EMPIRICAL ARTICLE

Semantic maturation during the comprehension-expression gap in late and typical talkers

Eva Jiménez | Thomas T. Hills

Department of Psychology, University of Warwick, Coventry, UK

Correspondence
Eva Jiménez, Department of Psychology, University of Warwick, University Road, Coventry CV4 7AL, UK.
Email: eva.jimenez-mesa@warwick.ac.uk

Funding information
Alan Turing Institute; Royal Society Wolfson Research Merit Award, Grant/Award Number: WM160074; Leverhulme Trust

Abstract

This study investigates the influence of semantic maturation on early lexical development by examining the impact of contextual diversity—known to influence semantic development—on word promotion from receptive to productive vocabularies (i.e., comprehension-expression gap). Study 1 compares the vocabularies of 3685 American-English-speaking typical talkers (TTs) and late talkers (LTs; 16–30 months old; 1257 females, 1021 gender unknown; ethnicity unknown; data downloaded in 2018) and finds that LTs, with a longer preverbal phase, produced nouns with lower contextual diversity (R^2 = .80), but verbs with higher contextual diversity (R^2 = .13). Study 2 compares computational network growth models of semantic maturation and finds that verbs require more semantic maturation than nouns, and TTs produce words that are more semantically mature than LTs.

Toddlers understand more words than they can say. This is the well-known *comprehension-expression gap* (Goldin-Meadow et al., 1976). When children spontaneously produce a word for the first time (like *dog* or *drink*), they already know something about the meaning of that word (*dog*: it is animated, furry, has a tail, etc.; *drink*: you can drink things like milk or water, you can drink from bottles or glasses, etc.). However, what researchers understand less well is how necessary *semantic maturation* (preproductive learning of a word's meaning) is for word production. Moreover, what paths does this semantic maturation take for different populations of young learners? In particular, the comprehension-expression gap is larger for late-talking toddlers (Thal & Bates, 1988). This means that the words late talkers (LTs) produce reside in their preproductive receptive vocabularies (i.e., comprehension) for longer. Could this mean

that LTs have the opportunity to form richer semantic representations of words before they produce them? Or is an atypical learning strategy by LTs responsible for creating weaker semantic representations, leading to delayed production?

One way to examine this is to focus on a well-known influence on early word learning, *contextual diversity* (Hills et al., 2009a; see also work on *semantic diversity*, Hoffman et al., 2013; Hsiao & Nation, 2018; Jones et al., 2012). Contextual diversity is a property of a word in its language environment and relates to the diversity of the contexts in which it is used. Words with high contextual diversity (appearing in many different linguistic contexts, and consequently with many other words) are easier to learn by children (e.g., Rosa et al., 2017), presumably because contextual diversity enriches a word's semantic representation (Vergara-Martínez et al., 2017).

In our examples above, the children would enrich the concepts of dog and milk by encountering the noun *dog* close to words like *furry* and *tail*, and the verb *drink* close to words like *water* or *bottle* within speech. Although evidence shows that toddlers learn words with high contextual diversity earlier (e.g., Hills et al., 2010), LTs have been hypothesized to produce more low-contextually diverse words, leading to productive lexicons with weaker semantic links between words (Beckage et al., 2011). Learning low-contextually diverse words might be related to low levels of semantic maturation, as illustrated in Figure 1. Might the higher production of low-contextually diverse words be a symptom of differences in the role of semantic maturation between typical and LTs?

In this paper, we shed light on these questions by, first, examining the impact of contextual diversity on the productive vocabularies of a large sample of typical and LT, and second, by examining the performance of four computational models to disentangle the causes underlying different developmental pathways of semantic maturation. Study 1 aims to confirm whether LTs show an atypical word learning based on contextual diversity. Although we expect to confirm previous hypothesis posed by Beckage et al. (2011) by replicating the same network measures, the current study also holds exploratory components as we also investigate contextual diversity and examine verbs and nouns separately, for which no hypothesis were formed for the latter. Study 2 explores the role of semantic maturation in word development, seeking to explain the results from Study 1. The computational models were specifically developed to understand the still unknown influence of the preverbal comprehension vocabularies on later word production making Study 2 exploratory in nature.

Studies from psycholinguistics and cognitive science have identified distinct differences between production and comprehension, although recent views defend the idea that these two systems are interwoven and support each other (for a review, see Pickering & Garrod, 2013). Sahni and Rogers (2008, see also Stokes et al., 2019), examining both production and comprehension, found that word production showed more influence of phonological density, whereas word comprehension showed more influence of semantic density (which they based on shared perceptual features). Semantic density of a word relates to contextual diversity in that both the word's perceptual features and the words that co-occur with it in the linguistic context define the word's concept. Though the operational definition of semantics varies across studies, the results in Sahni and Rogers's (2008) study nonetheless suggest that the promotion of words from receptive (i.e., comprehension) to productive vocabularies is not based entirely on seniority (i.e., the length of time the word has been in the child's receptive vocabulary before production). Some words in the receptive vocabulary become part of the productive vocabulary before words that have been in the receptive vocabulary for longer. If the factors influencing word promotion into and out of the receptive vocabulary differ across late and typical talkers (TTs), this should be reflected in the relative contextual diversity of the words in their productive vocabularies.

