

How children develop their ability to combine words: a network-based approach

Adaptive Behavior
2019, Vol. 27(5) 307–330
© The Author(s) 2019



Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/1059712319847993
journals.sagepub.com/home/adb



Lluís Barceló-Coblijn¹ , Aritz Irurtzun², Cristina Real Puigdollers³,
Emilio López-Navarro¹ and Antoni Gomila¹

Abstract

A new way to study and represent early syntactic development is introduced that offers a promising avenue to improve on standard cumulative approaches to language learning. The analysis is inspired by complex network theory and explores an important issue in psycholinguistics: how children combine words syntactically. To this end, the article exploits the longitudinal data of multiple individual studies stored and publicly available in the CHILDES database. This analysis proves useful in two regards: in coincidence with previous approaches to syntactic development, it reveals a similar linear–nonlinear pattern of syntactic development (a combination of linear periods interrupted by abrupt transitions) in typical children regardless the language they are acquiring. It also provides a straightforward objective measure of syntactic impairment in atypically developing children. A striking difference between typical and atypical children lies in the connectivity of functional words, which suddenly become hubs but only in typical development. Our technique offers a formal measure of impairment with respect to typical syntactic development, which goes beyond the standard notion of retardation in development, for the Down syndrome condition.

Keywords

Language acquisition, analysis of corpora, complex networks, syntax, Down syndrome

Handling Editor: Alberto Antonioni, University College London, UK

1. Introduction

In this study, we explore the possibility of establishing different language development endophenotypes based on a network analysis of spontaneous speech corpora (Barceló-Coblijn, Benítez-Burraco, & Irurtzun, 2015). To do this, we establish a comparison between typically developed and atypically developed children. We compare typical language samples from four Romance languages (Italian, French, Catalan, Spanish), three Germanic languages (English, German, Dutch), and a language isolate (Basque), uncovering a common schema in development despite the typological differences between those languages. Next, we compare the patterns of these typical populations with the patterns issued from the analysis of the speech of atypical populations (Dutch and English-speaking people with Down syndrome (DS)) observing clearly divergent developmental paths.

The structure of the article is the following one: in the remaining of section 1, we frame our study within

the field of language development and clinical linguistics. Section 2 very briefly introduces complex networks, and section 3 presents previous applications of complex network analyses to language and speech. Section 4 presents the scope and goals of our study, and section 5 the methods we propose for the analysis of spontaneous speech. Section 6 is devoted to the results and section 7 to their discussion. Finally, in section 8 we discuss the consequences, merits, and limits of our study, and comment on the future research venues that it opens.

¹University of the Balearic Islands, Palma, Spain

²Centre National de la Recherche Scientifique (CNRS), Centre IKER (UMR 5478), Bayonne, France

³Universitat Autònoma de Barcelona, Barcelona, Spain

Corresponding author:

Lluís Barceló-Coblijn, University of the Balearic Islands, cra. de Valldemossa, km 7.5, 07122 Palma, Spain.
Email: lluis.barcelo@uib.cat

1.1. Differing views on language development

The development of new tools and approaches for language analysis has been an important target of many branches in cognitive science. One of the pitfalls that applied linguistics faces concerns the definition of the target of study itself, since it is common to find radically opposed views regarding some aspects of human cognition. Some issues are traditionally discussed dichotomically: does the ability for language develop linearly or non-linearly? How do biological parameters and the environment surrounding the speaker contribute in language development within the individual (cf. the “nature vs. nurture” debate)? Especially, contentious was Chomsky’s claim on syntactic development that domain-general learning is unable to account for language acquisition (cf. *i.a.* Chomsky, 1975), and that some innate knowledge of grammar has to be hypothesized to account for this process. Authors like Mehler (1974, et seq.) have even hypothesized that learning a language amounts to “unlearning” others (see also Baker, 2005; Lenneberg, 1967; Piattelli-Palmarini, 1989, among many others).

The opposite view to biological-centered approaches (or bio-approaches) concentrates on the *use* of language: how speakers use language, when, and in which contexts. A common assumption in *usage-based* approaches to language acquisition has been that syntactic development is a linear, cumulative process (“gradual” and “piecemeal,” in Tomasello’s, 2003, terms). If the biological-centered viewpoint has used the switchboard metaphor, according to which children select a parameter value from a limited set of possibilities (ideally a binary parameter, “on/off,” say, head-first vs head-last), usage-based approaches consider that syntactic categories and structures are mastered “one after another,” so that it should be possible to index the moment a particular category and/or rule is acquired. While bio-approaches consider that children transfer categorial information from one word to another, usage-based approaches are reluctant (if not contrary) to this kind of learning.

1.2. Language development and clinical linguistics

In consonance with the current black-or-white scenario, research in clinical linguistics is often approached dichotomically as well. For example, although there is a research view that considers that linguistic skills in pathologies like DS do progress in a divergent way over the life course, another view held for a long time that DS development was a slowed down process (this view is still held nowadays by some authors). Thus, individuals affected by some pathologies are taken to lag behind their typical peers (often under the term *delay*, for example, Abbedutto, Warren, & Connors, 2007; Galeote, Soto, Sebastián, Checa, & Sánchez-Palacios,

2014; Mundy, Kasari, Sigman, & Ruskin, 1995; Roberts, Price, & Malikin, 2007).

In many works, individuals affected by a pathological condition are compared to controls not in terms of their biological age, but in terms of their alleged *Mental Age* (MA). Nevertheless, MA is a measure that can be calculated in several different ways and is not exempt of problems and discussion as Brock (2007) pointed out. That is, atypically developing (AD) individuals are often compared to typically developing (TD) individuals regarding the scores obtained in the particular MA test selected by the scholars. Those problems and different ways to test the MA of an individual have highlighted the necessity for looking for other tools whose measures can complement the information that a researcher is already able to obtain from more traditional systems (such as tests, tasks, recordings, scans, etc.). If one is interested in knowing the syntactic ability of a speaker, one will see that here too the systems used to calculate that ability are indeed many (and sometimes they yield contradictory results).

A basic aspect of the available measures is *how* they were conceived, that is, whether the creators of that particular measure/device conceived a priori that the language ability is something that develops linearly or not (see Müller and Ball, 2013, for an overview). Research methods can be divided in language production and comprehension methods. Among the former, for example, there are measures such as the *Assessment of Intelligibility of Dysarthric Speech* (Yorkston & Beukelman, 1981), the *Mean Length of Utterance* (MLU, Rice et al., 2010), or the *Index of Productive Syntax* (IPS; Scarborough, 1990), which are used to track syntactic development. These are implicitly committed to a linear view of the ontogeny of syntax. These measures, though, are viewed as imperfect proxies for syntactic development due to the well-known problems of the variability in morphological structure (words may comprise just one or several morphemes), and the critical role of functional, closed-class words for mastering syntax (Leonard & Finneran, 2003; Rollins, Snow, & Willett, 1996). Informative as it is, a crucial weakness of the MLU measure is that sentences of an equal number of words may be very different from a syntactic point of view (they may imply radically different syntactic structures involving hypotaxis vs parataxis, different sorts of dependencies and “displacement” strategies, etc.). As a consequence, these linear measures turn out to be of limited usefulness in characterizing atypical development. Another tool often used nowadays by neurologists (among other researchers) is the *Boston Naming Test* (Kaplan, Goodglass, & Weintraub, 1983), although the accuracy of this test is still not completely satisfactory since sometimes individuals reach a score that would suggest the patient has no problem at all, even when the individual has been diagnosed on the contrary. Other ways

employed to assess the syntactic proficiency of AD individuals have been the *Words and Sentences Scale* of the MacArthur–Bates Communicative Development Inventory (MCIDI), by asking the parents about what the child is able to perform (Fenson et al., 1993), or making the participants to repeat a series of sentences. In general, results of works trying to compare different clinical profiles rely on this type of tools. In the case of asking the parents, the researcher needs to trust what they think the AD child is able to perform, and the results of sentence repetition tasks can vastly vary between pathologies (e.g. people affected by Williams syndrome seem to perform orally better than individuals with DS). Comprehension techniques such as the picture-matching task do not show in and of themselves whether the individual has captured all the intricacies of a specific syntactic construct, from lexical syntax to sentential syntax (cf. Roeper, 2007; Schmitt & Miller, 2010, among others). The above-mentioned IPS seems to address some of the weaknesses in the MLU. The IPS score is obtained from a corpus of 100 utterances, within which 56 specific language structures must be found. IPS has been implemented into a new tool (Sagae, Davis, Lavie, MacWhinney, & Winter, 2010) by combining it with Natural Language Processing techniques. In particular, the new IPS approach introduces dependency structures to identify grammatical relationships (Hudson, 1990). In general, these methodologies have sought to capture in one way or another the level of syntactic complexity reached by the speaker.

In the following section, we present a new approach, focused on an Evo-Devo perspective, which is able to capture aspects of the complexity of the speaker's linguistic system which were unapproachable with previous techniques. This approach is born from the combination of studies on complexity theory and network science on one hand and linguistic (syntactic) theory on the other hand, and its results are interpreted under the light of an Evo-Devo approach to language development.

1.2.1. An Evo-Devo approach to clinical linguistics. Here, we want to propose a new technique seeking to bypass the dichotomous view of language development. This technique is based on the emerging viewpoint which builds upon Evo-Devo approaches and conceptual work developed in biology (e.g. West-Eberhard, 2003). The Evo-Devo perspective avoids the debate between two apparently opposed viewpoints, the first considering ontogeny focused on genes only and the counter viewpoint considering language acquisition to be something that has nothing to do with biology. Between genes and diseases there exist cognitive, neuroanatomical, neurophysiological, endocrine, or biochemical quantifiable components that can define endophenotypes

(Gottesman & Gould, 2003). Accordingly, this new way to observe language development takes into account that the development of an organism tends to follow a particular developmental path. In other words, the psychological ability for language would have been somehow *canalized* during evolution (Balari & Lorenzo, 2013; Benítez-Burraco & Boeckx, 2014) and clinical linguistics should take into account the many levels at which the development of an organism can be intervened or affected (Benítez-Burraco & Murphy, 2016). This third way recognizes the biological predisposition of humans to develop language in typical conditions and the important role of the environment surrounding the individual. *De-canalization* (Gibson, 2009) on the other way could explain “the high prevalence of complex diseases (including language disorders) among human populations” (Benítez-Burraco, 2016, p. 264). As we said, between genes and pathologies there exist cognitive, neuroanatomical, neurophysiological, endocrine, or biochemical quantifiable components that can define endophenotypes. Recently, Barceló-Coblijn et al. (2015) have put forward an endophenotype for language disorders which integrates up to 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. Thus, our reflections and conclusions on the results obtained in the present works have to be envisaged from the Evo-Devo viewpoint, as syntactic networks are considered here proxies of syntactic endophenotypes.

Before starting the linguistic analysis by means of networks, and to better understand a linguistic approach that combines syntactic analyses and complex networks, it is necessary to explain what the concept of complexity is in network science and how it can be integrated in psycholinguistics.

2. Computational approaches to linguistic complexity

The term *complexity*, as several others in science, is not easy to define. John Locke in his *Essay Concerning Human Understanding* (ch. XII) says that “*Ideas* thus made up of several simple ones put together, I call *Complex*”; and paraphrasing the physicist Neil Fraser Johnson (2007), a neutral, acceptable explanation could be that a complex behavior emerges from the interaction of a collection of objects (but see Mitchell, 2009, ch. 7, for discussion). We consider that complex systems theory offers a promising new approach to language, as it does in many other areas of scientific inquiry (Watts & Strogatz, 1998). The most important tenet of *network science* (Newman, 2010) is that two networks could contain the same number of nodes (or vertices), but a

different number of edges (or connections) and therefore, both networks may show a completely different behavior and radically different scores regarding network complexity measures. In other words, two systems having the same number of elements behave differently due to the different relationships holding between their elements. An interesting aspect of some networks is the presence of *hubs*. Hubs are highly connected nodes, usually few in number within the network. A network of n nodes can develop the characteristics of a *small-world* if it has a high *cluster coefficient* (indicating how many neighbors of a given node are also neighbors of each other) and a low *average shortest path length* (a small number of links along the shortest paths that must be traversed when navigating the shortest paths in the network). In a nutshell, the number of nodes provides an important descriptor of the size of a network, but even more important is the particular connectivity pattern that the network has. Importantly, the network-based viewpoint is highly versatile, as it can be applied to very different examples, ranging from urban networks (Neal, 2018), to biological networks (Arita, 2005), or to linguistic networks (e.g. among many others, Solé, Corominas-Murtra, Valverde, & Steels, 2010).

3. Networks and language

The application of networks to linguistic studies is not brand new and has been applied to semantics, phonology, or to the analysis of texts. At the end of this section, we will show that syntax and the development of the syntactic ability can also be approached with network analyses. Phonology is, probably, the area of linguistics that has gotten most attention in network studies. One such approach is Vitevitch (2008), where the target was to capture relevant aspects of the mental lexicon. A key concept in that work was *neighborhood density*, which accounts for the number of words that sound as a selected word (thus a word A is a “phonological neighbour” of B, if a single phonological change to A converts A into B). In these regards, the similarities between languages as different as Spanish, Mandarin, Hawaiian, and Basque point into the direction of commonly shared psycholinguistic mechanisms used in the architecture of the mental lexicon (Arbesman, Strogatz, & Vitevitch, 2010). Besides, it has to be noted that the usefulness of this type of approach is also being expanded to clinical studies as well (Kenett et al., 2013; Vitevitch & Castro, 2015).

