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

Tim Sainburg[a,b], Anna Mai[c], and Timothy Q Gentner[a,d,e,f]
[a]Department of Psychology
[b]Center for Academic Research & Training in Anthropogeny
[c]Department of Linguistics
[d]Neurosciences Graduate Program
[e]Neurobiology Section
[f]Kavli Institute for Brain and Mind
UC San Diego
La Jolla, CA, 92093
{tsainburg, acmai, tgentner}@ucsd.edu

August 19, 2020

Abstract

To convey meaning, human language relies on hierarchically organized, long-range relationships spanning words, phrases, sentences, and discourse. The strength of the relationships between sequentially ordered elements of language (e.g., phonemes, characters, words) decays following a power law as a function of sequential distance. To understand the origins of these relationships, we examined long-range statistical structure in the speech of human children at multiple developmental time points, along with non-linguistic behaviors in humans and phylogenetically distant species. Here we show that adult-like power-law statistical dependencies precede the production of hierarchically-organized linguistic structures, and thus cannot be driven solely by these structures. Moreover, we show that similar long-range relationships occur in diverse non-linguistic behaviors across species. We propose that the hierarchical organization of human language evolved to exploit pre-existing long-range structure present in much larger classes of non-linguistic behavior, and that the cognitive capacity to model long-range hierarchical relationships preceded language evolution. We call this the Statistical Scaffolding Hypothesis for language evolution.

Keywords language · hierarchy · power law · evolution

1 Significance Statement

Human language is uniquely characterized by semantically meaningful hierarchical organization, conveying information over long timescales. At the same time, many non-linguistic human and animal behaviors are also often characterized by richly hierarchical organization. Here, we compare the long-timescale statistical dependencies present in language to those present in non-linguistic human and animal behaviors as well as language production throughout childhood. We find adult-like, long-timescale relationships early in language development, before syntax or complex semantics emerge, and we find similar relationships in non-linguistic behaviors like cooking and even housefly movement. These parallels demonstrate that long-range statistical dependencies are not unique to language and suggest a possible evolutionary substrate for the long-range hierarchical structure present in human language.

2 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, sequential relationships are calculated as a function of the sequential distance between events. For example, in the sequence $a \to b \to c \to d \to e \to f$, 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 grow weaker. 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 (Eq. 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 human language syntax [5], semantics [3], and discourse structure [4]. To understand the link between hierarchical organization in language and a power-law decay in sequential dependencies, it is helpful to consider both the latent and surface structure of a sequence (Fig. 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. 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 (Fig. 1). For sequences generated by a Markov process, the strength of the relationship between elements decays exponentially (Eq. 5) as sequential distance increases [5, 21] (Fig. 1A). In the PCFG model, however, linear distances in the sequence are coupled to logarithmic distances in the latent structure of the hierarchy (Fig. 1B-C). While information continues to decay exponentially as a function of the distance in the latent hierarchy (Fig. 1D), this log-scaling results in a power-law decay when MI is computed over corresponding sequential distances (Fig. 1E).

In language, long-range relationships convey meaning across hierarchical levels of organization. This latent linguistic structure is thought to underlie the power-law relationships observed across texts and speech [2–5]. The presence of power-law sequential and temporal relationships in natural phenomena is not restricted to human language, however. Here, we demonstrate that the power law underlying long-range statistical relationships in human speech precedes complex morphosyntactic production in language and is part of a larger set of natural behaviors exhibiting similar temporal relationships. The potentially numerous generative mechanisms for these phenomena remain to be established; however their existence evinces a substrate that may have been exploited in the evolution of a cognitive capacity to represent long-range signals prior to the evolution of language.

Beyond language, power-law temporal relationships are observed in both human-unique behaviors like music production [22] and stock market turbulence [23, 24] as well as behaviors that are shared with other animals such as sleep patterns in infants [25] and heart rates in healthy adults [26, 27]. In fact, the ubiquity of power laws in the physical and biological sciences spreads beyond temporal and sequential relationships and is well documented across a variety of phenomena. $1/f$ noise, a power law in the spectral density of a stochastic process, is observed in signals ranging from neural oscillations to flocking patterns in birds [28–31]. The relationship between biological variables often scale following a power law, for example, the allometric scaling laws observed between an organisms size and metabolic rate [32]. A variety of natural distributions such as word frequencies are well described by power-law distributions, a phenomenon termed Zipfs law [33–37]. Power-law distributions are also observed in the connectivity of many biological and social networks, a property called scale-freeness [38–41]. Over much of the past several decades, heated debates have arisen over claims of universal organizing principles of natural phenomena characterized by power laws [28, 31, 34, 41–44].

Across the diverse phenomena described by power-law relationships in the natural sciences, one commonality is that the origins of the observed power law are still not fully understood and mechanistic implications of power laws are often overstated [28, 31, 34, 41, 43, 44]. Although mechanisms have been proposed to

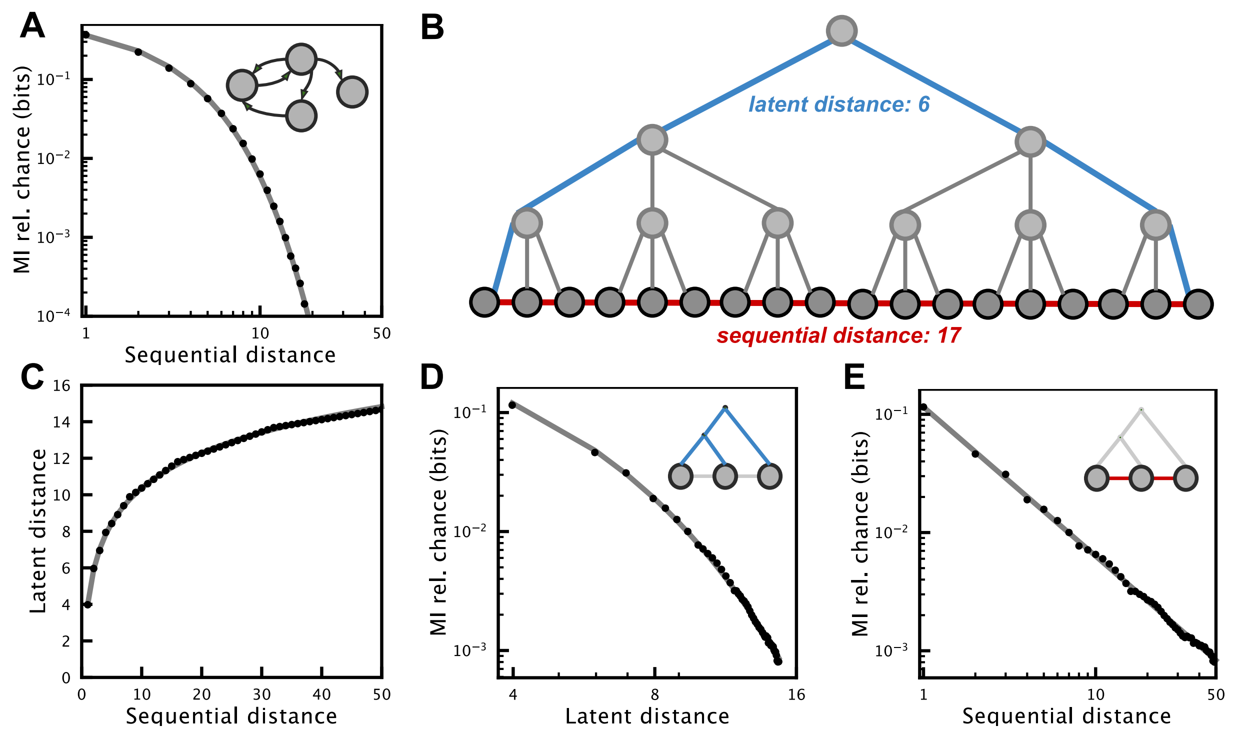

Figure 1: Comparison between sequences with deep latent relationships and iteratively generated sequences. (A) The MI between elements in an iteratively (Markov model) generated sequence decays exponentially as a function of sequential distance. (B) An example sequence with hierarchical latent structure. 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: $a * log_{x*b} + c$ 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 2 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.

account for the various forms of power laws observed in natural phenomena, the presence alone of a power law provides little insight into the underlying generative mechanism [31, 34, 42–44]. This is true of language as well. While the power laws characterized in language are consistent with generative mechanisms posited in syntactic theory [5, 45], they are not confirmatory. The presence of a power law in language does confirm, however, that relationships spanning long distances exist in the signal. Given the presence of power-law sequential relationships in human language, the question remains whether the power law is a product of linguistic structure, or whether these relationships originate in lower-level phenomena that are not unique to human language. If long-range relationships predate the evolution of language, they may have influenced the structure of temporal relationships that evolved with language.

Beyond human language, numerous other human behaviors [46–51], animal behaviors [52–57], animal vocalizations [37, 58–66], and other biologically-generated processes [25–27, 31, 67–70] have been described as being hierarchically organized or display long-timescale organization. Such behaviors range from the seemingly non-complex patterns of behavior exhibited by fruit flies [52, 56] to tool usage in great apes [53, 54]. For this reason, it has been argued that hierarchical organization is an inherent property of biological processes, including human behavior [50, 71, 72] and that the hierarchical structure of behavior is inherited from the lower-level organization of neurophysiological mechanisms that produce it [73–76], which themselves can be characterized by power-law relationships in temporal sequencing [29, 30, 77]. The developmental and/or evolutionary dependence of linguistic structure on underlying, domain-general, cognitive and neural processes has been posited by several researchers [50, 51, 76, 78].

Despite the numerous observations of hierarchical structure and long-range dependencies in non-human animal behaviors, few studies have examined the statistical dynamics of these behaviors quantitatively. Those that do have found power-law dynamics in the communication and behaviors of animals that are phylogenetically distant from humans [2, 79–81]. This, along with the prevalence of long-range power-law relationships in other natural phenomena [28, 31], supports the generality of these organizing principles across all behaviors. On the other hand, sequential organization in the vocal communication signals of non-human primates may extend over only a few elements [82, 83], and descriptions of hierarchical non-vocal behaviors in non-human primates tend to only be a few elements long [53, 54, 84], supporting at most a very shallow hierarchical structure. Thus, the extent to which a power-law decay provides a unified description of long-range statistical dependencies in behavior has yet to be determined. This question has particular relevance to human language, where it is unknown whether power-law relationships in sequential organization are present throughout language development, or emerge as linguistic structure develops. Understanding the ubiquity of power-law relationships across non-linguistic and non-human behavior, as well as across human language acquisition, may help to explain the origins of this organizing principle in language.

2.1 Present work

In the present work, we perform three groups of analyses exploring whether non-linguistic and pre-linguistic long-range statistical relationships parallel the long-range statistical relationships present in adult language. First, we analyze a series of language development corpora of children learning English, starting at six months of age [85–98], to determine whether long-range relationships are present in human vocalizations prior to the production of hierarchically-organized linguistic structure. Second, we analyze the long-range statistical dependencies of a human non-linguistic corpus of transcribed actions taken by humans while cooking [99], to determine whether power-law relationships are present in the sequential organization of non-linguistic human behaviors. Finally, we analyze the long-range sequential relationships in datasets of freely moving fruit flies (*Drosophila melanogaster*) [56] and zebrafish (*Danio rerio*) behavior [100], both of which have been previously characterized as being hierarchically organized, to determine whether a power law is present in the sequential organization of non-human non-linguistic behavior.

We show that both human non-linguistic and non-human non-linguistic behavior exhibits long-range power-law statistical dependencies like those observed in mature human language. In child language datasets, we observe a power-law as early as 6 to 12 months of age, while children are still in the "babbling" stage of language development. In the animal behavior datasets, we observe long-range power-law decays spanning many minutes (>6 minutes in *Drosophila* and >20 minutes in zebrafish).

3 Results

3.1 Language acquisition

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 [85, 86].

We first examined MI decay in sequences of words over nine datasets of natural speech from English speaking children included in the CHILDES repository [86, 91–98] and three datasets of sequences of phonemes from the PhonBank repository [85, 87–89], both of which are part of the TalkBank repository [86]. 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 6 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 analyzed as sequences of phonemes, where phoneme distributions for each transcript are treated independently to account for variation

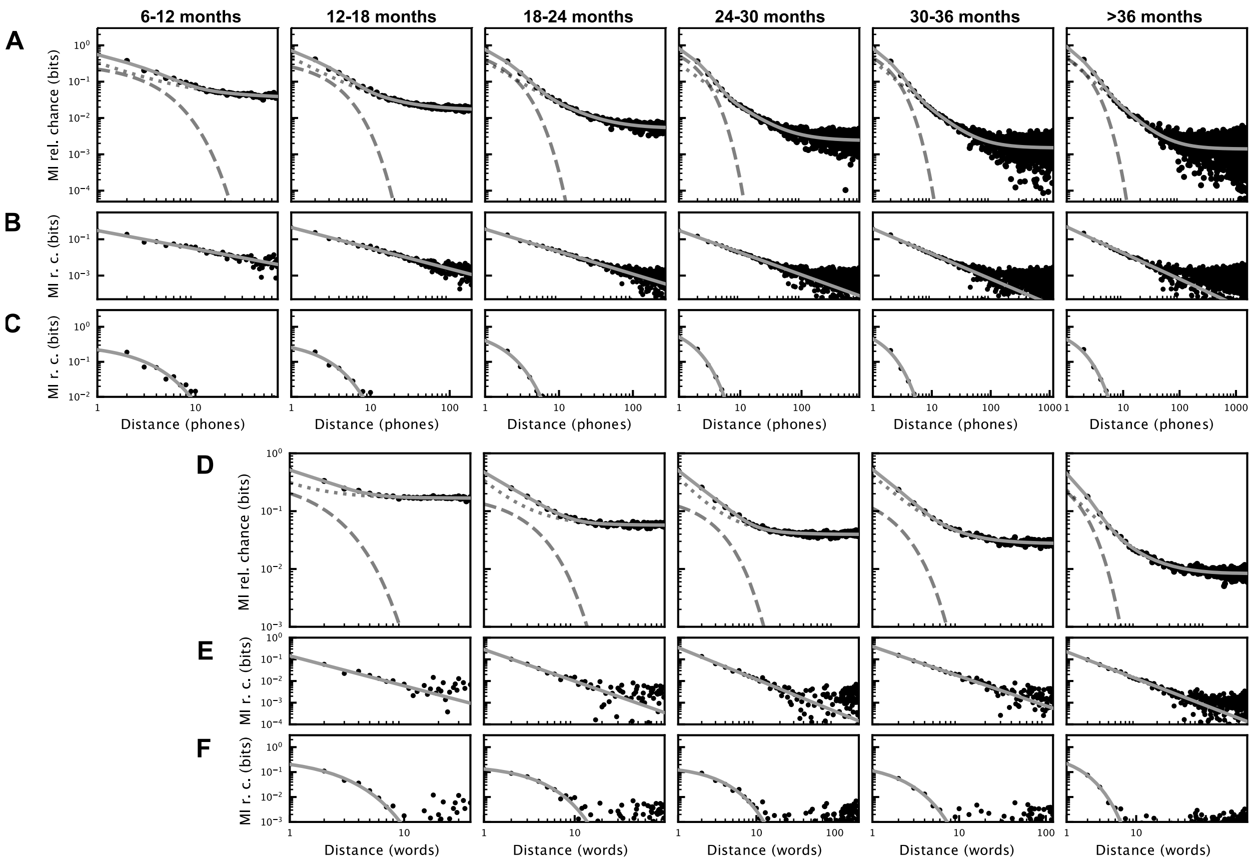

Figure 2: Mutual Information decay over words and phonemes during development. (A) MI decay over phonemes for each age group. MI decay is best fit by a composite model (solid grey line) for all age groups across phonemes and words. Exponential and power-law decays are shown as a dashed and dotted grey 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 same as in (B), but with the power-law component subtracted to show exponential component of the decay. (D-F) The same analyses as A-C, but for words.

in acquired vocabulary across individuals during development. Because transcript lengths varied between age groups (Fig. S1), we analyzed MI at sequential distances up to the median transcript length for each age group. Across all age groups, the decay in MI over sequences of phonemes is best fit by a composite power-law and exponential decay model (Fig. 2A-C; relative probabilities 0.897 to >0.999; Table S2). In each age group, we observe both a clear power law prominent over long distances (Fig. 2B) and a clear exponential decay at short word distances (Fig. 2C), consistent with prior results on adult speech [2].

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. The MI decay between words is best fit by a composite model of power-law and exponential decay (Eq. 7; relative probability $= 0.989$ for 12-18 months and > 0.999 for all other age groups; Fig. 2D-F; Table S1).

We also computed the MI decay over control sequences of words and phonemes that had been shuffled to isolate sequential relationships at different levels of organization (e.g. phoneme, word, utterance, transcript; Figs. S2, S3, S4). Consistent with Sainburg et al., [2], we observe that short-range relationships captured by exponential decay are largely carried within words and utterances, while long-range relationships captured by a power-law decay are carried across longer timescales between words and utterances. In particular, long-range relationships are eliminated when between-utterance structure is removed by randomly shuffling the order of utterances within a transcript (Figs. S2E, S3C) and retained when within-utterance structure is removed by shuffling words or phonemes within utterances (Figs. S2D, S3B) or phonemes within words (Fig. S2C). When MI decay is computed over part-of-speech labels for the words in CHILDES, we find a transition from MI decay that is best fit by a power-law decay alone at 12-24 months of age, to MI decay that is best fit by a composite model of power-law and exponential decay after 24 months (Fig S3D). Shuffling word order eliminates all long-range sequential relationships while preserving short timescale exponential relationships (Figs. S2B, S3E), and shuffling phoneme order within transcripts removes all sequential relationships (Figs. S2F). Across each shuffling analysis, we observe that short-range information content captured by exponential decay is largely captured within words and utterances, while long-range information is carried between utterances, even during early language production.

