Journal of Experimental Psychology: General

Manuscript version of

The Temporal Distribution of Speech to Infants Highlights Relatedness Among Words

Lillian R. Masek, Catalina Suarez-Rivera, Ori Ossmy, Catherine S. Tamis-LeMonda

Funded by:

- Eunice Kennedy Shriver National Institute of Child Health and Human Development
- LEGO Foundation
- National Science Foundation

Abstract

Categories are fundamental to cognition. By age two, infants form categories—animals, foods, and so on—and leverage them to learn new words. What enables them to do so? Prior work has focused on mechanisms internal to infants (e.g., core concepts and biases) and perceptual features of stimuli that facilitate learning (e.g., shape). Here we test the hypothesis that a key property of language input itself—the timing of words to infants—provides a reliable structure that informs on the semantic connections among words. We quantified the temporal structure of infants' speech exposure from video-recordings of infant-mother interactions (*N* = 78) during 1-2 hours of everyday activities (130 cumulative hours). Analysis of time-locked transcriptions yielded 1 billion+ latencies among 350,000+ words. Words from the same superordinate category (i.e., animals, vehicles, food, body parts, and clothing) occurred within a tighter time window relative to one another than did words overall and concrete nouns specifically. Thus, temporal parameters of speech offer reliable data that infants can leverage in the learning of taxonomic categories.

Keywords: taxonomic categories, language development, parent-child interaction, temporal structure, infancy

Word Count: 5393

Public Significance Statement

Learning language is more than adding words to a vocabulary. In the second year of life, infants begin to understand that words like apple, orange, and grapes refer to the category of foods, and that words like hats, shoes, and socks refer to the category of clothing. How might characteristics of speech aid infants' learning of categories? We tested whether the 'timing' of words to infants during everyday activities at home might inform on category membership. Transcriptions of videorecords of infant-caregiver interactions generated hundreds of thousands of words for analyses on when words were spoken relative to other words. Findings revealed that words within five early learned categories—animals, vehicles, foods, clothing, and body parts—occurred closer in time than did words generally and concrete nouns specifically. Thus, everyday exposure to ‘like words that hang together in time’ may offer a reliable cue to category membership.

The Temporal Distribution of Speech to Infants Highlights Relatedness Among Words

Categorization is fundamental to human cognition. In a world rich with information, categories organize experiences into meaningful and manageable units. The ability to categorize allows people to establish order, structure and retain information, draw inferences, reason and make predictions, and generalize from prior to new experiences (Perszyk & Waxman, 2018; Rakison & Oakes, 2003; Waxman & Gelman, 2009).

Language is vital for representing categories. With language, people can learn, think, and communicate about several things simultaneously. For example, 'dog' refers to all canines, from Chihuahuas to Great Danes. Words can refer to entities at varying levels of abstractness. Basic- level categories (e.g., dog) have high overlap in attributes and functions. A Chihuahua and a Great Dane differ in size, but both provide companionship; have fur, four feet, and a tail; and bark when the doorbell rings. In contrast, superordinate categories (e.g., animal) are abstract, with high variation among category members: Dogs and fish are both animals, but have different covering and appendages, and interact differently with people (Quillian, 1966; Rogers & McClelland, 2004; Rosch et al., 1976).

Impressively, infants begin to form superordinate categories in the second year (e.g., Wojcik, 2018). After being primed with a target word (e.g., cow) from a given superordinate category (e.g., animal), 21- to 24-month-olds shift attention faster and look longer to images of a word from the same category (e.g., sheep) than to images from a different category (Arias-Trejo & Plunkett, 2009; Styles & Plunkett, 2009). In turn, infants leverage their knowledge of superordinate categories to learn new words. Two-year-old infants who knew many words for vehicles and few words for clothing were faster to infer that a 'draisine' was a vehicle than a 'banyan' was an article of clothing (Borovsky et al., 2016).

What enables infants to connect like words and form superordinate categories? Unlike adults, who can draw on complex, hierarchical linguistic networks to incorporate new words into existing categories, novice infants do not have the vocabulary to do so. Researchers have long appealed to various internal mechanisms to explain infants' ability to form categories, including innate knowledge structures (e.g., Spelke & Kinzler, 2007), biases to attend to certain features over others (e.g., Mandler, 2004), and general learning mechanisms (e.g., Oakes, 2009).

Of course, internal processes necessarily operate on information in the environment. In particular, speech directed to infants likely facilitates the learning of taxonomic categories. Here, we consider a potentially rich source of linguistic information—the timing of words relative to one another. Regularity in the *temporal structure* of the words that caregivers direct to infants may communicate critical information about which basic-level entities belong together. That is, if words within superordinate categories (e.g., vehicles, body parts) occur closer in time relative to words generally, their temporal closeness may function to highlight word-to-word connections. To test this possibility, we videorecorded infant-mother dyads in the ecologically valid home environment (1-2 hours per visit). We quantified latencies among every word in a massive audiovisual corpus (350,000+ words timed to the msec yielded 1+ billion latencies) and compared the distributions of latencies between words within superordinate categories to the distribution of latencies between all words generally and concrete nouns specifically.

Building Taxonomic Categories

What features of input facilitate infants' learning of taxonomic categories? Perceptual features, including visual complexity and variability, certainly matter (Perry et al., 2010; Ross et al., 1986; Son et al., 2008). For example, infants leverage information about shape to categorize familiar and novel objects (Anderson et al., 2018; Clerkin & Smith, 2022; Oakes & Rakison, 2019; Rakison & Butterworth, 1998). Eighteen- to 24-month-old infants were more likely to generalize a label to a novel object of the same shape when taught that the label corresponded to simple objects with few features rather than complex objects with many features (Son et al., 2008).

However, infants cannot solely rely on perceptual features to learn abstract superordinate categories: Planes and birds both have wings, but planes are vehicles and birds are animals (Mandler & McDonough, 1993). Linguistic input is a critical and complementary tool for the conceptual organization of words. Already in the second year, infants capitalize on information in language input to draw connections among related things. When researchers labelled a superordinate category as applying to exemplars within the category (e.g., 'ong' paired with a picture of a guinea pig, koala bear, and a ram), infants generalized the label to a novel exemplar for that category (e.g., a meerkat; Lany & Saffran, 2010). Infants exposed to a consistent label across a set of objects were more likely to form a single category that encompassed all objects than were infants exposed to a different label for each object (LaTourrette & Waxman, 2019; Plunkett et al., 2008; Waxman & Braun, 2005). However, caregivers infrequently use superordinate labels compared to basic-level words (e.g., 'dog' rather than 'animal'; e.g., Leung et al., 2021; Poulin-Dubois et al., 1995). Thus, characterizing infants' exposure to category exemplars offers a fuller understanding of the language input on which infants build categories.

Speech to Infants in the Everyday Home Environment

Infants build language by extracting regularities in the input (Smith et al., 2018). That is, the speech that caregivers direct to infants may contain sufficient structure to highlight which words belong together. Understanding what goes into the system (i.e., the input on which infants learn language) is fundamental to understanding what comes out of the system (e.g., category knowledge). To date, lab-based observations, characterized by brief sessions with scripted input, dominate the study of taxonomic categories. In reality, language learning occurs in rich physical and social contexts. During everyday interactions at home, infants are exposed to thousands of words a day (Bergelson et al., 2019; Gilkerson et al., 2017; Tamis-LeMonda et al., 2017), across different activities, in different spaces, while engaging with a variety of objects that provide moment-to-moment exposure to speech about different categories (e.g., foods, clothing, and so on; Custode & Tamis-LeMonda, 2020; Hoff-Ginsberg, 1991; Roy et al., 2015; Tamis-LeMonda et al., 2019). Accordingly, infant-caregiver spontaneous language interactions at home are a primary, yet largely untapped source of big data on the temporal structure of speech to infants.

The Temporal Distribution of Words to Infants

Language interactions, like all behaviors, unfold over time. Infants are exposed to different words at different moments in time, with utterances spaced across the day. Over 100 years of research in cognitive psychology has converged on the importance of timing for memory (e.g., Ebbinghaus, 1913), attention (e.g., Yu & Smith, 2016), and learning (e.g., Trueswell et al., 2016). For example, two events are likely to be connected in memory when they occur within a brief time window (Rovee-Collier, 1995). Similarly, tight temporal clustering of words is associated with greater learning of novel words by one-to-two-year-old infants after controlling for the frequency of object labels (Slone et al., 2023). Moreover, the timing of input (e.g., massed versus spaced presentation of nonce words) differentially affects the retention of words by 16-month-old versus 20-month-old infants (Vlach & Johnson, 2013), such that retention intervals for linking two events increases across the second year.