However, the application of the *network science* is not limited to phonology. Semantic networks have also been studied in the last years, yielding interesting results and hypotheses. For example, Hills, Mouene, Mouene, Sheya, and Smith (2009) tested three different models of word acquisition, based on *preferential attachment* (according to which “rich become richer,” that is, new

nodes tend to connect to highly connected nodes¹), *preferential acquisition* (early learned words seem to be highly connected because they stand out in the learning environment and are more noticeable), and the *lure of associates* (“unknown words may be highlighted by known words to which they are related”). In that work semantic networks were built on the basis of 130 nouns. Thus, nodes represented nouns and links represented relationships between nouns (e.g. both *ball* and *apple* share the feature *round*). Interestingly, the authors covered a sensitive period for language development, from 16 to 30 months. Their results called into question that the principle of preferential attachment is the key of language development (and their results also highlighted the relevance of the connectivity of a word in the environment).

Besides, Beckage, Smith, and Hills (2011) investigated the growth of a semantic network by means of word co-occurrence. Their study analyzed both typical and late talkers, and their results support the intuition that small-worldness is related to the child’s lexical development. Until then, most works in semantics were based on English, German, Dutch, and Spanish, but Kenett, Kenett, Ben-Jacob, and Faust (2011) took a step further working on Hebrew—a Semitic language famous for its morphological roots and patterns (e.g. Ravid & Schiff, 2006)—and investigated the networks both at a local and global level and how these two levels interact. Thus, at the local level, these authors observed that the network was organized into sub-cliques, based on semantic categories.

Recent work on semantics and networks has paid attention to the ontogeny of the mental lexicon (Stella, Beckage, Brede, & De Domenico, 2018). These authors have taken into account several properties of lexicon (semantics, taxonomy, and phonology), showing that these properties indeed interact detecting a cluster of highly frequent words. Moreover, Stella and collaborators conclude that emergence of this core of the lexicon occurs abruptly during ontogeny (around age 7).

On another front, it has already been shown that language corpora exhibit a complex system behavior, and that they can be represented in terms of a graph of word interactions (Corominas-Murtra, Valverde & Solé, 2009; Ferrer-i-Cancho & Solé, 2001).² In these approaches, however, words are not connected by syntactic or semantic relations between them, but taking into account the “distance” to each other within the sentence. The maximal distance between words—typically between 2 and 3—is assessed by the researcher. These calculations have been used mainly in English. Nevertheless, if applied to other languages such as German or Dutch, which split many verbs into two components, the result could be that these components apparently have no link or relation. Consider, for instance, the following sentences: (1) is an example of

split of the verb in present perfect in German and (2) is an example of a phrasal verb in Dutch:

- (1) Heute *bin* ich mit meinem Bruder um zehn Uhr mit dem Bus nach Hause *gefahren*.
 Today am_[AUX] I with my brother at ten hour with the bus to house ridden_[PAST PARTICIPLE]
 Today I *have gone* home with my brother at ten o'clock by bus.
- (2) Ik *sta* elke dag om seven uur *op*.
 I stay_[VERB] every day at seven hour up_[PARTICLE].
 I *get up* every day at seven o'clock.

Notice that, even if “mit dem Bus” is removed in (1), the distance between *bin* and *gefahren* still is much longer than 3. This way to consider connectivity is directly based upon the *linearity* a printed text imposes. By linearity, linguists understand that—in oral languages—one word comes after the other, due to the constraints of the oral channel. In this case, one counts the “positions” and hence the distance between words. This type of approach has two sets of main problems: on one hand, it is built on the idea of counting word positions but as has been argued repeatedly, natural language does not seem to count (see *i.a.* Chomsky, 1975; Moro, 2008; Newmeyer, 2005), and on the other hand, it is blind to a core property of natural language which is its possibility to establish long-distance syntactic relations (such as the relation between the auxiliary *bin* and the lexical verb *gefahren* in (1) or the relation between the particle and the verb of the Dutch particle verb *opstaan* in (2), or the relation between the interrogative direct object “what” and its corresponding verb “eat” in an English sentence such as “What do you normally like to eat?”). However, although it is true that co-occurrence links are not based on true relations but on distance between words, these networks have anyway captured some aspects that have been also found in syntactic networks, like the central role played by functional items as hubs (Barceló-Coblijn, Corominas-Murtra, & Gomila, 2012; Corominas-Murtra et al., 2009). Network-based approaches have evolved to be able to capture these particularities, as we will see next.

3.1. Longitudinal syntactic networks

We have already mentioned longitudinal network-based approaches to language development which are focused on semantics. But there are also some studies arguing that the same can be done in syntax. An outstanding example is van Geert’s approach to language development based on dynamic theory (Ruhland & van Geert, 1998; van Geert, 1994). Van Geert’s approach to longitudinal corpora shows that the language capacity develops non-linearly. Also in a series of works Ninio (2006) has uncovered non-linear effects in syntactic development. Likewise, Corominas-Murtra et al.

(2009) developed a first approximation to combining syntactically hand-annotated sentences with network representations and then tracking the longitudinal stages in linguistic development. In that work, two longitudinal English corpora were syntactically analyzed by hand and then the analyzed sentences of each file were represented by a network program, where syntactic dependency relations constituted the basis of the links or edges in the networks. First, during the syntactic analysis, the linguist syntactically analyzed the sentence deciding which word was connected to which word, using DGA annotator (Popescu, 2003), a program based on Dependency Grammar (Hudson, 1990). Both corpora had 10 chronologically ordered files, containing spontaneous speech of children. This way, their objective was to follow the child’s linguistic development. Corominas-Murtra and collaborators saw that in both cases the ability to combine words represented by the network developed similarly. In that work, two different stages were detected (a phase characterized by tree-like networks and another characterized by scale-free networks with the small-world characteristic), which were separated by an abrupt transition. That transition coincided in time with the emergence of hubs, which interestingly corresponded to *functional words*, such as “a,” “the,” “that,” and “this” and modal verbs. Such words count with a strong morpho-syntactic load, though they are devoid of a rich denotational semantics. These results are in consonance with results from dynamic approaches, showing the relevance and non-linear emergence of functional words in general (Ruhland & van Geert, 1998), the particular use of prepositions (van Dijk & van Geert, 2007), or determiners (Bassano et al., 2011).

Very similar results were found by Barceló-Coblijn et al. (2012), who analyzed three longitudinal corpora of Spanish, Dutch, and German. In that study, adopting the perspective of dynamic systems, particular attention was paid to the emergence of determiners as hubs in all three languages, showing that morphological diversity needs time to be mastered (see, in particular, Table 5 on the analysis of German determiners). Another relevant difference was that, while Corominas-Murtra et al. (2009) included corpora with 10 files, Barceló-Coblijn et al. (2012) analyzed corpora including up to 17 files. The inclusion of more segments of the child’s first months of life apparently uncovered a third phase. Interestingly, all three children underwent the same three stadia in very similar periods to those found in French by the Labrell et al.’s (2014) study based on dynamic theory.

In a nutshell, complex networks offer us a very useful way for the representation of the ontogeny of this process; words (syntactic items) are represented by nodes, while edges between nodes represent relations of syntactic dependencies. In this sense, the analysis that

we will present in the following is similar to that of Sagae et al. (2010), since both use dependency relationships to syntactically analyze a sentence.

The present article also adopts the perspective of considering the cognitive capacity of language as a dynamic, developing system. This study seeks to contribute to this trend of analysis by enlarging the analyzed corpora to include several Romance and isolate languages that were unexplored with any network technique so far (thus allowing typological (crosslinguistic) comparisons), and also by assessing the suitability of the network technique in clinical studies by analyzing the patterns emerging from atypical populations (DS).

4. Comparing different languages and TD and AD populations

Van Geert (2004) made an appealing suggestion of applying the dynamic systems perspective of language development to Specific Language Impairment or SLI. In this article, like in Barceló-Coblijn et al. (2012), we also conceive language ontogeny from the perspective of dynamic systems. Specifically, we propose a new implementation of the network analysis with a comparative analysis of languages from the Romance family (Italian, French, Catalan, and Spanish), the Germanic family (English, Dutch, and German), and a language isolate (Basque). A second analysis will concern two groups of people with DS, 12 English speakers and 20 Dutch speakers. By comparing between different speakers of different languages it should be possible to (1) see whether the type of language has an influence on the ability to combine words (similarly to Bassano et al., 2011). By comparing between different groups of speakers (TD vs AD), it should be possible to (2) reinforce or to discard potential hypotheses about language development like those presented in previous sections. In van Geert's (2004) words, if typical development follows some general principles (like discontinuity, U-shape, S-shape, particular network structures, etc.), there is no reason that they should not be applied to the development of impairment and disorder, as van Geert (2004) did with SLI.

Such comparison also shows the welcome consequences of the application of the network analysis in the crosslinguistic study of language development in both typical and atypical populations. Typically, network-based approaches tend to compare control networks in size, so that it is possible to compare different systems made of the same number of elements. However, in the case of language ontogeny, such an option loses part of the ecological value of the original source. Analyzing language ontogeny with real data (spontaneous speech) shows an important difference: each period of the child's lifetime that is analyzed produces a different graph with different nodes (people say

different words in different sessions), but also a different number of words (child's lexicon grows, and her ability to combine words becomes more developed). Here, networks are not always the same in number, and hence the approach is rather different, since what we observe is not one network in comparison to another made of the same elements. The focus of our attention is how much "linguistic signal" the speaker produces in a limited amount of time (a session of spontaneous speech of a limited number of minutes), and what type of complexity can be observed in its underlying linguistic system. If all TD children produce more or less the same linguistic pattern, it will be possible to say that this might be the typical linguistic pattern of development. However, if a speaker greatly differs from the rest of her pairs, this opens the possibility for considering her linguistic ability as atypically developed. Hence, in the present approach we do not put the same quantity of text from the transcriptions (we cannot), since the set of AD children differ from TD children precisely in the quantity and quality of speech. Note that including, say, only 100 lines from a recorded session in both typical and atypical networks would mask the fact that perhaps atypically developed speakers need much more time until they produce the same quantity of words and sentences than TD children. It will be shown that, in general, and coinciding with previous results in the literature, DS speakers produce less words. Moreover, we will show that DS speakers produce less syntactic connections and, hence, produce atypical syntactic networks, in spite of having a similar recording session length (other differences found relate to the nature of lexical hubs).

In the first part of our study, we analyze the syntactic development of three children acquiring a (first) Romance language: Italian, French (Demuth & Tremblay, 2008; Tonelli, Dressler, & Romano, 1995), and Catalan (from the corpus Serra-Solé), and then we compare them to the German, Dutch, and Spanish corpora in Barceló-Coblijn et al. (2012). Moreover, to overcome a possible Indo-European or typological bias we also analyze two corpora of Basque (Soto Valle & Alonso, 2015)—a language isolate. This is the first time that such a language is analyzed with this technique. The inclusion of Basque allows us to overcome two potential issues: first, that the results we obtain could be biased by the phylogenetic relationships between the languages under study (Romance and Germanic languages belong to the same Indo-European family) and second, that this technique is only useful for non-agglutinative languages. In general, although there exist important differences, Romance and Germanic languages share a range of typological properties; Basque, however, is very different broadly speaking (see, for instance, the *World Atlas of Language Structures* (WALS) database).³ Basque is unique within the group because it is the only language that has a basic SOV

word order and agglutinative properties (see *i.a.* Laka, 1996; Trask, 1998). It displays a complex morphological system that exploits a lexical process called “agglutination” whereby words are formed by joining affix morphemes to the stem, where to a high degree each morpheme contributes one meaning and vice-versa; that is, each affix is a bound morpheme for one unit of meaning. For example, in Basque “*etxe*” (*house*) becomes by affixation “*etxea*” (*the house*), “*etxeak*” (*the houses*), “*etxeko*” (from *house/homemade*), and “*etxetik*” (from *the house/out of house*). In Basque it is possible to agglutinatively chain postpositions, obtaining clusters of suffixes and words with a complex morphology and a complex meaning, for example, “*etxe-ra-ko-ak*” (*etxe* + *ra* (allative) + *ko* (genitive locative) + *ak* (plural determiner); “those for the house”). Likewise, “*gizon*” (*man*) can participate in the generation of complex words like “*gizon-a-ren-a-gatik*” (*gizon* + *a* (determiner.singular) + *ren* (genitive) + *a* (determiner.singular) + *gatik* (because); “because of that of the man.”)⁴ Importantly, this typological fact makes word-based MLU-type measurements (MLU-w) non-optimal for crosslinguistic comparison, for example, the number or “words” that are required for the expression of the same sentence varies significantly from language to language. As a consequence, some researchers prefer the use of morphemes as the counting unit for MLU (see Ezeizabarrena & Garcia Fernandez, 2018, for discussion). In our approach, on the contrary, we decompose each sentence in each syntactic unit, separating adpositions from nominals and thus obtaining a much more accurate crosslinguistic comparison.

