

Transitions between discrete developmental stages during language acquisition through syntax network morphospace

Lluís Barceló-Coblijn ^{1,2,3,*} Víctor M. Eguíluz ⁴ and Luís F. Seoane ^{5,†}

¹*Health Research Institute of the Balearic Islands (IdISBa), Palma de Mallorca, Spain*

²*Research Institute of Health Sciences (IUNICS), Palma de Mallorca, Spain*

³*Universitat de les Illes Balears (UIB), Palma de Mallorca, Spain*

⁴*Instituto de Física Interdisciplinar y Sistemas Complejos IFISC (CSIC-UIB), E07122 Palma de Mallorca, Spain*

⁵*BIT Lab, Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Barcelona, Spain*



(Received 2 December 2025; accepted 11 March 2026; published 18 May 2026)

Syntax is an aspect of human language that sorts linguistic structures hierarchically. It is summarized by dependency trees with words as nodes and edges reflecting syntactic subordination. By merging trees from several sentences, we obtain syntax graphs or networks, which display distinct shapes depending on whether the language capability is well formed, still developing, or pathological. Such graphs make syntactic capacity quantifiable at a systemic level. What is the structure of syntax networks during ontogeny in typically developing children? Do cognitively challenged children develop language through alternative routes? We quantify and portray typical development of syntax networks in Dutch, and find that children affected by Down syndrome, hearing impairment, and specific language impairment track the typical developmental path but halt along the way. Through a broad dataset and a thorough mathematical quantification we (1) confirm and refine a proposed sharp transition in language development, (2) correlate network traits with syntax maturation, and (3) quantify aspects that fall short in atypical development—suggesting potential diagnostic tools. We find grounds to hypothesize a gap in syntax maturation that typically developing children bridge over seamlessly, but which divides atypically developing children regardless of their condition. Our quantitative analysis enables a rigorous visualization of linguistic development trajectories, an old (yet mostly qualitative) theme in linguistics. Similar works should allow to test specific linguistic concepts (such as development trajectories) on solid quantitative grounds—as done here. Future efforts should generalize to other languages and clinical conditions, seeking patterns and universalities in language development.

DOI: [10.1103/d93c-hm6w](https://doi.org/10.1103/d93c-hm6w)

I. INTRODUCTION

Language is a cognitive capability that characterizes *Homo sapiens* [1,2], carried out by dedicated circuits of both cortical and subcortical brain regions [3–10], often referred to as the *language network*. It is primarily housed in the left hemisphere, although some evidence suggests an initial potential for language to be hosted on either side [11]. A symmetry-breaking process [12,13] must occur as the language network matures, eventually leading to salient anatomical differences between the two hemispheres. Critical periods exist during which an infant must be in active contact with language. Failing to do so may result in a seemingly lifelong inability to acquire complete language skills [14–19]. Language equips *H. sapiens* with the capacity to organize thoughts, produce

new messages, and convey acquired knowledge, doing so in a way that no other species can [20,21]. Some convergences exist between the mathematical properties of animal and human communication codes [22–24]. However, despite sustained training, and to the best of our understanding, other animals are unable to produce full-fledged language [25]. All this suggests that, while language is inherently human and seemingly a biologically wired capacity, the final stages of its development require a social context and are not inescapable.

Groundbreaking research is achieving an impressive mapping of the language network within the brain—uncovering its constituent parts, what each of them does individually, and how they come together [7–10]. While generated in the brain, language is then externalized, resulting in speech, signing, or written texts. The resulting *surface form* of language presents an intricate structure that can be extremely rich and prone to ambiguity—regarding both meaning [26–28] and how words are parsed and linked together [29,30]. Despite these intricacies, the surface form is rather instantly resolved by language users. This surface form spans several linguistic levels, from phonetics to syntax and semantics. Syntax represents the backbone that summarizes connections, within a sentence, between words (the building blocks of language) by means of hierarchical relationships. These relationships

*Contact author: lluis.barcelo@uib.cat

†Contact author: brigan@gmail.com

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.

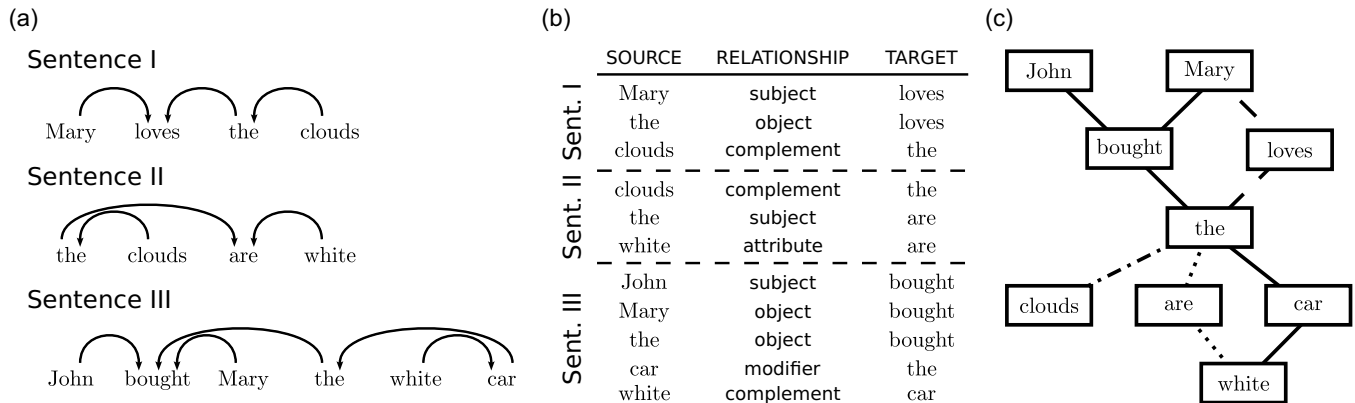


FIG. 1. Example of syntactic analysis and reconstruction of one syntax network. (a) We start out with sentences uttered by children in a session of spontaneous conversation with their parents or caretakers. Each sentence was syntactically analyzed within the Dependency Grammar framework. This results in a syntactic tree—a rather simple graph. (b) Each word-to-word relationship was collected as a line of a plain-text file. Each file contains all the data from a single conversation session and will become a network (hence a data point) for our analysis. (c) Each line in a file becomes an edge of the corresponding network. Each word that appears repeated across sentences in a single session (e.g., “the”) constitutes a single node. Every two nodes are connected if there is at least one syntactic relationship between them in that session. It is still possible to observe the syntactic tree of each individual sentence embedded within the emerging network—dashed edges for sentence I, dotted for sentence II, and solid for sentence III. The connection between “the” and “clouds” is present in sentences I and II, hence dash-dotted.

can be captured by graphs or networks. Such graphs can be reconstructed from utterances, signing, or written text. In this framework, each lexical item is taken as a node or vertex, and a link or edge connects two nodes if a syntactic dependency relation exists between them (Fig. 1).

Network theory is a very fruitful branch of mathematics [31]. A network, or graph, can recollect interactions between different components of complex systems, such as genes within a cell [32], words in a semantic space [33], species in an ecosystem [34], etc. This abstraction allows us to apply powerful computational and analytical methods derived from statistical mechanics [35], the study of graph structure [36–39], dynamical systems [40], etc.; and thus extract conclusions that might apply broadly across fields, to any networked system. Following Refs. [41,42], here we apply such abstractions to study syntax networks generated by children as they speak.

Similarly to the brain’s language capacity, syntax networks appear to mature as children grow up [41–44], typically reaching their final form within the first 4 yr. They start as broad trees, with large distances between arbitrary pairs of nodes. At some point, syntax networks become *small world* in which typical nodes are close to each other. Between these stages, the most connected words and other network properties change abruptly, similarly to a physical phase transition [41–43]. Theoretical considerations [45] offer strong grounds to believe that actual phase transitions might mediate the learning process. An enticing possibility is that the development of syntax networks reflects the maturation of the language organ within the brain. If this is the case, the readily accessible surface form of language could provide a window into the structures that neuroscientists are painstakingly examining. Thus, failures in syntactic development might reflect specific alterations in the brain circuits dedicated to language. Fully achieving such a perspective remains distant for now, but it constitutes the big picture driving this work.

We examine how syntax networks mature (as abstract mathematical objects) in typically developing (TD) children as well as in children with developmental challenges. We provide a more detailed analysis compared to earlier studies [41–44]: We investigate additional aspects (i.e., more distinct measures) that differentiate networks from one another, and using a larger empirical dataset. This allows us to tackle a series of relevant questions: What does the typical development of syntax networks (as a mathematical *graph* object) look like? Specifically so, what developmental path do they follow, from infancy to full-fledged language, across a bigger, abstract space containing all possible networks?

The notion of language developing along a specific trajectory is an old theme in linguistics [46–49]. In our big picture, anatomical alterations in the language neurological structures might result in alternative linguistic development trajectories which may or may not, eventually, become capable of indistinguishable full-fledged language. We expect such alternatives to also be reflected in the developmental path of syntax networks across the mathematical space of possible graphs [50]. We lack large databases that pair the graph shape of surface syntax with the language organ structure within the brain—since, at the moment, this neural network is reconstructed as averages across individuals [10]. However, atypically developing children (affected, e.g., by mutations, chromosomal trisomy, gene deletion or translocation, etc.) offer a narrow window of access. In such cases, the affected individuals may develop a visibly different phenotype. For example, while capable of complex communication, children with Down’s syndrome (DS) produce linguistic output that is perceived as strange or incorrect by speakers who have undergone the expected developmental steps [51].

Thus, back to questions that we can approach from our mathematical space of developing syntax graphs: Are challenged children eventually able to produce syntax networks similar to those of TD children? If not, where in the abstract

space of possible graphs do atypically developed networks dwell? How far are they from typically developed ones? Is their developmental trajectory across network space similar for all children? Bringing these questions into a quantitative framework might also offer diagnostic tools for atypical development.

In this paper, we address all these questions from a rigorous mathematical framework. We do so by studying a large database containing Dutch language produced by (1) 33 TD children at different stages of language development (i.e., of different early ages); (2) three TD children studied at various points in time (i.e., through longitudinal studies as they aged, producing 18, 17, and 14 data points each, respectively); and (3) three groups of atypically developing children, one of 20 DS children, another one of 20 children affected by hearing impairment (HI), and another one of 20 children affected by specific language impairment (SLI). From that database, we reconstructed syntax networks for each child and measured a series of properties that characterize graph structure. These measurements place individual syntax networks within an abstract mathematical space. We show how the space of TD syntax networks is shaped by a sharp shift—already identified in earlier studies [41–44]—between treelike graphs and mature syntax networks. We measure with great accuracy at what age the transition happens and the peak rate of change. Our abstract space allows us to visualize syntax network maturation and to compare it against graphs from atypically developing children. We demonstrate how atypically developing syntax tracks typical development, but stops at some point along the path—often around or right before the sharp transition. Our results hint at two different interpretations: (1) that children affected by different conditions are stopped at points along the developmental path (from earlier to later: DS, HI, SLI), versus (2.a) that some children from either condition manage to develop syntax networks indistinguishable from TD children, (2.b) that a gap separates them from those who get halted along the way, and (2.c) that all who halt do so at a roughly similar stage independently of condition. We contextualize our findings within the broader literature in Sec. IV.

II. METHODS

A. The CHILDES data bank

The CHILDES database [52] contains numerous datasets and resources for the study of language in children of different ages, clinical conditions, and tongues. A prominent corpus of data created by Bol and Kuiken [53] focuses on the Dutch language and includes both typically and atypically developing children. It also includes both longitudinal (which track the same child's progress over time) and cross-sectional (which track several children of different conditions at a similar stage in time) studies. In this paper, we restrict our study to a subset of this data collection.

Our raw data consist of transcriptions of spontaneous linguistic interactions between a child and their parent or caregiver. Each such session usually lasts between 30 and 45 min. Holophrases (a single-word utterance used by young children to express a complex idea or a full sentencelike meaning, relying on context and intonation to convey additional

information—such as “up” to mean “pick me up”), words, or sentences were recorded. Each sentence (or equivalent unit) was stored as the line of a plain-text file. Each unique session resulted in a single file, with as many lines as sentences, words (with full sentence meaning), or holophrases uttered. Each such single file was analyzed syntactically and processed (as described below) to produce a syntax network and a data point in an abstract mathematical space.

Our dataset contains 142 individual sessions (which will result in as many syntax networks and as many data points). Of these, 82 correspond to TD children, while 60 correspond to children with atypical development.

