

BRIEF RESEARCH REPORT

Frequent vs. infrequent words shape toddlers' real-time sentence comprehension

Christine E. POTTER^{1,2}  and Casey LEW-WILLIAMS²

¹Department of Psychology, University of Texas at El Paso, USA

²Department of Psychology, Princeton University, USA

Corresponding author: Christine E. Potter; Email: cepotter2@utep.edu

(Received 28 February 2022; revised 22 May 2023; accepted 27 May 2023)

Abstract

We examined how noun frequency and the typicality of surrounding linguistic context contribute to children's real-time comprehension. Monolingual English-learning toddlers viewed pairs of pictures while hearing sentences with typical or atypical sentence frames (*Look at the...* vs. *Examine the...*), followed by nouns that were higher- or lower-frequency labels for a referent (*horse* vs. *pony*). Toddlers showed no significant differences in comprehension of nouns in typical and atypical sentence frames. However, they were less accurate in recognizing lower-frequency nouns, particularly among toddlers with smaller vocabularies. We conclude that toddlers can recognize nouns in diverse sentence contexts, but their representations develop gradually.

Keywords: language development; language processing; statistical learning; word representations; frequency effects

Introduction

Theories of statistical learning posit that infants track basic information about the words they hear, such as their frequency and their co-occurrence with nearby words, which provide the foundation for more complex knowledge (e.g., Lany, 2014; Saffran, 2003; Thiessen & Erickson, 2013). If this is true, then children must be able to navigate the complexities of incoming speech in real time. Speakers have many options available to them when talking to children, and children therefore encounter words more or less frequently in a variety of sentence contexts. For example, a parent trying to direct a child's attention might say *Look at the horse*, but theoretically they could also say *Examine the pony*; *Behold the equine*; or *This thing is a beast*, among endless options. Yet, what we know about children's early language processing is based almost entirely on frequent, prototypical sentences. By exploring how comprehension may differ for typical vs. atypical phrases, we will be positioned better to evaluate how statistical regularities in children's environments shape their learning and understanding of language.

From a young age, infants show a capacity to learn how words co-occur in their input. Infants detect and recognize frequent combinations of words (Mintz, 2003; Skarabela

© The Author(s), 2023. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

et al., 2021), and children and adults are better able to process, remember, and repeat frequent phrases (Arnon & Clark, 2011; Arnon & Snider, 2010; Bannard & Matthews, 2008). This suggests that learners are proficient in using their knowledge of common patterns from their language experiences, and in fact, infants who show greater sensitivity to statistical patterns have been shown to have stronger language skills (Lany, 2014). It has also been found that highly typical sentence frames, such as *Look at the...*, support toddlers' ability to recognize subsequent words, compared to when words appear in isolation or occur after the word *Look* by itself (Fernald & Hurtado, 2006; Morini & Newman, 2019). Sentences frames have the potential to provide the child with phonetic, syntactic, and/or semantic information about an upcoming referent, and past studies have shown that young children experience disruptions in their processing when provided with inconsistent cues (e.g., Lew-Williams & Fernald, 2007; Mahr et al., 2015; Reuter et al., 2019). This work shows that infants use co-occurrence statistics to inform their comprehension in real time and reveals how they may benefit from hearing words in contexts that are most typical of their prior experience.

While infants also show broad sensitivity to frequency, there is less direct evidence concerning how they use frequency information in real time. Children learn frequent words at younger ages (Ambridge et al., 2015; Goodman et al., 2008), showing that they benefit from simply encountering words often. Both children and adults consistently show more efficient processing of frequent words (e.g., Garlock et al., 2001), perhaps because the representations of these words are more accessible, or because they have heard these words in more diverse contexts (see Goodman et al., 2008). In addition, bilingual toddlers tend to show enhanced understanding of words produced in the language they hear more often, and they have greater difficulty understanding words in their less-frequently-heard language (Hurtado et al., 2014; Morini & Newman, 2019). More specifically, bilingual toddlers have greater difficulty understanding words in their weaker language when those words appear in atypical sentence contexts (Potter et al., 2019). This pattern of results illustrates how children's comprehension can be influenced by both the frequency of individual words and the typicality of the context in which they appear. Thus, a statistical learning framework that spans multiple words in a sentence offers a tool for understanding children's real-time processing of diverse sentences (Lany et al., 2018).