How the linguistic context assists the infant in learning various aspects of their native language has been studied in the literature from different perspectives,
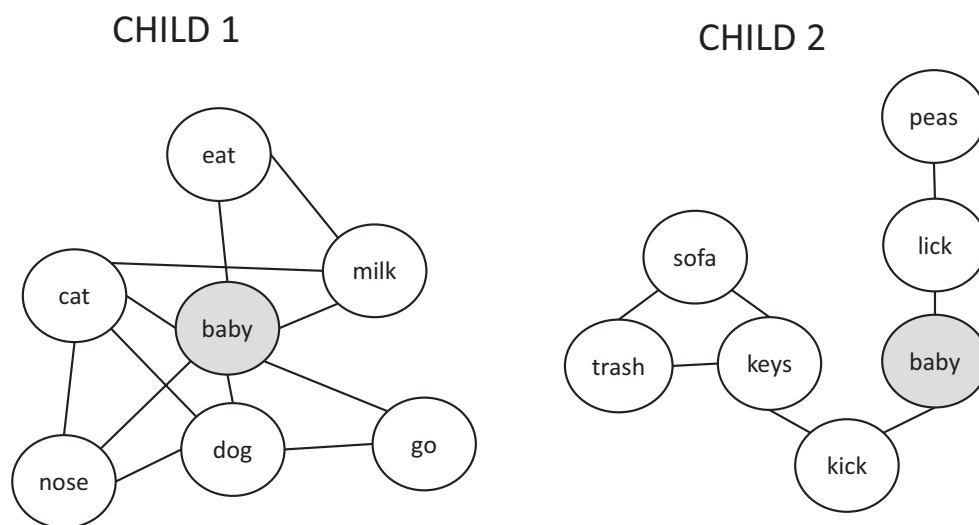

FIGURE 1 Vocabularies learned by two hypothetical children. *Note*: The word baby is learned by child 1 and child 2 along with six other words. Child 1 learned words with high contextual diversity and child 2 learned words with low contextual diversity. Baby has more semantic links with other words in child 1's vocabulary than in child 2's vocabulary. This leads the concept baby to have higher semantic maturation in child 1 compared to child 2, signifying that child 1 has a more complex understanding of what baby is (in this example, child 1 knows that babies drink milk, go to places, eat food, have noses and that babies can be dogs and cats). Also, the acquisition of low-contextually diverse words leads child 2's vocabulary to have different network properties compared to child 1's: It has fewer links, fewer clusters, and the average distance between any pair of nodes is larger.

such as syntactic bootstrapping (Gleitman, 1990), construction grammar (Goldberg, 2003), or variation sets (e.g., Küntay & Slobin, 2002). The common component of these theories is that those elements in the linguistic environment of a lexeme which are subject to change, such as morphemes, words, or grammar constructions, support learning (e.g., categorical learning, label learning, or semantic learning). Contextual diversity is an extension of this branch of investigation that solely focuses on the acquisition of individual words and allows the examination of the impact of contextual variation in word acquisition through a word feature (i.e., contextual diversity). Although most of these studies investigate the linguistic context, there are other environmental factors that also co-occur with a word of interest, such as location and time (see Roy et al., 2015). Nevertheless, on many occasions, the language environment provides this additional information (e.g., "baby gave the ducky a bath this morning").

Here, we follow the approach of Adelman et al. (2006) and Jones et al. (2012) in considering the word's contextual diversity a general feature of the word in its environment, measuring its likelihood of use in various contexts. Our approach to measuring contextual diversity in children's vocabularies therefore involves averaging a measure of the context diversity of the words a child knows as those words appear in the language a child is likely to hear. This approach is in accordance with findings that demonstrated that children tend to produce words that are well-connected with other words within their language environment (i.e., the *preferential acquisition* [PA] principle of growth) rather than producing words that have more connections with the words that the child already knows (i.e., *preferential attachment* principle of growth; Hills et al., 2009a). Specifically, Hills et al. (2009a) observed that the growth of early vocabularies clearly reflects a sensitivity to the contextual diversity in the structure of the language learning environment, and predicts age of acquisition above and beyond frequency, phonology, and a host of other word properties (see also Fourtassi et al., 2019; Hills et al., 2010; Stella et al., 2017).

Focusing on how contextual diversity influences semantic maturation processes during lexical development might be crucial to elucidate unexplained early language delays. This is the case of LTs, who do not present any disability or developmental disorder, and usually show normal levels of comprehension despite exhibiting a delay in production (Thal, Marchman, & Tomblin referred to these children as "late producers," 2013). Beckage et al. (2011) reported that TTs and LTs show well-connected small-world semantic networks among the words in their productive vocabularies, but late-talking children showed this to a lesser degree. Based on this, Beckage et al. proposed that LTs might learn words based on an inverted pattern of PA (Hills et al., 2009a): LTs might be producing words with fewer connections

with other words *in* the learning environment. Recent work by Horvath and Arunachalam (2021) measured TTs' and LTs' initial representations of novel verbs when learned either in a consistent or a varying context. The authors found that both groups benefited more from experiencing the novel verb in a consistent context, and this effect was even more pronounced for the late talker group. The evidence for a lower production of high contextual diversity words and the stronger effect of learning verbs in consistent contexts in LTs might indicate an atypical use of contextual diversity cues in the environment during semantic maturation.

Critically, different word classes have different patterns of contextual diversity, which suggests further evidence for differences in words' semantic maturation. Work of Gentner (1982) found that verbs have many tokens but few types in child-directed speech. This is in comparison with nouns, which have relatively fewer tokens but more types. In the context of word co-occurrences, this may mean that verbs have more chances to form links with more word types than nouns, enhancing the contextual diversity of verbs. There is also evidence to suggest that contextual diversity is better correlated with age of acquisition for verbs than for nouns (Hills, 2012). If sensitivity to contextual diversity were similar for nouns and verbs, then theoretically verbs should be acquired earlier for their high contextual diversity, yet they are generally produced and understood later (Bates et al., 1994). Of course, words have other characteristics that influence their learning, such as abstractness, which is higher in verbs and might explain their delayed onset (e.g., Hirsh-Pasek & Golinkoff, 2010). Since word class may have an impact on a word's contextual diversity, by using a larger dataset the present work offers an opportunity to extend the result of Beckage et al. (2011) to separate the analysis by word class. Furthermore, contextual diversity facilitates the discovery of the semantic features of words. However, words like function words do not appear to have this advantage (e.g., Hills, 2012; Hills et al., 2010). Thus, we focus on nouns and verbs here, which show the highest impact of contextual diversity. In addition, early lexicons are mostly composed of nouns and verbs (Bates et al., 1994).