We will show that the results of our linguistic analyses of TD children acquiring Catalan, French, or Italian are fully convergent with previous works that took into account the perspective of dynamic systems and complexity. Despite procedural differences in the original composition of the corpora and the heterogeneity of the data, a robust common pattern seems to emerge, apparently characterized by three stages of linear development separated by two non-linear phase transitions. In particular, it is intriguing that around 28 months of age a small-world network emerges, with a ratio of word/edges close to 1:2. Such is the case in each corpus, regardless of the language analyzed (this fact could suggest that the child has reached a particular developmental threshold, and that the pattern of word combination here represented by a small-world network was not dependent of the type of language the child acquired). This type of network indicates in all cases the maturity of the cognitive capacity for combining words. It reveals a stage where the child has already discovered the unbounded combinatoriability of the syntactic items in her language via some sort of syntactic “reverse combinatoriality” (in the spirit of the

Fodorian “reverse compositionality” (see Fodor, 1998; Fodor & Lepore, 2001))⁵ whereby a learner who has mastered the availability of syntactic combinations such as $[\alpha^{\wedge}\beta]$, $[\alpha^{\wedge}\gamma]$, $[\alpha^{\wedge}\delta]$, and $[\varepsilon^{\wedge}\beta]$ and $[\varepsilon^{\wedge}\gamma]$ (where α and ε pertain to the same syntactic category (say, determiners), and β , γ , and δ to a different one (say, nouns)) generalizes in terms of their syntactic category and goes on to freely combining elements of the $\{\alpha, \varepsilon\}$ type and elements of the $\{\beta, \gamma, \delta\}$ type generating novel combinations such as $[\varepsilon^{\wedge}\delta]$ and so on. In a nutshell, the network reveals the cohesion level of the child’s linguistic system. In addition, mastering of syntactic combination is related to a proper acquisition of functional words: as we said, the most connected nodes of the network (the “hubs”) are typically functional words and the set appears abruptly at this developmental moment. In sum, our results suggest that the three levels, common in all longitudinal analyses, are a reflection of the developmental path of TD children.

5. Method

5.1. Typical corpora

In the first part of the present work, five sets of corpora of longitudinal studies of typical language acquisition have been analyzed: one for Catalan, one for French, one for Italian, and two for Basque. Corpora were obtained from the international CHILDES database (MacWhinney, 2000). Each corpus contains at least 17 files chronologically ordered. Each file corresponds to a different recording session, which typically lasted between 30 and 60 min of spontaneous linguistic interaction between the child and their parents/caregivers. Children’s age extends from 17 to 32 months of age. Age is typically indicated as follows: “years;months.-days,” for example 2;06.25 means 2 years, 6 months and 25 days of life. Age has then been represented in averaged calculation.

The Basque corpora we employed are partially different from the rest, since we could not find corpora extending exactly for the same age periods as the other ones; in fact, there were just these two cases available at the CHILDES database. Both Basque corpora start at 2.6 years of life (30 months). To see the connectivity of Basque postpositions, we syntactically decomposed complex morpho-phonological words and we analyzed how many times a postposition appears, taking into account that repetitions of the same combinations of lexical items are excluded. For example, if the target is the postposition *-tik* (equivalent to the preposition “from” in English) and it appears in the text 10 times with the noun *etxe* (“house”), this counts as only one appearance type (10 tokens of the same type), but if it also appears with *herri* (“town”), then it counts as two appearance types.

5.2. Atypical corpora

The set of clinical corpora are also obtained from CHILDES. The first set contains data of 12 subjects of different age, all of them diagnosed with DS and speaking English (Rondal, 1978a, 1978b). There are two files per individual, with the same age regarding “year” and “month,” and no information regarding the “day.” We decided to average the values of both files to get the final score. The second set contains data of 20 subjects of different ages affected by DS and speaking Dutch (Bol & Kuiken, 1990). However, these files correspond to one session; hence, just one network for each participant could be construed in this case. The group is heterogeneous regarding ages, ranging from 3 to 18 years. Nevertheless, the analysis of a multiplicity of cases (cross-sectional) of each type reveals a common pattern which we would like to argue is characteristic of the impairment in question (see section 6.2).

5.3. Procedure

A team of linguists analyzed each corpus sentence by sentence, following the basic guidelines of Dependency Grammar (Hudson, 1990). This kind of grammatical analysis is known and used in computational approaches to language because it provides essential information about the syntactic structure of the sentence, and the coded sentences are easy to export for further analyses to a network-analysis program. To carry out the syntactic analysis, data were prepared following the same procedure as in previous studies on syntactic networks (Figure 1; Barceló-Coblijn et al. 2012; Corominas-Murtra et al., 2009).

Corpora were downloaded from the CHILDES database, each file was converted into *.xml* so that it could be read by the syntactic-annotation program, by means of a Perl script. The script also filtered the file, pruning the data that were not produced by the child (e.g. comments of the transcriber, phonetic transcriptions, what the parents say, etc.). The new *.xml* file was then sent to *DGA annotator* (Popescu, 2003), a program specifically created for syntactic analysis by means of dependency relationships.

The analysis was carried out by a linguist manually line by line, sentence by sentence, determining the syntactic dependency relationships between words by means of a dependency arrow. Other criteria followed by the linguists’ team are gathered in a public protocol (Corominas-Murtra, 2007) applied in previous studies.

After analyzing all files, another Python script specifically created for this kind of studies turned them into *.txt* files, which is a format accepted by the network-analysis program *Cytoscape* (Shannon et al., 2003).

Once *Cytoscape* reads the data, it produces a graph. In this study at least 17 graphs were obtained for each typical child acquiring an Indo-European language.

The first of these graphs corresponds to a date of the child’s life of around 18 months, and the last graph corresponds to about the 30th month. This way, the evolution through time of the network—that represents how words are combined—can be readily visualized. In the case of Basque, two small corpora (representing two children) of five files each have been analyzed.

One graph was obtained for each session (i.e. 17 graphs per child, 5 in the Basque cases). Each word in a graph corresponds to a node in the network, while edges reproduce the syntactic dependencies between words. It is important to keep in mind that the dependency arrow reproduced by *Cytoscape* respects the direction of the arrow created by the *DGA annotator*. *Cytoscape* also allowed to visualize the arrow point. However, in spite of the careful linguistic analysis, information about the kind of syntactic relationships could not be integrated. Therefore, since it was not possible to see the kind of link between nodes, it was decided to be cautious and conservative and the analysis of edges was made as undirected edges.

We would like to stress that our approach is not frequentist, but a different, complementary way of observing linguistic data. In a graph, one can see how many *different* words and how many *different* dependency relationships have been employed by the child. This fact implies that repetitions—of either lexical items or dependency relationships—are not taken into account in this approach. If the child repeats “Pluto barks” 10 times, it counts as only one *type of dependency relation*: one node for “Pluto,” one node for “barks,” and a dependency relationship from “Pluto” to “barks.” It is the breath of vocabulary and the word combinations effectively produced which are used as a measure of linguistic development (i.e. we measure the knowledge that a speaker has obtained of the combinatorial properties of the items in her lexicon, as represented in her speech). Hence, it is the number of *different* linguistic expressions—and not the number of productions—the focus of this study. However, it is true that the high number of connections of the most connected words—the hubs—also reflects their condition as the most used words.

In a graph one can see several disconnected networks, depending on how well connected the words were. Each piece is called a “connected component.” The typical graph of a network analysis of a corpus contains one large network (sometimes two), several small networks with two or three nodes, and a number of isolated nodes (words that were uttered in isolation and hence were not connected with other words, such as, for example, “no” in a plain denial answer, or “cat” when pointing to a cat).

Cytoscape (or any other network program) also makes it possible to calculate multiple mathematical measures for each network; the most relevant for our purposes are

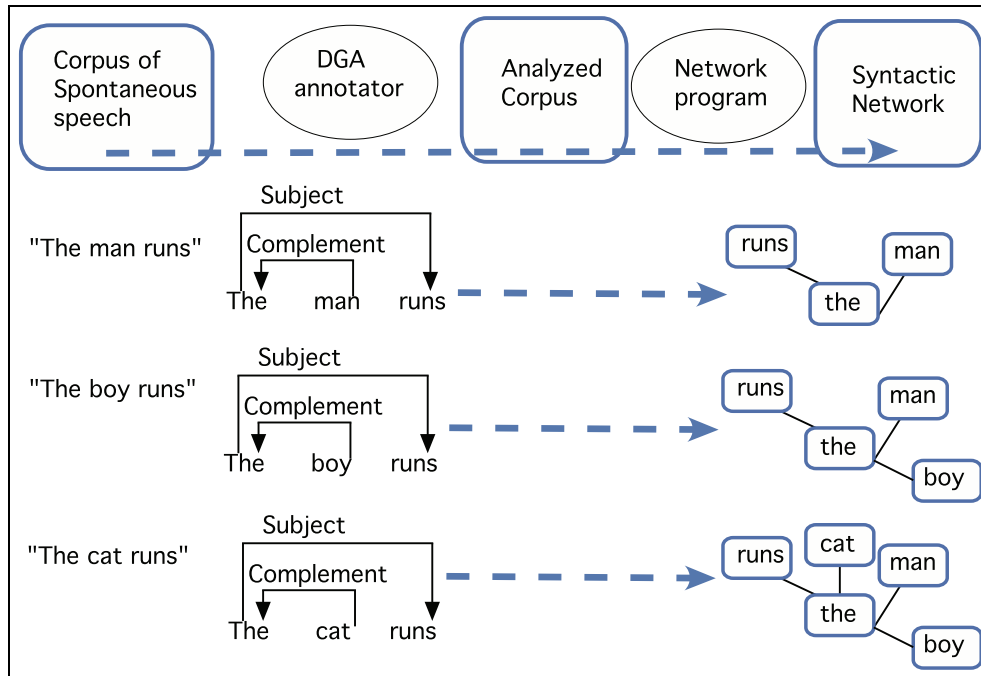


Figure 1. Steps and procedures in order to obtain a syntactic network from a corpus of spontaneous speech.

First, the corpus of spontaneous speech is created (on the basis of a spontaneous speech session; in this work, corpora have been downloaded from CHILDES database). Then each corpus is analyzed (word by word) with the DGA annotator, by establishing the syntactic relations between words. Once the corpus is analyzed, it is adapted by a script in order to be readable by the network program. Finally, the network program reads the annotated corpus and reflects the data as a network, where nodes are words and edges are syntactic relations.

1. the relative size of the largest connected component (or LCC), which indicates the size of the largest component regarding the whole graph;
2. the average clustering coefficient (C), which indicates the extent to which two connected nodes are also connected to a third one, creating a triangle (this is a real number between 0 and 1 (when there is no clustering the value is 0, while 1 indicates maximal clustering));
3. the average shortest path length (L), which indicates the mean number of intermediate nodes that connect any two of them; and finally
4. the average degree of the LCC ($\langle k \rangle$), which refers to the average number of connections per node; in our study, it indicates the average number of dependency relations for each particular word.

6. Results

6.1. Typical development

The study of TD corpora by analyzing longitudinal corpora of different languages aims to answer the following questions: (1) are there differences between German and Romance languages? and (2) what would happen in the case of typologically different languages?

Once analyzed and compared, the three new corpora (Catalan, Italian, and French) show a similar developmental pattern (Table 1 and Figure 2) to the one

observed in German, Dutch, and Spanish (Barceló-Coblijn et al. 2012). Thus, so far, all TD children exhibit the same pattern of syntactic development, regardless of the particular language they are learning. This pattern is apparently characterized by three different phases:

1. In the first phase, the networks emerging from their speech have a *tree-like* structure, characterized by a higher number of nodes (words) than edges (word combinations), and which appears with a relatively small LCC;
2. In the second phase, the child's pattern is characterized by networks in which edges overcome the number of words and, as expected, some hubs appear (these are "promiscuous" words that combine with several others, thus providing cohesion to the network, and increasing the *relative size* of the LCC);
3. In the third phase, a final phase, the LCC almost equals the whole graph. The characteristic path length decreases and a high clustering appears, showing a strong cohesion in the network. An interesting aspect is that at this phase the tendency is evolving to a ratio of nodes (words)/edges (syntactic links) of 1:2, that is, two connections for each word.

Interestingly, that final topology is reached more or less around the same age, 800 days of life, in spite of the heterogeneity of the original data in terms of

Table 1. Values of the indicators extracted from the syntactic networks from the TD corpora: mean path length (L), average degree of the giant connected component or GCC $\langle k \rangle$, clustering coefficient (C), relative size of the GCC (R size), number of nodes of the GCC, number of edges of the GCC, and ratio nodes/edges.

