

# Language-like efficiency and structure in house finch song

Mason Youngblood<sup>1</sup>

Communication needs to be complex enough to be functional while minimizing learning and production costs. Recent work suggests that the vocalizations and gestures of some songbirds, cetaceans, and great apes may conform to linguistic laws that reflect this trade-off between efficiency and complexity. Studies of non-human communication face an additional challenge, though, in that clustering signals into types cannot be done a priori, and decisions about the appropriate grain of analysis may affect statistical signals in the data. The aim of this study was to assess the evidence for language-like efficiency and structure in house finch (*Haemorrhous mexicanus*) song across three levels of granularity in syllable clustering. The results show strong evidence for Zipf's rank-frequency law, Zipf's law of abbreviation, and Menzerath's law. Additional analyses show that house finch songs have small-world structure, thought to reflect systematic structure in syntax, and the mutual information decay of sequences is consistent with a combination of Markovian and hierarchical processes. These statistical patterns are robust across three grains of analysis, pointing to a limited form of scale invariance. In sum, it appears that house finch song has been shaped by pressure for efficiency, possibly to offset the costs of female preferences for complexity.

## 1 Introduction

### 1.1 Efficiency & Complexity

Vocal communication is costly. Individuals learn and produce signals in noisy environments, all while avoiding predation from species adapted to detect them. As a result, communication systems tend to be optimized for efficiency, or the benefit that they bestow relative to the costs of learning and producing them:

$$\text{efficiency} \propto \frac{\sum \pi}{C_L + \sum C_P} \quad (1.1)$$

where  $\sum \pi$  is the lifetime benefit of the communication system,  $C_L$  is the cost of learning it, and  $\sum C_P$  is the lifetime cost of producing it (Gruber et al., 2022). Note that efficiency here is optimized over the course of cultural and biological evolution, as  $\sum \pi$ ,  $C_L$ , and  $\sum C_P$  are incomputable by the individual. This framing is consistent with understandings of efficiency in linguistics (Gibson et al., 2019): the famous “principle of least effort” can be thought of as a minimization of these costs (Zipf, 1949).

On the other hand, simpler sounds that are easier to learn and produce vary across fewer dimensions and are less distinctive from one another (Miton & Morin, 2021). Signals need to be distinguishable to be functional (De Boer, 2000), and complexity expands the possibilities within a signal space in a way that enhances functionality (e.g. to communicate concepts in language, or attract mates in birdsong) (Fitch, 2000). But there is, of course, a limit. Simulations show that cultural evolution plateaus when complex behaviors are too costly (Mesoudi, 2011). Evidence from linguistics, animal behavior, and cultural evolution suggest that communication systems generally evolve to balance this complexity-efficiency trade-off (Gibson et al., 2019; Gruber et al., 2022; Hahn et al., 2020; Jaeger & Tily, 2011; Youngblood et al., 2023).

Notions of complexity vary widely and are sometimes uncorrelated across levels of analysis (Benedict & Najar, 2019; Mikula et al., 2018). In the birdsong literature, for example, complexity is usually described at one of three levels: syllables

---

<sup>1</sup> Institute for Advanced Computational Science, Stony Brook University, [masonyoungblood@gmail.com](mailto:masonyoungblood@gmail.com)

(individual sounds within songs), songs (sequences of syllables), or repertoires (full set of syllables or songs that a bird produces). Syllable and repertoire complexity are approximated with measures of production cost (e.g. frequency bandwidth, number of transitions in pitch (Benedict & Najar, 2019; Youngblood & Lahti, 2022)) and learning cost (e.g. diversity of unique syllables or songs (Garamszegi et al., 2005; Soma & Garamszegi, 2011)), respectively. Song complexity, on the other hand, is often characterized by measures of hierarchical or combinatorial structure (Kershenbaum et al., 2014; Sainburg et al., 2019). Structured signals are more compressible and learnable (Gibson et al., 2019; Raviv et al., 2021; Verhoef, 2012), which allows more information to “pass through the bottleneck” of cultural transmission (Gruber et al., 2022; Kirby, 2017). The former increases both production and learning costs, whereas the latter decreases learning costs. Both notions of complexity, however, should boost the benefits of communication by expanding the signal space. Syllable complexity allows for greater diversity in syllable types, and song complexity allows those syllable types to be combined into a wider array of song types.

The effect of this complexity-efficiency trade-off has been confirmed by experiments showing that artificial languages gain efficiency (Fay & Ellison, 2013; Fedzechkina et al., 2012; Roberts et al., 2015) and structure (Kirby et al., 2008; Kirby et al., 2015; Nölle et al., 2018; Raviv et al., 2019) as they are culturally transmitted and used by participants. Efficiency gains can occur from language use alone (Motamedi et al., 2019), but the emergence of structure seems to require both use and cultural transmission (Carr et al., 2017; Kirby et al., 2015). Most experimental work has focused on compositional structure, or the lexical relationships that determine meaning. Birdsong has no mapping between signals and meanings and thus no compositional structure (Miyagawa et al., 2013), but iterated learning may still lead to combinatorial and hierarchical structure (Kirby et al., 2014). Studies of continuous whistled communication systems that resemble birdsong have also detected increases in hierarchical and combinatorial structure (Tamariz & Kirby, 2016; Verhoef, 2012; Verhoef et al., 2014, 2016).

Birdsong experiments yield similar results. Zebra finches raised in isolation sing atypical songs, but over several generations they converge towards wild-type song by shortening the longest syllables and gaining spectral structure (Fehér et al., 2009). In a replication of Fehér et al. (2009), isolate song also increased in efficiency: song bouts and motifs became shorter and denser, and syllables became shorter and lower in amplitude (Diez & MacDougall-Shackleton, 2020). Zebra finches also appear to learn songs more accurately when they have more acoustic variation and less extreme features, possibly as a way to balance the diversity of sequences and efficiency of syllables (Tchernichovski et al., 2021). Even foraging behaviors that are culturally transmitted become more efficient over time, a tendency that is accelerated by population turnover because new group members are more likely to adopt the more efficient behavior (Chimento et al., 2021).

## 1.2 Linguistic Laws & Structure

Pressure for compression and efficiency lead to regularities in organization that are so universal in human language that they are referred to as linguistic laws. The three most commonly-studied of these are Zipf’s rank-frequency law, Zipf’s law of abbreviation, and Menzerath’s law (Sample et al., 2022).

Zipf’s rank-frequency law predicts that the frequency of an item will be proportional to the inverse of its rank, a relationship that holds for most, if not all, of the world’s languages (Piantadosi, 2014). In this study, I will focus on Mandelbrot’s more flexible parameterization of Zipf’s rank-frequency law (see Analysis) (Mandelbrot, 1953, 1962), which is its most common form in contemporary linguistics (Piantadosi, 2014). Non-human animal communication systems, including bird and cetacean vocalizations (Allen et al., 2019; Briefer et al., 2010; Cody et al., 2016; Hailman et al., 1985; McCowan et al., 1999) and lizard courtship displays (Martins, 1994), exhibit more redundancy than languages, leading to a convex rank-frequency relationship that is better captured by the Zipf-Mandelbrot distribution. Zipf and Mandelbrot both interpreted this rank-frequency law as resulting from a minimization of production and perception costs (Mandelbrot, 1953; Zipf, 1949), and there are models showing that it can be derived from communicative efficiency (Ferrer-i-Cancho, 2016; Manin, 2009; Salge et al., 2015). However, some of these models assume that signals map to objects or concepts which is not the case in birdsong, and other causes are still debated (Piantadosi, 2014) (Allegrini et al., 2004; Corral & Serra, 2020; Simon, 1955). Even though there is still uncertainty about its causes, the presence of Zipf’s rank-frequency law in non-human communication systems has been interpreted as evidence for both communicative efficiency and information content (Genty & Byrne, 2010; Kershenbaum et al., 2021; McCowan et al., 1999; Stepanov et al., 2023).

Zipf’s law of abbreviation predicts that common items will tend to be shorter than rare items because their production cost is lower (Zipf, 1949). This negative correlation between frequency and duration is widespread in both written (Bentz & Ferrer-i-Cancho, 2016) and spoken language (Petrini et al., 2023), and has also been observed in writing systems (Koshevoy et al., 2023) and non-human communication like chimpanzee gestures (Heesen et al., 2019) and bird and primate vocalizations (Favaro et al., 2020; Huang et al., 2020; Semple et al., 2010). In some cases, like hyrax vocalizations, Zipf’s law of abbreviation is absent for duration but is present for other measures of production cost such as amplitude (Demartsev et al., 2019). The explanation for Zipf’s law of abbreviation is simple: when common items have a lower production cost than rare items then the overall production cost of signals goes down (Ferrer-i-Cancho et al., 2013). Shorter signals carry other benefits as well, such as reduced predation risk and reverberation in the environment (Ferrer-i-Cancho et al., 2013).

Menzerath’s law predicts that longer sequences will be comprised of shorter items to balance production costs (Menzerath, 1954). This negative correlation between sequence length and item length is found at various levels of analysis in language (e.g. clauses in sentences, morphemes in words) (Cramer, 2005; Eroglu, 2013; Stave et al., 2021) as well as in music (Boroda & Altmann, 1991). In non-human communication, Menzerath’s law appears to be present in chimpanzee gesture (Heesen et al., 2019) and in some primate and bird vocalizations (Favaro et al., 2020; Fedurek et al., 2017; Gustison et al., 2016; Huang et al., 2020; James et al., 2021), although its detection may depend on the type of vocalization (Clink & Lau, 2020) and level of analysis (Clink et al., 2020). The explanation for Menzerath’s law is a simple extension of Zipf’s law of abbreviation: when production costs are increased in one domain (e.g. song sequence length) they should be decreased in another (e.g. syllable duration).

Beyond linguistic laws, there are two proxy measure of linguistic structure that have recently been investigated in non-human communication: small-world structure (Watts & Strogatz, 1998) and mutual information decay (Sainburg et al., 2019).

Small-world networks are highly clustered and have short average path lengths, so that it only takes a few steps to jump between any pair of nodes (think “six degrees of separation”) (Humphries & Gurney, 2008; Watts & Strogatz, 1998). These sorts of networks are quite common in biological and social systems (Humphries & Gurney, 2008; Watts & Strogatz, 1998), including language. For example, networks of neighboring words in sentences and co-occurring words in thesauruses exhibit small-world structure (Cancho & Solé, 2001; Motter et al., 2002; Steyvers & Tenenbaum, 2005). One explanation for this is that short distances between concepts make language easier to process (Beckage et al., 2011). More broadly, though, small-worldness is thought to reflect general systematic structure and recurrence, which in turn improve the compressibility and learnability of information (Allen et al., 2019; Kirby et al., 2015; Raviv et al., 2021). It is also hypothesized to reflect the emergence of syntactic structure over time (Solé et al., 2010), as both nightingales and children in more advanced stages of vocal development have greater small-world structure in their transition networks (Beckage et al., 2011; Weiss et al., 2014). Humpback whales (Allen et al., 2019) and several songbird species (Beckage et al., 2011; Cody et al., 2016; Hedley, 2016; Sasahara et al., 2012; Weiss et al., 2014) all exhibit small-worldness in their song syntax to a similar degree:  $SWI \sim 1.69 - 4.7$ . The small-worldness index ( $SWI$ ) is based on the level of clustering and the distances between nodes in the network of signal transitions, where values  $> 1$  are consistent with small-world structure (Humphries & Gurney, 2008) (see Analysis).

Mutual information is a measure of dependency, or the amount of information that the presence of one thing has about another. In the context of language, past words provide information that can help predict future words above chance levels (i.e. “do you need anything from the”  $\rightarrow$  “store”). Many large language models, for example, are optimized for precisely this task (Schrimpf et al., 2021; Shlegeris et al., 2022). Intuitively, a word contains more information about the very next word than the one that comes after it (Pothos & Juola, 2007), but correlations can be detected even at very long distances of hundreds of words (Altmann et al., 2012; Alvarez-Lacalle et al., 2006). The rate at which the mutual information of words decreases with increasing distance can provide clues about underlying syntactic structure. Simple models of grammar that assume Markov processes (i.e. next word depends only on current word) lead to exponential decay in mutual information with distance, whereas hierarchical processes (i.e. word order comes from syntactic rules, common view going back to Chomsky (1957)) lead to power-law decay (Alvarez-Lacalle et al., 2006; Li, 1990; Lin & Tegmark, 2017). Some past studies have shown that mutual information decay in real language often follows a power-law (Ebeling & Pöschel, 1994; Li, 1990; Lin & Tegmark, 2017), but other work suggests a more complicated picture somewhere between these accounts (Frank et al., 2012; Melnyk et al., 2005). In German, Italian, Japanese, and English, mutual information decay is

exponential at short distances and fits a power-law at long distances, suggesting that sequences are generated by a combination of Markovian and hierarchical processes (Sainburg et al., 2019; Sainburg et al., 2022). This pattern has also been documented in four bird species, suggesting that sequential organization of song is more complex than previously thought (Sainburg et al., 2019).

Table 1.1: The linguistic laws and structural properties explored in this study.