The present studies

Here, we sought to answer the following research questions.

Q.1: *Do LTs produce fewer high-contextual diversity words than vocabulary-matched TTs?* In Study 1, we answer this question by examining the average contextual diversity of the nouns and verbs produced by a large sample of late- and typically talking toddlers. In addition, we examine the network properties of the two talker groups because the use of a contextual diversity strategy is associated with well-connected lexical networks. Based on

Beckage et al.'s findings, we expect to find LTs' networks to show a lower degree of connectedness for both nouns and verbs.

Q.2: *Do different word types go through different periods of semantic maturation before being produced?* We explore this question in Study 2 by comparing three computational models that tease apart the potential impact of a developing comprehension phase on posterior productive lexicons. Specifically, we are interested in determining whether verbs follow a different path of semantic maturation since verbs are presumed to have higher contextual diversity than nouns (a difference that we also test in our analysis).

Q.3: *Is contextual diversity a source of word information that mainly predominates in the preverbal phase (before the child produces her first word)?* Stella et al. (2017) suggested developmental changes in the sensitivity to contextual diversity during early stages of word learning. We address this question by letting the model change its sensitivity to contextual diversity during the preverbal and verbal phases of development.

Q.4: *Do LTs and TTs differ in the amount of semantic maturation necessary for words to move from comprehension to production?* We answer this question by examining the best free parameter values of our best computational model.

STUDY 1: TYPICAL TALKERS VERSUS LATE TALKERS

Methods

Sample, identification of late talkers and words for analysis

A total of 5520 vocabularies of American children aged 16 to 30 months were downloaded during April 2018 from *Wordbank* (Frank et al., 2017). "Full Child-by-Word data" were selected under data, "Words & Sentences" under forms, and "(American) English" under language. Various researchers contributed to gathering this data collected using the parental checklist MacArthur-Bates Communicative Development Inventory (CDI; Fenson et al., 1993). Because of challenges associated with making inferences about comprehension from the CDI Words & Gestures (e.g., Moore et al., 2019), we focus our analysis here on production data from the CDI Words & Sentences. For the identification of LTs, we used the vocabulary norms provided by the CDI for children aged 16–30 months and classified children at or below the 20th percentile of vocabulary size for their age as LTs. We selected this percentile threshold following previous work to facilitate the comparison of our results (Beckage et al., 2011). For nouns, we included the following CDI categories: "Animals", "Vehicles", "Toys", "Food

& drinks", "Clothing", "Body parts", "Household", "Furniture & rooms", "Outside" and "People". For verbs, we included all words categorized as "Action words". In order to measure contextual diversity and for clarity of interpretation, it was necessary to exclude homographs, for example, "swing" (noun) and "swing" (verb); concepts composed of two words, for example, "rocking chair"; and words that do not occur in the child-directed speech extracted from CHILDES (MacWhinney, 2000), which we describe in the next subsection. Our final word selection was composed of 286 nouns and 96 verbs. For a given word class (nouns or verbs), TTs were vocabulary-size matched to LT children. This necessarily excludes TTs with larger vocabulary sizes than our LT sample. That is, only TTs with noun vocabularies up to 190 words (which is the maximum noun vocabulary size found in our LT group) were included in the noun analysis, and only TTs with verb vocabularies up to 70 words (which is the maximum verb vocabulary size found in our LT group) were included in the verb analysis. We set a minimum of 10 words to avoid the high variation produced by averaging small vocabularies. Productive noun vocabulary sizes ranged between 10 and 190 words; productive verb vocabulary sizes ranged between 10 and 70 words.

The final subset of children comprises 3685 unique children, from which 3211 were included in the noun analysis, and 1949 children were included in the verb analysis. In the group of children for the noun analysis, 626 were identified as LTs; in the group of children for the verb analysis, 183 were identified as LTs. All children aged 16 to 30 months (see Table 1). Some children with small vocabularies only produced nouns, and therefore they were included in the noun analysis only; similarly, children with large vocabularies and more than 190 nouns produced were excluded for the noun analysis, although they were included in the verb analysis since their verb vocabulary size was within the verb range analyzed. A total of 1475 children had both their verb and noun vocabularies analyzed. In our noun analysis, within the LT group, 25.9% of children were female, 43.8% were male, and 30.4% had an unknown gender; within the TT group, 34.7% were female, 36.9% were male, and 28.4% had an unknown gender. In our verb analysis, within the LT group, 30.1% of children were female, 50.8% were male, and 19.1% had an unknown gender; within the TT group, 35.0% were female, and 36.4% had an unknown gender 28.6% were male. Maternal education was available: 22% of LTs' mothers completed or did some graduate education, 48% completed or did some college education, and 30% completed secondary or primary education; 20% of TTs' mothers completed or did some graduate education, 53% completed or did some college education, and 27% completed secondary or primary education. Since the current sample was drawn from a public repository where contributors were only allowed to submit a limited set of characteristics about their sample, no information

TABLE 1 Distribution of typical talkers (TT) and late talkers (LT) across productive vocabulary sizes

