Journal of Child Language (2021), 1–15
doi:10.1017/S0305000921000040

CAMBRIDGE UNIVERSITY PRESS

BRIEF RESEARCH REPORT

An investigation of iconic language development in four datasets

David M. SIDHU[a], Jennifer WILLIAMSON[a], Velina SLAVOVA[b], and Penny M. PEXMAN[a]

[a]University of Calgary, Canada and [b]New Bulgarian University, Bulgaria
Corresponding author: David M. Sidhu E-mail: dmsidhu@gmail.com

(Received 8 April 2020; revised 19 November 2020; accepted 3 January 2021)

Abstract

Iconic words imitate their meanings. Previous work has demonstrated that iconic words are more common in infants' early speech, and in adults' child-directed speech (e.g., Perry et al., 2015; 2018). This is consistent with the proposal that iconicity provides a benefit to word learning. Here we explored iconicity in four diverse language development datasets: a production corpus for infants and preschoolers (MacWhinney, 2000), comprehension data for school-aged children to young adults (Dale & O'Rourke, 1981), word frequency norms from educational texts for school aged children to young adults (Zeno et al., 1995), and a database of parent-reported infant word production (Frank et al., 2017). In all four analyses, we found that iconic words were more common at younger ages. We also explored how this relationship differed by syntactic class, finding only modest evidence for differences. Overall, the results suggest that, beyond infancy, iconicity is an important factor in language acquisition.

Introduction

One of the fundamental assumptions of language is that the linguistic sign is arbitrary (Hockett, 1963). Until recently it was largely believed that the relationship between a linguistic form and its meaning was based on convention. However, non-arbitrary mappings between form and meaning are far more widespread than previously believed (Aryani, Conrad, Schmidtke & Jacobs, 2018; Blasi, Wichmann, Hammarstrom, Stadler & Christiansen, 2016; Monaghan, Shillcock, Christiansen & Kirby, 2014; Sidhu, Westbury, Hollis, & Pexman, in press; see also Perniss, Thompson & Vigliocco, 2010). One example of non-arbitrariness is iconicity: the resemblance between form and meaning in language (for reviews see Dingemanse, Blasi, Lupyan, Christiansen & Monaghan, 2015; Dingemanse, Perlman & Perniss, 2020). The most common instance of iconicity in spoken language is onomatopoeia, in which the form of a word directly resembles the sound to which it refers (e.g., *woof*, *mumble*). Spoken language can also be iconic by way of SOUND SYMBOLISM: associations between certain phonemes and particular perceptual and/or semantic features (e.g., between /b/ or /u/ and roundness; for reviews see Lockwood & Dingemanse, 2015; Sidhu & Pexman, 2018a). That is, when a word evokes

associations that resemble its meaning, it could also be considered iconic (e.g., *balloon* and roundness). In this paper, we use *iconicity* to refer to any resemblance-based mapping between a word's form and its meaning (Dingemanse et al., 2015).

There is considerable evidence for the existence of iconicity in languages (Blasi et al., 2016; Dingemanse et al., 2015; Perniss et al., 2010). There is also evidence that this iconicity has consequences for language processing and for language acquisition. For example, experiments have demonstrated that both adults and children respond differently to iconic vs. arbitrary words (e.g., Kanero, Imai, Okuda, Okada & Matsuda, 2014; Sidhu, Vigliocco, & Pexman, 2020). Other experiments have found that adults, children, and pre-linguistic infants are sensitive to sound symbolism (Asano, Imai, Kita, Kitajo, Okada & Thierry, 2015; Imai, Kita, Nagumo & Okada, 2008; Imai, Miyazaki, Yeung, Hidaka, Kantartzis, Okada & Kita, 2015; Kovic, Plunkett & Westermann, 2010; Maurer, Pathman & Mondloch, 2006; Ozturk, Krehm & Vouloumanos, 2013; Yang, Asano, Kanazawa, Yamaguchi & Imai, 2019), while Perniss, Lu, Morgan and Vigliocco, (2018) found that iconicity was elaborated in sign language directed to children. Importantly, there is evidence that iconicity is more prevalent in early-acquired signed and spoken language (Perry et al., 2015; Thompson, Vinson, Woll & Vigliocco, 2012; Vinson, Cormier, Denmark, Schembri & Vigliocco, 2008). For example, Perry, Perlman and Lupyan (2015) collected iconicity ratings of ~600 English words that are commonly produced in the first three years of life. Participants rated the extent to which the words' sounds matched their meanings (see Methods for further details). The authors found a negative association between iconicity and each word's production frequency at 30 months, even after excluding onomatopoeia from the analysis and controlling for other factors related to age of acquisition. Thus, the earliest-learned words also tended to be the most iconic.

Several studies have replicated and extended this work (Hinojosa, Haro, Magallares, Duñabeitia & Ferr, 2020; for a review see Laing, 2019). For example, Perry, Perlman, Winter, Massaro and Lupyan (2018) found that conversations between adults were characterized by the use of low-iconicity, highly arbitrary words, whereas early parent-child conversations were rich in highly iconic words. Further, low-iconicity words increased in frequency as a function of age, suggesting that children gradually shift toward an adult style of speech by using more arbitrary words. Meanwhile, Massaro and Perlman (2017) examined the average iconicity of children's vocabularies during the first four years of life and found that iconicity gradually decreased with age, starting at 12 months. Thus, children's speech is initially characterized by frequent use of high-iconicity words, but arbitrary words start to dominate after 12 months. Finally, Perlman, Fusaroli, Fein and Naigles (2017) performed a longitudinal analysis of parent-child dyads, and found that the average iconicity of both children's and parents' speech was initially high but gradually decreased between the ages of 12 and 69 months. However, children tended to increase their use of arbitrary words, while parents decreased their use of iconic words. Changes in iconicity were also moderated by individual differences in children's cognitive skills, suggesting that parents tailored their use of iconicity to their child's skill level.