Law/Property	Level	Description	Explanation
Zipf's rank-frequency law	Syllables	Frequency follows a power-law with rank	Unclear
Zipf's law of abbreviation	Syllables	Common items are shorter	Production cost
Menzerath's law	Both	Longer sequences are comprised of shorter items	Production cost
Small-worldness index	Songs	Sequences have general systematic structure	Learning cost
Mutual information decay	Songs	Sequences have Markovian or hierarchical structure	Learning cost

### 1.3 Granularity

In languages, the boundaries between signals at different levels of hierarchy (e.g. phonemes, words) are apparent to the humans who use them. In non-human communication systems, categorizing signals into types is its own challenge. For relatively small datasets it is possible to manually inspect recorded signals and assign types. Increasingly, though, researchers are turning to automated methods such as hierarchical clustering (Ju et al., 2019; Youngblood & Lahti, 2022), dimension reduction (Sainburg et al., 2020), and machine learning (Merino Recalde, 2023; Rivera et al., 2023) that classify signals into types based on their acoustic features. These methods reduce subjectivity in classification and enable people to work with much more data, but they also require tuning. For example, hierarchical clustering, which is my preferred classification method, requires the user to choose an appropriate threshold below which the “branches” of the “tree” are combined into types (see Figure 2.2). This threshold effectively controls the granularity of clustering: higher values over-lump signals into fewer categories, while lower values over-split signals into more categories. The granularity of an analysis may influence the kinds of patterns that can be detected. Philosophers of cultural evolution call this the “grain problem”—some statistical patterns may be more apparent at certain levels of analysis (Charbonneau & Bourrat, 2021). In Gibbon vocalizations, for example, evidence for language-like compression is present for notes but not for phrases (Clink et al., 2020). Some have even proposed that adherence to Zipf's laws could be used to assess whether classification algorithms have identified relevant units in animal communication systems (Kershenbaum et al., 2016).

### 1.4 Aim & Model

The aim of this study is to assess the evidence for language-like efficiency and structure in house finch (*Haemorrhous mexicanus*) song across three levels of granularity in syllable clustering. By doing so, I hope to (1) identify which features of birdsong may be most subject to the complexity-efficiency trade-off, and (2) determine how clustering decisions affect the manifestation of linguistic laws in non-human communication systems. The data for this study come from a large corpus of house finch songs collected between 1975 and 2019 (Ju et al., 2019; Mundinger, 1975; Youngblood & Lahti, 2022). House finch song is an excellent model for these questions for several reasons. First, house finch song is socially learned (Mann et al., 2020) and culturally evolves (Mundinger, 1975), and thus should be subject to information compression. Second, male house finches are more likely to learn complex syllables (Youngblood & Lahti, 2022)—a content bias that may be an adaptation to female preferences for complexity. In house finches, males that sing longer songs at a faster rate are more attractive (Nolan & Hill, 2004) and have higher reproductive performance (Mennill et al., 2006), and courtship songs are longer and contain more syllable types (Ciaburri & Williams, 2019). Because these measures of complexity relate to production and learning costs, they may increase pressure for efficiency in other domains such as duration. Finally, house finch song is known to be subject to efficiency constraints. When house finches tutored by canaries reproduce the trills of their foster parents they are slower and much shorter (Mann et al., 2020). In the presence of urban noise, house finches increase the frequency of their vocalizations to minimize competition with the lower frequency sounds (Bermúdez-Cuamatzin et al., 2009; Bermúdez-Cuamatzin et al., 2023; Fernández-Juricic et al., 2005), in spite of the fact that urban birds evolve longer bills for foraging that should lead to lower frequencies (Giraudeau et al., 2014).

## 2 Analysis

Unless otherwise stated, all models were fit in STAN using the brms package in R (Bürkner, 2017), with 20,000 iterations across four MCMC chains. Model comparison was conducted using the widely-applicable information criterion (WAIC), a more robust alternative to AIC (Watanabe, 2021), and a Bayesian form of  $R^2$  (Gelman et al., 2019). Prior specifications, model diagnostics, and full model output for all analyses can be found in the [Supplementary Information](#). I use the diagnostic heuristics recommended by Gelman et al. (2014): effective sample size  $> 100$ , and  $1 < \hat{R} < 1.1$ .

### 2.1 Data

The recordings used in this study were collected in 1975 (Mundinger, 1975), 2012 (Ju et al., 2019), and 2019 (Youngblood & Lahti, 2022) in the New York metropolitan area (Brooklyn, Bronx, Queens, and Nassau County). Recordings from 1975 were collected with a Nagra III reel-to-reel tape recorder and a Sennheiser 804 shotgun microphone and converted to digital files (32 bit, 96 kHz) by the Cornell Lab of Ornithology in 2013 (Ju et al., 2019). Recordings from 2012 (16 bit, 44.1 kHz) and 2019 (32 bit, 48 kHz) were collected with a Marantz PD661 solid-state recorder and a Sennheiser ME66 shotgun microphone. The recordings from 1975 were downsampled to 48 kHz prior to analysis (Luscinia processes both 44.1 kHz and 48 kHz files). In all three years special precautions were taken to avoid recording the same bird twice (Ju et al., 2019). Each site was visited only once. Within a site, only one individual was recorded within a 160 m radius until they stopped singing or flew away.

All songs were analyzed by Youngblood & Lahti (2022) using Luscinia, a database and analysis program developed specifically for birdsong (<https://rflachlan.github.io/Luscinia/>). Songs were analyzed with a high-pass threshold of 2000 Hz, a maximum frequency of 9000 Hz and 5 dB of noise removal. 965 songs (26.2%) were excluded from the analysis due to high levels of noise or overlap with other bird species. Continuous traces with more than 20 ms between them were classified as syllables (Mundinger, 1975). Some syllables were further separable into elements, or traces with less than 20 ms between them. For this study, I used the mean frequency trace (mean frequency in each 1 ms bin) for each syllable in every song analyzed by Youngblood & Lahti (2022).

Figure 2.1 shows an example of the first ten syllables of a house finch song recorded by Mundinger (1975) and analyzed in Luscinia, where the blue line is the mean frequency over time.

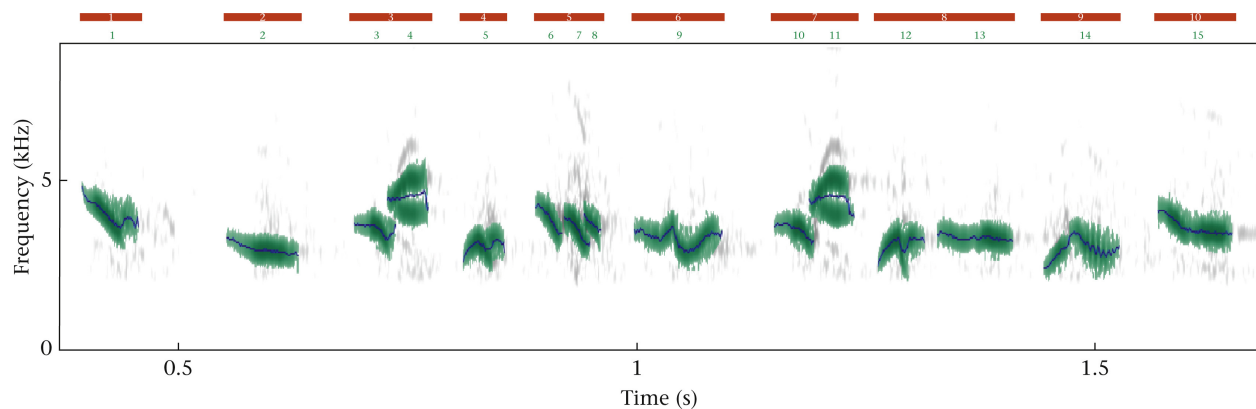


Figure 2.1: The first 10 syllables from a song recorded by Mundinger (1975) and analyzed in Luscinia. Each red bar corresponds to a syllable, and each green number corresponds to an element within that syllable. The blue traces represent mean frequency. In this song, syllable 3 and syllable 7 were classified as the same syllable type during dynamic time warping, and all other syllables are unique types. Reprinted from Youngblood & Lahti (2022).

The data used in this study comes directly from Youngblood and Lahti's repository (Youngblood & Lahti, 2022).



## 2.2 Clustering

Mean frequency was log-transformed prior to clustering (Lachlan et al., 2018). First, the normalized distances between all of the syllables were calculated via dynamic time warping with a window size of 10 (10% of the average signal length) using the dtwclust package in R (Sardá-Espinosa, 2019). A window size of 10% of the signal length is commonly used in speech processing research and seems to be a practical upper limit for many applications (Ratanamahatana & Keogh, 2005). Infinite distances (0.19% of cases) caused by comparisons of syllables with extreme signal length differences were assigned the maximum observed distance value. Next, hierarchical clustering and dynamic tree cut were used to cluster the syllables into types Roginek (2018). Hierarchical clustering was conducted with the UPGMA method implemented in the fastcluster package in R (Müllner, 2013), and dynamic tree cut was conducted with the dynamicTreeCut package in R (Langfelder et al., 2008).

For dynamic tree cut, I ran the hybrid algorithm with a minimum cluster size of 1 to maximize the representation of rare syllable types. The deep split parameter ( $DS$ ) determines the granularity of clustering by controlling the value of two other parameters: the maximum core scatter ( $MCS$ ), which controls the maximum within-group variation, and the minimum gap ( $MG$ ), which controls the minimum between-group variation (Ju, 2016; Langfelder et al., 2008).  $DS = \{0, 1, 2, 3, 4\}$  correspond to  $MCS = \{0.64, 0.73, 0.82, 0.91, 0.95\}$ , and  $MG = (1 - MCS) * 0.75$ . In their analyses of house finch song, Ju et al. (2019) and Roginek (2018) manually set  $MCS = 1$  and  $MG = 0.5$  while Youngblood & Lahti (2022) used  $DS = 3$  (corresponding to  $MCS = 0.91$  and  $MG = 0.0675$ ), all of which led to a similar number of syllable types. I will follow Youngblood & Lahti (2022) in using the simpler deep split parameter to control granularity in clustering, as this approach is recommended by the creators of dynamic tree cut (Langfelder et al., 2008) and has been widely used for a variety of applications (Liu et al., 2022; Zhao et al., 2020) including vocal analysis (Burkett et al., 2015).

The results of clustering can be seen in Figure 2.2.

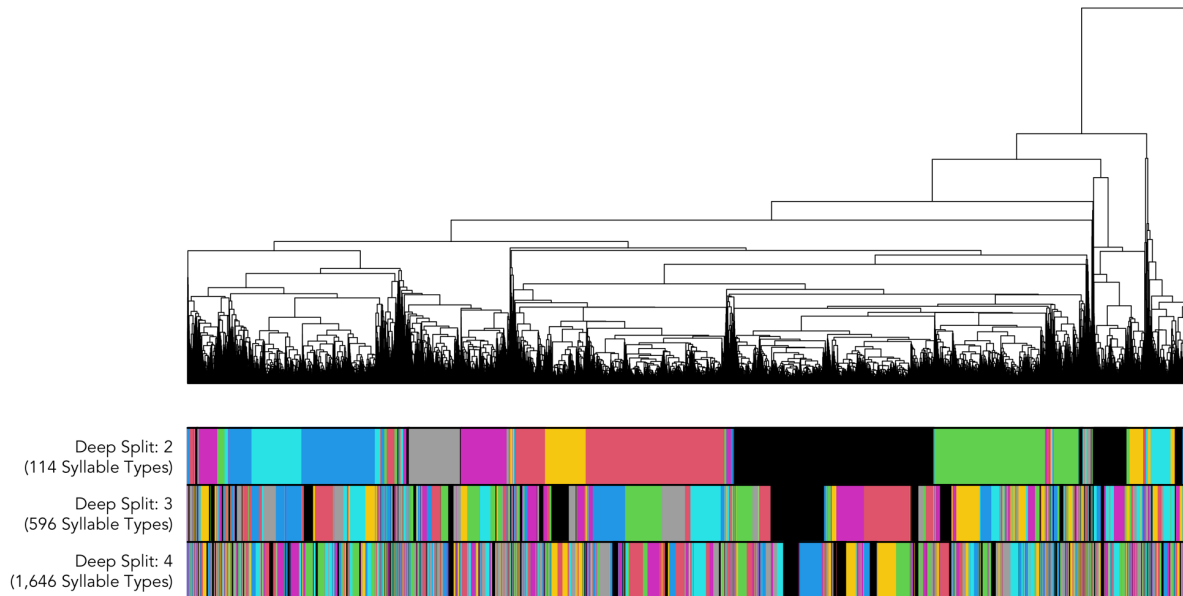


Figure 2.2: The results of hierarchical clustering. The colored bars below the dendrogram correspond to the categories assigned to each syllable when deep split is 2 (over-lumping), 3 (baseline), and 4 (over-splitting).

Once syllable types were identified by hierarchical clustering, I followed Ju et al. (2019) and Youngblood & Lahti (2022) in calculating the following syllable-level acoustic features for further analysis: average frequency (Hz), minimum frequency (Hz), maximum frequency (Hz), bandwidth (Hz), duration (ms), concavity (changes in the sign of the slope of the mean frequency trace/ms) and excursion (cumulative absolute change in Hz/ms). Concavity and excursion are both indicators of syllable complexity (Ju et al., 2019; Podos et al., 2016). Concavity was calculated after smoothing the mean

frequency trace using a polynomial spline with a smoothing parameter of 5 (Youngblood & Lahti, 2022). These acoustic features were averaged across all of the observations of each syllable type.

## 2.3 Zipf's Rank-Frequency Law

Mandelbrot's generalization of Zipf's rank-frequency law takes the following form (Mandelbrot, 1953, 1962):

$$f(r) = \frac{c}{(r + \beta)^\alpha} \quad (2.1)$$

where  $f(r)$  is the normalized frequency at each rank  $r$ , and  $\alpha$  and  $\beta$  are parameters that control slope and convexity (respectively). According to Izsák (2006), the bounds of (2.1) are  $\alpha > 1$ ,  $\beta > -1$ , and  $c > 1$ . When  $\beta = 0$ , this function simplifies to the original form of Zipf's rank-frequency law:  $f(r) \propto 1/r^\alpha$ .  $c$  is usually a normalization constant (Izsák, 2006; Mačutek, 2022; Mouillot & Lepretre, 2000) defined as:

$$c = \sum_{i=1}^{\infty} \frac{1}{(r + \beta)^\alpha} \quad (2.2)$$

In practice, this form of Zipf's rank-frequency law is notoriously difficult to fit to data due to strong correlations between  $\alpha$  and  $\beta$ , which in turn determine  $c$  (Izsák, 2006; Mačutek, 2022; Mouillot & Lepretre, 2000). Here, I use a simplified version of (2.1) that treats  $c$  as a third parameter that is estimated alongside  $\alpha$  and  $\beta$  (Ausloos, 2014), as has been done in studies of chickadee calls (Ficken et al., 1994; Freeberg & Lucas, 2012; Hailman, 1994), which should be interpreted as an approximation of the Zipf-Mandelbrot distribution.