Vocabulary size	Group TT n	Age M [range]	LT n	Age M [range]	Total n	Age M [range]
Nouns						
(9,29]	519	16.6 [16, 20]	272	20.9 [16, 30]	791	18.1 [16, 30]
(29,49]	383	17.7 [16, 23]	89	24.2 [20, 30]	472	18.9 [16, 30]
(49,69]	274	18.8 [16, 26]	68	25.5 [23, 30]	342	20.1 [16, 30]
(69,89]	227	19.8 [16, 27]	68	26.5 [24, 30]	295	21.4 [16, 30]
(89,109]	213	20.9 [16, 27]	40	27.3 [25, 30]	253	21.9 [16, 30]
(109,129]	219	21.4 [16, 28]	36	28.4 [27, 30]	255	22.4 [16, 30]
(129,149]	216	22.1 [16, 28]	25	28.6 [27, 30]	241	22.8 [16, 30]
(149,169]	244	23.8 [16, 30]	23	29.1 [27, 30]	267	24.2 [16, 30]
(169,190]	290	24.5 [16, 30]	5	29.4 [29, 30]	295	24.6 [16, 30]
Total	2585	20.1 [16, 30]	626	24.0 [16, 30]	3211	20.8 [16, 30]
Verbs						
(9,29]	698	20.9 [16, 29]	135	27.1 [17, 30]	833	20.9 [16, 30]
(29,49]	509	23.8 [16, 30]	41	28.4 [20, 30]	550	23.8 [16, 30]
(49,70]	559	25.2 [16, 30]	7	29.0 [28, 30]	566	25.2 [16, 30]
Total	1766	23.1 [16, 30]	183	27.5 [16, 30]	1949	23.5 [16, 30]

about the number of languages spoken at home, racial/ethnic information or other demographic characteristics of the sample were available.

Contextual diversity of words and network analysis

To compute contextual diversity, we analyzed American child-directed speech, taken from the CHILDES corpus (MacWhinney, 2000). We included all adults' speech directed to children up to 5 years old and computed the contextual diversity of each word. All children's speech was removed, and no free spaces were left between the adults' utterances. Following a surface proximity approach (see Evert, 2008), we determined the frequency in which each distinct word (node) in the corpus co-occurred with other words (collates). To do this, a matrix was populated by moving a window of size 10 word-by-word through the corpus. A window size of 10 was selected since it best predicted age of acquisition (see model details in Supporting Information). The word at the start of the window was used to index the row [i,], and any word encountered downstream and within the window of the starting word was used to index the column [, j]. When two words co-occurred, a value of one was added to position [i, j] in the matrix. The resulting weighted matrix was transformed into a binary matrix measuring contextual diversity with word-types by setting [i, j] = 1 for all [i, j] > 0. Finally, we extracted from this matrix a smaller matrix with nouns and verbs only. We then calculated the contextual diversity value for each word by adding the sum of the row and the sum of the column of this submatrix. Therefore, the contextual diversity value of a word reflects the number of semantic links that the word in question has with other verbs and nouns in our sample.

For the network analysis, we used two submatrices—one each for nouns and verbs—that we extracted from the matrix of co-occurrences described above. Words in the child's lexicon are represented as nodes, and the edges between nodes indicate semantic relatedness, inferred from the co-occurrence of each word with all other words within the speech stream. Network analysis in cognitive psychology (see Vitevitch, 2019) has been extremely successful in detecting structural differences in language acquisition (Bilson et al., 2015; Hills et al., 2009b) and lexical processing (e.g., Vitevitch et al., 2014). Network statistics were computed using R and the *igraph* package, version 1.0.1 (Csárdi & Nepusz, 2006). Once all the words were connected in each vocabulary using the binary matrix described above, the clustering coefficient and average path length were calculated for each child's undirected network. Clustering coefficient measures the degree to which nodes in a network tend to cluster together. Specifically, we calculated the local clustering coefficient, which measures the number of links that the neighbors of a node have among themselves. The average path length is the mean shortest path between all pairs of words in a network, describing the level of global access.

We decided to utilize a binary matrix for two reasons. First, this follows the approach as in previous work by Beckage et al. (2011) and Hills et al. (2010), which both predict order of word acquisition and identifiy differences between LTs and TTs. Second, this focuses more exclusively on contextual diversity (how many other word types a word appears with) because a weighted matrix is confounded with frequency. Using the same matrix to build the vocabulary networks as to compute the words' contextual diversity also allows us to compare the network statistics with the average contextual diversity of children's vocabularies.

We conducted correlational analyses of the following words' features: contextual diversity, frequency, concreteness, and word length. For this, we computed word frequency from our CHILDES sample, calculated word length (phonemes), and assigned concreteness values to each word in the analysis, which were taken from Brysbaert et al. (2014).

Statistical analyses used generalized additive models (GAM) using the *mgcv* package and the *bam* functions in R (see Wood, 2011). GAM was chosen over simpler models because, first, the dependent variables of interest (average contextual diversity, clustering coefficient, and average path length) were highly correlated with vocabulary size and the talker groups differed in the number of words produced; second, homogeneity of variance was violated, resulting in the need to randomize vocabulary size. In our GAMs, the independent variable talker type was set as a fixed term, and vocabulary size was set as a smooth term. Statistical assumptions were verified using the *gam. check* function. The Akaike information criterion (AIC) of the models presented here were significantly higher than the AICs of the same models without talker type as a fixed term. An interaction term between vocabulary size and type of talker did not significantly reduce the models' AIC. The source code for the two studies is available at the Supporting Information.

Results

Word characteristics

Following Gentner (1982) and justifying the distinction between verbs and nouns we make in our analysis below, verbs present higher contextual diversity than nouns (verbs: $Mdn = 268$, nouns: $Mdn = 170.50$, $W = 5854.50$, $p < .001$, $d = .95$, 95% CI [−111.00, −72.00]). Furthermore, in our best multilevel binomial regression model (AICs compared), contextual diversity significantly predicted word class ($b = 1.35$, $z = 5.58$, $p < .001$) in a model where frequency was randomized. We also computed the contextual diversity of every unique word in the CHILDES corpus, and our correlational analysis showed that, as Hills et al. (2010) reported, high contextual diversity words are produced earlier ($r = .46$; See Supporting Information for details).