Of the 82 TD sessions, 33 of them correspond to unique sessions, each with a different child (Table III), while 49 of them correspond to three longitudinal studies that obtained several sessions from three different children (Table IV). Those from unique sessions have ages between 1;07.11 (1 yr, 7 months, and 11 days) and 3;04.20. For visualization, we divide the data into three age-based subgroups, with 11 data points each: from 1 to 2 yr, from 2 to 3 yr, and from 3 to 4 yr. Longitudinal studies were performed on three children labeled Daan (18 sessions, from 1;09.09 to 2;06.25), Abel (17 sessions, from 1;10.30 to 2;10.00), and Matthijs (14 sessions, from 1;11.10 to 2;05.26).

Data from atypically developing children (Table V) come from three groups of equal size (20 sessions each) of children with Down's syndrome (from 4;04.21 to 18;11.5), children with hearing impairment (from 3;11.14 to 9;00.15), and children with specific language impairment (from 4;01.16 to 8;01.17).

Appendix A provides summary tables and additional descriptions of each developmental condition.

B. Syntactic analysis and network reconstruction

From each of our 142 transcriptions, we produced a syntax network. First, each individual sentence was analyzed adhering to the principles of Dependency Grammar [54,55]. This procedure extracts the dependency structure between words or morphemes with each other (Fig. 1). From each sentence, a small graph (a tree) is obtained. We collapse the trees pertaining to all sentences from an individual session by assigning each word to a node, and connecting two nodes if there is at least one syntactic dependency in one of the syntactic trees of the session between the corresponding words. The resulting graph is typically much larger than the individual trees. We term this collapsed graph a syntax network, which constitutes our unit of study.

All syntactic analyses were initially conducted manually and later validated using Netlang, a corpus annotation software [56]. Principles from syntactic theory (e.g., the Determiner Phrase [57,58]) have been taken into account. The resulting syntax networks were stored in plain-text files as a list of edges. Although syntactic dependencies are directional, we ignore directionality in this study—i.e., we restrict our analysis to undirected networks. This already allows us to tackle our main research questions. Expanding our analysis to directed graphs (while out of our current scope) is a promising line of research that should allow us to extract further insights.

C. Topological measurements on syntax networks

Network science has provided us with a large tool kit to study complex networks. Many strategies are available, and our choices do not exhaust what can be extracted from this dataset. We tried to quantify as many aspects as we could think of about graphs. We include broadly: (1) sizes, such as vocabulary (i.e., number of nodes) and number of edges; (2) properties pertaining to node connectivity such as their degree or clustering; (3) properties pertaining to a relationship between nodes and the network at large, such as the graph's diameter, assortativity, or k cores; (4) characterization of cycles present in the graph; (5) maximal constraints that can be imposed onto each network, such as the coloring number; or (6) measurements of similarity and diversity between nodes, such as Jaccard similarities. In Table I, we summarize names, symbols, and literature about each of the measured properties. We provide a more thorough description of each property in Appendix B. We discuss briefly what these abstract measurements might mean for syntax. Some of these properties might seem redundant as they are correlated—e.g., some properties increase if a graph contains longer linear chains. However, each quantity presents some potential subtlety, and all of them together are more capable of describing the different ways in which a network can be complex.

D. Dimensionality reduction

For each available syntax network, we measured the 18 topological properties from Table I. These quantities characterize several ways in which graphs can be different from each other, from the obvious number of nodes (i.e., vocabulary size) and average degree (i.e., how many different syntactic relationships a word participates of), to more abstract aspects such as coloring number (which here relates to a minimal number of word classes needed to tile a network without having two words of a same class in contact) or assortativity (which gives us a hint of whether a word relates syntactically to others similar to itself).

Each graph in our dataset is represented by an array containing its 18 topological measures. All networks together populate an 18-dimensional mathematical space with a cloud of points. Many of these properties are correlated—e.g., a larger number of edges likely results in a graph with more available paths, more cycles, and reduced network diameter, etc. Hence, several dimensions may be redundant, and the data occupy a subspace of much smaller dimension. We found a representation of reduced dimensionality, a strategy in vogue across disciplines from cell biology [64], through neuroscience [65–67], to network topology [39]. A series of techniques, such as autoencoders [68], umap [69], or nonlinear manifold embeddings [70], perform this job, each with some subtleties and specific challenges. We opted for principal component analysis (PCA) [71], the most straightforward and intuitive method.

In a nutshell, PCA finds the directions across the 18-dimensional space along which syntax networks present more variability. Each of these directions is a linear combination of the measured properties. These directions, ranked in order of decreased variability, constitute the principal components

(PCs) of our data. They are given by the eigenvectors of the correlation matrix between network properties, sorted in decreasing order of their eigenvalue [71]. If the data indeed occupy a smaller-dimensional subspace, then a few PC suffice to provide a satisfying description. Some bootstrapping methods are often employed to determine the optimal number of dimensions that should be kept. We forego such analysis here, as we find that a study of the first two PC allows us to tackle our research questions.

E. Null models

We built two kinds of null models to compare our real syntax networks to. We call our null models *rewire* (Rwr.) and *random* (Rnd.). To build each of them, we took each real syntax network from TD children and produced five bootstrapped graphs.

Rwr. networks were built by randomly rewiring existing connections. These graphs were derived from uttered sentences, from which syntactic relationships were extracted. We represent connections in these networks as a list of edges given as a tuple of words (w_1, w_2), meaning that a syntactic dependency exists between them. To build a Rwr. network, we listed all edges, took all words in the first position and shuffled them, then took all words in the second position and shuffled them. With this bootstrapping, each word roughly keeps its original degree. Other properties might correlate with the original values as well.

Rnd. networks simply have the same average connectivity as their real-world counterpart. We took a TD syntax graph, then computed its average degree and produced a Rnd. graph by generating a random network of the same size and degree. The resulting graphs might present some correlation with some properties in the original graph, but we expect them to diverge more than Rwr. from their source.

III. RESULTS

We applied PCA to the 18 topological properties from Table I measured on all syntax networks—including both typically and atypically developing children. These properties characterize graph structure as thoroughly as we could, capturing aspects such as degree, centrality, clustering, navigability, cycles, etc. PCA gave us the main axes of variability across syntax networks. This means that, if atypically developing children are consistently outstanding with respect to typically developing participants, then their patterns will stand out along a PC or as trajectories within the resulting visualizations. Figure 2(a) shows syntax networks from all (except two longitudinal studies omitted here for clarity) typically and atypically developing children projected onto our first two PC (which together explain 80% of the variance in the data; see Fig. 18 in the Supplemental Material [72]).

This representation is a morphospace. Morphospaces were first introduced by Raup [73] to chart possible sea-snail shells by identifying dimensions of a 3D space with shape parameters. By moving along each dimension, shells changed in a specific way (e.g., one dimension changed length, another one increased the number of twists around the shell's axis). This way, different morphological traits become segregated. A

TABLE I. Topological properties in our analysis and speculative interpretation in linguistic terms. Lengthier explanations with more mathematical formulations can be found in the SI.

Property kind	Property name	Network-theory meaning	Linguistic interpretation	Refs.
Sizes	Vocabulary (voc.) size, $N^W \equiv V $.	Number of nodes.	Lexical productivity. Number of distinct words in a conversation.	[31]
	Connected voc. fraction, $F_{CC} \equiv V_{CC} / V $.	Fraction of nodes in the largest connected component.	Words in integrated discourse, rather than isolated productions.	—
	Number of edges, $N^E \equiv E $.	Number of edges connecting nodes.	Distinct syntactic relations instantiated during a session.	[31]
Local connectivity	Degree, $\langle k \rangle$.	Mean number of neighbors per node.	Flexibility of words to associate syntactically with others.	[31]
	Clustering, C .	Fraction of completed triangles among all existing neighbors.	Tendency for syntactic partners of a word to also combine with each other.	[31]
	Assortativity, r .	Correlation between node degree and neighbor degree.	Correlates with whether words of similar syntactic valence preferentially combine.	[59]
Global organization	Diameter, D .	Longest shortest path between nodes.	Global compactness of syntax space. How easily accessible are words through syntax association.	[31]
	2-core size, $ K_2 $.	Nodes surviving recurrent degree-2 pruning.	Presence of a minimally connected syntactic backbone (i.e., no words with one syntactic connection only).	[60,61]
	3-core size, $ K_3 $.	Nodes surviving recurrent degree-3 pruning.	Presence of a more connected syntactic backbone.	[60,61]
	Largest k core, κ .	Highest nonempty k core.	Size of largest cluster of mutually or recurrently connecting words.	[60,61]
Cycles	Cycle basis size, $ C $.	Number of independent cycles.	Alternative pathways to connect words. Richness of different syntactic structures.	[62]
	Average cycle length, L_C .	Mean length of cycles in the basis.	Typical length of syntactic structures between words that close cycles (i.e., hubs such as articles and pronouns).	[62]
Structural constraints	Maximal matching size, $ M^{\max} $.	Largest set of edges that do not share a node.	Prevalence of linear chains or treelike syntax (long sentences without syntactic repetition or cycles).	—
	Maximal independent set, $ I^{\max} $.	Largest set of nodes that do not share an edge.	Presence of peripheral lexical items (e.g., stars formed by repeated “I want X” constructions).	—
	Coloring number, q .	Minimum number of labels that can be imposed without having equally labeled neighbors.	Constraints to hypothetical word classes if we imposed that two words of a same kind are not mutually connected.	[42]
Node diversity	Jaccard similarity, s_J .	Neighborhood overlap between nodes.	Similarity of words that two syntactically related words can connect to.	[63]
	Jaccard entropy, H_J .	Entropy of Jaccard similarity distribution.	Heterogeneity in whether words connect syntactically to same words as neighbors.	—
	Pagerank entropy, H_P .	Entropy of node centrality distribution.	Heterogeneity in whether words are central or peripheral in syntax constructions.	—

similar strategy has been deployed to study a host of complex systems [74–80], including language networks [28]. Here, by moving along the horizontal or vertical axes, we find differently structured networks—from simple graphs such as stars at the bottom-left [Figs. 2(b) and 2(c)], through expanded trees or starlike motifs connected by protruding branches around

the center [Figs. 2(d)–2(g)], to very densely connected graphs with a clear core-periphery structure toward the right half of the plane [Figs. 2(h) and 2(i)].

All results in this study were elaborated including all available networks. In the next subsection, we focus on TD children. Comparative analyses follow in later subsections.