I fit (2.1) to the rank-frequency distributions of the syllable classifications from each level of deep split in two batches. First, I fit all three parameters to approximate Mandelbrot's versions of the rank-frequency law. Then, I set  $\beta = 0$  to approximate Zipf's original formulation.

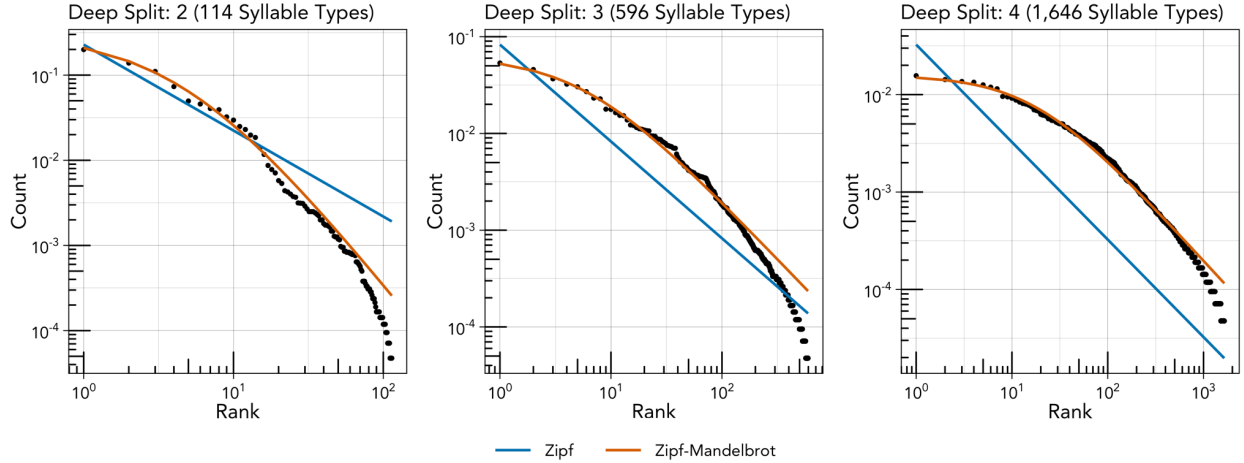


Figure 2.3: The relationship between rank (x-axis) and count (y-axis) at each level of deep split (left, center, and right). The blue and orange lines denote the expected distributions according to Zipf's rank-frequency law (blue) and Mandelbrot's extension of it (orange).

As can be seen in Figure 2.3, the observed frequency distribution is consistent with the Zipf-Mandelbrot distribution at all three levels of granularity ( $R^2 = \{0.995, 0.995, 0.995\}$  at  $DS = \{2, 3, 4\}$ ), but is only weakly consistent with the original form of Zipf's rank frequency law ( $R^2 = \{0.945, 0.777, 0.523\}$  at  $DS = \{2, 3, 4\}$ ).  $R^2$  is the most commonly reported goodness-of-fit measure for both forms of Zipf's rank-frequency law, and is generally  $> 0.8$  in studies of human language (Linders & Louwerse, 2023). The WAIC values and fitted parameters for all six models can be found in the [Supplementary Information](#).

## 2.4 Zipf's Law of Abbreviation

Zipf's law of abbreviation predicts that common items will be shorter in duration than rarer ones. Rather than focusing on duration alone, I explored whether the frequency of syllables is negatively correlated with four different measures of production cost: duration (ms), bandwidth (Hz), concavity (changes in the sign of the slope of the mean frequency trace/ms) and excursion (cumulative absolute change in Hz/ms). As mentioned in [Clustering](#), concavity and excursion are both indicators of syllable complexity (Ju et al., 2019; Podos et al., 2016).

For each level of deep split and each measure of production cost, I constructed a log-normal model with the measure in question as the outcome variable, count as the predictor variable, and syllable type as a varying intercept. The alternative formulation, a Poisson model with count as the outcome variable, does not allow for the correct random effects structure (e.g. counts are identical across observations of the same syllable type). Zipf's law of abbreviation predicts that there will be a strong effect of count on each measure of production cost.

*Table 2.1: The estimated effect of count on each measure of production cost, using the syllable classifications from each level of deep split. 95% credible intervals that do not overlap with 0, which indicate a strong effect, are marked with an asterisk.*

Model	DS	Est.	Err.	2.5%	97.5%	
duration ~ count	2	-0.36	0.14	-0.63	-0.10	*
	3	-0.46	0.06	-0.58	-0.34	*
	4	-0.42	0.03	-0.48	-0.35	*
bandwidth ~ count	2	-0.55	0.11	-0.75	-0.34	*
	3	-0.68	0.06	-0.79	-0.57	*
	4	-0.64	0.03	-0.69	-0.58	*
concavity ~ count	2	-0.02	0.10	-0.21	0.17	
	3	-0.06	0.04	-0.15	0.02	
	4	-0.07	0.03	-0.12	-0.02	*
excursion ~ count	2	-0.26	0.07	-0.41	-0.11	*
	3	-0.33	0.04	-0.41	-0.25	*
	4	-0.32	0.02	-0.36	-0.27	*

The results of the log-normal models can be seen in Table 2.1. Duration, bandwidth, and excursion had strong negative effects on count at all three levels of granularity. Concavity had no effect on count at the first two levels of granularity, but had a very weak negative effect at the third level of granularity.



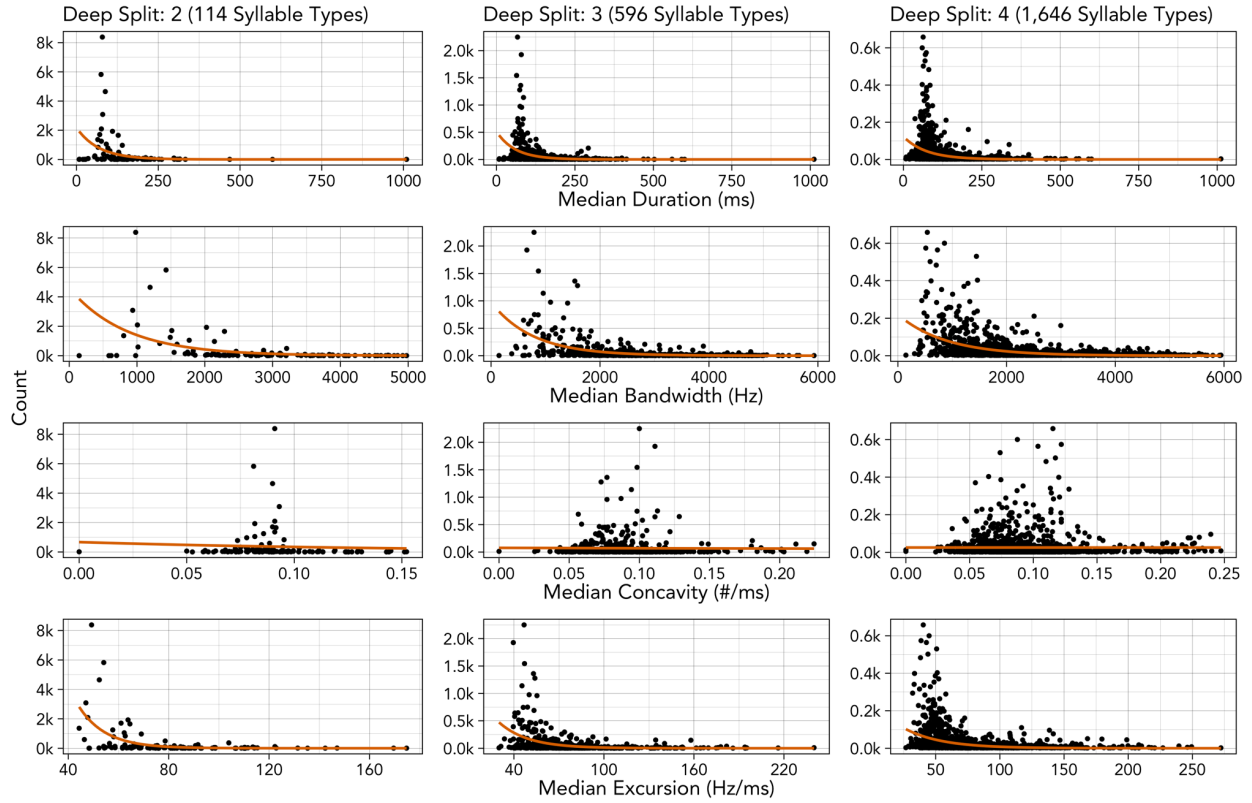


Figure 2.4: The relationship between four measures of production cost (x-axis) and count (y-axis) for each level of deep split (left, center, right). Each point shows the median value for a syllable type, so the orange best fit lines are from a simple Poisson model ( $\text{count} \sim \text{cost}$ ) rather than the full log-normal model.

These patterns are apparent in the plots of median production costs and counts in Figure 2.4. The lack of effect for concavity may be due in part to the weaker right-skew.

## 2.5 Menzerath's Law

Menzerath's law predicts that longer sequences will be comprised of smaller items. Importantly, Menzerath's law is sometimes detected in random sequences from null models (Ferrer-i-Cancho et al., 2014; G. Torre et al., 2021; Tanaka-Ishii, 2021). I think that there are two sorts of null models that make sense in this context: (1) random sequences with the same number of syllables as real songs, and (2) pseudorandom sequences that match both the number of syllables and cumulative duration in time. James et al. (2021) interpret the latter as approximating simple motor constraints—Menzerath's law resulting from efficiency in production alone—where stronger effects in the real data would indicate additional mechanisms (e.g. communicative efficiency). In this study, I compare the real data against each of these to assess both the presence of Menzerath's law and whether it is beyond what would be expected from production constraints.

The production constraint model of James et al. (2021) works as follows. For each iteration of the model, a pseudorandom sequence was produced for each real song in the dataset. Syllables were randomly sampled (with replacement) from the population until the duration of the random sequence exceeded the duration of the real song. If the difference between the duration of the random sequence and the real song was  $< 50\%$  of the duration of the final syllable, then the final syllable was kept in the sequence. Otherwise, it was removed. Each iteration of the model produces a set of random sequences with approximately the same distribution of durations as the real data.

To estimate the strength of Menzerath's law, I constructed a log-normal model with syllable duration as the outcome variable, song length in number of syllables as the predictor variable, and the song as a varying intercept (Safryghin et al., 2022). The production constraint model removes variation accounted for by the year and individual bird, so song is the

only varying intercept that is appropriate for comparison. Syllable durations do not differ across years. This model was used to estimate the strength of the effect of song length on syllable duration in both the real data and 10 simulated datasets from both null models. The `brm_multiple` function from the `brms` package in R was used to fit a single model to the the 10 simulated datasets from each null model and produce a combined posterior distribution (Bürkner, 2017).

This analysis differs from James et al. (2021) in two key ways: (1) I use the actual duration of syllables rather than a single median value for each song, and (2) I compare the full posterior distributions rather than point estimates of effects. These decisions should yield more conservative conclusions about the differences between the real and simulated data. Note that syllable type is also not incorporated into the modeling, so this is the only analysis that is not conducted across multiple levels of deep split.

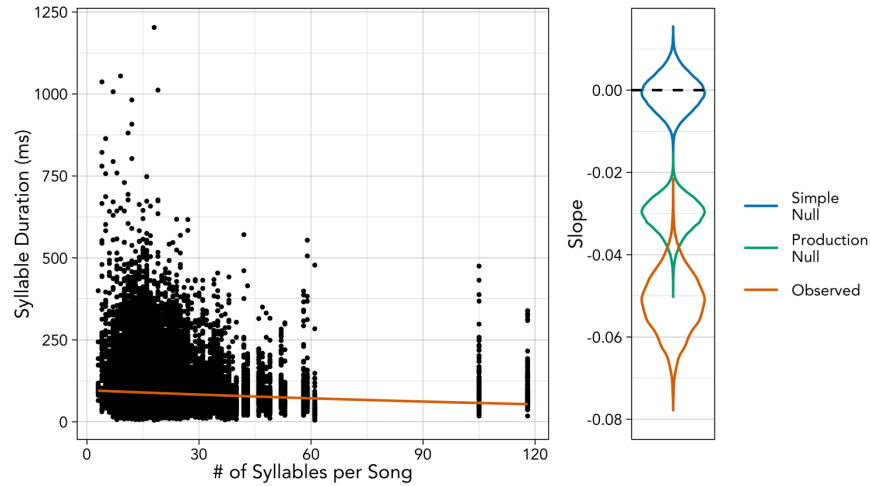


Figure 2.5: The left panel shows the relationship between song length in number of syllables ( $x$ -axis) and syllable duration ( $y$ -axis), with a best fit line from the fitted lognormal model. The right panel shows the posterior distribution of the effect of song length on duration for the real data (orange) compared to 10 simulated datasets from the simple (blue) and production (green) null models. The combined posteriors for the simulated datasets are based on 2,500 posterior samples from each of the 10 models.