Nonetheless, studies of everyday speech to infants in the home have not yet examined the time-locked spacing of words relative to one another. Consider a mother who speaks 20 different words to her toddler in 10 minutes (ordered 1, 2, 3, and so on). If the mother said each of the 20 words at evenly spaced intervals, the latency between the first and second word would equal the latency between the second and third word, third and fourth word, and so on. In such a scenario, the distribution of latencies among words would approximate a triangle (Figure 1 Panel A). In contrast, if the mother produced words in bursts (i.e., with certain words appearing closer in time than others), the temporal distribution would show a distinct peak at short latencies, resulting in a right-skewed distribution (Figure 1 Panel B). Although these two distributions contain the same number of words, differences in the timing of words may critically impact learning. In the case of taxonomic categories, tight temporal timing may inform on which basic-level entities comprise the same superordinate.

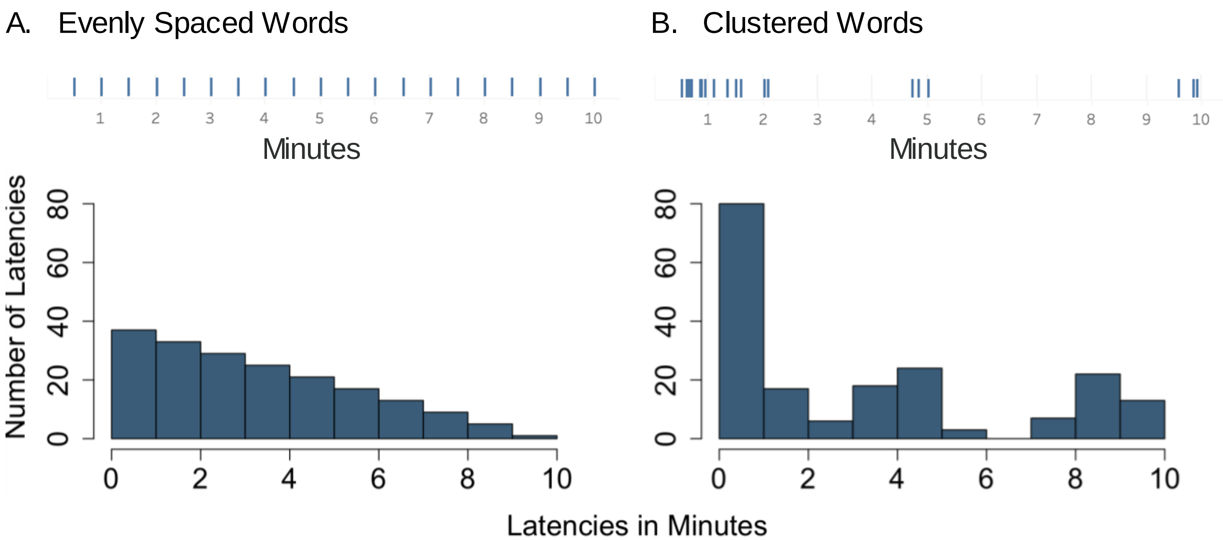


Figure 1. Two hypothetical distributions of temporal patterns. Panel A shows a hypothetical distribution of latencies based on evenly spaced words in the input. Panel B shows a hypothetical distribution of latencies based on words clustered in the input. The distribution of evenly spaced words in Panel A yields a triangular shape. The distribution of temporally clustered words in Panel B yields a sharp peak at short latencies. Note: these hypothetical (not real) data serve as a visualization for understanding distinct temporal patterns of word input.

Several studies suggest that words within superordinate categories may occur in tighter temporal proximity than words more generally. Everyday activities shape the words that caregivers direct to infants (e.g., Hoff-Ginsberg, 1991; Soderstrom & Wittebolle, 2013): Mothers name foods during mealtime, clothing items during dressing, animals while reading, and so on (Tamis-LeMonda et al., 2019), and mothers use specific words in specific rooms and locations, such as saying ‘diaper’ to an infant on a changing table (Custode et al., 2020; Roy et al., 2015).

Furthermore, computers trained on interactions between caregivers and infants reliably generate semantic connections among words from ordinal data (referred to as co-occurences)—such as two words being adjacent or separated by only a few words (e.g., ‘dog’ and ‘cat’ being closer in order to each other than ‘dog’ and ‘spoon’), or ordinally connected to a third, common word (e.g., ‘dog’ and ‘cat’ appearing in order with ‘furry’; Fourtassi et al., 2020; Huebner & Willits, 2018; Jiang et al., 2022; Unger & Fisher, 2021). Natural language processing models applied to a corpus of child-directed speech found that words for taxonomic categories—including animals, vehicles, body parts, clothing items, and food—were more likely to co-occur within three, seven, and 11 words compared to words with similar frequencies from different taxonomic categories (Unger et al., 2023).

Nonetheless, ordinal data do not provide critical precision on the “temporal closeness” of words. For example, the two distributions of words in Figure 1 have identical ordinals, but differ in their temporal separation. In Figure 1 Panel A, every adjacent word is 30 second apart. In contrast, the clustered words in Figure 1 Panel B vary from 2 seconds to 4 minutes apart. Thus, accurate quantification of the temporal properties of speech offers a unique, yet unexamined window into the input around taxonomic categories.

Current Study

We generated a massive dataset on word-to-word latencies to test whether natural speech to infants is characterized by a tight temporal structure for words from five superordinate categories—animals, vehicles, clothing, body parts, and food. These categories are common in children’s vocabulary by age two and half (Fourtassi & Frank, 2019) and frequently tested in studies of early category learning (Borovsky et al., 2016; Lany & Saffran, 2010; Mandler & McDonough, 1993; Waxman & Braun, 2005).

We hypothesized that words within superordinate categories would yield a higher proportion of short latencies—and thus a more pronounced left-peaked distribution (Figure 1 Panel B)—than latencies among all words that mothers direct to infants. Furthermore, we expected the tight temporal structure among words from a superordinate category to apply to the speech of mothers in aggregate and individually. To ensure that such differences were not driven by word class (i.e., words from the five tested categories were all nouns), we further compared latencies between words within superordinate categories to latencies between all concrete nouns that mothers directed to their infants. In contrast to the tight timing of words within superordinate categories, we expected the temporal distributions of all words and all nouns to approximate the triangular shape of an evenly spaced distribution (Figure 1 Panel A).

Finally, in exploratory analyses we examined the frequency of mothers' use of superordinate labels (i.e., the words animal, vehicle, clothes/clothing, body, and food), and the timing of superordinate labels relative to exemplars from the same category.

Methods

Participants

Participants were 78 English-speaking mothers and their first-born infants. Infants were 12 to 24 months old (*M* = 17.57, *SD* = 4.11, *range* = 11.93 – 24.43); 36 were reported to be female and 42 to be male. All infants were born at term and all mothers spoke English to their infants. Mothers averaged 35 years old (*SD* = 4.7). Most mothers (89.7%) and fathers (84.6%) had earned at least a bachelor's degree and 42.3% and 38.5%, respectively, had post-bachelor education. Most mothers (70.5%) and fathers (88.5%) worked full- or part-time outside the home. From a set of options, 62 mother (79.5%) identified as White, three (3.8%) as Black or African American, two (2.6%) as Asian, five (6.4%) as other (i.e., a race not listed), three (3.8%) as more than one race, and three (3.8%) mothers did not report their race. No mothers reported their race as American Indian or Alaska Native or as Native Hawaiian or Other Pacific Islander. Sixteen mothers further identified as Hispanic or Latino, 57 identified as not Hispanic or Latino, and five mothers did not report their ethnicity. All information about race and ethnicity was reported by the mother. Families were recruited from a large city through local hospital records, referrals, and brochures.