Iconicity and word learning

Taken together, the results from these studies demonstrate that high-iconicity words are both acquired earlier and used more often by infants and young children. Imai and Kita

(2014) suggested that iconicity might help infants map speech sounds onto referents and constrain a word's potential meaning (i.e., Imai & Kita, 2014; Perniss & Vigliocco, 2014). This is what Nielsen and Dingemanse (2020) termed LOCAL LEARNING ENHANCEMENT (i.e., iconic words are easier to learn), in contrast to GENERAL LEARNING ENHANCEMENT (i.e., learning iconic words makes all words easier to learn). Their characterization of the literature is that there is a good deal of evidence for local enhancement, but no clear evidence for general enhancement.

If iconicity does provide local enhancement of language acquisition, there should be evidence of a learning advantage for iconic words, especially in early vocabulary (Perlman et al., 2017). Indeed, word-learning experiments have demonstrated that, in some cases, iconic words have a learning and processing advantage over non-iconic words (Imai et al., 2008; Kanero et al., 2014; Kovic et al., 2010; Laing, 2017; Lockwood, Dingemense & Hagoort, 2016; Nielsen & Rendall, 2012). Meanwhile, indirect evidence for the role of iconicity in word learning comes from the high proportion of iconic words in early parent-child conversations and the fact that iconic words tend to be acquired earlier (Perlman et al., 2017; Perry et al., 2018).

While there is evidence of a learning advantage for iconic words, there remains a great deal of work to be done in understanding the developmental trajectory of this advantage (see Nielsen & Dingemanse, 2020). For instance, a study by Massaro and Perlman (2017) found that the first ~60 words understood by children are not rated as being high in iconicity. In addition, Tzeng, Nygaard and Namy (2017) reported that 3-year-olds were not sensitive to the sound symbolism of existing foreign words. This suggests that effects of iconicity may not be consistent throughout development and also highlights the need for various kinds of data beyond production.

One goal of the present study was to help piece together the role of iconicity across different stages in development. While previous research has examined the frequency of iconic words in language used by and produced for infants and preschoolers, here we also examined the relative frequency of iconic words in vocabulary acquisition during the school years. Thus, we were able to investigate the role of iconicity in vocabulary acquisition beyond infancy and the preschool years. In addition, we examined this using a variety of measures (including production, comprehension and input) to gain a fuller appreciation of the different ways that effects of iconicity manifest throughout development.

Iconicity and part of speech

In addition to the observed relationship with age of acquisition, iconicity has been shown to vary by syntactic class. Perry et al. (2015) found that onomatopoeia and interjections were rated as the most iconic word classes, followed by adjectives, adverbs, and verbs, with nouns and function words rated as least iconic. Work by Winter, Perlman, Perry and Lupyan (2017) has shown that words with a greater amount of associated sensory experience tend to be more iconic. This makes sense because iconicity often involves a resemblance of some perceptual experience. Hence, since adjectives and adverbs depict sensorimotor experiences, their greater iconicity is consistent with what one would expect. Subsequently, Winter et al. (2017) replicated the different iconicity among syntactic classes using ratings for an additional ~ 2,400 items, and also found that words with a greater amount of auditory and tactile information tended to be more iconic (see also Perlman, Little, Thompson & Thompson, 2018). This finding may help explain the higher iconicity of verbs

relative to nouns, given the greater amount of associated auditory experience among verbs than nouns (e.g., Sidhu & Pexman, 2018b). Thus, the meanings of words from certain syntactic classes may be easier to depict through iconicity compared to those from other classes.

Differences in iconicity between syntactic classes may also be affected by SEMANTIC NEIGHBOURHOOD DENSITY, or the number of words with similar meanings. Dense semantic neighbourhoods may impede iconicity due to the greater risk for confusion among similar forms with similar meanings. Indeed, Sidhu and Pexman (2018b) found that iconic words tend to have sparser semantic neighbourhoods. Further, Imai and Kita (2014) speculated that verbs might have sparser semantic neighbourhoods than nouns[1] – since verbs tend to make less fine-grained semantic distinctions than do nouns. These sparser neighbourhoods may in turn contribute to the greater iconicity of verbs.

Importantly, the relationship between age of acquisition and iconicity could be moderated by syntactic class. Investigating this possibility was another goal of the present study. It is possible that differences in semantic density between syntactic classes are not uniform throughout development. While nouns' semantic neighbourhoods may be especially dense in adult vocabulary, this may not yet be the case for children.[2] This would mean that early vocabularies are still able to support iconic nouns, without running into the difficulty of dense semantic neighbourhoods. If this is true, one would predict that the average iconicity of nouns in infants' vocabulary would decline faster than it does for other syntactic classes. Further, iconicity may aid in the learning of other syntactic classes even after noun space has become too crowded. Indeed, Imai and Kita (2014) speculated that iconicity could facilitate the learning of verbs past the stage at which it is facilitative for nouns, due to verbs' sparser neighbourhoods overall.

Present study

In the present study, we investigated iconicity in development by making use of existing language development datasets. In particular, we examined whether there is evidence that the acquisition of iconic words is facilitated across a variety of measures, including children's early production, comprehension, and input frequency. In addition, we explored whether this facilitation exists across various stages of development beyond infancy. Finally, in a largely exploratory set of analyses, we examined whether the developmental trajectory of iconic words differs based on syntactic classes.