The results of the log-normal model indicate that song length has a negative effect on syllable duration (Mean Estimate: -0.052; 95% CI: [-0.066, -0.038]) (left panel of Figure 2.5). The posterior distribution for this effect from the model fit to the actual data (orange) is more negative than the posterior distributions from 10 simulated datasets from the production constraint null model of James et al. (2021) (green) and a simple random null model (blue), although there is notable overlap between the lower tail of the production constraining model and the upper tail of the actual data (right panel of Figure 2.5).

## 2.6 Small-Worldness Index

The small-worldness index ( $SWI$ ) is a ratio based on the clustering coefficient ( $C$ ) and average path length ( $L$ ) (Humphries & Gurney, 2008):

$$SWI = \frac{C/C_{rand}}{L/L_{rand}} \quad (2.3)$$

where  $C_{rand}$  and  $L_{rand}$  are calculated from random networks of the same size as the real network. Values of  $SWI > 1$  are consistent with small-world structure (Humphries & Gurney, 2008), which is thought to reflect general systematic structure and thus compressibility (Allen et al., 2019; Raviv et al., 2021).

In this study, I followed Allen et al. (2019) in calculating  $SWI$  from the unweighted directed network of all syllable transitions in the population (lower panel of Figure 2.6A) to allow comparison with previous studies of non-human song.

Note that I am using the abbreviation *SWI* rather than the more commonly used *S* to differentiate it from entropy ( $\hat{S}$ , see [Mutual Information](#)).

Importantly, if the frequency distribution of syllable types is very skewed, then random sequences could exhibit small-world structure simply because of clustering around common types. To avoid this confound, *SWI* was calculated 1,000 times from the real data and 1,000 times from random sequences with the same distribution of types. Variation in observed *SWI* is due to  $C_{rand}$  and  $L_{rand}$  being computed from different random networks. It is also possible to compute *S* with a weighted undirected network ([Muldoon et al., 2016](#)), but I opted for the unweighted form to allow for comparison with other whale and birdsong studies.

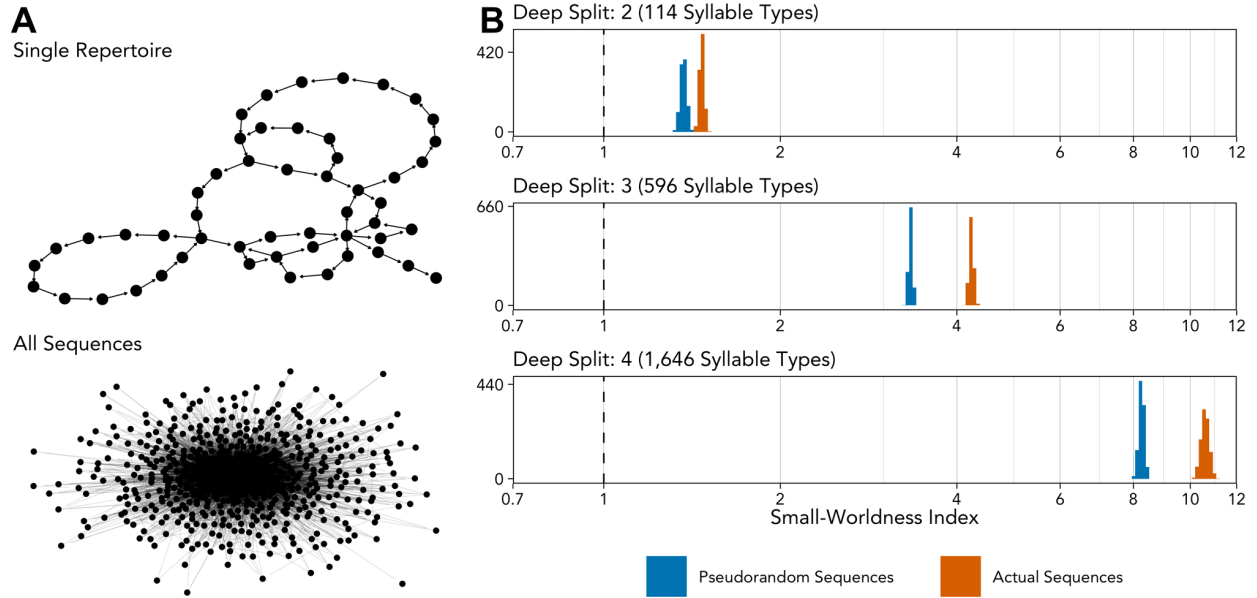


Figure 2.6: The left panel shows transition network between syllables types for a single male (#22 from 1975) above the global transition network for the entire dataset. The right panel shows the estimated small-worldness index calculated 1,000 times from the real data (blue) compared to pseudorandom sequences with the same frequency distribution of syllable types (orange). The dashed vertical line at 1 corresponds to the standard threshold for a network having small-world structure ([Humphries & Gurney, 2008](#)).

Figure 2.6B shows the distribution of *SWI* calculated from random sequences with the same distribution of types (orange) and the actual sequences (blue). At all three levels of granularity, the observed small-worldness of the real songs is above both the standard threshold ( $SWI > 1$ ) and the pseudorandom sequences.

## 2.7 Mutual Information

The rate at which the mutual information between items decays with distance reflects whether underlying syntactic structure is Markovian, hierarchical, or a composite of the two ([Sainburg et al., 2019](#)). Mutual information (*MI*) was calculated from pairs of sequential syllables separated by a certain distance, using the method of [Sainburg et al. \(2019\)](#). [Sainburg et al. \(2019\)](#) concatenated all songs recorded in a particular day before calculating *MI*. In our case, each day of recording represents several individuals recorded in several locations so concatenating sequences within days would lead to spurious pairs. Instead, I compute *MI* on concatenated sequences of songs from each individual. For example, if there are two individuals with two concatenated sequences of songs,  $a \rightarrow b \rightarrow c \rightarrow d$  and  $e \rightarrow f \rightarrow g$ , and the target distance is 2, then the pairs used for *MI* calculated would be  $((a, c), (b, d), (e, g))$ . *MI* would then be calculated using:

$$\hat{I}(X, Y) = \hat{S}(X) + \hat{S}(Y) - \hat{S}(X, Y) \quad (2.4)$$

where  $\hat{I}$  denotes information and  $\hat{S}$  denotes entropy.  $X$  is the distribution of first syllables ( $a, b, e$ ),  $Y$  is the distribution of second syllables ( $c, d, g$ ), and  $XY$  is the joint distribution of pairs  $((a, c), (b, d), (e, g))$ .  $\hat{S}$  was calculated using the method of Grassberger (2008) and Grassberger (2022) used by Lin & Tegmark (2017):

$$\hat{S} = \ln(N) - \frac{1}{N} \sum_{i=1}^K N_i \psi(N_i) \quad (2.5)$$

where  $N$  is the total number of tokens in the distribution,  $N_i$  is the number of tokens within each type  $i$  out of the total number of types  $K$ , and  $\psi$  is the digamma function. The actual  $MI$  is then measured using:

$$MI = \hat{I} - \hat{I}_{sh} \quad (2.6)$$

where  $\hat{I}_{sh}$  is the estimated lower bound of  $MI$  calculated from shuffled sequences, created by randomly permuting the concatenated sequences of individuals' songs.

Sainburg et al. (2019) and Sainburg et al. (2022) simulated data from the hierarchical model of Lin & Tegmark (2017), the Markov model of Katahira et al. (2011), and their composite model in which Markov chains are nested within a larger hierarchical structure.

To determine whether the observed  $MI$  decay was consistent with a Markov process, hierarchical process, or both, I fit the following three decay models to the data:

$$\text{exponential decay: } y = ae^{-xb} \quad (2.7)$$

$$\text{power-law decay: } y = cx^d \quad (2.8)$$

$$\text{composite decay: } y = ae^{-xb} + cx^d \quad (2.9)$$

where  $y$  is the estimated  $MI$  and  $x$  is the distance between syllables (Sainburg et al., 2019; Sainburg et al., 2022). Sainburg et al. (2019) included an intercept  $f$  in all three models, but I removed it because it led to overfitting issues (e.g. exponential model with an extra “knee” after the initial decay), did not significantly improve model fit ( $\Delta WAIC < 2$ ), and was not included in Lin & Tegmark (2017). I focused my analysis on distances of up to 100 syllables to enable easy comparison with Sainburg et al. (2019). I followed Sainburg et al. (2019) in comparing the fit of each model at increasing distances from 100 to 1,200 (the longest individual sequence is 1,219), and found that the composite model outperforms both the exponential and power-law models at all distances (see [Supplementary Information](#)). Note that  $MI$  decay becomes noisy and begins to slowly increase past ~400 syllables. This is because only a small number of individuals have concatenated sequences longer than that (~4%), and  $MI$  is known to be systematically biased upwards at low sample sizes (Treves & Panzeri, 1995).

Table 2.2: The WAIC and R-Squared value for each model at each level of deep split.

DS	Model	WAIC	R-Sq
2	Exponential	-829	0.951
	Power-Law	-766	0.910
	Composite	-840	0.952
3	Exponential	-617	0.982
	Power-Law	-487	0.927
	Composite	-668	0.986
4	Exponential	-599	0.990
	Power-Law	-432	0.926
	Composite	-688	0.993

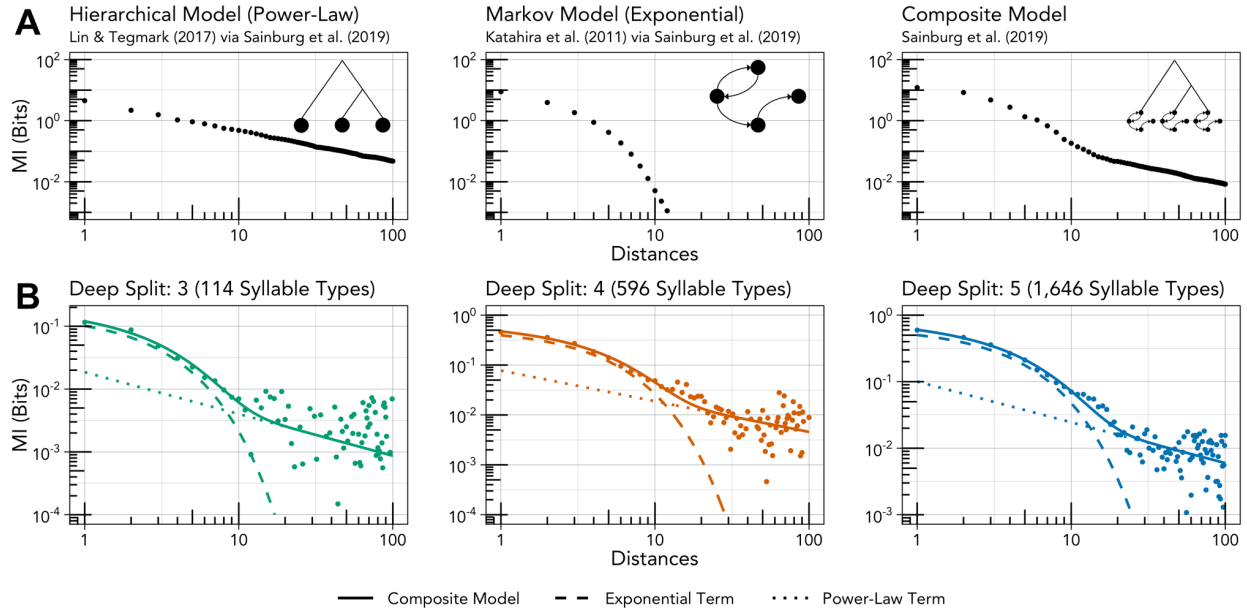


Figure 2.7: The top panel shows simulated mutual information decay curves from the Lin & Tegmark (2017) hierarchical model (left), the Katahira et al. (2011) Markov model (center), and the Sainburg et al. (2019) composite model (right) (data and inspiration for diagrams from Sainburg et al., 2019). The bottom panel shows the computed mutual information decay curves for the observed data at different levels of deep split (left, center, right). The solid line corresponds to the full composite model, while the dashed and dotted lines correspond to the exponential and power-law terms, respectively.

At all three levels of granularity, the composite model has a lower  $WAIC$  and higher  $R^2$  than both the exponential and power-law models (Table 2.2 and Figure 2.5), suggesting that mutual information decay in house finch song is more consistent with a combination of Markovian and hierarchical processes rather than with one or the other.

### 3 Discussion

All three linguistic laws considered here are present in house finch song. Three out of the four measures of production cost, most importantly duration, are consistently and strongly negatively correlated with frequency, providing robust evidence for Zipf’s law of abbreviation. Menzerath’s law also found solid support, with a steeper negative relationship between song length and syllable duration than predicted by a model of production constraints (James et al., 2021). Together, these results show clear evidence for efficiency—syllables that are difficult to produce are less common and more likely to appear in shorter songs. Mandelbrot’s form of Zipf’s rank-frequency law provided a good fit to the data and has a more convex shape than the original form, which is consistent with studies of other non-human communication systems that have more redundancy than human language (Allen et al., 2019; Briefer et al., 2010; Cody et al., 2016; Hailman et al., 1985; Martins, 1994; McCowan et al., 1999). I will limit my interpretation of this result, given ongoing debates about the cause of the rank-frequency law, but it is notable that the goodness-of-fit of the Zipf-Mandelbrot distribution to house finch song ( $R^2 > 0.99$ ) is within the range of human language (Linders & Louwerse, 2023; Piantadosi, 2014).