Procedures

A trained researcher video-recorded mothers and infants for 1 to 2 hours during everyday home activities using a handheld video camera. In aggregate, we recorded 129.58 hours of everyday activities across participants (recording time *M* = 99.7 minutes, *SD* = 28.2). Consent was obtained from the mother prior to the recording. During the recording, the researcher kept the infant in view, recorded the mother when she was near the infant, but kept focus on the infant when they separated. Visits were scheduled during weekday daytime hours when the infant was not napping and only the mother and infant were present. Mothers were instructed to go about their daily activities as if the researcher was not present, but to remain inside the home.

Transcription Process

Interactions were transcribed using Datavyu (https://datavyu.org), an open-source, computerized coding tool designed for behavioral researchers. Trained research assistants transcribed the interaction using a protocol established in PLAY-project.org, adapted from guidelines in Codes for the Human Analysis of Transcripts (CHAT; https://talkbank.org/manuals/CHAT.pdf) conventions. Each utterance was time-locked to the msec, allowing for an estimation of when mother said each word. Transcripts were exported into the Computerized Language Analysis program (CLAN), and generated 355,334 total words (i.e., word tokens) and 8,064 unique words (i.e., word types) aggregated across mothers.

From the full list of words, we manually coded lexical class. Words that could be coded as either an action verb or a concrete noun (e.g., ‘drink’ and ‘hug’) were coded as the verb. Words with multiple common meanings (e.g., ‘can’ and ‘back’) and non-English words were excluded. From the list of words, we categorized concrete nouns into one of five superordinate categories (see Table 1). Words that did not fit one of the superordinate categories were classified as other concrete nouns (e.g., names of household objects; see Figure 2). Two researchers coded the words and disagreements were settled through discussion.

Category	Definition	Tokens	Types	Most frequent words
Animals	Animal names, sounds, features	5775	215	dog, bear, bunny
Food	Food names	4012	205	milk, food, cheese
Body Parts	Body parts, features	4020	82	hand, foot, mouth
Clothing	Clothing items/materials	1362	80	diaper, shoe, sock
Vehicles	Vehicle names, sounds, parts	1522	52	car, choo, bus

Table 1. Categories included in the analysis. Latencies between words from five superordinate categories were used to test the hypothesis that temporal structure of words provides reliable structure which may be used by infants to learn categories. Categories are ordered from most frequent in the input to least frequent.

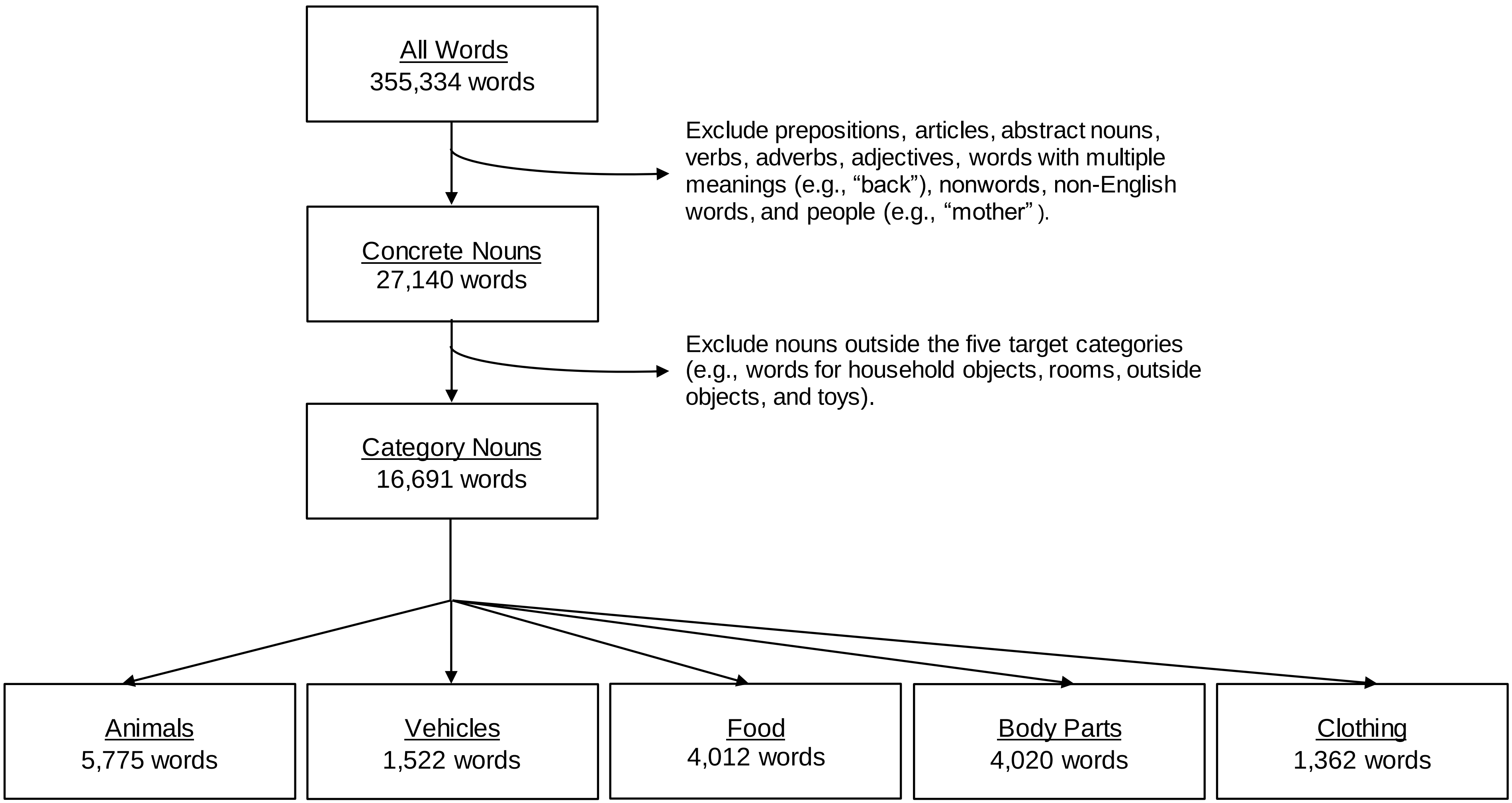


Figure 2. Words included in analyses. To generate data for analyses, a supercomputer calculated the latencies in milliseconds between 355,334 word tokens spoken by 78 mothers across 130 hours of parent-child interaction yielding 1,082,450,983 latencies for analysis (the baseline distribution). To generate data for the concrete noun comparison, we filtered out prepositions, articles,

abstract nouns, verbs, adverbs, adjectives, words with multiple meanings (e.g., “back”), nonwords, non-English words, and types of people from the baseline distribution, yielding 27,140 concrete noun tokens and 6,525,400 latencies between concrete nouns. To generate data on the five superordinate categories, we further filtered out nouns that were not from one of the five superordinate categories and latencies between words from different superordinate categories.

Computing Word Latencies and Categorization of Words

Latencies among all words spoken by mothers were generated by a supercomputer with 665 nodes, 31,584 Intel cores, and 163 terabytes of RAM. The program took a week to run nonstop. The supercomputer calculated over a billion latencies to the msec between every word the mother said and every subsequent word she said for each of the transcripts. For example, if the mother said “Get the ball. Over there.” the computer computed the number of milliseconds between the words: 1) “get” and “the”, 2) “get” and “ball”, 3) “get” and “over”, 4) “get” and “there”, 5) “the” and “ball”, 6) “the” and “over”, 7) “the” and “there”, 8) “ball” and “over”, 9) “ball” and “there”, and 10) “over” and “there”. Timing of words was determined at the utterance level. Thus, words within an utterance had a latency of 0 milliseconds. In the example above, latencies are 0 for pairs 1, 2, 5, and 10 since these words occurred within the same utterance. The full matrix of latencies (on all words, regardless of word class) yielded a ‘baseline’ distribution.

After generating the full set of word latencies, we filtered data to include latencies among all concrete nouns—including nouns from the five superordinate categories—to create a comparison set of latencies for words within the same word class. Finally, we selected latencies among words within each of the five superordinate categories (Table 1), including repetitions (e.g., ‘dog’ to ‘cat’ and ‘dog’ to ‘dog’). See https://nyu.databrary.org/volume/1762 for full lists of words.

