

Research



Cite this article: Sainburg T, Mai A, Gentner TQ. 2022 Long-range sequential dependencies precede complex syntactic production in language acquisition. *Proc. R. Soc. B* **289**: 20212657.
<https://doi.org/10.1098/rspb.2021.2657>

Received: 7 December 2021

Accepted: 1 February 2022

Subject Category:

Behaviour

Subject Areas:

behaviour, cognition, developmental biology

Keywords:

language acquisition, hierarchy, power law, language evolution

Author for correspondence:

Tim Sainburg

e-mail: timsainb@gmail.com

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.5862335>.

Long-range sequential dependencies precede complex syntactic production in language acquisition

Tim Sainburg^{1,2}, Anna Mai³ and Timothy Q. Gentner^{1,4}

¹Department of Psychology, ²Center for Academic Research & Training in Anthropogeny, ³Department of Linguistics, and ⁴Neurosciences Graduate Program, Neurobiology Section, Kavli Institute for Brain and Mind, UC San Diego, La Jolla, CA 92093, USA

TS, 0000-0003-4223-2689; AM, 0000-0002-8343-9216; TQG, 0000-0002-4516-9841

To convey meaning, human language relies on hierarchically organized, long-range relationships spanning words, phrases, sentences and discourse. As the distances between elements (e.g. phonemes, characters, words) in human language sequences increase, the strength of the long-range relationships between those elements decays following a power law. This power-law relationship has been attributed variously to long-range sequential organization present in human language syntax, semantics and discourse structure. However, non-linguistic behaviours in numerous phylogenetically distant species, ranging from humpback whale song to fruit fly motility, also demonstrate similar long-range statistical dependencies. Therefore, we hypothesized that long-range statistical dependencies in human speech may occur independently of linguistic structure. To test this hypothesis, we measured long-range dependencies in several speech corpora from children (aged 6 months–12 years). We find that adult-like power-law statistical dependencies are present in human vocalizations at the earliest detectable ages, prior to the production of complex linguistic structure. These linguistic structures cannot, therefore, be the sole cause of long-range statistical dependencies in language.

1. Introduction

Since Shannon's original work characterizing the sequential dependencies present in language, the structure underlying long-range information in language has been the subject of a great deal of interest in linguistics, statistical physics, cognitive science and psychology [1–20]. Long-range information content refers to the dependencies between discrete elements (e.g. units of spoken or written language) that persist over long sequential distances spanning words, phrases, sentences and discourse. For example, in Shannon's original work, participants were given a series of letters from an English text and were asked to predict the letter that would occur next. Using the responses of these participants, Shannon derived an upper bound on the information added by including each preceding letter in the sequence. More recent investigations compute statistical dependencies directly from language corpora using either correlation functions [3,4,7,8,10,12,13] or mutual information (MI) functions [2,5,6,14] between elements in a sequence. In both cases, dependencies are calculated as a function of the sequential distance between pairs of elements. For example, in the sequence $a \rightarrow b \rightarrow c \rightarrow d \rightarrow e \rightarrow f$, at a distance of three elements, relationships would be calculated over the pairs a and d , b and e and c and f .

On average, as the distance between elements increases, statistical dependencies weaken. Across many different sequence types, including phonemes, syllables and words in both text and speech, the decay of long-range correlations and MI in language follows a power law (equation (2.6)) [2–14,18,19]. This power-law relationship is thought to derive at least in part from the hierarchical organization of language, and has been variously attributed to

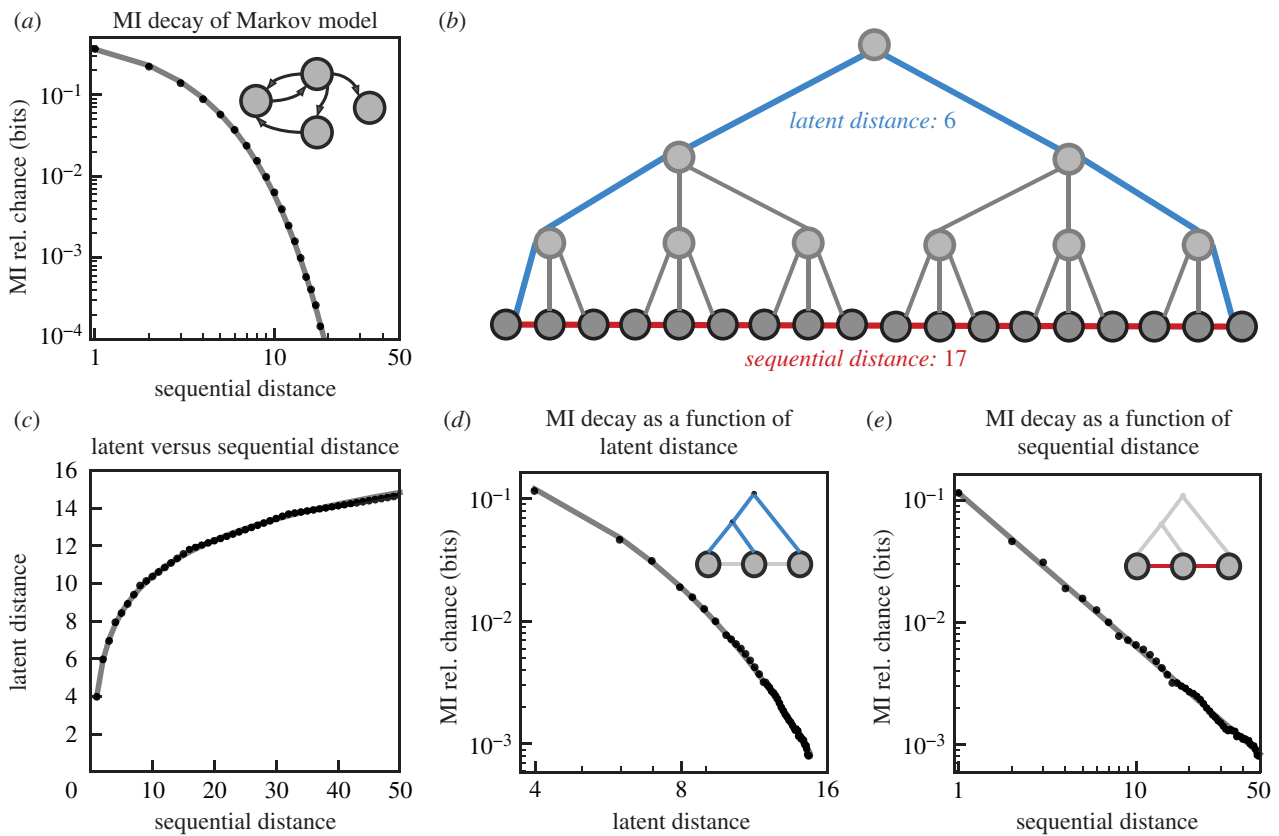


Figure 1. Comparison of long-range statistical dependencies between sequences with and without deep latent relationships. (a) The MI between elements in an iteratively (Markov model) generated sequence decays exponentially as a function of sequential distance. (b) An example sequence with hierarchical latent structure. The latent distance between the two end elements in the sequence is 6 (blue), while the sequential distance is 17 (red). (c) In sequences with hierarchical latent structure, the sequential distance between elements is logarithmically related to the latent distance (fit model: $a \cdot \log_x b + c$ where x is sequential distance). (d) Like sequential distance in (a), the MI between elements in a hierarchically generated sequence decays exponentially as a function of latent distance. (e) The MI between elements in a hierarchically generated sequence decays following a power law as a function of sequential distance, which is related to the exponential MI decay seen in (d) and the logarithmic relationship between sequential and latent distance seen in (c). In (a), the probabilistic Markov model used to generate the empirical data has two states with a self-transition probability of 0.1. In (c–e), a probabilistic context-free grammar [5] with the same transition probability is used.

hierarchical structure in human language syntax [5], semantics [3] and discourse structure [4].

(a) Mechanisms for long-range organization

To understand the link between hierarchical sequential organization in language and long-range sequential dependencies, it is helpful to consider both the latent and surface structure of a sequence (figure 1). When only the surface structure of a sequence is available, as it is for language corpora, a power-law decay in the MI between sequence elements gives evidence of an underlying hierarchical latent structure [5]. This phenomenon can be demonstrated by comparing the MI between elements in a sequence generated from a hierarchically structured language model, such as a probabilistic context-free grammar (PCFG), to the MI between elements in a sequence generated by a non-hierarchical model, such as a Markov process (figure 1). For sequences generated by a Markov process, the strength of the relationship between elements decays exponentially (equation (2.5)) as sequential distance increases [5,21] (figure 1a). In the PCFG model, however, linear distances in the sequence are coupled to logarithmic distances in the latent structure of the hierarchy (figure 1b,c). While information continues to decay exponentially as a function of the distance in the latent hierarchy (figure 1d), this log-scaling results in a

power-law decay when MI is computed over corresponding sequential distances (figure 1e).