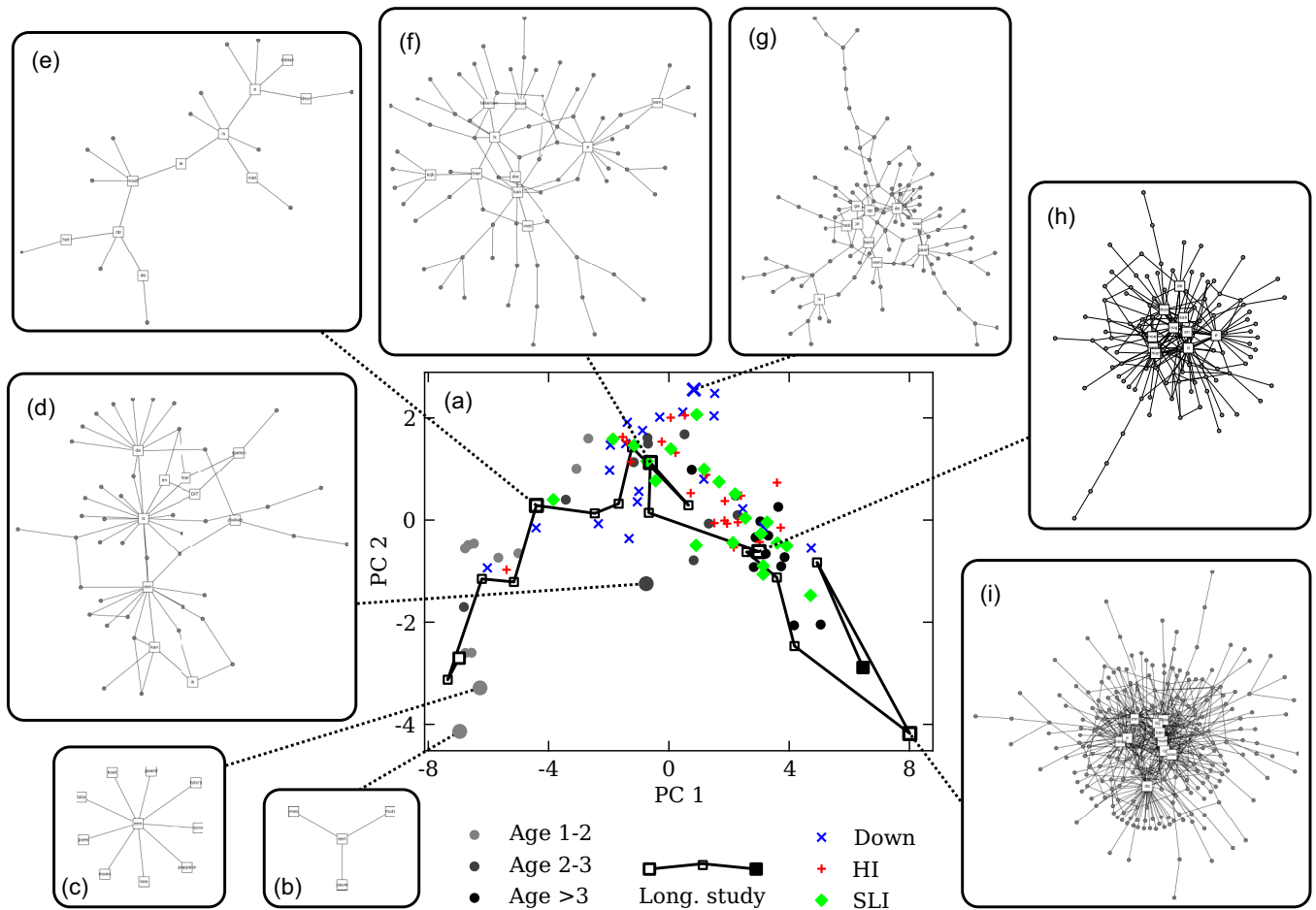


FIG. 2. The syntax network morphospace. (a) Quantitative visualization of the syntax network development trajectory. Color codes for children of different age and condition. (b)–(h) Examples of real syntax graphs across the syntax network morphospace. The location of each graph in the morphospace is marked by slightly larger symbols. Boxes mark the ten words with the most connections—all other nodes appear as filled circles. (b) TD child; age 1 yr, 8 months, and 23 days (~628 days). (c) TD child; age 1 yr, 7 months, and 11 days (~586 days). (d) TD child; age 2 yr, 6 months, and 10 days (~920 days). (e) TD child; age 2 yr, 0 months, and 4 days (~734 days). (f) TD child; age 2 yr, 2 months, and 16 days (~806 days). (g) Atypical development (DS) child; age 18 yr, 2 months, and 29 days (~6659 days). (h) TD child; age 2 yr, 4 months, and 0 days (~850 days). (i) TD child; age 2 yr, 4 months, and 0 days (~905 days). Networks [panels (e), (f), (h), and (i)] are from the same child.

A. A sharp transition shapes the morphospace of typically developing syntax networks

Figure 3(a) reproduces the morphospace but including only all TD children. Any position in the plane of Figs. 2(a) and 3(a) may correspond to one (or many) graph(s), with graphs close in the plane likely having a similar shape. Syntax networks occupy this space unevenly, meaning that they do not realize all possible graph structures. In other words, syntax networks have a specific shape among all possible graphs. Our data trace an arc from left to right, with the age of children increasing throughout. This is a quantitative visualization of the trajectory of typical syntactic capacity development—i.e., of how syntax networks change shape as a child develops full-fledged language.

The developmental trajectory becomes clearer when PC-1 is plotted as a function of age. This component captures 70% of the variance in the data. Figure 3(b) shows that PC-1, for typically developing children, grows more or less

monotonously—as we would expect from network properties that increase over time until they saturate at their *mature* value.

TABLE II. Fit of first PC to sigmoid Eq. (1). The mid column shows parameter estimates and uncertainties for fits to all typically development data. The rightmost column shows the same for fits to longitudinal studies only.

Parameter	All TD	Longitudinal TD
A	10.21(80)	15.3(2.9)
B	−6.41(62)	−8.8(2.0)
$A + B$	3.8(1.4)	6.5(4.9)
$t_{1/2}$ (days)	795.5(9.2)	796(17)
$t_{1/2}$ (yr)	2.179(25)	2.181(47)
k (days ^{−1})	0.0241(47)	0.0152(42)
$Ak/4$ (days ^{−1})	0.061(52)	0.058(78)
$100 \cdot k/4$ (%·days ^{−1})	0.60(12)	0.38(10)

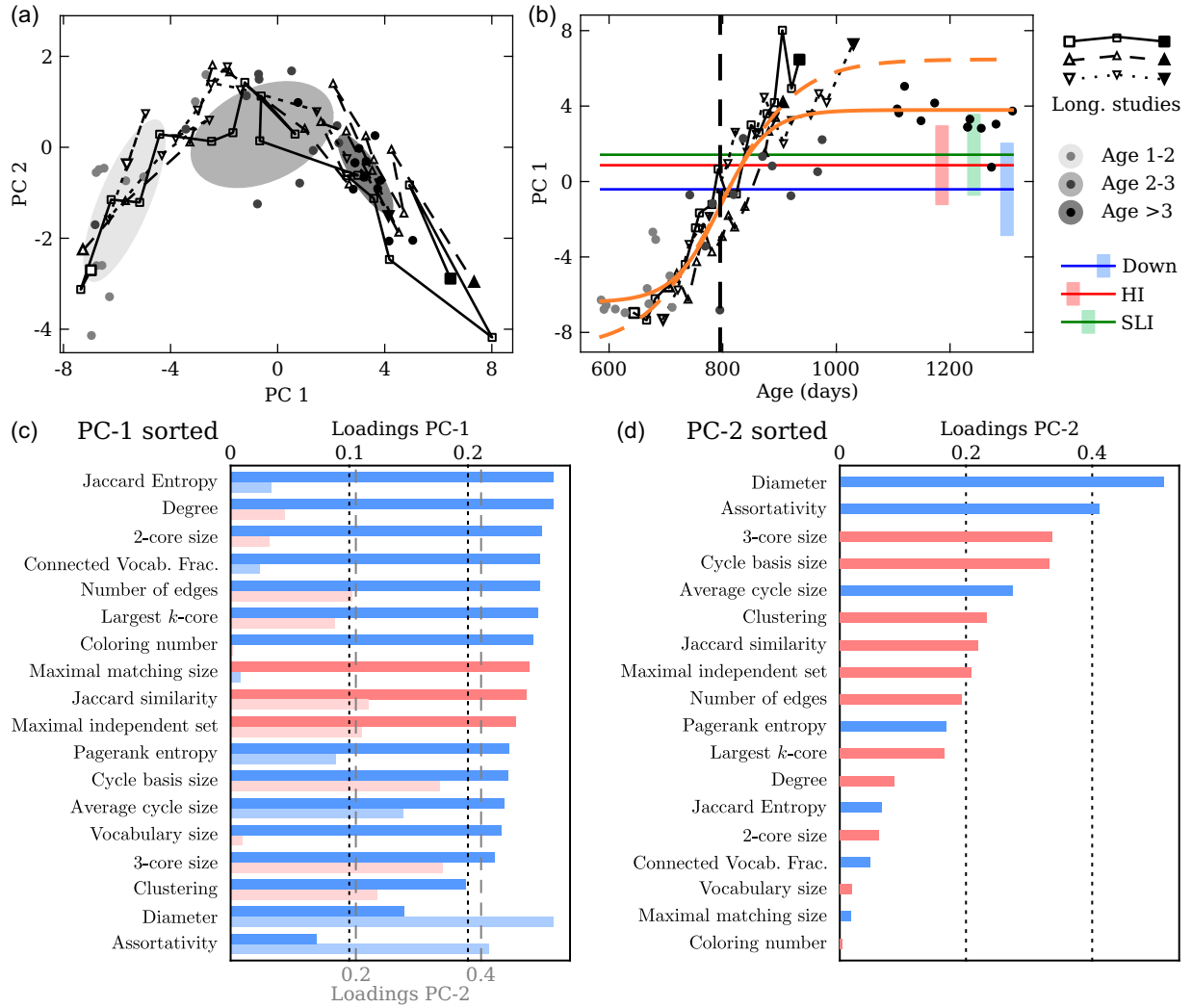


FIG. 3. Analysis of TD children. (a) First two PC of the syntax morphospace with graphs from typically developing children alone. Ellipsoids are centered on the mean PC-1 and PC-2 for age groups 1–2, 2–3, and >3 yr from lighter to darker gray. Ellipsoid axes align with, and are proportional to, the variation along the principal directions within each group—i.e., they depict dispersion within each group. Networks from the same longitudinal study are joined by solid, dashed, or dotted curves. (b) Time evolution of the first PC for typically developing children. Horizontal lines and shaded rectangles indicate the mean and standard deviation of PC-1 values in atypically developing children. Orange curves are the result of fitting sigmoids Eq. (1) to PC-1 values of all typically developing children (solid curve) or only of longitudinal studies (dashed curve). Fit parameters with estimated uncertainties for all data are found in Table II. (c) Absolute values of loadings of each property in PC-1 (darker) alongside those in PC-2 (lighter), sorted in decreasing order of PC-1 loadings. Color indicates sign: blue positive, red negative. (d) Loadings of each property in PC-2 sorted in decreasing order.

We expect a similar behavior, e.g., for the number of words uttered within a conversation (Fig. S3(a) in the Supplemental Material [72]). Since sessions are finite, there is an average maximal number of words in any mature conversation. Figures S3–S9 of the Supplemental Material [72] show the temporal evolution of all properties for typically developing children. It is illuminating to see which network aspects follow the same saturating behavior as PC-1. Noteworthy among them are the fraction of vocabulary in the network’s largest connected component, the number of edges (indicating that mature syntax networks reach an average maximal number of syntax relationships within finite conversations), the degree (closely related), k -core sizes, or the average size of cycles (suggesting that typical phrases reach a *mature* size as well).

Our PC-1 traces a much neater saturating curve than any individual property. This suggests that the main axis of variation across syntax networks captures the maturation process. Fluctuations (expected and observed in individual properties) are smoothed out in PC-1. The resulting trajectory is relatively sharp as well. Orange curves in Fig. 3(b) are fits of sigmoids with the form

$$\sigma(t) = \frac{A}{1 + e^{-k(t-t_{1/2})}} + B \quad (1)$$

to PC-1 values from all data in TD children (solid) and to longitudinal cases only (dashed). $A + B$ represents the saturating value of PC-1; $t_{1/2}$ is the time at which the maturation process reaches its midpoint, which is also when development happens at its fastest rate; and $Ak/4$ is the slope of the curve

TABLE III. Summary of TD participants, individual sessions. Data collected at 1 (left), 2 (center), and 3 (right) yr of age. Age format is year;month.day. Participants are sorted by increasing age.

1–2 yr	2–3 yr	3–4 yr
Barry 01;07.11	Dieuwertje 02;00.12	Bert 03;00.12
Tijn 01;07.16	Katja 02;01.09	Katja 03;00.15
Marco 01;07.20	Marco 02;01.22	Maaïke 03;00.25
Corine 01;08.06	Tijn 02;02.05	Floor 03;01.24
Gea 01;08.23	Channe 02;02.29	Frederique 03;02.18
Thom 01;10.02	Karel 02;03.16	Marcel 03;04.16
Ronnie 01;10.05	Rianne 02;04.20	Otje 03;04.20
Els 01;10.12	Els 02;05.07	Chantal 03;05.10
Suzanne 01;10.17	Maaïke 02;06.10	Jelle 03;05.28
Rogier 01;11.16	Floor 02;07.27	Martijn 03;06.06
Marieke 01;11.12	Frederique 02;08.05	Bert 03;07.05

at $t_{1/2}$ (i.e., the maximum rate of change). Table II shows fit results for all parameters with uncertainties.

Such a sharp development has been compared to a *phase transition* in the topology of maturing syntax networks [41,42]. Our work corroborates this result: Language development is sudden in infants and saturates to a completion—which should be similar to its adult form. Our inclusion of more topological properties allows us to obtain better defined transition curves than before—thus to better estimate the transition point, $t_{1/2}$. We locate it at $t_{1/2} \simeq 2.179 \pm 0.025$ yr (i.e., 2 yr and 65.3 $\pm$ 9.1 days) for all TD data and at $t_{1/2} \simeq 2.181 \pm 0.047$ yr (i.e., 2 yr and 66 $\pm$ 17 days) for longitudinal studies only. These values are within uncertainty ranges, even though children from longitudinal studies saturate at higher PC-1 values. This higher saturation value might relate to

TABLE IV. Summary of TD participants, longitudinal studies. Data for Daan (left, 18 sessions), Abel (center, 17 sessions), and Matthijs (right, 14 sessions) over multiple observations. Age format is year;month.day. Participants are sorted by decreasing number of sessions.