To ensure that differences in the number of latencies between the five categories of interest and the comparison sets (i.e., baseline distribution and noun distribution) did not affect findings, we randomly sampled an equal number of words from the overall baseline distribution and noun distribution as in the target superordinate category. For example, the category of animals had 5,775 words (aggregated across mothers), resulting in 390,555 latencies; thus, for comparison purposes, we randomly selected 390,555 latencies from the baseline distribution and 390,555 latencies from the noun distribution to create distributions for comparison. For each comparison, we ran 100 iterations to test the robustness of the effect. All iterations yielded equivalent results (see https://nyu.databrary.org/volume/1762 for the distribution of effects for all 100 iterations). Thus, we present results from the first iteration only.

Two metrics compared temporal distributions of words within superordinate categories to the temporal distribution of words generally and the temporal distribution of all concrete nouns. A Kolmogorov–Smirnov test analyzed the overall shape of the distributions to determine whether they differed in temporal structure. To determine whether any observed differences in temporal structure were driven by words occurring close together in time, a two-proportion *z*-test analyzed the proportion of short latencies (words occurring within 60 seconds). One minute—a researcher-determined cutoff—was used as the cut off for short latencies to align with prior work on everyday activities around eating, bathing, and so forth (Tamis-LeMonda et al., 2019) and to allow sufficient time to capture conversations that involve naming multiple exemplars in a category. However, we acknowledge that the 60-second cutoff (as any cutoff) is arbitrary. Tests were run on the pooled data followed by data from individual mothers. Descriptive statistics were calculated at the visit level to provide the true number of words available for analysis.

Transparency and Openness

We report all participant recruitment criteria, all data exclusions (if any), all procedures, all variables, and all analyses in the study. All data, analysis R scripts, and study materials are publicly available on Databrary (https://nyu.databrary.org/volume/1762). The transcription manual is freely available online (https://play-project.org). Statistical analyses were run in R (v4.1.1; R Core Team, 2021) via RStudio (v2023.12.1.402; Posit Team, 2024) using the following packages: dplyr (v1.0.10; Wickham et al., 2022), here (v1.0.1; Müller, 2020), tidyr (v1.2.1; Wickham & Girlich, 2022), ggplot2 (v3.4.1; Wickham, 2016), moments (v0.14.1; Komsta & Novomestky, 2022), dgof (v1.2; Arnold & Emerson, 2011), forcats (v0.5.1; Wickham, 2021), and progress (v1.2.3; Csárdi & FitzJohn). This study was not preregistered.

Results

Individual mothers averaged 4,556 (*SD* = 2,554) word tokens and 635 (*SD* = 263) word types per visit but varied enormously from a low of 262 tokens and 117 types to a high of 10,874 tokens and 1,498 types. Within this full baseline set of words, mothers also varied in their use of concrete nouns, averaging 348 (*SD* = 220) word tokens, ranging from 11 to 852 and 111 (SD = 58) word types, ranging from 5 to 258.

Across the five superordinate categories, mothers used on average 214 (*SD* = 146) word tokens and 64 (*SD* = 35) word types. All mothers (*N*=78) used words for food, 76 used words for body parts and animals, 75 used words for clothing, and 71 used words for vehicles. Mothers used a variety of words within each category, averaging between 17 tokens (clothing) and 74 tokens (animals) per category. These data confirm that such words are frequent in input to infants, as documented by others (Bergelson et al., 2019; Custode et al., 2020; Nelson, 1973; Tamis-LeMonda et al., 2019), and thus provide ample data to quantify the temporal structure of the input.

Words Within Categories Clustered More than All Words and Concrete Nouns

As hypothesized, words within each superordinate category were more likely to occur close in time compared to words generally and nouns specifically. That is, for all categories, the distribution of latencies between words differed significantly in shape, as indicated by a Kolmogorov–Smirnov test, from the distribution of randomly sampled latencies for all words, *d*'s = 0.101 (food) – 0.232 (vehicles); all *p*'s < .001, and all concrete nouns, *d*'s = 0.087 (food) – 0.219 (vehicles); all *p*'s < .001. The distributions of latencies for words within categories showed an amplified right skew (left peak of short latencies; see Figure 3), resulting in a significant difference in the proportion of latencies under one minute compared to randomly sampled latencies for all words, X^2's = 1452 (clothing) – 18003 (animals), Δ proportion = .043 (food) - .147 (vehicles), all *p*'s < .001, and all concrete nouns, X^2's = 1098 (clothing) – 11222 (animals), Δ proportion = .037 (food) - .137 (vehicles), all *p*'s $< .001$. Notably, unlike the amplified right-skewed distributions of latencies for superordinate categories, distributions based on baseline words and concrete nouns generally adhered to a triangular pattern with a very slight peak at short latencies (See Figure 4).

Differences in the temporal distribution of words within superordinate categories compared to both all words and concrete nouns suggest that timing is a salient cue in the data available to infants learning superordinate categories. Moreover, the temporal distribution of words within superordinate categories does not reflect a pattern characteristic of nouns generally. To ensure this effect was not driven by words that occurred in the same utterance (i.e., 0 msec latencies; 0.3% of all category latencies) or repeated words (e.g., "dog" to "dog"; 9.9% of all category latencies), we re-ran these analyses excluding all 0 msec latencies and excluding all latencies between repeated words. Results remained overwhelmingly significant, suggesting a robust effect. See supplementary materials for additional analyses.

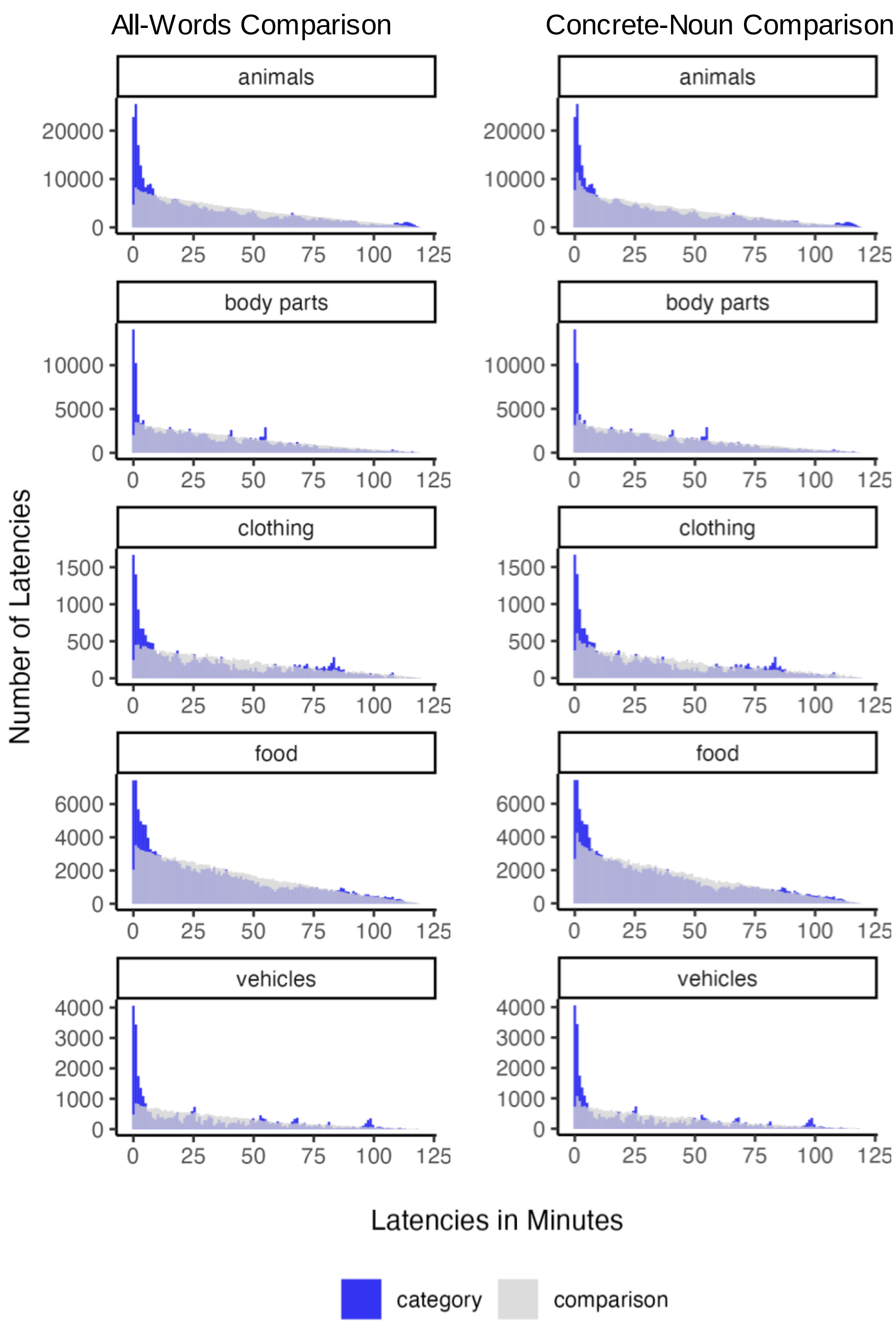


Figure 3. Distributions of category latencies and comparison latencies matched for frequency. The five category distributions yielded an exaggerated peak at short latencies relative to baseline latencies and relative to concrete nouns. The overlapping histograms compare the distribution of category latencies (blue) against the same number of latencies