To test relative contributions of frequency and co-occurrence statistics, we investigated monolingual English-learning toddlers' comprehension of frequent vs. infrequent target words that appeared in typical vs. atypical sentence contexts. Using the looking-while-listening procedure, we monitored toddlers' eye movements as they viewed pairs of familiar images and heard sentences labeling one object (Fernald et al., 2008). Target nouns were high- or low-frequency labels for the same objects (e.g., *horse* vs. *pony*; *plate* vs. *dish*). Toddlers were tested on their understanding of these nouns following typical or atypical sentence frames (e.g., *Look at the...* vs. *Examine the...*; *Do you like the...* vs. *Would you prefer the...*). Our hypothesis was that toddlers' comprehension would be affected by both word frequency and the typicality of the preceding sentence context. Specifically, we predicted that toddlers would show weaker comprehension of less (vs. more) frequent words, particularly if they occurred in atypical (vs. typical) sentence contexts. We chose to use sentences that included plausible, but unlikely, phrases for toddlers to hear, reasoning that if toddlers' comprehension depends on the statistics of their input, low-probability sentence frames might derail their comprehension. To understand how toddlers' growing knowledge of words might affect their comprehension, we also tested whether toddlers' vocabulary skills predicted the strength of these effects.

Method

Participants

The final sample included 34 full-term typically developing monolingual English-learning toddlers of ages 21 to 27 months (15 female, $M = 23.7$ months, $SD = 2.1$) from the Northeastern United States, matched to the age of bilingual participants in prior studies (Morini & Newman, 2019; Potter et al., 2019). Parents of all participants provided informed consent, and all procedures were approved by the Princeton University Institutional Review Board. An *a priori* power analysis (Faul et al., 2007) indicated this sample size would have .81 power to detect a medium main effect ($f = .25$). Eighteen additional toddlers were tested but excluded for reported language delay (1), fussiness (9), equipment error (2), failure to contribute at least two usable trials in all conditions (5), or an ill-timed fire alarm (1).

Stimuli

Auditory stimuli

Auditory stimuli were produced by a female native English speaker using child-directed speech. Each sentence included a Typical or Atypical sentence frame (Table 1) and a High- or Low-Frequency target noun.

Table 1. Sentence frames

Typical	Atypical
<i>Do you like the...?</i>	<i>Would you prefer the...?</i>
<i>Look at the...!</i>	<i>Examine the...!</i>
<i>Can you find the...?</i>	<i>Could you locate the...?</i>
<i>Show me the...!</i>	<i>Identify the ...!</i>

Target nouns were selected based on what children of this age would be expected to know, according to norms for the MacArthur-Bates Communicative Development Inventory: Words and Sentences (MCDI; Fenson et al., 2007). We chose items that were reportedly produced by at least 30% of 24-month-olds and more than 50% of 30-month-olds to ensure that these were likely to be familiar words, and parental reports confirmed that the majority of children understood both the High- and Low-Frequency labels. For all items, data from the CHILDES corpus indicated that High-Frequency labels were at least twice as frequent as Low-Frequency labels in North American child-directed speech (Table 2).

Each of 16 target nouns occurred once following a Typical frame and once following an Atypical frame. Sentences were edited using Praat (Boersma & Weenink, 2016) in order to match sentences in intensity (65db) and to standardize the duration of all frames (1151ms) and target nouns (665ms), such that they were identical in length.

Visual stimuli

Visual stimuli were pairs of images of familiar objects presented side-by-side on grey backgrounds (Figure 1). Images always occurred in yoked pairs matched for animacy and salience (*horse/bird, plate/light, candy/pants, sheep/chicken*). Each visual image could be

Table 2. Target Nouns

High-Frequency Nouns				Low-Frequency Nouns				
Label	Parent-reported comprehension (%) ^a	WordBank Norm (%) ^b	CHILDES Freq ^c	Label	Parent-reported comprehension (%) ^a	WordBank Norm (%) ^b	CHILDES Freq ^c	CHILDES ratio ^d
Horse	100	79	403	Pony	59	31	55	7.3
Bird	100	88	772	Goose	74	35	85	9.1
Plate	100	57	84	Dish	65	33	37	2.3
Light	100	77	564	Lamp	74	31	21	26.9
Candy	65	66	68	Lollipop	68	32	23	3.0
Pants	100	69	264	Jeans	65	34	22	12.0
Sheep	97	60	351	Lamb	79	39	149	2.4
Chicken	97	66	320	Rooster	76	32	113	2.8
<i>Average</i>	<i>95</i>	<i>70</i>	<i>353</i>		<i>70</i>	<i>33</i>	<i>63</i>	<i>8.2</i>