Method

Iconicity was quantified using subjective ratings collected by Perry et al. (2015; Experiment 1) and later supplemented by Winter et al. (2017). In total, iconicity ratings were available for 3,001 English words. Raters in these studies were shown words one at a time and asked to rate the extent to which their sounds matched

[1]We confirmed this using the semantic neighbourhood density values from Shaoul and Westbury (2010). Verbs ($n = 11{,}794$) had a lower semantic neighbourhood density ($M = 0.38$, $SD = 0.16$) than nouns ($n = 23{,}257$; $M = 0.41$, $SD = 0.15$), $t(35{,}049) = 19.93$, $p < .001$.

[2]We confirmed this using semantic neighbourhood density values from Shaoul and Westbury (2010). Part of speech (particularly verbs compared to nouns) interacted with age of acquisition (Kuperman et al., 2012) in the prediction of semantic neighbourhood density ($b = -0.03$, $p < .001$). The pattern was such that in earlier acquired words, verbs had denser neighbourhoods; in later acquired words nouns had denser neighbourhoods.

their meanings. Ratings were made on a scale ranging from -5 ("words that sound like the opposite of what they mean") to +5 ("words that sound like what they mean"), with 0 indicating words that "do not sound like what they mean or the opposite" (Perry et al., 2015; p. 12). We used each word's average rating as the measure of its iconicity.

This study involved analysis of data from four datasets: the CHILDES corpus (MacWhinney, 2000), the Wordbank child production data (Frank, Braginsky, Yurovsky & Marchman, 2017), the Dale and O'Rourke (1981) vocabulary comprehension data, and the Zeno word frequency norms (Zeno, Ivens, Millard & Duvvuri, 1995). Note that in all cases we only analyzed nouns, verbs and adjectives. Part of speech was based on Brysbaert, New and Keuleers (2012).

CHILDES corpus

The CHILDES database (MacWhinney, 2000) includes over 5,000 transcriptions of children's speech. These transcriptions can be used to investigate changes in child speech by examining the average lexical and semantic properties of utterances at different ages.

Transcripts of 620 dialogues, containing a total of 125,353 child sentence-utterances were downloaded from CHILDES together with the provided grammatical annotation. These were from typically developing children aged between 9 and 62 months, from all English language corpora (32 in total) that were available at the time of download in 2015–2016. Using information provided in the database we decomposed children's sentences into words (309,116 word-occurrences were found) and assigned their part of speech (POS). A subset of dialogues (comprising 5% of the word-occurrences in the sample) was checked for accuracy, identifying instances where the CHILDES annotation was problematic for mapping accurately to words in the iconicity norms, including compound nouns, contractions, POS, parsing errors, and typos. This process found an error rate of 19.7%. Half of these errors were contractions which are decomposed in CHILDES annotation. We did not want to include in our analyses the iconicity for the constituent words. Another common issue was the CHILDES annotation of compound words, for which we again did not want to include the iconicity values for the constituent words but only for the compound itself.

After identifying these types of errors in the subset of dialogues, the entire database was cleaned to correct them (e.g., spelled out contractions and compound nouns were restored, such that the iconicity of their constituent words were not included in the analyses; plural forms included only where there was a plural form in the Perrt et al., 2015; or Winter et al., 2017 iconicity norms). After data cleaning, 75% of the children's words were found to have corresponding iconicity norms. The final sample consisted of 530 separate dialogues from 157 children. We computed the average iconicity of words uttered in each dialogue, using the ratings collected by Perry et al. (2015) and Winter et al. (2017).

Wordbank data

This dataset combines data from the Macarthur-Bates Communicative Development Inventory (CDI) across multiple sites (see Frank et al., 2017). In particular, we made use of data regarding the percentage of children producing a given word at ages ranging from 16 to 30 months. Using this, we derived a measure of age of acquisition for each word, defined using the same criterion as Dale and O'Rourke

(1981): the youngest age at which more than 67% of children produced a given word. Values on all predictors were available for 387 words.

Dale and O'Rourke data

The Dale and O'Rourke (1981) data estimate age of acquisition (AoA) by objective means, testing children's vocabulary knowledge across school grades and including over 31,000 unique words. Each word's AoA is equal to the lowest grade in which it is known to an estimated 50–70% of students, based on three-alternative multiple choice tests and corrected for guessing. Brysbaert and Biemiller (2017) updated and expanded the data so that they offer vocabulary estimates for grades 2, 4, 6, 8, 10, 12, 13, and 16 (the latter two are college levels). Brysbaert and Biemiller reported that the correlation between the Dale and O'Rourke test-based AoA values and the subjective ratings of AoA from Kuperman, Stadthagen-Gonzalez and Brysbaert (2012; in which adult participants estimated the age at which they had learned each word) was strong at $r = .757$.

In the present analyses, we used the same set of 1,860 items in the analyses of these and the Zeno et al. norms. This was the maximum number of items that were present in both sets of items and had values available for all predictors.

Zeno et al. norms

These norms estimate the vocabulary encountered in schooling for American students in grades 1 through college (Zeno et al., 1995). Using 6,300 text sources and a corpus of over 17,000,000 tokens (154,000 word types), the norms estimate word frequency grade by grade. In our analysis, we examined the frequency of a word in each grade level.

Results

Analyses were carried out using the statistical software R (R Core Team, 2018). Mixed effects models were computed using the "lme4" package (Bates, Maechler, Bolker & Walker, 2015b). We developed each model's random effects structure using the approach suggested by Bates, Kliegl, Vasishth and Baayen (2015a). This involved first fitting a model with all random intercepts and slopes (for fixed effects of interest), then using principal components analysis to determine the number of random effects that the data could support, using the "RePsychLing" package (Baayen, Bates, Kliegle & Vasishth, 2015). Finally, we tested the inclusion of correlations among random effects and the effects themselves using likelihood ratio tests. Continuous predictors were always standardized. Data and code for all analyses are available here: https://osf.io/nz8e3/.