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 parallels the results across transcripts from Fig. 2 (Figs. S5, S6).

3.2 Human behavior

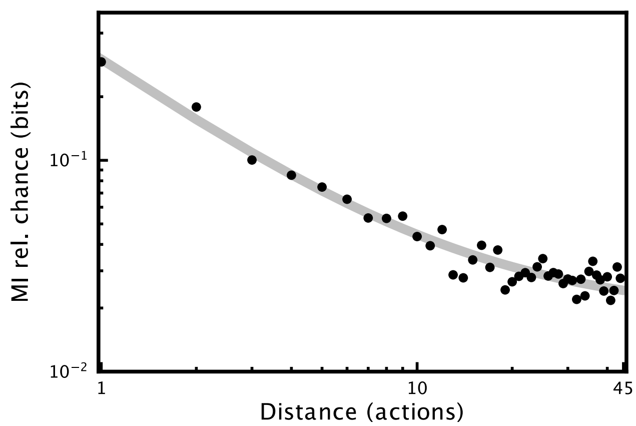

Figure 3: Mutual Information decay over actions in the Epic Kitchens dataset [99]. Data is fit by a power-law decay model (Eq. 6).

To contrast the long-range statistical structure of human language with non-linguistic human behaviors, we require a relatively large dataset of long, discrete, sequences of behavior. We chose the Epic Kitchens dataset [99], as it was the largest available segmented dataset of long sequences of individual actions, and because cooking has previously been described as having complex hierarchical syntactic structure [101].

The Epic Kitchens dataset consists of a series of videos in which each section of the video is labeled with an action and noun, for example *open door* $\rightarrow$ *turn-on light* $\rightarrow$ *close door* $\rightarrow$ *open fridge* $\rightarrow \ldots$. We calculate MI only over the sequences of verb classes, of which there are 119 unique classes. We computed the MI up to a distance of the median sequence length of 45 actions.

In contrast with the speech datasets, we found that the Epic Kitchens dataset was best fit by a power-law decay model with no exponential component (Eq. 6; Fig. 3; relative probability = 0.597; Table S3). We additionally looked at the MI decay of the longest cooking transcripts and found the MI decay of individual sequences were similar to MI decay across the entire dataset (Fig S7).

3.3 Animal behavior

The datasets of animal behavior used in our analyses were videos of zebrafish [100] and *Drosophila* [56] movements that had been transcribed in an unsupervised manner, i.e without external reference to *a priori* state labels. In both datasets, raw data recorded from individual animals were projected into a low-dimensional space and were then clustered into discrete states. These states were then labelled *post hoc* with human-interpretable descriptions such as "slow", "side leg", or "anterior" for *Drosophila*, and "O-bend" or "J-turn" for zebrafish. *Drosophila* behavior has a long history of being described in hierarchical terms [52, 56, 102], and the dataset used here, in particular, demonstrates long-range relationships extending over hundreds to thousands of states [56]. The zebrafish dataset used here has also previously been shown to contain sequential information that unfolds over multiple timescales [100, 103]. Both datasets were chosen because they contain large sets of discrete behaviors from individuals over long periods of time.

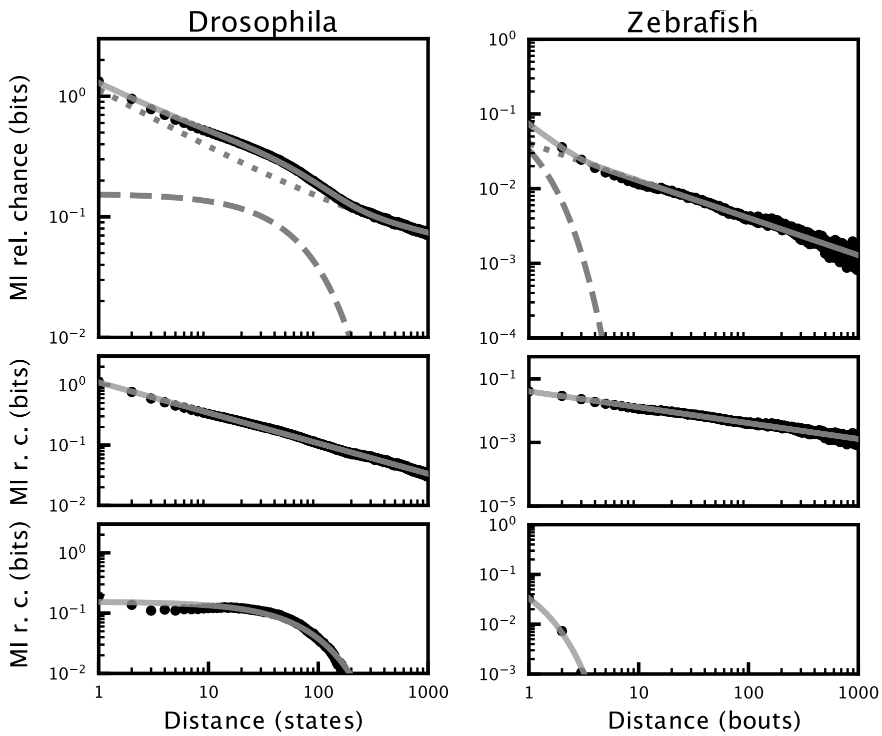

Figure 4: Mutual Information decay over Zebrafish and *Drosophila* behavior. Data is displayed in the same manner as Fig. 2.

In both the zebrafish and *Drosophila* datasets, we observe an MI decay that is best fit by a composite power-law and exponential decay model (Fig. 4; relative probabilities > 0.999; Table S3). The shape of the MI decay differs somewhat between the two datasets, however. In the case of the zebrafish, the relative contributions of the exponential and power-law components of the decay mirror the results obtained in speech. That is, an exponential component to the decay is observed at short distances under 10 elements, which gives way to a power-law at longer distances. In the case of the *Drosophila*, the power-law component of the decay is dominant throughout the signal, and the exponential component of the decay only captures a small portion of the variance at a distance of around 10-200 elements.

We additionally looked at a subset of the longest individual transcripts of *Drosophila* (Fig. S8) and zebrafish (Fig. S9) behavior and found that MI decay at the individual level varies between individual transcripts but matches the long-range decay observed across the datasets.

4 Discussion

We analyzed the long-range sequential information present in language production during development, and several sequentially organized and putatively hierarchical non-linguistic behaviors in other species. In all cases, the information between behavioral elements decays following a power law as sequential distance increases. For language, we find that that the long-range statistical relationships characteristic of adult usage [2] are present as early as 6 to 12 months in phonemes and 12-18 months in words, preceding the production of complex linguistic structure [84]. We see similar long-range power-law structure in the sequential organization of human food preparation and cooking. Cooking is a relatively modern and human-unique behavior [104], however, and may have arisen after humans developed more deeply hierarchical and highly planned tool usage behaviors [84, 105]. Yet, we also observe similar long-range organization in the movement patterns of *Drosophila* and zebrafish, consistent with previous reports for birdsong [2]. Long-range statistical relationships are present developmentally in speech before hierarchical linguistic structures are produced, and exist in widely varying animal species. Thus, the long-range statistical relationships present in language are not unique to linguistic behaviors or to humans.

These results compel reconsideration of the mechanisms that shape long-range statistical relationships in human language. 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 development provides a natural experiment in which one can examine human vocal communication absent the production of complex syntactic and semantic structures. Remarkably, even at a very early age, prior to the production of mature syntactic structures, vocal sequences show adult-like long-range dependencies. This does not rule out the possibility that long-range dependencies in adult language are driven in part by linguistic structure, but this hierarchical organization alone cannot explain our observations. What seems most reasonable to us, is that multiple mechanisms impose long-range dependencies on human speech and language, and that these operate on different developmental timescales. We take our observations of similar power laws in diverse non-linguistic behaviors to reinforce the idea that multiple mechanisms impose power-law dynamics on behavioral sequences. Indeed, power-laws are found in natural phenomena as distant from language as the sequential organization of earthquakes [106] and river water levels [107]. It may be that the power-law structure of human language reflects a very deep embedding of multiple, hierarchically structured complex systems, at varying levels of abstraction from linguistic, to motor control, to even more general underlying processes. Understanding the various power-law relationships in natural phenomena, and their origins, remains an area of active research [28, 31, 42].

Regardless of any deeper understanding of underlying mechanisms, our results demonstrate clear patterns in the information conveyed across time in both linguistic and non-linguistic behaviors. These patterns exist. Thus, they are potentially available and useful to any cognitive agent that engages with them. For example, in the movement patterns of a housefly, evolutionary fitness may be conferred to individuals (e.g. predators or mates) that can better anticipate the behavior of others by integrating long-range statistical dependencies. For human language, these selective advantages and abilities seem clear, as sensitivity to long-range organization has obvious benefit for comprehension. Outside of language, evidence for long-range sensitivities is more sparse, but humans do show scale invariance in retrospective memory tasks [108] and attention to power-law timescales in anticipation of future events in cognitive tasks [109]. The extent to which non-human animals are sensitive to the long-range dynamics (power-law or otherwise) of information in the environment is unknown. If non-human animals can model the long-range statistical dependencies present in their environment, this capacity would constitute a component of the broad faculty of language [110], that is, a necessary, but not uniquely-human, component of language. The presence of long-range statistical dependencies in non-linguistic behaviors and a generalized perceptual sensitivity to them would provide a scaffold on which language could evolve, and where hierarchical syntax and semantics can be understood as later additions that exploit existing long-range structures and sensitivities. We refer to this idea as the Statistical Scaffolding Hypothesis.

5 Methods

5.1 Mutual information

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

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

$$I(X,Y) = S(X) + S(Y) - S(X,Y) \tag{1}$$

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 [111] method which accounts for under-sampling true entropy from finite samples:

$$\hat{S} = \log_2(N) - \frac{1}{N}\sum_{i=1}^{K} N_i \psi(N_i) \tag{2}$$

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 ($\hat{I}_{sh}$):

$$MI = \hat{I} - \hat{I}_{sh} \tag{3}$$

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:

$$\hat{I}_{sh}(X,Y) = \hat{S}\left(X_{sh}\right) + \hat{S}\left(Y_{sh}\right) + \hat{S}\left(X_{sh}, Y_{sh}\right) \tag{4}$$

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 $S(Y_{sh})$. $\hat{I}_{sh}$ is related to the Expected Mutual Information [112–114] 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 "mutual information function", to distinguish it as the functional form of mutual information, which measures the dependency between two random variables [14]. In the mutual information 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 non-linear relationships [14].

5.2 Fitting mutual information decay

We fit the three following models:

An exponential decay model:

$$MI = a * e^{-x*b} + f \tag{5}$$

A power-law model:

$$MI = c * x^{d} + f \tag{6}$$

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

$$MI = a * e^{-x*b} + c * x^{d} + f \tag{7}$$

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 models 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 [115] 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, 116]. The parameters for each best-fit model for Figs 2, 3, and 4 can be found in Table 4.

5.3 Shuffling controls

The speech datasets are organized hierarchically into transcripts, utterances, words, and phonemes allowing us to shuffle the dataset at multiple levels of organization. In the Epic Kitchens, *Drosophila*, and zebrafish datasets no levels of organization were available beyond individual transcripts. To ensure that our MI decay results are a direct result of the sequential organization of each dataset, we performed a control in each dataset in which we shuffled behavioral elements within each individual transcript. In each case, the MI decay is flat confirming that the observed MI decay is a result of sequential organization (Figs S2F, S2E, S10). To ensure that long-range relationships were not due to trivial repetitions of behaviors, we looked in each dataset at MI decay over sequences in which repeated elements were removed. Removing repeats does not qualitatively alter the pattern of long-range relationships between elements (Fig. S4).

5.4 Data Availability

The five datasets can be acquired from the TalkBank repository [86], PhonBank repository [85], Berman et al. [56], Damen et al., [99], and Marques et al., [100]. We performed analyses over these transcripts without any modification. Example transcripts for each dataset are displayed in the Supplementary Information. The distribution of sequence lengths of each dataset is shown in Fig. S1. The code necessary for reproducing our results is available on GitHub [117].

5.5 Acknowledgements

Work 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 NIH DC0164081 and DC018055 to T.Q.G.

References

[1] Claude E Shannon. Prediction and entropy of printed english. *Bell system technical journal*, 30(1):50–64, 1951.

[2] Tim Sainburg, Brad Theilman, Marvin Thielk, and Timothy Q Gentner. Parallels in the sequential organization of birdsong and human speech. *Nature communications*, 10, 2019.

[3] Enrique Alvarez-Lacalle, Beate Dorow, J-P Eckmann, and Elisha Moses. Hierarchical structures induce long-range dynamical correlations in written texts. *Proceedings of the National Academy of Sciences*, 103(21):7956–7961, 2006.

[4] Eduardo G Altmann, Giampaolo Cristadoro, and Mirko Degli Esposti. On the origin of long-range correlations in texts. *Proceedings of the National Academy of Sciences*, 109(29):11582–11587, 2012.

[5] Henry Lin and Max Tegmark. Critical behavior in physics and probabilistic formal languages. *Entropy*, 19(7):299, 2017.

[6] Peter Grassberger. Estimating the information content of symbol sequences and efficient codes. *IEEE Transactions on Information Theory*, 35(3):669–675, 1989.

[7] Alain Schenkel, Jun Zhang, and Yi-Cheng Zhang. Long range correlation in human writings. *Fractals*, 1(01):47–57, 1993.

[8] Werner Ebeling and Thorsten Pöschel. Entropy and long-range correlations in literary english. *EPL (Europhysics Letters)*, 26(4):241, 1994.

[9] Paolo Allegrini, Paolo Grigolini, and Luigi Palatella. Intermittency and scale-free networks: a dynamical model for human language complexity. *Chaos, Solitons & Fractals*, 20(1):95–105, 2004.

[10] SS Melnyk, OV Usatenko, VA Yampolskii, and VA Golick. Competition between two kinds of correlations in literary texts. *Physical Review E*, 72(2):026140, 2005.

[11] Marcelo A Montemurro and Damián H Zanette. Entropic analysis of the role of words in literary texts. *Advances in complex systems*, 5(01):7–17, 2002.

[12] Marcelo A Montemurro and Damián H Zanette. Towards the quantification of the semantic information encoded in written language. *Advances in Complex Systems*, 13(02):135–153, 2010.

[13] Marcelo A Montemurro and Pedro A Pury. Long-range fractal correlations in literary corpora. *Fractals*, 10(04):451–461, 2002.

[14] Wentian Li. Mutual information functions versus correlation functions. *Journal of statistical physics*, 60(5-6):823–837, 1990.

[15] Edwin B. Newman and Louis J. Gerstman. A new method for analyzing printed english. *Journal of experimental psychology*, 44(2):114–125, 08 1952.

[16] N. G. Burton and C. R. Licklider. Long-range constraints in the statistical structure of printed english. *American Journal of Psychology*, 68(4):650, Dec 01 1955.

[17] Edwin B. Newman. The pattern of vowels and consonants in various languages. *The American Journal of Psychology*, 64:369–379, 1951.

[18] Thomas Cover and Roger King. A convergent gambling estimate of the entropy of english. *IEEE Transactions on Information Theory*, 24(4):413–421, 1978.

[19] Huitao Shen. Mutual information scaling and expressive power of sequence models. *arXiv preprint arXiv:1905.04271*, 2019.

[20] Richard Futrell, Kyle Mahowald, and Edward Gibson. Large-scale evidence of dependency length minimization in 37 languages. *Proceedings of the National Academy of Sciences*, 112(33):10336–10341, 2015.

[21] Wentian Li. Power spectra of regular languages and cellular automata. *Complex Systems*, 1(1):107–130, 1987.

[22] Daniel J Levitin, Parag Chordia, and Vinod Menon. Musical rhythm spectra from bach to joplin obey a 1/f power law. *Proceedings of the National Academy of Sciences*, 109(10):3716–3720, 2012.

[23] Rosario N Mantegna and H Eugene Stanley. Stock market dynamics and turbulence: parallel analysis of fluctuation phenomena. *Physica A: Statistical Mechanics and its Applications*, 239(1-3):255–266, 1997.

[24] Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4):394–419, 1963.

[25] E Canessa and A Calmetta. Physics of a random biological process. *Physical Review E*, 50(1):R47, 1994.

[26] Masanori Kobayashi and Toshimitsu Musha. 1/f fluctuation of heartbeat period. *IEEE transactions on Biomedical Engineering*, (6):456–457, 1982.

[27] C-K Peng, J Mietus, JM Hausdorff, Shlomo Havlin, H Eugene Stanley, and Ary L Goldberger. Long-range anticorrelations and non-gaussian behavior of the heartbeat. *Physical review letters*, 70(9):1343, 1993.

[28] Miguel A Munoz. Colloquium: Criticality and dynamical scaling in living systems. *Reviews of Modern Physics*, 90(3):031001, 2018.

[29] Klaus Linkenkaer-Hansen, Vadim V. Nikouline, J. Matias Palva, and Risto J. Ilmoniemi. Long-range temporal correlations and scaling behavior in human brain oscillations. *The Journal of Neuroscience*, 21(4):1370–1377, February 2001.

[30] Biyu J He, John M Zempel, Abraham Z Snyder, and Marcus E Raichle. The temporal structures and functional significance of scale-free brain activity. *Neuron*, 66(3):353–369, 2010.

[31] T Gisiger. Scale invariance in biology: coincidence or footprint of a universal mechanism? *Biological Reviews*, 76(2):161–209, 2001.

[32] Geoffrey B West, James H Brown, and Brian J Enquist. A general model for the origin of allometric scaling laws in biology. *Science*, 276(5309):122–126, 1997.

[33] Aaron Clauset, Cosma Rohilla Shalizi, and Mark EJ Newman. Power-law distributions in empirical data. *SIAM review*, 51(4):661–703, 2009.

[34] Mark EJ Newman. Power laws, pareto distributions and zipf's law. *Contemporary physics*, 46(5):323–351, 2005.