The two structural properties of language considered here are also found in house finch song. First, the syllable network has a small-world structure, characterized by high levels of clustering and low average path lengths, which is thought to reflect systematic structure and efficient recall in human language (Cancho & Solé, 2001; Motter et al., 2002; Steyvers & Tenenbaum, 2005). The small-worldness index of house finch song is within the range seen in humpback whales (Allen et al., 2019) and other songbirds at the first two levels of granularity (Beckage et al., 2011; Cody et al., 2016; Hedley, 2016; Sasahara et al., 2012; Weiss et al., 2014) ( $SWI \sim 1.69 - 4.7$ ), but is much higher when syllables are over-split into types ( $SWI \sim 10 - 11$ ). Second, the decay in mutual information between syllables with increasing distance has a similar shape

to that found in several human languages and songbird species (Sainburg et al., 2019), which is associated with Markov chains nested within a larger hierarchical structure. Historically, birdsong sequences were assumed to follow simple Markov chains (Berwick et al., 2011; Katahira et al., 2011; Leonardo & Konishi, 1999). This result lends support to an emerging consensus that song sequences are actually much more complex, containing long-range dependencies and hierarchical processes that parallel human language (Kershenbaum et al., 2014; Morita et al., 2021; Sainburg et al., 2019; Searcy et al., 2022). In combination, the small-worldness and mutual information decay of house finch song sequences suggest that they exhibit the kind of systematic structure that is thought to maximize expressivity while reducing learning costs (Allen et al., 2019; Kirby et al., 2015; Raviv et al., 2021), making it easier for more information to “pass through the bottleneck” of social learning (Gruber et al., 2022).

Notably, these patterns are consistently found across three levels of granularity in syllable clustering, pointing to a limited level of scale invariance in house finch song. Scale invariance refers to self-similarity in organization across levels of analysis, and is hypothesized to be a key indicator of complex systems (Stanley et al., 2000) including language and speech (Gervain & Geffen, 2019; Mehri & Jamaati, 2021), behavior (Bialek & Shaevitz, 2023), and cognition (Chater & Brown, 1999). I use the term “limited”, though, because here it is consistency across syllable classifications rather than across levels of hierarchy (e.g. phrases, songs, bouts). I have heard others studying the cultural evolution of birdsong refer to this idea as “fractal equivalency”—different resolutions of clustering should show similar statistical signals of organization (D. C. Lahti, personal communication, 2020). A practical conclusion of this finding is that information theoretic measures may not be reliable indicators of clustering quality in non-human communication systems (Kershenbaum et al., 2016), although this should be explored in other species.

A long-standing critique of Zipf’s laws is that they may be statistical artifacts of other processes (Caplan et al., 2020), starting with Miller’s observation that randomly typing on keyboards can produce similar patterns (Miller, 1957). That being said, random typing accounts are not realistic causal descriptions of how communication systems emerge, and there are good empirical reasons to doubt that they undermine efficiency accounts (Piantadosi, 2014). Randomly-generated texts produce rank-frequency distributions that differ from those in real corpora (Ferrer-i-Cancho & Elvevåg, 2010), random typing models are not truly neutral as they can be mathematically reframed as minimizing costs (Ferrer-i-Cancho, 2016; Ferrer-i-Cancho et al., 2022), and there is direct experimental evidence that Zipfian abbreviation emerges from pressure for efficient communication (Kanwal et al., 2017). In my view, the most important contribution of the random typing account is to highlight that the problem of equifinality—different processes leading to similar outcomes (Barrett, 2019)—means that patterns resembling Zipf’s laws are not sufficient to make conclusions about efficiency (Kanwal, 2017; Semple et al., 2022). Multiple lines of evidence should be presented alongside other work demonstrating that efficiency is shaping the system (e.g. physical (Mann et al., 2020) and environmental (Bermúdez-Cuamatzin et al., 2023) constraints), as I have done here. See Semple et al. (2022), Piantadosi (2014), and Kanwal (2017) for more complete summaries of this debate.

Outside of linguistics, efficiency and complexity are often discussed in relation to cumulative cultural evolution (CCE). Definitions of CCE vary and a full review is outside of the scope of this study, but for convenience we will use the definition of Williams et al. (2022): “the accumulation of sequential changes within a single socially learned behavior that results in improved function”. Discussions of CCE often focus on increasing complexity over time (Wilks & Blakey, 2018), which was once thought to be a hallmark of human culture (Tomasello, 1999) but has now been observed in several non-human communication systems including humpback whale (Garland et al., 2022) and Savannah sparrow song (Williams et al., 2022). Gruber et al. (2022) make a convincing argument that efficiency deserves more attention in CCE, as increases in complexity in one domain require increases in efficiency in another (see Equation (1.1) in the Introduction). House finch song may be a good research model for how the interplay between efficiency and complexity drives CCE, as male house finches have a social learning bias for more complex syllables (Youngblood & Lahti, 2022), possibly as an adaptation to female preferences for more complex songs (Ciaburri & Williams, 2019; Mennill et al., 2006; Nolan & Hill, 2004), and there appears to be pressure for efficiency at the level of both syllables and songs. That being said, CCE may not be the best framework for understanding the interaction between efficiency and complexity in birdsong, as its logic is more difficult to apply to “aesthetic” behavior (Sinclair et al., 2022) especially when it is optimized for female preferences that evolve to maximize inclusive fitness rather than the specific properties of songs that males sing (Geller & Lahti, in press).



## Acknowledgments

This work was inspired in large part by conversations with Olivier Morin, Alexey Koshevoy, and other attendees of the Communicative Efficiency Workshop hosted by the Max Planck Institute for Geoanthropology in April 2023.

## Data & Code Availability

The data for this study come from Youngblood & Lahti (2022) (<https://github.com/masonyoungblood/TransmissionBias>). The analysis code for this study is available in the source R Markdown file in the GitHub repository ([https://github.com/masonyoungblood/linguistic\\_efficiency](https://github.com/masonyoungblood/linguistic_efficiency)).

## References

- Allegrini, P., Grigolini, P., & Palatella, L. (2004). Intermittency and scale-free networks: a dynamical model for human language complexity. *Chaos, Solitons & Fractals*, 20(1), 95–105. [https://doi.org/10.1016/S0960-0779\(03\)00432-6](https://doi.org/10.1016/S0960-0779(03)00432-6)
- Allen, J. A., Garland, E. C., Dunlop, R. A., & Noad, M. J. (2019). Network analysis reveals underlying syntactic features in a vocally learnt mammalian display, humpback whale song. *Proceedings of the Royal Society B: Biological Sciences*, 286(1917), 20192014. <https://doi.org/10.1098/rspb.2019.2014>
- Altmann, E. G., Cristadoro, G., & Esposti, M. D. (2012). On the origin of long-range correlations in texts. *Proceedings of the National Academy of Sciences*, 109(29), 11582–11587. <https://doi.org/10.1073/pnas.1117723109>
- Alvarez-Lacalle, E., Dorow, B., Eckmann, J.-P., & Moses, E. (2006). Hierarchical structures induce long-range dynamical correlations in written texts. *Proceedings of the National Academy of Sciences*, 103(21), 7956–7961. <https://doi.org/10.1073/pnas.0510673103>
- Ausloos, M. (2014). Zipf–Mandelbrot–Pareto model for co-authorship popularity. *Scientometrics*, 101(3), 1565–1586. <https://doi.org/10.1007/s11192-014-1302-y>
- Barrett, B. J. (2019). Equifinality in empirical studies of cultural transmission. *Behavioural Processes*, 161, 129–138. <https://doi.org/10.1016/j.beproc.2018.01.011>
- Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS ONE*, 6(5), e19348. <https://doi.org/10.1371/journal.pone.0019348>
- Benedict, L., & Najar, N. A. (2019). Are commonly used metrics of bird song complexity concordant? *The Auk*, 136(1), uky008. <https://doi.org/10.1093/auk/uky008>
- Bentz, C., & Ferrer-i-Cancho, R. (2016). Zipf's law of abbreviation as a language universal. *Proceedings of the Leiden Workshop on Capturing Phylogenetic Algorithms for Linguistics*, 1–4. <https://doi.org/10.15496/publikation-10057>
- Bermúdez-Cuamatzin, E., García, C. M., Ríos-Chelén, A. A., & Gil, D. (2009). Strategies of song adaptation to urban noise in the house finch: syllable pitch plasticity or differential syllable use? *Behaviour*, 146(9), 1269–1286. <https://doi.org/10.1163/156853909X423104>
- Bermúdez-Cuamatzin, E., Slabbekoorn, H., & Macías García, C. (2023). Spectral and temporal call flexibility of House Finches (*Haemorrhous mexicanus*) from urban areas during experimental noise exposure. *Ibis*, 165(2), 571–586. <https://doi.org/10.1111/ibi.13161>
- Berwick, R. C., Okanoya, K., Beckers, G. J. L., & Bolhuis, J. J. (2011). Songs to syntax: the linguistics of birdsong. *Trends in Cognitive Sciences*, 15(3), 113–121. <https://doi.org/10.1016/j.tics.2011.01.002>
- Bialek, W., & Shalvitz, J. W. (2023). Long time scales, individual differences, and scale invariance in animal behavior (arXiv:2304.09608). arXiv. <http://arxiv.org/abs/2304.09608>
- Boroda, M., & Altmann, G. (1991). Menzerath's law in musical texts. *Musikometrika*, 3, 1–13.
- Briefer, E., Osiejuk, T. S., Rybak, F., & Aubin, T. (2010). Are bird song complexity and song sharing shaped by habitat structure? An information theory and statistical approach. *Journal of Theoretical Biology*, 262(1), 151–164. <https://doi.org/10.1016/j.jtbi.2009.09.020>
- Burkett, Z. D., Day, N. F., Peñagarikano, O., Geschwind, D. H., & White, S. A. (2015). VoICE: A semi-automated pipeline for standardizing vocal analysis across models. *Scientific Reports*, 5(1), 10237. <https://doi.org/10.1038/srep10237>
- Bürkner, P.-C. (2017). brms: an R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1). <https://doi.org/10.18637/jss.v080.i01>
- Cancho, R. F. i., & Solé, R. V. (2001). The small world of human language. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 268(1482), 2261–2265. <https://doi.org/10.1098/rspb.2001.1800>
- Caplan, S., Kodner, J., & Yang, C. (2020). Miller's monkey updated: communicative efficiency and the statistics of words in natural language. *Cognition*, 205, 104466. <https://doi.org/10.1016/j.cognition.2020.104466>
- Carr, J. W., Smith, K., Cornish, H., & Kirby, S. (2017). The cultural evolution of structured languages in an open-ended, continuous world. *Cognitive Science*, 41(4), 892–923. <https://doi.org/10.1111/cogs.12371>
- Charbonneau, M., & Bourrat, P. (2021). Fidelity and the grain problem in cultural evolution. *Synthese*, 199(3–4), 5815–5836. <https://doi.org/10.1007/s11229-021-03047-1>
- Chater, N., & Brown, G. D. A. (1999). Scale-invariance as a unifying psychological principle. *Cognition*, 69(3), B17–B24. [https://doi.org/10.1016/S0010-0277\(98\)00066-3](https://doi.org/10.1016/S0010-0277(98)00066-3)
- Chimento, M., Alarcón-Nieto, G., & Aplin, L. (2021). Population turnover facilitates cultural selection for efficiency. *Current Biology*, 31, 2477–2483. <https://doi.org/10.1016/j.cub.2021.03.057>
- Chomsky, N. (1957). *Syntactic Structures*. Mouton & Co.
- Ciaburri, I., & Williams, H. (2019). Context-dependent variation of house finch song syntax. *Animal Behaviour*, 147, 33–42. <https://doi.org/10.1016/j.anbehav.2018.11.001>
- Clink, D. J., Ahmad, A. H., & Klinck, H. (2020). Brevity is not a universal in animal communication: evidence for compression depends on the unit of analysis in small ape vocalizations. *Royal Society Open Science*, 7(4), 200151. <https://doi.org/10.1098/rsos.200151>