Age days	L	$\langle k \rangle$	C	R size	Nodes GCC	Edges GCC	Ratio nodes/edges
Catalan							
593	1	1	0	0.25	2	1	2
613	0	0	0	0.037037037	1	0	—
634	2.894	1.8	0	0.27027027	10	9	1.1111111111111111
640	1.364	1.5	0	0.19047619	4	3	1.3333333333333333
678	2.626	2	0	0.20289855	14	14	1
712	1.893	1.6	0.091	0.166666667	6	5	1.2
785	1.821	2	0.161	0.25	7	7	1
799	4.145	2.44	0.022	0.43269231	45	55	0.8181818181818181
879	3.292	2.788	0.038	0.44594595	66	93	0.709677419354839
938	3.223	2.565	0.067	0.73015873	46	61	0.754098360655738
976	3.46	3.692	0.076	0.79194631	240	449	0.534521158129176
1022	3.78	3.176	0.045	0.83783784	187	298	0.62751677852349
1067	3.827	2.807	0.029	0.8071066	161	229	0.703056768558952
1127	5.111	2.133	0.006	0.61971831	45	48	0.9375
1265	4.207	2.883	0.011	0.76243094	145	209	0.69377990430622
1309	3.896	3.4	0.057	0.75	180	311	0.578778135048232
1405	3.647	3.827	0.038	0.78059072	185	356	0.519662921348315
French							
552	1	1	0	0.111111111	2	1	2
571	0	0	0	0	1	0	—
584.5	0	0	0	0	1	0	—
600.5	0	0	0	0	1	0	—
621	0	0	0	0	1	0	—
662.5	0	0	0	0	1	0	—
694	0	0	0	0	1	0	—
728.5	3.106	2.424	0.025	0.402439024390244	33	41	0.804878048780488
748	2.426	2	0.058	0.25	10	10	1
819	2.928	2.627	0.073	0.3777777777777778	51	69	0.739130434782609
832.5	3.28	2.769	0.046	0.456140350877193	52	77	0.675324675324675
846.5	3.46	2.571	0.038	0.563758389261745	84	110	0.763636363636364
861	3.208	2.867	0.072	0.535714285714286	60	89	0.674157303370787
875	3.33	2.831	0.073	0.550847457627119	65	95	0.68421052631579
895.5	3.238	3.25	0.077	0.6181818181818181	136	231	0.588744588744589
909.5	3.486	3.41	0.093	0.674033149171271	122	216	0.564814814814815
Italian							
522.5	1.4	1.5	0	0.042105263157895	4	4	1
536.5	1.977	1.778	0	0.1	9	8	1.125
551	1.45	1.6	0	0.027472527472528	5	4	1.25
571	3.421	2.0625	0	0.228571428571429	32	33	0.96969696969697
585.5	3.031	2.095	0.006	0.106598984771574	21	22	0.954545454545455
598.5	2.012	1.846	0	0.053719008264463	13	12	1.083333333333333
613	3.231	2.074	0.006	0.094405594405594	27	28	0.964285714285714
627	1.167	1.333	0	0.090909090909091	3	2	1.5
641.5	1.711	1.667	0	0.024	6	5	1.2
655.5	4.148	1.917	0	0.07843137254902	24	23	1.04347826086957
669.5	2.268	1.882	0	0.057046979865772	17	16	1.0625
683	4.885	2.08	0	0.152439024390244	50	52	0.961538461538462
697	4.932	2.078	0.016	0.184115523465704	51	53	0.962264150943396
717.5	4.881	2.174	0.009	0.347432024169184	115	126	0.912698412698413
732	5.13	2.4	0.009	0.421286031042129	190	228	0.833333333333333
746	5.346	2.312	0.015	0.495176848874598	154	180	0.855555555555556
759	4.578	2.486	0.016	0.615	245	314	0.780254777070064
773.5	4.145	2.764	0.027	0.715492957746479	254	355	0.715492957746479
789.5	4.446	2.426	0.019	0.622516556291391	94	114	0.824561403508772
794	4.036	2.905	0.022	0.705093833780161	263	383	0.68668407310705
825.5	4.692	2.525	0.004	0.705202312138728	244	311	0.784565916398714
838.5	4.477	2.788	0.03	0.77521613832853	269	379	0.70976253298153
852.5	3.955	2.994	0.034	0.772234273318872	356	538	0.661710037174721
867	4.266	2.719	0.022	0.744471744471744	303	413	0.73365617433414

(continued)

Table 1. Continued

Age days	<i>L</i>	$\langle k \rangle$	<i>C</i>	<i>R</i> size	Nodes GCC	Edges GCC	Ratio nodes/edges
Basque A							
939	6.022	2.065	0.002	0.442857142857143	62	65	0.953846153846154
952.5	2.667	1.882	0	0.129770992366412	17	16	1.0625
967.5	3.494	2	0	0.184873949579832	22	22	1
981	3.767	1.92	0	0.233644859813084	25	24	1.04166666666667
1015.5	3.866	3.029	0.082	0.73015873015873	138	220	0.627272727272727
1029.5	4.571	2.133	0.015	0.441176470588235	60	65	0.923076923076923
Basque M							
935	3.959	2.458	0.023	0.588652482269503	83	105	0.79047619047619
949.5	3.974	2.364	0.015	0.582781456953642	88	106	0.830188679245283
963.5	4.194	2	0.002	0.367647058823529	50	50	1
984	3.829	2.286	0.021	0.538461538461538	42	51	0.823529411764706
998	5.704	2.127	0.003	0.508064516129032	63	67	0.940298507462687
1025.5	4.665	2.452	0.023	0.67906976744186	146	179	0.815642458100559

TD: typically developing.

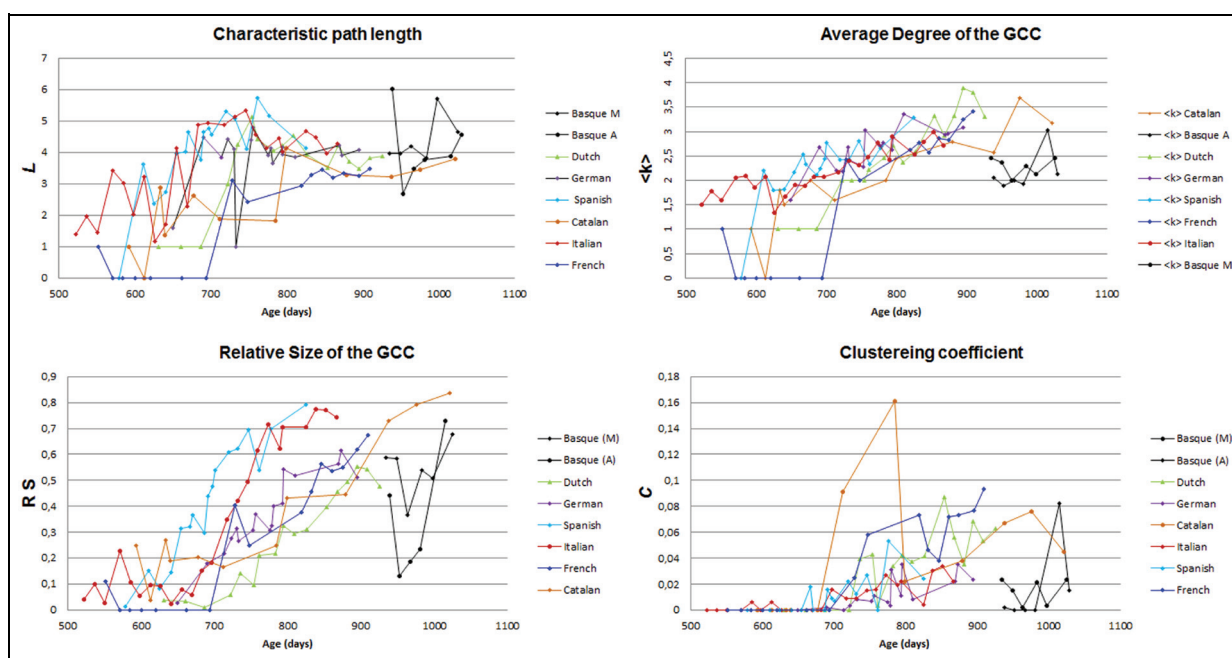


Figure 2. Summary of results for the relevant measures of the resulting networks for eight typically developing individuals. Each point displayed represents a single moment and not a mean score. We have integrated the five new corpora of this study with our previous results, in German, Dutch, and Spanish (Barceló-Coblijn, Corominas-Murtra, & Gomila, 2012). The same developmental pattern was found by Corominas-Murtra, Valverde, and Solé (2009), with two English learners.

sampling time, procedures, periodicity, and so on. The Basque cases (shorter in number of recordings) show a slight delay, in part, due to the lack of data from previous ages, but it could also be related to its more agglutinative character. Nevertheless, networks with a high clustering and a low path length are also obtained in this language.

Furthermore, the hubs that abruptly give rise to the small-world networks tend to belong to the same lexical class: these are the so-called functional words (mostly determiners; see Supplementary Material for a sample

of the evolution of hubs in Dutch and for a sample of the evolution of connectivity of determiners in Catalan), a result that coincides with the previous results. This pattern is found in all cases late and abruptly, coinciding with an important change in the network topology (see Table 1 in comparison to Table 5). Determiners (e.g. “the,” “a”) emerge as hubs in all languages so far analyzed following this late, abrupt transition pattern. In the previous work on networks in English (Corominas-Murtra et al., 2009), the hubs were “the,” “a,” “that,” and “it.” These nodes have at

Table 2. Affixes in Basque from the Maider corpus (each one is treated as an independent node).

Bare form(0)	-a(1)	-(a)k(2)	-k(3)	-(r)en(4)	-(r)(e)kin(5)	-n(6)	-ra(7)	-dik(8)	A, V(9)	-ik(10)	-ta(11)	-tu(12)	-ten/-tzen(13)	-tzeko(14)
73	19	6	10	8	1	14	1	1	39	1	3	2	5	1

Under each suffix form, a number within parenthesis refers to the class of suffix: (0) bare form with no morphological suffix; (1) singular article, absolutive; (2) plural article, absolutive; (3) ergative; (4) possessive; (5) comitative; (6) inessive; (7) allative; (8) ablative; (9) auxiliary verb; (10) suffix for the generation of manner adverbs, akin to English "-ly"; (11) result-denoting suffix (akin to English "-ed"); (12) participial; (13) imperfect; (14) future. The bottom row indicates the number of other nodes each of the suffixes combines with.

30 months of life between 30 and 40 links, almost 80 in the case of "a." Articles in Dutch, German, and Spanish also obtained similar numbers, producing around 30 and 40 syntactic dependency links. Some language-specific differences are always expected: the use of "it" in English does not correspond with the use of the equivalents in Dutch (*het*) or in German (*es*). Spanish and Catalan have three and two demonstratives respectively (some Catalan dialects show three demonstratives, but they are not represented in the sample), while Dutch and English have two and German has only one. However, German has a declension system, complicating a little bit the proper use of each form. The Romance languages here analyzed, Italian, French, and Catalan, follow the very same pattern, which is the pattern previously observed in Spanish.

One of the benefits of making a longitudinal analysis is that one can observe the evolution of the graph over a long period of time. As said, each graph contains several "components" (or networks of one or more nodes). Interestingly, in all languages analyzed until now (Catalan, Italian, French, Spanish, Basque, German, English, and Dutch) the number and size of components behaved in a very similar way: when the child began to speak (tree-like network stage), there were many small components of two and three nodes, and sometimes a couple of larger networks of four or five nodes (plus a large number of isolated nodes, representing unconnected words which were uttered in isolation). As the child grew and developed her linguistic ability to combine words, we observed the emergence of a LCC of a dozen or more nodes, sometimes followed by one or two small networks of five to seven nodes (plus the mini networks of one, two, or three nodes). When the child got to the second phase, small networks larger than four nodes seldom appeared. Instead, there was a clear and dominant LCC, and a number of networks of one and two nodes, followed by a smaller number of components with three nodes. Finally, when children got to the last stage, we observed an enormous LCC. In this stage, there was no independent network larger than three nodes, and even the presence of three-node networks is not guaranteed.

The second main question that concerned typical development and network analysis is related to the agglutinative features of languages like Basque, and whether this represents an obstacle for its application. From the two corpora of Basque that we analyzed, we obtained networks very similar to those of the non-agglutinative languages. We also performed a separate analysis of postpositions to see their connectivity pattern (Table 2).

Even in Basque, where determiners can be both free lexical units (e.g. *bat* "one/a," as in *etxe bat* "a house") or suffixes attached to the nominal stem (e.g. *-a* like in *etxea*, "the house"), these elements play the role of

Table 3. Number of links of the determiners and other functional words in English, from networks of subjects affected by DS.

Age, months	Subject	MLU	a	that	the	it	there	and	I
54	Abby	1.55	9	7	4	0	2.5	0	0
38	Baxter	1.04	2	0.5	1.5	1	1.5	0	0
56	Bob	1.72	4.5	2.5	4	2.5	0	1	2
57	Cheryl	1.79	5	2.5	3.5	3.5	1.5	2.5	11
84	Carrie	2.09	8.5	2.5	11.5	3.5	0	4	7.5
56	Dan	1.96	19	11	17	8.5	2.5	2.5	9.5
55	Janet	1.75	13	5	5	8	3	2.5	3.5
52	Kevin	1.12	3	2.5	0	0	0	0	0.5
146	Mat	2.93	19.5	4	30.5	2.5	1	24.5	13.5
100	Rhoda	2.70	13.5	12	17	5	2	10.5	8
37	Selma	1.11	0	0.5	0	0	0	0	0
94	Silvia	2.03	15	14	15	2.5	1	1.5	11.5

MLU: Mean Length of Utterance.