^aPercentage of children in the current study who were reported to understand the label

^bPercentage of 24-month-old American English-learning infants reported to produce the label, according to MCDI norms (Frank et al., 2017)

^cFrequency (per million words) in the CHILDES corpus for North American English (MacWhinney, 2000)

^dRatio of occurrence of the High- vs. Low-frequency label in the CHILDES corpus

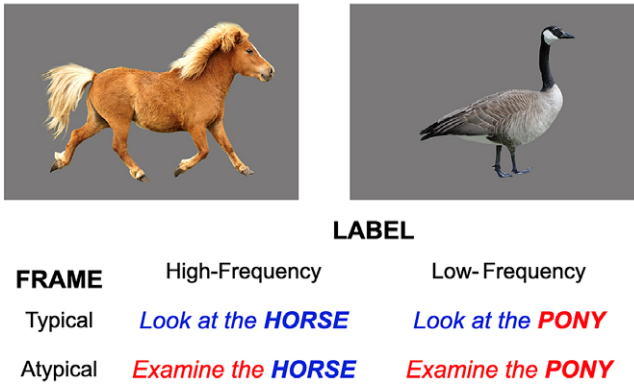


Figure 1. Sample stimuli.

Note. Sample sentences provide examples of all four trial types where *horse* is the target image. Blue text represents Typical sentences frames and High-Frequency nouns; red indicates Atypical sentences frames and Low-Frequency target nouns.

labeled with equal plausibility by a High- and Low-Frequency noun. All tokens appeared equally often on the right and left side and equally often as the target and distracter. All stimuli are available through OSF (<https://osf.io/frvx7/>).

Procedure

Toddlers sat on their parents' laps in a sound-attenuated, darkened room. Parents wore opaque glasses and were asked not to interfere. Visual stimuli appeared on a 55" monitor, and speech stimuli played over a loudspeaker. On each trial, images were presented in silence for 2s. Participants then heard a sentence labeling one object, followed by a sentence used to maintain attention (e.g., *Look at the horse! Do you see it?*) and 1.5s of silence, for a total trial duration of 6.7s.

Testing sessions were divided into four 8-trial blocks. Within each block, toddlers heard one type of sentence frame (Typical or Atypical), and half of trials involved High-Frequency target nouns. Toddlers heard alternating blocks of Typical vs. Atypical sentences, and we counterbalanced across participants which type of frame occurred first. Between each block, toddlers saw a filler trial to engage interest. Trial orders were pseudorandomized such that there were never more than three consecutive trials with the same type of target noun (High- or Low-Frequency), and toddlers were randomly assigned to one of four experimental orders.

Following testing, all parents reported whether their children understood and said each target noun. Thirty two of 34 parents also filled out the MCDI to provide a measure of children's vocabulary ($M = 299$ words, range = 11-582, $SD = 176$).

Coding

Trained coders, blind to condition, recorded at 33ms intervals whether the child was looking right, left, or at neither image. We excluded trials where the child was not looking at either image at noun onset or for more than 500ms continuously during our

analysis window, resulting in an average of 19.9 usable trials per participant. Twenty percent of sessions were re-coded by a second coder. Coders agreed on the gaze location on 99% of frames and agreed within a single frame on 98% of frames surrounding shift events.

Results

To evaluate children's recognition of words and their referents, we computed toddlers' mean accuracy, defined as the proportion of time looking at the target image divided by time viewing either image during the window from 367-2000ms following the onset of the target noun. All data and supplemental analyses can be found at <https://osf.io/frvx7/>.

For all trial types, single-sample two-tailed *t*-tests revealed that toddlers significantly exceeded chance (.5) in looking to the target ($M = .66$, all $p < .00001$, Table 3), demonstrating successful comprehension, on average, of both frequent and infrequent labels across different sentences.

We then tested the influences of sentence typicality and word frequency using a 2x2 within-subjects ANOVA (Sentence Frame: Typical vs. Atypical, Target Noun: High- vs. Low-Frequency, Figure 2). The effect of Frame was not significant, $F(1,33) = 0.25$, $p = .62$, and there was no interaction between Frame and Noun, $F(1,33) = 0.001$, $p = .97$, indicating that toddlers showed no significant difference in looking to the target following Typical or Atypical sentence frames. However, there was a significant main effect of Target Noun, $F(1,33) = 4.36$, $p = .04$, $\eta_p^2 = .12$; accuracy was higher for High-Frequency nouns. Exploratory analyses revealed no significant order effects or changes across blocks (see Figures S1-S3 in supplementary materials). Thus, while toddlers recognized both High- and Low-Frequency labels in real time, they showed better understanding of frequent words.