(b) Hierarchy in language and behaviour

The thought that human language syntax is generated by CFGs [22] has led many to speculate that the long-range dependencies observed in language corpora are the product of abstract linguistic structure [2–5]. Although the long-range statistical dependencies in language corpora are clearly tied to linguistic structure [3,4], it does not follow that syntactic structure is necessarily the only source for long-range dependencies in language. Indeed, hierarchical organization is unique to neither CFGs nor human language and diverse classes of mechanisms, many of which are decidedly not language-like [23–28], are capable of generating power-law relationships. Many non-linguistic human behaviours [29–34], animal behaviours [35–40], animal vocalizations [41–50] and other biologically generated processes [28,51–57] are organized hierarchically. Likewise, long-range, power-law distributed dependencies are observed in sequential behaviours, including whale song [46], birdsong [2,58] and *Drosophila* [39] and zebrafish motility [59]. Instead, long-range dependencies in language and other human behaviour [33,60,61] may reflect more general biological processes inherited from the organization of underlying neurophysiological mechanisms [62–65]

that are, in turn, characterized by power-law relationships in temporal sequencing [66–68]. When viewed as an instance of this more general class of sequentially organized behaviour, one might reasonably predict that human speech should display long-range statistical dependencies independent of linguistic structure.

(c) General hierarchical organization, linguistic hierarchy and hierarchical description

While many biological, behavioural and other natural signals can be and have been described using hierarchical terms, an important distinction exists between the many different ways in which the term hierarchy is used. For example, the generative mechanisms underlying some signals are either known or widely thought to be organized hierarchically, such as the cascade of motor programs originating in cortex and ending in effector neurons or the production rules underlying formal language theory. In other signals, hierarchicality is less clear, and what is often used to determine whether a signal is hierarchical is whether that signal is well described using hierarchical terms, without explicit hypotheses about the source of that hierarchy. For example, compression via tree-based grammars like the one in figure 1 are commonly used to describe hierarchical organization of non-linguistic structures such as animal behaviours and genomic and protein sequences [59,69,70].

In this work, we specifically explore the relationship between long-range statistical dependencies in speech and the emergence of hierarchical linguistic organization (i.e. syntactic, semantic and discourse structure). We ask whether the long-range statistical dependencies present in speech originate alongside the presence of syntactically complex linguistic productions, or whether they precede the production of this specific form of hierarchical structure during development. The developmental emergence of hierarchical organization in linguistic productions is well studied [71–73]. Meylan *et al.* [72], for example, demonstrate that linguistic productivity is initially low and rapidly develops at around 24 months of age, suggesting that young children's speech lacks the rich grammatical structure that enables language's remarkable productivity when they are first producing language. Hierarchical organization in semantic and discourse structure also grow in hierarchical complexity throughout ontogeny [71]. By 5–6 years of age, children are capable of producing well-formed narratives, which continue to increase in complexity to adulthood [74,75]. Thus, if long-range statistical dependencies are observed at an age prior to the emergence of complex linguistic structure, at both syntactic and discourse levels, those dependencies are not driven by those levels of linguistic structure and are therefore not likely to be driven fully by those linguistic structures in adult speech either.

(d) Origins of long-range statistical dependencies in language

To test whether long-range statistical dependencies occur independently of complex linguistic structure in speech, we used MI decay as a measure of long-range dependencies over several speech corpora from children ranging from six months of age to adults [76–89]. Because complex linguistic productions emerge during language acquisition, we use these corpora to determine whether long-range relationships are present in

human vocalizations prior to the production of linguistically complex speech, or whether they emerge alongside linguistically complex productions. If long-range dependencies were to emerge over the course of development alongside complex utterances, we could conclude that abstract linguistic structure plays a dominant role in the sequential statistical structure of speech. However, if long-range statistical dependencies are observed in infant speech prior to the production of structurally complex utterances, then it is likely that the long-range dependencies observed in adult speech are not solely governed by abstract linguistic structure. Indeed, we find that human speech exhibits long-range power-law statistical dependencies like those observed in mature human language early in development, at 6–12 months of age, while children are still in the 'babbling' stage of language development.

2. Methods

(a) Datasets

We examined MI decay in sequences of words over nine datasets of natural speech from English speaking children included in the CHILDES repository [77,82–89] and three datasets of sequences of phonemes from the PhonBank repository [76,78–80], both of which are part of the TalkBank repository [77]. Each dataset within CHILDES and PhonBank was collected in a slightly different manner. In our analyses, we included only transcripts of spontaneous speech that were collected from typically developing children (usually at an in-home setting with family or an experimenter). The subset of CHILDES we used includes word-level transcripts of speech from children aged 12 months to 12 years of age. The subset of PhonBank we used includes phonetic transcriptions of speech given in the International Phonetic Alphabet (IPA) from children aged six months to four years of age. Between the phoneme and word-level datasets, a large range of speech and language development is covered.

For the MI analysis on phonemes, we binned transcripts into five 6-month age groups (6–12, 12–18, 18–24, 24–30, 30–36) and one age group from 3 years to 4 years. Each transcript was analysed as sequences of phonemes, where phoneme distributions for each transcript are treated independently to account for variation in acquired vocabulary across individuals during development. Because transcript lengths varied between age groups (electronic supplementary material, figure S1), we analysed MI at sequential distances up to the median transcript length for each age group. For the MI analysis on words, we binned transcripts into four 6-month age groups (12–18, 18–24, 24–30, 30–36) and one age group from 3 years to 12 years. We analysed words in the same manner as phonemes. No 6–12 month age group was used in word-level analyses because of the sparsity of word-level productions at that age. We additionally repeated our analysis on a set of French transcripts gathered from PhonBank [90–95].

(b) Transition entropy and Sequitur hierarchical compressibility

Before performing the MI decay analysis we looked at the transition structure and hierarchical compressibility for the 25 longest PhonBank transcripts for each age group to determine whether any sequential structure is present in speech at each age group and if so to quantify and compare how well each speech dataset can be described hierarchically.

To quantify whether non-random transition structure is present in the vocalizations of each age group, we computed a transition matrix between phonemes in each corpus (i.e. a first-order Markov model). We then compared the non-randomness

of the sequence on the basis of the mean transition entropy across states. This transition entropy captures the level of uncertainty about the next state, given the current state. For example, an element (A) that transitions to another element (B) 100% of the time will have a transition entropy of 0. Thus, if the mean transition entropy of a transcript is lower than the mean transition entropy of a randomized transcript, there is some degree of predictability in that transcript beyond randomness. As speech at all ages is dictated both by motor constraints and experience with language starting *in utero* [96], we expect non-random structure to be observed at all ages.

We then quantify how hierarchically compressible each transcript is relative to a transcript of equal length generated by a Markov model. To quantify the hierarchical compressibility, we use the Sequitur algorithm [69] which infers a hierarchical re-write system from symbolic sequences. Sequitur infers a form of deterministic grammar that is a restriction on context-free grammars and does not allow recursion or the capacity for a symbol to take more than one expansion from a non-terminal [97]. Sequitur has been used to losslessly and compressively represent data spanning linguistic, biological, behaviour and generally hierarchical signals (see [69] for the algorithm) and can be used to compare the relative compressibility of sequences [70]. We compare the hierarchical compressibility of transcripts to Markov-generated sequences as controls for each transcript.

(c) Mutual information analysis over sequences

For each dataset, we calculate the sequential MI over the elements of the sequence dataset (e.g. words produced by a child). Each element in each sequence is treated as unique to that transcript to account for different distributions of behaviours across different transcripts within datasets.

Given a sequence of discrete elements $a \rightarrow b \rightarrow c \rightarrow d \rightarrow e$. We calculate MI as:

$$I(X, Y) = S(X) + S(Y) - S(X, Y), \quad (2.1)$$

where X and Y are the distributions of single elements at a given distance. For example, at a distance of two, X is the distribution $[a, b, c]$ and Y is $[c, d, e]$ from the set of element-pairs $(a - c, b - d)$ and $(c - e)$. $\hat{S}(X)$ and $\hat{S}(Y)$ are the marginal entropies of the distributions of X and Y , respectively, and $\hat{S}(X, Y)$ is the entropy of the joint distribution of X and Y .

To estimate entropy, we employ the Grassberger [98] method which accounts for under-sampling true entropy from finite samples:

$$\hat{S} = \log_2(N) - \frac{1}{N} \sum_{i=1}^K N_i \psi(N_i), \quad (2.2)$$

where ψ is the digamma function, K is the number of categories of elements (e.g. words or phones) and N is the total number of elements in each distribution.

