

The Applications and Limitations of the Burstiness Metric in Investigating the Temporal Distribution of Words in Child-Centered Audio

Zeynep Marasli (marasli2@illinois.edu)

Department of Psychology, 603 E Daniel St
Champaign, IL 61801 USA

Jessica L. Montag (jmontag@illinois.edu)

Department of Psychology, 603 E Daniel St
Champaign, IL 61801 USA

Abstract

As the use of naturalistic speech data of children's language experiences increases, the temporal dynamics of the speech environment becomes a more obvious aspect of speech to investigate. To connect the temporal dynamics in child-centered speech to existing experimental work showing that the temporal presentation of items has measurable effects on learning, it is important to develop measures that quantify temporal patterns in speech. The present work explores one such measure, the burstiness metric, and investigates word burstiness and its relationship with frequency, its behavior across different timescales, and whether it can be used to quantify constructs of massed and spaced orders. Our findings suggest that while related, word burstiness is not an index of frequency, that burstiness varies across both words of different lexical classes and timescales, and that it does not appropriately capture massed and spaced temporal patterns. We discuss implications for how this measure may be used for child-centered audio.

Keywords: language development; daylong audio; corpus analysis

Introduction

Language unfolds across time; children do not experience language in aggregate, but rather, there is considerable moment-to-moment variability in the speech they hear. This fact may be important for linking the language environment to language learning.

Classic and more recent experimental work with children and adults show that the temporal distribution of training (i.e., massed/spaced presentation order) affects the kinds of information that are attended to and encoded (Estes, 1955; Vlach et al., 2011; Vlach & Johnson, 2013; Carvalho & Goldstone, 2015; 2017; Benitez et al., 2020). For example, in category learning, spaced/ interleaved presentations allow for learners to attend more to and better encode for cross-category properties, and blocked/ massed presentations allow for learners to better attend to and encode for within-category properties (Carvalho & Goldstone, 2017). Experimental work has suggested that spaced presentations allow for better learning and generalization of novel words for children as young as 20-months, while younger infants perform better in

massed presentations (Vlach et al., 2008; Vlach et al., 2011; Vlach & Johnson, 2013), suggesting that different temporal presentations are more/less salient at various developmental stages. A separate but related line of experimental work also shows that the visual, physical and linguistic contexts in which words occur also has implications on word learning and generalization (Samuelson & Horst, 2007; Vlach & Sandhofer, 2011; Goldenberg & Sandhofer, 2013).

In naturalistic lines of work, the temporal dynamics of language input is an emerging feature of interest. Observational and corpus studies demonstrate how different everyday conversational contexts in children's language environments contain measurable lexical and syntactic variability (Snow & Beals, 2006; Montag et al., 2015; Tamis-LeMonda & Custode, 2018), and this variability has been shown to predict numerous outcome measures (Weizman & Snow, 2001; Snow & Beals, 2006; Mol et al., 2008).

Due to this context-to-context variability, the distribution of lexical items in children's language environments are not random, and instead exhibit a dynamic temporal structure in which auditory and linguistic input has a nested, non-uniform organization (Abney et al., 2017; Mendoza & Fausey, 2021; Warlaumont et al., 2022; Karmazyn-Raz & Smith, 2022). Visualizations of when individual words occur across the day collected from child-centered daylong audio recordings show that the temporal distribution of words can occur in "bursts", or short clusters of repetition and longer intervals of no occurrences, and that this variation in temporal distribution appears across both items and individuals (Montag, 2020). Analyses of naturalistic spoken and written language corpora and of non-verbal communicative behaviors also illustrate the bursty nature of speech (e.g., Altmann et al., 2009; Abney et al., 2017; Karmazyn-Raz & Smith, 2022).

Despite the evidence suggested by laboratory-based investigations of the temporal distribution of training, the temporal distribution of children's daily language input has not been well-linked to trajectories of learning. This is in part due to (1) a lack of time-stamped transcriptions of spontaneous child-centered speech, and (2) challenges associated with operationalizing temporal distributions in

naturalistic language samples that has led to largely approaching the temporal dimension of speech in a descriptive, qualitative manner and less so quantitatively (deBarbaro & Fausey, 2022; Clerkin & Smith, 2022; Karmazyn-Raz & Smith, 2022). In recent years, however, researchers in the domain have adopted the burstiness metric, which was first introduced out-of-domain by Goh and Barabási (2008).