[35] Ramon Ferrer i Cancho and Ricard V Solé. Least effort and the origins of scaling in human language. *Proceedings of the National Academy of Sciences*, 100(3):788–791, 2003.

[36] Ryuji Suzuki, John R Buck, and Peter L Tyack. The use of zipf's law in animal communication analysis. *Animal Behaviour*, 69(1):F9–F17, 2005.

[37] Ramon Ferrer-i Cancho and Brenda McCowan. The span of correlations in dolphin whistle sequences. *Journal of Statistical Mechanics: Theory and Experiment*, 2012(06):P06002, 2012.

[38] Anna D Broido and Aaron Clauset. Scale-free networks are rare. *Nature communications*, 10(1):1–10, 2019.

[39] Albert-László Barabási. Scale-free networks: a decade and beyond. *science*, 325(5939):412–413, 2009.

[40] John C Doyle, David L Alderson, Lun Li, Steven Low, Matthew Roughan, Stanislav Shalunov, Reiko Tanaka, and Walter Willinger. The robust yet fragile nature of the internet. *Proceedings of the National Academy of Sciences*, 102(41):14497–14502, 2005.

[41] Evelyn Fox Keller. Revisiting scale-free networks. *BioEssays*, 27(10):1060–1068, 2005.

[42] Wentian Li. Expansion-modification systems: a model for spatial 1/f spectra. *Physical Review A*, 43(10):5240, 1991.

[43] Michael PH Stumpf and Mason A Porter. Critical truths about power laws. *Science*, 335(6069):665–666, 2012.

[44] Guido Boffetta, Vincenzo Carbone, Paolo Giuliani, Pierluigi Veltri, and Angelo Vulpiani. Power laws in solar flares: self-organized criticality or turbulence? *Physical review letters*, 83(22):4662, 1999.

[45] Noam Chomsky. On certain formal properties of grammars. *Information and control*, 2(2):137–167, 1959.

[46] Richard Cooper and Tim Shallice. Contention scheduling and the control of routine activities. *Cognitive neuropsychology*, 17(4):297–338, 2000.

[47] W Tecumseh Fitch and Mauricio D Martins. Hierarchical processing in music, language, and action: Lashley revisited. *Annals of the New York Academy of Sciences*, 1316(1):87–104, 2014.

[48] Andrew Whiten, Emma Flynn, Katy Brown, and Tanya Lee. Imitation of hierarchical action structure by young children. *Developmental science*, 9(6):574–582, 2006.

[49] Matthew M Botvinick. Hierarchical models of behavior and prefrontal function. *Trends in cognitive sciences*, 12(5):201–208, 2008.

[50] Karl Spencer Lashley. *The problem of serial order in behavior*, volume 21. Bobbs-Merrill, 1951.

[51] Valeri Aleksandrovich Kozhevnikov and Liudmila Andreevna Chistovich. Speech: Articulation and perception. 1965.

[52] Marian Dawkins and Richard Dawkins. Hierachical organization and postural facilitation: Rules for grooming in flies. *Animal Behaviour*, 24(4):739–755, 1976.

[53] Jill D Pruetz and Paco Bertolani. Savanna chimpanzees, pan troglodytes verus, hunt with tools. *Current biology*, 17(5):412–417, 2007.

[54] Richard W Byrne and Jennifer ME Byrne. Complex leaf-gathering skills of mountain gorillas (gorilla g. beringei): variability and standardization. *American Journal of Primatology*, 31(4):241–261, 1993.

[55] Louis Lefebvre. Grooming in crickets: timing and hierarchical organization. *Animal Behaviour*, 29(4):973–984, 1981.

[56] Gordon J Berman, William Bialek, and Joshua W Shaevitz. Predictability and hierarchy in drosophila behavior. *Proceedings of the National Academy of Sciences*, 113(42):11943–11948, 2016.

[57] Louis Lefebvre. The organization of grooming in budgerigars. *Behavioural processes*, 7(2):93–106, 1982.

[58] Arik Kershenbaum, Ann E Bowles, Todd M Freeberg, Dezhe Z Jin, Adriano R Lameira, and Kirsten Bohn. Animal vocal sequences: not the Markov chains we thought they were. *Proceedings of the Royal Society of London B: Biological Sciences*, 281(1792):20141370, 2014.

[59] Tina C Roeske, Damian Kelty-Stephen, and Sebastian Wallot. Multifractal analysis reveals music-like dynamic structure in songbird rhythms. *Scientific Reports*, 8(1):4570, 2018.

[60] Jeffrey E Markowitz, Elizabeth Ivie, Laura Kligler, and Timothy J Gardner. Long-range order in canary song. *PLoS Computational Biology*, 9(5):e1003052, 2013.

[61] Richard W Hedley. Composition and sequential organization of song repertoires in Cassin's vireo (Vireo cassinii). *Journal of Ornithology*, 157(1):13–22, 2016.

[62] Kazutoshi Sasahara, Martin L Cody, David Cohen, and Charles E Taylor. Structural design principles of complex bird songs: a network-based approach. *PLoS One*, 7(9):e44436, 2012.

[63] Ryuji Suzuki, John R Buck, and Peter L Tyack. Information entropy of humpback whale songs. *The Journal of the Acoustical Society of America*, 119(3):1849–1866, 2006.

[64] Xinjian Jiang, Tenghai Long, Weicong Cao, Junru Li, Stanislas Dehaene, and Liping Wang. Production of supra-regular spatial sequences by macaque monkeys. *Current Biology*, 28(12):1851–1859, 2018.

[65] Julia Hyland Bruno and Ofer Tchernichovski. Regularities in zebra finch song beyond the repeated motif. *Behavioural Processes*, 2017.

[66] Takashi Morita, Hiroki Koda, Kazuo Okanoya, and Ryosuke O Tachibana. Measuring long context dependency in birdsong using an artificial neural network with a long-lasting working memory. *bioRxiv*, 2020.

[67] H Eugene Stanley, Viktor Afanasyev, Luis A Nunes Amaral, SV Buldyrev, AL Goldberger, Steve Havlin, Harry Leschhorn, P Maass, Rosario N Mantegna, C-K Peng, et al. Anomalous fluctuations in the dynamics of complex systems: from dna and physiology to econophysics. *Physica A: Statistical Mechanics and its Applications*, 224(1-2):302–321, 1996.

[68] Wentian Li and Kunihiko Kaneko. Long-range correlation and partial $1/f\alpha$ spectrum in a noncoding dna sequence. *EPL (Europhysics Letters)*, 17(7):655, 1992.

[69] C-K Peng, Sergej V Buldyrev, Ary L Goldberger, Shlomo Havlin, Francesco Sciortino, Michael Simons, and HE Stanley. Long-range correlations in nucleotide sequences. *Nature*, 356(6365):168, 1992.

[70] Gandhimohan M Viswanathan, V Afanasyev, SV Buldyrev, EJ Murphy, PA Prince, and H Eugene Stanley. Lévy flight search patterns of wandering albatrosses. *Nature*, 381(6581):413, 1996.

[71] Richard Dawkins. Hierarchical organisation: A candidate principle for ethology. 1976.

[72] Herbert A Simon. The architecture of complexity. In *Facets of systems science*, pages 457–476. Springer, 1991.

[73] Dietmar Todt and Henrike Hultsch. How songbirds deal with large amounts of serial information: retrieval rules suggest a hierarchical song memory. *Biological cybernetics*, 79(6):487–500, 1998.

[74] Etienne Koechlin, Chrystele Ody, and Frédérique Kouneiher. The architecture of cognitive control in the human prefrontal cortex. *Science*, 302(5648):1181–1185, 2003.

[75] Matthew M Botvinick, Yael Niv, and Andrew C Barto. Hierarchically organized behavior and its neural foundations: A reinforcement learning perspective. *Cognition*, 113(3):262–280, 2009.

[76] Michael T Ullman. A neurocognitive perspective on language: The declarative/procedural model. *Nature reviews neuroscience*, 2(10):717–726, 2001.

[77] J Matias Palva, Alexander Zhigalov, Jonni Hirvonen, Onerva Korhonen, Klaus Linkenkaer-Hansen, and Satu Palva. Neuronal long-range temporal correlations and avalanche dynamics are correlated with behavioral scaling laws. *Proceedings of the National Academy of Sciences*, 110(9):3585–3590, 2013.

[78] Morten H Christiansen and Nick Chater. Language as shaped by the brain. *Behavioral and brain sciences*, 31(5):489–509, 2008.

[79] Shouwen Ma, Andries Ter Maat, and Manfred Gahr. Power-law scaling of calling dynamics in zebra finches. *Scientific reports*, 7(1):8397, 2017.

[80] G. M. Viswanathan, V. Afanasyev, S. V. Buldyrev, E. J. Murphy, P. A. Prince, and H. E. Stanley. Lévy flight search patterns of wandering albatrosses. *Nature*, 381(6581):413–415, May 1996.

[81] Luiz G. A. Alves, Peter B. Winter, Leonardo N. Ferreira, Renée M. Brielmann, Richard I. Morimoto, and Luís A. N. Amaral. Long-range correlations and fractal dynamics in c. elegans : Changes with aging and stress. *Physical Review E*, 96(2), August 2017.

[82] Christopher I Petkov and Erich Jarvis. Birds, primates, and spoken language origins: behavioral phenotypes and neurobiological substrates. *Frontiers in evolutionary neuroscience*, 4:12, 2012.

[83] Christopher I Petkov and Benjamin Wilson. On the pursuit of the brain network for proto-syntactic learning in non-human primates: conceptual issues and neurobiological hypotheses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367(1598):2077–2088, 2012.

[84] Patricia M Greenfield. Language, tools and brain: The ontogeny and phylogeny of hierarchically organized sequential behavior. *Behavioral and brain sciences*, 14(4):531–551, 1991.

[85] Yvan Rose and Brian MacWhinney. The phonbank project: Data and software-assisted methods for the study of phonology and phonological development. 2014.

[86] Brian MacWhinney. The childes project: Tools for analyzing talk: Volume i: Transcription format and programs, volume ii: The database, 2000.

[87] Barbara L Davis and Peter F MacNeilage. The articulatory basis of babbling. *Journal of Speech, Language, and Hearing Research*, 38(6):1199–1211, 1995.

[88] Jennifer M Parsons. *Positional effects in phonological development: a case study.* PhD thesis, Memorial University of Newfoundland, 2006.

[89] Jae Yung Song, Katherine Demuth, Karen Evans, and Stefanie Shattuck-Hufnagel. Durational cues to fricative codas in 2-year-olds' american english: Voicing and morphemic factors. *The Journal of the Acoustical Society of America*, 133(5):2931–2946, 2013.

[90] Roger Brown. *A first language: The early stages.* Harvard U. Press, 1973.

[91] Edward C Carterette and Margaret Hubbard Jones. *Informal speech: Alphabetic & phonemic texts with statistical analyses and tables.* Univ of California Press, 1974.

[92] Susan R Braunwald. Mother-child communication: the function of maternal-language input. *Word*, 27(1-3):28–50, 1971.

[93] Marty J Demetras, Kathryn Nolan Post, and Catherine E Snow. Feedback to first language learners: The role of repetitions and clarification questions. *Journal of child language*, 13(2):275–292, 1986.

[94] Elise F Masur and Jean B Gleason. Parent–child interaction and the acquisition of lexical information during play. *Developmental Psychology*, 16(5):404, 1980.

[95] Ernst Moerk. Factors of style and personality. *Journal of psycholinguistic research*, 1(3):257–268, 1972.

[96] Ronald Bradley Gillam and Nils A Pearson. *TNL: test of narrative language*. Pro-ed Austin, TX, 2004.

[97] Maura Jones Moyle, Susan Ellis Weismer, Julia L Evans, and Mary J Lindstrom. Longitudinal relationships between lexical and grammatical development in typical and late-talking children. *Journal of Speech, Language, and Hearing Research*, 2007.

[98] Johanna G Nicholas and Ann E Geers. Communication of oral deaf and normally hearing children at 36 months of age. *Journal of Speech, Language, and Hearing Research*, 40(6):1314–1327, 1997.

[99] Dima Damen, Hazel Doughty, Giovanni Maria Farinella, Sanja Fidler, Antonino Furnari, Evangelos Kazakos, Davide Moltisanti, Jonathan Munro, Toby Perrett, Will Price, and Michael Wray. Scaling egocentric vision: The epic-kitchens dataset. In *European Conference on Computer Vision (ECCV)*, 2018.

[100] João C Marques, Simone Lackner, Rita Félix, and Michael B Orger. Structure of the zebrafish locomotor repertoire revealed with unsupervised behavioral clustering. *Current Biology*, 28(2):181–195, 2018.

[101] Hilde Kuehne, Ali Arslan, and Thomas Serre. The language of actions: Recovering the syntax and semantics of goal-directed human activities. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 780–787, 2014.

[102] Andrew M Seeds, Primoz Ravbar, Phuong Chung, Stefanie Hampel, Frank M Midgley Jr, Brett D Mensh, and Julie H Simpson. A suppression hierarchy among competing motor programs drives sequential grooming in drosophila. *Elife*, 3:e02951, 2014.

[103] Marcus Ghosh and Jason Rihel. Hierarchical compression reveals sub-second to day-long structure in larval zebrafish behaviour. *bioRxiv*, page 694471, 2019.

[104] Richard Wrangham. *Catching fire: how cooking made us human*. Basic Books, 2009.

[105] Dietrich Stout, Thierry Chaminade, Andreas Thomik, Jan Apel, and A Aldo Faisal. Grammars of action in human behavior and evolution. *bioRxiv*, page 281543, 2018.

[106] Kim Christensen, Leon Danon, Tim Scanlon, and Per Bak. Unified scaling law for earthquakes. *Proceedings of the National Academy of Sciences*, 99(suppl 1):2509–2513, 2002.

[107] M Sadegh Movahed and Evalds Hermanis. Fractal analysis of river flow fluctuations. *Physica A: Statistical Mechanics and its Applications*, 387(4):915–932, 2008.

[108] Elizabeth A Maylor, Nick Chater, and Gordon DA Brown. Scale invariance in the retrieval of retrospective and prospective memories. *Psychonomic Bulletin & Review*, 8(1):162–167, 2001.

[109] Damian G Stephen, Nigel Stepp, James A Dixon, and MT Turvey. Strong anticipation: Sensitivity to long-range correlations in synchronization behavior. *Physica A: Statistical Mechanics and its Applications*, 387(21):5271–5278, 2008.

[110] Marc D Hauser, Noam Chomsky, and W Tecumseh Fitch. The faculty of language: what is it, who has it, and how did it evolve? *Science*, 298(5598):1569–1579, 2002.

[111] Peter Grassberger. Entropy estimates from insufficient samplings. *arXiv preprint physics/0307138*, 2003.

[112] Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. *Journal of Machine Learning Research*, 11(Oct):2837–2854, 2010.

[113] Lawrence Hubert and Phipps Arabie. Comparing partitions. *Journal of classification*, 2(1):193–218, 1985.

[114] Nguyen Xuan Vinh, Julien Epps, and James Bailey. Information theoretic measures for clusterings comparison: is a correction for chance necessary? In *Proceedings of the 26th annual international conference on machine learning*, pages 1073–1080, 2009.

[115] Matthew Newville, Till Stensitzki, Daniel B Allen, Michal Rawlik, Antonino Ingargiola, and Andrew Nelson. Lmfit: non-linear least-square minimization and curve-fitting for Python. *Astrophysics Source Code Library*, 2016.

[116] Kenneth P. Burnham, David R. Anderson, and Kathryn P. Huyvaert. Aic model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology and Sociobiology*, 65(1):23–35, Jan 2011.

[117] Tim Sainburg. Code for "long-range sequential dependencies are phylogenetically pervasive in behavior and precede complex syntactic production in language". https://github.com/timsainb/LongRangeSequentialOrgPaper, 2020.

6 Supplementary Materials

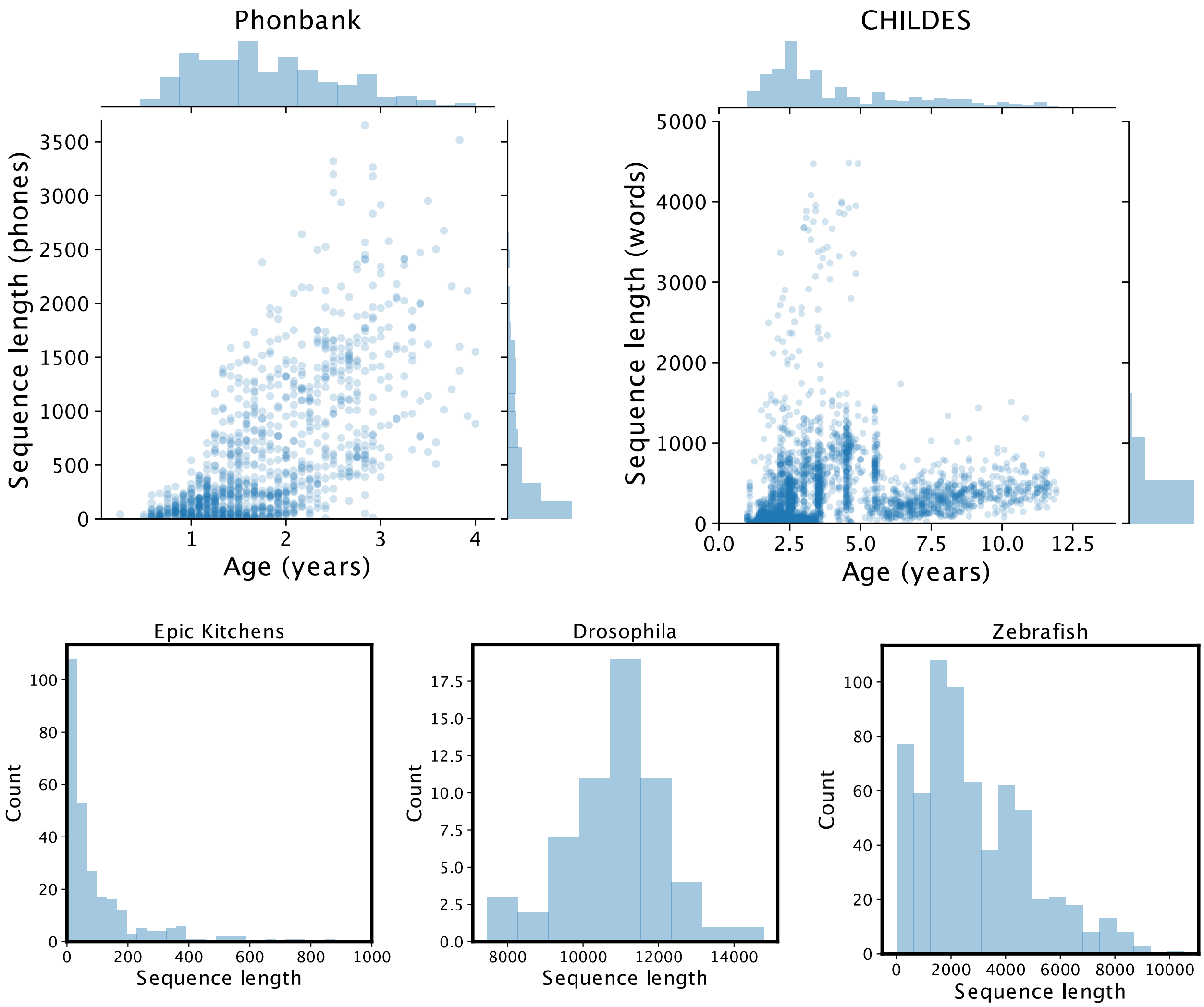

Figure S1: Distribution of sequence lengths for each dataset.