(grey) randomly sampled from all words (left panel) and concrete nouns (right panel). The scale for the y-axis varied across categories to match the variation in the number of latencies occurring between words of different categories.

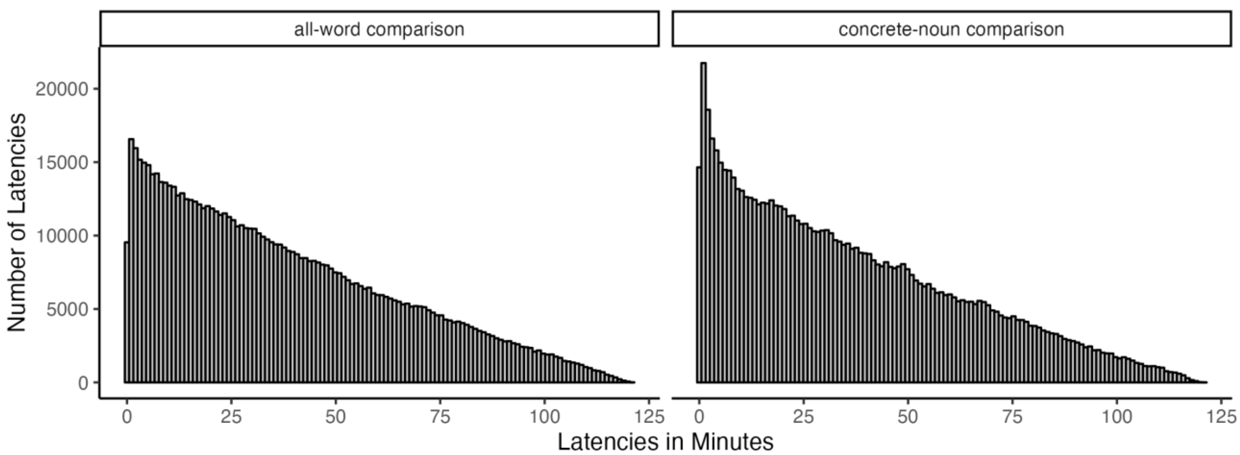


Figure 4. Distribution of latencies included in the all-word comparison and the concrete-noun comparison. Relative to the distribution of latencies within superordinate categories, latencies across all words and across concrete nouns approximated a triangle shape with a slight peak at short latencies, suggesting concrete nouns and words overall are evenly spaced with some, but minimal, clustering of words.

Tight Temporal Proximity for Words within Categories Applies to the Speech of All Mothers

With very rare exceptions, findings for individual mothers mirrored those for pooled data, providing rigorous support for the hypothesis that close temporal proximity among words within superordinate categories applies to the speech exposure of individual infants. Although individual mothers varied in the specificity of their word timing, of the 290 comparisons (i.e., comparisons of every mother for each of the superordinate categories they used), most (approximately 95%) showed statistically significant divergence in the temporal structure of their words within superordinate categories compared to all words and all nouns (see Figure 5A and 6A). Similarly, for comparisons of proportion of latencies under a minute, 92% showed statistically significant divergence of superordinate categories compared to all words and 84% showed statistically significant divergence compared to all nouns (see Figure 5B and 6B). Notably, every mother showed a difference for most superordinate categories. Thus, the temporal closeness of words within superordinate categories could not be explained by a handful of mothers: Individual infants were commonly exposed to words that hung together in meaning and time.

A.

B.

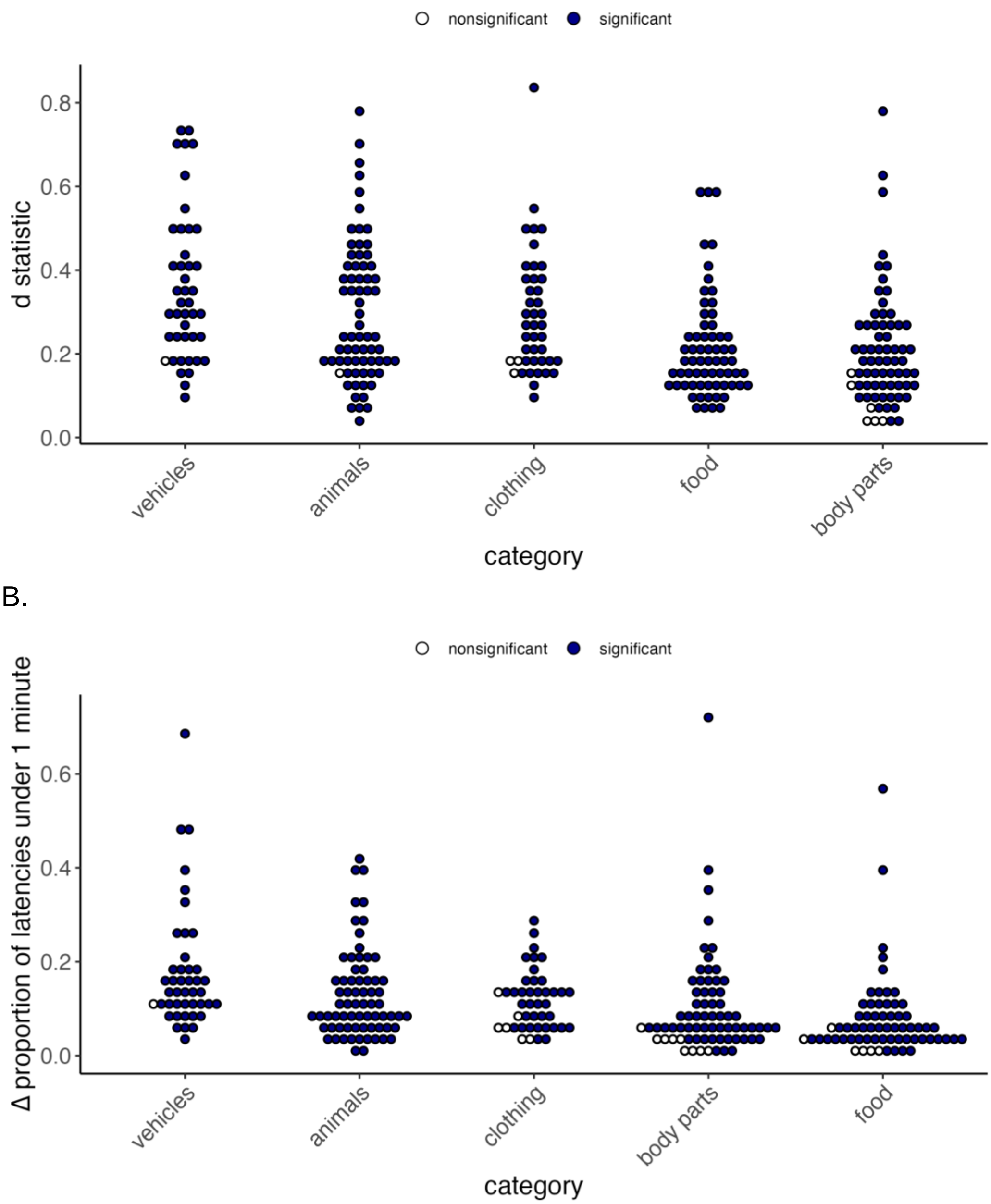
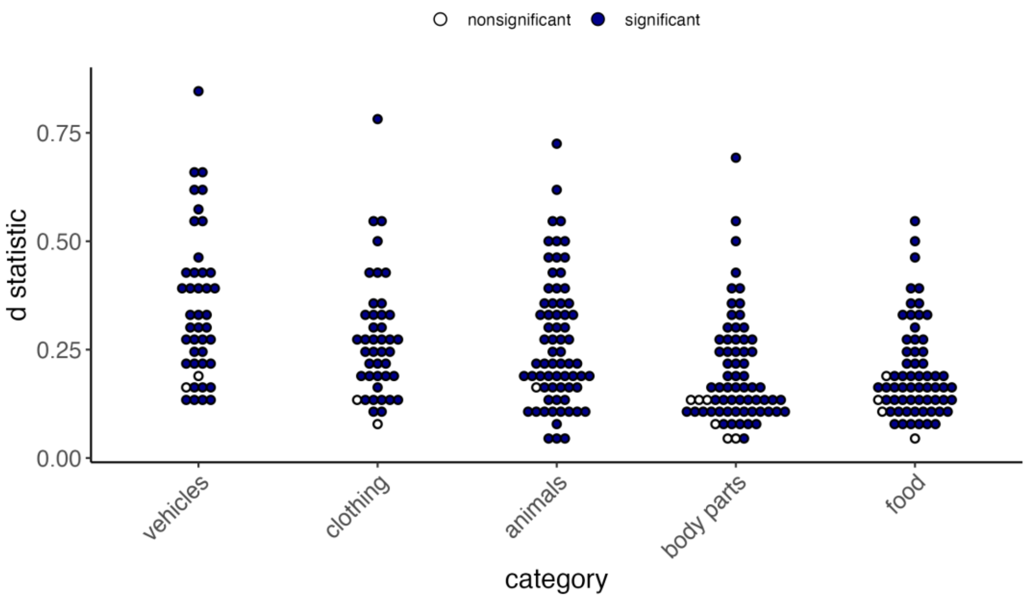