We then adjust the estimated MI to account for chance. To do so, we subtract a lower bound estimate of chance MI (\hat{I}_{sh}):

$$MI = \hat{I} - \hat{I}_{sh} \quad (2.3)$$

This sets chance MI at zero. We estimate MI at chance (\hat{I}_{sh}) by calculating MI on permuted distributions of labels X and Y :

$$\hat{I}_{sh}(X, Y) = \hat{S}(X_{sh}) + \hat{S}(Y_{sh}) - \hat{S}(X_{sh}, Y_{sh}). \quad (2.4)$$

X_{sh} and Y_{sh} refer to random permutations of the distributions X and Y described above. Permuting X and Y effects the joint entropy $S(X_{sh}, Y_{sh})$ in I_{sh} , but not the marginal entropies $S(X_{sh})$ and $S(Y_{sh})$.¹ \hat{I}_{sh} is related to the expected mutual information [99–101] which accounts for chance using a hypergeometric model of randomness.

Importantly, MI calculated over a sequence as a function of distance is referred to as a ‘MI function’, to distinguish it as the functional form of MI, which measures the dependency between two random variables [14]. In the MI function, samples from the distributions X and Y are taken from the same sequence, thus they are not independent. MI as a function of distance acts as a generalized form of the correlation function that can be computed over symbolic sequences and captures nonlinear relationships [14].

(d) Fitting mutual information decay

We fit the three following models: An exponential decay model:

$$MI = a * e^{-x*b} + f. \quad (2.5)$$

A power-law model:

$$MI = c * x^d + f. \quad (2.6)$$

A composite model of the power-law and exponential models:

$$MI = a * e^{-x*b} + c * x^d + f, \quad (2.7)$$

where x represents the inter-element distance between units (e.g. phones or syllables).

To fit the model on a logarithmic scale, we computed the residuals between the log of the MI and the log of the model’s estimation of the MI. We scaled the residuals during fitting by the log of the distance between elements to emphasize fitting the decay in log-scale because distance was necessarily sampled linearly as integers. Models were fit using the `lmfit` Python package [102] using Nelder–Mead minimization. We compared model fits on the basis of AICc and report the relative probability of each model fit to the MI decay [2,103].

(e) Mutual information decay controls

Datasets are organized hierarchically into transcripts, utterances, words and phonemes allowing us to shuffle the dataset at multiple levels of organization. To ensure that our MI decay results are a direct result of the sequential organization of each dataset, we performed a control on each dataset shuffling behavioural elements within each individual transcript at each hierarchical level. In addition, to ensure that long-range relationships were not due to trivial repetitions of behaviours, for each dataset we looked at MI decay over sequences in which repeated elements were removed. Finally, we analysed transcripts from a subset of the longest individual transcripts to confirm that our results were not the product of mixing together multiple datasets and transcripts.

3. Results

(a) Transition entropy and hierarchical compression

We analysed the transition entropy and hierarchical compressibility of the 25 longest transcripts for each age group in the PhonBank datasets. An example Markov model inferred from three of these age groups is given in electronic supplementary material, figure S2. Across each age group, as expected, we found non-random transition structure indicated by the mean transition entropy (figure 2a).

We then compared the Sequitur hierarchical compressibility of each transcript to a length-controlled Markov-generated transcript as well as a randomly permuted version of the original transcript. Because in our datasets transcripts vary in size, with younger age groups typically having shorter transcripts available, we artificially varied the length of

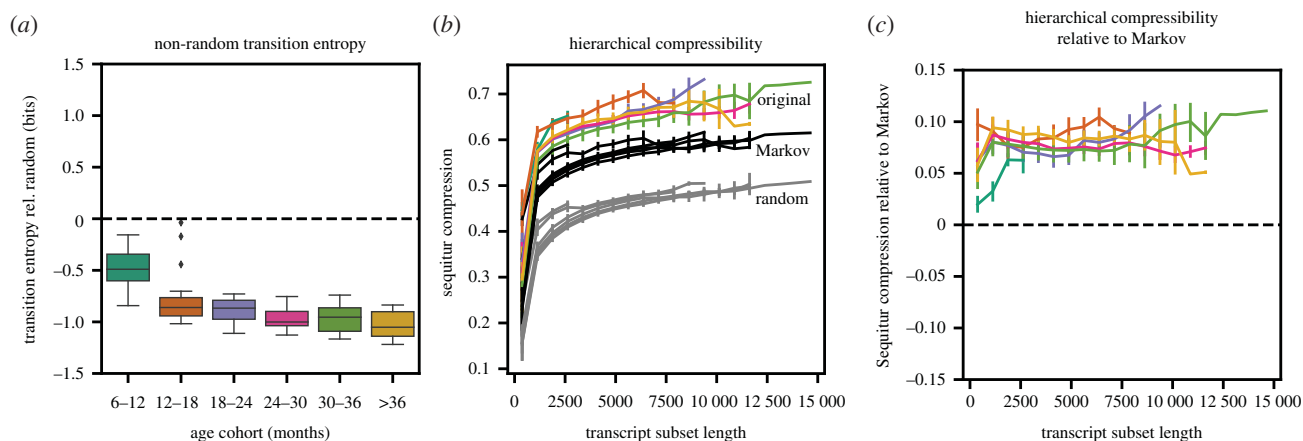


Figure 2. Transition entropy and hierarchical compressibility of children's speech. (a) The average transition entropy of the 25 largest phoneme corpora for each age group, relative to the transition entropy of a permutation of that corpora. The dashed line indicates equal transition entropy. Values below the dashed line indicate transitions are more predictable than random sequences. (b) Sequitur hierarchical compressibility (compressed sequence length divided by original sequence length) for the same transcripts in (a), the permutations in (a), and matched-lengthed transcripted generated from Markov models trained on the transcripts in (a). (c) The same Sequitur compression results as in (b) plotted relative to the Markov results in (b). Values above the dashed line indicate greater hierarchical compressibility than a first-order Markov model. (Online version in colour.)

each transcript between 1 and 100% of the full transcript and computed the compressibility of the transcript at each length to enable comparison between age-groups.

We find that, across all age groups, transcripts are more hierarchically compressible than their Markov or randomized counterparts (figure 2b). We also find that the degree to which these transcripts are compressible over their Markov counterparts is similar across each age group (figure 2c). These results indicate non-random and hierarchical structure across corpora from all age groups. In the following MI analyses, we measure the sequential relationships underlying these signals.

(b) Mutual information

Although much work has explored the information content and long-range sequential organization of human language, relatively few studies have examined these properties in speech [2] or language development directly. Here we investigate the long-range information present in speech during language development using datasets from the TalkBank project [76,77].

We examined MI decay in sequences of phones over three datasets of natural speech from English-speaking children included in the PhonBank repository. Across all age groups, starting at 6–12 months of age, the decay in MI over sequences of phonemes is best fit by a composite power-law and exponential decay model (figure 3a–c; relative probabilities 0.897 to > 0.999; electronic supplementary material, table S1). In each age group, we observe both a clear power law, prominent over long distances (figure 3b), and a clear exponential decay at short, word-length distances (figure 3c), consistent with prior results adult speech [2]. We then examined MI decay in sequences of words over nine datasets of natural speech from English-speaking children included in the CHILDES repository. As with phonemes, the MI decay between words is best fit by a composite model of power-law and exponential decay (equation (2.7); relative probability = 0.989 for 12–18 months and greater than 0.999 for all other age groups; figure 3d–f; electronic supplementary material, table S2). To ensure that our results are not specific to English, we repeated this analysis over French corpora also taken from PhonBank and

found the same power-law and exponential components of the decay in each age group (electronic supplementary material, figure S3). The parameters for each best-fit model for figure 3 can be found in electronic supplementary material, table S3. We additionally plotted the decay parameters of the PhonBank datasets for individual corpora across age groups varying corpus size in electronic supplementary material, figure S4.

(c) Mutual information decay controls

As controls, we also computed the MI decay over sequences of words and phonemes that had been shuffled to isolate sequential relationships at different levels of organization (e.g. phoneme, word, utterance, transcript). A subset of these controls over the PhonBank dataset are shown in figure 4 while the remainder are given in electronic supplementary material, figures S5 and S6.

To aid in interpreting these controls, we additionally performed equivalent shuffling controls with the PCFG model from figure 1 extended with Markov-generated endpoints (figure 4a; as in Sainburg *et al.* [2]). In this model, we replace each terminal state generated from the PCFG with a Markov-generated sequence. These two separate models (hierarchical and finite-state) capture the distinction between the finite-state, Markovian, organization of phonological relationships at short distances (e.g. within and between words) [104] and the hierarchical organization which extends beyond (e.g. syntax, discourse). We have previously found that this model captures the observed interplay between the exponential and power-law MI decay found in speech [2], where the exponential decay is driven by the Markovian model and the power-law is driven by the context-free grammar.