Daan	Abel	Matthijs
01;09.09	01;10.30	01;11.10
01;10.01	01;11.26	01;11.24
01;10.16	02;00.11	02;00.09
01;11.21	02;01.02	02;00.24
02;00.04	02;01.16	02;01.07
02;00.22	02;02.19	02;01.21
02;00.29	02;03.02	02;02.09
02;01.21	02;03.23	02;02.20
02;02.02	02;04.09	02;03.01
02;02.16	02;04.23	02;03.19
02;03.04	02;05.06	02;04.24
02;04.00	02;05.27	02;05.01
02;04.14	02;06.11	02;05.13
02;04.28	02;07.15	02;05.26
02;05.11	02;07.29	—
02;05.25	02;08.13	—
02;06.11	02;10.00	—
02;06.25	—	—

increased engagement, since these children were repeatedly interviewed under similar circumstances. The maximal development rate, $Ak/4$, is also quite close—at 0.061 ± 0.052 and 0.058 ± 0.078 , respectively, within uncertainty ranges. This is a maturation rate in the abstract units of PC-1. It becomes more interpretable as a percentage over the total change from the earliest graphs to maturity, $100 \cdot k/4$. When language is developing at its fastest, syntax progresses around $\simeq 0.60\%$ each day (0.38% for longitudinal studies, outside uncertainty ranges). This quantitative characterization supports the presence of a cognitive phase-transition-like behavior that is the same for all TD children—despite possible differences in engagement.

Figure 3(c) shows the loadings of PC-1. Loadings tell us how much each original measured property (Table I) correlates with PC-1, which seems to capture development. We find that almost every variable has very large loadings, meaning that most network traits follow a monotonic trend toward their mature value. Three of these properties (maximal matching size, Jaccard similarity, and maximal independent set) decrease monotonically, meaning that their values decrease as syntax graphs mature. For Jaccard similarity, this means that, in early syntax networks, words have more similar neighborhoods than later (see Appendix B for definitions and interpretations of topological properties).

While all topological traits are rather monotonic, PC-2 is not—neither with respect to PC-1 [Fig. 3(a)] nor to time (Fig. S2(b) in the Supplemental Material [72]). This means that, while the development process (and, specifically, the maturation of most topological properties separately) is rather straightforward, there is a combination of factors in the data that starts with some low value, peaks during maturation, and then wanes. This raise-and-fall signature is the second most important feature of the data (since it is captured by PC-2, which explains 10% of all variance in the data), but it could easily be missed if we were to look at any individual property alone.

If we think about the maturation process portrayed by a sigmoid in PC-1, such raise-and-fall behavior is displayed precisely by its derivative: It is small for $t \ll t_{1/2}$ and $t \gg t_{1/2}$ and it peaks at $t_{1/2}$. It is an enticing possibility that PC-1 shows the level of syntax maturation, while PC-2 shows a development rate. But our data did not enable us to establish this. A limitation is that PC-2 over time is much noisier than PC-1 (Figs. S2(a) and S2(b) in the Supplemental Material [72]).

Figure 3(d) shows the loadings of PC-2. This component is made up mainly by the two properties least correlated with PC-1 (diameter and assortativity) plus combinations of all other monotonic properties with changing signs (which is needed to craft something that rises and falls out of similar monotonic trends).

Network diameter, the property with largest loading on PC-2, is particularly interesting (Fig. S5(a) in the Supplemental Material [72]). It starts out necessarily small, as syntax networks begin with few words. It grows monotonically as children utter more sentences. But then it stops growing and even decreases slightly (which is difficult to appreciate in Fig. S5(a) of the Supplemental Material [72]). This behavior becomes more clear if we normalize by the number of nodes in the largest connected component—i.e., if we present the

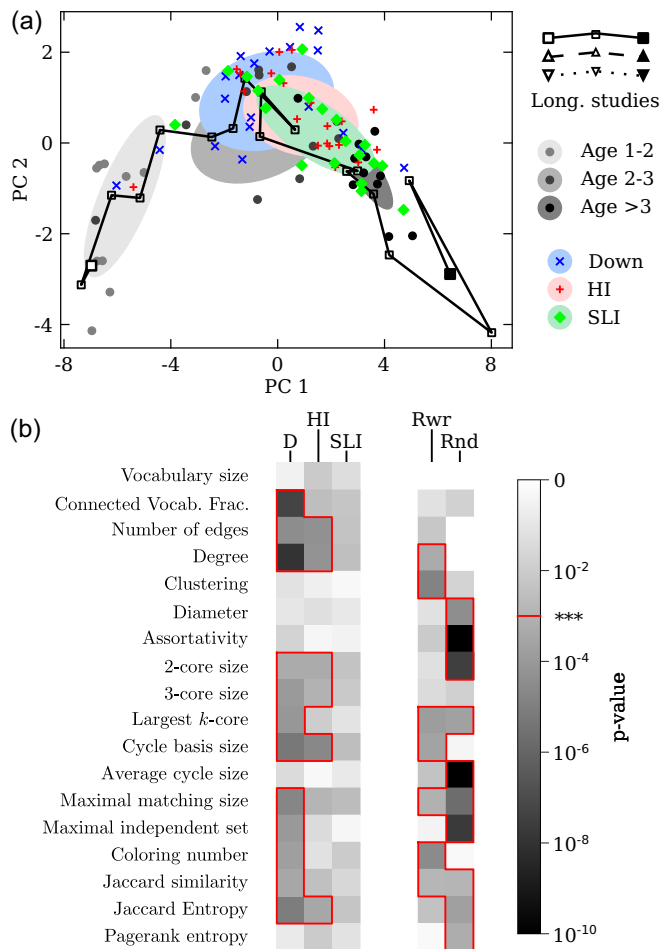


FIG. 4. Analysis of atypically developing children. (a) Same as in Fig. 3(a), but including data from atypically developing children. The added ellipsoids are centered on the mean PC-1 and PC-2 for children affected by Down syndrome (blue), hearing impairment (red), and specific language impairment (green). Their axes align with, and are proportional to, the variation along the principal directions within each group. (b) Reports of p -values from a t -test comparing each graph property from each atypically developing group to the same property in the group of the oldest, typically developing children (ages 3–4). Properties with a p -value below 10^{-3} are enclosed in red. Same tests are reported for null model networks Rwr. and Rnd.

diameter as a ratio to the number of connected vocabulary (Fig. S10 in the Supplemental Material [72]). We see that this number drops drastically after the transition point. Beyond this point, syntactic networks are able to grow without increasing their diameter. In other words, while vocabulary grows, the graph remains well connected, accessible, and easy to navigate.

B. Atypical development tracks, and stops along, the typical development trajectory

One possibility was that atypically developing children might follow alternative maturation paths, and that they might end up in different corners of the morphospace. This does not happen. Figure 4(a) shows almost all syntax graphs, including

those from atypically developing children (two developmental studies are omitted again for clarity). Data from atypically developing children are distributed close to and alongside the TD trajectory—i.e., they retrace the developmental trajectory of TD children. But they do not accumulate at the end of the arc, hinting that atypically developing children might cease maturing at different points along the way. Ages of HI and SLI children are between 5 and 10 yr, while most DS children are well into their teens (Table V). Compared to TD children, they had more time to reach their final linguistic stages—but no guarantee exists that they did. Colored ellipsoids in Fig. 4(a) are centered at the mean PC-1 and PC-2 values of each group. DS children (blue) halt their progress earlier than HI children (red), who in turn stop slightly earlier than SLI children (green).

Figure 4(b) compares each graph property in atypically developing children to the same property in the TD group with the oldest children (ages 3–4). The gray scale shows p -values of a double-tailed t -test from each comparison, with $p < 10^{-3}$ marked in red. The smaller a p -value, the less likely that the differences between atypical and typical syntax networks are due to chance. We must be careful with analyses based on p -values, but for economy of language we will say that whenever $p < 10^{-3}$ then two properties are distinguishable between typical and atypical children. If $p > 10^{-3}$, we will say that property is indistinguishable or similar in both groups, and that differences are likely due to chance.

The SLI group presents a p -value larger than 10^{-3} for all properties. This group is more similar to TD children. Differences between SLI and TD children aged 3–4 yr are perhaps due to chance. In turn, HI children present certain properties with $p < 10^{-3}$, and DS children present those same properties and a few additional ones. This recovers the order in which each group halted along the development trajectory.

It is informative to study what properties differ one by one (see Appendix B for definitions and linguistic interpretations). All children reach similar vocabulary sizes, but only DS speakers fail to integrate a similar fraction in the largest connected component. Their linguistic output contains more disconnected structures. Both DS and HI children have fewer edges and a lower average degree—they form fewer syntactic connections—than other children. 2- and 3-cores are typically smaller in both DS and HI (even though HI children have similar largest k cores than TD children). Hence, DS and HI children often fail to establish dense syntactic cores. Their networks contain more terminal branches (linear chains that cannot form k cores). This is consistent with reduced size in cycle bases also in both DS and HI, although the average cycle length is indistinguishable—i.e., they close less cycles but those completed are similar in length to SLI or TD cycles. Only DS children have different constraint properties from others (larger maximal matching and independent sets, smaller coloring number). This is again compatible with networks with terminal branches or linear chains. Interestingly, DS children also have higher Jaccard similarity than others, suggesting that DS nodes connect to more similar pools of words than in other children. Finally, again both DS and HI children present a lower Jaccard entropy.

We generated null models by shuffling all the connections in a given real graph, Rwr., or by producing graphs that had the

TABLE V. Summary of participants with atypical development. Data collected from children with DS (left), HI (center), and SLI (right). Age format is year;month.day. Within each column, participants are sorted by increasing age.

Down's syndrome (DS)	Hearing impairment (HI)	Specific Language Impairment (SLI)
Marcel 04;04.21	Thomas 03;11.14	Rinanda 04;01.16
John 08;01.27	Nita 04;05.12	Pierre 04;07.20
Chantal 08;04.07	Berry 04;07.20	Renzo 04;08.21
Tineke 08;04.07	Nanda 04;11.28	Monique 04;09.08
Fiona 10;05.03	Bjorn 05;02.27	Wilma 05;01.02
Sven 10;08.10	Judith 05;06.18	Diana 05;01.04
Sanne 13;00.02	Jenny 05;08.23	Lieneke 05;03.07
Antoinette 13;05.22	Wim 05;09.28	Pascal 05;04.28
Bert 13;08.17	Danielle 05;10.28	Mark 05;11.22
Iris 15;01.09	Jelle 05;11.19	Bertus 06;00.10
Pieter 15;02.07	Sander 06;02.08	Pim 06;00.13
Eline 15;06.16	Charlotte 06;03.20	Hessel 06;00.24
Ronnie 16;03.16	Jeroen 06;05.23	Jelle 06;01.13
Erik 16;07.17	Sander 06;08.27	Joep 06;01.26
Anita 16;10.01	Mike 06;09.25	Ramon 06;02.10
Martijn 17;01.09	Corine 06;11.17	Joost 06;07.22
Michael 18;02.29	Merel 07;02.12	Pjotr 07;00.18
Marius 18;05.02	Marco 07;09.13	Sjouke 07;01.26
Lisa 18;07.18	Kalinka 08;08.11	Kees 07;04.19
Katrijn 18;11.05	Peter 09;00.15	Lennart 08;01.17

same average connectivity as real syntax networks, Rnd. (see Sec. II E). The Rnd. null model checks whether syntax networks occupy the morphospace in a similar fashion to rather arbitrary graphs. Figure S1 in the Supplemental Material [72] shows that this is the case early along the developmental trajectory, until a bit before the transition point. Afterward, Rnd. graphs diverge from empirical and Rwr. networks. This indicates that, as it matures (but not necessarily earlier), syntax instantiates a specific class of graphs within all possible arbitrary networks.

We can also think of null models as specific kinds of pathological syntax networks. It is then interesting to compare which properties differ with respect to those of mature syntax graphs [rightmost columns, Fig. 4(b)] These p -values indicate that null models are different from mature syntax networks in distinct ways from networks from atypically developing children, showing that the morphospace of possible graphs is more varied than that explored by syntax.

Also, compare this to the differences between mature TD and those of atypically developing children. The set of distinguishable properties in the latter case hints us at a hierarchy: SLI is indistinguishable, HI differs in certain properties, and DS differs in a larger set of properties that contains those of HI. If, similarly to null models, the set of distinguishable properties for HI would not be contained within that of DS (and vice versa), that would suggest orthogonal linguistic phenotypes. Instead, we observe a parsimonious degradation of graph properties, suggesting a progression $DS \rightarrow HI \rightarrow SLI \rightarrow TD$.

C. Discrete developmental stages

The distribution of data points seems irregular along the syntax development trajectory [Fig. 5(a)]. They are scarce in the first third, but denser both near its peak and at a point along

its final third. This aligns well with both classic [46,81] and network-based [41,43,44] studies on language acquisition that strongly suggest discrete developmental stages. We compare our findings to those of the previous studies in Sec. IV.