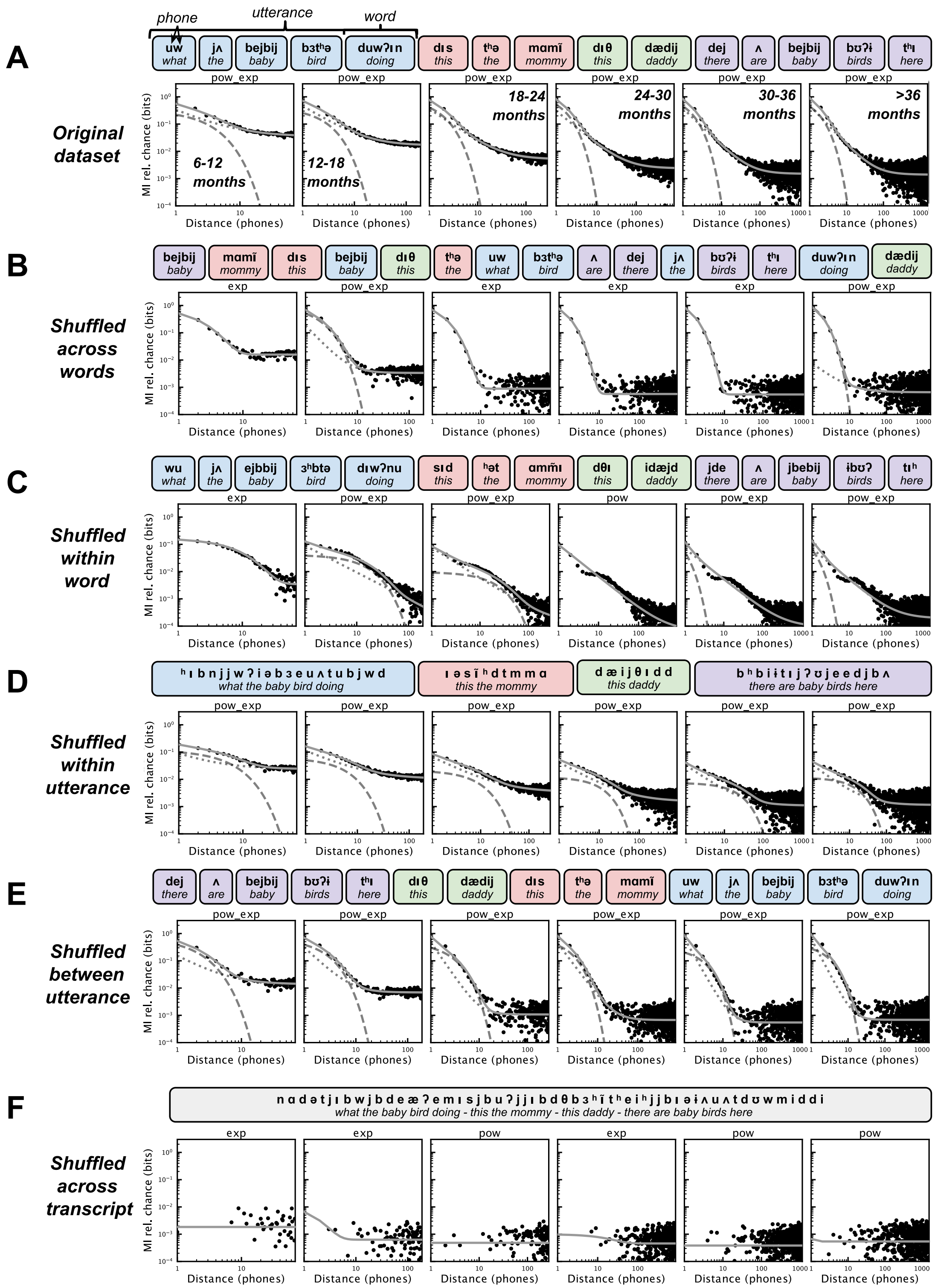

Figure S2: MI decay between phones under different shuffling conditions. (A) MI decay for each age group from the entire dataset, as in Fig. 2A. The sequence above the MI decay shows an example set of utterances of the corpus to illustrate the shuffling conditions. Utterances are grouped by color, words are grouped by rounded rectangles, and phones are displayed in bold above orthographic transcriptions. (B) Words are shuffled within each transcript. (C) Phones are shuffled within words. (D) Phones are shuffled within utterances. (E) Utterances are shuffled within each transcript. (F) Phones are shuffled within each transcript. The best fit model is printed above each plot, and is plotted as grey lies alongside the data and in Fig. 1.

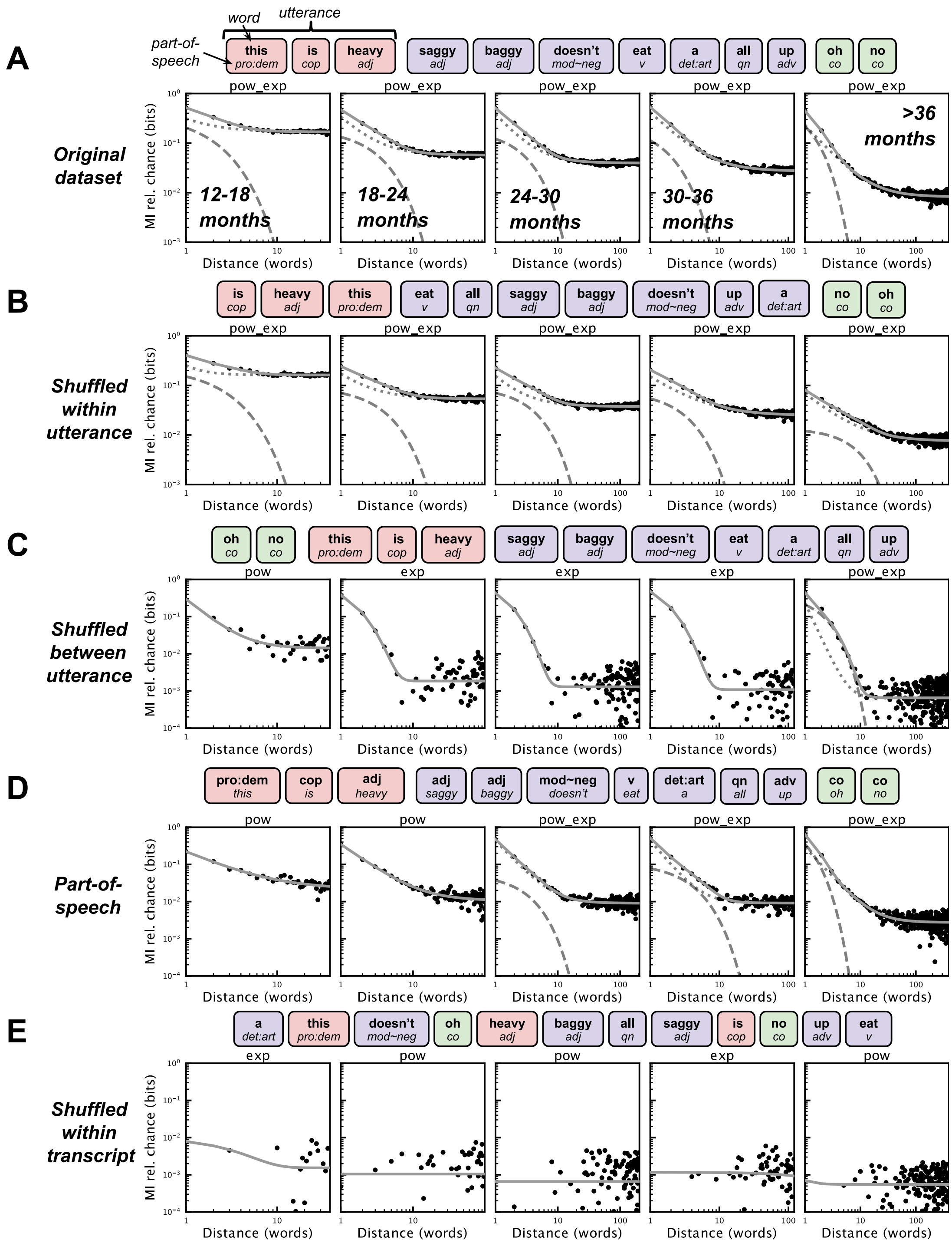

Figure S3: MI decay between words under different shuffling conditions. (A) MI decay for each age group from the entire dataset, as in Fig. 2D. (B) Words are shuffled within each utterance. (C) Utterances are shuffled within each transcript. (D) MI is calculated over part-of-speech transcriptions of words. (E) Words are shuffled within each transcript. (F) Words are shuffled within each transcript. The best fit model is printed above each plot, and is plotted as grey lies alongside the data and in Fig. 1.

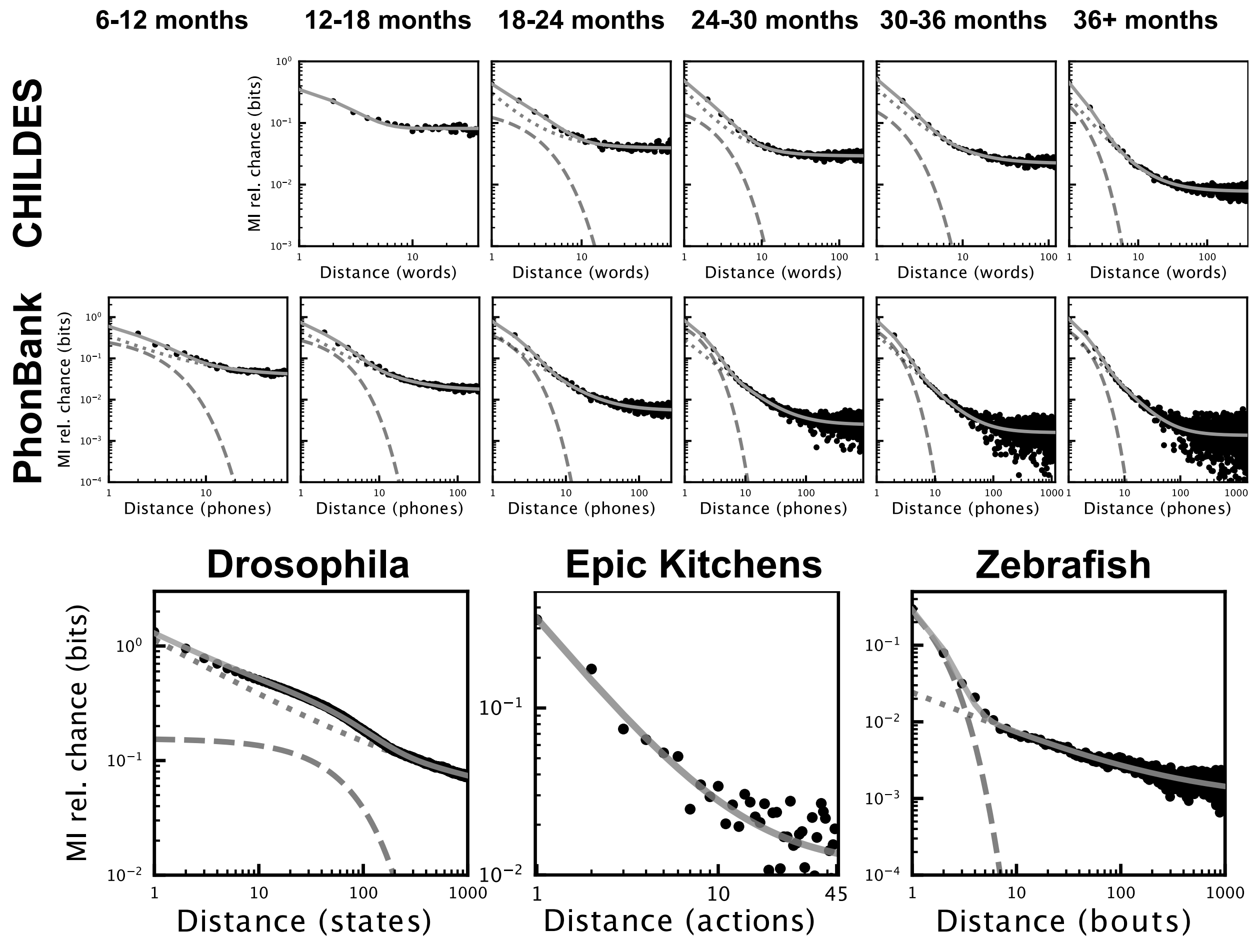

Figure S4: MI decay with repeated elements removed across each dataset.

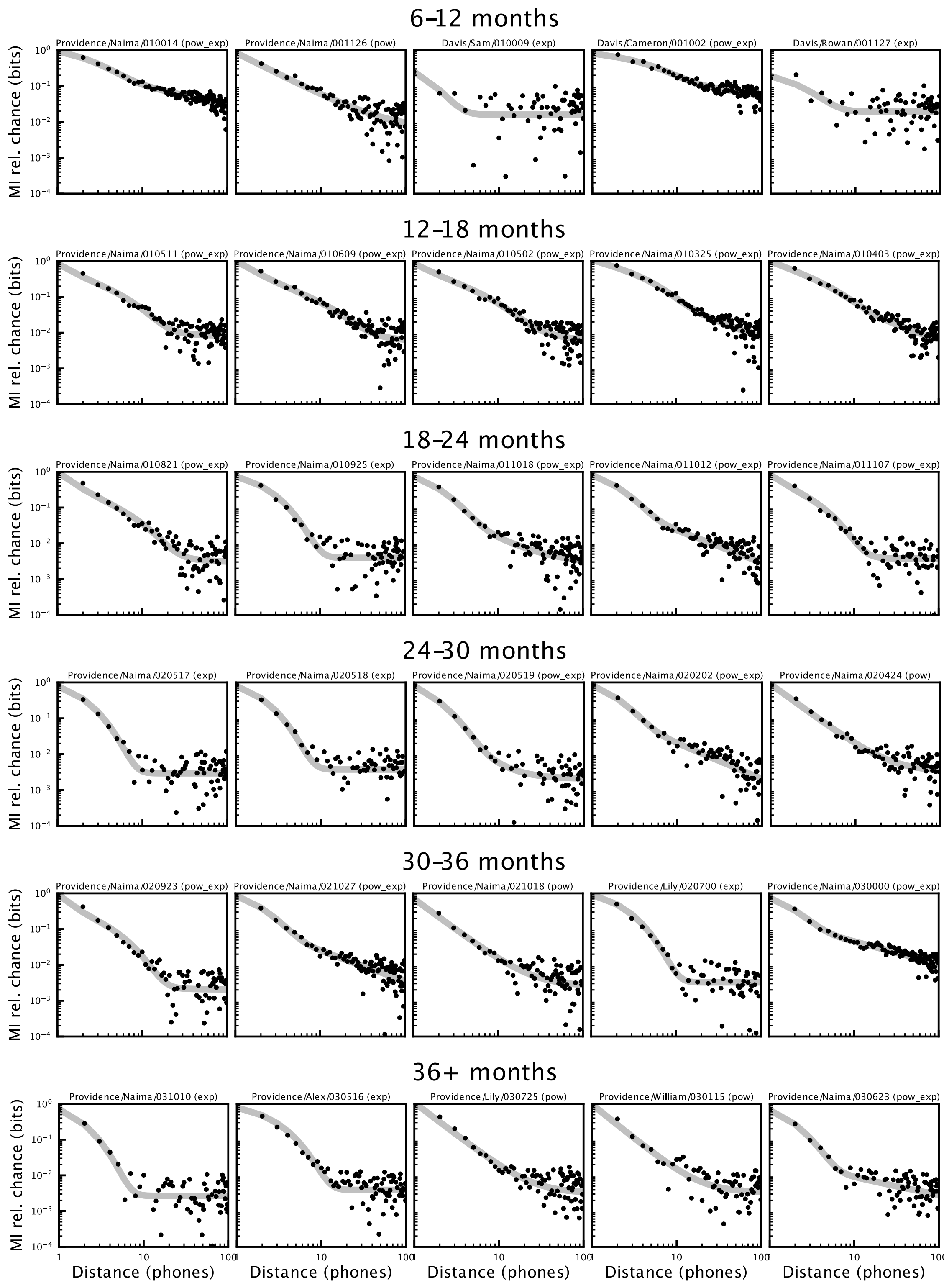

Figure S5: MI decay and best fit model of five largest transcripts for each age group across PhonBank. Transcript identity and best fit model are displayed above each plot.