CHILDES corpus

We began with an analysis including all nouns, verbs and adjectives. This consisted of a linear mixed effects model in which the average iconicity of words uttered in that dialogue was the dependent variable. Our predictor of interest was the child's age in each dialogue.[3] Here, and in subsequent analyses, we included the following control

[3]In these analyses age is a predictor. This is because we were testing the hypothesis that a child's age affected the iconicity of their speech. The reverse–that the iconicity of a child's speech affected their

Table 1 Linear mixed effects regression model predicting average dialogue iconicity in the CHILDES corpus.

Predictor	*B*	*SE*	*T*	*p*
Intercept	0.70	0.01	51.45	<.001***
Age	−0.09	0.01	−6.50	<.001***
Frequency	0.07	0.02	4.07	<.001***
Length	−0.08	0.02	−4.05	<.001***
OLD	0.09	0.02	5.37	<.001***
Concreteness	0.08	0.02	4.39	<.001***
Random Effect		s^2		
Subject Intercept		0.005		

Notes. OLD is orthographic Levenshtein distance (Yarkoni et al., 2008).
* $p < .05$, ** $p < .01$, *** $p < .001$.

variables: log transformed frequency in child-directed speech (in CHILDES)[4], word length, orthographic Levensthein distance (*OLD*; Yarkoni, Balota & Yap, 2008), and concreteness (Brysbaert, Warriner & Kuperman, 2014). In the present analyses these were averaged within each dialogue. We also included a random subject intercept. This revealed a significant effect of age ($b = -0.09$, $p < .001$) such that dialogues from younger children tended to contain words with higher average iconicity (see Table 1). It also revealed that more iconic dialogues tended to contain words that were more frequent, shorter, more orthographically distinct, and more concrete.

We next calculated separate dialogue iconicity averages for nouns ($n_{\text{Tokens}} = 37{,}399$), verbs ($n_{\text{Tokens}} = 50{,}512$), and adjectives ($n_{\text{Tokens}} = 6{,}041$), and included part of speech as a predictor (dummy coded; nouns as reference category) along with its interaction with age. We found that there was no interaction between part of speech and age for verbs ($p = .91$) or for adjectives ($p = .26$; see Figure 1). Thus, the relationship between age and iconicity did not differ by syntactic class.

Wordbank data

These analyses consisted of linear regressions in which the dependent measure was production-based AoA. The predictor of interest was iconicity. This analysis revealed a significant effect of iconicity ($b = -0.32$, $p = .02$) such that words higher in iconicity were produced earlier (see Table 2). The analysis also revealed that words produced at a younger age were more frequent and more concrete. Next, we included part of speech (dummy coded; nouns as a reference group) as a predictor, along with its interaction with iconicity. We found that there was no interaction between part of speech and iconicity for verbs ($p = .27$), or for adjectives ($p = .43$; see Figure 2).

age–was not causally possible. This is also why age (grade) is a sensible predictor in the analyses of the Zeno et al. database. In the analyses of the Dale and O'Rourke, and Wordbank data, age is used as a dependent variable because it represents the age at which a word was learned. In those norms age (at which a word was learned) could be causally affected by the iconicity of a word.

[4]This is based on Ping Li's processing of the CHILDES database, available here: https://childes.talkbank.org/derived/parentfreq.cdc

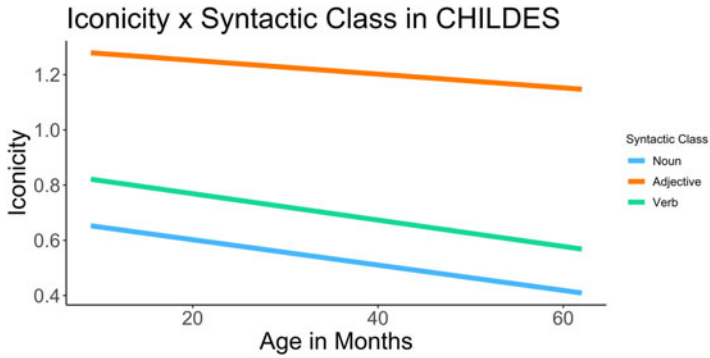

Figure 1. The average iconicity of dialogues at different ages as a function of syntactic class.

Table 2 Linear regression model predicting age at which words are produced, in the Wordbank data (Frank et al., 2017).

Predictor	*B*	*SE*	*T*	*p*
Intercept	25.98	0.13	195.88	<.001***
Iconicity	−0.32	0.14	−2.30	.02*
Frequency	−1.18	0.16	−7.23	<.001***
Length	0.18	0.15	1.20	.23
OLD	0.13	0.16	0.78	.43
Concreteness	−1.50	0.15	−9.84	<.001***

Notes. OLD is orthographic Levenshtein distance (Yarkoni et al., 2008).
* $p < .05$, ** $p < .01$, *** $p < .001$.

Given the pattern observable in Figure 2, these nonsignificant interactions may be due to the small number of items in the analysis ($n_{\text{Nouns}} = 259$, $n_{\text{Verbs}} = 87$, $n_{\text{Adjectives}} = 41$). Nevertheless, we conclude that the relationship between iconicity and age did not differ by syntactic class.