Broadly, across each age group, we find results that are consistent with the model from figure 4a–e. We observe that short-range relationships captured by exponential decay are largely carried within words and utterances, while long-range relationships are captured by a power-law decay carried across longer timescales between words and utterances. In the full model, as in figure 3a, the MI decay (figure 4b) is well explained by an exponential decay occurring at shorter

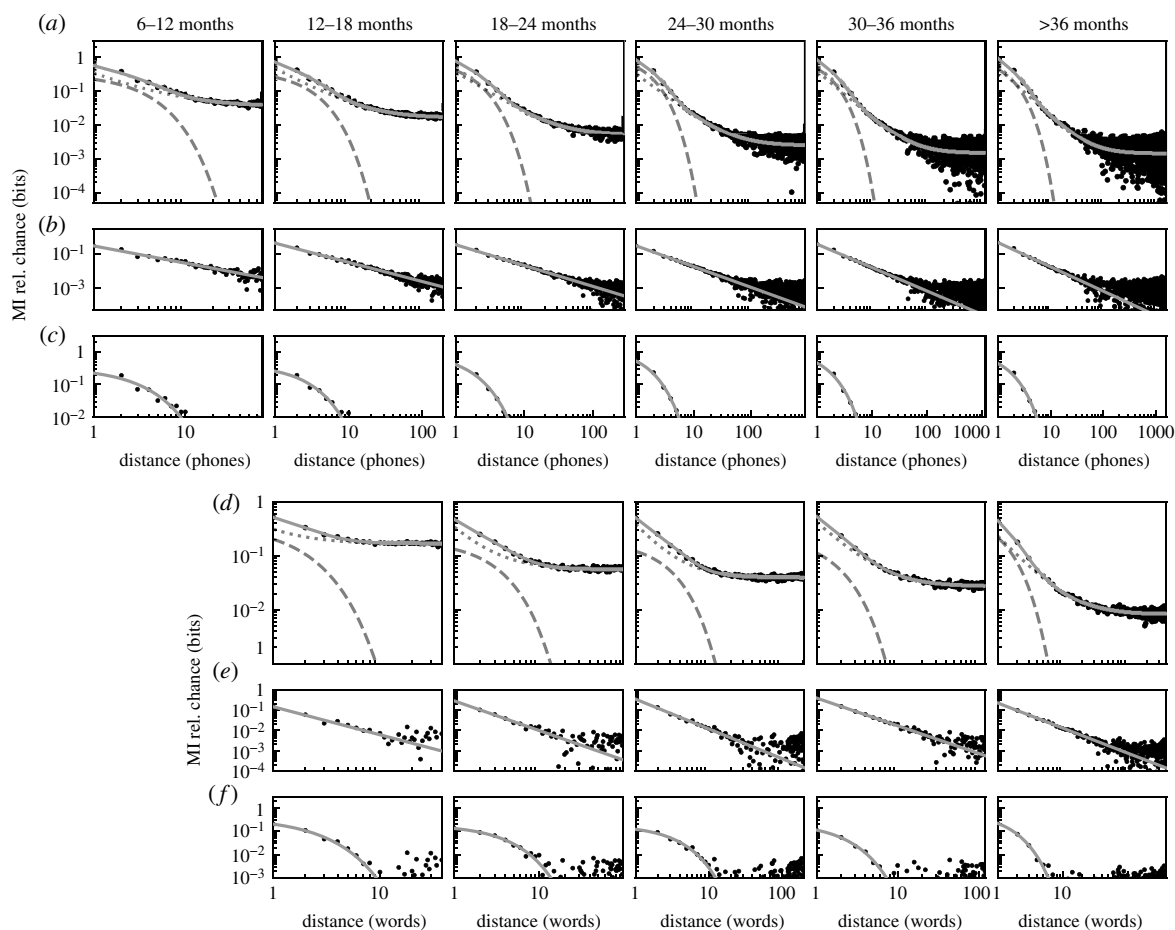


Figure 3. MI decay over words and phonemes during development. (a) MI decay over phonemes for each age group. MI decay is best fit by a composite model (solid grey line) for all age groups across phonemes and words. Exponential and power-law decays are shown as a dashed and dotted grey lines, respectively. (b) The MI decay (as in (a)) with the exponential component of the fit model subtracted to show the power-law component of the decay. (c) The MI decay (as in (a)) with the power-law component subtracted to show the exponential component of the decay. (d–f) The same analyses as (a–c), but for words. No 6–12 month age group was used in word-level analyses because of the sparsity of word-level productions at that age.

distances (dashed line) and a power-law decay at longer distances (dotted line). Shuffling at different levels of organization modifies the relative contribution of these exponential and power-law components of the decay.

Shuffling Markov sequences within the transcript (figure 4c) relates to shuffling words within the transcript (figure 4g). Because the hierarchical structure is eliminated but the Markovian structure is preserved, the power-law component of the decay is effectively eliminated (figure 4c, dashed line). Likewise, because within-word structure is largely preserved but between word structure is destroyed, the power-law component of the decay is also eliminated (figure 4g).

On the opposite end of the hierarchy, when Markovian structure is destroyed by shuffling the sequence within the Markov-generated sequences (figure 4d), the exponential component of the decay is lost (dashed line) and the decay is dominated by the power law. Similarly, when phones are shuffled within utterances (figure 4h), the power-law component of the decay is largely retained while the exponential component is reduced (relative to figure 3a). Consistent with the analyses over words in the CHILDES transcripts (figure 3f), the exponential component of the decay is not entirely destroyed when shuffling within utterances (figure 4h) or words (electronic supplementary material, figure S5c), indicating that a boundary between Markovian and hierarchical organization cannot be fully assigned at either level of organization.

Finally, in an intermediary shuffling, as in the mid-level branch sequences (figure 4e), the model is still well explained by both the exponential and power-law decay components, however, the power-law component (dotted line) now falls between what it did when all of the hierarchical structure was destroyed (figure 4c) and when the full signal was present (figure 4b). Again, we observe similar results with the transcript data in figure 4i (in comparison to figure 3a and figure 4g). A complete set of shuffling results at more levels of organization as well as with the CHILDES dataset are given in electronic supplementary material, figures S5 and S6.

As an additional control, to ensure that the observed MI decay patterns are not the product of mixing datasets from multiple individuals, we also computed the MI decay of the longest individual transcripts comprising each age cohort across both phonemes and words. The decay of the longest individual transcripts parallel the results across transcripts shown in figure 3 (electronic supplementary material, figures S7 and S8). We also analysed the MI decay of transcripts when repeated elements were removed to ensure long-range relationships were not the product of behavioural repetitions (as in e.g. [41]). Removing repeats does not qualitatively alter the pattern of long-range relationships between elements (electronic supplementary material, figure S9).

One reasonable hypothesis is that these long-range relationships in child speech are driven by interaction. Child speech in