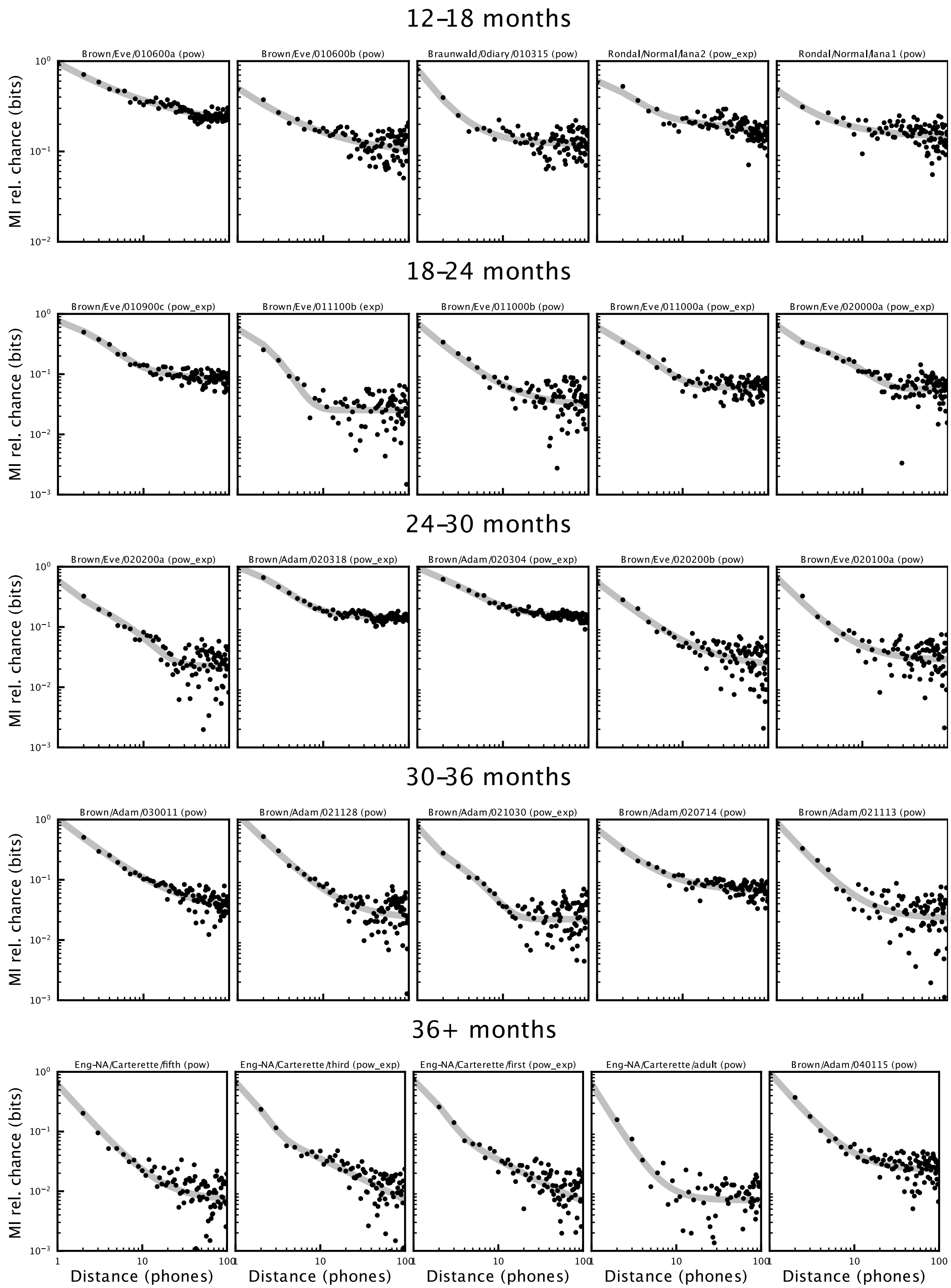

Figure S6: MI decay and best fit model of five largest transcripts for each age group across CHILDES. Transcript identity and best fit model are displayed above each plot.

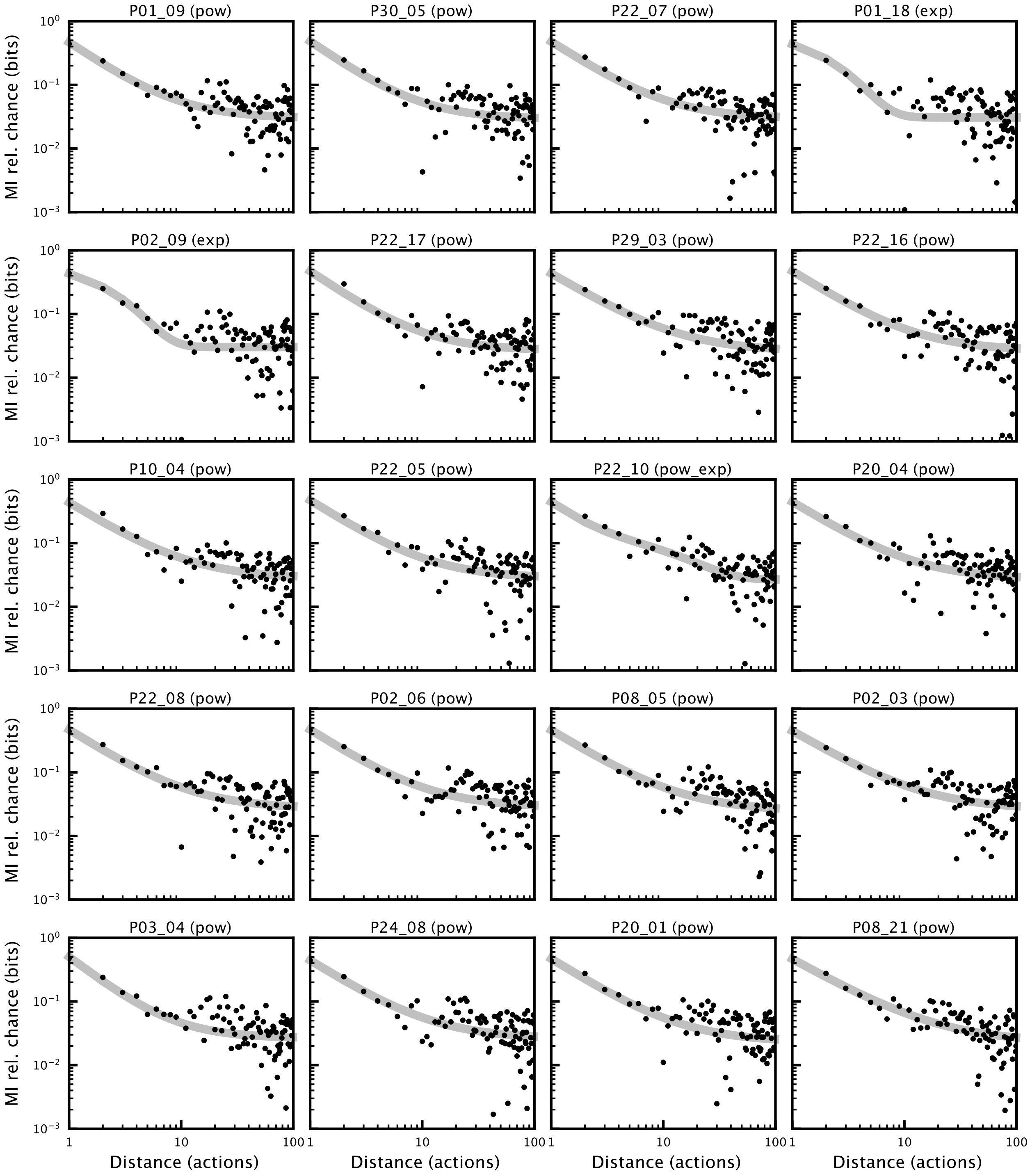

Figure S7: MI decay over the 20 longest Epic kitchens cooking sequences. Transcript identity and best fit model are displayed above each plot.

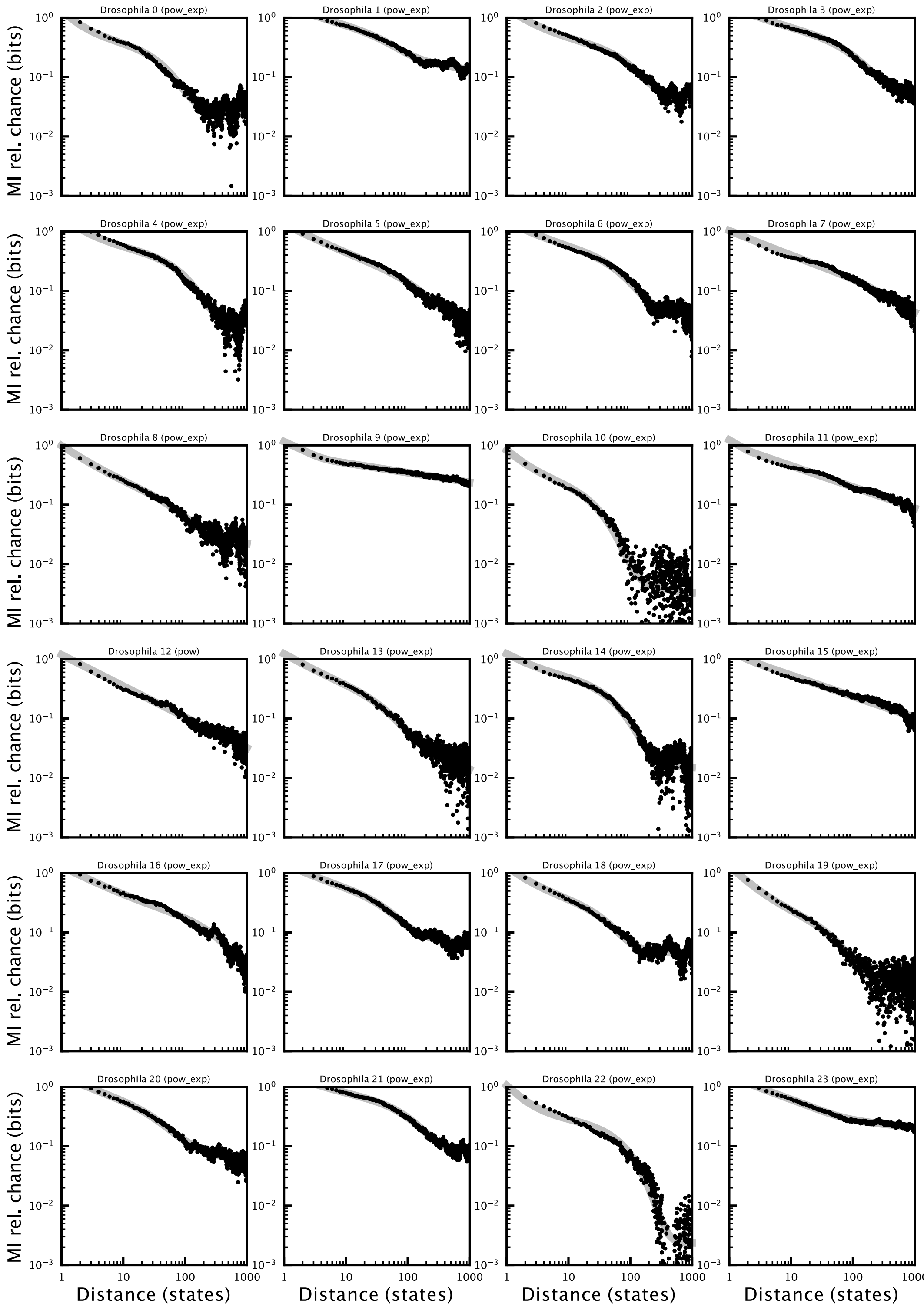

Figure S8: MI decay of example individual Drosophila behavioral sequences over one hour. Transcript identity and best fit model are displayed above each plot.

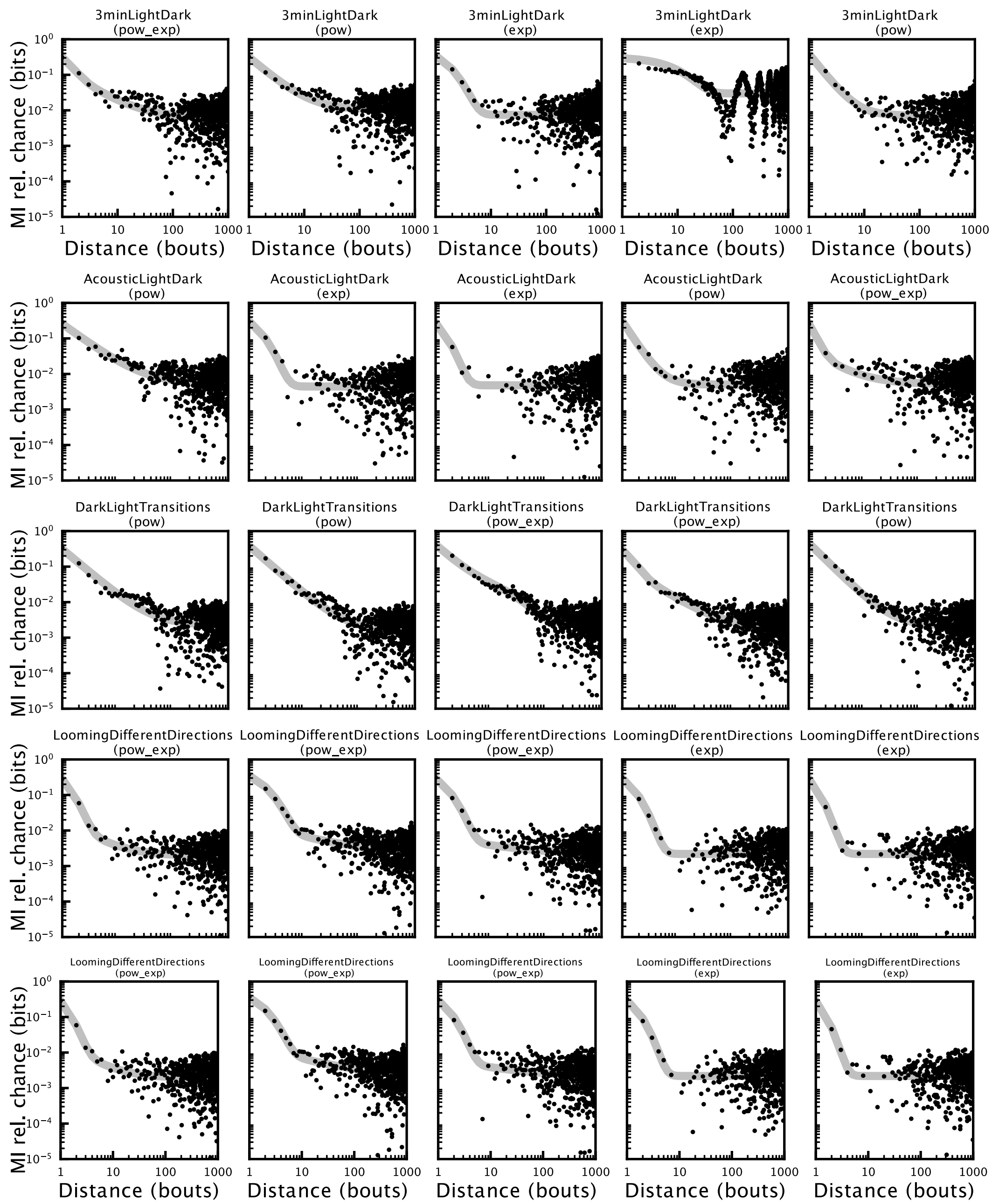

Figure S9: MI decay of several individual Zebrafish behavioral sequences. Each plot corresponds to the continuous behavior of a single Zebrafish. Each row corresponds to a different behavioral setting. The behavioral setting is written above the plot alongside the best fit model.

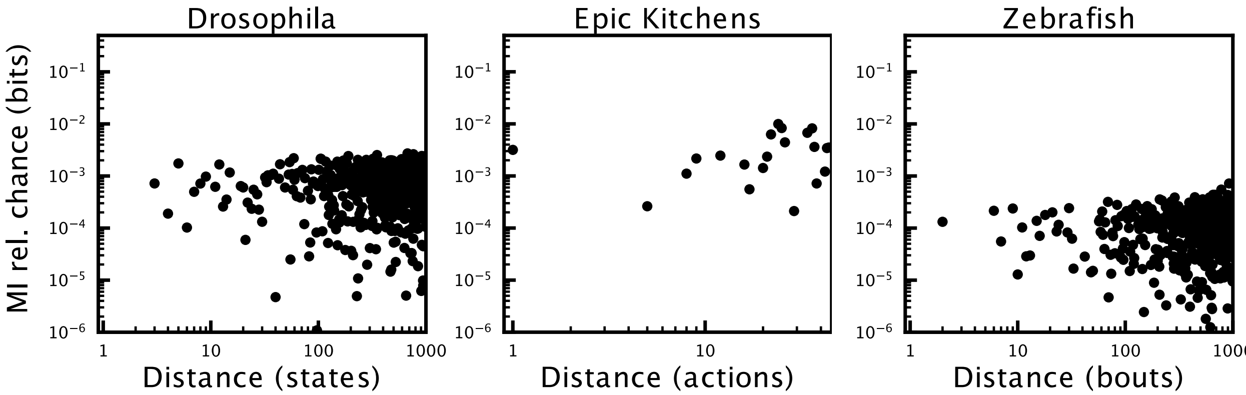

Figure S10: MI decay of shuffled sequences for Drosophila, Zebrafish, and Epic Kitchens datasets. No information decay is seen between elements of any sequence.