In this table the number of links of the typical hubs in English has been summarized. In the first study with two 30-month-old unimpaired children, hubs where “a”, “the”, “that,” and “it” and the scores reached between the 30 and 40 links. Instead, in DS cases of English the scores are much reduced. In each case, values come from averaging the scores of two networks (see Supplementary Material).

Table 4. Number of links of the determiners and other functional words in Dutch, from networks of subjects affected by DS.

Age, months	Individual	MLU	De the	Het the	het ^{pron} it	een a	Deze this	Dit this	Dat that	Die that	En and	Ik I
196	Anita	3.3	2	0	0	5	0	0	1	3	4	3
125	Antoniette	3.2	7	2	4	10	0	0	2	1	4	7
221	Bert	2.4	6	0	0	0	0	0	0	0	4	11
182	Chantal	2.8	7	0	0	0	1	1	2	4	2	2
181	Eline	2.5	12	2	0	7	1	0	0	1	25	3
219	Erik	1.8	2	1	0	3	0	0	0	1	3	1
162	Fiona	3.4	13	0	4	9	0	1	6	3	7	14
156	Iris	3.0	0	0	1	4	1	1	3	2	10	12
129	John	2.0	6	0	0	3	0	0	1	1	7	1
187	Katrijn	1.9	7	3	2	6	0	0	1	0	1	2
165	Lisa	1.8	6	2	0	10	0	0	1	2	2	2
202	Marcel	1.2	0	0	0	1	5	0	1	0	1	0
98	Martijn	4.7	9	3	3	6	0	0	3	10	3	1
227	Marius	3.3	8	2	1	7	2	1	1	5	10	8
205	Michael	3.0	17	0	2	8	0	0	1	1	3	2
100	Pieter	3.5	14	2	3	7	2	1	2	5	21	7
100	Ronnie	5.2	14	0	3	21	0	0	8	7	16	17
224	Sanne	3.0	4	0	2	2	0	0	4	2	6	10
200	Sven	2.4	12	2	0	7	3	0	3	1	6	2
53	Tineke	2.4	0	2	0	5	5	0	3	3	6	0

MLU: Mean Length of Utterance.

In a single conversation in Dutch, a 900-day-old individual typically produces about 30 links for each of the hubs that are determiners (cf. Barceló-Coblijn, Corominas-Murtra, & Gomila, 2012).

hubs in the small-world network. Table 2 displays the patterns of connectivity of different suffixes showing that determiners are much more connected than postpositions.

6.2. Atypical development

The third main question was whether there are differences between typically and atypically developed children. As expected, the answer is positive. One large group of 32 speakers with DS was analyzed. The group

is divided into two subgroups, depending on the language they spoke, Dutch or English. The networks that characterize the syntactic performance reached by DS children of older ages are notably dissimilar. Results are shown in Tables 3–5. English DS children became able to produce a very low number of words and to barely combine them (the speaker “Mat” stands out in comparison to the group, though he is still far from TD younger speakers; see Figure 3).

Then, the number of words increased over time, but very slowly, given the ages of the speakers. In the case

Table 5. Information about the Dutch and English DS speaker's networks.

	Months	Nodes	Edges	Ratio N/E	Clustering	Path length	Rel_Size	<k>
Dutch								
Ronnie	196	223	356	0.6264044944	0.042	4.039	0.8479087452	3.175
Fiona	125	150	236	0.6355932203	0.034	3.907	0.7075471698	3.107
Marius	221	130	194	0.6701030928	0.025	4.083	0.6914893617	2.969
Pieter	182	128	165	0.7757575758	0.027	4.843	0.6844919786	2.5625
Iris	181	119	158	0.753164557	0.034	4.814	0.650273224	2.639
Michael	219	116	138	0.8405797101	0.033	5.737	0.5979381443	2.379
Antoniette	162	98	133	0.7368421053	0.061	4.407	0.5414364641	2.673
Sanne	156	110	126	0.873015873	0.028	5.217	0.597826087	2.273
Sven	129	72	84	0.8571428571	0.002	4.68	0.5217391304	2.306
Eline	187	74	80	0.925	0.068	3.351	0.4431137725	2.135
Bert	165	62	77	0.8051948052	0.05	4.141	0.4881889764	2.484
Anita	202	57	61	0.9344262295	0	5.085	0.5643564356	2.14
John	98	47	56	0.8392857143	0.128	4.075	0.4234234234	2.298
Katrijn	227	54	56	0.9642857143	0.06	6.361	0.357615894	2.074
Martijn	205	45	56	0.8035714286	0.015	4.017	0.6	2.4
Tineke	100	51	56	0.9107142857	0	4.69	0.504950495	2.196
Chantal	100	41	42	0.9761904762	0.013	3.979	0.4270833333	2.049
Lisa	224	35	37	0.9459459459	0	3.376	0.4117647059	2.114
Erik	200	10	9	1.1111111111	0	2.533	0.0862068966	1.8
Marcel	53	6	5	1.2	0	1.867	0.0789473684	1.667
English								
Mat	146	199.5	263	0.7585551331	0.041	4.3885	0.772826087	2.6425
Silvia	94	116.5	178	0.654494382	0.025	4.3545	0.5907146524	2.4795
Dan	56	111.5	160	0.696875	0.0385	3.82	0.523478836	2.859
Rhoda	100	110.5	137	0.8065693431	0.0595	4.055	0.699741435	2.993
Cheryl	57	79	98.5	0.8020304569	0.051	4.184	0.5575624886	2.399
Carrie	84	70.5	77	0.9155844156	0.031	5.3375	0.4257745999	2.161
Abby	54	33.5	38.5	0.8701298701	0.0285	3.389	0.3578307728	2.3095
Janet	55	34.5	35.5	0.9718309859	0.0145	3.931	0.2365130219	2.0185
Bob	56	38.5	33.5	1.1492537313	0.0295	4.098	0.3779050736	2.093
Baxter	38	10	9	1.1111111111	0	2.8085	0.1423001949	1.6875
Kevin	52	8.5	8.5	1	0.0775	2.474	0.1214285714	1.6665
Selma	37	2.5	1.5	1.6666666667	0	1.1665	0.0275012444	1.1665

The table has been ordered on the number of edges/syntactic links produced by the speaker. Age has also been calculated in months. *Rel_Size* indicates the relative size of the giant connected component (GCC) or largest network within the graph; <*k*> indicates the averaged number of edges/syntactic links of the GCC; and ratio N/E indicates the averaged number of nodes/edges.

of Dutch DS children, their vocabulary is still small but there seems not to be an analogous difficulty for combining words syntactically (even though only one case reached a small-world topology), and in spite of possible morpho-syntactic deficits (Bol & Kuiken, 1990). Interestingly, network measures clearly showed that their respective developmental paths differ substantially from the one followed by neurotypicals.

Crucially, our data showed that up to eight DS speakers produced a larger number of different words than the neurotypical children (it has to be noted that our DS speakers are significantly older than our TD speakers) but the number of syntactic connections was noticeably smaller for DS speakers, and only three DS speakers produced as many edges as TD younger children did. That is, vocabulary size does not directly map onto higher network complexity: in our case TD children with smaller vocabularies showed a richer command of their combinatorial properties than AD children with larger vocabularies. This is a clear

outcome of our study, which may go unnoticed to the naked eye, but was uncovered by the network analysis.

Furthermore, it has been mentioned that an indicator of the cohesion of networks is the clustering coefficient (*C*). It is interesting to see whether networks that are similar in the number of nodes and edges differ regarding *C*, because this would be the clearest indicator that there is a substantial difference in the cohesion of the network (see Table 5). For example, only one of the DS Dutch speakers, "Ronnie," reached the threshold of the small-world characterized by a ratio of word/syntactic relations of 1:2. Does this mean that there is no difference between this DS speaker's linguistic combinatorial ability and that of TD speakers? Let's compare data: the networks of the TD Dutch speaker "Daan" at 926 days have the following values: $N = 223$, $E = 368$, $C = 0.063$, while the DS Dutch Ronnie at the much older 5963 days reaches $N = 223$, $E = 356$, $C = 0.036$; thus, two networks quite similar regarding the n of nodes and edges show a notable

difference regarding the *C*. By contrast, the TD German speaker “Simone” (in purple in Figure 3) reaches the DS Dutch Ronnie’s *C* with barely half of the elements: at 873 days ($N = 139$, $E = 206$, $C = 0.035$). In other words, the 3-year-old TD German speaker, with much less lexical elements, shows a pattern of linguistic cohesion which is similar to the pattern attested in the big DS linguistic network. These empirical data confirm again what Watts and Strogatz (1998) theoretically proposed: in networks, nodes are important, of course, but connectivity is a crucial factor that can sharply distinguish between two apparently similar networks. Again, a network smaller in the number of nodes can be more cohesive (i.e. have a higher *C* and a lower *L*) than a network with a larger number of nodes; the size matters, but is not the only important indicator. In our study, we interpret this as suggesting that a more connected network shows a richer command of the combinatorial properties of lexical items; a more disconnected one, on the other hand, seems to treat combinations more as extensional exceptions than as a general intensional rule that can be recursively applied.

Summarizing, our network analysis proves to be a useful clinical tool and clearly shows the difference between population groups, providing objective and formal indicators of complexity in performance reached by different individuals/groups.

On a different note, the comparison of the hubs found provides also an important source of clinical information (Tables 3 and 4). While hubs in the atypicals largely belong to the same functional class of hubs in TD networks, the number of combinations in AD is much smaller. Although it can be said that these are still hubs, the number of links is considerably smaller in comparison to that of healthy individuals. English speakers affected by DS produced a few more than a dozen syntactic links with the word “the.” Only the oldest individual (“Mat”) produced 30 combinations. In the case of “it,” this item is seldom produced (almost disappeared) and it does not constitute any hub. In the study of DS in Dutch, results also show striking differences in the scores of hubs. The number of links is in the best case (“Ronnie”) half of that of TD children. However, there appears a different hub, represented by the word *en* (“and”), suggesting that these speakers exploit coordination (parataxis) as a linguistic strategy maybe to avoid the complexities of structural embedding (hypotaxis). Interestingly, the number of links of the set of demonstratives is severely reduced in both DS groups, which in several cases amounts to zero links. See Table 3 for the DS performance in English, Table 4 for the Dutch cases, and Table 5 for the network values of DS syntactic networks.

Finally, an additional, interesting question is whether or not this network-based technique has convergent validity. Given that the MLU measure was

available for both groups of DS individuals in the CHILDES database and in the original works (Bol & Kuiken, 1990; Rondal, 1978a, 1978b), we further investigated the relationship between this measure and our basic ones: the number of nodes (words) and edges (dependency relationships) of the LCC, the largest network in each graph (Figure 3). First of all, we checked for normality and homoscedasticity of the data. Both normality and homoscedasticity were met. Then we wanted to know whether both MLU and linguistic networks measured the same linguistic aspect. Hence, a Pearson correlation was performed and it revealed a similar level of correlation between MLU and both network measures ($r = .704$ for nodes and $r = .709$ for edges, in both cases with $p < .001$), suggesting that both measures keep track of the same growth process.

An important question was whether the type of language could affect the network analysis. For this reason, three ANOVAs (analyses of variance) were performed setting language as independent variable, and MLU, nodes, and edges as dependent variables. Analyses showed that for MLU, the language variable accounts for 27% of the differences between groups ($F(1, 30) = 11.03$; $p = .002$; $\eta = .27$; CI (η)95%: .025–.971). On the contrary, the type of language only accounts for 1.4% in the case of nodes ($F(1, 30) = 0.44$; $p = .512$; $\eta = .014$; CI 95%: .025–.487) and 1.3% in the case of edges ($F(1, 30) = 0.38$; $p = .542$; $\eta = .013$; CI 95%: .025–.458)—even if the latter are non-significant. This outcome suggests that complex networks can help overcoming the language-specific problems of morphology and closed-class words that the word-based MLU measurement faces, becoming a useful complementary methodology to assess the linguistic competence of DS speakers.

7. Discussion

The present work has explored incipiently the dynamics of language ontogeny in both typical and atypical conditions. This has been possible thanks to the combination of linguistic theory and network science. The sentences produced by the speakers have first been analyzed syntactically following syntactic theory: syntactic items were detected and their dependency relationships, if any, labeled with a dependency link. In contrast to other network approaches based on word co-occurrence, here networks have been created on the basis of large collections of sentences, each of them syntactically analyzed by a linguist, word by word, item by item. These analyses have been transmitted to a network program, which interpreted words as nodes and syntactic links as edges, creating a network from each file (Figure 1). This procedure takes a considerable amount of time, though we think it is worth it, given the results we obtained, which uncovered a pattern of crosslinguistic



Figure 3. Number nodes (top) and edges (bottom) produced by typical and atypical individuals. Age is presented in days. Typical individuals (on the left of each graphic) are represented by three linked points, letting see their progress through time. Each point represents a different kind of network. The Down syndrome cases (12 English speakers represented by dots and 20 Dutch speakers represented by triangles) represent different individuals of different ages.

unity and cross-population variation (we observed similar developmental paths traversing the same phases at similar ages across genetically and typologically different languages, and radically diverging developmental

paths in DS populations). Admittedly, the procedure followed during the analyses of corpora could have affected or influenced in the resultant structure (see Gruenenfelder & Pisoni, 2009). Barceló-Coblijn,

Duguine, and Irurtzun (2019) analyzed twice the same language corpora of three languages (Basque, French, and Dutch), adopting different linguistic theoretical hypotheses each time (related to the role of determiners as *governors* or as dependent elements of nouns). Results clearly show that the same corpus produces two different networks and that the degree of edges of determiners is severely altered depending on the adopted hypothesis. However, clustering (*C*) was also affected in an important way. In the present work, we have adopted and applied always the same theoretical decisions to both typical and DS corpora. Differences between these two groups of children are clear, separating typical children from DS children. The absence of mixed scores between typical and atypical reinforces our intuition that these first, explorative results could be on the right track.