The burstiness metric is used in complex time series analyses to quantify the temporal pattern of events and has been used in a wide range of applications such as for communication networks, gene expression patterns, and seismic activity (So et al., 2011; Jo et al., 2012; Salditch et al., 2020). The burstiness metric, B , characterizes a temporal pattern from periodic ($B = -1$) to random/Poisson ($B = 0$) to theoretically maximally bursty ($B = 1$). The burstiness metric is calculated using the distribution of interevent intervals, i.e., the time between two consecutive occurrences of an event. Within the word learning domain, Slone and colleagues (2023), using 30 parent-toddler dyads, recorded parents' speech about novel objects during a short play session, and assessed the toddlers' learning of the objects in a subsequent object-naming task. Toddlers were better able to learn the names of objects that parents had talked about in a bursty manner than those that were not. These results demonstrate that the burstiness metric, as a measure of the temporal distribution of training events, appears to be capturing meaningful variation in individual words' temporal patterns in child-centered speech.

With the growing interest in naturalistic, long-form temporal data within the field, researchers' goals are to understand the structures that underlie their data and ensure that they are equipped with ecologically valid measures. The present work aims to expand upon our understanding of this burstiness metric as it applies to child-centered speech. The goal of this work is two-fold: first, to better quantify the temporal pattern of words in spontaneous speech; and second, to better understand the measure itself and to shed insights on the potential strengths and limitations of the metric when applied to child-centered speech data. Using a dataset of 5 transcribed child-centered daylong audio recordings, we present a series of analyses exploring the qualities of word burstiness. We first investigate the relationship between a word's burstiness and its frequency. We then compare how burstiness behaves in longform (daylong audio recordings) and shortform speech data (CHILDES) within and across different lexical classes. We then use word burstiness to predict vocabulary norms. Finally, we assess how well bursty patterns may be capturing massed/spaced presentation orders and contextually bound words. We do not aim to make conclusions about how word burstiness may play a role in learning, but rather we intend this work to serve as a fundamental first step in understanding this theoretically important, but nonetheless relatively underexplored dimension of speech.

Methods

Transcripts of Child-Centered Recordings

Analyses were conducted using 5 transcribed daylong audio recordings (Montag et al., in prep). The recordings were collected with LENA recorders within English-speaking homes of two children 10-12 months of age on separate days. All recordings were manually transcribed using ACLEW conventions (Soderstrom et al., 2021). All utterances were diarized (assigned a speaker), segmented, such that each transcribed utterance is annotated with a start and end timestamp, and the content of the utterance was transcribed. The recordings have a mean audio length of 11.9 hours ($SD = 0.56$ hours). The recordings were collected within the home only, so some recordings contain large periods where no speech was transcribed (indicating either naps, or intervals where the child left the home); intervals of silence greater than 30 minutes were removed prior to analysis. The mean length of speech captured across the recordings is 7.65 hours ($SD = 3.5$ hours).

As a comparison to the daylong audio recordings, we also analyzed transcripts from the CHILDES North American English corpus (MacWhinney, 2000). The raw transcript files were originally downloaded from the website on 6/12/24. We selected files that contained speech available to 24-month-old target children. Transcripts that had fully annotated utterance timestamps and that were less than 30 minutes were considered. 193 total transcripts from CHILDES were included in the presented analyses. These transcripts are of recordings with a mean audio length of 13.9 minutes ($SD = 5.49$ minutes), representing a very different audio length compared to the daylong recordings.

Calculating Burstiness Values

All speech regardless of whether it was child-directed or not was included in the analyses, and all word forms were preserved as is and were not lemmatized. Within each transcript, burstiness values were calculated for each unique word by first identifying the timestamps of all utterances in which a given word occurred. Interevent intervals (IEIs) were generated by calculating the difference in time in milliseconds for two successive occurrences. If a word occurred more than once in an utterance, the resulting interevent time was determined to be 0. The mean and standard deviation for each word's interevent interval distribution was used to calculate its burstiness value, B , using the metric defined from Kim & Jo (2016), a modified version of the original burstiness metric which takes into account finite size effects (see Panel A of Figure 1). Since burstiness values were calculated for each unique word within in each transcript, words that occurred in one or more transcripts had multiple burstiness values (i.e., "you" occurs across all 5 daylong transcripts and has 5 burstiness values in the daylong transcript sample) that were not averaged across in the presented analyses unless specified otherwise. Words with a frequency of less than 2 were not included in analyses