To recover discrete developmental stages, we fitted a k -means model (with $k = 3$) to our TD children data [Fig. 5(b)]. While only PC-1 and PC-2 are plotted, we used all components for this fit. We also tried out variations: fitting only to PC-1 and PC-2, including all typically and atypically developing children, fitting to atypically developing children only. Our results were consistent in all cases.

Our 3-means [orange circles in Figs. 5(a) and 5(b)] identify the scarcely sampled first third of the development trajectory (Stage I), the densely populated region near the trajectory's peak (Stage II), and the other densely populated region toward the end (Stage III). Stage II corresponds to a point right before the moment of fastest development, $t_{1/2}$. Stage III corresponds to a point on the fully-matured side of the trajectory.

While these three stages were derived from typically developing children, atypically developing participants fell seamlessly into the two final phases (save one DS and one HI child stuck in the first one). They show similar accumulation points right around the peak of the trajectory and after it. Children at each stage become indistinguishable (in a statistical sense) from TD children at the same phase. This is so for children of any condition (DS, HI, and SLI).

This suggests an alternative interpretation of our results. In the previous section, we assumed a smooth development along a continuous trajectory. But phase transitions are discrete. Appearances of continuity come from noise or flimsy transitory states caught as they are traversed irreversibly. Figure 5(b) suggests that children of any condition (typically or atypically developing, and among them those with either impairment) can instead be grouped according to whether they managed to

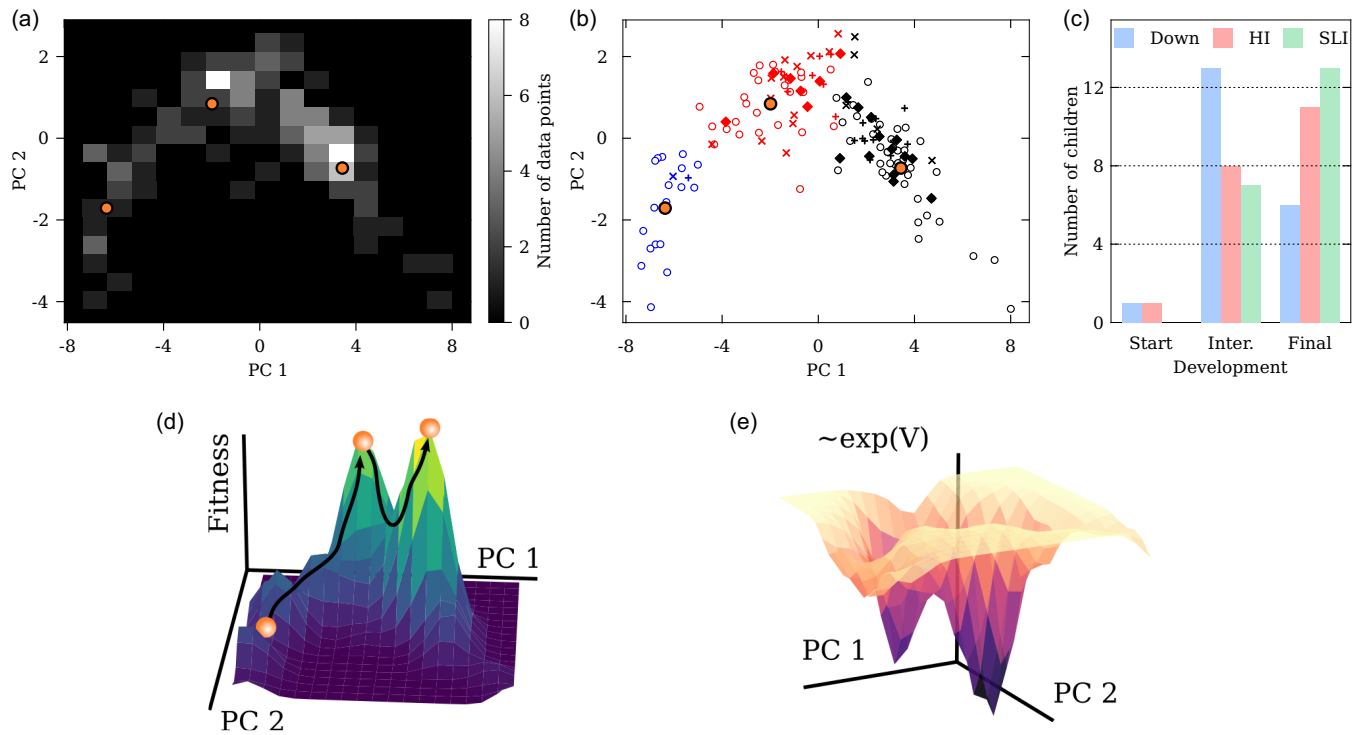


FIG. 5. Identifying discrete developmental stages. (a) A histogram with the number of graphs found along the developmental trajectory suggests discrete maturation stages. Orange circles indicate the centroids of a 3-means model fitted to the data. (b) All data points are colored according to the 3-means model. Empty circles are typically developing children. Atypically developing children are shown with their original markers: crosses (DS), plus signs (HI), and diamonds (SLI). (c) Number of atypically developing children in each 3-means-derived stage. Approximation of a fitness landscape (d) and an energy potential (e) from empirical data. These plots are for illustrative purposes only—They are not mathematically rigorous, but they convey these concepts and were built from the actual empirical data. The fitness landscape consists of the histogram in panel (a) smoothed with a Gaussian filter of unit standard deviation in every direction. Balls rolling over the fitness landscape represent a made-up child following a developmental trajectory. The energy potential (rather, its exponential) is the negative of the fitness landscape.

reach the full-fledged stage or not, with some obvious noise and variation.

According to this interpretation, each DS, HI, or SLI child would not halt at different lengths along the development trajectory. Rather, individual children of each condition either move from Stage II to III, or not. Figure 5(c) shows the number of children from each condition at each stage. It seems more difficult for DS children to make the final transition, while this is easier for HI and SLI children. Since there are more DS children stuck in the second stage, when averaging over all DS participants, it seems that they, as a whole, halt earlier and at a gradual point along the developmental trajectory. Due to the fraction of children who move on to Stage III for HI and SLI children, these groups as a whole seem to halt at different gradual positions along the trajectory—even at the purported boundary between Stages II and III.

Our data are suggestive of these two different interpretations: either (1) halting development at gradual points along a rather continuous developmental trajectory or (2) a discrete developmental trajectory with attractors at either side of a phase transition, with a certain appearance of continuity due to noise. Our current data do not allow us to decide between the two interpretations. The boundaries between Stages II and III in Figure 5(b) are blurred. This is likely because of some of the mechanisms suggested above that can make

discontinuous phase transitions look smooth (i.e., noise or transients caught in the act). Our second interpretation, which proposes three discrete developmental stages, is more appealing. It also aligns better with the existing literature on language acquisition and corroborates the findings of other studies on network science applied to syntax development.

IV. DISCUSSION

We analyzed the largest database to date to quantitatively chart the developmental trajectory of human syntax networks. Our study produced a morphospace of syntax graphs—i.e., a mathematical space in which the structure of different networks maps onto points in a plane, and moving along specific directions sees certain traits of syntax graphs vary [Fig. 2(a)]. For example, moving from left to right changes: (1) from smaller to bigger networks; (2) from nodes (words) with less neighbors on average to others better connected, which are also part of denser cores; (3) from graphs with less and shorter cycles to others with more and longer loops, etc. Moving from bottom to top shifts from (1) networks with smaller diameters (whether they have less nodes, at the left of the plane; or more, but are tightly connected, at the right) to others with longer diameters; (2) from more assortative graphs to others that are more disassortative, etc.

Empirical syntax networks do not occupy this morphospace uniformly. They trace an arc that rises and falls. The age of TD children increases along this arc—suggesting that the main morphospace dimension, PC-1, tracks development. Indeed, PC-1 over time unfolds like a sigmoid [Fig. 3(b)], with a sharp increase at some time, $t_{1/2}$, and saturating for $t \gg t_{1/2}$. This is expected of traits that mature to completion. The sigmoid sharpness is consistent with a phase transition in the topology of syntax networks, as identified in earlier works [41–44]. We put forward the hypothesis that, while PC-1 traces accumulated maturation, PC-2 might be capturing the rate (as change per unit time) at which this maturation process takes place. However, noise in PC-2 over time prevents us from establishing this on a sound basis.

An outstanding aspect of our work is its quantitative power. We measured $t_{1/2}$ to a remarkable accuracy ($t_{1/2} \simeq 2.179 \pm 0.025$). That is, we pinpointed the transition point (which parents often identify as their children become suddenly fluent) to 2 yr and 65 days; and measured that, at that time, children progress about 0.6% of the total maturation process each day. This is unprecedented in the literature, which often assigned linguistic phases loosely (e.g., McNeill’s fall suspiciously into integer or half years). This quantitative basis should enable linguistic theories to be tested rigorously and should be expanded upon.

We also studied 20 children affected by Down’s syndrome, 20 affected by hearing impairment, and 20 affected by specific language impairment. People affected by these conditions often produce linguistic output identified as odd by TD speakers. A possibility was that they would produce radically different syntax networks, perhaps inhabiting remote morphospace corners. This does not happen. All nontypical groups appear to fall close along the TD developmental trajectory, and seemingly halt their development at different points along this process. The order in which each group halts depends on the clinical condition: The DS group stops earlier than HI, which in turn stops earlier than SLI.

An enticing possibility is that syntax networks reflect maturation of the brain’s language organ—hence that our syntax graphs serve as proxies for syntactic endophenotypes [50]. Then, length along the development trajectory should correlate with differences in brain structure—which exist, as evidenced by scans on clinical populations [82,83]. Studies pairing surface syntax of individuals with neuroimaging of their language organ (currently a technical challenge) are needed. We looked at network properties that differed in atypically and TD children, suggesting diagnostic criteria. These differences must stem from the distinct ways in which children build sentences. DS children use mechanisms that are less common in TD children [44]. They abuse coordination, while typical syntactic hubs (determiners, pronouns) fail to establish enough connections. This might result in fragmented graphs (DS children have less connected vocabulary, less average edges, and lower degree; among others). The abuse of coordination (e.g., long sentences connected by “and”) might result in lengthy terminal branches sprawling out (explaining discrepancies such as maximal matching and independent sets). We are unaware of coping mechanisms in HI or SLI children, but they should be explored within the context of the syntax network shape.

An alternative interpretation of our results suggests three distinct developmental Stages. From this perspective, atypically developing children would not halt at gradual lengths along the trajectory. Instead, each child either gets stuck at Stage II or jumps to Stage III, where TD children with full-fledged language dwell. Children at each Stage present similar networks, regardless of their impairment. For example, Stage III SLI children exhibit syntax graphs indistinguishable from TD full-fledged language. These discrete Stages seem evident in our data [Figs. 5(d) and 5(e)], but noise and intermediate cases are present. Further study is necessary—discrete Stages in language development touch upon hotly debated issues of language acquisition and contradicting views of genetic determinism versus Evo-Devo and phenotypic plasticity [84,85].

A classic view championed by Tomasello [86,87] posits that language acquisition is “lineal,” without discrete jumps toward its full-fledged state. But both classic [46,81] and network-based [41–44] approaches suggest discrete steps. McNeill [81] identifies: a phase consisting of holophrases likely undetected in our data, another of two-word combinations (at 1.5–2 yr, akin to our Stage I), and a more complex phase with chains of more than two lexical items (at 2–3 yr, which may fit our Stage II). Two earlier network-based studies [41,42] identified a sudden regime shift compared to physical phase transitions. This better fits our observations if these works merged Stages I and II in a phase before the transition, and if the transition captures our sudden shift to Stage III. Other network-based studies by one of us [43,44] already hinted, with less conclusive data, at three linguistic phases: a first one in which words had on average one edge or less, a second one with more than one edge but less than two, and a third one with an average degree of 2 or more. Our current analysis renders a more clear picture, accumulating evidence in favor of a discrete progression.

A recent statistical physics framework also suggests that a phase transition mediates language learning [45]. This *random language model* builds the ensemble of all context-free grammars and studies its behavior as two parameters change. Their surface form (i.e., sentences the model produces) varies abruptly as parameters cross a threshold, changing from unsorted collections of words to well-ordered (*meaning-carrying*, as argued in Ref. [45]) sequences. In the latter, symmetry is broken (some tokens always follow certain others) and aspects of human syntax emerge. In Ref. [45], this transition is discussed within the principles and parameters framework [88], according to which children are born with a general language organ compatible with a large array of grammar rules. The learning process then switches “on” rules present in the observed tongue, and sets all others “off.” This powerful framework inspired us, but we appreciate key differences. We never register a random phase—all observed utterances carry meaning. The emerging picture suggests a minimal, universal, and versatile syntactic operation present since the beginning (children form syntactic connections in Stage I). This operation would get refined or expanded as language matures. This is closer to the minimalist program [89], which aimed at substituting principles and parameters. In Ref. [45], it is already discussed how rules could be navigated in a hierarchy, moving the random language model

toward a minimal operation that becomes richer with successive symmetry-breaking choices.