A preliminary conclusion is that the apparent existence of several developmental phases crosslinguistically could imply that linguistic development is not a process that follows either a simple linear or a non-linear progress, but a “saltational” process that combines both kinds of progresses (and thus overcoming the black-or-white scenario mentioned above): when the system reaches a particular level, it takes some time until the next phase transition; during this period a pattern of linear progress is also observed in the network growth. These results are in agreement with Labrell et al. (2014) who found a developmental path characterized by several spurts and plateaus and also with recent findings in the ontogeny of English syntactic networks (Corominas-Murtra, Sánchez Fibla, Valverde, & Ricard Solé, 2018). However non-linearity is evident when the network radically changes its topology, from one type of network into another of a completely different type. Until now, in the present work this kind of dual progression has been confirmed in seven longitudinal corpora of seven different languages. Although the group size is still too small for a statistic generalization, and therefore the nature of the present work remains explorative, we think that our results suggest that this is not happenstance. Interestingly, our results and the results found in Barceló-Coblijn et al. (2012) have both found non-linearity of developmental patterns, coinciding with other works based on dynamic systems theory (Corominas-Murtra et al., 2018; Labrell et al. 2014; Ruhland & van Geert, 1998).

Changes in topology coincide with the increase of the degree of functional words, as if they were they core of the syntactic capability. This is not something new, as it has been pointed out in empirical works in linguistics (Bassano et al., 2011; Ruhland & van Geert, 1998; van Dijk & van Geert, 2007), though the present work, in spite of its exploratory aim, provides a global perspective of how these words emerge and occupy their place within the system. The perspective that the linguistic system evolves so that several stages must be

traversed during ontogeny, and that the system changes when important, structural elements are incorporated, filling structural gaps, has already been put forth in recent works on semantics and lexicon ontogeny (Stella, Beckage, & Brede, 2017). More recently, Stella et al. (2018) have shown how the lexicon could evolve during ontogeny through an abrupt transition at age 7, and that the key of that developmental abruptness is the acquisition of the core set of vocabulary. Interestingly, the results of network-based approaches on both syntax and semantics suggest that the language capacity develops following quite similar patterns regarding syntax and semantics, something that nicely fits Elisabeth Bates’ classic hypothesis about the “inseparability of grammar and the lexicon” and that “the emergence of grammar is highly dependent upon vocabulary size” (Bates & Goodman, 1997).

With respect to atypical development, our study might have uncovered a pattern that is unobservable by other means even if the paucity of linguistic structures in AD children is considered symptomatic and related to their atypical condition. The present approach provides a framework to change our conception of AD: in agreement with previous findings, atypical cases are not just delayed stages in the course of typical development. And in coincidence with previous works (e.g. Chapman, Hesketh & Kistler, 2002; Chapman, Sindberg, Bridge, Gigstead, & Hesketh, 2006), a large degree of variability has been found within the DS group. We would like to stress that several DS children could reach a high score in their lexical production and that notwithstanding their syntactic performance—their capability to syntactically combine lexical items—clearly differed from their typically developed peers. This suggests that focusing on word counting only could mask important information related to the individual’s language capacity. In our study, most of the times the DS speakers did not reach the TD speakers’ topology (only 1 speaker out of 32 did, and at a significantly older age), nor their use and connectivity of functional words was the same, which could be interpreted as being due to a specific difficulty in grasping the syntactic rules and mastering the combinatoriability of functional, closed-class words in combination to the general difficulty to learn new vocabulary (Assen & Bol, 2012). This strongly suggests that the *connectivity* in atypical cases like DS is essentially different. Note that, for example, up to eight DS speakers have a number of nodes (words) larger than several neurotypicals. The clearest example is the German TD speaker (“Simone,” in purple, in Figure 3)—a child qualified as “exceptionally gifted” according to IQ tests (see CHILDES online manuals). If it were all about the number of words, the TD German child should have a less developed linguistic skill than the group of DS speakers that have produced more words than her (or their production should be more complex than hers).

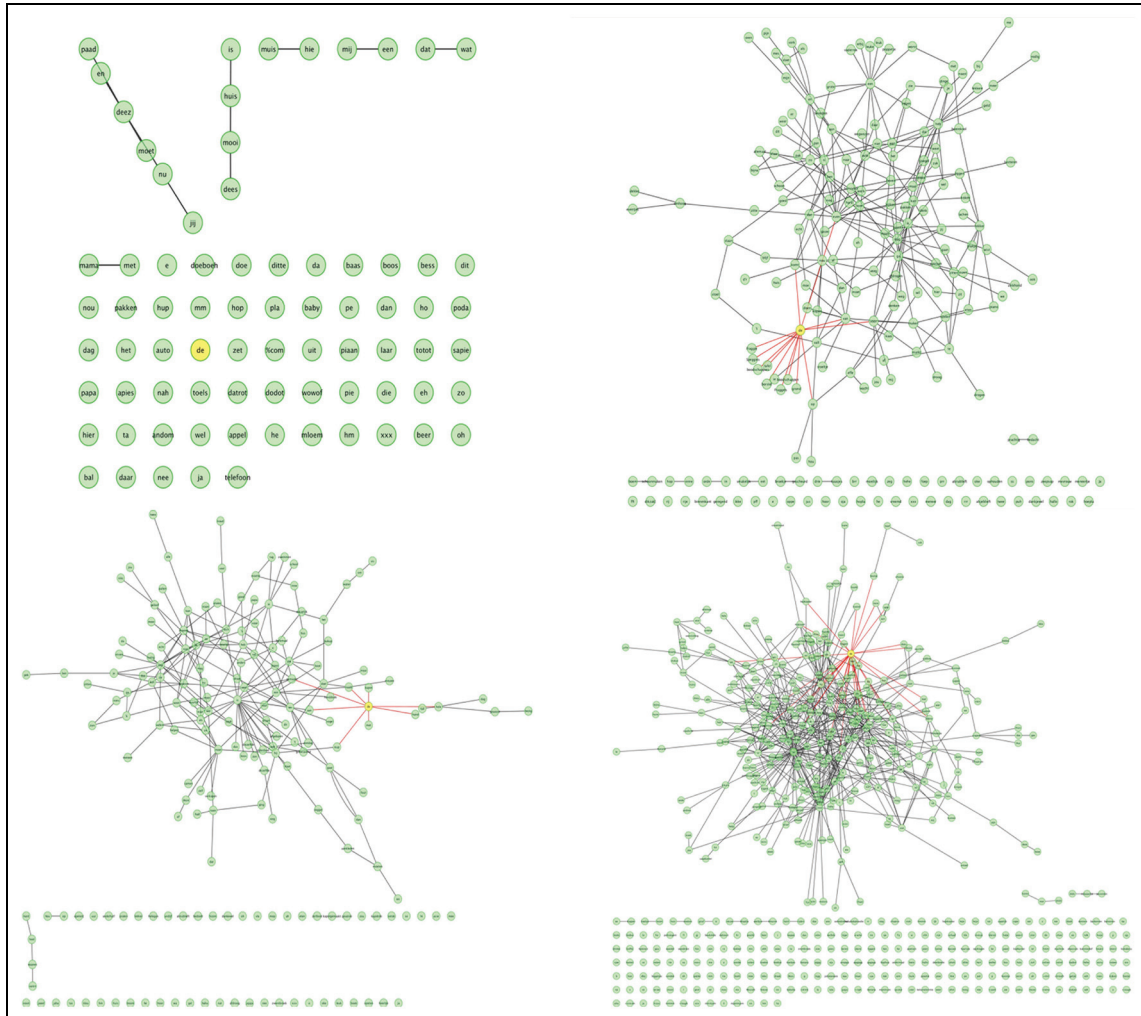


Figure 4. Four graphs from Dutch corpora—three samples of linguistic performance affected by Down syndrome: top left (age 04;04.21), top right (age 10;05.03), and bottom left (age 18;05.02). Bottom right, typical child's performance (age 02;06.25). The Dutch determiner “de” has been highlighted.

But, as repeatedly said in this work, the difference is not the number of words *only*, because two structurally different networks could have a similar number of words/nodes. To us, the crucial difference is in the way nodes are interconnected, the number of edges (syntactic connections), and in the connectivity of hubs, most of the times represented by functional words. Here, the performance of Simone, reflected in the network, undergoes an explosion of syntactic edges and they are connected in such a way that the network develops the structure and topology of a network with a high clustering coefficient and a low average path length (thus suggesting small-worldness). Like the rest of TD children, Simone's hubs were functional words and the relation between words and syntactic links is near to 1:2 (for each word, two syntactic relationships). Differences in the network topology are the reflection of their different linguistic performance (see Figure 4 for comparison). In DS networks, there are less connections in number, but also the type of long-distance

connections—allegedly the promoter of small-worldness—is considerably reduced, if not completely absent. Moreover, the hubs of the network also show an atypical behavior, with very low scores or with lexical differences. These results coincide again with Bates & Goodman (1997), who pointed out that some structures like function words are especially vulnerable to brain damage in neurological intact speakers under perceptual degradation (see Barceló-Coblijn et al. 2017 for a first syntactic network analysis of linguistic development of a couple of twins, one of them affected by a focal brain injury).

Regarding the use and applicability of networks in psycholinguistics, statistics also indicate that this technique has convergent validity and hence that it can complement other measures such as the MLU. As a complement to MLU or other techniques, a network analysis is not their substitute. A network analysis extracts formal indicators from a corpus that otherwise cannot be obtained. Other techniques are more focused

on the analysis of sentences as independent entities. From a tree-like network, one could track the path of one of the sentences, but it is certainly a challenge for humans to track a sentence in a small-world network with hundreds of nodes and edges. Although the creation of specific software could perhaps track a sentence within a small-world network, this would be of little use for the macroscopic view we took in this work. Tracking the original sentences is not the function of networks. In our view, what they are really useful for is for providing formal indicators of the pattern of connectivity extracted from a linguistic corpus. An outstanding example of this central idea, applied to semantic association, is Kenett, Gold, and Faust's (2015) approach to semantic organization of people with Asperger syndrome (AS). Their networks suggest that AS speakers have a particular (hyper-modularized), more compartmentalized, mental lexicon.

With respect to syntactic theory, we think that the present work contributes to solving at least part of the issues raised by Ninio (2014). We think that some parameters are more useful than others, depending on the nature behind the network that is being analyzed. The small-world topology is a first indicator of efficiency (Martin, 2012), but in our case, the linguistic context is really important (Čech, Mačutek, & Liu, 2016). It is a cognitive context, and it is a (developing) organism what is under study and hence a dynamical process (van Geert 1994; Barceló-Coblijn, Corominas-Murtra, & Gomila 2012). Nonetheless, this is not enough. Importantly, a crucial characteristic of AD speech is the atypical, low connectivity of the typical hubs, or even the emergence of other words as hubs, something that has already been pointed out in other network-based approaches to clinical studies (Crossley et al., 2014). Moreover, in our approach, syntactic connections/edges are not random but established following syntactic theory. The linearity imposed by speech is not a problem for our approach, since the analysis is based on syntactic dependency relations and syntactic theory, which are not limited by linear distance between words within a sentence. The absence of any linearity constraints is more important than it may seem at first glance, because, among other reasons (e.g. remind Zweig's (2016) observations regarding the true nature of co-occurrence networks), our syntactic analysis goes beyond the transcription and takes benefit of advances in contemporary syntactic theory. In this aspect, we agree with Čech, Mačutek, and Liu's (2016) observation that answers about the role of syntax in networks should take into account linguistic theory, and we think that the speaker's syntactic ability affects several relevant aspects, like the nature and connectivity of hubs. However, theoretical syntax itself can also affect hubs' connectivity, depending on the theoretical choice one applies—for example, the degree $\langle k \rangle$ (n of edges) of determiners (see Barceló-Coblijn et al., 2019, and

Supplementary Material in the present work). Importantly, notice that acknowledging syntactic relations as something real—that is, not as mere theoretical labels—would open the door to analyses of directed networks.