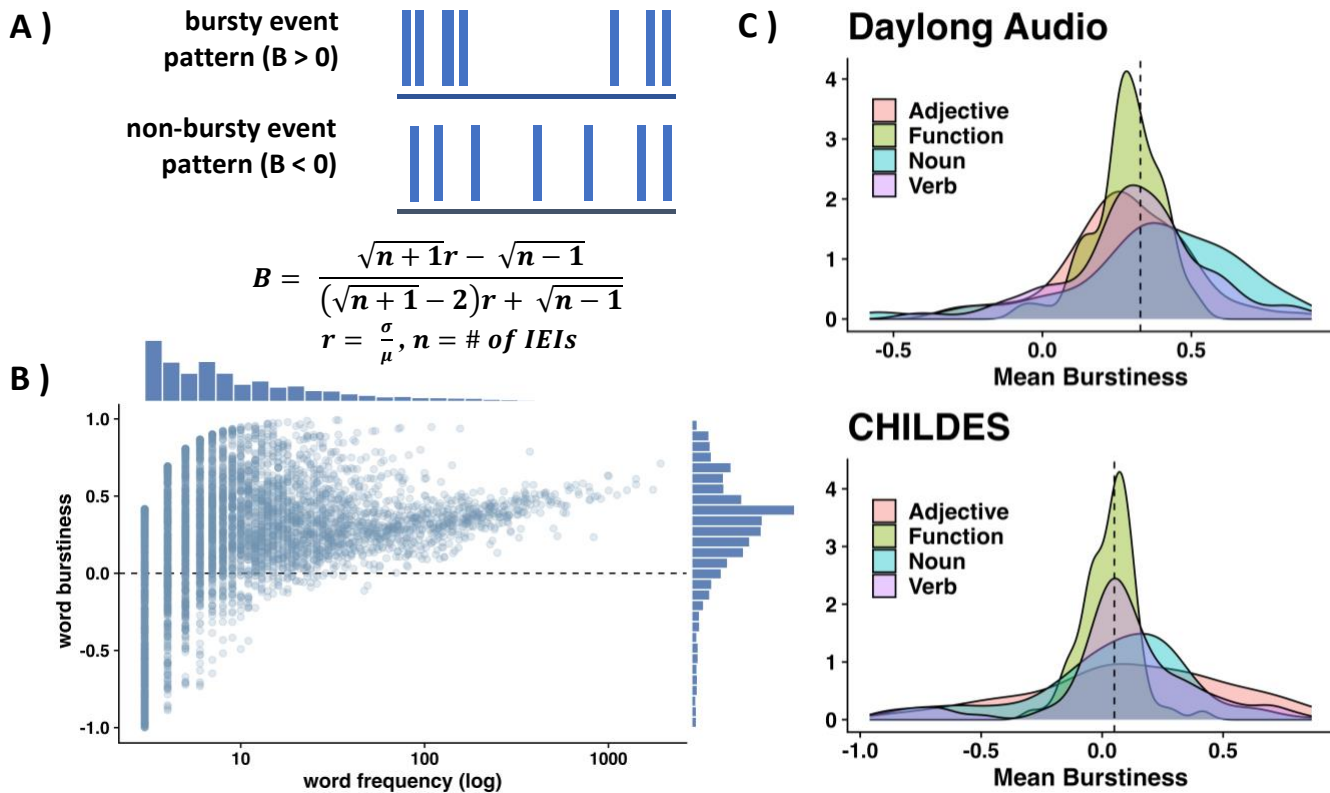


Figure 1: *Panel A* The burstiness metric as defined by Kim & Jo (2016) and illustrations of bursty and non-bursty event patterns. *Panel B* Scatter plot with marginal histograms of word frequency and burstiness of all words in the daylong audio transcripts. *Panel C* Distribution of average burstiness and frequency values of CDI word items grouped by lexical class occurring in the daylong audio and CHILDES transcripts; a black dotted line is included at the mean for interpretability.

due to an insufficient number of interevent intervals needed to calculate a burstiness value. In total, burstiness values for 5,154 word tokens were calculated from the daylong transcripts. Analyses were implemented using Python and R; all analysis code will be available upon publication.

Comparing Words Burstiness in Longform and Shortform Audio Data

To compare across longform audio (daylong audio recordings) and shortform audio (CHILDES recordings), burstiness values of word items from the MacArthur-Bates Communicative Development Inventory Words and Sentences (CDI) (Fenson, 2007) were used in these analyses. Homonyms (i.e., “drink”) and multi-word items (i.e., “peanut butter”) were excluded. Words items were then manually coded by the author into one of four lexical categories: nouns, verbs, adjectives, and function words. Burstiness values for each individual item that occurs in each transcript was first calculated and then averaged across all transcripts in their respective longform and shortform audio groups.

Results

We first present analyses investigating burstiness’s relationship to frequency, and then comparisons of burstiness values in longform and shortform data. We then show

regression analyses using word burstiness and frequency to predict for vocabulary norms. Last, we present our assessments on the kinds of event distributions the metric categorizes as bursty and non-bursty, and how well they align with our intuitions of massed and spaced temporal orders.

Word Burstiness and Frequency

Intuitively, burstiness is related to frequency such that higher frequency words may necessarily appear across more different contexts. We first show that burstiness is related to, but distinct from word frequency.