Contextual diversity was positively correlated with word frequency ($r[369] = .47$, $p < .001$), and negatively correlated with concreteness ($r[369] = −.43$, $p < .001$). Concreteness and frequency were negatively correlated ($r[369] = −.45$, $p < .001$). Word length was not correlated with either frequency or contextual diversity (frequency: $r(369) = −.0008$, $p > .05$; contextual diversity: $r(369) = −.005$, $p > .05$), but it was weakly correlated with concreteness ($r(369) = .10$, $p < .05$).

Contextual diversity

As shown in Figure 2, LT and TT children differ in the contextual diversity of the nouns and verbs they produce (fitted values for nouns and verbs come from two separate models). Regarding nouns, LTs produced vocabularies with lower average contextual diversity than TTs, adjusted $R^2 = .80$, $F(1, 3201.23) = 20.54$, $p < .001$, 95% CI [1.35, 3.49]. In the case of verbs, LTs
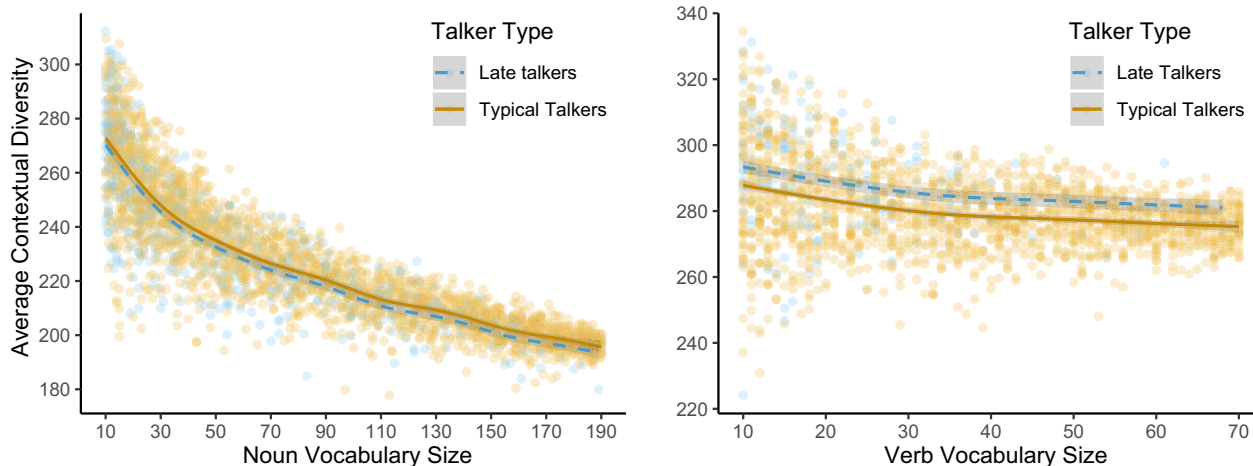

FIGURE 2 Average contextual diversity in the noun (left panel) and verb (right panel) vocabularies of late talkers and typical talkers across vocabulary sizes. *Note*: Shadows around the curves represent 95% confidence intervals. Each individual dot represents a child's vocabulary.

produced verb vocabularies with higher average contextual diversity than TTs, adjusted R^2 = .13, $F(1, 1943.91)$ = 35.08, $p < .001$, 95% CI [−7.48, −3.70]. In addition, both types of talkers show a decrease in contextual diversity as they produce more words. This is true for nouns, $F(7.77, 8.60)$ = 1385, $p < .001$, as well as for verbs, $F(3.09, 3.83)$ = 53.57, $p < .001$.

We re-ran our analysis with males only, and excluded children whose mother's education were below college level since maternal education was found to be a reliable predictor of children's language (e.g., Reilly et al., 2007; Reilly et al., 2010; Nouns TT: n = 541, age in months M = 21.0, age SD = 3.4, age range = 16–30; Nouns LT: n = 165, age M = 23.6, age SD = 3.9, age range = 16–30; Verbs TT: n = 429, age M = 23.1, age SD = 3.6, age range = 16–30; Verbs LT: n = 41, age M = 27.5, age SD = 2.4, age range = 20–30). The same differences emerged between the groups in the noun analysis, with LTs producing noun vocabularies with lower contextual diversity compared to TTs, adjusted R^2 = .79, $F(1, 697.9544)$ = 19.75, $p < .001$, 95% CI [2.67, 7.12]. Similarly, LTs produced verb vocabularies with higher contextual diversity, adjusted R^2 = .10, $F(1, 466.9994)$ 15.22, $p < .001$, 95% CI [−12.00, −3.87].

Given the high correlation of concreteness and frequency with contextual diversity, we proceeded to re-run our analysis including these two word features as smoothing terms in our GAMs to adjust for these variables. A potential concern for comparing the model's coefficient estimates was predictor concurvity (a generalization of collinearity in GAMs), however it does not affect the overall fit of the model and allows us to identify those variables that can explain a significant part of the variance. The resulting models showed that the talker type remained a significant predictor for the verb model (adjusted R^2 = .67, $F(1, 1940.944)$ = 29.75, $p < .001$, 95% CI [−4.36, −2.02]), but not for the noun model (adjusted R^2 = .92, $F(1, 3189.305)$ = 0.182, $p > .05$, 95% CI [−0.53, 0.82]). Although we continue investigating the noun and verb differences in our computational models in Study 2, we consider the influence of concreteness and frequency in our noun results in the discussion.