- Clink, D. J., & Lau, A. R. (2020). Adherence to Menzerath's Law is the exception (not the rule) in three duetting primate species. *Royal Society Open Science*, 7(11), 201557. <https://doi.org/10.1098/rsos.201557>
- Cody, M. L., Stabler, E., Sánchez Castellanos, H. M., & Taylor, C. E. (2016). Structure, syntax and “small-world” organization in the complex songs of California Thrashers (*Toxostoma redivivum*). *Bioacoustics*, 25(1), 41–54. <https://doi.org/10.1080/09524622.2015.1089418>
- Corral, Á., & Serra, I. (2020). The brevity law as a scaling law, and a possible origin of Zipf's law for word frequencies. *Entropy*, 22(2), 224. <https://doi.org/10.3390/e22020224>
- Cramer, I. (2005). The parameters of the Altmann-Menzerath law. *Journal of Quantitative Linguistics*, 12(1), 41–52. <https://doi.org/10.1080/09296170500055301>
- De Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics*, 28(4), 441–465. <https://doi.org/10.1006/jpho.2000.0125>
- Demartsev, V., Gordon, N., Barocas, A., Bar-Ziv, E., Ilany, T., Goll, Y., Ilany, A., & Geffen, E. (2019). The “law of brevity” in animal communication: sex-specific signaling optimization is determined by call amplitude rather than duration. *Evolution Letters*, 3(6), 623–634. <https://doi.org/10.1002/evl3.147>
- Diez, A., & MacDougall-Shackleton, S. A. (2020). Zebra finches go wild! Experimental cultural evolution of birdsong. *Behaviour*, 157(3–4), 231–265. <https://doi.org/10.1163/1568539X-00003588>
- Ebeling, W., & Pöschel, T. (1994). Entropy and long-range correlations in literary English. *Europhysics Letters*, 26(4). <https://doi.org/10.1209/0295-5075/26/4/001>
- Eroglu, S. (2013). Menzerath–Altmann law for distinct word distribution analysis in a large text. *Physica A: Statistical Mechanics and Its Applications*, 392(12), 2775–2780. <https://doi.org/10.1016/j.physa.2013.02.012>
- Favaro, L., Gamba, M., Cresta, E., Fumagalli, E., Bandoli, F., Pilega, C., Isaja, V., Mathevon, N., & Reby, D. (2020). Do penguins' vocal sequences conform to linguistic laws? *Biology Letters*, 16(2), 20190589. <https://doi.org/10.1098/rsbl.2019.0589>
- Fay, N., & Ellison, T. M. (2013). The cultural evolution of human communication systems in different sized populations: usability trumps learnability. *PLoS ONE*, 8(8), e71781. <https://doi.org/10.1371/journal.pone.0071781>
- Fedurek, P., Zuberbühler, K., & Semple, S. (2017). Trade-offs in the production of animal vocal sequences: insights from the structure of wild chimpanzee pant hoots. *Frontiers in Zoology*, 14(1), 50. <https://doi.org/10.1186/s12983-017-0235-8>
- Fedzechkina, M., Jaeger, T. F., & Newport, E. L. (2012). Language learners restructure their input to facilitate efficient communication. *Proceedings of the National Academy of Sciences*, 109(44), 17897–17902. <https://doi.org/10.1073/pnas.1215776109>
- Fehér, O., Wang, H., Saar, S., Mitra, P. P., & Tchernichovski, O. (2009). De novo establishment of wild-type song culture in the zebra finch. *Nature*, 459(7246), 564–568. <https://doi.org/10.1038/nature07994>
- Fernández-Juricic, E., Poston, R., Collibus, K. D., Morgan, T., Bastain, B., Martin, C., Jones, K., & Treminio, R. (2005). Microhabitat selection and singing behavior patterns of male house finches (*Carpodacus mexicanus*) in urban parks in a heavily urbanized landscape in the western U.S. *Urban Habitats*, 3(1).
- Ferrer-i-Cancho, R. (2016). Compression and the origins of Zipf's law for word frequencies. *Complexity*, 21(S2), 409–411. <https://doi.org/10.1002/cplx.21820>
- Ferrer-i-Cancho, R., Bentz, C., & Seguin, C. (2022). Optimal coding and the origins of Zipfian laws. *Journal of Quantitative Linguistics*, 29(2), 165–194. <https://doi.org/10.1080/09296174.2020.1778387>
- Ferrer-i-Cancho, R., & Elvevåg, B. (2010). Random texts do not exhibit the real Zipf's law-like rank distribution. *PLoS ONE*, 5(3), e9411. <https://doi.org/10.1371/journal.pone.0009411>
- Ferrer-i-Cancho, R., Hernández-Fernández, A., Baixeries, J., Dębowski, L., & Mačutek, J. (2014). When is Menzerath–Altmann law mathematically trivial? a new approach. *Statistical Applications in Genetics and Molecular Biology*, 13(6). <https://doi.org/10.1515/sagmb-2013-0034>
- Ferrer-i-Cancho, R., Hernández-Fernández, A., Lusseau, D., Agoramoorthy, G., Hsu, M. J., & Semple, S. (2013). Compression as a universal principle of animal behavior. *Cognitive Science*, 37(8), 1565–1578. <https://doi.org/10.1111/cogs.12061>
- Ficken, M. S., Hailman, E. D., & Hailman, J. P. (1994). The chick-a-dee call system of the Mexican chickadee. *The Condor*, 96(1), 70–82. <https://doi.org/10.2307/1369065>
- Fitch, W. T. (2000). The evolution of speech: a comparative review. *Trends in Cognitive Sciences*, 4(7), 258–267. [https://doi.org/10.1016/S1364-6613\(00\)01494-7](https://doi.org/10.1016/S1364-6613(00)01494-7)
- Frank, S. L., Bod, R., & Christiansen, M. H. (2012). How hierarchical is language use? *Proceedings of the Royal Society B: Biological Sciences*, 279(1747), 4522–4531. <https://doi.org/10.1098/rspb.2012.1741>
- Freeberg, T. M., & Lucas, J. R. (2012). Information theoretical approaches to chick-a-dee calls of Carolina chickadees (*Poecile carolinensis*). *Journal of Comparative Psychology*, 126(1), 68–81. <https://doi.org/10.1037/a0024906>
- G. Torre, I., Dębowski, L., & Hernández-Fernández, A. (2021). Can Menzerath's law be a criterion of complexity in communication? *PLoS ONE*, 16(8), e0256133. <https://doi.org/10.1371/journal.pone.0256133>
- Garamszegi, L. Z., Balsby, T. J. S., Bell, B. D., Borowiec, M., Byers, B. E., Draganoiu, T., Eens, M., Forstmeier, W., Galeotti, P., Gil, D., Gorissen, L., Hansen, P., Lampe, H. M., Leitner, S., Lontkowski, J., Nagle, L., Nemeth, E., Pinxten, R., Rossi, J.-M., ... Moller, A. P. (2005). Estimating the complexity of bird song by using capture-recapture approaches from community ecology. *Behavioral Ecology and Sociobiology*, 57(4), 305–317. <https://doi.org/10.1007/s00265-004-0866-6>
- Garland, E. C., Garrigue, C., & Noad, M. J. (2022). When does cultural evolution become cumulative culture? A case study of humpback whale song. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1843), 20200313. <https://doi.org/10.1098/rstb.2020.0313>
- Geller, F. C., & Lahti, D. C. (in press). Is sexiness cumulative? Arguments from birdsong culture. *Animal Behaviour*, 205.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2014). *Bayesian Data Analysis* (3rd ed.). CRC Press.
- Gelman, A., Goodrich, B., Gabry, J., & Vehtari, A. (2019). R-squared for Bayesian regression models. *The American Statistician*, 73(3), 307–309. <https://doi.org/10.1080/00031305.2018.1549100>
- Genty, E., & Byrne, R. W. (2010). Why do gorillas make sequences of gestures? *Animal Cognition*, 13(2), 287–301. <https://doi.org/10.1007/s10071-009-0266-4>
- Gervain, J., & Geffen, M. N. (2019). Efficient neural coding in auditory and speech perception. *Trends in Neurosciences*, 42(1), 56–65. <https://doi.org/10.1016/j.tins.2018.09.004>
- Gibson, E., Futrell, R., Piantadosi, S. P., Dautriche, I., Mahowald, K., Bergen, L., & Levy, R. (2019). How efficiency shapes human language. *Trends in Cognitive Sciences*, 23(5), 389–407. <https://doi.org/10.1016/j.tics.2019.02.003>
- Giraudeau, M., Nolan, P. M., Black, C. E., Earl, S. R., Hasegawa, M., & McGraw, K. J. (2014). Song characteristics track bill morphology along a gradient of urbanization in house finches (*Haemorrhous mexicanus*). *Frontiers in Zoology*, 11(1), 83. <https://doi.org/10.1186/s12983-014-0083-8>
- Grassberger, P. (2008). *Entropy estimates from insufficient samplings* (arXiv:physics/0307138). arXiv. <https://doi.org/10.48550/arXiv.physics/0307138>
- Grassberger, P. (2022). On generalized Schürmann entropy estimators. *Entropy*, 24(5), 680. <https://doi.org/10.3390/e24050680>

- Gruber, T., Chimento, M., Aplin, L. M., & Biro, D. (2022). Efficiency fosters cumulative culture across species. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 377(1843), 20200308. <https://doi.org/10.1098/rstb.2020.0308>
- Gustison, M. L., Semple, S., Ferrer-i-Cancho, R., & Bergman, T. J. (2016). Gelada vocal sequences follow Menzerath's linguistic law. *Proceedings of the National Academy of Sciences*, 113(19). <https://doi.org/10.1073/pnas.1522072113>
- Hahn, M., Jurafsky, D., & Putrell, R. (2020). Universals of word order reflect optimization of grammars for efficient communication. *Proceedings of the National Academy of Sciences*, 117(5), 2347–2353. <https://doi.org/10.1073/pnas.1910923117>
- Hailman, J. P. (1994). Constrained permutation in “chick-a-dee”-like calls of a black-lored tit *Parus xanthogenys*. *Bioacoustics*, 6(1), 33–50. <https://doi.org/10.1080/09524622.1994.9753270>
- Hailman, J. P., Ficken, M. S., & Ficken, R. W. (1985). The “chick-a-dee” calls of *Parus atricapillus*: a recombinant system of animal communication compared with written English. *Semiotica*, 56(3–4). <https://doi.org/10.1515/semi.1985.56.3-4.191>
- Hedley, R. W. (2016). Composition and sequential organization of song repertoires in Cassin's Vireo (*Vireo cassinii*). *Journal of Ornithology*, 157(1), 13–22. <https://doi.org/10.1007/s10336-015-1238-x>
- Heesen, R., Hobäiter, C., Ferrer-i-Cancho, R., & Semple, S. (2019). Linguistic laws in chimpanzee gestural communication. *Proceedings of the Royal Society B: Biological Sciences*, 286, 20182900. <https://doi.org/10.1098/rspb.2018.2900>
- Huang, M., Ma, H., Ma, C., Garber, P. A., & Fan, P. (2020). Male gibbon loud morning calls conform to Zipf's law of brevity and Menzerath's law: insights into the origin of human language. *Animal Behaviour*, 160, 145–155. <https://doi.org/10.1016/j.anbehav.2019.11.017>
- Humphries, M. D., & Gurney, K. (2008). Network “small-world-ness”: a quantitative method for determining canonical network equivalence. *PLoS ONE*, 3(4), e0002051. <https://doi.org/10.1371/journal.pone.0002051>
- Izsák, J. (2006). Some practical aspects of fitting and testing the Zipf-Mandelbrot model: a short essay. *Scientometrics*, 67(1), 107–120. <https://doi.org/10.1007/s11192-006-0052-x>
- Jaeger, T. F., & Tily, H. (2011). On language “utility”: processing complexity and communicative efficiency. *WIREs Cognitive Science*, 2(3), 323–335. <https://doi.org/10.1002/wcs.126>
- James, L. S., Mori, C., Wada, K., & Sakata, J. T. (2021). Phylogeny and mechanisms of shared hierarchical patterns in birdsong. *Current Biology*, 31(13), 2796–2808.e9. <https://doi.org/10.1016/j.cub.2021.04.015>
- Ju, C. (2016). *FinchCatcher* (Beta 3) [Computer software]. <http://finchcatcher.net>
- Ju, C., Geller, F. C., Mundinger, P. C., & Lahti, D. C. (2019). Four decades of cultural evolution in house finch songs. *The Auk: Ornithological Advances*, 136, 1–18. <https://doi.org/10.1093/auk/uky012>
- Kanwal, J. (2017). *Word length and the principle of least effort: language as an evolving, efficient code for information transfer* [Doctor of Philosophy]. School of Philosophy, Psychology, and Language Sciences, University of Edinburgh.
- Kanwal, J., Smith, K., Culbertson, J., & Kirby, S. (2017). Zipf's law of abbreviation and the principle of least effort: language users optimise a miniature lexicon for efficient communication. *Cognition*, 165, 45–52. <https://doi.org/10.1016/j.cognition.2017.05.001>
- Katahira, K., Suzuki, K., Okanoya, K., & Okada, M. (2011). Complex sequencing rules of birdsong can be explained by simple hidden Markov processes. *PLoS ONE*, 6(9), e24516. <https://doi.org/10.1371/journal.pone.0024516>
- Kershenbaum, A., Blumstein, D. T., Roch, M. A., Akçay, Ç., Backus, G., Bee, M. A., Bohn, K., Cao, Y., Carter, G., Căsar, C., Coen, M., DeRuiter, S. L., Doyle, L., Edelman, S., Ferrer-i-Cancho, R., Freeberg, T. M., Garland, E. C., Gustison, M., Harley, H. E., ... Zamora-Gutierrez, V. (2016). Acoustic sequences in non-human animals: a tutorial review and prospectus. *Biological Reviews*, 91, 13–52. <https://doi.org/10.1111/brv.12160>
- Kershenbaum, A., Bowles, A. E., Freeberg, T. M., Jin, D. Z., Lameira, A. R., & Bohn, K. (2014). Animal vocal sequences: not the Markov chains we thought they were. *Proceedings of the Royal Society B: Biological Sciences*, 281(1792), 20141370. <https://doi.org/10.1098/rspb.2014.1370>
- Kershenbaum, A., Demartsev, V., Gammon, D. E., Geffen, E., Gustison, M. L., Ilany, A., & Lameira, A. R. (2021). Shannon entropy as a robust estimator of Zipf's Law in animal vocal communication repertoires. *Methods in Ecology and Evolution*, 12(3), 553–564. <https://doi.org/10.1111/2041-210X.13536>
- Kirby, S. (2017). Culture and biology in the origins of linguistic structure. *Psychonomic Bulletin & Review*, 24(1), 118–137. <https://doi.org/10.3758/s13423-016-1166-7>
- Kirby, S., Cornish, H., & Smith, K. (2008). Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proceedings of the National Academy of Sciences*, 105(31), 10681–10686. <https://doi.org/10.1073/pnas.0707835105>
- Kirby, S., Griffiths, T., & Smith, K. (2014). Iterated learning and the evolution of language. *Current Opinion in Neurobiology*, 28, 108–114. <https://doi.org/10.1016/j.conb.2014.07.014>
- Kirby, S., Tamariz, M., Cornish, H., & Smith, K. (2015). Compression and communication in the cultural evolution of linguistic structure. *Cognition*, 141, 87–102. <https://doi.org/10.1016/j.cognition.2015.03.016>
- Koshevoy, A., Miton, H., & Morin, O. (2023). Zipf's law of abbreviation holds for individual characters across a broad range of writing systems. *Cognition*, 238, 105527. <https://doi.org/10.1016/j.cognition.2023.105527>
- Lachlan, R. F., Ratmann, O., & Nowicki, S. (2018). Cultural conformity generates extremely stable traditions in bird song. *Nature Communications*, 9. <https://doi.org/10.1038/s41467-018-04728-1>
- Lahti, D. C. (2020). [Personal communication].
- Langfelder, P., Zhang, B., & Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: the Dynamic Tree Cut package for R. *Bioinformatics*, 24(5), 719–720. <https://doi.org/10.1093/bioinformatics/btm563>
- Leonardo, A., & Konishi, M. (1999). Decrystallization of adult birdsong by perturbation of auditory feedback. *Nature*, 399(6735), 466–470. <https://doi.org/10.1038/20933>
- Li, W. (1990). Mutual information functions versus correlation functions. *Journal of Statistical Physics*, 60(5–6), 823–837. <https://doi.org/10.1007/BF01025996>
- Lin, H., & Tegmark, M. (2017). Critical behavior in physics and probabilistic formal languages. *Entropy*, 19(7), 299. <https://doi.org/10.3390/e19070299>
- Linders, G. M., & Louwerse, M. M. (2023). Zipf's law revisited: Spoken dialog, linguistic units, parameters, and the principle of least effort. *Psychonomic Bulletin & Review*, 30(1), 77–101. <https://doi.org/10.3758/s13423-022-02142-9>
- Liu, S., Lu, Y., & Geng, D. (2022). Molecular Subgroup Classification in Alzheimer's Disease by Transcriptomic Profiles. *Journal of Molecular Neuroscience*, 72(4), 866–879. <https://doi.org/10.1007/s12031-021-01957-w>
- Mačutek, J. (2022). Why do parameter values in the Zipf-Mandelbrot distribution sometimes explode? *Journal of Quantitative Linguistics*, 29(4), 413–424. <https://doi.org/10.1080/09296174.2021.1887613>
- Mandelbrot, B. (1953). An informational theory of the statistical structure of language. *Communication Theory*, 486–502.
- Mandelbrot, B. (1962). *On the theory of word frequencies and on related Markovian models of discourse*. 190–219.
- Manin, D. Yu. (2009). Mandelbrot's model for Zipf's law: can Mandelbrot's model explain Zipf's law for language? *Journal of Quantitative Linguistics*, 16(3), 274–285. <https://doi.org/10.1080/09296170902850358>