Dale and O'Rourke data

These analyses and those of the Zeno et al. norms (reported next) used the same set of items (1,256 nouns, 332 verbs, and 272 adjectives). The present analyses consisted of linear regressions in which the dependent measure was test-based AoA. The predictor of interest was iconicity. This analysis revealed a significant effect of iconicity ($b = -0.09$, $p = .006$) such that words higher in iconicity were acquired earlier (see Table 3). As would be expected, the analysis also revealed that words comprehended in an earlier grade are more frequent, shorter, and more concrete. Next, we included part of speech (dummy coded; nouns as a reference group) as a predictor, along with its interaction with iconicity. We found that there was no interaction between part of speech and iconicity for verbs ($p = .48$) or for adjectives

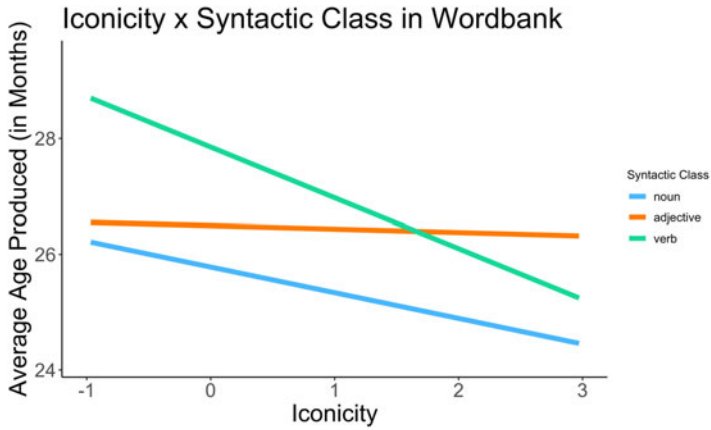

Figure 2. The average age at which words of different iconicity are produced, as a function of syntactic class.

Table 3 Linear regression model predicting grade at which competency is achieved, in the Dale & O'Rourke data.

Predictor	*B*	*SE*	*t*	*p*
Intercept	3.13	0.04	90.58	<.001***
Iconicity	−0.10	0.04	−2.73	.006
Frequency	−0.80	0.04	−21.46	<.001***
Length	0.16	0.07	2.31	.02*
OLD	0.05	0.07	0.75	.45
Concreteness	−0.40	0.04	−11.27	<.001***

Notes. OLD is orthographic Levenshtein distance (Yarkoni et al., 2008).
* $p < .05$, ** $p < .01$, *** $p < .001$.

($p = .11$; see Figure 3). Thus, the relationship between iconicity and age did not differ by syntactic class.

We performed an additional analysis that removed the earliest acquired words ($n = 462$). This was operationalized as those appearing in the Macarthur Bates Developmental Inventory (Fenson, 2007), which was intended to represent the words acquired in the first three years. With these items, the effect of iconicity was marginally significant ($p = .051$). When an interaction with syntactic class was included the effect of iconicity was significant ($p = .004$) and did not interact with syntactic class for verbs ($p = .25$) or adjectives ($p = .08$).

Zeno et al. norms

The analysis of Zeno et al. word frequency values consisted of a mixed effects linear regression. The dependent variable was the log transformed frequency of each word +1 at each grade. Our predictor of interest was an interaction between iconicity and

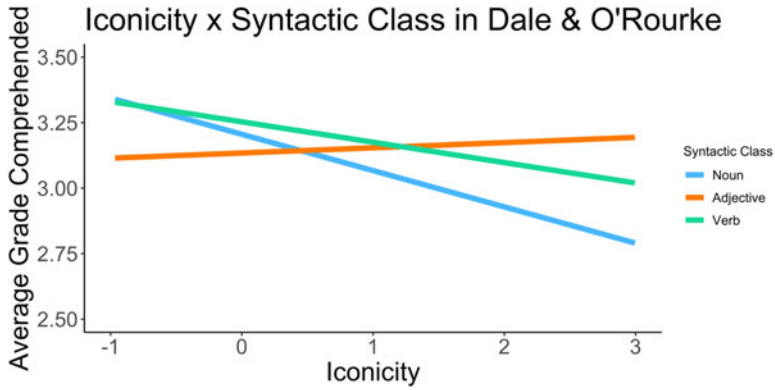

Figure 3. The average grade at which words of different iconicity are comprehended, as a function of syntactic class.

Table 4. Linear mixed effects regression model predicting word frequency at each grade in Zeno et al. norms.

Predictor	*B*	*SE*	*t*	*p*
Intercept	2.87	0.03	113.06	<.001***
Iconicity	−0.21	0.03	−8.13	<.001***
Frequency	1.27	0.02	53.85	<.001***
Length	−0.04	0.04	−1.02	.31
OLD	−0.23	0.04	−5.23	<.001***
Concreteness	−0.17	0.02	−7.78	<.001***
Grade	−0.05	0.01	−4.36	<.001***
Iconicity x Grade	−0.10	0.01	−7.83	<.001***
Random Effect		s^2		*r*
Item Intercept		1.19		
Item Grade Slope		0.28		.52

Notes. OLD is orthographic Levenshtein distance (Yarkoni et al., 2008).
* $p < .05$, ** $p < .01$, *** $p < .001$.

grade. The model also included a random item intercept, and a random item slope for grade. This revealed a significant interaction between iconicity and grade ($b = -0.10$, $p < .001$; see Table 4). The nature of this interaction was that iconic words were more frequent in earlier grades, but less frequent in later grades. The results were unchanged by removing words that are outliers in each grade (+/− 3 SDs from mean grade frequency). In addition, words that were more frequent in educational materials were more frequent in child-directed speech, less orthographically distinct, and less concrete. Next we included a three-way interaction between iconicity, grade, and part of speech (dummy coded; nouns as a reference category). This revealed a