Network properties

The results of the network analyses also indicate differences between LTs and TTs. With regard to nouns, LTs exhibited lower clustering coefficient, adjusted R^2 = .79, $F(1, 3202.10)$ = 18.73, $p < .001$, 95% CI [0.0029, 0.0078], and higher average path length compared to TTs, adjusted R^2 = .82, $F(1, 3201.93)$ = 32.05, $p < .001$, 95% CI [−0.020, −0.097]. With regard to verbs, LTs presented higher clustering coefficient, adjusted R^2 = .22, $F(1, 1945.19)$ = 24.25, $p < .001$, 95% CI [−0.012, −0.005] and lower average path length than TTs, $F(1, 1946.00)$ = 22.72, $p < .001$, adjusted R^2 = .21, 95% CI [−0.007, −0.017]. These network results agree with our

contextual diversity results, since vocabularies with high average contextual diversity are expected to show higher clustering coefficients and lower path length than vocabularies with low average contextual diversity.

In sum, both contextual diversity and network analysis of LTs and TTs indicate differences in their underlying semantic maturation driving noun and verb production. The larger effect for verbs is, as we note in the discussion, potentially predicted by prior work on LTs. In Study 2, we focus on addressing whether or not semantic maturation added to prior models of lexical acquisition can produce the different patterns of contextual diversity we see here for verbs and nouns.

STUDY 2: COMPREHENSION-EXPRESSION MODELS

The differences between nouns and verbs and TTs and LTs offer a unique opportunity to investigate a potential role for semantic maturation in word development. To do this, we develop a series of models that each test-specific hypotheses about word development. The examination of the parameters of the best model will help to elucidate the origin of the opposing differences that we have observed between LTs and TTs depending on the type of word analyzed. Currently, the two best existing semantic network models are Preferential Acquisition (PA) and Lure of the Associates (LA), introduced by Hills et al. (2009a, 2010), which perform best when their learning rule is based on contextual diversity, as opposed to, for example, frequency, phonology, or perceptual features (e.g., Bilson et al., 2015; Stella et al., 2017). In what follows, we first describe the PA model and the LA model and then introduce a series of extensions to evaluate the following: (a) semantic maturation in the preverbal phase (progressive preferential acquisition [PPA]), and (b) feedback from productive vocabulary to comprehension (progressive lure of associates [PLA]). These two models further allow us to evaluate (c) differences in semantic maturation between nouns and verbs, and (d) differences in growth rules between preverbal and verbal development. For completeness, we also verified that models based on frequency did not fit the observed data as well as contextual diversity did, which, following prior work (e.g., Hills et al., 2009a, 2010), is the learning cue we use here. All models are shown in Figure 3.

Methods

Preferential Acquisition model

Preferential acquisition proposes that children leverage contextual diversity in the language learning environment to learn new words. A word i is selected to be learned out of a pool of words to be learned W (nouns
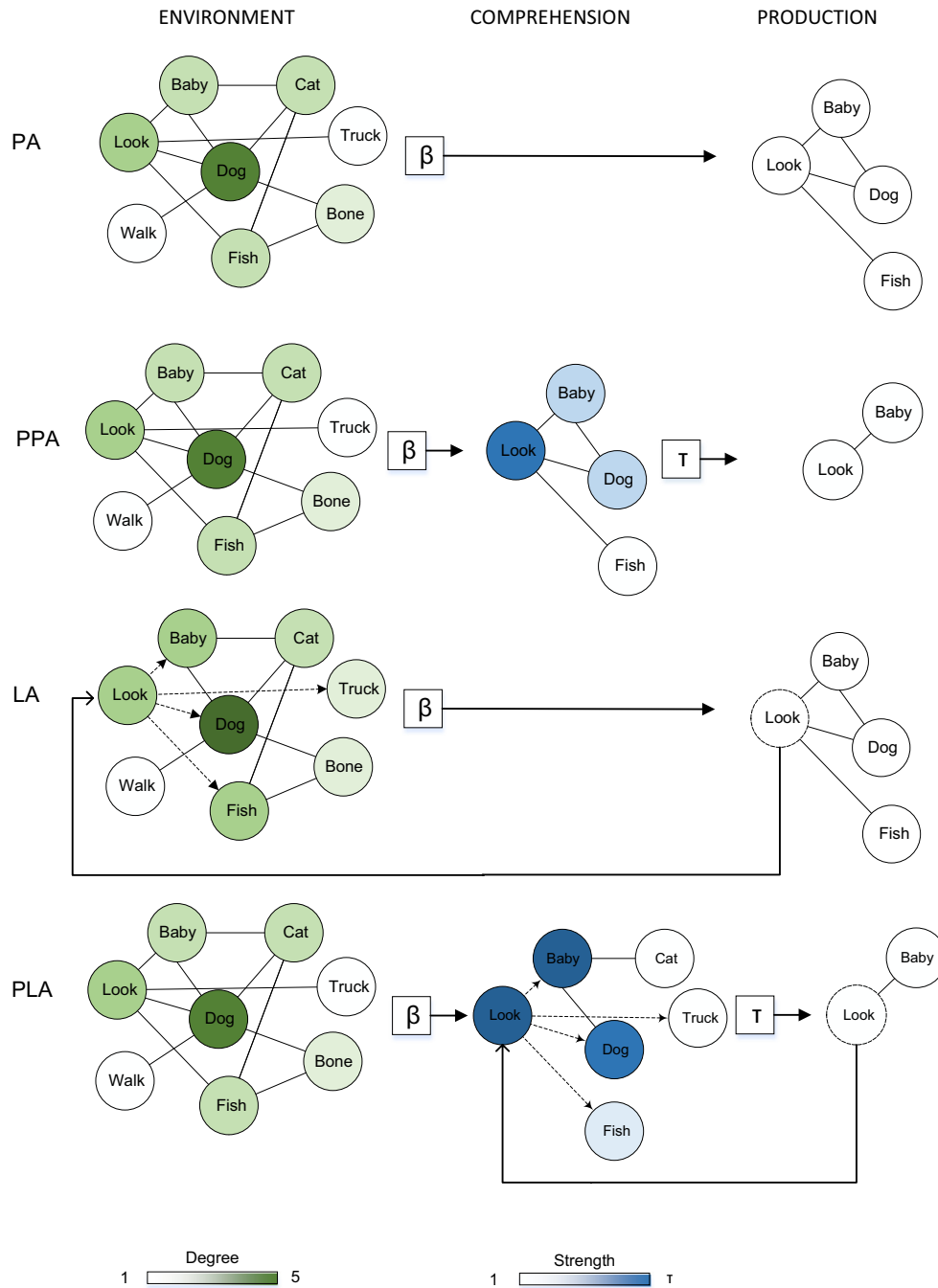
FIGURE 3 The four primary models introducing semantic maturation in a comprehension layer and feedback from production to comprehension. *Note*: From top to bottom: Preferential acquisition (PA), progressive preferential acquisition (PPA), lure of the associates (LA), and progressive lure of the associates (PLA). For each model, words are sampled from the learning environment according to Equation (1) with sensitivity parameter (*β*). After each sample, the word is either added to production (PA and LA) or its semantic maturation is boosted in the comprehension network (PPA and PLA). In PPA and PLA, when words exceed a maturation threshold (*τ*) in the comprehension network, they move into the productive vocabulary. In the LA and PLA models, movement into the productive vocabulary leads to an additional boost for associates either in the environment network (LA) or in the comprehension network (PLA). Degree is synonymous with contextual diversity and it is determined by the number of links each word has with other words in the learning environment.