- Mann, D. C., Lahti, D. C., Waddick, L., & Mundinger, P. C. (2020). House finches learn canary trills. *Bioacoustics*, 1–17. <https://doi.org/10.1080/09524622.2020.1718551>
- Martins, E. P. (1994). Structural complexity in a lizard communication system: the *Sceloporus graciosus* “push-up” display. *Copeia*, 1994(4), 944. <https://doi.org/10.2307/1446717>
- McCowan, B., Hanser, S. F., & Doyle, L. R. (1999). Quantitative tools for comparing animal communication systems: information theory applied to bottlenose dolphin whistle repertoires. *Animal Behaviour*, 57(2), 409–419. <https://doi.org/10.1006/anbe.1998.1000>
- Mehri, A., & Jamaati, M. (2021). Statistical metrics for languages classification: a case study of the Bible translations. *Chaos, Solitons & Fractals*, 144, 110679. <https://doi.org/10.1016/j.chaos.2021.110679>
- Melnik, S. S., Usatenko, O. V., Yampol’skii, V. A., & Golick, V. A. (2005). Competition between two kinds of correlations in literary texts. *Physical Review E*, 72(2), 026140. <https://doi.org/10.1103/PhysRevE.72.026140>
- Mennill, D. J., Badyaev, A. V., Jonart, L. M., & Hill, G. E. (2006). Male house finches with elaborate songs have higher reproductive performance. *Ethology*, 112(2), 174–180. <https://doi.org/10.1111/j.1439-0310.2006.01145.x>
- Menzerath, P. (1954). *Die Architektur des Deutschen Wortschatzes*. Dümmler.
- Merino Recalde, N. (2023). pykanto: A python library to accelerate research on wild bird song. *Methods in Ecology and Evolution*, 14(8), 1994–2002. <https://doi.org/10.1111/2041-210X.14155>
- Mesoudi, A. (2011). Variable Cultural Acquisition Costs Constrain Cumulative Cultural Evolution. *PLOS ONE*, 6(3), e18239. <https://doi.org/10.1371/journal.pone.0018239>
- Mikula, P., Petrusková, T., & Albrecht, T. (2018). Song complexity—no correlation between standard deviation of frequency and traditionally used song complexity metrics in passerines: A comment on Pearce et al. (2018). *Evolution*, 72(12), 2832–2835. <https://doi.org/10.1111/evo.13634>
- Miller, G. A. (1957). Some effects of intermittent silence. *The American Journal of Psychology*, 70(2), 311–314. JSTOR. <https://doi.org/10.2307/1419346>
- Miton, H., & Morin, O. (2021). Graphic complexity in writing systems. *Cognition*, 214, 104771. <https://doi.org/10.1016/j.cognition.2021.104771>
- Miyagawa, S., Berwick, R. C., & Okanoya, K. (2013). The emergence of hierarchical structure in human language. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00071>
- Morita, T., Koda, H., Okanoya, K., & Tachibana, R. O. (2021). Measuring context dependency in birdsong using artificial neural networks. *PLOS Computational Biology*, 17(12), e1009707. <https://doi.org/10.1371/journal.pcbi.1009707>
- Motamedi, Y., Schouwstra, M., Smith, K., Culbertson, J., & Kirby, S. (2019). Evolving artificial sign languages in the lab: From improvised gesture to systematic sign. *Cognition*, 192, 103964. <https://doi.org/10.1016/j.cognition.2019.05.001>
- Motter, A. E., De Moura, A. P. S., Lai, Y.-C., & Dasgupta, P. (2002). Topology of the conceptual network of language. *Physical Review E*, 65(6), 065102. <https://doi.org/10.1103/PhysRevE.65.065102>
- Mouillot, D., & Lepretre, A. (2000). Introduction of relative abundance distribution (RAD) indices, estimated from the rank-frequency diagrams (RFD), to assess changes in community diversity. *Environmental Monitoring and Assessment*, 63, 279–295. <https://doi.org/10.1023/A:1006297211561>
- Muldoon, S. F., Bridgeford, E. W., & Bassett, D. S. (2016). Small-world propensity and weighted brain networks. *Scientific Reports*, 6(1), 22057. <https://doi.org/10.1038/srep22057>
- Müllner, D. (2013). fastcluster: Fast hierarchical, agglomerative clustering routines for R and python. *Journal of Statistical Software*, 53(9), 1–18. <https://doi.org/10.18637/jss.v053.i09>
- Mundinger, P. C. (1975). Song dialects and colonization in the house finch, *Carpodacus mexicanus*, on the east coast. *The Condor*, 77(4), 407–422. <https://doi.org/10.2307/1366088>
- Nolan, P. M., & Hill, G. E. (2004). Female choice for song characteristics in the house finch. *Animal Behaviour*, 67(3), 403–410. <https://doi.org/10.1016/j.anbehav.2003.03.018>
- Nölle, J., Staib, M., Fusaroli, R., & Tylén, K. (2018). The emergence of systematicity: How environmental and communicative factors shape a novel communication system. *Cognition*, 181, 93–104. <https://doi.org/10.1016/j.cognition.2018.08.014>
- Petrini, S., Casas-i-Muñoz, A., Cluet-i-Martinell, J., Wang, M., Bentz, C., & Ferrer-i-Cancho, R. (2023). The optimality of word lengths. *Theoretical foundations and an empirical study* (arXiv:2208.10384). arXiv. <https://doi.org/10.48550/arXiv.2208.10384>
- Piantadosi, S. T. (2014). Zipf’s word frequency law in natural language: A critical review and future directions. *Psychonomic Bulletin & Review*, 21(5), 1112–1130. <https://doi.org/10.3758/s13423-014-0585-6>
- Podós, J., Moseley, D. L., Goodwin, S. E., McClure, J., Taft, B. N., Strauss, A. V. H., Rega-Brodsky, C., & Lahti, D. C. (2016). A fine-scale, broadly applicable index of vocal performance: frequency excursion. *Animal Behaviour*, 116, 203–212. <https://doi.org/10.1016/j.anbehav.2016.03.036>
- Pothos, E. M., & Juola, P. (2007). Characterizing linguistic structure with mutual information. *British Journal of Psychology*, 98(2), 291–304. <https://doi.org/10.1348/000712606X1122760>
- Ratanamahatana, C. A., & Keogh, E. (2005). Three myths about dynamic time warping data mining. *Proceedings of the 2005 SIAM International Conference on Data Mining*, 506–510. <https://doi.org/10.1137/1.9781611972757.50>
- Raviv, L., De Heer Kloots, M., & Meyer, A. (2021). What makes a language easy to learn? A preregistered study on how systematic structure and community size affect language learnability. *Cognition*, 210, 104620. <https://doi.org/10.1016/j.cognition.2021.104620>
- Raviv, L., Meyer, A., & Lev-Ari, S. (2019). Compositional structure can emerge without generational transmission. *Cognition*, 182, 151–164. <https://doi.org/10.1016/j.cognition.2018.09.010>
- Rivera, M., Edwards, J. A., Hauber, M. E., & Woolley, S. M. N. (2023). Machine learning and statistical classification of birdsong link vocal acoustic features with phylogeny. *Scientific Reports*, 13(1), 7076. <https://doi.org/10.1038/s41598-023-33825-5>
- Roberts, G., Lewandowski, J., & Galantucci, B. (2015). How communication changes when we cannot mime the world: experimental evidence for the effect of iconicity on combinatoriality. *Cognition*, 141, 52–66. <https://doi.org/10.1016/j.cognition.2015.04.001>
- Roginek, E. W. (2018). *Spatial variation of house finch (Haemorhous mexicanus) song along the American Southwest coast*. Queens College.
- Safryghin, A., Cross, C., Fallon, B., Heesen, R., Ferrer-i-Cancho, R., & Hobaiter, C. (2022). Variable expression of linguistic laws in ape gesture: a case study from chimpanzee sexual solicitation. *Royal Society Open Science*, 9, 220849. <https://doi.org/10.1098/rsos.220849>
- Sainburg, T., Mai, A., & Gentner, T. Q. (2022). Long-range sequential dependencies precede complex syntactic production in language acquisition. *Proceedings of the Royal Society B: Biological Sciences*, 289, 20212657. <https://doi.org/10.1098/rspb.2021.2657>
- Sainburg, T., Theilmann, B., Thielk, M., & Gentner, T. Q. (2019). Parallels in the sequential organization of birdsong and human speech. *Nature Communications*, 10(1), 3636. <https://doi.org/10.1038/s41467-019-11605-y>
- Sainburg, T., Thielk, M., & Gentner, T. Q. (2020). Finding, visualizing, and quantifying latent structure across diverse animal vocal repertoires. *PLOS Computational Biology*, 16(10), e1008228. <https://doi.org/10.1371/journal.pcbi.1008228>
- Salge, C., Ay, N., Polani, D., & Prokopenko, M. (2015). Zipf’s law: balancing signal usage cost and communication efficiency. *PLOS ONE*, 10(10), e0139475. <https://doi.org/10.1371/journal.pone.0139475>
- Sardá-Espinosa, A. (2019). Time-series clustering in R using the dtwclust package. *The R Journal*, 11(1), 22–43. <https://doi.org/10.32614/RJ-2019-023>