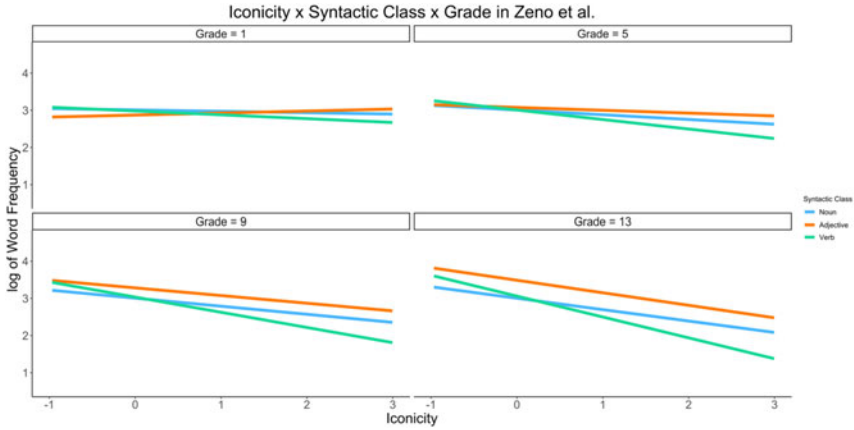

Figure 4. The log word frequency of words of different iconicity at grades 1, 5, 9 and 13, as a function of syntactic class.

marginally significant three-way interaction for verbs ($b = -0.06$, $p = .06$) but not for adjectives ($p = .30$; see Figure 4). We thus conducted separate analyses on the nouns, verbs and adjectives. We found a significant interaction between iconicity and grade for nouns ($b = -0.08$, $p < .001$), verbs ($b = -0.16$, $p < .001$), and adjectives ($b = -0.12$, $p < .001$), though it was largest for verbs.

We again performed an additional analysis in which we removed words from the Macarthur Bates Developmental Inventory (Fenson, 2007). The direction and significance of the results did not change, except that the interaction between iconicity and part of speech was now significant for verbs ($p = .02$).

Discussion

Previous work has shown that there is a negative relationship between age and iconicity, such that earlier acquired words tend to be higher in iconicity (e.g., Perry et al., 2015). We examined this relationship across four different kinds of data: examples of child language production between the ages of nine and 63 months (the CHILDES corpus; MacWhinney, 2000); parent-reported estimates of children's word production between 16 and 30 months (the Wordbank data; Frank et al., 2017); word comprehension data from school age children to young adults between grades 2 and 13 (Dale & O'Rourke, 1981); and word frequency norms from school materials written for school age children to adults between grades 1 and 13 (Zeno et al., 1995). We found that earlier words tended to be higher in iconicity in each of these databases.

There are several novel contributions of this work. The first is the finding that the relationship between age and iconicity continues beyond infancy. Previous work has shown a relationship between iconicity and production frequency at 30 months (Perry et al., 2015), as well as production frequency across the ages of 12 to 69 months (Perry et al., 2018). Here we found that a relationship could be observed in language datasets for children ranging in age from six to 19 years (i.e., Dale & O'Rourke, 1981; Zeno et al., 1995). In particular, in the Dale and O'Rourke data, iconic words were comprehended earlier than arbitrary words. It should be noted

that the majority of the words in our sample from the Dale and O'Rourke data were learned between grades two and four. Nevertheless, this still explores an age range of seven to ten years, thus going well beyond infancy. Importantly, this relationship remained ($p = .051$) even after removing the earliest acquired words; thus, it was not simply a carryover from infancy. This suggests that the benefit of iconicity to word learning is not restricted to infancy. Further, while Perlman et al. (2018) supposed that iconicity might benefit word leaning by providing referential insight and/or simply making words easier to learn, our results from later childhood suggest that the mechanism cannot be referential insight alone.

While the Dale and O'Rourke data provided a glimpse at a period that was largely focused between grades two and four, the Zeno et al. norms provided insight into a much longer time span. The results suggested that the transition towards arbitrariness plays out over a period extending well beyond infancy. The implication is that the lexicon does not arrive at an arbitrary state after infancy, but that this happens gradually. While there are benefits to iconicity, arbitrariness is also important: for example, arbitrariness has been suggested to improve the discriminability of words, particularly in crowded lexicons (see Monaghan, Christiansen & Fitneva, 2011; Sidhu & Pexman, 2018b). Arbitrariness also allows language to refer to meanings that cannot be imitated. As a child's lexicon develops it becomes more dense, and must refer to a wider array of more complex meanings. Both of these factors may direct it towards arbitrariness.

Another important aspect of this work is the variety of ways in which lexical acquisition was operationalized. Previous work has largely focused on child language production corpora (e.g., Perry et al., 2015; 2018), as did the present work with the CHILDES and Wordbank datasets. However, we went beyond this by including a measure of comprehension (Dale & O'Rourke, 1981). Thus, not only are iconic words PRODUCED more often earlier on, they are also COMPREHENDED more often earlier on. This provides direct support for the proposal that iconic words are easier to learn and understand. Additionally, we found that iconic words were included more frequently in educational materials for younger children vs. older children and young adults. Previous work has shown that child- vs. adult-direct speech contains a greater proportion of iconic words (Perlman et al., 2017; Perry et al., 2018; for a review, see Imai & Kita, 2014). Here we extended that work using written input that was specifically designed for children of different ages. The findings support Perry et al.'s (2018) proposal that adults adjust their language use for younger individuals to aid comprehension.

It may be that authors of educational material are either implicitly or explicitly aware of the learning benefit for iconic words and make greater use of them for younger readers. Of course, it is also possible that the greater frequency of iconic words in earlier grades is simply a function of authors using words that tend to be acquired earlier, and these are more likely to be iconic. This would still beg the question of why iconic words tend to be acquired earlier, and the easier acquisition of these words is a reasonable suggestion.

The final contribution of this work is exploring how the relationship between iconicity and age plays out differently across syntactic classes. We did not find evidence for the hypothesized sharp decline in iconic noun frequency. It may be that this change happens in production at an age past the highest age tested in CHILDES. We did, however, observe a marginal interaction involving verbs in the Zeno et al. norms. In this analysis we observed a marginal three-way interaction

between iconicity, frequency and syntactic class (comparing nouns to verbs). While the relationship between iconicity and frequency changed over time for all words, with iconic words becoming less frequent, this change was perhaps most pronounced for verbs. The fact that this interaction only approached significance, however, means that it should not be overinterpreted.