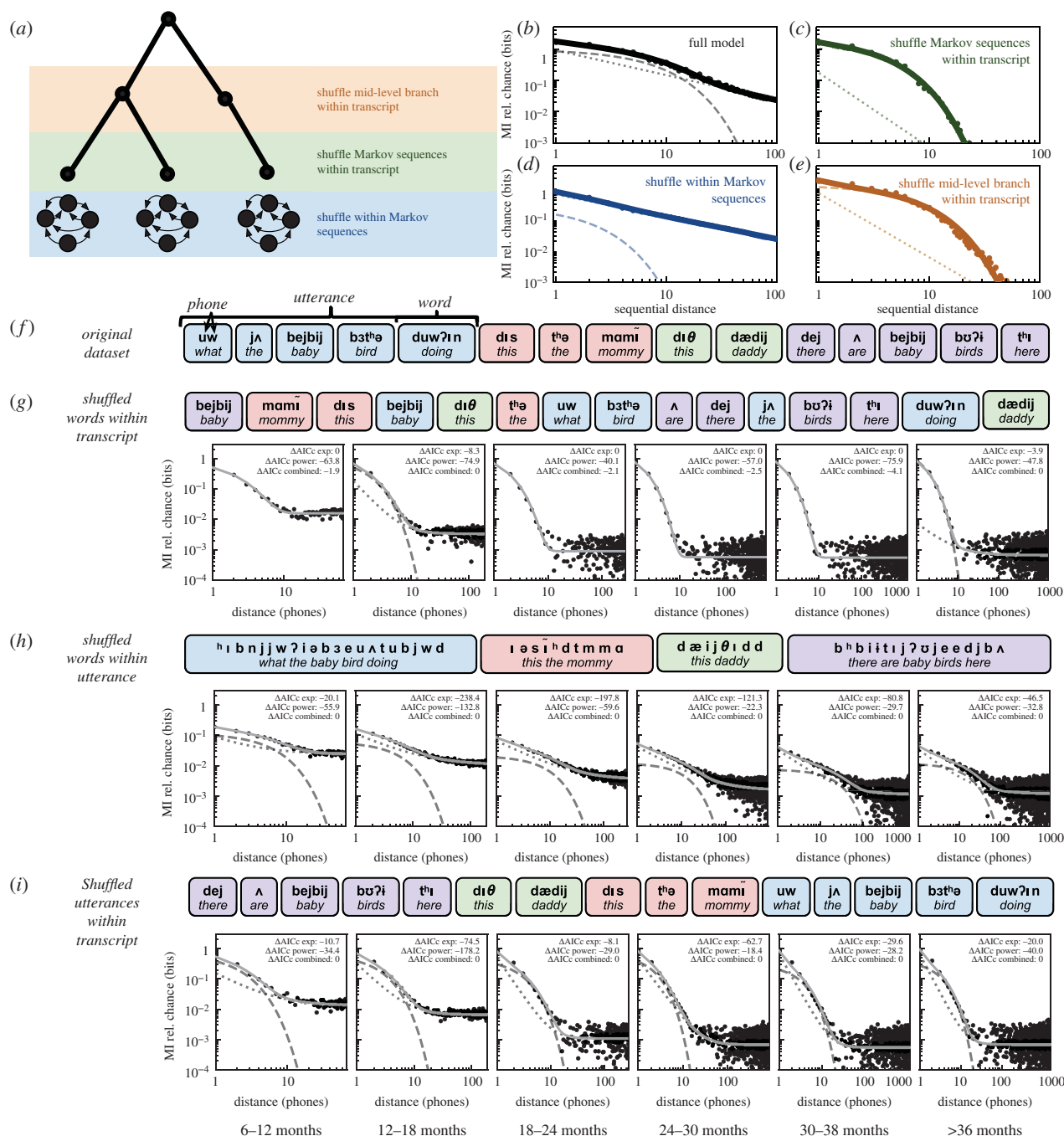


Figure 4. MI decay between phones under different shuffling conditions. (a) An example PFSG model with endpoints generated from Markov models as in [2]. Shuffling controls are given at three levels of organization in the model: within Markov-generated sequences (bottom/blue), between Markov-generated sequences (middle/green) and at a mid-level branch in the generated hierarchy (top/orange). (b) The MI decay curve for the model from (a). (c) The MI decay curve for the shuffle Markov sequences condition from (a). (d) The MI decay curve for the shuffle within Markov sequences condition from (a). (e) The MI decay curve for the shuffle within mid-level branch condition from (a). (f) An example sequence of utterances from the PhonBank dataset. Utterances are grouped by colour, words are grouped by rounded rectangles, and phones are displayed in bold above orthographic transcriptions. (g) MI decay, as in figure 3 when words are shuffled within each transcript. $\Delta AICc$ values (relative to best fit model) are shown for each fit model in each panel. (h) MI decay when phones are shuffled within each utterance. (i) MI decay when utterances are shuffled within each transcript. For each panel, as in figure 3a,d, the dashed and dotted lines represent the exponential and power-law components of the fit composite decay, respectively. When the best-fit model is not the combined model, a solid line is shown for the model fit without the dashed or dotted components.

the datasets used here is produced in an interactive context with adults, which affects discourse structure in children's speech [105] thus, adult speech could be driving the long-range relationships observed in child speech. If this were the case, one could argue that the complex hierarchical structure underpinning the adult's speech was driving the long-range dependencies found in infant speech. To test whether this is the case, for each corpus where adult speech was transcribed ($n_{\text{CHILDES}} = 1630$, $n_{\text{PhonBank}} = 309$) we tested the effect of

non-subject engagement (the proportion of speech not produced by the child) on the improvement in model fit ($\Delta AICc$) of a power-law model over exponential model alone. In both datasets, we observe that adult involvement provides no additional predictive information about the improvement in fit of the power-law model over the exponential model, when controlling for the dataset, child's age and length of the transcript (CHILDES: $F_{1,1620} = 1.49$, $p = 0.22$; PhonBank: $F_{1,306} = 0.21$, $p = 0.65$). Although our results do not provide irrefutable

evidence that the long-range relationships observed are driven by adult speech, these results do not rule out the possibility. Our analyses were based on the natural variability in adult speech across corpora and are not explicitly controlled.

4. Discussion

We analysed the developmental time course of long-range sequential information present in speech throughout early childhood. We observed adult-like long-range statistical relationships [2] present as early as 6–12 months in phoneme sequences, and at 12–18 months in word sequences. Our findings compel reconsideration of the mechanisms that shape long-range statistical relationships in human language. Adult speech is characterized by both long-range and short-range statistical relationships characterized by power-law and exponential decays in MI, respectively. Traditionally, the power-law decay in information between the elements of language (phonemes, words, etc.) has been thought to be imposed by the hierarchical linguistic structure of syntax, semantics, and discourse [3–5]. Early speech development provides a natural experiment in which one can examine human vocal communication absent the production of complex syntactic and discourse structure [71–75]. Remarkably, even at a very early age, prior to the production of mature linguistic structures, vocal sequences show adult-like long-range dependencies.

The long-range dependencies observed in adult language corpora have been clearly tied to linguistic structure [3–5], but the absence of these organizational components in productions of the youngest children rule out the possibility that linguistic structure is the only driver of long-range dependencies in human speech. One reasonable hypothesis consistent with our results is that multiple mechanisms impose long-range dependencies on human speech and language and that these operate on different developmental timescales. Below we consider non-linguistic mechanisms that could generate long-range dependencies in speech at early ages. Limitations of the current dataset prevent quantitative exploration of these mechanisms, but we note how they could be pursued in future research.

One possibility is that the long-range structure we observe in young children is driven by proximal environmental factors, such as the long-range statistics in a child's linguistic environment. In animal behaviour, long-range statistical relationships between behavioural states can be affected by seemingly inconsequential variables such as lighting environment [59]. In the child's environment, linguistic behaviour could be influenced by similar environmental variables. Furthermore, communication is an inherently interactive behaviour with other humans. While our results did not provide evidence that long-range relationships are driven by interactions with adult speakers, we cannot rule out language interaction or other exogenous variables as possible drivers for the observed long-range relationships. It would be valuable to explore the impact of environment on the parameters of the observed information decay (e.g. the exponential and power-law decay parameters and transition between exponential and power-law decay). These factors cannot be properly adjudicated in the present study because our study relies on datasets where corpus variability across environments and age groups makes it difficult to compare model fits directly (e.g. electronic supplementary material, figure S4).

Another possible non-linguistic source of the observed long-range relationships in early speech is the general hierarchical organization of the motor control systems that produce speech and all other overt behaviours. Consistent with this, we note the observation of similar long-timescale dynamics in recent computational ethology studies examining comparatively simple behaviours such as *Drosophila* motility [39]. From this perspective, our results are consistent with the notion that specialized, hierarchical linguistic structure is overlaid on a more general (non-linguistic) hierarchical motor control structure. Accordingly, long-range dependencies relevant to language may emerge more slowly with the development of adult-like linguistic competence. Little is known at present, however, about the neural and motor control underpinnings of these long-range statistical dependencies in non-vocal behavioural sequences, and/or whether these relationships are driven by environmental structure.

The observation of similar power laws in diverse non-linguistic behaviours reinforces the idea that multiple mechanisms can shape the sequential dynamics of behaviour, including speech. There are many potential sources for long-range correlations in biological and physical systems that do not necessarily guarantee an underlying hierarchical structure [23–28]. For example, $1/f$ long-range statistical relationships are characteristic of both physiological and environmental signals. Although $1/f$ signals are often associated with hierarchical organization, their origins in physical systems remains an area of active research [27,106–108]. The presence of long-range correlations characterized as $1/f$ 'noise' in human neural signals is associated with healthy brain states and disappears in disease states and age-related impairments in working memory [108–110]. $1/f$ power spectra produce similar power-law correlations and MI decay as those we observe in speech; thus, it is possible that long-range statistical relationships in speech may originate from such physiological sources as those observed to generate $1/f$ power spectra. In electronic supplementary material, figure S10 we show that behavioural states tied to $1/f$ noise would produce the same power-law MI decay as the context-free grammar model from figure 1*b,e*. Comparing measurements of long-range statistical relationships in speech in both healthy and disease states in relation to the degradation of $1/f$ characteristics in neural signals would indicate whether long-range dependencies in speech and $1/f$ noise in neural systems are related.