To assess any relationship between a word’s burstiness and frequency, panel B of Figure 1 presents a scatter plot with marginal histograms of individual words’ log frequencies and their respective burstiness values across the daylong audio transcripts. This figure illustrates a few key points about the relationship between a word’s burstiness and its frequency. Burstiness values vary widely within smaller frequency bands. In other words, words that have a low to moderate frequency display a wide range of burstiness values. Comparatively, the highest frequency words tend to be among the highest burstiness values, but not the highest within their respective transcripts. For instance, in one transcript, the most frequent word is “you” with 1,926 occurrences and a burstiness value of 0.71; the most bursty word is “bananas” with a burstiness value of 0.97 and a

frequency of 14. Conversely, non-bursty ($B < 0$) words tend to also be less frequent ($M = 5.1, SD = 4.2$) compared to bursty ($B > 0$) words' frequency ($M = 34.8, SD = 94.8$). These results demonstrate that while there is a relationship between a word's frequency and its burstiness, burstiness is not necessarily an index of frequency. Especially at lower frequencies, burstiness values vary widely for words of similar frequency.

Burstiness at Long and Short Timescales

We next demonstrate how the length of a recording has effects on the calculated burstiness values. We do this to illustrate the profound effects of the burstiness value based on the type of data from which the value is calculated.

The burstiness metric is calculated using the distance between consecutive event occurrences, so it is likely that the length of the timescale across which events are counted affects the distribution of resulting burstiness values. This is particularly important to consider when working with naturalistic speech data, as the length of the examined interval can substantially change the characteristics of a word's event distribution. For example, food-related nouns may appear to be more widely distributed across one interval of mealtime. However, at the daylong timescale, these words may be better characterized as contextually bound or more massed because each recording contains contexts other than mealtimes and multiple discrete instances of mealtimes. Still, it is also possible that certain categories of words, regardless of timescale, tend to be more/less bursty than others, akin to how some words tend to be ubiquitously more frequent (e.g., function words) regardless of data length or size.

To compare burstiness at different timescales, average burstiness for CDI Words & Sentences word items (Fenson, 2007) were obtained from the long timescale (daylong audio) and short timescale (CHILDES) transcripts. Additionally, words were categorized as either a noun, adjective, verb or function word. Overall, the CDI word items in the daylong audio transcripts are more bursty ($M = 0.33, SD = 0.25$) than in the CHILDES transcripts ($M = 0.05, SD = 0.31$). As shown in Panel C of Figure 1, the distribution of burstiness values differs across and within groups. A Pearson correlation analysis indicates that word burstiness between the two corpora is not significant overall, $r(326) = -.02, p = .762$, and weak for any one lexical class except for function words, $r(71) = 0.32, p < .05$. This contrasts with word frequency, which is significantly correlated between the daylong audio and CHILDES transcripts both overall, $r(326) = .83, p < .001$, and for each lexical class. This suggests that across corpora of different interval lengths, burstiness is not necessarily comparable.

Burstiness as a Predictor

An eventual aim of this work is using the burstiness metric to connect the temporal distribution of children's daily speech input to their learning outcomes. To this end, we conducted multiple linear regressions to predict for CDI 24-month-old

vocabulary norms (Frank et al., 2017) as a demonstration that the metric may be capturing meaningful variance in the data.

For each corpus, regressions with word burstiness and frequency as predictors were first conducted for all words regardless of lexical class and then separately for each individual class to discern any group differences. As shown in Table 1, the models across all classes were significant overall, with significant main effects of both word burstiness and word frequency and a significant interaction effect. When using CHILDES transcripts, word burstiness was a significant predictor for function words; when using the daylong audio transcripts, word burstiness was a significant predictor for verbs and significantly interacted with frequency for both verbs and adjectives. These results demonstrate that word burstiness may be a meaningful predictor for word learning, and the variation in significant results within lexical classes across the two corpora reaffirms that burstiness is not comparable across different timescales.

Table 1.

Summary of Multiple Linear Regressions Analyses using CHILDES (above) and Daylong Audio (below)

CHILDES	Burstiness	Frequency	Burstiness x Frequency
All Classes ($R^2 = .068$)	$\beta = .286$ **	$\beta = -.032$ ***	$\beta = .093$ *
Nouns ($R^2 = .053$)	$\beta = .085$	$\beta = .035$ *	$\beta = .015$
Verbs ($R^2 = .055$)	$\beta = .202$	$\beta = -.01$	$\beta = .037$
Adjectives ($R^2 = .045$)	$\beta = -.053$	$\beta = .017$	$\beta = .003$
Functions ($R^2 = .119$)	$\beta = .485$ *	$\beta = .015$	$\beta = .07$
Daylong Audio	Burstiness	Frequency	Burstiness x Frequency
All Classes ($R^2 = .12$)	$\beta = -.309$ ***	$\beta = -.123$ ***	$\beta = .265$ *
Nouns ($R^2 = .034$)	$\beta = -.196$	$\beta = .01$	$\beta = .1$
Verbs ($R^2 = .125$)	$\beta = -.31$ •	$\beta = -.09$ **	$\beta = .256$ **
Adjectives ($R^2 = .209$)	$\beta = -.297$	$\beta = -.067$	$\beta = .31$ *
Functions ($R^2 = .169$)	$\beta = -.4$	$\beta = .045$	$\beta = .047$