In addition to exploring the prevalence of iconicity in early language, the corresponding shift towards arbitrariness is also important to understand. It has been proposed that arbitrariness provides benefits to language by enhancing discriminability (see Monaghan et al., 2011; Sidhu & Pexman, 2018b), and allowing reference to meanings that cannot be imitated (see Dingemanse et al., 2015). These benefits might affect the shift towards arbitrariness, as language begins to address topics that are vulnerable to confusion or are difficult to imitate.

In conclusion, we replicated and extended previous findings of iconicity providing a benefit to word learning using a variety of child language data sources. We also observed a gradual shift towards arbitrariness over the school years. These data demonstrate that the balance between iconicity and arbitrariness plays a role throughout language development and is important to the understanding of language as a whole.

Acknowledgements. This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada under Discovery Grant # RGPIN/03860-2018 to PMP.

References

Aryani, A., Conrad, M., Schmidtke, D., & Jacobs, A. (2018). Why 'piss' is ruder than 'pee'? The role of sound in affective meaning making. *PloS ONE*, *13*(6).

Asano, M., Imai, M., Kita, S., Kitajo, K., Okada, H., & Thierry, G. (2015). Sound symbolism scaffolds language development in preverbal infants. *Cortex*, *63*, 196–205.

Baayen, H., Bates, D., Kliegle, R., & Vasishth, S. (2015). RePsychLing: Data sets from Psychology and Linguistics experiments. R package version 0.0.4.

Bates, D., Kliegl, R., Vasishth, S., & Baayen, H. (2015a). Parsimonious mixed models. arXiv:1506.04967 [stat]. Retrieved from http://arxiv.org/abs/1506.04967

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015b). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, *67*, 1–48.

Blasi, D. E., Wichmann, S., Hammarstrom, H., Stadler, P. F., & Christiansen, M. H. (2016). Sound-meaning association biases evidenced across thousands of languages. *Proceedings of the National Academy of Sciences, USA*, *113*, 10818–10823.

Brysbaert, M., & Biemiller, A. (2017). Test-based age-of-acquisition norms for 44 thousand English word meanings. *Behavior Research Methods*, *49*, 1520–1523.

Brysbaert, M., New, B., & Keuleers, E. (2012). Adding part-of-speech information to the SUBTLEX-US word frequencies. *Behavior Research Methods*, *44*, 991–997.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911.

Dale, E., & O'Rourke, J. (1981). *The living word vocabulary, the words we know: A national vocabulary inventory*. Chicago: World Book.

Dingemanse, M., Blasi, D. E., Lupyan, G., Christiansen, M. H., & Monaghan, P. (2015). Arbitrariness, iconicity, and systematicity in language. *Trends in Cognitive Sciences*, *19*, 603–615.

Dingemanse, M., Perlman, M., & Perniss, P. (2020). Construals of iconicity: experimental approaches to form–meaning resemblances in language. *Language and Cognition*, *12*, 1–14.

Fenson, L. (2007). *MacArthur-Bates communicative development inventories*. Baltimore, MD: Paul H. Brookes Publishing Company.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: an open repository for developmental vocabulary data. *Journal of Child Language*, *44*, 677–694.

Hinojosa, J. A., Haro, J., Magallares, S., Duñabeitia, J. A., & Ferr, P. (2020). Iconicity ratings for 10,995 Spanish words and their relationship with psycholinguistic variables. *Behavior Research Methods*, 1–14. https://doi.org/10.3758/s13428-020-01496-z

Hockett, C. (1963). The problem of universals in language. In J. Greenberg (Ed.), *Universals of language* (pp. 1–22). Cambridge, MA: MIT Press.

Imai, M., & Kita, S. (2014). The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 20130298. doi:10.1098/rstb.2013.0298

Imai, M., Kita, S., Nagumo, M., & Okada, H. (2008). Sound symbolism facilitates early verb learning. *Cognition*, *109*, 54–65.

Imai, M., Miyazaki, M., Yeung, H. H., Hidaka, S., Kantartzis, K., Okada, H., & Kita, S. (2015). Sound symbolism facilitates word learning in 14-month olds. *PLoS ONE*, *10*, e0116494. doi:10.1371/journal.pone.0116494.

Kanero, J., Imai, M., Okuda, J., Okada, H., & Matsuda, T. (2014). How sound symbolism is processed in the brain: A study on Japanese mimetic words. *PLoS ONE*, *9*, e97905. doi:10.1371/journal.pone.0097905

Kovic, V., Plunkett, K., & Westermann, G. (2010). The shape of words in the brain. *Cognition*, *114*, 19–28.

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*, 978–990.

Laing, C. E. (2017). A perceptual advantage for onomatopoeia in early word learning: Evidence from eye-tracking. *Journal of Experimental Child Psychology*, *161*, 32–45.

Laing, C. E. (2019). A role for onomatopoeia in early language: Evidence from phonological development. *Language and Cognition*, *11*, 173–187.

Lockwood, G., & Dingemanse, M. (2015). Iconicity in the lab: A review of behavioral, developmental, and neuroimaging research into sound-symbolism. *Frontiers in Psychology*. doi: 10.3389/fpsyg.2015.01246

Lockwood, G., Dingemense, M., & Hagoort, P. (2016). Sound symbolism boosts novel word learning. *Journal of Experimental Psychology*, *42*, 1274–1281.

MacWhinney, B. (2000). *The CHILDES Project: Tools for analyzing talk* (3rd edn.). Mahwah, NJ: Lawrence Erlbaum Associates.