Regardless of any further understanding of the specific mechanisms that underlie the sequential dependencies in speech, clear patterns in the information conveyed across time exist in human vocal behaviour at very early ages. In principle, this structure is available to listeners and can provide predictive information to any cognitive agent that engages with it. Humans are necessarily sensitive to long-range relationships in language, and although more sparse, evidence for long-range sensitivities outside language has also been reported, such as scale-invariance in retrospective memory tasks [111] and attention to power-law timescales in anticipation of future events in cognitive tasks [112]. Among non-human animals, the evidence supporting sensitivity to the long-range dynamics (power-law or otherwise) of information in the environment is not well studied, especially at long intervals. If non-human animals can perceive the long-range statistical dependencies present in their environment, this capacity may constitute a broad faculty of language

[113], that is, a necessary, but not uniquely human, component of language. Indeed, the presence of long-range statistical dependencies in non-linguistic behaviours and a generalized perceptual sensitivity to them could provide a scaffold for language to evolve, and where hierarchical syntax and semantics can be understood as later additions that exploit existing long-range structures and sensitivities.

Data accessibility. All scripts used in this study are openly accessible at <https://github.com/timsainb/LongRangeSequentialOrgPaper>. The data are provided in the electronic supplementary material [114]. The datasets can be acquired from the TalkBank repository [77] and PhonBank repository [76]. We performed analyses over these transcripts without any modification. Example transcripts for each dataset are displayed in the electronic supplementary material, Supplementary Information. The distribution of sequence lengths of each dataset is shown in electronic supplementary material, figure S1.

Authors' contributions. T.S.: conceptualization, data curation, formal analysis, funding acquisition, investigation, methodology, project administration, resources, software, validation, visualization,

writing—original draft, writing—review and editing; A.M.: conceptualization, writing—original draft, writing—review and editing; T.Q.G.: conceptualization, funding acquisition, project administration, supervision, writing—original draft, writing—review and editing.

All authors gave final approval for publication and agreed to be held accountable for the work performed therein.

Competing interests. We declare we have no competing interests.

Funding. This work was supported by NSF GRF 2017216247 and an Annette Merle-Smith Fellowship to T.S., NIMH training fellowship T32MH020002 and William Orr Dingwall Dissertation Fellowship to A.M. and T.S., and NIH DC0164081 and DC018055 to T.Q.G.

Endnote

¹The joint entropy is computed over pairs of X and Y , whereas the marginal entropy is computed over X independently and Y independently. Permuting X and Y shuffles the pairing between elements in X and Y , but does not change the frequency of elements in the distributions X or Y .

References

- Shannon CE. 1951 Prediction and entropy of printed English. *Bell Syst. Tech. J.* **30**, 50–64. (doi:10.1002/j.1538-7305.1951.tb01366.x)
- Sainburg T, Theilman B, Thielk M, Gentner TQ. 2019 Parallels in the sequential organization of birdsong and human speech. *Nat. Commun.* **10**, 3636. (doi:10.1038/s41467-019-11605-y)
- Alvarez-Lacalle E, Dorow B, Eckmann J-P, Moses E. 2006 Hierarchical structures induce long-range dynamical correlations in written texts. *Proc. Natl Acad. Sci. USA* **103**, 7956–7961. (doi:10.1073/pnas.0510673103)
- Altmann EG, Cristadoro G, Degli Esposti M. 2012 On the origin of long-range correlations in texts. *Proc. Natl Acad. Sci. USA* **109**, 11 582–11 587. (doi:10.1073/pnas.1117723109)
- Lin H, Tegmark M. 2017 Critical behavior in physics and probabilistic formal languages. *Entropy* **19**, 299. (doi:10.3390/e19070299)
- Grassberger P. 1989 Estimating the information content of symbol sequences and efficient codes. *IEEE Trans. Inf. Theory* **35**, 669–675. (doi:10.1109/18.30993)
- Schenkel A, Zhang J, Zhang Y-C. 1993 Long range correlation in human writings. *Fractals* **1**, 47–57. (doi:10.1142/S0218348X93000083)
- Ebeling W, Pöschel T. 1994 Entropy and long-range correlations in literary English. *Euro Phys. Lett.* **26**, 241. (doi:10.1209/0295-5075/26/4/001)
- Allegrini P, Grigolini P, Palatella L. 2004 Intermittency and scale-free networks: a dynamical model for human language complexity. *Chaos Solitons Fractals* **20**, 95–105. (doi:10.1016/S0960-0779(03)00432-6)
- Melnyk SS, Usatenko OV, Yampol'skii VA, Golick VA. 2005 Competition between two kinds of correlations in literary texts. *Phys. Rev. E* **72**, 026140. (doi:10.1103/PhysRevE.72.026140)
- Montemurro MA, Zanette DH. 2002 Entropic analysis of the role of words in literary texts. *Adv. Complex Syst.* **5**, 7–17. (doi:10.1142/S0219525902000493)
- Montemurro MA, Zanette DH. 2010 Towards the quantification of the semantic information encoded in written language. *Adv. Complex Syst.* **13**, 135–153. (doi:10.1142/S0219525910002530)
- Montemurro MA, Pury PA. 2002 Long-range fractal correlations in literary corpora. *Fractals* **10**, 451–461. (doi:10.1142/S0218348X02001257)
- Li W. 1990 Mutual information functions versus correlation functions. *J. Stat. Phys.* **60**, 823–837. (doi:10.1007/BF01025996)
- Newman EB, Gerstman LJ. 1952 A new method for analyzing printed English. *J. Exp. Psychol.* **44**, 114–125. (doi:10.1037/h0055693)
- Burton NG, Licklider CR. 1955 Long-range constraints in the statistical structure of printed English. *Am. J. Psychol.* **68**, 650–653. (doi:10.2307/1418794)
- Newman EB. 1951 The pattern of vowels and consonants in various languages. *Am. J. Psychol.* **64**, 369–379. (doi:10.2307/1418998)
- Cover T, King R. 1978 A convergent gambling estimate of the entropy of English. *IEEE Trans. Inf. Theory* **24**, 413–421. (doi:10.1109/TIT.1978.1055912)
- Shen H. 2019 Mutual information scaling and expressive power of sequence models. arXiv 1905.04271 (<https://arxiv.org/abs/1905.04271>)
- Futrell R, Mahowald K, Gibson E. 2015 Large-scale evidence of dependency length minimization in 37 languages. *Proc. Natl Acad. Sci. USA* **112**, 10 336–10 341. (doi:10.1073/pnas.1502134112)
- Li W. 1987 Power spectra of regular languages and cellular automata. *Complex Syst.* **1**, 107–130.
- Chomsky N. 1959 On certain formal properties of grammars. *Inf. Control* **2**, 137–167. (doi:10.1016/S0019-9958(59)90362-6)
- Newman MEJ. 2005 Power laws, Pareto distributions and Zipf's law. *Contemp. Phys.* **46**, 323–351. (doi:10.1080/00107510500052444)
- Stumpf MPH, Porter MA. 2012 Critical truths about power laws. *Science* **335**, 665–666. (doi:10.1126/science.1216142)
- Fox Keller E. 2005 Revisiting 'scale-free' networks. *BioEssays* **27**, 1060–1068. (doi:10.1002/bies.20294)
- Boffetta G, Carbone V, Giuliani P, Giuliani P, Veltri P, Vulpiani A. 1999 Power laws in solar flares: self-organized criticality or turbulence? *Phys. Rev. Lett.* **83**, 4662. (doi:10.1103/PhysRevLett.83.4662)
- Munoz MA. 2018 Colloquium: criticality and dynamical scaling in living systems. *Rev. Mod. Phys.* **90**, 031001. (doi:10.1103/RevModPhys.90.031001)
- Gisiger T. 2001 Scale invariance in biology: coincidence or footprint of a universal mechanism? *Biol. Rev.* **76**, 161–209. (doi:10.1017/S1464793101005607)
- Cooper R, Shallice T. 2000 Contention scheduling and the control of routine activities. *Cogn. Neuropsychol.* **17**, 297–338. (doi:10.1080/026432900380427)
- Fitch WT, Martins MD. 2014 Hierarchical processing in music, language, and action: Lashley revisited. *Ann. NY Acad. Sci.* **1316**, 87–104. (doi:10.1111/nyas.12406)
- Whiten A, Flynn E, Brown K, Lee T. 2006 Imitation of hierarchical action structure by young children. *Dev. Sci.* **9**, 574–582. (doi:10.1111/j.1467-7687.2006.00535.x)
- Botvinick MM. 2008 Hierarchical models of behavior and prefrontal function. *Trends Cogn. Sci.* **12**, 201–208. (doi:10.1016/j.tics.2008.02.009)
- Spencer Lashley K. 1951 *The problem of serial order in behavior*. Oxford, UK: Bobbs-Merrill.