Note. • $p < .1$. * $p < .05$. ** $p < .01$. *** $p < .001$

Burstiness as a Construct

Another main aim of this work is to better understand burstiness itself as a construct. Specifically, we aim to

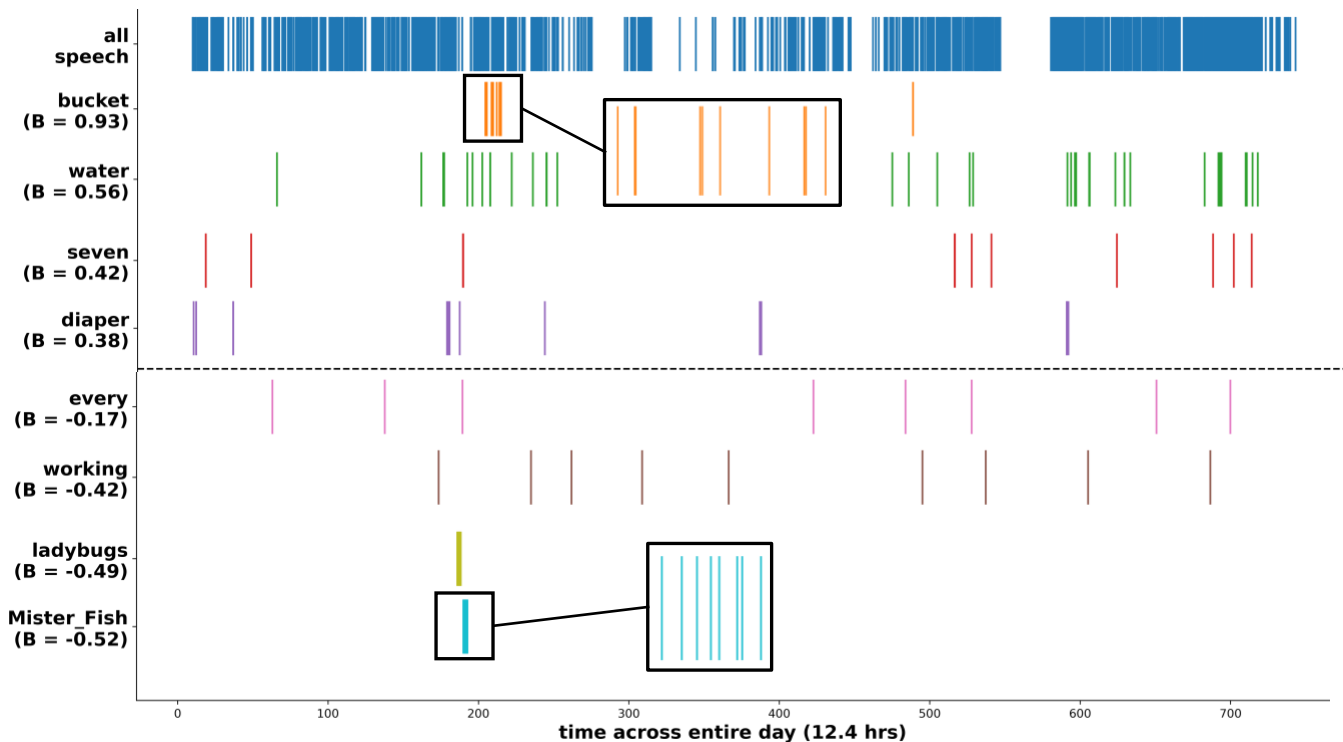


Figure 2: Event plots of words from a daylong transcript with positive burstiness values (above the dotted line) and negative burstiness values (below the dotted line). The presented bursty words occur 12-15 times, except for “water” which occurs 47 times; the non-bursty words presented occur 8-10 times. To better discern differences between event patterns, portions of the most bursty presented word (“bucket”) and the least bursty (“Mister Fish”) are enlarged.

understand how burstiness relates to and differs from constructs of temporal distributions primarily used in experimental paradigms (i.e., massed/spaced presentation order and context-dependent speech) and whether the metric could serve as a naturalistic analogue to these experimental constructs. To this end, we assessed the kinds of distributions the burstiness metric was categorizing as bursty ($B > 0$) and non-bursty ($B < 0$) and how well they align with massed and spaced presentations. One may assume that because bursty patterns are characterized as short successive occurrences (short IEIs) and longer intervals of no occurrences (long IEIs), that more bursty words ought to be more massed, and non-bursty words ought to be more spaced. When examining visualizations of the distribution of word occurrences, however, it becomes clear that this is not always the case.