and verbs) with a probability proportional to its contextual diversity in the learning environment:

$$P(i) = \frac{(K_i + 1)^{\beta}}{\sum_{i \in W} (K_i + 1)^{\beta}}. \tag{1}$$

The probability of word *i* being selected depends on its contextual diversity value K_i, computed as described in Study 1. The denominator sums these values over all possible words, so the total probability sums to 1. Sensitivity to contextual diversity, represented by the *β* in Figure 3, is fit to the data and is the primary factor influencing how

words move from the learning environment into the productive vocabulary. When β is greater than 0 there is a preference to add more contextually diverse words to the network. When β is 0, all words are treated equally with respect to contextual diversity. Words are chosen based on Equation (1), one word at a time, until no unknown words remain to be learned. Each time that the model samples from the environment it represents a situation where the child has experienced a word in numerous different linguistic contexts.

Lure of the Associates model

The LA model proposes that the words that the child knows influence the next words to be learned: the words in the language environment that have semantic connections with known words have an increased probability of being acquired earlier. The LA model further leverages the principle of mutual exclusivity underlying PA by allowing known words that appear together with unknown words to further facilitate their acquisition (Hills et al., 2009b). Hills et al. (2009b, 2010) found roughly equivocal performance between PA and lure of associates, both outperforming a variety of other models. The additional data and differences between TTs and LTs and nouns and verbs in our present analysis allow us to further tease apart these model differences.

The LA model follows the PA model but extends it one step further by introducing a mechanism for known words to enhance the learning probability of related words, such that words associated or linked to a newly produced word gain an additional *boost* to their probability in the environment. This is represented for the LA model in Figure 3 by the arrow from production to environment and by the dashed arrows in the environment network radiating out from produced words to near associates. For completion, we also developed a version of the LA model starting with a uniformly random selection (i.e., all the contextual diversity values of the environment were equal to zero at the beginning of the simulation). However, this version of the LA model showed a worse fit to the observed data than the original version described above (mean squared error [MSE] for nouns = 534.8; MSE for LTs = 373.2; note the lower MSE values of the original version in Table 3), and for this reason, we decided to only include the original LA model in our analysis below.

Progressive Preferential Acquisition model

In the PPA model, word choice follows the PA model, but words move first into a comprehension layer. Each time a word is chosen using Equation (1), its value in the comprehension layer receives a maturational *boost*, increasing the strength of the words' semantic representation in the comprehension network. We used a value of 0.5 for *boost*, but a value of 1 produced similar results. Words continue to mature in the comprehension network until their maturation reaches a threshold represented by the τ in Figure 3, at which point it moves into the productive vocabulary. Words are sampled using Equation (1) until no unknown words remain to be learned. PPA adapts to the PA framework the maturational proposal in the prior simulation work of McMurray (2007) and Nematzadeh et al. (2014).

Each time a word is sampled from the learning environment, the semantic maturation of the word is increased in the child's word representation. The PPA model assumes that the child recognizes the label of the referents to be able to identify them in the language environment and make use of contextual diversity. This means that the comprehension vector exclusively represents what the child has learned about a word through contextual diversity.

Progressive Lure of Associates model

The PLA model assumes that words move into comprehension and production exactly as defined for the PPA model. However, PLA adds an additional mechanism for boosting the comprehension of words, as in the LA model (*boost* = 0.5). This is represented for the PLA model in Figure 3 by the arrow from production to comprehension and by the dashed arrows in the comprehension network radiating out from produced words to near associates. The PLA model follows the rationale that a produced word provides semantic information about associated words in the language environment. This boosts all associates in the receptive vocabulary, including those which currently have a comprehension value of zero. Like the PPA model, this approach has not been proposed in the past, but it extends the logic of the LA model to a period of semantic maturation.