- Sasahara, K., Cody, M. L., Cohen, D., & Taylor, C. E. (2012). Structural design principles of complex bird songs: a network-based approach. *PLoS ONE*, 7(9), e44436. <https://doi.org/10.1371/journal.pone.0044436>
- Schrimpf, M., Blank, I. A., Tuckwell, G., Kauf, C., Hosseini, E. A., Kanwisher, N., Tenenbaum, J. B., & Fedorenko, E. (2021). The neural architecture of language: Integrative modeling converges on predictive processing. *Proceedings of the National Academy of Sciences*, 118(45), e2105646118. <https://doi.org/10.1073/pnas.2105646118>
- Searcy, W. A., Soha, J., Peters, S., & Nowicki, S. (2022). Long-distance dependencies in birdsong syntax. *Proceedings of the Royal Society B: Biological Sciences*, 289(1967), 20212473. <https://doi.org/10.1098/rspb.2021.2473>
- Semple, S., Ferrer-i-Cancho, R., & Gustison, M. L. (2022). Linguistic laws in biology. *Trends in Ecology & Evolution*, 37(1), 53–66. <https://doi.org/10.1016/j.tree.2021.08.012>
- Semple, S., Hsu, M. J., & Agoramoorthy, G. (2010). Efficiency of coding in macaque vocal communication. *Biology Letters*, 6(4), 469–471. <https://doi.org/10.1098/rsbl.2009.1062>
- Shlegeris, B., Roger, F., Chan, L., & McLean, E. (2022). *Language models are better than humans at next-token prediction* (arXiv:2212.11281). arXiv. <http://arxiv.org/abs/2212.11281>
- Simon, H. A. (1955). On a class of skew distribution functions. *Biometrika*, 42(3/4), 425–440.
- Sinclair, N. C., Ursell, J., South, A., & Rendell, L. (2022). From Beethoven to Beyoncé: do changing aesthetic cultures amount to “cumulative cultural evolution?” *Frontiers in Psychology*, 12, 663397. <https://doi.org/10.3389/fpsyg.2021.663397>
- Solé, R. V., Corominas-Murtra, B., Valverde, S., & Steels, L. (2010). Language networks: their structure, function, and evolution. *Complexity*, NA–NA. <https://doi.org/10.1002/cplx.20305>
- Soma, M., & Garamszegi, L. Z. (2011). Rethinking birdsong evolution: meta-analysis of the relationship between song complexity and reproductive success. *Behavioral Ecology*, 22(2), 363–371. <https://doi.org/10.1093/beheco/arq219>
- Stanley, H. E., Amaral, L. A. N., Gopikrishnan, P., Ivanov, P. C., Keitt, T. H., & Plerou, V. (2000). Scale invariance and universality: organizing principles in complex systems. *Physica A: Statistical Mechanics and Its Applications*, 281(1–4), 60–68. [https://doi.org/10.1016/S0378-4371\(00\)00195-3](https://doi.org/10.1016/S0378-4371(00)00195-3)
- Stave, M., Paschen, L., Pellegrino, F., & Seifart, F. (2021). Optimization of morpheme length: a cross-linguistic assessment of Zipf's and Menzerath's laws. *Linguistics Vanguard*, 7(s3), 20190076. <https://doi.org/10.1515/lingvan-2019-0076>
- Stepanov, A., Zhivomirov, H., Nedelchev, I., & Stateva, P. (2023). *Bottlenose dolphins' broadband clicks are structured for communication* [Preprint]. *Animal Behavior and Cognition*. <https://doi.org/10.1101/2023.01.11.523588>
- Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1), 41–78. [https://doi.org/10.1207/s15516709cog2901\\_3](https://doi.org/10.1207/s15516709cog2901_3)
- Tamariz, M., & Kirby, S. (2016). The cultural evolution of language. *Current Opinion in Psychology*, 8, 37–43. <https://doi.org/10.1016/j.copsyc.2015.09.003>
- Tanaka-Ishii, K. (2021). Menzerath's law in the syntax of languages compared with random sentences. *Entropy*, 23(6), 661. <https://doi.org/10.3390/e23060661>
- Tchernichovski, O., Eisenberg-Edidin, S., & Jarvis, E. D. (2021). Balanced imitation sustains song culture in zebra finches. *Nature Communications*, 12(1), 1–14. <https://doi.org/10.1038/s41467-021-22852-3>
- Tomasello, M. (1999). *The Cultural Origins of Human Cognition*. Harvard University Press.
- Treves, A., & Panzeri, S. (1995). The upward bias in measures of information derived from limited data samples. *Neural Computation*, 7(2), 399–407. <https://doi.org/10.1162/neco.1995.7.2.399>
- Verhoef, T. (2012). The origins of duality of patterning in artificial whistled languages. *Language and Cognition*, 4(4), 357–380. <https://doi.org/10.1515/langcog-2012-0019>
- Verhoef, T., Kirby, S., & de Boer, B. (2014). Emergence of combinatorial structure and economy through iterated learning with continuous acoustic signals. *Journal of Phonetics*, 43, 57–68. <https://doi.org/10.1016/j.wocn.2014.02.005>
- Verhoef, T., Kirby, S., & de Boer, B. (2016). Iconicity and the emergence of combinatorial structure in language. *Cognitive Science*, 40(8), 1969–1994. <https://doi.org/10.1111/cogs.12326>
- Watanabe, S. (2021). Information criteria and cross validation for Bayesian inference in regular and singular cases. *Japanese Journal of Statistics and Data Science*, 4(1), 1–19. <https://doi.org/10.1007/s42081-021-00121-3>
- Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of “small-world” networks. *Nature*, 393, 440–442. <https://doi.org/10.1038/30918>
- Weiss, M., Hultsch, H., Adam, I., Scharff, C., & Kipper, S. (2014). The use of network analysis to study complex animal communication systems: a study on nightingale song. *Proceedings of the Royal Society B: Biological Sciences*, 281(1785), 20140460. <https://doi.org/10.1098/rspb.2014.0460>
- Wilks, C. E. H., & Blakey, K. H. (2018). In the jungle of cultural complexity. *Evolutionary Anthropology: Issues, News, and Reviews*, 27(5), 180–183. <https://doi.org/10.1002/evan.21724>
- Williams, H., Scharf, A., Ryba, A. R., Ryan Norris, D., Mennill, D. J., Newman, A. E. M., Doucet, S. M., & Blackwood, J. C. (2022). Cumulative cultural evolution and mechanisms for cultural selection in wild bird songs. *Nature Communications*, 13(1), 4001. <https://doi.org/10.1038/s41467-022-31621-9>
- Youngblood, M., & Lahti, D. (2022). Content bias in the cultural evolution of house finch song. *Animal Behaviour*, 185, 37–48. <https://doi.org/10.1016/j.anbehav.2021.12.012>
- Youngblood, M., Miton, H., & Morin, O. (2023). Statistical signals of copying are robust to time- and space-averaging. *Evolutionary Human Sciences*, 5, e10. <https://doi.org/10.1017/ehs.2023.5>
- Zhao, H., Zhang, S., Shao, S., & Fang, H. (2020). Identification of a Prognostic 3-Genes Risk Prediction Model for Thyroid Cancer. *Frontiers in Endocrinology*, 11. <https://www.frontiersin.org/articles/10.3389/fendo.2020.00510>
- Zipf, G. (1949). *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Addison-Wesley.

## 4 Supplementary Information

### 4.1 Zipf's Rank-Frequency Law

Table 4.1: Prior specification for the Zipf and Zipf-Mandelbrot models fit across the three levels of granularity.

Parameter	Class	Prior	Lower Bound
a	b	normal(0, 10)	1
b	b	normal(0, 10)	-1
c	b	normal(0, 10)	0
	sigma	student_t(3, 0, 2.5)	0

Table 4.2: WAIC values from the Zipf and Zipf-Mandelbrot models fit across the three levels of granularity.

DS	Zipf	Zipf-Mandelbrot
1	-799	-1,075
2	-5,328	-7,817
3	-17,566	-25,617

#### 4.1.1 Deep Split: 2

Table 4.3: Estimates and diagnostics for the Zipf model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
a	1.01	1.00	1.03	1	5,510	3,580
c	0.23	0.22	0.24	1	7,223	6,440

Table 4.4: Estimates and diagnostics for the Zipf-Mandelbrot model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
a	2.20	2.03	2.39	1	2,129	2,702
b	4.66	4.00	5.41	1	2,132	2,653
c	9.43	5.22	16.68	1	2,102	2,515

#### 4.1.2 Deep Split: 3

Table 4.5: Estimates and diagnostics for the Zipf model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
a	1.00	1.00	1.00	1	6,264	4,253
c	0.08	0.08	0.09	1	7,540	5,954



Table 4.6: Estimates and diagnostics for the Zipf-Mandelbrot model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
a	1.20	1.18	1.22	1	2,132	1,965
b	5.69	5.42	5.98	1	2,187	2,477
c	0.52	0.48	0.56	1	2,094	2,123

### 4.1.3 Deep Split: 4

Table 4.7: Estimates and diagnostics for the Zipf model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
a	1.00	1.00	1.00	1	6,266	4,117
c	0.03	0.03	0.03	1	8,434	7,188

Table 4.8: Estimates and diagnostics for the Zipf-Mandelbrot model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
a	1.08	1.07	1.09	1	2,079	2,879
b	17.29	16.79	17.79	1	2,076	3,021
c	0.34	0.32	0.36	1	2,032	2,774

## 4.2 Zipf's Law of Abbreviation

Table 4.9: Prior specification for all four models of Zipf's law of abbreviation.

Class	Prior	Lower Bound
b	normal(0, 0.5)	
Intercept	normal(0, 4)	0
sd	normal(0, 0.5)	0
sigma	normal(0, 0.5)	0

Table 4.10: The model diagnostics for each model of Zipf's law of abbreviation.

Outcome	DS	Predictor	Rhat	Bulk ESS	Tail ESS
Duration	2	Intercept	1	712	913
		Count	1	794	1,064
	3	Intercept	1	451	915
		Count	1	566	1,069
	4	Intercept	1	268	580
		Count	1	302	597
Bandwidth	2	Intercept	1	835	1,801
		Count	1	1,047	2,088

Outcome	DS	Predictor	Rhat	Bulk ESS	Tail ESS
Concavity	3	Intercept	1	184	448
		Count	1	230	586
	4	Intercept	1	170	406
		Count	1	230	525
	2	Intercept	1	1,059	2,059
		Count	1	1,354	2,695
	3	Intercept	1	1,307	2,561
		Count	1	1,546	3,209
Excursion	4	Intercept	1	723	1,469
		Count	1	775	1,620
	2	Intercept	1	1,301	2,456
		Count	1	1,508	2,661
	3	Intercept	1	735	1,398
		Count	1	855	1,592
	4	Intercept	1	403	771
		Count	1	479	956

### 4.3 Menzerath's Law

Table 4.11: Prior specification for the model of Menzerath's law.

Class	Prior	Lower Bound
b	normal(0, 0.1)	
Intercept	normal(0, 3)	0
sd	normal(0, 0.5)	0
sigma	normal(0, 0.5)	0

Table 4.12: Estimates and diagnostics for the model of Menzerath's law applied to the real data.

Param.	Estimate	1-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	4.48	4.47	4.49	1	6,740	7,103
Song Length	-0.05	-0.07	-0.04	1	4,794	6,044

Table 4.13: The R-hat values from the two null models applied to each of the 10 simulated datasets.

Dataset	Simple Null Model		Production Null Model	
	Intercept	Song Length	Intercept	Song Length
1	0.9998875	0.9996983	0.9997456	1.000656
2	0.9997542	0.9997727	0.9999726	1.000328
3	1.0000627	0.9997787	0.9998281	1.001598
4	0.9997391	0.9997015	1.0005752	1.000895

	Simple Null Model		Production Null Model	
Dataset	Intercept	Song Length	Intercept	Song Length
5	0.9999400	1.0004971	0.9997567	1.000817
6	0.9999405	0.9998756	0.9997009	1.000332
7	0.9999780	0.9998101	0.9997015	1.001533
8	1.0005872	0.9998251	0.9998900	1.000225
9	0.9999938	0.9997752	0.9998704	1.001099
10	0.9997444	0.9998224	1.0001865	1.002680

## 4.4 Mutual Information

Table 4.14: Prior specification for all three models of mutual information decay.

Parameter	Prior	Lower Bound
a	normal(0, 1)	0
b	normal(0, 1)	0
c	normal(0, 1)	0
d	normal(0, 1)	0

Table 4.15: The WAIC values for each model, at each level of deep split, for increasing maximum distances between syllables.

	2			3			4		
	Exp	PL	Comp	Exp	PL	Comp	Exp	PL	Comp
100	-829	-765	-841	-617	-486	-668	-598	-432	-688
200	-1,557	-1,513	-1,560	-1,232	-1,068	-1,281	-1,201	-967	-1,267
300	-2,101	-2,086	-2,114	-1,635	-1,555	-1,695	-1,626	-1,468	-1,686
400	-2,683	-2,676	-2,712	-2,008	-1,977	-2,131	-2,082	-1,950	-2,142
500	-2,984	-2,988	-3,073	-2,078	-2,093	-2,276	-2,260	-2,217	-2,387
600	-3,114	-3,176	-3,248	-2,264	-2,299	-2,527	-2,454	-2,449	-2,639
700	-3,291	-3,430	-3,471	-2,493	-2,539	-2,808	-2,650	-2,669	-2,901
800	-3,620	-3,734	-3,766	-2,479	-2,598	-2,798	-2,209	-2,240	-2,485
900	-3,615	-3,785	-3,801	-2,108	-2,379	-2,467	-1,568	-1,804	-1,886
1000	-3,712	-3,931	-3,942	-1,958	-2,350	-2,408	-1,235	-1,608	-1,654
1200	-3,175	-3,244	-3,246	-1,483	-1,711	-1,731	-823	-1,084	-1,104

### 4.4.1 Deep Split: 2

Table 4.16: Estimates and diagnostics for exponential model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
a	0.17	0.16	0.19	1	4,336	4,595
b	0.39	0.35	0.44	1	4,511	4,996

Table 4.17: Estimates and diagnostics for power-law model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
c	0.13	0.12	0.14	1	6,738	7,059
d	1.08	1.01	1.15	1	6,437	6,285

Table 4.18: Estimates and diagnostics for composite model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
a	0.16	0.12	0.18	1	3,354	3,685
b	0.43	0.38	0.48	1	4,592	4,387
c	0.02	0.00	0.04	1	3,021	2,579
d	0.66	0.29	0.96	1	3,237	2,798

#### 4.4.2 Deep Split: 3

Table 4.19: Estimates and diagnostics for exponential model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
a	0.60	0.57	0.64	1	4,478	5,210
b	0.27	0.26	0.29	1	4,453	5,343

Table 4.20: Estimates and diagnostics for power-law model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
c	0.52	0.49	0.56	1	6,895	5,998
d	0.95	0.90	1.00	1	7,133	7,055

Table 4.21: Estimates and diagnostics for composite model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
a	0.53	0.47	0.59	1	2,820	3,907
b	0.30	0.28	0.32	1	4,896	4,646
c	0.08	0.04	0.13	1	2,882	2,864
d	0.61	0.42	0.77	1	2,978	3,096

#### 4.4.3 Deep Split: 4

Table 4.22: Estimates and diagnostics for exponential model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
a	0.74	0.71	0.77	1	3,787	4,572
b	0.24	0.23	0.25	1	4,287	5,362

Table 4.23: Estimates and diagnostics for power-law model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
c	0.69	0.64	0.73	1	6,835	6,785
d	0.92	0.87	0.97	1	6,638	6,848

Table 4.24: Estimates and diagnostics for composite model.

Param.	Estimate	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
a	0.65	0.59	0.71	1	2,387	2,851
b	0.26	0.25	0.27	1	4,566	4,959
c	0.10	0.05	0.15	1	2,433	2,760
d	0.61	0.45	0.74	1	2,360	2,726