To illustrate this, Figure 2 presents event patterns of bursty and non-bursty words from one individual daylong audio recording. This figure illuminates a few caveats about the burstiness metric. First, because words with positive burstiness values can take on a wider range of frequencies, the number of times a word occurs subsequently affects the nature of the temporal pattern it can take. For instance, despite “water” having a greater burstiness value ($B = 0.56$) than the words “seven” ($B = 0.42$) and “diaper” ($B = 0.38$), its occurrences still have a greater spread across the length of the transcript because it occurs 47 times whereas “seven” and “diaper” occur 14 and 15 times, respectively. This is even more apparent for words with even greater frequencies, such

as the word “you”, which occurs 1,752 times ($B = 0.54$). Since it is the most frequent word in this daylong recording, its occurrences were used to plot the “all speech” row in Figure 2; clearly, categorizing this distribution as massed would be a mischaracterization. The burstiness metric seems to depart from our intuitive understanding of “spaced” and “massed” learning in critical ways.

Examining distributions of non-bursty words further highlight how burstiness is not capturing the concept of massed/spaced distributions. While the event patterns of non-bursty words do indeed appear to be more periodic, as described by the metric, the events themselves can be widely spread out across the interval (see “every” and “working” in Figure 2) or clustered together (see “ladybugs” and “Mister Fish”). Because the burstiness metric considers the variation in interevent intervals and not necessarily the spread across an interval, the highly-contextually bound items can take on both positive and negative bursty values. For example, a word that occurs every day, only at breakfast time will appear very periodic if a single day is analyzed, and very bursty if two days are analyzed. Once again, the way burstiness is calculated means that burstiness values can vary widely based on the time interval that is used for computation.

Discussion

The present work aims to provide insights on the burstiness metric as it applies to naturalistic child-centered speech data. As the temporal dimension of children’s daily language

experiences becomes a greater focus of research, it is imperative that researchers have ecologically valid measures to operationalize constructs of theoretical interest. Initially an out-of-domain metric (Goh & Barabási, 2008), the burstiness metric provides researchers with a measure that is straightforward to implement and interpret. However, just as any other measure, it requires a thorough investigation in terms of how it is quantifying speech and how it aligns with and differs from constructs established within the field. This work aims to elucidate how the burstiness metric interacts with other measures (frequency), its behavior across different data types (shortform and longform timescales), and how it is similar and different from massed/spaced temporal order.

We observed that in relation to a word's frequency, a word's respective burstiness value can take on a wide range of values, and that burstiness is not necessarily a redundant measure of frequency (see Panel B of Figure 1). Non-bursty words tend to be less frequent; however, bursty words can have a wide range of frequencies such that an item's burstiness value is not completely conflated with its frequency.

We then used word burstiness and frequency to predict vocabulary norms to evaluate whether burstiness can be a useful predictor. Our model results show that burstiness can indeed capture significant variation, however results within individual lexical classes appear to vary due to the timescale upon which burstiness is calculated. These initial analyses aim to demonstrate that burstiness is potentially capturing some meaningful variation in words' temporal patterns. We do not aim to draw conclusions or propose mechanisms for how burstiness affects learning and why it may be more salient for some types of words over others, which certainly requires further and more detailed analyses. In the future, we aim to more closely tie these observed burstiness effects in language learning to the more mechanistic hypotheses for why the distribution of training may lead to different learning, as are described in the adult literature (Carvalho & Goldstone 2014; Brunmair & Richter, 2019; Raviv et al., 2022).

Further, we described that the burstiness metric behaves very differently across different timescales. Understanding how the metric is influenced by timescale is important as these differences will have consequences on how researchers can interpret their data. Just as Zipf's Law and Heap's Law are fundamental to understanding how frequency, type-token ratios, and corpus length interact with one another (Lü et al., 2010; Piantadosi, 2014; Montag et al., 2018), it is just as important to identify features of the corpora that affect how burstiness is computed or interpreted. Our analyses indicate that the distribution of burstiness changes significantly across corpora of different timescale lengths, and that the distribution of burstiness values also varies between different lexical classes. This is unlike word frequency, which is significantly correlated overall and within all individual lexical classes. This contrast makes sense. Longer form recordings capture more distinct contexts in a single recording, allowing the metric to capture words that are

bound to single contexts with longer lulls in between those contexts (e.g., mealtimes). Shorter form recordings often capture only a single context in isolation, making some context-dependence impossible to capture. Burstiness is influenced by timescale and the types of words that tend to be more/less bursty at longer timescales are not necessarily the same ones that tend to be similarly bursty at shorter timescales. Rather, burstiness appears to be capturing the nested temporal behavior of speech (deBarbaro & Fausey, 2022; Clerkin & Smith, 2022; Lancaster & Wass, 2024).