Finally, we would like to also frame our approach within contemporary debates in cognitive science. Here the psychological ability for language is considered a complex dynamic system. Complexity emerges when different elements within a system interact with each other. In the case of language, indicators of complexity had already been detected in phonology (e.g. Vitevitch, 2008) or semantics (Bilson, Yoshida, Tran, Woods, & Hills, 2015; Hills, Maouene, Riordan, & Smith, 2010; Hills et al., 2009), and the present work shows that the ability to combine words also shows clear indicators of a complex behavior. By comparing TD and AD speakers, we have seen that the latter do develop, although following a different developmental path. Our studies fit that third Evo-Devo approach to human ontogeny mentioned in the introduction, according to which the individual development follows both linear and non-linear progressions (e.g. see the classic Evo-Devo biology, for example, Alberch & Alberch, 1981; Alberch, Gould, Oster, & Wake, 1979). An Evo-Devo approach also provides a conceptual framework to make sense of our results on atypical development: they are to be viewed not as delays, but as different phenotypical paths (hence meeting Karmiloff-Smith's, 2007, and van Geert's, 2004, viewpoint). Therefore, we think that syntactic networks can be used as psychiatric endophenotypes (Barceló-Coblijn et al., 2015) since the network and its formal, objective indicators meet the criteria of a biomarker bridging genotype and phenotype. When a biologically driven disturbance constrains development, like in DS, the linguistic phenotype stabilizes differently. Such divergent development due to the effect of the trisomy in DS populations is also supported by studies of the *connectome* that show that the connectivity pattern of DS brains resembles a random network, while the connectivity pattern of neurotypicals is small-world (Ahmadlou, Gharib, Hemmati, Vameghi, & Sajedi, 2013). Future research will deal with linguistic samples of other syndromes that also affect language, to see whether or not it is possible to visualize a phenotypic morpho-space, as hypothesized by Barceló-Coblijn and Gomila (2014). On another front, more control is necessary during corpus creation. If session time is selected as a crucial feature, then it should be controlled both during recording sessions (all sessions with the same duration) and between sessions (the same intervals in days between recordings). Other aspects, like the intervention of other actors (relatives, friends, etc.) should also be strongly controlled. This could be solved with the development of a specific protocol. The same can be said regarding the hypothesis the linguist follows during the syntactic analysis. As in any other

discipline, syntactic theory contains several hypotheses. Syntactic network approaches would greatly benefit from a detailed syntactic protocol created and shared to the community.

8. Limitations and future directions

The present work contributes to the issue of the suitability of network analyses vis-à-vis language variation. It argues for such a suitability, provided that the network modeling is combined with a proper linguistic analysis. However, there still are some aspects that could be improved in the future. One of them is the integration of syntactic information from the syntactic links inside edges. Until now, the available software does not allow such integration. Ideally, information from the syntactic link such as “subject,” “complement,” and so on (Figure 1) should be encapsulated inside the edge so that it can be analyzed further. As mentioned before, this would make possible an analysis of directed networks, since the information “inside” the edge could be checked anytime. Information about lexical categories (i.e. noun, verb, adjective, determiner, etc.) is also lost during file transformation in current work. Ideally, in future developments such information should also be saved to see automatically which lexical categories vertebrate the network, for example, using a specific program for these matters (e.g. Barceló-Coblijn et al., 2017). Another important aspect to reinforce is the statistical information that can be obtained from this kind of approaches. The present work is exploratory due to the size of corpora. Future work including a larger number of participants and languages could draw a stronger conclusion on the developmental trajectories between TD and AD children. In relation to this, and as observed in the previous section, gathering new data from spontaneous speech controlling time session, intervals, and the number of intervening actors will discard potential biases affecting the analysis.

Authors' Note

Cristina Real Puigdollers is now affiliated with Universitat Pompeu Fabra, Barcelona, Spain.

Acknowledgements

We thank Paco Calvo and Manuel G Bedia for their support in the development of this work.

Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work has been supported by the following projects and grants: IT769-13 (Basque Government), FFI2017-87140-C4-1-P (MINECO), BIM (ANR), UV2 (ANR-DFG), PSI2014-62092-EXP, TIN2016-80347-R and PGC2018-096870-B-100 (MiCIU, AEI & FEDER, EU).

Supplemental material

Supplemental material for this article is available online.

ORCID iD

Lluís Barceló-Coblijn  <https://orcid.org/0000-0002-8765-6314>

Notes

1. See Barabási and Albert's (1999) seminal work.
2. For example, a text like *Moby Dick* shows precisely the connectivity pattern of the small-world network (Solé et al., 2010).
3. <http://wals.info/>
4. Besides, the Basque verb is highly inflected, showing agreement in number and person with subjects, direct objects, and indirect objects (plus optionally even the addressee) to the point that a sentence can be completely grammatical composed of just a synthetic verbal form such as *ba-d-akar-zki-zu-e-ke-t* [Pol-X-√bring-3PlAbs-2sgDat-2Pl-Pot-1sgErg] ‘‘I may bring them to you (pl.) .’’
5. “Reverse compositionality” is introduced by Fodor and Lepore (2001) as such.

Compositionality says, roughly, that its syntax and its lexical constituents determine the meaning of a complex expression; it's thus part of the explanation of why practically everybody who understands “dogs” and “bark” understands “dogs bark.” But it also needs explaining that you practically never find people who understand “dogs bark” but don't understand “dogs” or “bark.” What we'll call “reverse” compositionality explains this by assuming that each constituent expression contributes the whole of it's meaning to its complex hosts. If that's right, then if you understand “dogs bark,” it follows that you know everything you need to determine the meanings of “dog” and “bark”: in effect, the meanings of the parts of a complex expression supervene on the meaning of that expression. (pp. 365–366)

References

- Abbeduto, L., Warren, S. F., & Conners, F. A. (2007). Language development in Down syndrome: From the prelinguistic period to the acquisition of literacy. *Mental Retardation and Developmental Disabilities Research Reviews*, *13*, 247–261.
- Ahmadlou, M., Gharib, M., Hemmati, S., Vameghi, R., & Sajedi, F. (2013). Disrupted small-world brain network in children with Down syndrome. *Clinical Neurophysiology*, *124*, 1755–1764.
- Alberch, P., & Alberch, J. (1981). Heterochronic mechanisms of morphological diversification and evolutionary change in the neotropical salamander, *Bolitoglossa occidentalis* (Amphibia: Plethodontidae). *Journal of Morphology*, *167*, 249–264.
- Alberch, P., Gould, S. J., Oster, G. F., & Wake, D. B. (1979). Size and shape in ontogeny and phylogeny. *Paleobiology*, *5*, 296–317.
- Arbesman, S., Strogatz, S. H., & Vitevitch, M. S. (2010). The structure of phonological networks across multiple languages. *International Journal of Bifurcation and Chaos*, *20*, 679–685.

- Arita, M. (2005). Scale-freeness and biological networks. *Journal of Biochemistry*, *138*, 1–4.
- Assen, N., & Bol, G. (2012). Fast mapping bij jongeren met het syndroom van Down [Fast mapping in juveniles with Down syndrome]. *Stem-, Spraak- En Taalpathologie*, *17*(4), 1–19.
- Baker, M. C. (2005). The innate endowment for language: Overspecified or underspecified? In P. Carruthers, S. Lawrence, & S. Stich (Eds.), *The innate mind: Structure and contents* (pp. 156–174). Oxford, UK: Oxford University Press.
- Balari, S., & Lorenzo, G. (2013). *Computational phenotypes: Towards an evolutionary developmental biolinguistics*. Oxford, UK: Oxford University Press.
- Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, *286*, 509–512. doi:10.1126/science.286.5439.509
- Barceló-Coblijn, L., Benítez-Burraco, A., & Irurtzun, A. (2015). Syntactic networks as endophenotype of developmental language disorders: An Evo-Devo approach to clinical linguistics. *Biolinguistics*, *9*, 43–49.
- Barceló-Coblijn, L., Corominas-Murtra, B., & Gomila, A. (2012). Syntactic trees and small-world networks: Syntactic development as a dynamical process. *Adaptive Behavior*, *20*, 427–442. doi:10.1177/1059712312455439
- Barceló-Coblijn, L., Duguine, M., & Irurtzun, A. (2019). The emergence of hubs in complex syntactic networks and the DP hypothesis: The relevance of a linguistic analysis. In A. Bastardas, A. Massip, & G. Bel-enguix (Eds.), *Applications of complexity theory in language and communication sciences* (pp. 273–288). Dordrecht, The Netherlands: Springer. doi:10.1007/978-3-030-04598-2_15
- Barceló-Coblijn, L., & Gomila, A. (2014). Syntactic development in pheno-typic space. In E. A. Cartmil, S. Roberts, H. Lyn, & H. Cornish (Eds.), *The Evolution of Language: Proceedings of the 10th International Conference (EVO-LANG10)*, pp. 33–40. Singapore: World Scientific.
- Barceló-Coblijn, L., Serna Salazar, D., Isaza, G., Castillo Ossa, L. F., & Bedia, M. G. (2017). Netlang: A software for the linguistic analysis of corpora by means of complex networks. *PLoS ONE*, *12*(8), e0181341.
- Bassano, D., Maillochon, I., Korecky-Kröll, K., van Dijk, M., Laaha, S., Dressler, W. U., & van Geert, P. (2011). A comparative and dynamic approach to the development of determiner use in three children acquiring different languages. *First Language*, *31*, 253–279.
- Bates, E., & Goodman, J. (1997). On the inseparability of grammar and the lexicon: Evidence from acquisition, aphasia and real-time processing. *Language and Cognitive Processes*, *12*, 507–586.
- Beckage, N., Smith, L., & Hills, T. (2011). Small-world network growth in typical and late talkers. *PLoS ONE*, *6*(5), e19348. doi:10.1371/journal.pone.0019348
- Benítez-Burraco, A. (2016). A biolinguistic approach to language disorders: Towards a paradigm shift in clinical linguistics. In K. Fujita, & C. Boeckx (Eds.), *Advances in biolinguistics: The human language faculty and its biological basis* (pp. 256–272). London, England: Routledge.
- Benítez-Burraco, A., & Boeckx, C. (2014). Universal Grammar and biological variation: An Evo-Devo agenda for comparative biolinguistics. *Biological Theory*, *9*, 122–134.
- Benítez-Burraco, A., & Murphy, E. (2016). The oscillopathic nature of language deficits in autism: From genes to language evolution. *Frontiers in Human Neuroscience*, *10*, Article 120. doi:10.3389/fnhum.2016.00120
- Bilson, S., Yoshida, H., Tran, C. D., Woods, E. A., & Hills, T. T. (2015). Semantic facilitation in bilingual first language acquisition. *Cognition*, *140*, 122–134.
- Bol, G., & Kuiken, F. (1990). Grammatical analysis of developmental language disorders: A study of the morphosyntax of children with specific language disorders, with hearing impairment and with Down's syndrome. *Clinical Linguistics and Phonetics*, *4*, 77–86. doi:10.3109/02699209008985472
- Brock, J. (2007). Language abilities in Williams syndrome: A critical review. *Development and Psychopathology*, *19*, 97–127.
- Čech, R., Mačutek, J., & Liu, H. (2016). Syntactic complex networks and their applications. In A. Mehler, A. Lücking, S. Banish, P. Blanchard, & B. Frank-job (Eds.), *Towards a theoretical framework for analyzing complex linguistic networks* (pp. 167–183). Berlin, Germany: Springer-Verlag.
- Chapman, R. S., Hesketh, L. J., & Kistler, D. J. (2002). Predicting longitudinal change in language production and comprehension in individuals with Down syndrome: Hierarchical linear modeling. *Journal of Speech, Language, and Hearing Research*, *45*, 902–915.
- Chapman, R. S., Sindberg, H., Bridge, C., Gigstead, K., & Hesketh, L. J. (2006). Effect of memory support and elicited production on fast mapping of new words by adolescents with Down syndrome. *Journal of Speech, Language, and Hearing Research*, *49*, 3–15.
- Chomsky, N. (1975). *Reflections on language*. New York, NY: Pantheon Books.
- Corominas-Murtra, B. (2007). Network statistics on early English Syntax: Structural criteria. arXiv. Retrieved from <http://arxiv.org/abs/0704.3708>
- Corominas-Murtra, B., Sánchez Fibla, M., Valverde, S., & Solé, R. (2018). Chromatic transitions in the emergence of syntax networks. *Royal Society Open Science*, *5*, Article 181286. doi:10.1098/rsos.181286
- 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*, *12*, 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*, *137*, 2382–2395.
- Demuth, K., & Tremblay, A. (2008). Prosodically-conditioned variability in children's production of French determiners. *Journal of Child Language*, *35*, 99–127.
- Ezeizabarrena, M.-J., & Garcia Fernandez, I. (2018). Length of utterance, in morphemes or in words? MLU3-w, a reliable measure of language development in early Basque. *Frontiers in Psychology*, *8*, Article 2265. doi:10.3389/fpsyg.2017.02265
- Fenson, L., Dale, P., Reznick, S., Thal, D., Bates, E., Hartung, J. P., & . . . Reilly, J. S. (1993). *MacArthur Communicative Development Inventories: Technical manual*. San Diego, CA: Singular Publishing Group.

