static entities. On the contrary, they are able to compensate damages at different levels and throughout growth—this probably explains why the linguistic profile of affected people varies in specific ways across populations and throughout development.

In sum, we regard this combination of syntactic analysis, complexity studies, and evo-devo theories as a promising approach to clinical linguistics. Specifically, we expect it to contribute developing better tools for diagnosing these complex conditions.

References

Abney, Steven P. 1987. The English Noun Phrase in Its Sentential Perspective. Cambridge: MA: MIT PhD dissertation.

Barceló-Coblijn, Lluís, Bernat Corominas-Murtra & Antoni Gomila. 2012. Syntactic trees and small world networks: Syntactic development as a dynamical process. *Adaptive Behavior* 20, 427–442.

Barceló-Coblijn, Lluís & Antoni Gomila. 2014. Syntactic development in phenotypic space. In Erica A. Cartmil, Séan Roberts, Heidi Lyn & Hannah Cornish (eds.), *The Evolution of Language: Proceedings of the 10th International Conference (EVOLANG10)*, 33–40. Singapore: World Scientific.

Barceló-Coblijn, Lluís, Aritz Irurtzun, A. & Maia Duguine. In press. The emergence of hubs in complex syntactic networks and the DP hypothesis: The relevance of a linguistic analysis. In Albert Bastardas-Boada, À. Massip-Bonet & Gemma Bel-Enguix (eds.), *Complexity Applications in Language and Communication Sciences*. Dordrecht: Springer.

Barceló-Coblijn, Lluís, Cristina Real, Aritz Irurtzun, Emilio López-Navarro & Antoni Gomila. Submitted. Networks reveal the phenotypic space of both typical and atypical development of syntax.

Bartke, Susanne & Julia Siegmüller (eds.). 2004. *Williams Syndrome across Languages.* Amsterdam: John Benjamins.

Benítez-Burraco, Antonio. 2013. Genetics of language: Roots of specific language deficits. In Cedric Boeckx & Kleanthes K. Grohmann (eds.), *The Cambridge Handbook of Biolinguistics*, 375–412. Cambridge: Cambridge University Press.

Benítez-Burraco, Antonio & Cedric Boeckx. 2014. Universal Grammar and biological variation: An EvoDevo agenda for comparative biolinguistics. *Biological Theory* 9, 122–134.

Corominas-Murtra, Bernat, Sergi Valverde & Ricard Solé. 2009. The ontogeny of scale-free syntax networks: Phase transitions in early language acquisition. *Advanced Complex Systems* 12, 371–392.

Deacon, Terrence W. 2005. Language as an emergent function: Some radical neurological and evolutionary implications. *Theoria* 20, 269–286.

Goñi Joaquín, Andrea Avena-Koenigsberger, NievesVélez de Mendizábal, Martijn P. van den Heuvel, Richard F. Betzel & Olaf Sporns. 2013. Exploring the morphospace of communication efficiency in complex networks. *PLoS One* 8, e58070.

Gottesman, Irving I. & Todd D. Gould. 2003. The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry* 160, 636–645.

Gould, Todd D. & Irving I. Gottesman. 2006. Psychiatric endophenotypes and the

development of valid animal models. *Genes, Brain, and Behavior* 5, 113–119.
Hudson, Richard A. 1990. *English Word Grammar*. Oxford: Blackwell.
Longobardi, Giuseppe. 2000. The structure of DPs: Some principles, parameters, and problems. In Mark Baltin & Chris Collins (eds.), *The Handbook of Contemporary Syntactic Theory*, 562–601. London: Blackwell.
Martin, Gary E., Molly Losh, Bruno Estigarribia, John Sideris & Joanne Roberts 2013. Longitudinal profiles of expressive vocabulary, syntax and pragmatic language in boys with fragile X syndrome or Down syndrome. *International Journal of Language and Communication Disorders* 48, 432–443.
McGhee, George R. 2006. *The Geometry of Evolution: Adaptive Landscapes and Theoretical Morphospaces.* Cambridge: Cambridge University Press.
Neuhoff, Nina, Jennifer Bruder, Jürgen Bartling, Andreas Warnke, Helmut Remschmidt, Bertram Müller-Myhsok & Gerd Schulte-Körne. 2012. Evidence for the late MMN as a neurophysiological endophenotype for dyslexia. *PLoS One* 7, e34909.
Palmer, Margarita. 2014. L'anàlisi del llenguatge en afectats per la síndrome de Williams per mitjà de xarxes complexes: Un estudi pilot. [A complex network approach to the study of language abilities in people with Williams syndrome: A pilot study]. Palma de Mallorca: University of the Baleric Islands MA thesis.
Peter, Beate, Mark Matsushita & Wendy H. Raskind. 2012. Motor sequencing deficit as an endophenotype of speech sound disorder: A genome-wide linkage analysis in a multigenerational family. *Psychiatric Genetics* 22, 226–234.
Pietroski, Paul M. 2005. *Events and Semantic Architecture*. Oxford: Oxford University Press.
Stein, Catherine M., Barbara Truitt, Fenghua Deng, Allison Avrich Ciesla, Feiyou Qiu, Peronne Joseph, Rekha Raghavendra, Jeremy Fondran, Robert P. Igo, Jr., Jessica Tag, Lisa Freebairn, H. Gerry Taylor, Barbara A. Lewis & Sudha K Iyengar. 2014. Association between AVPR1A, DRD2, and ASPM and endophenotypes of communication disorders. *Psychiatric Genetics* 24, 191–200.

Lluís Barceló-Coblijn
University of the Balearic Islands
Psychology
Carretera de Valldemossa km 7,5
07122 Palma de Mallorca
Spain
lluis.barcelo@uib.cat

Antonio Benítez-Burraco
University of Huelva
Spanish Philology and Its Didactics
Avda. de las Fuerzas Armadas s/n
21071 Huelva
Spain
antonio.benitez@dfesp.uhu.es

Aritz Irurtzun
CNRS, IKER
UMR 5478
Paul Bert Plaza 15
64100 Bayonne
France
aritz.irurtzun@iker.cnrs.fr