A useful tool for conceptualizing phase transitions is energy potentials and fitness landscapes, which we derived informally [Figs. 5(d) and 5(e)]. In fitness landscapes, more functional syntax graphs have higher fitness. Developmental trajectories are conceptualized as balls rolling uphill. Both Stages II and III appear as fitness peaks in our landscapes, suggesting “cognitive attractors”—i.e., particularly fitter cognitive operational modes. From this perspective, escaping the attractors would be difficult due to valleys that need to be traversed between working configurations.

Adopting a modern Evo-Devo viewpoint, hoping to ignite debate, we launch two conjectures. First, similarly to how embryo development often retraces earlier evolutionary steps, we suggest that language organs might revisit previous evolutionary linguistic phases as they mature. Then, the “cognitive attractor” of Stage II could resemble an earlier linguistic phase [90], more primitive than our present full-fledged capacity. Its fitness peak would be a fossil of sorts. The debated possibility of a proto-language comes to mind [91,92]. As suggested by Bickerton [91], the study of infants would be an opportunity to understand language evolution. Second, Pere Alberch’s “logic of monsters” posits that evolutionary and developmental constraints surface in pathological phenotypes. It is compelling that atypically developing children who cannot reach Stage III seem captured by the Stage II attractor *irrespective of their condition*, suggesting an intrinsic logic to how syntax might go awry.

We focused on analyzing the first two PC, but our data are not exhausted. We could explore differences in the next PC, but their importance dwindles with rank (Fig. S18 in the Supplemental Material [72]). The pioneer work on syntax network ontogeny [41] studied how certain words progress to become network hubs. We have not looked at individual words. Such analysis should cast a light on specific syntactic mechanisms along the developmental trajectory and pathology. Finally, we dealt with Dutch, assuming that our results are general. This should be confirmed with similar studies on other tongues.

ACKNOWLEDGMENTS

L.B.-C. was supported by Grant No. PID2021-128404NA-I00 funded by MICIU/AEI/10.13039/50110001103 and by ERDF/EU. V.M.E. received partial support from project CSxAI (PID2024-157526NB-I00) funded by MICIU/AEI/10.13039/501100011033/FEDER, UE. L.F.S. received funding from the Occident Foundation (Grant No. FJSCNB-2022-12-B) and from the Spanish State Research Agency, AEI, through Project No. PID2023-153225NA-I00 funded by MICIU/AEI/10.13039/501100011033 and ERDF, EU. This work was started at IFISC, where LSF was benefiting from a fellowship at the María de Maeztu Program for Units of Excellence in R&D (MDM-2017-0711).

DATA AVAILABILITY

Our study is based on public data available in Ref. [53].

APPENDIX A: DETAILS OF PARTICIPANTS AND ATYPICAL DEVELOPMENT CONDITIONS

Down’s syndrome occurs in 1 in 700 live births. It is a syndromic condition of known etiology. The most typical genotype of DS carries an extra copy of chromosome 21 (Trisomy 21). Other profiles, such as translocation and mosaicism, might result in DS but are less common. It has been extensively studied how DS influences language development [53,93–95]. Several studies point to a less affected vocabulary and a more impaired syntactic capability [96].

Language development in children affected by hearing impairment is inevitably affected, since the input produced by others is not processed under the same conditions as in TD children [53,97]. Speakers with HI have been observed to experience difficulties in processing both sentential syntax and morphosyntactic structures [97,98].

Specific Language Impairment [99] (also known as Developmental Language Disorder) has been extensively studied since the 1980s and is one of the most researched clinical conditions in the field of clinical linguistics. SLI is a syndrome observed in children who exhibit atypical ontogenetic linguistic development without any apparent nonlinguistic causes. These causes could include, for example, neurological dysfunction, general mental or cognitive delay, hearing impairment, or inadequate or insufficient exposure to linguistic stimuli due to the socio-educational characteristics of the environment in which SLI children are raised [100].

APPENDIX B: DETAILED DESCRIPTION OF TOPOLOGICAL PROPERTIES

Let us introduce some useful notation: A graph or network, $\mathcal{G}$, consists of a set, V , containing $N^n \equiv |V|$ nodes or vertices $v_i \in V$; and a set, E , containing $N^e \equiv |E|$ unordered tuples of the kind $(v_i, v_j) \in E$ that indicate that nodes v_i and v_j are connected. We call the elements of E edges or links indistinctly. We will say $(v_i, v_j) \in E$ if either $(v_i, v_j) \in E$ or $(v_j, v_i) \in E$. For simplicity, we might note nodes v_i and v_j as just i and j at odd times—this notation will be clear from the context. It is convenient to introduce the adjacency matrix: $A \equiv \{a_{ij}; i = 1, \dots, N^n; j = 1, \dots, N^n\}$ with $a_{ij} = 1$ if $(v_i, v_j) \in E$. It is also convenient to introduce the neighborhood of a node: $V_i \equiv \{v_j \in V, (v_i, v_j) \in E\}$. In this paper, we only work with unweighted, undirected graphs with no self-loops [hence, $(v_i, v_i) \notin E$ for any i].

(1) Vocabulary size, N^W : A single graph, $\mathcal{G}$, extracted from a single conversation session with one of the participants will contain as many nodes as unique words were uttered. The vocabulary size is the number of nodes in $\mathcal{G}$: $N^W \equiv |V|$.

(2) Connected vocabulary fraction, F_{CC} : During a session, a child might utter words that constitute a sentence of their own—e.g., an interjection, or an isolated “yes.” Such particles have no syntactic dependencies on any other words. Unless they are uttered within longer sentences in the same session, as we build the network they will appear as isolated nodes. Similarly, a sentence might contain words that are uttered only once throughout a whole session. This would result in an isolated graph, a tree, disconnected from all others. These are ways in which a syntax graph might possess several connected

components. Let $\mathcal{G}_{CC}$ be the largest connected component of a graph. Let $V_{CC} \subset V$ be the number of nodes in the largest connected component. Then, $F_{CC} \equiv |V_{CC}|/|V|$ is the fraction of the vocabulary that belongs to the largest connected component.

(3) Number of edges, N^E : Note that, the way we constructed our graphs $\mathcal{G}$, we ignore repeated syntactic associations. The number of edges in $\mathcal{G}$, $N^E \equiv |E|$, then measured the unique distinct syntax relationships instantiated during a conversation session.

(4) Diameter, D : Between any two words in a syntax network (say nodes $v_i, v_j \in V$) we can find the shortest path connecting them. Let us call $d(i, j)$ the distance of this path. The diameter of a network is the largest shortest distance between any two nodes, $\min_{\{i, j\}}(d(i, j))$. Paths between distant words are made up of chains of valid syntactic relationships. These, when pieced together, make up sentences. Often, a network diameter gets smaller when the graph is dense and when there are several alternative ways to get from one node to another. In our context, this would mean several valid chains of syntactic relationships—i.e., a dense and richly webbed route map of possible, grammatical sentences. This indicates a richness of syntactic combinations. Opposite to this, when a diameter is large, it means that some nodes can be reached through a unique and long path. This, for us, means that the corresponding words are just connected by a singular sentence, indicating somehow poor syntactic combinatorics.

(5) Degree, $\langle k \rangle$: The degree of a node, k_i , is the number of other nodes to which it is connected, $k_i = \sum_j a_{ij}$. In our context, this refers to the number of unique syntax relationships that the corresponding word is involved in. To characterize the network at large, we are interested in the average degree, $\langle k \rangle \equiv \sum_i k_i / N^W$. Thus, a larger average degree indicates that words are more flexible, involved in more syntactic relationships than in networks with a smaller average degree. This again indicates us the combinatorial richness of the syntax behind a network.

(6) Clustering, C : The clustering coefficient of a network is the fraction of complete triangles among all possible triangles, $C = \sum_i \sum_{j, k \in v_i} a_{jk} / (2k_i(k_i - 1))$, where $j, k \in v_i$ are all couples of nodes in the neighborhood, v_i , of the i -th node. A triangle is completed when two nodes (say j and k) that are neighbor of a third one (say i) are also connected to each other—hence, the set E contains all three edges (i, j) , (j, k) , and (i, k) . Clustering coefficient talks to us about local density of connections. In our context, a high clustering means that if a word is connected to other two, those two words also have a syntactic relationship in some context. Note that this can be rare in syntax. Take the sentences “I like chocolate” and “I like oranges.” The word “like” is syntactically connected to both “chocolate,” and “oranges,” but it is unlikely that these two words will share a syntactic relationship in a naturally occurring sentence. The same reasoning applies for other syntactic relationships, which often involve words of different kinds (here, a verb and a noun), while words of the same kind take more effort to combine. These reflections will be important when considering the coloring number of our graphs.

(7) Assortativity, r : Assortativity quantifies whether a node is similar to its neighbors regarding some property—usually degree. If, in a network, nodes with large degrees

tend to have neighbors who are also well connected, while vertices with few links are connected to each other, then that graph is assortative. If well-connected nodes link to those with few connections, then the network is antiassortative. We use the Pearson correlation coefficient, r , between each node's degree and that of its neighbors to measure assortativity [59]. Linguistically, since a word's degree tells us about how many distinct syntactic connections that word can form, an assortative graph would mean that words with similar syntactic valence are linked together. That would be the case, e.g., if central connectors (such as pronouns or the verb to be) would preferentially link to each other. This is typically not the case in language networks. They are antiassortative, meaning that a few connector words tend to link together to nodes that have a lower capacity to combine with others.

(8) k cores: A graph's k core is found by iteratively removing all nodes with degree less than k until no more nodes can be removed [60,61]. Take 2-cores as an example: From a network, remove all nodes that have a degree less than 2. In the resulting graph, some nodes might now have degree 1. We remove these next. We repeat, until no node is left with less than two neighbors. If no node is left at the end of this process, then the network does not have a 2-core (a typical syntax tree often does not have a 2-core). We find that most of our syntax graphs have got 2- and 3-cores of varied sizes, but few have 4- or higher k -cores. We call K_k to the set of nodes in a k -core and take the sizes $|K_2|$ and $|K_3|$ of 2- and 3-cores to characterize each network. We also take

$$\kappa = \max_k \{K_k, |K_k| > 0\}, \quad (\text{B1})$$

i.e., k of the largest k core. k cores tell us whether a graph contains densely connected regions, as we can think of them as core sets of nodes that survive once peripheries are removed. Linguistically, the presence of a k core would be telling us that a graph contains a dense, varied subset of syntactic relationships that are recurrent—i.e., that involve the same nodes repeatedly and in different situations.

(9) Cycles: Characterizing cycles is particularly challenging because they grow combinatorially as soon as networks show recurrence—i.e., as loops are completed. We need to introduce some technicalities to account for cycles properly, but the definitions below just formalize the intuition that a cycle is a circuit of nodes within the network that return to a same point. Syntactically, to have cycles present in a graph, we need that some same words have been used in different sentences with different combinations of words. To provide an example: “Mary loves the clouds” and “Some clouds rained on Mary” would close a loop involving 6 different words: “Mary,” “loves,” “the,” “clouds,” “rained,” “on,” “Mary.” Having a large number of cycles in a graph says that a high linguistic combinatorics is at play. Cycles might be longer or shorter. Long cycles in syntax networks reflect long linear constructions (i.e., richer phrases) that are connected by some words at their extremes. Those connectors need to be reused in different phrases (such as “Mary” and “clouds” above). We wanted to capture two properties of cycles: their abundance and lengths. To this end, we need to introduce some notation: Given a network $\mathcal{G}$ with its sets V and E of nodes and edges, respectively, a cycle of this network is an

ordered collection of vertices, $c = \{v(1), \dots, v(n)\}$ such that $v(i) \in V$ for all $i = 1, \dots, n$ and such that $(v(i), v(i+1)) \in E$ for all $i = 1, \dots, n-1$ and $(v(n), v(1)) \in E$. [Note that $v(i)$ labels vertices within the cycle while we used v_i to label vertices within V .] The cycle length is given by n . Let us call $E(c) = \{(v(i), v(i+1)), i = 1, \dots, n\} \cup \{(v(n), v(1))\}$ to the set of edges involved in cycle c . We define the addition of two cycles as the operation resulting from the exclusive or (xor) of the corresponding sets of edges. This is, given two cycles c_1 and c_2 with edges $E(c_1)$ and $E(c_2)$, we say that $c_3 = c_1 + c_2$ with $E(c_3) = E(c_1) \oplus E(c_2)$. Nodes within c_3 might be ordered differently to those in c_1 and c_2 . With this, we say that a set, $C \equiv \{c_l, l = 1, \dots, |C|\}$, of cycles in $\mathcal{G}$ forms a basis of its cycles if any cycle in $\mathcal{G}$ can be generated by linear combinations (i.e., by summing) cycles in C . We refer to $|C|$ as the basis size. If n_j is the size of the j th cycle in C , then

$$L = \frac{1}{|C|} \sum_{j=1}^{|C|} n_j \quad (\text{B2})$$

is the average length of cycles in C . Among all bases, some of them are minimal, meaning that they involve the least number of cycles that can generate all others. But due to their combinatorial explosion, finding a minimal basis in a short time is prohibitive. Because of this, we evaluate cycle properties stochastically. Using a greedy algorithm [62] (implemented by Python's networkx library), we generate 100 cycle bases of each network. We take the mean of their sizes and the mean of the average length of their cycles to enter our analysis.