34. Kozhevnikov VA, Chistovich LA. 1965 *Speech: articulation and perception*. Moscow, Russia: Nauka.
35. Dawkins M, Dawkins R. 1976 Hierarchical organization and postural facilitation: rules for grooming in flies. *Anim. Behav.* **24**, 739–755. (doi:10.1016/S0003-3472(76)80003-6)
36. Pruetz JD, Bertolani P. 2007 Savanna chimpanzees, *Pan troglodytes* versus, hunt with tools. *Curr. Biol.* **17**, 412–417. (doi:10.1016/j.cub.2006.12.042)
37. Byrne RW, Byrne JME. 1993 Complex leaf-gathering skills of mountain gorillas (*Gorilla g. beringei*): variability and standardization. *Am. J. Primatol.* **31**, 241–261. (doi:10.1002/ajp.1350310402)
38. Lefebvre L. 1981 Grooming in crickets: timing and hierarchical organization. *Anim. Behav.* **29**, 973–984. (doi:10.1016/S0003-3472(81)80050-4)
39. Berman GJ, Bialek W, Shaevitz JW. 2016 Predictability and hierarchy in *Drosophila* behavior. *Proc. Natl Acad. Sci. USA* **113**, 11 943–11 948. (doi:10.1073/pnas.1607601113)
40. Lefebvre L. 1982 The organization of grooming in budgerigars. *Behav. Processes* **7**, 93–106. (doi:10.1016/0376-6357(82)90019-5)
41. Kershenbaum A, Bowles AE, Freeberg TM, Jin DZ, Lameira AR, Bohn K. 2014 Animal vocal sequences: not the Markov chains we thought they were. *Proc. R. Soc. B* **281**, 20141370. (doi:10.1098/rspb.2014.1370)
42. Roeske TC, Kelty-Stephen D, Wallot S. 2018 Multifractal analysis reveals music-like dynamic structure in songbird rhythms. *Sci. Rep.* **8**, 4570. (doi:10.1038/s41598-018-22933-2)
43. Markowitz JE, Ivie E, Kligler L, Gardner TJ. 2013 Long-range order in canary song. *PLoS Comput. Biol.* **9**, e1003052. (doi:10.1371/journal.pcbi.1003052)
44. Hedley RW. 2016 Composition and sequential organization of song repertoires in Cassin's vireo (*Vireo cassinii*). *J. Ornithol.* **157**, 13–22. (doi:10.1007/s10336-015-1238-x)
45. Sasahara K, Cody ML, Cohen D, Taylor CE. 2012 Structural design principles of complex bird songs: a network-based approach. *PLoS ONE* **7**, e44436. (doi:10.1371/journal.pone.0044436)
46. Suzuki R, Buck JR, Tyack PL. 2006 Information entropy of humpback whale songs. *J. Acoust. Soc. Am.* **119**, 1849–1866. (doi:10.1121/1.2161827)
47. Jiang X, Long T, Cao W, Li J, Dehaene S, Wang L. 2018 Production of supra-regular spatial sequences by macaque monkeys. *Curr. Biol.* **28**, 1851–1859. (doi:10.1016/j.cub.2018.04.047)
48. Bruno JH, Tchernichovski O. 2017 Regularities in zebra finch song beyond the repeated motif. *Behav. Processes* **163**, 53–59. (doi:10.1016/j.beproc.2017.11.001)
49. Ferrer-i-Cancho R, McCowan B. 2012 The span of correlations in dolphin whistle sequences. *J. Stat. Mech: Theory Exp.* **2012**, P06002. (doi:10.1016/j.tics.2008.02.009)
50. Morita T, Koda H, Okanoya K, Tachibana RO. 2020 Measuring long context dependency in birdsong using an artificial neural network with a long-lasting working memory. *bioRxiv*. (doi:10.1101/2020.05.09.083907)
51. Peng C-K, Mietus J, Hausdorff JM, Havlin S, Stanley HE, Goldberger AL. 1993 Long-range anticorrelations and non-Gaussian behavior of the heartbeat. *Phys. Rev. Lett.* **70**, 1343. (doi:10.1103/PhysRevLett.70.1343)
52. Stanley HE, et al. 1996 Anomalous fluctuations in the dynamics of complex systems: from DNA and physiology to econophysics. *Physica A* **224**, 302–321. (doi:10.1016/0378-4371(95)00409-2)
53. Li W, Kaneko K. 1992 Long-range correlation and partial $1/f_{\alpha}$ spectrum in a noncoding DNA sequence. *Euro Phys. Lett.* **17**, 655. (doi:10.1209/0295-5075/17/7/014)
54. Peng C-K, Buldyrev SV, Goldberger AL, Havlin S, Sciortino F, Simons M, Stanley HE. 1992 Long-range correlations in nucleotide sequences. *Nature* **356**, 168–170. (doi:10.1038/356168a0)
55. Viswanathan GM, Afanasyev V, Buldyrev SV, Murphy EJ, Prince PA, Stanley HE. 1996 Lévy flight search patterns of wandering albatrosses. *Nature* **381**, 413–415. (doi:10.1038/381413a0)
56. Kobayashi M, Musha T. 1982 1/f fluctuation of heartbeat period. *IEEE Trans. Biomed. Eng.* **BME-29**, 456–457. (doi:10.1109/TBME.1982.324972)
57. Canessa E, Calmetta A. 1994 Physics of a random biological process. *Phys. Rev. E* **50**, R47. (doi:10.1103/PhysRevE.50.R47)
58. Ma S, Maat AT, Gahr M. 2017 Power-law scaling of calling dynamics in zebra finches. *Sci. Rep.* **7**, 8397. (doi:10.1038/s41598-017-08389-w)
59. Ghosh M, Rihel J. 2019 Hierarchical compression reveals sub-second to day-long structure in larval zebrafish behaviour. *BioRxiv*, 694471.
60. Dawkins R. 1976 *Hierarchical organisation: a candidate principle for ethology*. Cambridge, UK: Cambridge University Press.
61. Simon HA. 1991 The architecture of complexity. In *Facets of systems science, International Federation for Systems Research International Series on Systems Science and Engineering*, vol 7, pp. 457–476. Boston, MA: Springer. (doi:10.1007/978-1-4899-0718-9_31)
62. Todt D, Hultsch H. 1998 How songbirds deal with large amounts of serial information: retrieval rules suggest a hierarchical song memory. *Biol. Cybern.* **79**, 487–500. (doi:10.1007/s004220050498)
63. Koehlin E, Ody C, Kouneiher F. 2003 The architecture of cognitive control in the human prefrontal cortex. *Science* **302**, 1181–1185. (doi:10.1126/science.1088545)
64. Botvinick MM, Niv Y, Barto AC. 2009 Hierarchically organized behavior and its neural foundations: a reinforcement learning perspective. *Cognition* **113**, 262–280. (doi:10.1016/j.cognition.2008.08.011)
65. Ullman MT. 2001 A neurocognitive perspective on language: the declarative/procedural model. *Nat. Rev. Neurosci.* **2**, 717–726. (doi:10.1038/35094573)
66. Linkenkaer-Hansen K, Nikouline VV, Palva JM, Ilmoniemi RJ. 2001 Long-range temporal correlations and scaling behavior in human brain oscillations. *J. Neurosci.* **21**, 1370–1377. (doi:10.1523/JNEUROSCI.21-04-01370.2001)
67. He BJ, Zempel JM, Snyder AZ, Raichle ME. 2010 The temporal structures and functional significance of scale-free brain activity. *Neuron* **66**, 353–369. (doi:10.1016/j.neuron.2010.04.020)
68. Palva JM, Zsigalov A, Hirvonen J, Korhonen O, Linkenkaer-Hansen K, Palva S. 2013 Neuronal long-range temporal correlations and avalanche dynamics are correlated with behavioral scaling laws. *Proc. Natl Acad. Sci. USA* **110**, 3585–3590. (doi:10.1073/pnas.1216855110)
69. Nevill-Manning CG, Witten IH. 1997 Identifying hierarchical structure in sequences: a linear-time algorithm. *J. Artif. Intell. Res.* **7**, 67–82. (doi:10.1613/jair.374)
70. Gomez-Marín A, Stephens GJ, Brown AEX. 2016 Hierarchical compression of *Caenorhabditis elegans* locomotion reveals phenotypic differences in the organization of behaviour. *J. R. Soc. Interface* **13**, 20160466. (doi:10.1098/rsif.2016.0466)
71. Greenfield PM. 1991 Language, tools and brain: the ontogeny and phylogeny of hierarchically organized sequential behavior. *Behav. Brain Sci.* **14**, 531–551. (doi:10.1017/S0140525X00071235)
72. Meylan SC, Frank MC, Roy BC, Levy R. 2017 The emergence of an abstract grammatical category in children's early speech. *Psychol. Sci.* **28**, 181–192. (doi:10.1177/0956797616677753)
73. Crain S, Thornton R. 2012 Syntax acquisition. *Wiley Interdiscip. Rev.: Cogn. Sci.* **3**, 185–203. (doi:10.1002/wcs.1158)
74. Ravid D. 2005 Emergence of linguistic complexity in later language development: evidence from expository text construction. In *Perspectives on language and language development* (DD Ravid, HBZ Shyldkrot), pp. 337–355. Boston, MA: Springer. (doi:10.1007/1-4020-7911-7_25)
75. Berman R, Verhoeven L. 2002 Cross-linguistic perspectives on the development of text-production abilities: speech and writing. *Writ. Lang. Lit.* **5**, 1–43. (doi:10.1075/wll.5.1.02ber)
76. Rose Y, MacWhinney B. 2014 The PhonBank project: data and software-assisted methods for the study of phonology and phonological development. In *The Oxford handbook of corpus phonology* (eds J Durand, U Gut, G Kristoffersen). Oxford, UK: Oxford University Press. (doi:10.1093/oxfordhb/9780199571932.013.023)
77. MacWhinney B. 2000 *The childes project: tools for analyzing talk: volume I: transcription format and programs, volume II: the database*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.
78. Davis BL, MacNeilage PF. 1995 The articulatory basis of babbling. *J. Speech Lang. Hear. Res.* **38**, 1199–1211. (doi:10.1044/jshr.3806.1199)
79. Parsons JM. 2006 Positional effects in phonological development: a case study. PhD thesis, Memorial University of Newfoundland.
80. Yung Song J, Demuth K, Evans K, Shattuck-Hufnagel S. 2013 Durational cues to fricative codas in 2-year-olds' American English: voicing and morphemic factors. *J. Acoust. Soc. Am.* **133**, 2931–2946. (doi:10.1121/1.4795772)