Massaro, D. W., & Perlman, M. (2017). Quantifying iconicity's contribution during language acquisition: Implications for vocabulary learning. *Frontiers in Communication*, *2*, 1–14. doi:10.3389/fcomm.2017.00004

Maurer, D., Pathman, T., & Mondloch, C. J. (2006). The shape of boubas: Sound-shape correspondence in toddlers and adults. *Developmental Science*, *9*, 316–322.

Monaghan, P., Christiansen, M. H., & Fitneva, S. A. (2011). The arbitrariness of the sign: Learning advantages from the structure of the vocabulary. *Journal of Experimental Psychology: General*, *140*, 325–347.

Monaghan, P., Shillcock, R. C., Christiansen, M. H., & Kirby, S. (2014). How arbitrary is language?. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*. doi:10.1098/rstb.2013.0299

Nielsen, A. K., & Dingemanse, M. (2020). Iconicity in word learning and beyond: A critical review. *Language and Speech*. doi:10.1177/0023830920914339

Nielsen, A., & Rendall, D. (2012). The source and magnitude of sound-symbolic biases in processing artificial word material and their implications for language learning and transmission. *Language and Cognition*, *4*, 115–125.

Ozturk, O., Krehm, M., & Vouloumanos, A. (2013). Sound symbolism in infancy: Evidence for sound-shape cross-modal correspondences in 4-month olds. *Journal of Experimental Child Psychology*, *114*, 173–186.

Perlman, M., Fusaroli, R., Fein, D., & Naigles, L. (2017). The use of iconic words in early child-parent interactions. In G. Gunzelmann, A. Howes, T. Tenbrink & E. Davelaar (Eds.), *Proceedings of the 39th Annual Conference of the Cognitive Science Society (CogSci 2017)* (pp. 913–918). Austin, TX: Cognitive Science Society.

Perlman, M., Little, H., Thompson, B., & Thompson, R. L. (2018). Iconicity in signed and spoken vocabulary: A comparison between American Sign Language, British Sign Language, English, and Spanish. *Frontiers in Psychology*, *9*, 1433. doi:10.3389/fpsyg.2018.01433

Perniss, P., Lu, J. C., Morgan, G., & Vigliocco, G. (2018). Mapping language to the world: The role of iconicity in sign language input. *Developmental Science*, *21*, e12551

Perniss, P., Thompson, R. L., & Vigliocco, G. (2010). Iconicity as a general property of language: Evidence from spoken and signed languages. *Frontiers in Psychology*, *1*, 227. doi:10.3389/fpsyg.2010.00227

Perniss, P., & Vigliocco, G. (2014). The bridge of iconicity: From a world of experience to the experience of language. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*, 20130300. doi:10.1098/rstb.2013.0300

Perrt, L. K., Perlman, M., & Lupyan, G. (2015). Iconicity in English and Spanish and its relation to lexical category and age of acquisition. *PLoS ONE*, *10*, e0137147. doi:10.1371/journal.pone.0137147

Perry, L. K., Perlman, M., Winter, B., Massaro, D. W., & Lupyan, G. (2018). Iconicity in the speech of children and adults. *Developmental Science*, *21*, e12572. doi:10.1371/journal.pone.0137147

R Core Team (2018). R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Online <https://www.R-project.org/>.

Shaoul, C., & Westbury, C. (2010). Exploring lexical co-occurrence space using HiDEx. *Behavior Research Methods*, *42*, 393–413.

Sidhu, D. M., & Pexman, P. M. (2018a). Five mechanisms of sound symbolic association. *Psychonomic Bulletin & Review*, *25*, 1619–1643.

Sidhu, D. M., & Pexman, P. M. (2018b). Lonely sensational icons: semantic neighbourhood density, sensory experience and iconicity. *Language, Cognition and Neuroscience*, *33*, 25–31.

Sidhu, D. M., Vigliocco, G., & Pexman, P. M. (2020). Effects of iconicity in lexical decision. *Language and Cognition*, *12*, 164–181.

Sidhu, D. M., Westbury, C., Hollis, G., & Pexman, P. M. (in press). Sound symbolism shapes the English language: The maluma/takete effect in English nouns. *Psychonomic Bulletin & Review*.

Thompson, R. L., Vinson, D. P., Woll, B., & Vigliocco, G. (2012). The road to language learning is iconic: Evidence from British Sign Language. *Psychological Science*, *23*, 1443–1448.

Tzeng, C. Y., Nygaard, L. C., & Namy, L. L. (2017). Developmental change in children's sensitivity to sound symbolism. *Journal of Experimental Child Psychology*, *160*, 107–118.

Vinson, D. P., Cormier, K., Denmark, T., Schembri, A., & Vigliocco, G. (2008). The British Sign Language (BSL) norms for age of acquisition, familiarity, and iconicity. *Behavior Research Methods*, *40*, 1079–1087.

Winter, B., Perlman, M., Perry, L. K., & Lupyan, G. (2017). Which words are most iconic? Iconicity in English sensory words. *Interaction Studies*, *18*, 433–454.

Yang, J., Asano, M., Kanazawa, S., Yamaguchi, M. K., & Imai, M. (2019). Sound symbolism processing is lateralized to the right temporal region in the prelinguistic infant brain. *Scientific Reports*, *9*, 1–10.

Yarkoni, T., Balota, D., & Yap, M. (2008). Moving beyond Coltheart's N: A new measure of orthographic similarity. *Psychonomic Bulletin & Review*, *15*, 971–979.

Zeno, S. M., Ivens, S. H., Millard, R. T., & Duvvuri, R. (1995). *The Educator's Word Frequency Guide*. New York, NY: Touchstone Applied Science Associates, Inc.

Cite this article: Sidhu DM, Williamson J, Slavova V, Pexman PM (2021). An investigation of iconic language development in four datasets. *Journal of Child Language* 1–15. https://doi.org/10.1017/S0305000921000040