Figure 5. Results of A) Kolmogorov–Smirnov for each category and B) Two-proportion z-test comparing category latencies to latencies between all words for individual mothers. Higher *d* statistics and higher differences in the proportion of words under 1 minute represent larger differences between category and baseline distributions. Categories are ordered by means. Each circle represents one Kolmogorov-Smirnov test (panel A) or two-proportion z-test (panel B) run for one mother and one category of words. Filled circles indicate $p < .05$; unfilled

circles indicate $p > .05$. Only mothers who produced at least 10 words in a given category are included in these tests.

A.



B.

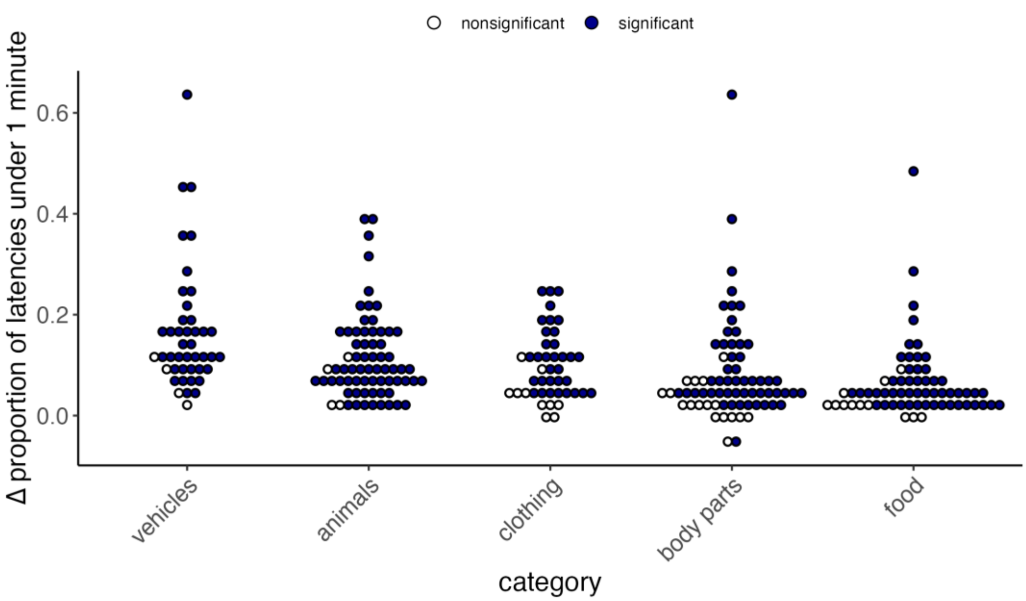


Figure 6. Results of A) Kolmogorov–Smirnov for each category and B) Two-proportion z-test comparing category latencies to latencies between concrete nouns for individual mothers. Higher *d* statistics and higher differences in the proportion of words under 1 minute represent larger differences between category and concrete noun distributions. Categories are ordered by means. Each circle represents one Kolmogorov-Smirnov test (panel A) or two-proportion z-test (panel B) run for one mother and one category of words. Filled circles indicate

p < .05; unfilled circles indicate *p* > .05. Only mothers who produced at least 10 words in a given category are included in these tests.

Mothers do not Consistently use Superordinate Labels

Exploratory analyses revealed relatively rare use of superordinate labels by mothers. Of the five superordinate labels, four were used by fewer than half of the mothers—'vehicle' was used by one mother (1.3%), 'body' was used by 9 mothers (11.5%), 'clothing' or 'clothes' by 21 mothers (26.9%), and 'animal' was used by 35 mothers (44.9%). The only superordinate label used by more than half of mothers was 'food', used by 46 mothers (59.0%). Interestingly, the categories for which fewer mothers used the superordinate label had the tightest temporal clustering as shown by the Kolmogorov–Smirnov test and the two-proportion *z*-tests. Thus, temporal structure may be a critical feature of language input that supports category formation when superordinate labels are infrequent.

When mothers used superordinate labels, most did so infrequently. Mothers used, on average, fewer than five superordinate labels throughout the observation, though individual differences were large. 'Vehicle' was used 3 times, 'body' was used 1.2 times (*SD* = 0.4), 'clothing' or 'clothes' was used 2.9 times (*SD* = 3.4), 'food' was used 3.2 times (*SD* = 3.2), and 'animal' was used 3.4 times (*SD* = 3.6). The overall low frequency mirrors prior lab-based research (e.g., Leung et al., 2021; Poulin-Dubois et al., 1995).

Although superordinate labels were relatively infrequent, we explored whether their timing relative to words from the same taxonomic category might provide further insight into word meaning. Of the 340 superordinate labels used by mothers across the corpus (i.e., total number of times 'animal', 'body', 'clothes', 'clothing', 'food', and 'vehicle' were produced), 298 occurred within one minute of a word from that category (e.g., dog)—yielding a temporal proximity rate of .876 compared to the .124 rate of superordinate labels used in isolation (*IRR* =

7.09, $p < .001$). Moreover, most superordinates were flanked by category words on either side ($n = 171$), with the remaining superordinates preceding ($n = 72$) or following ($n = 55$) category words. Timelines of the five mothers who used the most superordinates (Figure 7) illustrates that superordinates were primarily preceded and/or followed by category exemplars, although occasionally used in isolation. Thus, temporal structure of speech to infants may work together with superordinate labels to support learning (e.g., hearing ‘animal’ close in time to ‘dog’, ‘cat’, ‘cow’, and ‘horse’).