		12-18 months	18-24 months	24-30 months	30-36 months	3+ years
AICc	**exp.**	-313.876	-696.201	-1464.82	-735.697	-2314.6
	combined	-322.789	-819.742	-1737.37	-951.049	-2989.21
	power-law	-296.67	-746.061	-1623	-933.579	-2939.72
r^2	**exp.**	0.997	0.992	0.991	0.986	0.973
	combined	0.998	0.998	0.998	0.998	0.995
	power-law	0.995	0.995	0.996	0.997	0.994
Relative likelihood	**exp.**	0.012	<0.001	<0.001	<0.001	<0.001
	combined	>0.999	>0.999	>0.999	>0.999	>0.999
	power-law	<0.001	<0.001	<0.001	<0.001	<0.001
Relative probability	**exp.**	0.011	<0.001	<0.001	<0.001	<0.001
	combined	0.989	>0.999	>0.999	>0.999	>0.999
	power-law	<0.001	<0.001	<0.001	<0.001	<0.001

Table 1: CHILDES dataset model fit results for each decay model as shown in Fig. 2.

		6-12 months	12-18 months	18-24 months	24-30 months	30-36 months	3+ years
AICc	**exp.**	-5687.13	-4842.29	-4240.44	-1371.61	-1091.13	-417.621
	combined	-5998.03	-5302.25	-5025.96	-1903.5	-1522.77	-484.369
	power-law	-5993.7	-5288.72	-4971.83	-1836.16	-1369.5	-437.315
r^2	**exp.**	0.803	0.878	0.928	0.967	0.983	0.989
	combined	0.841	0.919	0.971	0.995	0.998	0.996
	power-law	0.841	0.918	0.969	0.994	0.996	0.992
Relative likelihood	**exp.**	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	combined	>0.999	>0.999	>0.999	>0.999	>0.999	>0.999
	power-law	0.115	0.001	<0.001	<0.001	<0.001	<0.001
Relative probability	**exp.**	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001
	combined	0.897	0.999	>0.999	>0.999	>0.999	>0.999
	power-law	0.103	0.001	<0.001	<0.001	<0.001	<0.001

Table 2: PhonBank dataset model fit results for each decay model as shown in Fig. 2.

		Cooking	Drosophila	Zebrafish
AICc	**exp.**	-236.312	-6513.67	-5125.71
	combined	-269.057	-11115.3	-7340.27
	power-law	-269.846	-8894.93	-6066.59
r^2	**exp**	0.98	0.952	0.918
	combined	0.991	0.999	0.991
	power-law	0.991	0.996	0.968
Relative likelihood	**exp.**	<0.001	<0.001	<0.001
	combined	0.674	>0.999	>0.999
	power-law	>0.999	<0.001	<0.001
Relative probability	**exp.**	<0.001	<0.001	<0.001
	combined	0.403	>0.999	>0.999
	power-law	0.597	<0.001	<0.001

Table 3: Epic Kitchens, Drosophila, and Zebrafish model fit results at 45, 1000, and 1000 elements of distance respectively.

Dataset	Age (yrs)	a	b	c	d	f
CHILDES	1-1.5	0.387±0.101	0.645±0.113	0.145±0.038	-1.382±0.345	0.168±0.003
	1.5-2.0	0.194±0.022	0.382±0.034	0.283±0.016	-1.461±0.083	0.057±0.001
	2-2.5	0.185±0.022	0.418±0.033	0.346±0.014	-1.464±0.04	0.04±0.0
	2.5-3.0	0.239±0.099	0.753±0.105	0.391±0.039	-1.367±0.053	0.027±0.0
	>3	0.639±0.065	1.082±0.047	0.223±0.022	-1.238±0.041	0.008±0.0
PhonBank	0.5-1	0.326±0.065	0.391±0.045	0.301±0.041	-1.013±0.087	0.035±0.002
	1-1.5	0.404±0.047	0.463±0.021	0.446±0.029	-1.137±0.027	0.016±0.0
	1.5-2	0.891±0.098	0.794±0.032	0.358±0.042	-1.234±0.044	0.005±0.0
	2-2.5	1.225±0.136	0.877±0.054	0.305±0.043	-1.219±0.046	0.002±0.0
	2.5-3	1.112±0.255	0.908±0.1	0.38±0.082	-1.381±0.07	0.001±0.0
	>3	1.019±0.371	0.857±0.137	0.476±0.132	-1.433±0.087	0.001±0.0
Drosophila	-	0.155±0.002	0.014±0.0	1.1±0.004	-0.506±0.002	0.04±0.001
Zebrafish	-	0.943±0.054	1.33±0.051	0.06±0.005	-0.661±0.052	0.0±0.001
Cooking	-	-	-	0.227±0.029	-1.133±0.18	0.023±0.003

Table 4: MI decay parameters for Figs 2, 3, and 4. The parameters correspond to Equation 7 ($a * e^{-x*b} + c * x^d + f$). a and b for the Cooking dataset are not shown because the best-fit model is the power-law model.

7 Example sequences from datasets

7.1 PhonBank

A random sample of the transcripts used in this manuscript at different ages. Each line corresponds to an utterance and each utterance is followed by an orthographic representation in parentheses. 'xxx' in orthographic transcription refers to unintelligible speech and 'yyy' refers to phonological coding. The meanings of other coding symbols such as '@' and '&' used in orthographic representations can be found in the TalkBank manuals for PhonBank and CHILDES.

7.1.1 Davis/Nate/001105.xml 11 months

hɛ (xxx)
je (xxx)
gɪg (xxx)
ɛ (xxx)
ʔe (xxx)
ʔɪʔe (xxx)
hɔ (xxx)
jæhɛʔ (xxx)
ʔæ (xxx)
hɛʔ (xxx)
he (xxx)
he (xxx)
ʔɪ (xxx)
hɪ (xxx)
hæ (xxx)
hɛ (xxx)
ʔɛ (xxx)
ʔɛːæ (xxx)
ɛæː (xxx)
ɛ (xxx)
æa (xxx)
ʔɛʔʔɪʔ (xxx)
ʔɛ (xxx)
ʔɛ (xxx)
dɪ (xxx)
ɛʔɛːæː (xxx)
jɛjĩ (j@l)
jæ̃jɛ̃ (xxx)
hæ (xxx)
hɛ (xxx)
hæh (xxx)
hɛ (xxx)
ʔɛ (xxx)
hæ (xxx)
bʷʌʔβː (xxx)
ʔɛ (xxx)
hʌjʌlalalajæ (xxx)
bababa (xxx)
ʔɛoʷː (xxx)
bɪː (xxx)
jae (xxx)
æ (xxx)
hɛ (xxx)
βː (xxx)
dejehɛ (xxx)
ejɛːhe (xxx)
æ (xxx)
dʷæ (xxx)
ʔʌːoʷ (xxx)
m̩ (xxx)
hæ (xxx)
ʔæʔʌʔdɪ (xxx)
pʰ (xxx)
m̩bʊʔ (xxx)
bʊbwɪ (xxx)
ʔɛː (xxx)
ɛ̃jæ̃ (xxx)
hʌː (xxx)
mʌ (xxx)
ɛ (xxx)
hɛjæ (xxx)
dæwu (xxx)
wɛ (xxx)
hɪ (xxx)
ʔɪʔɪhɛɛːʔɛʔɛ̃ (xxx)
hɛːjæɛ (xxx)
ʔeʔ (xxx)
ɛæːe (xxx)
ʔɪʔɛ (xxx)
jæwɛ (xxx)
ɛ (xxx)
ʔɪ (xxx)
ʔeː (xxx)
ʔɛ (xxx)
hɛ (xxx)
ɛ (xxx)
ʔɪ (xxx)
ʔæ̃ʔ (xxx)
ʔɛ (xxx)
ijɛ̃ː (xxx)
hæi (xxx)
hɛdʰ (xxx)
hɪ (xxx)
læ (xxx)
ʔʌ (xxx)
tɪtɪːdɛ (xxx)
sædɪ (xxx)
ʔʌːæo (xxx)
ʔæ (xxx)
ʔɛː (xxx)
ʔuʃ (xxx)
wɪ (xxx)
hɛː (xxx)
hɛ (xxx)
ʔæɪjɛː (xxx)
ʔʌɰoʔ (xxx)
ʔɪʔ (xxx)
ʔɪʔɛː (xxx)
ʔɛːæɛ (xxx)
æ (xxx)
ɛ (xxx)
ʔɛ (xxx)
ʔɛ (xxx)
gʊʧ (xxx)

7.1.2 Providence/William/011115.xml 23 months

wəʃ ˈdi (what's this)
ˈni (yyy)
ˈʌ di ˈkwomə (are yyy yyy)
uˈkwo ˈwɑː (yyy yyy)
əˈkwo ˈwɑ (yyy yyy)
ˈmɑˈmi (mommy)
ˈjɑmi (yummy)
ˈðus (juice)
ˈjʌmiː (yummy)
ˈgʊ ˈdʒus (good juice)
ˈjɑ (yah)
ˈaʊ ɪnə ˈtɑˈmeɪ (I wanna Thomas)
ˈʌwə ˈtɑmʊt (yyy Thomas)
ˈtɑm ɪˈɪʃɪʔ (Thomas yyy)
ˈbʌˈkeɪ (pocket)
ˈnoʔ ˈno ˈbɑgɪt (yyy no pocket)
ˈno ˈbʌgˈɛt (no pocket)
ˈno ˈbʌkɛt (no pocket)
ˈnu (no)
ˈokeɪ (okay)
ˈɔ (yyy)
ˈokeɪ (okay)
ˈokeɪ (okay)
ˈjɛ (yeah)

ˈogɛ (okay)
ˈeɪ (yyy)
ˈwaɪ (why)
wə ˈtow ɪz ɪt (what time is it)
ˈwaɪ ˈwaɪ (why why)
ˈno (no)
ˈokeɪ (okay)
ˈjɛ (yeah)
ˈnːoː (no:)
ˈopɛn (open)
ˈo (no)
oˈbɛn (open)
ˈdæˈɾi (daddy)
ˈdæˈɾi (daddy)
ˈdæˈɾi (daddy)
ˈdæˈɾi (daddy)
ˈdæɾi (daddy)
ˈdæˈɾi (daddy)
əˈdæɾi (daddy)
ˈdæː (daddy)
ˈno (no)
ˈno (no)
ˈbɑˈbʌs (yyy)
ˈno (no)
ˈno ˈdʒɪkə ˈbu bum (no ↪ chicka_boom_boom@si)
ˈʌ ˈno ɔ ˈdʌn (yyy no all done)
ˈaɪjə ˈnʌʔ ˈʌʔ ˈnu ˈgʌmə (yyy yyy ↪ yyy yyy yyy)
ˈwʌn ˈdæd ˈɑmɑ (wan dad yyy)
ˈno (no)
ˈno ˈɑ ˈwɑ (no ice pop)
ˈno (no)
ˈɛ ˈno (yyy no)
ˈæbəlæs (ambulance)
ˈhæmbəˈlɪnt (hi ambulance)
ˈæbəlæns (ambulance)
əˈwæˈwɪw (yyy)
ˈfaɪjəɛˈdʒɪnt (fire+engine)
ˈno (no)
ˈno ˈtʃrʌk (no truck)
ˈwɑ də ˈtivi (watch the tv)
ˈbɔni (Barney)
ˈbɔni (Barney)
ˈnʌʔˈo (no)
ˈmʊʔ (yyy)
ˈwʌ ˈhɪˈjʌ (right here)
ˈʌ ˈwʌɾ ˈiʔ (yyy what it)
ˈmʌ wʌɾ ˈɪz ˈɪt ˈtwʌk (yyy what is ↪ it truck)
ˈuː ˈu (ooh ooh)
ˈnoː (no)
ˈʌʔ ˈo (uhoh)
ðə ˈdʌmˈtrʌk (the dump+truck)
tʃ̂ˈʌk (truck)
ˈiˈnaɪt (night+night)
* ˈtʃrʌk (xxx truck)
* ˈtsʙʌk (xxx truck)
* ˈtrʌk (xxx truck)
* (xxx)
ˈdi jə ˈsiː (do you see)
ˈnɪni ˈditʃi ˈtʃrʌk (yyy yyy truck)
ˈmɪbɛbit * (yyy xxx)
ˈtʃʌk (truck)
ˈnɪ ˈnɪnəðəðə ˈtrʌk (yyy yyy truck)
* ˈtʃrʌk (xxx truck)
* ˈtrʌk (xxx truck)
* ˈtʃrʌk (xxx truck)
ˈdʌˈtʃrʌk (dump+truck)
ɪz ˈdæ ə ˈtʃrʌk (is that a truck)
ˈhɪʒ ˈtrʌk (a truck)
ˈtrʌk ˈdæt ˈtʃrʌk (truck that truck)
ˈtʃʌ * (truck xxx)
ˈoˈheɪ (okay)
ˈʌʔ ˈʌ ˈɪzə * ˈpʌzə (yyy yyy yyy ↪ xxx puzzle)
* (xxx)
* (xxx)
ˈdɑ (yeah)
ˈnoː ˈno ˈno ˈnop (no: no no no)
ˈeɪ ˈbi ˈsiːz (abcs)

—— (continued) ——

7.1.3 Goad/Julia/20510.xml 29 months

tʰɜʃ pʰɑpʰ (toast pop)
ʔʌ bɨlũːw (a balloon)
bɪŋkʰ babɔ (big bubble)
ə najɨn (a lion)
wʊhəs dat kʰijə duwɪn (what's ↪ that kid doing)
dʌn pʰɛŋkʰ (can 0of paint)
wʌs də mæn dowɛ̃ (what's the ↪ man doing)
kʰʌpfajə (campfire)
ə̃kʰɛpfaj tʰɛw maj mam (campfire ↪ tell my mom)
mej kʰæpfajʌ (make campfire)
tʰʌ̥muw tʰʌmɪn (camel coming)
dɪ hə bəlow (this is blue)
ʔawfɪt˺ (elephant)
ʔʌ bejbij əfɪtʰ (a baby elephant)
wʌs ə neːj duwɪŋ (what's the lady ↪ doing)
wijθ dow ɾɑf (wings fell off)
wəhɛ ə fax duwɛ̃n (what are frogs ↪ doing)
tʰɛkʰɪn dowɛ̃n (chicken doing)
wʌ ɛ̩ tʰʌkijn duwɪn (what the ↪ chicken doing)
wɪ ɾʌ bejbij tʰawkʰĩ bawtʰ (what ↪ the baby talkin about)
jɪs maj dæʔ ʃɪftʰ ʔɪs (yes my dad ↪ shaved his)
ɪn ə bɔks (in a books)
wʌs kʰæmʌ dʌʔɪn (what's camel ↪ doing)
jɛs aj dʉuw (yes I do)
wʌ tʰæmʌ dʌəĩ (what camel ↪ doing)
pʰɪw mamɪ sej (what mommy ↪ say)
wə ðə mamij sejɪŋ (what the ↪ mommy saying)
wʊ dædij duwɪŋ (what daddy ↪ doing)
ʔa dʊ dæ tʰuw (I do that too)
æn mij tʰʌ (and me too)
wʌ hɪm duwəɪn (what him doing)
hʌ bejbij tʰajɪŋ (the baby crying)
wʌ ðə mæn duwəjɪn (what the ↪ man doing)
ʔa dʊ dætʰ (I do that)
kʰɛ̃cĩj (sixteen)
ʔowh jɛ mʌ̃kʰĩn dowĩ (what the ↪ monkey doing)
jɛ mij tʰuw aj dow dæʔ tʰuw ↪ (yeah me too I do that too)
ʔa duw dæʔ tʰow (I do that too)
ʔæn nɪçlɪs tʰʌ (and Nicolas too)
ʔaj owpʰẽj maj dʌf (I open my ↪ mouth)
wʌ jʌ pʰejnʔfɪʃ duwəjĩ (what the ↪ peoples doing)
we ja pʰɜʃɪ (what the person)
lɪtʰɵ pʰɛpʰɪʃ dəjɪn (little peoples ↪ doing)
pʰɛtʰ ʔawʊ (pet owl)
ʔa duw dæt tʰuw awʊ (I do that ↪ too Owl)
nʌ fĩn (no thanks)
hʌʔ hɪm duwəjĩn (what him ↪ doing)
maj mʌm ʃow mij (my mom ↪ show me)
ʔɛn kʰeː tʰuw (and Kate too)
kʰɛt tʰowm (Kate too)
ʔɛsajkʰ (outside)
maj dæ duw dæt˺ (my dad do ↪ that)
ʔɛn maj mam duw dɛʔ (and my ↪ mom do that)

bejbij tʰajɛː (baby tired)
də bejbɪθ tʰajə (the baby's tired)
dowĩj t͡ʃʌ (drying himself)
hapʰij tʰə jʌː (happy to you)