Finally, we evaluated how burstiness as a construct relates to the experimental construct of massed/spaced item presentation orders. If we want to connect the existing literature to naturalistic data, it is important to evaluate whether the burstiness metric can be used to quantify these kinds of constructs, or whether it should be regarded as a wholly separate construct. Our evaluations of the kinds of temporal distributions bursty and non-bursty words exhibit show that the burstiness metric cannot be used to separate out massed and spaced presentations. The burstiness metric does not seem to map on to our intuitive notion of spaced and massed distributions that derive from the laboratory-based learning literature. Rather, we find that bursty patterns tend to be simultaneously massed and spaced (i.e., multiple clusters of successive occurrences separated by long intervals of no occurrences), and non-bursty patterns can exhibit either isolated massed clusters of occurrences or distribution that are widely but more evenly spread out across the interval. This corresponds with Slone & colleagues (2023) conclusions, which suggest that comparing massed presentations against spaced presentations as is common in experimental work may not be an appropriate construct for the naturalistic language environment.

We see this work as an important first step toward operationalizing burstiness, or a similar metric, that will allow us to quantify distributions of words over time, and how those distributions may relate to trajectories of learning. Since burstiness appears to capture temporal patterns of words that appear in various contexts, a future line of work could discern burstiness' relationship from measures of the related construct of lexical/contextual diversity. Its relationship to contextual diversity measures could be like that of with frequency; different ranges of burstiness measures could exhibit smaller or greater ranges of contextual diversity measures. To quantify massed/spaced temporal patterns, future work may also involve developing a metric that not only considers the variation of interevent intervals as burstiness does but also considers the spread of occurrences across an interval, which may also allow for more consistency across corpora of different interval lengths.

We intend for this work to inform researchers about the behavior of this metric and its potential utilities and limitations when working with naturalistic speech data. Ultimately, we hope that these insights will shape how we develop our tools and metrics for investigating the temporal distribution of naturalistic speech.