81. Brown R. 1973 *A first language: the early stages*. Cambridge, MA: Harvard University Press.
82. Carterette EC, Jones MH. 1974 *Informal speech: alphabetic and phonemic texts with statistical analyses and tables*. Berkeley, CA: University of California Press.
83. Braunwald SR. 1971 Mother-child communication: the function of maternal-language input. *Word* **27**, 28–50. (doi:10.1080/00437956.1971.11435613)
84. Demetras MJ, Post KN, Snow CE. 1986 Feedback to first language learners: the role of repetitions and clarification questions. *J. Child Lang.* **13**, 275–292. (doi:10.1017/S030500090008059)
85. Masur EF, Gleason JB. 1980 Parent-child interaction and the acquisition of lexical information during play. *Dev. Psychol.* **16**, 404. (doi:10.1037/0012-1649.16.5.404)
86. Moerk E. 1972 Factors of style and personality. *J. Psycholinguist. Res.* **1**, 257–268. (doi:10.1007/BF01074442)
87. Gillam RB, Pearson NA. 2004 *TNL: test of narrative language*. Austin, TX: Pro-ed.
88. Moyle MJ, Weismer SE, Evans JL, Lindstrom MJ. 2007 Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *J. Speech Lang. Hear Res.* **50**, 508–528. (doi:10.1044/1092-4388(2007/035))
89. Nicholas JG, Geers AE. 1997 Communication of oral deaf and normally hearing children at 36 months of age. *J. Speech Lang. Hear Res.* **40**, 1314–1327. (doi:10.1044/jslhr.4006.1314)
90. Goad H, Buckley M. 2006 Prosodic structure in child French: evidence for the foot. *Catalan J. Linguist.* **5**, 109–142. (doi:10.5565/rev/catjl.81)
91. Hunkeler H. 2005 Aspects of the evolution of the interactions mother-child: case study of two dizygotic twin children between 15 and 26 months. PhD thesis, University of Rouen.
92. Kern S, Davis B, Zink I. 2009 *From babbling to first words in four languages: common trends across languages and individual differences*. Philadelphia, PA: John Benjamins Publishing Company.
93. Demuth K, Tremblay A. 2008 Prosodically-conditioned variability in children's production of French determiners. *J. Child Lang.* **35**, 99–127. (doi:10.1017/S0305000907008276)
94. Morgenstern A, Sekali M. 2009 What can child language tell us about prepositions? In *Studies in language and cognition* (eds J Zlatev, M Johansson Falck, C Lundmark, M Andrén), pp. 261–275. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.
95. Vihman MM, Macken MA, Miller R, Simmons H, Miller J. 1985 From babbling to speech: a re-assessment of the continuity issue. *Language* **61**, 397–445. (doi:10.2307/414151)
96. Doupe AJ, Kuhl PK. 1999 Birdsong and human speech: common themes and mechanisms. *Annu. Rev. Neurosci.* **22**, 567–631. (doi:10.1146/annurev.neuro.22.1.567)
97. Nevill-Manning CG, Witten IH. 2000 On-line and off-line heuristics for inferring hierarchies of repetitions in sequences. *Proc. IEEE* **88**, 1745–1755. (doi:10.1109/5.892710)
98. Grassberger P. 2003 Entropy estimates from insufficient samplings. arXiv physics/0307138. (<https://arxiv.org/abs/physics/0307138>)
99. Vinh NX, Epps J, Bailey J. 2010 Information theoretic measures for clusterings comparison: variants, properties, normalization and correction for chance. *J. Mach. Learn. Res.* **11**, 2837–2854.
100. Hubert L, Arabie P. 1985 Comparing partitions. *J. Classif.* **2**, 193–218. (doi:10.1007/BF01908075)
101. Vinh NX, Epps J, Bailey J. 2009 Information theoretic measures for clusterings comparison: is a correction for chance necessary? In *Proc. of the 26th Annual Int. Conf. on Machine Learning*, pp. 1073–1080.
102. Newville M, Stensitzki T, Allen DB, Rawlik M, Ingargiola A, Nelson A. 2016 LMFIT: non-linear least-square minimization and curve-fitting for Python. *Astrophys. Source Code Libr.*
103. Burnham KP, Anderson DR, Huyvaert KP. 2011 AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behav. Ecol. Sociobiol.* **65**, 23–35. (doi:10.1007/s00265-010-1029-6)
104. Kaplan RM, Kay M. 1994 Regular models of phonological rule systems. *Comput. Linguist.* **20**, 331–378.
105. Blum-Kulka S, Snow CE. 2002 *Talking to adults: the contribution of multiparty discourse to language acquisition*. New York, NY: Psychology Press.
106. Chen YG. 2017 Fractal analysis: applications in health sciences and social sciences. In *Fractal analysis* (ed. F Brambila), pp. 141–164. Rijeka, Croatia: InTechpages.
107. Li W. 1991 Expansion-modification systems: a model for spatial 1/f spectra. *Phys. Rev. A* **43**, 5240. (doi:10.1103/PhysRevA.43.5240)
108. Diniz A *et al.* 2011 Contemporary theories of 1/f noise in motor control. *Hum. Mov. Sci.* **30**, 889–905. (doi:10.1016/j.humov.2010.07.006)
109. Voytek B, Kramer MA, Case J, Lepage KQ, Tempesta ZR, Knight RT, Gazzaley A. 2015 Age-related changes in 1/f neural electrophysiological noise. *J. Neurosci.* **35**, 13 257–13 265. (doi:10.1523/JNEUROSCI.2332-14.2015)
110. Reinhart RMG, Nguyen JA. 2019 Working memory revived in older adults by synchronizing rhythmic brain circuits. *Nat. Neurosci.* **22**, 820–827. (doi:10.1038/s41593-019-0371-x)
111. Maylor EA, Chater N, Brown GDA. 2001 Scale invariance in the retrieval of retrospective and prospective memories. *Psychon. Bull. Rev.* **8**, 162–167. (doi:10.3758/BF03196153)
112. Stephen DG, Stepp N, Dixon JA, Turvey MT. 2008 Strong anticipation: sensitivity to long-range correlations in synchronization behavior. *Physica A* **387**, 5271–5278. (doi:10.1016/j.physa.2008.05.015)
113. Hauser MD, Chomsky N, Fitch WT. 2002 The faculty of language: what is it, who has it, and how did it evolve? *Science* **298**, 1569–1579. (doi:10.1126/science.298.5598.1569)
114. Sainburg T, Mai A, Gentner TQ. 2002 Long-range sequential dependencies precede complex syntactic production in language acquisition. Figshare.