—— (continued) ——

7.1.4 Providence/Alex/021122.xml 36 months

ˈwo ˈwʌts ɪs ˈɛː (yyy what's this yyy)
ˈɪs ˈpwis ˈhɛt (yyy yyy yyy)
ˈuː (yyy)
* ˈprɪːɾi (xxx pretty)
ˈprʌi (yyy)
ˈwo ˈaɪ ˈlaɪk ˈðæt (whoa I like that)
ə ˈpɪsələ ˈkʊki ˈtʃwɛ ˈpʌ (yyy yyy yyy yyy yyy)
ə ˈpʌkɪn (a pumpkin)
ˈbuː (yyy)
ˈwu (yyy)
ˈwʌts ˈɪs (what's this)
ˈwʌts ˈɪs (what's this)
ˈwʌts zɪs (what's this)
ˈwʌts ɪs (what's this)
ˈʌi (yyy)
ˈuː (ooh)
ə ˈtʃwɔlo (a yyy)
ˈwoz ɜr də ˈwʌ (those are the yyy)
ˈðoz ə ðə ˈwɑɾə (those are the water)
ðə ˈwɑɾə ˈsli (the water yyy)
əˈtiho * (yyy xxx)
ˈʌm ˈwʌts ɪs (&-um what's this)
ˈgost ˈkukis ˈwʌts ˈɪs (ghost cookies what's this)
ə ˈkukis (a cookies)
ˈʌb (yyy)
ə ˈbʌg (a bug)
ˈwɔɑː ˈʌzə ˈtʃɪkɪn (yyy yyy chicken)
ˈʃɪki ˈaɪ ˈlaɪk ˈdæt ˈtʃɪkɪn (chicken I like that chicken)
ˈu (ooh)
ˈuː (ooh)
ˈuː (ooh)
ˈʌm (&-um)
ˈfwut (fruit)
ˈɑlɪvz (olives)
ˈweɪps (grapes)
ˈbluˈbɛvi (blueberry)
ˈwʌts ˈɪs (what's this)
ˈpʊpə ˈgweɪps (purple grapes)
ˈwɑː ˈpwɛsˆˈdʌ (yyy pretzels)
ˈðɪs (this)
ˈpwɛsə (pretzels)
ˈwaʊ (wow)
ˈtʃɑˆkələt (chocolate)
ˈʃːɑklət (chocolate)
ˈtʃɑklɪt ˈdʌŋk (chocolate yyy)
ˈuː ˈwʌts ɪs (ooh what's this)
ˈwʌts ˈðɪs (what's this)
ˈuː ə ˈbɪg ˈkeɪk (ooh a big cake)
ˈwʌts ˈɪs (what's this)
ˈʌb ˈwʌts ˈɪs (yyy what's this)
ˈdʒɔðəts (yyy)
ˈwʌz ˈɪz ˈðɪs (what is this)
ˈwʌts ˈðɪs (what's this)
ˈʌjʊ ˈiɾ ɪt (yyy eat it)
ˈspweɪkos (sprinkles)
ˈno ðə ˈsteɪzəs (no yyy yyy)
ˈdʒiˈdʒi (Gigi)
ˈaɪː kə ˈdu ə ˈðɪ (I can do yyy it)
oˈkeɪ (okay)
ˈo (oh)
* ˈmɑm (xxx Mom)
ˈjɛ (yeah)
ˈdʒiˈdʒi * (Gigi xxx)
ˈdʒwʌːʒi (yyy)
* ˈdʒiˈdʒi (xxx Gigi)
ˈno ˈmɑmi (no Mommy)
ˈwʌz ˈdɑɾi (where's Daddy)
ə ˈspwikəl ˈdonət (a sprinkle donut)
ˈaɪ ˈlaɪk ə ˈspwiŋkəl ˈdonət (I like a sprinkle donut)
ˈmɑmi (Mommy)
ˈaɪ ˈlaɪk ə ˈspweɪŋkəl ˈdonət (I like a sprinkle donut)
ˈjæ (yeah)
ə ˈdʌn ˈpleɪiŋ (are 0we done playing)
əi ˈdʌn ˈpleɪiŋ (are we done playing)
ˈmɑmi (Mommy)
əˈlɑkətʃʌ * (yyy xxx)
ˈɑl ˈteɪk ju * (I'll take you xxx)
* ˈteɪk * (xxx take xxx)
* ˈteɪk ju (xxx take you)
ˈaɪ ˈlaɪk ə ˈteɪk ju ˈmɑm (I like yyy take you Mom)
ˈaɪ ˈteɪk ju (I take you)
ˈʌ wi ˈɑl ˈdʌn (are we all done)
ˈno ˈno (no no)
ˈno (no)
ˈæpəˈsɔs ˈjɑ (applesauce yyy)
ˈɑt (yyy)
ˈkæ̃ndi (candy)
ˈdʒus (juice)
ˈwʊt (yyy)
ˈpiz (peas)
ˈuː (school)
ˈskuː (school)
əˈweɪŋ (swing)
ˈstɑː (star)
ˈflæg (flag)
ˈstɛz (stairs)
ˈʌvɪn (oven)
ˈbɛntʃ (bench)
ˈbɛɾəm (bedroom)
ˈbɛd (bed)
ˈtaʊː (towel)
ˈtweɪ (tray)
ˈtæʃ (trash)
ˈpleɪt (plate)
ˈpleɪt (plate)
ˈmɑp (mop)
ˈkom (comb)
ˈbwum (broom)
ˈlɛg (leg)
ˈhæ̃nd (hand)
ˈɪː (ear)
ˈtʃɪn (chin)
ˈsɑk (sock)
ˈʃu (shoe)
ˈnɛkləs (necklace)
ˈhæt (hat)
ˈkaɪː (sky)
ˈpɑːɾɾi (party)
ˈno (no)
ˈfwɛnd (friend)
ˈpɜsən (person)
ˈbaɪ (bye)
ˈhaɪ (hi)
ˈno (no)
ˈʃɑpi (shopping)
ˈθeɪg ju (thank you)
ˈkæwi (carry)
ˈtʃeɪs (chase)
ˈdʌmp (dump)
ˈfɪnɪs (finish)
ˈfɪt (fit)
ˈhʌg (hug)
ˈlɪθː (listen)
ˈlaɪk (like)
ˈpwiˈtɛːnd (pretend)
ˈrɪp (rip)
ˈʃeɪk (shake)
ˈteɪst (taste)
ˈdʒɛntə (gentle)
ˈwɪk (think)

ˈwɪʃ (wish)
ˈɪf (if)
ˈwʊd (would)
ˈnid (need)
ˈkʊd (could)
ˈmːʌtʃ (much)
ˈɑː (all)
ˈʌndɝ (under)
ˈdaʊn (down)
ˈbiˈsaɪd (beside)
ˈwɛː (where)
ˈʌs (us)
ˈðɪs (this)
ˈðɛm (them)
ˈaʊː (our)
təˈnaɪt (tonight)
əˈgɛː (yyy)
ˈæftɝ (after)
ˈwɛt (wet)
ˈtɑni (tiny)
ˈlæst (last)
ˈhɑt (hot)
ˈhæpi (happy)
ˈfæt (fast)
ˈkot^{h} (cold)
ˈɔ ˈgɑn (all gone)
ˈʃeɪps (shapes)
ə ˈtʃwaɪeɪgə (a triangle)
ə ˈbɪg ˈtʃwaɪeɪŋgə (a big triangle)
ˈtʃwaɪeɪgəː * (triangle xxx)
ˈtwaɪəgə (triangle)
ˈsʌ ə ˈbɪg * ə ˈbɪg ˈtʃraɪeɪŋgə (yyy a big xxx a big triangle)
ˈuː (ooh)
ə ˈbɪg ˈsɝkəl (a big circle)
ə ˈbɪg ˈtʃraɪ^ˈeɪgo ə ˈbɪg ˈskwɛː (a big triangle a big square)
ˈuː (ooh)
ə ˈbɪg ˈovəl (a big oval)
ˈoː (ooh)

—— **(continued)** ——

7.2 CHILDES

A random sample of the transcripts used in this manuscript at different ages. Each line corresponds to an utterance and each utterance is followed by transcribed part-of-speech tags.

7.2.1 Eng-NA/Braunwald/010511.xml 17 months

night_night (co)
night_night (co)
here (adv)
it is night_night (pro:per 0cop n)
Daddy (n:prop)
spiders (n)
oh (co)
me Dwww (pro:obj n:prop)
on (adv)
on (unk)
no (co)
buttons (unk)
uh ()
down (adv)
water (n)
water (n)
there (adv)
dance there (unk adv)
ahhah (co)
on (adv)
don't (mod~neg)
give (v)
I want (pro:sub v)
Daddy (n:prop)
dance (n)
on (adv)
I want that that that that (0pro:sub v pro:dem)
eh ()
go in there (v prep n)
uhoh (co)
uhoh (co)
uhoh (co)
yeah (co)
thank you (v pro:per)
thank you (v pro:per)
yeah (co)
on (adv)
Cee (n:prop)
spider (n)
Cee (n:prop)
down (adv)
byebye (co)
car (n)
car (n)
there (adv)
byebye (co)
car (n)
car (n)
baby (n)
night_night (co)
night_night (co)
Cee (n:prop)
cookie (unk)
spoon (n)
oh (co)
down (unk)
down (unk)
there (adv)
recorder (n)
aya (bab)
door (n)
key (n)
byebye (co)
car (n)
kitty (n)
outside (adv)
bow (on)
bye (co)
byebye (co)
bow (on)
bow (on)
she lives next door to us (pro:sub v adj n prep pro:obj)
bow (on)
recorder (n)
cookie (n)
no (co)
Deedee (n:prop)
here (adv)
cookie (n)
that that door (det:dem n)
that tata (comp chi)
nose (n)
eye (n)
ear (n)
Laura (n:prop)
toe (n)
tickle (n)
toe (n)
ah (co)
uh ()
toe (n)
recorder (unk)
toe (n)
ah (co)
toe (n)
my toe (det:poss n)
toe (n)
where (pro:rel)
here (adv)
no (co)
there (adv)

—— **(continued)** ——

7.2.2 Brown/Adam/020801.xml 32 months

this is heavy (pro:dem cop adj)
saggy baggy doesn't eat a all up (adj adj mod~neg v det:art qn adv)
oh no (co co)
le me (v pro:obj)
you going faster (pro:per part adj)
washer (n)
going go little (part v adv adj)
what is what de the in (pro:int det:art det:art prep n)
pocket (n)
dis this one (pro:dem pro:dem pro:indef)
booking (chi)
booking booking booking booking (chi chi chi chi)
booking booking (chi chi)
tease book tease (n n n)
tease (n)
tease (n)
tease tease (n n)
teasing teasing teasing (part part part)
teasing (part)
teasing (part)
tease a Cromer (v det:art n:prop)
what this is car (pro:int det:dem aux n)
pin (n)
yeah Mommy pin (co n:prop n)
car (n)
yeah (co)
red car (adj n)
yellow car (n n)
watch (n)
where horses go (pro:int n v)
where horses (pro:int n)
horse go yes Mommy (n v co n:prop)
did he (mod pro:sub)
there he is Mommy (adv pro:sub cop n:prop)
corral corral (n n)
baby horses (n n)
horses (n)
baby horses (n n)
ready me go (v pro:obj v)
ready me (v pro:obj)
go down dere there (v prep n n)
go down right side (v adv adj n)
switch (n)
doing switch (part n)
trick (n)
doin trick (part n)
doing chair tricks (part n n)
yeah funny (co adj)
chair trick laughing (n n part)
chair tricks (n v)
Mommy chair tricks (n:prop v n n)
chair tricks chair tricks chair tricks (n v n v n n)
press a button (v det:art n)
press a button (v det:art n)
yeah (co)
what a happen have a tail (pro:int det:art v v det:art n)
yeah (co)
press a button (v det:art n)
doing rope tricks (part n n)
rope tricks (n v)
watch it rope tricks (v pro:per n n)
yeah (co)
watch it (v pro:per)
car car (n n)
fell down Mommy's floor (v prep adj n)
throw dat that (v pro:dem pro:dem)
what dat that (pro:int adv adv)
tricks (n)
yep tricks (co v)
press a button (v det:art n)
okay de the horses tail (co det:art det:art n n)
okay horses (co n)
okay horses okay horses (co n adj n)
good night rope tricks (adj n n n)
good night my rope tricks (adj n det:poss n n)
yeah rope tricks (co n n)
rope trick fell down (n n v adv)
go tired go tired (v part v part)
Mommy Mommy (n:prop n:prop)
holler doesn't fit in (v mod~neg v prep n)
horse fit in   (n v adv prep adv)
ropes (n)
Mommy roller will stand up (n:prop n mod v adv)
try him dere there (v pro:obj adv adv)
Mommy Mommy (n:prop n:prop)
will fit in (mod n prep n)
see (v)
le me do rope tricks (v pro:obj v n n)
let me do ropes (v pro:obj v n)
hello hello hello (co n n)
what dat that Mommy cowboy (pro:int adv adv n:prop n)
hello cowboy (co n)
wh cowboy (pro:int v n)
wh cowboy (pro:int v n)
happen to him (v prep pro:obj)
wh him (pro:int v pro:obj)
yeah (co)
happen cow watching Rusty down   (v n part n:prop adv prep adv)
see him down there (v pro:obj prep n)

—— (continued) ——

7.2.3 Eng-NA/Carterette/first.xml 72 months

you mean uh um like England or something (pro:per v conj n:prop coord pro:indef)
when we walk home from school I walk home with two friends (conj pro:sub v n prep n pro:sub n n prep det:num n)
and sometimes we can't run home from school though (coord adv pro:sub mod~neg v adv prep n adv)
because um one girl where every time she wants to runs she gets the wheezes and stuff (conj det:num n pro:rel qn n pro:sub v inf v pro:sub v det:art v coord n)
and then she can't breathe very well and she gets sick (coord adv:tem pro:sub mod~neg v adv adv coord pro:sub v adj)
that's why we can't run (pro:dem~cop pro:int pro:sub mod~neg v)
I like to go to my grandmother's house (pro:sub v inf v prep det:poss adj n)

well because she gives us candy
↪ (co conj pro:sub v pro:obj n)
well um we eat there sometimes
↪ (co pro:sub v adv adv)
sometimes we sleep overnight
↪ there (adv pro:sub v adv
↪ adv)
sometime when I go to go to my
↪ cousin's I get to play softball
↪ or play badminton and all
↪ that (adv conj pro:sub v inf v
↪ prep det:poss adj pro:sub v
↪ prep n n coord n n coord qn
↪ pro:dem)
thing I hate to play is doctor (n
↪ pro:sub v prep n cop v)
oh (co)
I hate to play doctor or house or
↪ that (pro:sub v prep n n
↪ coord n coord pro:dem)
don't like it or stuff (mod~neg v
↪ pro:per coord n)
we've been learning a_lot_of
↪ Spanish words (pro:sub~aux
↪ aux part qn n:prop n)
our teacher speaks Spanish
↪ sometimes (det:poss n v
↪ n:prop adv)
so does my father (adv v det:poss
↪ n)
yyy ()
well my father doesn't know very
↪ much Spanish (co det:poss n
↪ mod~neg v adv adv n:prop)
but he doesn't know what gray is
↪ in Spanish (conj pro:sub
↪ mod~neg v pro:int adj aux
↪ prep n:prop)
and its (coord det:poss L2)
and he doesn't and he knows
↪ what blue is in Spanish
↪ (coord pro:sub v pro:int n
↪ cop prep n:prop)
and he knows what um red is
↪ (coord pro:sub v pro:int n
↪ cop)
in Spanish (prep n:prop)
and sometimes I like to go to
↪ Mexico but I've never been
↪ there before (coord adv
↪ pro:sub v inf v prep n:prop
↪ conj pro:sub~aux adv cop
↪ adv adv)
only when I was a little teeny
↪ baby I been there and I don't
↪ even remember it (adv conj
↪ pro:sub cop det:art adj adj n
↪ pro:sub cop adv coord
↪ pro:sub mod~neg adv v
↪ pro:per)
there this one night I couldn't get
↪ any food (adv pro:dem
↪ pro:indef n pro:sub
↪ mod~neg v qn n)
I mean there was this one day I
↪ couldn't get any food at
↪ home unless I asked it for
↪ Spanish (pro:sub v adv cop
↪ det:dem det:num n pro:sub
↪ mod~neg v qn n prep adv
↪ conj pro:sub v pro:per prep
↪ n:prop)
my um my mother and father is
↪ going to pretty soon take us
↪ to Philadelphia (det:poss
↪ det:poss n coord n aux part
↪ inf adj adv v pro:obj prep
↪ n:prop)
and we're going to see our
↪ grandmother there (coord
↪ pro:sub~aux part inf v
↪ det:poss n adv)
I wish we went to (pro:sub v
↪ pro:sub v prep)
uh we went to Mexico not
↪ Mexico San_Diego once
↪ (pro:sub v prep n:prop adv)
and they had a little um pool that
↪ was full of water and it was
↪ two feet (coord pro:sub v
↪ det:art adj n pro:rel cop adj
↪ prep n coord pro:per cop
↪ det:num n)
and then they and then they had
↪ another pool (coord adv:tem
↪ pro:sub v qn n)
it was five feet eight feet (pro:per
↪ cop det:num n det:num n)
Randy my brother went in eight
↪ feet and I went in five feet
↪ (n:prop det:poss n v prep
↪ det:num n coord pro:sub v
↪ prep det:num n)
and I think there was a three feet
↪ (coord pro:sub v adv cop
↪ det:art det:num n)
there was (pro:exist cop)
and I jumped off and I uh and I
↪ jumped off the edge of the
↪ swimming pool (coord
↪ pro:sub v prep det:art n prep
↪ det:art n:gerund n)
I got on the edge and I jumped
↪ off (pro:sub v prep det:art n
↪ coord pro:sub v adv)
and then I holded held on to a
↪ edge because I couldn't swim
↪ very well (coord adv:tem
↪ pro:sub v v adv prep det:art
↪ n conj pro:sub mod~neg v
↪ adv adv)
when I start when I started to
↪ swim I was always holding
↪ on to the edge (conj pro:sub
↪ v inf v pro:sub aux adv part
↪ adv prep det:art n)
I wouldn't dare to go more than
↪ this away from the edge or
↪ else I I'd I'd start jumping
↪ dancing into the water
↪ (pro:sub mod~neg v inf v qn
↪ prep pro:dem adv prep
↪ det:art n coord post
↪ pro:sub~mod v part part
↪ prep det:art n)
when my father wanted to take a
↪ picture of me with you_know
↪ one of those floating things
↪ one of those floating rings
↪ that you put around you but
↪ I don't wanna because
↪ you_know I know how to
↪ swim (conj det:poss n v inf v
↪ det:art n prep pro:obj prep
↪ co det:num prep det:dem
↪ part n pro:indef prep
↪ det:dem part n pro:rel
↪ pro:per v prep pro:per conj
↪ pro:sub mod~neg v~inf
↪ conj co pro:sub v pro:int inf
↪ v)
but when I took it off I almost
↪ drownded drowned (conj
↪ conj pro:sub v pro:per adv
↪ pro:sub adv part part)
and I was jumping up and down
↪ to see if I could swim or not
↪ (coord pro:sub aux part adv
↪ coord adv inf v conj pro:sub
↪ mod v coord neg)
and (coord)