Finally, we conducted exploratory analyses to test associations between word comprehension and vocabulary knowledge. MCDI scores were not significantly correlated with overall accuracy, $r = .17$, $p = .34$, 95% $CI = [-.19, .49]$, and toddlers showed above-chance comprehension even of the words that their parents did not report they knew, $M = .59$, $t(22) = 2.61$, $p = .02$. We then tested how toddlers' vocabulary related to their sensitivity to word frequency. To do so, we calculated a difference score for each toddler that captured the difference between their accuracy on trials with High- vs. Low-Frequency targets. Difference scores were significantly correlated with vocabulary size, $r = -.38$, $p = .03$, 95% $CI = [-.64, -.038]$, indicating that toddlers with larger vocabularies showed smaller discrepancies in recognizing High- vs. Low-Frequency words (see Figure S5).

Table 3. Mean performance and comparison to chance (.5) across conditions

Sentence Frame	Target Noun	Mean accuracy	SD	<i>t</i>	95% CI
Typical	High-Frequency	.689	.111	9.90	[.65, .73]
Typical	Low-Frequency	.645	.143	5.89	[.59, .69]
Atypical	High-Frequency	.677	.130	7.94	[.63, .72]
Atypical	Low-Frequency	.634	.142	5.50	[.58, .68]

Note: Performance across all trial types significantly exceeded chance, $p < .00001$

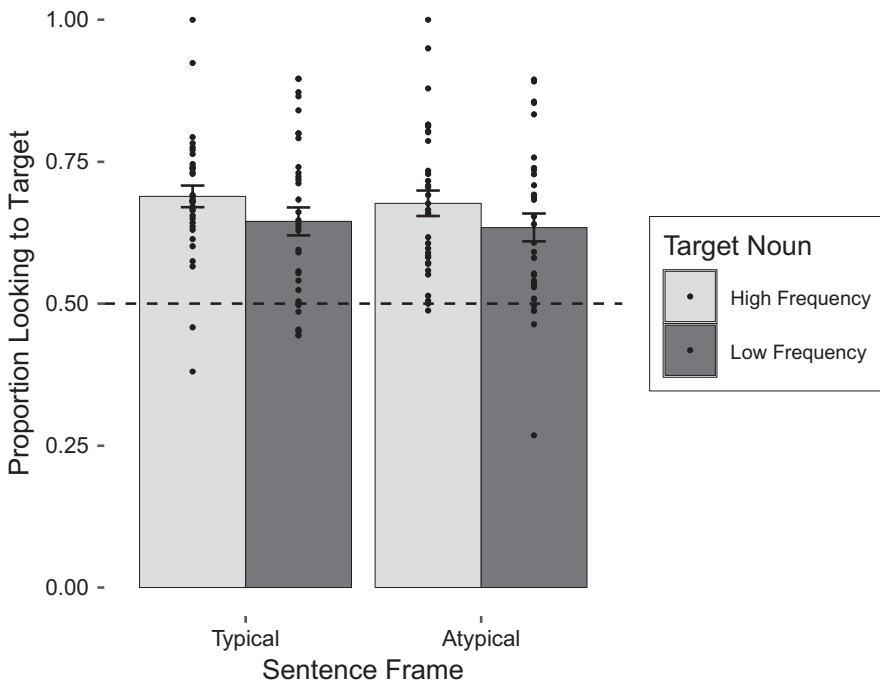


Figure 2. Toddlers' mean accuracy in looking to the target.

Note. Accuracy was computed from 367 to 2000ms following noun onset. Each point represents the mean for an individual subject. Dashed line indicates chance (.5). Error bars denote standard errors of the mean.

Discussion

This study tested monolingual toddlers' sensitivity to the statistical patterns in their language input by exploring their understanding of frequent vs. infrequent target nouns that appeared in typical vs. atypical sentence frames. Toddlers successfully demonstrated comprehension of all sentence types, but they were less accurate in recognizing low-frequency words, and differences were especially pronounced for children with smaller vocabularies. These results yielded two important insights. First, toddlers' real-time comprehension is flexible enough to contend with sentence contexts that involve relatively unfamiliar sequences of words – at least in a controlled lab setting. Second, frequency information affects children's ability to recognize nouns, indicating that the robustness of children's knowledge of a target word is a key factor in their real-time comprehension.