References

- Chalnick, A., & Billman, D. (1988). Unsupervised learning of correlational structure. *Proceedings of the tenth annual conference of the cognitive science society* (pp. 510-516). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Abney, D. H., Dale, R., Louwerse, M. M., & Kello, C. T. (2018). The bursts and lulls of multimodal interaction: Temporal distributions of behavior reveal differences between verbal and non-verbal communication. *Cognitive science*, 42(4), 1297-1316.
- Altmann, E. G., Pierrehumbert, J. B., & Motter, A. E. (2009). Beyond word frequency: Bursts, lulls, and scaling in the temporal distributions of words. *PLOS one*, 4(11), e7678.
- Benitez, V. L., Zettersten, M., & Wojcik, E. (2020). The temporal structure of naming events differentially affects children's and adults' cross-situational word learning. *Journal of Experimental Child Psychology*, 200, 104961.
- Brunmair, M., & Richter, T. (2019). Similarity matters: A meta-analysis of interleaved learning and its moderators. *Psychological bulletin*, 145(11), 1029.
- Carvalho, P. F., & Goldstone, R. L. (2015). The benefits of interleaved and blocked study: Different tasks benefit from different schedules of study. *Psychonomic bulletin & review*, 22, 281-288.
- Carvalho, P. F., & Goldstone, R. L. (2017). The sequence of study changes what information is attended to, encoded, and remembered during category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 43(11), 1699.
- Clerkin, E. M., & Smith, L. B. (2022). Real-world statistics at two timescales and a mechanism for infant learning of object names. *Proceedings of the National Academy of Sciences*, 119(18), e2123239119.
- de Barbaro, K., & Fausey, C. M. (2022). Ten lessons about infants' everyday experiences. *Current Directions in Psychological Science*, 31(1), 28-33.
- Estes, W. K. (1955). Statistical theory of distributional phenomena in learning. *Psychological review*, 62(5), 369.
- Fenson, L. (2007). MacArthur-Bates communicative development inventories.
- Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of child language*, 44(3), 677-694.
- Carvalho, P. F., & Goldstone, R. L. (2014). Putting category learning in order: Category structure and temporal arrangement affect the benefit of interleaved over blocked study. *Memory & cognition*, 42, 481-495.
- Goh, K. I., & Barabási, A. L. (2008). Burstiness and memory in complex systems. *Europhysics Letters*, 81(4), 48002.
- Goldenberg, E. R., & Sandhofer, C. M. (2013). Same, varied, or both? Contextual support aids young children in generalizing category labels. *Journal of Experimental Child Psychology*, 115(1), 150-162.
- Jo, H. H., Karsai, M., Kertész, J., & Kaski, K. (2012). Circadian pattern and burstiness in mobile phone communication. *New Journal of Physics*, 14(1), 013055.
- Karmazyn-Raz, H., & Smith, L. B. (2023). Sampling statistics are like story creation: a network analysis of parent-toddler exploratory play. *Philosophical Transactions of the Royal Society B*, 378(1870), 20210358.
- Kim, E. K., & Jo, H. H. (2016). Measuring burstiness for finite event sequences. *Physical Review E*, 94(3), 032311.
- Lancaster, K. L., & Wass, S. V. (2024). Finding order in chaos: influences of environmental complexity and predictability on development. *Trends in Cognitive Sciences*.
- Lü, L., Zhang, Z. K., & Zhou, T. (2010). Zipf's law leads to Heaps' law: Analyzing their relation in finite-size systems. *PloS one*, 5(12), e14139.
- MacWhinney, B. (2000). The CHILDES Project: Tools for Analyzing Talk. 3rd Edition.
- Mendoza, J. K., & Fausey, C. M. (2021). Everyday music in infancy. *Developmental Science*, 24(6), e13122.
- Mol, S. E., Bus, A. G., De Jong, M. T., & Smeets, D. J. (2008). Added value of dialogic parent-child book readings: A meta-analysis. *Early education and development*, 19(1), 7-26.
- Montag, J. L. (2020). New insights from daylong audio transcripts of children's language environments. In *CogSci*.
- Montag, J. L., Jones, M. N., & Smith, L. B. (2018). Quantity and diversity: Simulating early word learning environments. *Cognitive science*, 42, 375-412.
- Montag, J. L., Jones, M. N., & Smith, L. B. (2015). The words children hear: Picture books and the statistics for language learning. *Psychological science*, 26(9), 1489-1496.
- Montag, J. L., Marasli, Z., & Koch, K., (2025). Theoretical and methodological insights from five transcribed and annotated daylong audio recordings of young children's language environments. Manuscript in preparation.
- Piantadosi, S. T. (2014). Zipf's word frequency law in natural language: A critical review and future directions. *Psychonomic bulletin & review*, 21, 1112-1130.
- Raviv, L., Lupyan, G., & Green, S. C. (2022). How variability shapes learning and generalization. *Trends in cognitive sciences*, 26(6), 462-483.
- Salditch, L., Stein, S., Neely, J., Spencer, B. D., Brooks, E. M., Agnon, A., & Liu, M. (2020). Earthquake supercycles and long-term fault memory. *Tectonophysics*, 774, 228289.
- Samuelson, L. K., & Horst, J. S. (2007). Dynamic noun generalization: moment-to-moment interactions shape children's naming biases. *Infancy*, 11(1), 97-110.
- Slone, L. K., Abney, D. H., Smith, L. B., & Yu, C. (2023). The temporal structure of parent talk to toddlers about objects. *Cognition*, 230, 105266.
- Snow, C. E., & Beals, D. E. (2006). Mealtime talk that supports literacy development. *New directions for child and adolescent development*, 2006(111), 51-66.
- So, L. H., Ghosh, A., Zong, C., Sepúlveda, L. A., Segev, R., & Golding, I. (2011). General properties of transcriptional time series in *Escherichia coli*. *Nature genetics*, 43(6), 554-560.

- Soderstrom, M., Casillas, M., Bergelson, E., Rosemberg, C., Alam, F., Warlaumont, A. S., & Bunce, J. (2021). Developing a cross-cultural annotation system and metacorpus for studying infants' real world language experience. *Collabra: Psychology*, 7(1), 23445.
- Tamis-LeMonda, C. S., Custode, S., Kuchirko, Y., Escobar, K., & Lo, T. (2019). Routine language: Speech directed to infants during home activities. *Child development*, 90(6), 2135-2152.
- Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, 127(3), 375-382.
- Vlach, H. A., & Sandhofer, C. M. (2011). Developmental differences in children's context-dependent word learning. *Journal of Experimental Child Psychology*, 108(2), 394-401.
- Vlach, H. A., Sandhofer, C. M., & Kornell, N. (2008). The spacing effect in children's memory and category induction. *Cognition*, 109(1), 163-167.
- Warlaumont, A. S., Sobowale, K., & Fausey, C. M. (2022). Daylong mobile audio recordings reveal multitime-scale dynamics in infants' vocal productions and auditory experiences. *Current directions in psychological science*, 31(1), 12-19.
- Weizman, Z. O., & Snow, C. E. (2001). Lexical output as related to children's vocabulary acquisition: Effects of sophisticated exposure and support for meaning. *Developmental psychology*, 37(2), 265.