um I live in an apartment and we have a big pool and it's eight and a half in part and four and a half and three and a half (pro:sub v prep det:art n coord pro:sub v det:art adj n coord pro:per~cop det:num coord det:art n prep n coord det:num coord det:art n coord det:num coord det:art n)
and this summer I get to go swimming in it (coord det:dem n pro:sub v inf v part prep pro:per)
in the summer we go swimming (prep det:art n pro:sub v part)
and that's when my birthday is (coord pro:dem~cop conj det:poss n cop)
we don't go in spring or winter because it's too cold (pro:sub mod~neg v prep n coord n conj pro:per~cop adv adv)
my my brother can go swimming in the winter though because he gots got his tonsils out you_know (det:poss n mod v part prep det:art n adv conj pro:sub v v det:poss n adv co)
and he and he gets sick uh sick um once in a few years (coord pro:sub v adj adv prep det:art qn n)
I get sick just about every day (pro:sub v adj adv prep qn n)
there's just one thing I can't stand in my family (pro:exist~cop adj det:num n pro:sub mod~neg v prep det:poss n)
my baby makes too much noise (det:poss n v adv qn n)
I can't even get get to sleep for a minute (pro:sub mod~neg adv v prep n prep det:art n)
he won't stop jumping around in the bath (pro:sub mod~neg v part adv prep det:art n)
in the bath (prep det:art n)
no (co)
in the crib (prep det:art n)
he he keeps jumping around gets tired (pro:sub v part adv v part)
then he goes to bed then he finally gets to sleep (adv:tem pro:sub v prep n adv:tem pro:sub adv v prep n)
can't go to sleep in about a hour (mod~neg v inf v adv prep det:art n)
not with that in the house (neg prep pro:dem prep det:art n)
it would just take two minutes to get to sleep (pro:per mod adv v det:num n inf v prep n)
just about two minutes (adv prep det:num n)
if you just um why don't you get some cotton and plug it in your ears and then you can't hear him (pro:int mod~neg pro:per v qn n coord v pro:per prep det:poss n coord adv:tem pro:per mod~neg v pro:obj)
he makes so much noise he makes so much noise it probably sound effect through it (pro:sub v adv qn n pro:per adv adj n prep pro:per)
well what does the baby do (co pro:int v det:art n v)
come out get out crawl out_of his crib and then come along in your bed and pull out your ear (v adv v adv n prep det:poss n coord adv:tem v adv prep det:poss n coord v adv det:poss n)
once once he keep jump jumping jumping and then this thing slide down (adv pro:sub v part coord adv:tem det:dem n n adv)
and then he fell over to the other bed and he start crying (coord adv:tem pro:sub v adv prep det:art qn n coord pro:sub v part)
and I couldn't get to bed so I I hafta wake up put him back in my crib (coord pro:sub mod~neg v prep n conj pro:sub mod~inf v adv v pro:obj adv prep det:poss n)
in your crib (prep det:poss n)
no not in my crib (co neg prep det:poss n)
I don't have a crib (pro:sub mod~neg v det:art n)
you said put him back in your crib (pro:per v v pro:obj adv prep det:poss n)
I mean in his crib (pro:sub v prep det:poss n)
I don't have a crib (pro:sub mod~neg v det:art n)
uh sometimes I like to go to the I like to go to my grandmothers (adv pro:sub v inf v prep det:poss n)
I would like to sleep over her at her house every day because she lets me stay up late about ten o'clock or twelve thirty (pro:sub mod v inf v adv prep det:poss n qn n conj pro:sub v pro:obj v adv adv prep det:num n coord det:num det:num)
you're lucky (pro:per~cop adj)
I only get to stay up until eight (pro:sub adv v inf v adv prep det:num)
and I only get to stay up until nine (coord pro:sub adv v inf v adv prep det:num)
I get to stay up until um say about between ten o'clock and nine thirty (pro:sub v inf v adv prep v adv prep det:num n coord det:num det:num)
uh and sometimes sometimes I get to go to bed at twelve thirty (coord adv pro:sub v inf v prep n prep det:num det:num)
sometimes but most of the times I don't (adv conj qn prep det:art n pro:sub mod~neg)
on holidays and you_know like um weekends (prep n coord co prep n)
on holitinna holidays and I mean on holidays I get to stay up all night (prep n n coord pro:sub v prep n pro:sub v inf v adv qn n)
uh on weekends like when I'm not going to school (prep n prep conj pro:sub~aux neg part prep n)
see this day I'm going to school and then the next day you don't hafta (v det:dem n pro:sub~aux part prep n coord adv:tem det:art adj n pro:per mod~neg mod~inf)

I can stay up late because I the next day I can sleep all I want (pro:sub mod v adv adv conj pro:sub det:art adj n pro:sub mod n adv pro:sub v)
that's why we hafta go to bed early on school days (pro:dem~cop pro:int pro:sub mod~inf v prep n adv prep n n)
every holiday um um my my grandmother and my aunt come over (qn n det:poss n coord det:poss n v adv)
well you_know it's because well you_know it's just about becoming Easter (co co pro:per~cop conj adv co pro:per~aux adj adv part n:prop)
about just twenty days or twenty one (adv adv det:num n coord det:num det:num)
on Easter I hafta get all this gooshy egg (prep n:prop pro:sub mod~inf v qn det:dem adj n)

— (continued) —

7.3 *Drosophila*

One hour of behavioral state transitions from a single example *Drosophila*. There are 117 unique behavior states. Behavioral states do not have names but belong to broad categories (Posterior, Side Legs, Anterior, Locomotion, Idle, Slow).

59 43 11 21 11 51 52 46 52
60 59 65 46 27 32 33 40 52
43 39 43 76 106 76 52 43 9
4 9 21 9 21 11 21 69 59
46 42 52 43 9 21 4 9 10
52 46 80 69 80 84 103 60 43
9 21 4 21 52 69 66 46 52
43 21 43 52 53 60 59 68 46
52 40 52 39 43 21 10 21 43
52 43 52 76 52 31 9 10 9
10 9 4 43 52 48 59 32 65
38 45 52 45 33 46 33 40 52
39 4 43 52 65 53 60 52 43
4 9 4 10 21 51 43 52 53
65 46 55 52 43 21 9 10 21
4 43 40 32 33 49 46 15 33
39 51 4 9 43 52 53 59 65
59 65 45 52 43 52 60 62 65
62 60 52 48 21 9 51 43 52
53 50 46 68 59 50 46 27 69
80 65 68 59 49 57 66 59 65
49 44 41 44 46 48 53 59 66
65 66 59 67 77 60 43 52 59
65 59 69 77 53 55 59 64 54
65 44 46 65 50 65 49 32 59
50 44 49 47 50 65 69 53 52
43 51 21 51 57 39 43 52 65
52 45 65 66 43 53 65 80 53
43 21 39 71 52 43 52 55 66
46 55 53 52 43 52 43 52 60
77 60 67 71 84 106 98 87 84
93 108 93 67 87 67 60 52 53
59 65 59 48 52 39 21 9 11
21 11 31 52 45 65 59 52 43
52 53 59 69 27 46 27 15 32
34 39 43 52 43 52 60 53 59
46 66 27 47 49 35 47 49 1
38 14 38 50 19 25 49 7 38
46 15 22 32 38 44 46 15 38
35 38 32 44 65 49 44 46 47
69 59 52 43 39 21 10 4 9
11 4 9 4 10 4 39 40 33
19 27 46 27 32 33 45 40 33
46 33 65 71 79 71 87 84 69
79 46 54 32 22 46 15 27 44
27 35 49 20 19 46 27 15 29
14 20 28 35 15 44 28 50 47
49 57 41 37 52 51 61 49 65
43 51 21 39 52 66 68 65 49
46 19 40 31 21 10 21 4 21
39 20 28 20 32 33 22 35 28
46 19 38 36 46 65 66 65 68
45 49 47 49 44 50 46 68 69
87 77 87 84 87 77 87 79 46
27 20 30 38 46 49 65 49 41
32 45 65 56 49 65 49 57 44
46 27 23 34 31 39 21 39 19
38 19 40 34 33 32 15 35 38
36 46 44 66 35 49 28 15 47
15 14 27 46 49 14 1 2 14
19 15 14 38 15 13 19 38 46
20 15 38 20 38 65 49 27 46
32 33 21 10 9 21 9 21 9
11 9 10 9 11 9 21 43 52
34 32 49 46 27 32 23 33 40
39 21 9 21 9 21 43 52 53
68 49 46 27 32 39 43 21 43
52 48 40 44 49 44 32 46 45
65 59 80 46 33 32 52 49 52
45 65 52 45 49 32 46 38 46
29 38 20 28 35 27 35 27 20
38 15 46 15 32 44 27 19 46
49 47 49 35 49 47 49 44 32
49 44 35 49 44 38 5 6 14
35 22 14 20 28 35 49 35 19
35 49 44 49 20 49 1 15 14
38 28 14 38 25 20 25 49 25
35 27 44 27 25 20 46 49 35
27 49 47 49 35 49 57 65 44
56 46 35 47 65 50 59 41 49
44 22 29 25 14 27 14 27 1
2 1 2 1 15 20 38 27 46
19 27 35 38 46 49 25 49 28
14 38 20 6 38 46 15 35 49
44 15 7 15 38 14 8 7 38
46 25 38 25 38 28 14 19 25
15 14 38 27 14 1 2 15 38
14 38 14 19 14 19 38 19 27
38 19 49 46 49 65 49 65 69
44 46 20 38 15 33 45 55 59
41 36 79 38 46 20 14 15 32
13 15 38 29 84 46 90 105 84
115 87 55 59 75 98 103 93 75
90 46 99 87 107 115 65 59 32
46 20 38 15 13 23 33 34 40
39 31 52 48 59 65 59 46 44
109 105 93 76 87 103 93 84 65
98 59 45 53 65 46 45 33 52
3 10 9 11 21 11 9 11 9
11 9 3 11 9 3 10 4 9
21 4 10 21 9 21 9 10 9

— (continued) —

7.4 Zebrafish

Behavioral states for zebrafish. Several behavioral contexts are used in this dataset. The example behavioral sequence shown below is acquired during a phototaxis paradigm (SCS: Short Capture Swims; LCS: Long Capture Swims; BS: Burst type forward Swim with high tail-beat frequency; SLC: Fast C-start escape Swims;

RT: Routine Turns; LLC: Long Latency C-starts; AS: Approach Swims; SAT: Spot Avoidance Turn; HAT: High Angle Turn).

SAT RT S2 RT S1 S1 RT RT
HAT S1 RT RT RT RT RT RT
RT RT S1 S1 HAT RT SAT S2
S2 RT RT RT S2 RT S2 S2
HAT RT SAT RT S1 RT S1 S2
HAT S1 HAT S1 S1 S1 RT S1
HAT RT HAT HAT S2 RT HAT S2
S2 RT RT S1 S2 RT S2 RT
S1 RT SAT S2 SAT RT RT S2
S2 O-bend S1 S2 RT S2 RT S2
RT S2 S2 RT S2 S2 S2 RT
S2 S2 S2 S1 S1 RT RT HAT
RT S2 S1 S2 S2 S2 RT RT
S2 S1 RT RT S2 S2 S2 S2
RT S2 RT RT S2 RT RT S2
RT RT S2 S2 S2 S2 S2 S2
RT S2 HAT HAT RT S1 S2 RT
SAT S2 S2 S2 S2 RT S1 RT
S1 RT S1 S2 S2 S2 S1 S2
S2 S2 J-turn HAT S2 RT S2 S1
S2 RT RT S2 RT RT HAT S2
O-bend HAT S1 S2 S2 S2 S2 S2
S2 S2 S2 RT RT S2 RT HAT
S2 S1 S1 RT RT RT RT RT
RT HAT RT S2 RT RT HAT S1
S1 S1 RT S2 S2 RT S2 SAT
S2 S2 S1 S2 J-turn RT RT HAT
RT S2 S2 S2 HAT RT S2 S2
S2 S2 S2 HAT S1 RT HAT S1
S1 S2 AS HAT S1 S2 S1 RT
HAT RT S1 S1 RT S1 S2 S2
RT RT S2 S1 S2 S2 S1 J-turn
S2 S2 RT RT S1 S1 S2 RT
S2 S1 HAT S1 AS RT RT RT
S2 S2 HAT AS RT S2 RT S1
RT S2 RT S2 RT RT RT S1
S1 S1 S2 HAT S1 AS RT HAT
RT RT S2 S2 S2 S2 RT S2
RT HAT S2 RT S2 RT S2 S2
RT HAT S1 S1 S2 RT RT RT
HAT S1 HAT S2 S2 RT J-turn S2
S2 S2 RT S1 S2 S2 RT RT
HAT S1 S2 RT RT HAT HAT S1
S2 S2 S2 S2 S2 S2 S2 S2
RT S1 S1 S1 HAT HAT S2 HAT
S2 HAT S2 S2 S2 S2 S2 RT
HAT S1 S1 S2 S2 HAT S1 RT
SCS J-turn S2 HAT S1 S2 S2 S2
S2 RT S1 RT S1 AS J-turn RT
RT RT RT O-bend J-turn S1 RT
↪ RT
RT S2 S2 RT S2 RT O-bend S2
S2 S2 S2 S2 J-turn RT RT S2
S2 HAT S1 J-turn RT S2 S2 S2
S1 S2 S2 RT S2 S2 S2 RT
RT S1 S2 S2 S1 S2 HAT S1
RT S2 S2 S2 RT RT HAT S1
SAT HAT HAT S2 S2 HAT HAT
↪ S1
S2 S2 S2 S2 S1 S2 S1 S1
S2 S1 S1 RT S2 S2 RT RT
S1 S2 HAT S1 O-bend RT S1 S2
RT RT RT S1 S1 HAT SAT S1
S2 S2 S2 S2 S2 S2 S2 S2
S2 S2 RT HAT S1 S2 S1 RT
S1 S2 S2 S2 S2 RT S2 RT
RT HAT S1 RT RT S2 HAT S1
RT RT RT J-turn AS S2 S1 RT
S2 RT RT S1 S1 S1 S2 RT
HAT RT RT HAT S1 S1 S1 RT
S2 S2 HAT RT RT S1 HAT RT
S2 RT S2 S2 S2 S2 S2 SAT
S2 S2 S2 S2 RT S2 S2 RT
S2 S2 RT S2 S2 RT HAT S1
J-turn S2 RT S2 HAT S1 S2
↪ J-turn
RT S1 RT S2 J-turn HAT RT S2
RT SAT S2 RT HAT HAT S2 S2
S2 HAT S1 S1 S2 S2 RT RT
S2 HAT S1 HAT J-turn S1 RT S2
S2 HAT S2 RT J-turn J-turn SCS
↪ S2
J-turn J-turn S1 SAT S2 RT RT S2
S2 J-turn RT S2 RT S2 HAT HAT
S2 S2 S2 S2 SAT S1 S1 S2
S2 RT SAT S1 RT RT S1 S2
S1 S2 S1 S1 S1 S1 S1 S2
S1 RT S2 S2 RT RT S2 S2
S1 S2 S2 S2 S2 S2 S2 S2
S2 S2 RT S2 S2 RT RT RT
S1 RT RT S2 S2 HAT RT HAT
S1 S2 S2 S2 S2 S2 S2 S2
S2 S2 S2 RT RT S2 RT HAT
S1 RT S1 S2 RT S2 S1 RT
S2 S2 S2 S2 RT S2 S2 S2
RT RT S2 S2 HAT RT S1 HAT
SAT RT RT S2 S1 S1 S2 S2
S2 J-turn S1 HAT HAT S1 RT
↪ HAT
S2 RT S2 J-turn AS S1 S2 S1
S2 S2 S1 RT HAT S2 S2 S2
S2 HAT S1 S1 RT RT S2 RT
S1 RT J-turn HAT S1 S1 RT S2
S2 S2 S2 S2 S2 S2 S2 S1
S1 HAT HAT S2 S1 S1 S1 S1
HAT RT S1 RT S1 S1 S2 S2

—— **(continued)** ——

7.5 Epic Kitchens

Each transcript in Epic Kitchens contains a sequence of behaviors consisting of an action and object. One example sequence is shown below.

open door
turn-on light
close door
open fridge
take celery
take container
take tofu
close fridge
open fridge
take carrot
open drawer
close fridge
put-down vegetable
open cupboard
take board:cutting
put-down board:cutting
close cupboard
open drawer
take knife
take knife
put-down knife
close drawer
put-down knife
open tap
wash courgette
wash courgette
wash carrot
wash carrot
close tap
put-down vegetable
open cupboard
take grater
take pan
put-down pan
close cupboard
close cupboard

take courgette
cut courgette
turn-on hob
cut courgette
cut courgette
dice courgette
dice courgette
dice courgette
dice courgette
pour courgette
throw courgette
open drawer
close drawer
take spatula
stir courgette
take salt
open salt
pour salt
put-down salt
stir courgette
put-down spatula
take celery
wash celery
open tap
wash celery
close tap
put-down celery
cut celery
cut celery
pour celery
put-down board:cutting
take celery
throw celery
open fridge
put celery
close fridge
take spatula
stir spatula
put-down spatula
open container
take onion
take onion
put-down onion
close container
take spatula
take knife
cut onion
cut onion
cut onion
put-down knife
take kettle
open tap
pour water
pour water
close tap
turn kettle
take spatula
stir vegetable
stir vegetable
take glass
take glass
open cupboard
put glass
close cupboard

—— **(continued)** ——