Across different sentences, toddlers displayed better understanding of nouns that are more common in child-directed speech, illustrating how toddlers' ability to track frequency information has a measurable effect on their real-time comprehension. Interestingly, the relation we observed between vocabulary size and the difference in recognizing high- vs. low-frequency nouns tentatively suggests that when children are in earlier periods of vocabulary growth – or perhaps when they have received less input from caregivers – they may be particularly sensitive to how often they have encountered a given word. That is, frequency effects in word recognition may be especially pronounced when a child has not yet had the chance to form a robust representation. Crucially, high- and low-

frequency labels referred to identical visual tokens, ensuring that differences were attributable to familiarity with words, not objects (Kartushina & Mayor, 2019). This imbalance mirrors data from adults showing that word frequency reliably predicts comprehension across a variety of tasks (Ellis, 2002; Gerhand & Barry, 1998; Morton, 1969) and is consistent with studies showing that children's ability to correctly select a label's referent depends on their familiarity with that word (Kucker et al., 2018).

Unexpectedly, the typicality or atypicality of the sentence frame had no observable effect on toddlers' comprehension. That is, even when hearing an unusual frame, toddlers recognized the noun that followed, suggesting that they were relatively insensitive to the unexpectedness of the atypical sentences that we tested (see Figure S2). This does not mean that toddlers are *unaffected* by prior linguistic context. Indeed, studies suggest that toddlers exploit semantic, morphosyntactic, and phonetic cues in real-time processing (Borovsky et al., 2012; Fernald et al., 2008; Lew-Williams & Fernald, 2007; Mahr et al., 2015), so familiar linguistic contexts could have facilitated downstream recognition, or the atypical sentence frames could have impeded their comprehension. One possibility for why we did not find a significant effect of sentence frame is that the frames were not sufficiently different in this experimental environment. For our atypical sentences, we chose to use frames that are quite uncommon in child-directed speech, yet maintained reasonable naturalness. For example, all sentences included natural co-articulation and prosody, which play a role in supporting more seamless processing of upcoming words (e.g., De Carvalho et al., 2017; Mahr et al., 2015; Paquette-Smith et al., 2016). The noun was also preceded by the determiner *the* in every sentence, and there is good evidence that children are sensitive to determiners and can use them to facilitate processing (Höhle et al., 2004; Kedar et al., 2006; Lew-Williams & Fernald, 2007). Another explanation is that infants can perhaps 'listen through' uninformative or unusual parts of a sentence (Thorpe & Fernald, 2006) and attend primarily to target nouns, or other meaningful information that arrives downstream. None of our sentence frames, either Typical or Atypical, reliably predicted any particular referent, therefore any of the subsequent target nouns were equally likely or unlikely to follow. Future studies will continue to examine whether infants benefit from hearing words in predictable contexts, as well as exploring gradients of familiarity (i.e., highly probable vs. possible vs. unlikely vs. impossible sentences), which will inform current understanding of how they exploit different types of prior experience and what kinds of regularities most strongly shape infants' comprehension.

Importantly, while toddlers showed reduced understanding of low-frequency nouns, they still recognized more than one label for the identical visual token. This contrasts with views of early word learning proposing that young children are unwilling to accept multiple names for a single referent (Markman & Wachtel, 1988) and provides additional evidence that toddlers, like older children, recognize that the same object can be labeled in more than one way (Bergelson & Swingley, 2013; Waxman & Hatch, 1992).

These results also converge with recent research on early bilingualism. Monolingual toddlers' difficulty understanding low-frequency words is analogous to bilingual toddlers' reduced comprehension of words in their weaker and/or less-frequently-heard language (Hurtado et al., 2014; Potter et al., 2019). The current results, in conjunction with findings from bilingual participants, illustrate that children from different language backgrounds construct their knowledge through experience with words encountered more and less frequently, and suggest that a statistical learning framework can unite research on early learning across monolingual and bilingual environments.