(10) Constraints: We took three graph measurements that might be interpreted as finding maximal constraints on a network: the *maximal matching size*, the *maximal independent set*, and the *coloring number*. This was inspired by earlier studies on syntax networks, which appreciated the importance of the coloring number [42].

(a) Maximal matching size: A matching of a graph is a set of edges without common vertices. The maximal matching size is the size of the largest matching. This property will be larger in graphs that present long linear chains or sparse trees, while a star of a clique will have maximal matching sizes of 1. Thus, we expect that this property decays as a network becomes more small world. We compute this quantity as the fraction of edges with respect to the total.

(b) Maximal independent set: This property is the converse of the previous one—with nodes instead of edges. An independent set is a set of nodes without common vertices. The maximal independent set is the largest independent set. Contrary to the previous one, this property will be largest in a star and also large in a linear chain or a tree; but it can still be small in most small-world graphs. This set is found with a stochastic greedy algorithm, so we run it 100 times. We take the quantity as the fraction of nodes in the independent set with respect to the total.

(c) Coloring number: This quantity represents the smallest number of colors that you would need to color the nodes of a graph such that two contiguous nodes are colored differently. This quantity is smallest for linear chains or stars, while maximal for a clique. Thus, all three

constraint properties taken together can explore across a range of complex graphs in which extremes we find cliques, stars, or linear chains.

(11) Jaccard similarity: The Jaccard similarity, s_J , between two nodes tells us if their neighborhoods are similar [63]. In a syntax network, this quantity will be large between words that combine with similar words—e.g., those in the same grammatical class. Let $V_i \subset V$, be the neighborhood of node i : $V_i \equiv \{v_k \in V, (v_i, v_k) \in E\}$. Then

$$s_J(v_i, v_j) = \frac{|V_i \cap V_j|}{|V_i \cup V_j|}. \quad (\text{B3})$$

(12) Jaccard entropy: The property above averages Jaccard similarity over all nodes. There might be some nodes that tend to have neighborhoods that are more similar to those of others (think words that can link to any other), while others are less promiscuous in the kind of syntactic relationships they form. We wanted to capture a possible heterogeneity within the network. Since the Jaccard similarity is always a positive number, we captured this possible heterogeneity by normalizing the Jaccard similarity of all nodes, thus each value giving us a fraction within $[0,1]$, and all of them adding to 1. We then computed an entropy, H_J . This entropy will be high if all nodes have similar Jaccard entropy, and smaller if there are some words with much higher values than others.

(13) Pagerank entropy: Similarly, we wanted to capture somehow a possible heterogeneity in the centrality of nodes within the network. To that end, we computed the eigenvector centrality of each node [31]. Since this is a positive number, we again normalized it and computed an entropy, H_P .

1. Further reflections on linguistic aspects of graph properties

Our approach is to measure as many graph properties as possible, then apply dimensionality reduction to distill relevant syntax network patterns. There are many good reasons behind this line of action.

Our data are relatively scarce. While more abundant than in earlier studies, we still have relatively few data points. They are hence noisy, and variability is increased due to the complexity of syntax, with potential nonlinear dependencies (see, e.g., clustering coefficients in Fig. S4 in the Supplemental Material [72]). By combining several, non-fully overlapping properties, more parsimonious patterns emerge (e.g., the saturating curve of PC-1).

The same graph property captures different features on different networks [39]. This affects even keystone graph theoretical properties such as clustering and betweenness centrality. Clustering counts the completed triangles among a node's neighbors. In hubs, potential triangles grow combinatorially, thus becoming impossible to complete. For large networks of moderately high degree, clustering drops due to mathematical (not structural) reasons. Betweenness centrality can capture bridges and/or hubs depending on the network structure. Hence, single metrics are usually not consistent nor exhaustive to compare networks (particularly of different sizes).

As a consequence, we cannot know beforehand what properties are more revealing about networks. This is more pressing in syntax networks, where we hope that graph

properties also capture linguistic facets. On the downside, results from this approach might be difficult to interpret or fail to capture all relevant dimensions. These issues should be alleviated as similar studies accumulate.

Our measurements fall roughly into six categories. We tried to be exhaustive and capture nonoverlapping linguistic traits. Redundancy helps patterns emerge and is otherwise sorted out by PCA:

(1) Sizes: Capture lexical productivity (*vocabulary size*); whether we observe a unique, coordinate syntax network versus unconnected, off-hand constructions (*fraction of connected vocabulary*); and syntactic productivity (unique syntax connections) *number of edges*.

(2) Local connectivity: Capture syntax structure near individual words. *Degree* (related to the number of edges) also measures syntactic productivity. *Clustering coefficient* measures whether neighbors of a word tend also to be connected with each other (rare in syntax). *Assortativity* tells whether syntactic hubs connect among themselves, or rather link words that form less relationships.

(3) Global organization: We move beyond individual words to the general network structure dependent on emergent relationships. *Diameter* distinguishes sprawling networks (e.g., from few unique long sentences) from compact “small world” (as in Stage III). *k* cores speak about the robustness of such compact graphs. Sprawling networks have low *k* cores: Their structure depends on a few syntactic relationships instantiated a few times, resembling a feeble use of syntax

combinatorics. Graphs with robust cores contain repeated use of similar syntax relationships in different contexts—a competent use of productive language.

(4) Cycles: They present mathematical challenges due to their combinatorial nature. They also come in different kinds, each capturing different syntactic aspects. Triangles are the smallest cycles, thus overlapping with clustering. Longer cycles might require: (1) long phrases made of nonrepeating words and (2) key words repeated in different contexts (e.g., hubs like “it,” “the,” etc.). More, longer cycles form when hubs are competently used (i.e., in Stage III).

(5) Structural constraints: Their usefulness was noted earlier [42], and comes naturally with the intuition that syntax is about rules of how to connect classes of words. Say it were barren to connect two words of the same class (not so strange—e.g., determinants do not go well with determinants, etc.; even only special verbs link other verbs). Then, word classes would impose a graph coloring. *Coloring number* identifies the smallest such set of classes. Other constraint properties segregate networks of different structures, as exemplified above.

(6) Node diversity: Aim at capturing whether words form similar syntactic connection patterns. If all words are linked to any others, neighborhoods will be similar, and the *Jaccard* index would be large. *Jaccard entropy* captures whether words are uniform in this sense. If syntax networks were unstructured (e.g., random), their eigenvector centrality would be rather uniform and have low *eigenvector entropy*.

-
- [1] A. D. Friederici and S. M. Gierhan, The language network, *Curr. Opin. Neurobiol.* **23**, 250 (2013).
 - [2] R. C. Berwick and N. Chomsky, *Why Only Us* (MIT Press, Cambridge, MA, 2016).
 - [3] N. Geschwind, Language and the brain, *Sci. Am.* **226**, 76 (1972).
 - [4] P. Lieberman, On the nature and evolution of the neural bases of human language, *Am. J. Phys. Anthropol.* **119**, 36 (2002).
 - [5] M. Catani, D. K. Jones, and D. H. Ffytche, Perisylvian language networks of the human brain, *Ann. Neurol.* **57**, 8 (2005).
 - [6] E. Fedorenko and N. Kanwisher, Neuroimaging of language: Why hasn't a clearer picture emerged? *Lang. Ling. Compass* **3**, 839 (2009).
 - [7] E. Fedorenko, A. Nieto-Castanon, and N. Kanwisher, Lexical and syntactic representations in the brain: An fMRI investigation with multi-voxel pattern analyses, *Neuropsychologia* **50**, 499 (2012).
 - [8] E. Fedorenko and S. L. Thompson-Schill, Reworking the language network, *Trends Cognit. Sci.* **18**, 120 (2014).
 - [9] I. Blank, Z. Balewski, K. Mahowald, and E. Fedorenko, Syntactic processing is distributed across the language system, *NeuroImage* **127**, 307 (2016).
 - [10] S. Malik-Moraleda, D. Ayyash, J. Gallée, J. Affourtit, M. Hoffmann, Z. Mineroff, O. Jouravlev, and E. Fedorenko, An investigation across 45 languages and 12 language families reveals a universal language network, *Nat. Neurosci.* **25**, 1014 (2022).
 - [11] O. A. Olulade, A. Seydell-Greenwald, C. E. Chambers, P. E. Turkeltaub, A. W. Dromerick, M. M. Berl, W. D. Gaillard, and E. L. Newport, The neural basis of language development: Changes in lateralization over age, *Proc. Natl. Acad. Sci. USA* **117**, 23477 (2020).
 - [12] L. F. Seoane and R. Solé, Modeling brain reorganization after hemispherectomy, *bioRxiv*, (2020)–12, doi:10.1101/2020.12.25.424412.
 - [13] L. F. Seoane, Optimality pressures toward lateralization of complex brain functions, *Phys. Rev. X* **13**, 031028 (2023).
 - [14] S. Curtiss, *Genie: A Psycholinguistic Study of a Modern-Day Wild Child* (Academic Press, New York, 2014).
 - [15] V. Fromkin, S. Krashen, S. Curtiss, D. Rigler, and M. Rigler, The development of language in genie: A case of language acquisition beyond the “critical period”, *Brain Lang.* **1**, 81 (1974).
 - [16] B. T. Woods and H.-L. Teuber, Changing patterns of childhood aphasia, *Ann. Neurol.* **3**, 273 (1978).
 - [17] B. T. Woods and S. Carey, Language deficits after apparent clinical recovery from childhood aphasia, *Ann. Neurol.* **6**, 405 (1979).
 - [18] A. C. Marcotte and D. A. Morere, Speech lateralization in deaf populations: Evidence for a developmental critical period, *Brain Lang.* **39**, 134 (1990).
 - [19] L. Hertz-Pannier, C. Chiron, I. Jambaqué, V. Renaux-Kieffer, P.-F. Van de Moortele, O. Delalande, M. Fohlen, F. Brunelle, and D. L. Bihan, Late plasticity for language in a child's non-