- Ferrer-i-Cancho, R., & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society B*, 268, 2261–2265. doi:10.1098/rspb.2001.1800
- Fodor, J. (1998). *In critical condition*. Cambridge: MIT Press.
- Fodor, J., & Lepore, E. (2001). Why compositionality won't go away: Reflections on Horwich's "deflationary" theory. *Ratio*, 14, 350–368. doi:10.1111/1467-9329.00169
- Galeote, M., Soto, P., Sebastián, E., Checa, E., & Sánchez-Palacios, C. (2014). Early grammatical development in Spanish children with Down syndrome. *Journal of Child Language*, 41, 111–131.
- Gibson, G. (2009). Decanalization and the origin of complex disease. *Nature Reviews Genetics*, 10, 134–140. doi:10.1038/nrg2502
- Gottesman, I. I. & T. D. Gould. (2003). The endophenotype concept in psychiatry: Etymology and strategic intentions. *American Journal of Psychiatry*, 160, 636–645.
- Gruenenfelder, T. M., Pisoni, D. B. (2009). The lexical restructuring hypothesis and graph theoretic analyses of networks based on random lexicons. *Journal of Speech Language and Hearing Research*, 52, 596–609.
- Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, 63, 259–273.
- Hills, T. T., Mouene, M., Mouene, J., Sheya, A., & Smith, L. B. (2009). Longitudinal analysis of early semantic networks. *Psychological Science*, 20, 729–739. doi:10.1111/j.1467-9280.2009.02365.x
- Hudson, R. A. (1990). *English word grammar*. Hoboken, NJ: Blackwell Publishing.
- Johnson, N. F. (2007). *Two's company, three is complexity: A simple guide to the science of all sciences*. Oxford, UK: Oneworld.
- Kaplan, E., Goodglass, H., & Weintraub, S. (1983). *Boston Naming Test*. Philadelphia, PA: Lea & Febiger.
- Karmiloff-Smith, A. (2007). Atypical epigenesis. *Developmental Science*, 10, 84–88.
- Kenett, Y. N., Gold, R., & Faust, M. (2015). The hypermodular associative mind: A computational analysis of associative responses of persons with Asperger syndrome. *Language and Speech*, 59(Pt. 3), 297–313.
- Kenett, Y. N., Kenett, D. Y., Ben-Jacob, E., & Faust, M. (2011). Global and local features of semantic networks: Evidence from the Hebrew mental lexicon. *PLoS ONE*, 6(8), e23912.
- Kenett, Y. N., Wechsler-Kashi, D., Kenett, D. Y., Schwartz, R., Ben-Jacob, E., & Faust, M. (2013). Semantic organization in children with cochlear implants: Computational analysis of verbal fluency. *Frontiers in Psychology*, 4, Article 543. doi:10.3389/fpsyg.2013.00543
- Labrell, F., van Geert, P., Declercq, C., Baltazart, V., Caillies, S., Olivier, M., & Le Sourn-Bissaoui, S. (2014). "Speaking volumes": A longitudinal study of lexical and grammatical growth between 17 and 42 months. *First Language*, 34, 97–124.
- Laka, I. (1996). *A brief grammar of Euskara, the Basque language*. Bilbao, Spain: University of the Basque Country UPV/EHU. Retrieved from <http://www.ehu.es/eu/web/eins/a-brief-grammar-of-euskara>
- Lenneberg, E. H. (1967). *Biological foundations of language*. New York, NY: Wiley.
- Leonard, L. B., & Finneran, D. (2003). Grammatical morpheme effects on MLU: "The same can be less" revisited. *Journal of Speech Language and Hearing Research*, 46, 878–888.
- MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. The database*. Mahwah, NJ: Lawrence Erlbaum.
- Martin, G. (2012). Network analysis and the connectopathies: Current research and future approaches. *Nonlinear Dynamics, Psychology, and Life Sciences*, 16, 79–90.
- Mehler, J. (1974). Connaître par déssaprentissage. In E. Morin, & M. Piattelli-Palmarini (Eds.), *L'unité de l'homme II: le cerveau humain [Knowing by forgetting]* (pp. 25–37). Paris, France: Editions du Seuil.
- Mitchell, M. (2009). *Complexity: A guided tour*. Oxford, UK: Oxford University Press.
- Moro, A. (2008). *The boundaries of babel*. Cambridge: MIT Press.
- Müller, N., & Ball, M. J. (Eds.) (2013). *Research methods in clinical linguistics and phonetics: A practical guide*. Oxford, UK: Wiley-Blackwell.
- Mundy, P., Kasari, C., Sigman, M., & Ruskin, E. (1995). Non-verbal communication and early language acquisition in children with Down syndrome and normally developing children. *Journal of Speech and Hearing Research*, 38, 157–167.
- Neal, Z. (2018). Is the urban world small? The evidence for small world structure in urban networks. *Networks and Spatial Economics*, 1–17. doi:10.1007/s11067-018-9417-y
- Newman, M. E. J. (2010). *Networks: An introduction*. Cambridge, MA: Oxford University Press.
- Newmeyer, F. (2005). *Possible and probable languages: A generative perspective on linguistic typology*. Oxford, UK: Oxford University Press.
- Ninio, A. (2006). *Language and the learning curve: A new theory of syntactic development*. Oxford, UK: Oxford University Press.
- Ninio, A. (2014). Syntactic networks, do they contribute valid information on syntactic development in children? Comment on "Approaching human language with complex networks" by J. Cong and H. Liu. *Physics of Life Reviews*, 11, 632–634.
- Piattelli-Palmarini, M. (1989). Evolution, selection and cognition: From "learning" to parameter setting in biology and in the study of language. *Cognition*, 31(1), 1–44. doi:10.1016/0010-0277(89)90016-4
- Popescu, M. (2003). Dependency Grammar annotator. In H. Florentina, & M. Popescu (Eds.), *Building awareness in language technology* (pp. 17–34). Bucharest, Romania: Editura Universității din București.
- Ravid, D., & Schiff, R. (2006). Roots and patterns in Hebrew language development: Evidence from written morphological analogies. *Reading and Writing*, 19, 789–818.
- Rice, M. L., Smolik, F., Perpich, D., Thompson, T., Rytting, N., & Blossom, M. (2010). Mean length of utterance levels in 6-month intervals for children 3 to 9 years with and without language impairments. *Journal of Speech Language and Hearing Research*, 53, 333–349. doi:10.1044/1092-4388(2009/08-0183)
- Roberts, J. E., Price, J., & Malikin, C. (2007). Language and communication development in Down syndrome. *Mental Retardation and Developmental Disabilities Research Reviews*, 13, 26–35.

- Roeper, T. (2007). *The prism of grammar: How child language illuminates humanism*. Cambridge: MIT Press.
- Rollins, P. R., Snow, C. E., & Willett, J. B. (1996). Predictors of MLU: Semantic and morphological developments. *First Language*, 16, 243–259. doi:10.1177/014272379601604705
- Rondal, J. (1978a). Maternal speech to normal and Down's syndrome children matched for mean length of utterance. *Monograph of the American Association of Mental Deficiency*, 3, 193–265.
- Rondal, J. (1978b). Patterns of correlations for various language measures in mother-child interactions for normal and Down's syndrome children. *Language and Speech*, 21, 242–252.
- Ruhland, R. & van Geert P. (1998). Jumping into syntax: Transitions in the development of closed class words. *British Journal of Developmental Psychology*, 16, 65–95.
- Sagae, K., Davis, E., Lavie, A., MacWhinney, B., & Winter, S. (2010). Morphosyntactic annotation of CHILDES transcripts. *Journal of Child Language*, 37, 705–729.
- Scarborough, H. S. (1990). Index of productive syntax. *Applied Psycholinguist*, 11, 1–22. doi:10.1017/S0142716400008262
- Schmitt, C., & Miller, K. (2010). Using comprehension methods in language acquisition research. In E. Blom, & S. Unsworth (Eds.), *Experimental methods in language acquisition research* (pp. 32–56). Amsterdam, The Netherlands: John Benjamins.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., & . . . Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13, 2498–2504. doi:10.1101/gr.1239303
- Solé, R. V., Corominas-Murtra, B., Valverde, S., & Steels, L. (2010). Language networks: Their structure, function, and evolution. *Complexity*, 15, 20–26.
- Soto Valle, R., & Alonso, G. A. (2015). Apuntes sobre la adquisición de la morfosintaxis del euskera en niños de 2 y medio a 5 años [Notes on the acquisition of Basque morpho-syntax in children between 2 and a half and 5 years]. *Revista de Logopedia, Foniatría y Audiología*, 35, 30–49. doi:10.1016/j.rlfa.2014.06.002
- Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. *Scientific Reports*, 7, Article 46730. doi:10.1038/srep46730
- Stella, M., Beckage, N. M., Brede, M., & De Domenico, M. (2018). Multiplex model of mental lexicon reveals explosive learning in humans. *Scientific Reports*, 8, Article 2259. doi:10.1038/s41598-018-20730-5
- Tomasello, M. (2003). *Constructing a language: A usage-based theory of language acquisition*. Cambridge, MA: Harvard University Press.
- Tonelli, L., Dressler, W. U., & Romano, R. (1995). Frühstufen des Erwerbs der italienischen Konjugation: zwei Fallstudien [First steps in the acquisition of Italian conjugation: two case studies]. *Suvremena Lingvistika*, 21, 3–15.
- Trask, R. L. (1998). The typological position of Basque: Then and now. *Language Sciences*, 20, 313–324. doi:10.1016/S0388-0001(98)00006-0
- van Dijk, M., & van Geert, P. (2007). Wobbles, humps and sudden jumps: A case study of continuity, discontinuity and variability in early language development. *Infant and Child Development*, 16, 7–33.
- van Geert, P. (1994). *Dynamic systems of development: Change between Complexity and Chaos*. New York, NY: Prentice Hall.
- van Geert, P. (2004). A dynamic systems approach to diagnostic measurement of SLI. In L. Verhoeven, & H. Van Balkom (Eds.), *Classification of developmental language disorders* (pp. 327–348). Mahwah, NJ: Lawrence Erlbaum.
- Vitevitch, M. S. (2008). What can graph theory tell us about word learning and lexical retrieval? *Journal of Speech Language Hearing Research*, 51, 408–422.
- Vitevitch, M. S., & Castro, N. (2015). Using network science in the language sciences and clinic. *International Journal of Speech-Language Pathology*, 17, 13–25.
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393, 440–442. doi:10.1038/30918
- West-Eberhard, M. J. (2003). *Developmental plasticity and evolution*. Oxford, UK: Oxford University Press.
- Yorkston, K. M., & Beukelman, D. R. (1981). *Assessment of Intelligibility of Dysarthric Speech*. Tigard, OR: CC Publications.
- Zweig, K. A. (2016). Are word-adjacency networks networks? In A. Mehler, A. Lücking, S. Banish, P. Blanchard, & B. Frank-job (Eds.), *Towards a theoretical framework for analyzing complex linguistic networks* (pp. 153–163). Berlin, Germany: Springer-Verlag.

About the Authors



Lluís Barceló-Coblijn is head of the Laboratory of Research on Complexity and Experimental Linguistics at the University of the Balearic Islands (UIB). He has got a bachelor's degree in Catalan philology (2006, UIB) and another one in general linguistics (2007, University of Barcelona (UB)). After that, he made a master's degree in cognitive science and language (2008, UB) and another master's in human evolution and cognition (2010, UIB). His PhD thesis (2012) is on both language evolution and language evolution in ontogeny (first language acquisition). Currently, he is developing a research agenda on language ontogeny from the viewpoint of complex, dynamic systems, analyzing corpora by means of complex networks.



Aritz Irurtzun is *chargé de recherche* (research associate) in linguistics at the CNRS, working at the IKER lab (UMR 5478) in Bayonne, France. Before that, he was adjunct professor at the University of the Basque Country (UPV/EHU), the same institution where he obtained his doctoral degree with a dissertation titled “The grammar of focus at the interfaces.” He works on the general architecture of grammar and the nature of the syntax-semantics and syntax-phonology interfaces, with particular interest on phenomena such as focus, interrogatives, event semantics, or prosody and intonation.

Cristina Real Puigdollers is a postdoctoral researcher in the Department of Translation and Language Sciences at the Universitat Pompeu Fabra (UPF) in Barcelona, from 2016. She holds a PhD in linguistics from the Universitat Autònoma de Barcelona (2013). Her research interests are mainly in linguistic theory, specifically, in the semantics–syntax interface, with particular interest in the semantic and syntactic primitives that compose words. She is currently a researcher in the Group of Formal Linguistics (GLiF) at UPF. Her current work focuses on the lexicalization of events and locations in Romance languages and the theoretical models to derive crosslinguistic variation. She has published in journals including *Acta Linguistica Hungarica*, *The Linguistic Review*, and *Linguistic Analysis*.



Emilio López-Navarro is a member of Human Evolution and Cognition Research Group (EVOCOG), an associated unit to Spanish CSIC. He got his graduate in psychology in the University of Oviedo, followed by a master’s degree in human evolution and cognition in the UIB. In 2013, he began his scientific career collaborating with EVOCOG and with clinical research groups from the UIB to address the challenges of third-wave therapies for psychotic disorders treatment. His PhD thesis titled “Person-centered approximation to psychoses: effectiveness and outcomes” was advised by Enric Munar and Paul Chadwick, within the Doctoral Program of Human Cognition and Evolution. Currently, he is working in the attentional foundations of mindfulness applied to emotional distractors.



Antoni Gomila is a full professor of psychology at the psychology department of the UIB. His background is in philosophy (PhD in 1990, “After the computational mind”), and his interests have centered on issues at the “no man’s land” of the frontiers between philosophy of mind, psychology, and cognitive science: representation and meaning, expression and intersubjectivity, intentional explanation and rationality, always taking an evolutionary perspective into account. He is a researcher of the human evolution and cognition research group. His last book is *Verbal Minds: Language and the Architecture of Cognition* (Elsevier, 2011).