While the current study provides evidence that toddlers are sensitive to overall frequency information, we relied on general estimates of how often words are used, and normative data may not reflect a given child's experience. That is, some children inevitably had more experience than others with certain words. Future studies could provide a stronger test of the types of statistics that children use to guide their language comprehension by either collecting detailed measures of their experiences or by controlling the frequency of words in different contexts in an experimental design. With such tailored measures and structured manipulations, it might be possible to capture subtle effects, such as differences in reaction time, that could provide insight into the efficiency with which infants use different cues (see Figure S6 for exploratory analyses involving reaction times). Moreover, if we had been able to include more trials for each item, it might have been easier to detect influences of linguistic context and of particular nouns (see Table S1). Future studies should collect sufficient data to explore individual differences at both the child and the word level, perhaps by testing children across multiple sessions. More generally, future studies that are tailored to children's interests and to their experiences with uncommon words and constructions could help us understand the idiosyncratic nature of their developing knowledge.

Broadly, these results demonstrate how the distributional statistics of children's language input shape real-time comprehension and offer a potential mechanism underlying reported links between the amount of language children hear in their everyday lives and their language skills (Weisleder & Fernald, 2013). We often think of input in an aggregated sense, but when a child hears more TOTAL words compared to another child, they will also likely hear individual words more often (at least on average). This increased word-level experience could support the development of robust representations that enhance comprehension. Thus, we suggest that statistical learning principles can explain how children understand frequent and infrequent words in their input. Namely, over time, children's representations strengthen as they encounter the same words in new contexts.

Acknowledgments. We would like to thank the participating families and members of the Princeton Baby Lab, especially Kennedy Casey, Ellie Breitfeld, Katie Vasquez, Rinat Tal, and Ariella Cohen. This work was funded by grants from NICHD (F32 HD093139, R01HD095912) and the James S. McDonnell Foundation. Stimuli, data, and supplemental materials are available at <https://osf.io/frvx7/>.

Competing interest. The authors declare none.

References

- Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(2), 239–273.
- Arnon, I., & Clark, E. V. (2011). Why brush your teeth is better than teeth—Children's word production is facilitated in familiar sentence-frames. *Language Learning and Development*, *7*(2), 107–129.
- Arnon, I., & Snider, N. (2010). More than words: Frequency effects for multi-word phrases. *Journal of Memory and Language*, *62*(1), 67–82.
- Bannard, C., & Matthews, D. (2008). Stored word sequences in language learning: The effect of familiarity on children's repetition of four-word combinations. *Psychological Science*, *19*(3), 241–248.
- Bergelson, E., & Swingle, D. (2013). Young toddlers' word comprehension is flexible and efficient. *PLoS one*, *8*(8), e73359.
- Boersma, P., & Weenink, D. (2016). Praat: Doing phonetics by computer (Version 6.0.14).