- dominant hemisphere: A pre-and post-surgery fMRI study, *Brain* **125**, 361 (2002).
- [20] H. M. Manrique, D. W. Read, and M. J. Walker, On some statistical and cerebral aspects of the limits of working memory capacity in anthropoid primates, with particular reference to Pan and Homo, and their significance for human evolution, *Neurosci. Biobehav. Rev.* **158**, 105543 (2024).
- [21] D. W. Read, H. M. Manrique, and M. J. Walker, On the working memory of humans and great apes: Strikingly similar or remarkably different? *Neurosci. Biobehav. Rev.* **134**, 104496 (2022).
- [22] I. Arnon, S. Kirby, J. A. Allen, C. Garrigue, E. L. Carroll, and E. C. Garland, Whale song shows language-like statistical structure, *Science* **387**, 649 (2025).
- [23] A. Whiten and M. Youngblood, Convergent evolution in whale and human vocal cultures, *Science* **387**, 581 (2025).
- [24] M. Youngblood, Language-like efficiency in whale communication, *Sci. Adv.* **11**, eads6014 (2025).
- [25] H. S. Terrace, L. A. Petitto, R. J. Sanders, and T. G. Bever, Can an ape create a sentence? *Science* **206**, 891 (1979).
- [26] R. F. I. Cancho and R. V. Solé, Least effort and the origins of scaling in human language, *Proc. Natl. Acad. Sci. USA* **100**, 788 (2003).
- [27] R. V. Solé and L. F. Seoane, Ambiguity in language networks, *Linguist. Rev.* **32**, 5 (2015).
- [28] L. F. Seoane and R. Solé, The morphospace of language networks, *Sci. Rep.* **8**, 10465 (2018).
- [29] J. Fortuny, Structural ambiguity and the architecture of language 1, *Stud. Linguist.* **78**, 35 (2024).
- [30] L. Steels and E. G. Casademont, Ambiguity and the origins of syntax, *Linguist. Rev.* **32**, 37 (2015).
- [31] M. Newman, *Networks* (Oxford University Press, Oxford, UK, 2018).
- [32] G. Palla, I. Derényi, I. Farkas, and T. Vicsek, Uncovering the overlapping community structure of complex networks in nature and society, *Nature (London)* **435**, 814 (2005).
- [33] J. Goñi, G. Arrondo, J. Sepulcre, I. Martincorena, N. Vélez de Mendizábal, B. Corominas-Murtra, B. Bejarano, S. Ardanza-Trevijano, H. Peraita, D. P. Wall, *et al.*, The semantic organization of the animal category: Evidence from semantic verbal fluency and network theory, *Cogn. Process.* **12**, 183 (2011).
- [34] J. M. Montoya, S. L. Pimm, and R. V. Solé, Ecological networks and their fragility, *Nature (London)* **442**, 259 (2006).
- [35] R. Albert and A.-L. Barabási, Statistical mechanics of complex networks, *Rev. Mod. Phys.* **74**, 47 (2002).
- [36] M. E. J. Newman, The structure and function of complex networks, *SIAM Rev.* **45**, 167 (2003).
- [37] M. E. J. Newman, Finding community structure in networks using the eigenvectors of matrices, *Phys. Rev. E* **74**, 036104 (2006).
- [38] T. P. Peixoto, *Descriptive vs. Inferential Community Detection in Networks: Pitfalls, Myths and Half-Truths* (Cambridge University Press, 2023).
- [39] L. F. Seoane, Topological communities in complex networks, *arXiv:2409.02317*.
- [40] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, Complex networks: Structure and dynamics, *Phys. Rep.* **424**, 175 (2006).
- [41] B. Corominas-Murtra, S. Valverde, and R. Solé, The ontogeny of scale-free syntax networks: Phase transitions in early language acquisition, *Adv. Complex Syst.* **12**, 371 (2009).
- [42] B. Corominas-Murtra, M. Sánchez Fibla, S. Valverde, and R. Solé, Chromatic transitions in the emergence of syntax networks, *R. Soc. Open Sci.* **5**, 181286 (2018).
- [43] L. Barceló-Coblijn, B. Corominas-Murtra, and A. Gomila, Syntactic trees and small-world networks: Syntactic development as a dynamical process, *Adapt. Behav.* **20**, 427 (2012).
- [44] L. Barceló-Coblijn, A. Irurtzun, C. Real Puigdollers, E. López-Navarro, and A. Gomila, How children develop their ability to combine words: A network-based approach, *Adapt. Behav.* **27**, 307 (2019).
- [45] E. DeGiuli, Random language model, *Phys. Rev. Lett.* **122**, 128301 (2019).
- [46] R. Brown, *A First Language: The Early Stages* (Harvard University Press, Cambridge, MA, 1973).
- [47] N. Chomsky, A review of BF skinner's verbal behavior, *Language* **35**, 26 (1959).
- [48] B. F. Skinner, *Verbal Behavior* (Appleton-Century-Crofts, New York, 1957).
- [49] P. K. Kuhl, A new view of language acquisition, *Proc. Natl. Acad. Sci. USA* **97**, 11850 (2000).
- [50] L. Barceló-Coblijn, A. Benítez-Burraco, and A. Irurtzun, Syntactic networks as an endophenotype of developmental language disorders: An Evo-Devo approach to clinical linguistics, *Biolinguistics* **9**, 043 (2015).
- [51] A. Irurtzun, The “globularization hypothesis” of the language-ready brain as a developmental frame for prosodic bootstrapping theories of language acquisition, *Front. Psychol.* **6**, 1817 (2015).
- [52] B. MacWhinney, The CHILDES Project: Tools for Analyzing Talk, Volume I: *Transcription Format and Programs* (Lawrence Erlbaum Associates, Mahwah, NJ, 2014).
- [53] G. Bol and F. Kuiken, 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, *Clin. Linguist. Phon.* **4**, 77 (1990).
- [54] R. A. Hudson, *A Word Grammar of English* (Basil Blackwell, Oxford, 1990).
- [55] R. A. Hudson, *Language Networks: The New Word Grammar* (Oxford University Press, USA, 2007).
- [56] L. Barceló-Coblijn, D. Serna Salazar, G. Isaza, L. F. Castillo Ossa, and M. G. Bedia, *Netlang*: A software for the linguistic analysis of corpora by means of complex networks, *PLoS One* **12**, e0181341 (2017).
- [57] S. P. Abney, The English noun phrase in its sentential aspect, Ph.D. thesis, Massachusetts Institute of Technology, 1987.
- [58] G. Longobardi, The structure of DPs: Some principles, parameters, and problems, in *The Handbook of Contemporary Syntactic Theory*, edited by M. Baltin and C. Collins (Blackwell Publishers Inc. Malden, MA, 2001), pp. 562–603.
- [59] M. E. J. Newman, Assortative mixing in networks, *Phys. Rev. Lett.* **89**, 208701 (2002).
- [60] V. Batagelj and M. Zaveršnik, Fast algorithms for determining (generalized) core groups in social networks, *Adv. Data Analysis Classification* **5**, 129 (2011).

- [61] L. Hébert-Dufresne, J. A. Grochow, and A. Allard, Multi-scale structure and topological anomaly detection via a new network statistic: The onion decomposition, *Sci. Rep.* **6**, 31708 (2016).
- [62] K. Paton, An algorithm for finding a fundamental set of cycles of a graph, *Commun. ACM* **12**, 514 (1969).
- [63] P. Jaccard, The distribution of the flora in the alpine zone. 1, *New Phytol.* **11**, 37 (1912).
- [64] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell, Dimensionality reduction for visualizing single-cell data using umap, *Nat. Biotechnol.* **37**, 38 (2019).
- [65] J. A. Gallego, M. G. Perich, L. E. Miller, and S. A. Solla, Neural manifolds for the control of movement, *Neuron* **94**, 978 (2017).
- [66] R. J. Gardner, E. Hermansen, M. Pachitariu, Y. Burak, N. A. Baas, B. A. Dunn, M.-B. Moser, and E. I. Moser, Toroidal topology of population activity in grid cells, *Nature (London)* **602**, 123 (2022).
- [67] E. R. Sebastian, J. P. Quintanilla, A. Sánchez-Aguilera, J. Esparza, E. Cid, and L. M. de la Prida, Topological analysis of sharp-wave ripple waveforms reveals input mechanisms behind feature variations, *Nat. Neurosci.* **26**, 2171 (2023).
- [68] M. A. Kramer, Nonlinear principal component analysis using autoassociative neural networks, *AIChE J.* **37**, 233 (1991).
- [69] J. Healy and L. McInnes, Uniform manifold approximation and projection, *Nat. Rev. Methods Primers* **4**, 82 (2024).
- [70] L. Van der Maaten and G. Hinton, Visualizing data using t-SNE, *J. Mach. Learn. Res.* **9**, 11 (2008).
- [71] K. Pearson, LIII. On lines and planes of closest fit to systems of points in space, *London Edinburgh Dublin Philos. Mag. J. Sci.* **2**, 559 (1901).
- [72] See Supplemental Material at <https://link.aps.org/supplemental/10.1103/d93c-hm6w> for additional plots.
- [73] D. M. Raup, Geometric analysis of shell coiling: General problems, *J. Paleontol.* **40**, 1178 (1966).
- [74] K. J. Niklas, Computer models of early land plant evolution, *Annu. Rev. Earth Planet. Sci.* **32**, 47 (2004).
- [75] B. Corominas-Murtra, J. Goñi, R. V. Sole, and C. Rodríguez-Caso, On the origins of hierarchy in complex networks, *Proc. Natl. Acad. Sci. USA* **110**, 13316 (2013).
- [76] J. Goñi, A. Avena-Koenigsberger, N. Velez de Mendizabal, M. P. van den Heuvel, R. F. Betzel, and O. Sporns, Exploring the morphospace of communication efficiency in complex networks, *PLoS One* **8**, e58070 (2013).
- [77] A. Avena-Koenigsberger, J. Goñi, R. Solé, and O. Sporns, Network morphospace, *J. R. Soc. Interface* **12**, 20140881 (2015).
- [78] X. D. Arsiwalla, R. Sole, C. Moulin-Frier, I. Herreros, M. Sánchez-Fibla, and P. Verschure, The morphospace of consciousness: Three kinds of complexity for minds and machines, *NeuroSci.* **4**, 79 (2023).
- [79] A. Ollé-Vila, L. F. Seoane, and R. Solé, Ageing, computation and the evolution of neural regeneration processes, *J. R. Soc. Interface* **17**, 20200181 (2020).
- [80] R. Solé and L. F. Seoane, Evolution of brains and computers: The roads not taken, *Entropy* **24**, 665 (2022).
- [81] D. McNeill, *The Acquisition of Language: The Study of Developmental Psycholinguistics* (Harper & Row, Publishers, Inc., New York, 1970).
- [82] J. O. Edgin, C. A. Clark, E. Massand, and A. Karmiloff-Smith, Building an adaptive brain across development: Targets for neurorehabilitation must begin in infancy, *Front. Behav. Neurosci.* **9**, 232 (2015).
- [83] M. Ahmadlou, M. Gharib, S. Hemmati, R. Vameghi, and F. Sajedi, Disrupted small-world brain network in children with Down syndrome, *Clin. Neurophysiol.* **124**, 1755 (2013).
- [84] M. West-Eberhard, *Developmental Plasticity and Evolution* (Oxford University Press, New York, 2003).
- [85] M. J. West-Eberhard, Developmental plasticity and the origin of species differences, *Proc. Natl. Acad. Sci. USA* **102**, 6543 (2005).
- [86] M. Tomasello, The item-based nature of children's early syntactic development, *Trends Cognit. Sci.* **4**, 156 (2000).
- [87] H. Behrens, Constructivist approaches to first language acquisition, *J. Child Lang.* **48**, 959 (2021).
- [88] N. Chomsky, *Lectures on Government and Binding: The Pisa Lectures* (Mouton de Gruyter, The Hague, 1993), Vol. 9.
- [89] N. Chomsky, A minimalist program for linguistic theory, (1993).
- [90] S. Balari and G. Lorenzo, Computational phenotypes: Where the theory of computation meets Evo-Devo, *Biolinguistics* **3**, 002 (2009).
- [91] D. Bickerton, *Bastard Tongues: A Trailblazing Linguist Finds Clues to Our Common Humanity in the World's Lowliest Languages* (Hill and Wang, New York, 2008).
- [92] D. Bickerton, *More Than Nature Needs* (Harvard University Press, Cambridge MA, 2014).
- [93] G. E. Martin, J. Klusek, B. Estigarribia, and J. E. Roberts, Language characteristics of individuals with Down syndrome, *Top. Lang. Disord.* **29**, 112 (2009).
- [94] J. Rondal, Patterns of correlations for various language measures in mother-child interactions for normal and Down's syndrome children, *Lang. Speech* **21**, 242 (1978).
- [95] J. A. Rondal, *Maternal Speech to Normal and Down's Syndrome Children Matched for Mean Length of Utterance* (Xerox University Microfilms, Ann Arbor, Michigan, 1976).
- [96] G. Laws and D. V. Bishop, A comparison of language abilities in adolescents with Down syndrome and children with specific language impairment, *J. Speech Lang. Hear. Res.* **46**, 1324 (2003).
- [97] A. R. Taboh, D. E. Shalom, B. Alvares, and C. A. Gattei, Sentence comprehension in Spanish-speaking children with hearing loss: On the integration of morphosyntactic and lexico-semantic cues, *J. Speech Lang. Hear. Res.* **68**, 602 (2025).
- [98] L. Tuller, H. Delage, C. Monjauze, A.-G. Piller, and M.-A. Barthez, Clitic pronoun production as a measure of atypical language development in French, *Lingua* **121**, 423 (2011).
- [99] L. B. Leonard, Facilitating linguistic skills in children with specific language impairment, *Appl. Psycholinguist* **2**, 89 (1981).
- [100] S. Reilly, B. Tomblin, J. Law, C. McKean, F. K. Mensah, A. Morgan, S. Goldfeld, J. M. Nicholson, and M. Wake, Specific language impairment: A convenient label for whom? *Int. J. Lang. Commun. Disord.* **49**, 416 (2014).