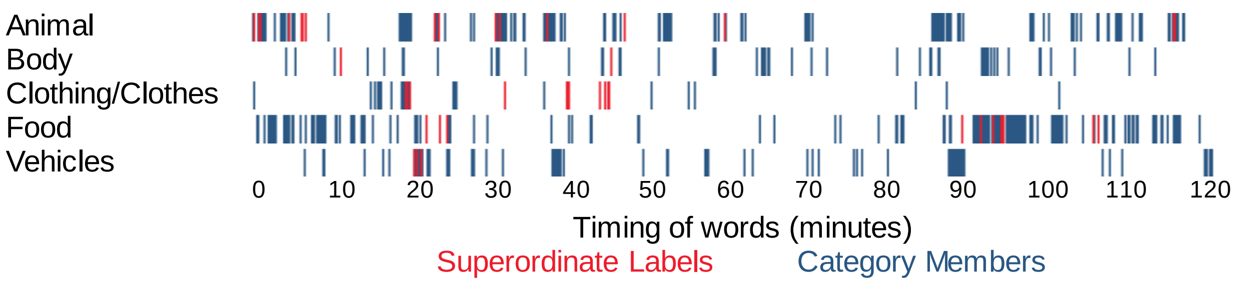


Figure 7. Temporal distribution of words within categories (blue) and superordinate labels (red). Each row represents the speech to one infant. Each vertical line represents a word, and blocks of color represent clusters of words that occur close in time (due to resolution constraints). For illustrative purposes, timelines represent mothers who used the most superordinate labels for each category. For timelines of all mothers, see Figures S3-S7 in the supplementary materials.

Discussion

Growing a vocabulary is more than accumulating words. As infants build their vocabulary, they create rich semantic networks—connections among words based on meaning (Beckage et al., 2010; Collins & Loftus, 1975; Wojcik, 2018). Superordinate categories—words of a similar kind such as apple, pasta, and cheese that fall under the broader group of foods—are one type of semantic network. Infants begin to construct superordinate categories in the second year, yet how infants form such categories remains debated. Prior work has focused on mechanisms internal to the infant and available perceptual information, leaving open questions about the nature of the linguistic input—the data from which infants build superordinate categories.

Timing of Words within Categories

We characterized the temporal structure of words to infants during everyday activities. Unlike lab studies that constrain mothers' speech to infants, our examination of interactions at home generated ecologically valid data on the timing of words that infants spontaneously experience as they go about their day. Our analyses, based on an enormous amount of data—over 100 hours of mother-child interaction across 78 mothers, hundreds of thousands of words, and over 1 billion latencies—revealed robust and reliable patterns in the structure of speech to infants. Words within categories clustered closer in time than did either words generally or concrete nouns specifically. Moreover, temporal closeness characterized the speech of every mother, providing robust evidence that the timing of word input may cue infants into the words that belong together.

Quantifying the timing of events (here words) in ecologically valid settings (here the home) is fundamental to theories of attention and memory. Input interacts with cognitive mechanisms to shape learning (e.g., Masek et al., 2021). For example, regularities in the prosody of speech entrain fluctuations in attention at a neural level, which may in turn facilitate learning of the input (e.g., Jones, 1976; Obleser & Kayser, 2019). Similarly, timing in the presentation of novel words affects retention of those words (e.g., Vlach & Johnson, 2013). As such, a focus on the timing of everyday experience can inform the design of controlled lab-based manipulations, and reciprocally, lab-based phenomena can guide inquiry into how events unfold in natural settings.

Timing as One Cue of Many

Of course, word timing does not occur in isolation. Rather, infant language exposure is fundamentally multimodal (e.g., Schroer & Yu, 2022; Suárez-Rivera et al., 2019; Yu et al., 2008). Words accompany contexts rich with a variety of cues that support category learning—including activity structure, superordinate labels, visual inputs, and spatial information (Anderson et al., 2018; Clerkin & Smith, 2022; Lany & Saffran, 2010; LaTourrette & Waxman, 2019; Oakes & Rakison, 2019; Perry et al., 2010; Plunkett et al., 2008; Rakison & Butterworth, 1998; Ross et al., 1986; Roy et al., 2015; Son et al., 2008; Tamis-LeMonda et al., 2019; Waxman & Braun, 2005). For example, caregivers might talk about a variety of foods using both basic-level labels (e.g., 'pasta', 'beef', and 'peas') and superordinate-level labels (i.e., 'food'); the specified food items may be in spatial proximity to each other on a plate, on a kitchen table, during mealtime. Indeed, our exploratory analyses revealed that superordinate labels occurred in close temporal proximity to words within the category. Such regularity indicates that the environment is rife with cues that support learning rather than impoverished (at one extreme) or overly complex (at the other).

Timing as a Broad Informational Cue

Notably, researchers can take different approaches to quantifying the temporal structure of words in speech. For example, a bottom-up approach to word timing (e.g., cluster analysis or natural language processing) could reveal how words from different categories relate to each other thematically. When listening to the sentence "the girl drinks milk," two-year-olds shift their gaze to the image of milk after hearing the word ‘drink’ and before hearing the word ‘milk’ (Mani & Huettig, 2012). Does the temporal structure of words in speech encode associations between foods and feeding verbs or other associative information such as between clothing and body parts? Similarly, the temporal clustering of words could help toddlers acquire more granular categories than we examined, such as ‘fruit’ or ‘reptile’ (rather than food and animal). Questions about how information is encoded in temporal structure extend beyond semantics. How does the timing of speech input convey information about the syntactic structure or phonology of language? Word timing may also facilitate learning about the world more generally, such as science, mathematics, and literacy.

Notably, temporal clustering of like words is an artifact of natural communication. Coherent discourse requires the use of related words in a short period of time, resulting in a pattern of related words clustering close together. Our results indicate that the temporal closeness of related words characterizes speech to infants and may serendipitously create a structure that supports learning. A critical next step is to experimentally manipulate the timing of novel words to test how temporal structure may impact category learning.

Constraints on Generality

Our findings on the temporal structure of infant-directed speech are limited to the population from which our participants were drawn: English-speaking mothers and toddlers in an urban area of the United States. We observed 12-to-24-month-olds based on existing evidence that infants at this age are beginning to form superordinate categories. Mothers were chosen as the target caregiver because they were the primary caregivers of the infants. Studies with children of other ages—either older or younger—or with other caregivers—such as fathers, teachers, or grandparents—may yield different results, although the characteristics of natural, coherent discourse imply high likelihood of replication. Across cultural contexts, toddlers experience wide variation in language input and may also organize words into different categories than examined here. A fuller understanding of early category learning and accompanying linguistic cues would benefit from a deep dive into the taxonomic categories that comprise different languages across different cultural communities.

Conclusion

By testing the temporal latencies among words, we demonstrate that speech to infants provides a reliable structure that may inform on category membership. Our big-data approach to quantifying the temporal structure of language input, offers novel insights and spurs new questions about how the timing of words—together with other behavioral and contextual cues—may facilitate language learning specifically and cognitive development broadly.

References

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. https://doi.org/10.1111/cogs.12612

Anderson, N., Meagher, K., Welder, A., & Graham, S. A. (2018). Animacy cues facilitate 10-month-olds' categorization of novel objects with similar insides. *PloS one*, *13*(11), e0207800. https://doi.org/10.1371/journal.pone.0207800

Arias-Trejo, N., & Plunkett, K. (2009). Lexical–semantic priming effects during infancy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1536), 3633-3647. https://doi.org/10.1098/rstb.2009.0146

Beckage, N., Smith, L., & Hills, T. (2010). Semantic network connectivity is related to vocabulary growth rate in children. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 32, No. 32). https://escholarship.org/uc/item/9xm7v9nq

Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, *22*(1), e12715. https://doi.org/10.1111/desc.12715

Borovsky, A., Ellis, E. M., Evans, J. L., & Elman, J. L. (2016). Lexical leverage: Category knowledge boosts real-time novel word recognition in 2-year-olds. *Developmental Science*, *19*(6), 918-932. https://doi.org/10.1111/desc.12343

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. https://doi.org/10.1073/pnas.2123239119

Collins, A. M., & Loftus, E. F. (1975). A spreading-activation theory of semantic processing. *Psychological Review, 82*(6), 407–428. https://doi.org/10.1037/0033-295X.82.6.407

Custode, S. A., & Tamis-LeMonda, C. (2020). Cracking the code: Social and contextual cues to language input in the home environment. *Infancy*, *25*(6), 809-826. https://doi.org/10.1111/infa.12361

Ebbinghaus, H. (1913). Retention and obliviscence as a function of the time (H. A. Ruger & C. E. Bussenius, Trans.). In H. Ebbinghaus & H. A. Ruger, C. E. Bussenius (Trans.), *Memory: A contribution to experimental psychology* (pp. 62–80). Teachers College Press. https://doi.org/10.1037/10011-007