- Borovsky, A., Elman, J. L., & Fernald, A.** (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, *112*(4), 417–436.
- De Carvalho, A., Dautriche, I., Lin, I., & Christophe, A.** (2017). Phrasal prosody constrains syntactic analysis in toddlers. *Cognition*, *163*, 67–79.
- Ellis, N. C.** (2002). Frequency effects in language processing: A review with implications for theories of implicit and explicit language acquisition. *Studies in second language acquisition*, *24*(2), 143–188.
- Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A.** (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191.
- Fenson, L., Bates, E., Dale, P. S., Marchman, V. A., Reznick, J. S., & Thal, D. J.** (2007). *MacArthur-Bates Communicative Development Inventories*. Paul H. Brookes Publishing Company.
- Fernald, A., & Hurtado, N.** (2006). Names in frames: Infants interpret words in sentence frames faster than words in isolation. *Developmental Science*, *9*(3).
- Fernald, A., Zangl, R., Portillo, A. L., & Marchman, V. A.** (2008). Looking while listening: Using eye movements to monitor spoken language. *Developmental Psycholinguistics: On-line methods in children's language processing*, *44*, 97.
- 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.
- Garlock, V. M., Walley, A. C., & Metsala, J. L.** (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language*, *45*(3), 468–492.
- Gerhand, S., & Barry, C.** (1998). Word frequency effects in oral reading are not merely age-of-acquisition effects in disguise. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(2), 267.
- Goodman, J. C., Dale, P. S., & Li, P.** (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531.
- Höhle, B., Weissenborn, J., Kiefer, D., Schulz, A., & Schmitz, M.** (2004). Functional elements in infants' speech processing: The role of determiners in the syntactic categorization of lexical elements. *Infancy*, *5*(3), 341–353.
- Hurtado, N., Grüter, T., Marchman, V. A., & Fernald, A.** (2014). Relative language exposure, processing efficiency and vocabulary in Spanish–English bilingual toddlers. *Bilingualism: Language and Cognition*, *17*(1), 189–202.
- Kartushina, N., & Mayor, J.** (2019). Word knowledge in six-to nine-month-old Norwegian infants? Not without additional frequency cues. *Royal Society Open Science*, *6*, 180711.
- Kedar, Y., Casasola, M., & Lust, B.** (2006). Getting there faster: 18- and 24-month-old infants' use of function words to determine reference. *Child Development*, *77*(2), 325–338.
- Kucker, S. C., McMurray, B., & Samuelson, L. K.** (2018). Too much of a good thing: How novelty biases and vocabulary influence known and novel referent selection in 18-month-old children and associative learning models. *Cognitive Science*, *42*, 463–493.
- Lany, J.** (2014). Judging words by their covers and the company they keep: Probabilistic cues support word learning. *Child Development*, *85*(4), 1727–1739.
- Lany, J., Shoaib, A., Thompson, A., & Graf Estes, K.** (2018). Infant statistical-learning ability is related to real-time language processing. *Journal of Child Language*, *45*(2), 368–391.
- Lew-Williams, C., & Fernald, A.** (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, *18*(3), 193–198.
- MacWhinney, B.** (2000). The CHILDES project: Tools for analyzing talk (third edition): Volume I: Transcription format and programs, Volume II: The database. *Computational Linguistics*, *26*(4), 657.
- Mahr, T., McMillan, B. T. M., Saffran, J. R., Ellis Weismer, S., & Edwards, J.** (2015). Anticipatory coarticulation facilitates word recognition in toddlers. *Cognition*, *142*, 345–350.
- Markman, E. M., & Wachtel, G. F.** (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive Psychology*, *20*(2), 121–157.
- Mintz, T. H.** (2003). Frequent frames as a cue for grammatical categories in child directed speech. *Cognition*, *90*(1), 91–117.
- Morini, G., & Newman, R. S.** (2019). Dónde está la ball? Examining the effect of code switching on bilingual children's word recognition. *Journal of Child Language*, *46*(6), 1238–1248.
- Morton, J.** (1969). Interaction of information in word recognition. *Psychological Review*, *76*(2), 165.

- Paquette-Smith, M., Fecher, N., & Johnson, E. K.** (2016). Two-year-olds' sensitivity to subphonemic mismatch during online spoken word recognition. *Attention, Perception, & Psychophysics*, **78**(8), 2329–2340.
- Potter, C. E., Fourakis, E., Morin-Lessard, E., Byers-Heinlein, K., & Lew-Williams, C.** (2019). Bilingual toddlers' comprehension of mixed sentences is asymmetrical across their two languages. *Developmental Science*, **22**(4), e12794.
- Reuter, T., Borovsky, A., & Lew-Williams, C.** (2019). Predict and redirect: Prediction errors support children's word learning. *Developmental Psychology*, **55**, 1656–1665.
- Saffran, J. R.** (2003). Statistical language learning: Mechanisms and constraints. *Current Directions in Psychological Science*, **12**, 110–114.
- Skarabela, B., Ota, M., O'Connor, R., & Arnon, I.** (2021). 'Clap your hands' or 'take your hands'? One-year-olds distinguish between frequent and infrequent multiword phrases. *Cognition*, **211**, 104612.
- Thiessen, E. D., & Erickson, L. C.** (2013). Beyond word segmentation: A two-process account of statistical learning. *Current Directions in Psychological Science*, **22**(3), 239–243.
- Thorpe, K., & Fernald, A.** (2006). Knowing what a novel word is not: Two-year-olds 'listen through' ambiguous adjectives in fluent speech. *Cognition*, **100**, 389–433.
- Waxman, S. R., & Hatch, T.** (1992). Beyond the basics: Preschool children label objects flexibly at multiple hierarchical levels. *Journal of Child Language*, **39**, 153–166.
- Weisleder, A., & Fernald, A.** (2013). Talking to children matters: early language experience strengthens processing and builds vocabulary. *Psychological Science*, **24**(11), 2143–2152. <https://doi.org/10.1177/0956797613488145>

Cite this article: Potter C.E., & Lew-Williams C. (2023). Frequent vs. infrequent words shape toddlers' real-time sentence comprehension. *Journal of Child Language* 1–11, <https://doi.org/10.1017/S0305000923000387>