Fourtassi, A., Bian, Y., & Frank, M. C. (2020). The growth of children's semantic and phonological networks: Insight from 10 languages. *Cognitive Science*, *44*(7), e12847. https://doi.org/10.1111/cogs.12847

Gilkerson, J., Richards, J. A., Warren, S. F., Montgomery, J. K., Greenwood, C. R., Kimbrough Oller, D., ... & Paul, T. D. (2017). Mapping the early language environment using all-day recordings and automated analysis. *American journal of speech-language pathology*, *26*(2), 248-265. https://doi.org/10.1044/2016_AJSLP-15-0169

Hoff-Ginsberg, E. (1991). Mother-child conversation in different social classes and communicative settings. *Child development*, *62*(4), 782-796. https://doi.org/10.1111/j.1467-8624.1991.tb01569.x

Huebner, P. A., & Willits, J. A. (2018). Structured semantic knowledge can emerge automatically from predicting word sequences in child-directed speech. *Frontiers in Psychology*, *9*, 285125. https://doi.org/10.3389/fpsyg.2018.00133

Jiang, H., Frank, M. C., Kulkarni, V., & Fourtassi, A. (2022). Exploring Patterns of Stability and Change in Caregivers' Word Usage Across Early Childhood. *Cognitive Science*, *46*(7), e13177. https://doi.org/10.1111/cogs.13177

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83*(5), 323–355. https://doi.org/10.1037/0033-295X.83.5.323

Lany, J., & Saffran, J. R. (2010). From statistics to meaning: Infants' acquisition of lexical categories. *Psychological Science*, *21*(2), 284-291. https://doi.org/10.1177/0956797609358570

LaTourrette, A., & Waxman, S. R. (2019). A little labeling goes a long way: Semi-supervised learning in infancy. *Developmental Science*, *22*(1), e12736. https://doi.org/10.1111/desc.12736

Leung, A., Tunkel, A., & Yurovsky, D. (2021). Parents fine-tune their speech to children's vocabulary knowledge. *Psychological Science*, *32*(7), 975-984. https://doi.org/10.1177/0956797621993104

Mandler, J. M. (2004). *The foundations of mind: Origins of conceptual thought*. Oxford University Press.

Mandler, J. M., & McDonough, L. (1993). Concept formation in infancy. *Cognitive Development*, *8*(3), 291-318. https://doi.org/10.1016/S0885-2014(93)80003-C

Mani, N., & Huettig, F. (2012). Prediction during language processing is a piece of cake—But only for skilled producers. *Journal of Experimental Psychology: Human Perception and Performance, 38*(4), 843–847. https://doi.org/10.1037/a0029284

Masek, L. R., McMillan, B. T., Paterson, S. J., Tamis-LeMonda, C. S., Golinkoff, R. M., & Hirsh-Pasek, K. (2021). Where language meets attention: How contingent interactions promote learning. *Developmental Review*, *60*, 100961. https://doi.org/10.1016/j.dr.2021.100961

Nelson, K. (1973). Structure and strategy in learning to talk. *Monographs of the society for research in child development*, 1-135. https://doi.org/10.2307/1165788

Oakes, L. M. (2009). The "Humpty Dumpty problem" in the study of early cognitive development: Putting the infant back together again. *Perspectives on Psychological Science*, *4*(4), 352-358. https://doi.org/10.1111/j.1745-6924.2009.01137.x

Oakes, L. M., & Rakison, D. H. (2019). *Developmental cascades: Building the infant mind*. Oxford University Press.

Obleser, J., & Kayser, C. (2019). Neural entrainment and attentional selection in the listening brain. *Trends in cognitive sciences*, *23*(11), 913-926. https://doi.org/10.1016/j.tics.2019.08.004

Poulin-Dubois, D., Graham, S., & Sippola, L. (1995). Early lexical development: The contribution of parental labelling and infants' categorization abilities. *Journal of Child Language*, *22*(2), 325-343. https://doi.org/10.1017/S0305000900009818

Plunkett, K., Hu, J. F., & Cohen, L. B. (2008). Labels can override perceptual categories in early infancy. *Cognition*, *106*(2), 665-681. https://doi.org/10.1016/j.cognition.2007.04.003

Quillian, M. R. (1966). *Semantic memory* (pp. 227-270). Air Force Cambridge Research Laboratories, Office of Aerospace Research, United States Air Force.

Rakison, D. H., & Butterworth, G. E. (1998). Infants' use of object parts in early categorization. *Developmental Psychology*, *34*(1), 49-62. https://doi.org/10.1037/0012-1649.34.1.49

Rakison, D. H., & Oakes, L. M. (Eds.). (2003). *Early category and concept development: Making sense of the blooming, buzzing confusion*. Oxford University Press.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*. MIT press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive psychology*, *8*(3), 382-439. https://doi.org/10.1016/0010-0285(76)90013-X

Rovee-Collier, C. (1995). Time windows in cognitive development. *Developmental Psychology, 31*(2), 147–169. https://doi.org/10.1037/0012-1649.31.2.147

Roy, B. C., Frank, M. C., DeCamp, P., Miller, M., & Roy, D. (2015). Predicting the birth of a spoken word. *Proceedings of the National Academy of Sciences*, *112*(41), 12663-12668. https://doi.org/10.1073/pnas.1419773112

Schroer, S. E., & Yu, C. (2022). The real-time effects of parent speech on infants' multimodal attention and dyadic coordination. *Infancy*, *27*(6), 1154-1178. https://doi.org/10.1111/infa.12500

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. https://doi.org/10.1016/j.cognition.2022.105266

Soderstrom, M., & Wittebolle, K. (2013). When do caregivers talk? The influences of activity and time of day on caregiver speech and child vocalizations in two childcare environments. *PloS One*, *8*(11), e80646. https://doi.org/10.1371/journal.pone.0080646

Spelke, E. S., & Kinzler, K. D. (2007). Core knowledge. *Developmental Science*, *10*(1), 89-96. https://doi.org/10.1111/j.1467-7687.2007.00569.x

Styles, S. J., & Plunkett, K. (2009). How do infants build a semantic system?. *Language and Cognition*, *1*(1), 1-24. https://doi.org/10.1515/LANGCOG.2009.001

Suarez-Rivera, C., Smith, L. B., & Yu, C. (2019). Multimodal parent behaviors within joint attention support sustained attention in infants. *Developmental Psychology, 55*(1), 96–109. https://doi.org/10.1037/dev0000628

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. https://doi.org/10.1111/cdev.13089

Tamis-LeMonda, C. S., Kuchirko, Y., Luo, R., Escobar, K., & Bornstein, M. H. (2017). Power in methods: Language to infants in structured and naturalistic contexts. *Developmental science*, *20*(6), e12456. https://doi.org/10.1111/desc.12456

Trueswell, J. C., Lin, Y., Armstrong III, B., Cartmill, E. A., Goldin-Meadow, S., & Gleitman, L. R. (2016). Perceiving referential intent: Dynamics of reference in natural parent–child interactions. *Cognition*, *148*, 117-135. https://doi.org/10.1016/j.cognition.2015.11.002

Unger, L., Yim, H., Savic, O., Dennis, S., & Sloutsky, V. M. (2023). No frills: Simple regularities in language can go a long way in the development of word knowledge. *Developmental science*, *26*(4), e13373. https://doi.org/10.1111/desc.13373

Vlach, H. A., & Johnson, S. P. (2013). Memory constraints on infants' cross-situational statistical learning. *Cognition*, *127*(3), 375-382. https://doi.org/10.1016/j.cognition.2013.02.015

Wojcik, E. H. (2018). The development of lexical–semantic networks in infants and toddlers. *Child Development Perspectives*, *12*(1), 34-38. https://doi.org/10.1111/cdep.12252

Yu, C., & Smith, L. B. (2016). The social origins of sustained attention in one-year-old human infants. *Current biology*, *26*(9), 1235-1240. https://doi.org/10.1016/j.cub.2016.03.026

Yu, C., Smith, L. B., & Pereira, A. F. (2008). Grounding word learning in multimodal sensorimotor interaction. In *Proceedings of the annual meeting of the cognitive science society* (Vol. 30, No. 30). https://escholarship.org/uc/item/4j4897rt