In the base versions of PPA and PLA, we allow β to change dynamically between the preverbal, β_p, and verbal, β_v, phase, such that $\beta_p \neq \beta_v$. In addition, the threshold for nouns, τ_n, and verbs, τ_v, are allowed to vary independently as well, such that $\tau_n \neq \tau_v$.

Extending the models

The base PA and LA models, and their extensions, the PPA and PLA models, offer a framework for further examining additional hypotheses about developmental changes in semantic maturation. Furthermore, they also allow us to test some of their underlying assumptions. In particular, we can test the validity of the dynamic change in sensitivity to contextual diversity underlying semantic maturation between preverbal and verbal development, as well as differences in

semantic maturation between nouns and verbs. We do this as follows.

Changes between preverbal and verbal development

Recent work suggests there may be developmental changes in sensitivity to contextual diversity during early word learning (Stella et al., 2017). To test this, we compare the dynamic $\beta_p \neq \beta_v$ models described above (PPA^0 and PLA^0) with models that use a single, β, such that $\beta_p = \beta_v$ (PPA^1 and PLA^1), which assume that the sensitivity to contextual diversity is the same during preverbal and verbal development.

Differences in semantic maturation between nouns and verbs

To evaluate the potential for differences in semantic maturation between nouns and verbs, we extend the base models described above, for which $\tau_n \neq \tau_v$, by comparing them with a single τ model (PPA^2 and PLA^2), which assumes that nouns and verbs mature at a similar rate.

Parameter estimation and model comparison

Though maximum likelihood models were initially described for parameter estimation and model comparison of the generative network growth model underlying PA and LA (see Hills et al., 2010), similar analytical solutions for PPA and PLA are computationally intractable due to their complicated dependency structures (see Hills et al., 2010). We solve this problem using two approaches that trade-off search efficiency (grid search) versus model complexity penalization (Approximate Bayesian Computation [ABC]). Both of these methods confirm the findings of the other regarding model comparisons. Both models are also approximations with respect to parameter estimation—a common challenge for rugged high-dimensional landscapes. Our aim here is to capture the qualitative switch between nouns and verbs in LT and TT populations, while also providing the best quantitative fit to the data. All data and code to replicate our results are available at the Supporting Information, but we describe our approaches in words as follows.

To identify the best fitting parameters, we used a grid search, which provided uniform coverage of the search space. We used 500 vocabulary growth simulations for each set of parameters and for each model, with the optimal parameters identified as those that minimized the MSE between the observed and simulated data. To compute the MSE, we split the simulated ordered vocabulary into a noun ordering and a verb ordering (following Figure 2). The average contextual diversity was computed for the words in each vocabulary size network, creating a vector, V_i, of contextual diversity scores corresponding to a single developmental trajectory. The

mean of these trajectories, $\overline{V}$, was computed for the 500 simulations and compared with a similar vector for the observed data, V_o, which was computed using the data presented in Study 1. The optimal parameters that minimized the MSE between $\overline{V}$ and V_o are presented in Table 3.

Though comparison of the MSE between models is consistent with the model comparison we describe below, MSE does not take into account model complexity. Thus, we turn to ABC, which is well-adapted for complex model comparisons where maximum likelihoods are unavailable, and which has growing popularity in fields such as evolutionary biology and genetics (e.g., Fraïsse et al., 2018). ABC randomly samples model parameters from prior distributions and computes posterior likelihoods via model simulation by comparing simulated and observed data using summary statistics (Hartig et al., 2011; see also the *abc* R package, Csillery et al., 2012). For each model's free parameters, we used uniform priors set to bound the optimal parameter values identified using the grid search described above. Prior distributions were identical across models: $\beta_p = [0,2]$; $\beta_v = [0,1]$; $\tau_n = [2,14]$; and $\tau_v = [2,18]$. We then randomly sampled a set of parameters from the prior distributions and used them to simulate vocabulary growth trajectories, iteratively sampling from the pool of available words ($n = 382$). Following the suggestion of Beaumont et al. (2002) regarding reducing summary statistic complexity, we averaged across progressive size networks for each simulated growth trajectory in bins of size 20 to produce a simulated growth vector for that parameter set, $\overline{V}_s$, and a corresponding vector for the observed data, V_o. In total, we simulated a total of 100,000 simulations for each model.

The posterior distribution is composed of those samples for which the Euclidian distance between the simulated and observed summary statistics, $\rho(V_s, V_o) = |V_s - V_o|$, is below a threshold tolerance, formally: $\rho(V_s, V_o) < \epsilon$. Accepted parameters approximate the posterior distribution (Hartig et al., 2011). The model comparison and Bayes factors were calculated based on the posterior distributions using the "abc" R package using the local regression method, with similar results for the best model using the neural network method (see Csillery et al., 2012 for details).

We further query whether a poor language input could produce different results in our simulations. To do this, we artificially impoverished the language input by randomly deleting links between words from our matrix of word co-occurrences. Specifically, we deleted 10,000 links (1 s) out of the original 37,915 links. This reduced the average contextual diversity values for words in the learning environment from ($M = 198.51$, $SD = 92.36$) to ($M = 146.15$. $SD = 68.37$). These rich and poor environments provide a means to test the robustness of our models to impoverished amounts of contextual diversity in the environment.

Results

Table 2 displays the posterior probabilities of all the models in comparison with one another, split by talker type. The posterior probabilities indicate that the base PPA has the highest likelihood to have generated both the LT and TT target observed data. For completeness, the best parameter values for each of the four base models are shown in Table 3, including the most probable PPA. Critically, for the PPA model, the attention parameters are not different between the LT and TT populations. However, semantic maturation thresholds for nouns and verbs are different from one another, and also different between talker type. LTs have a smaller gap between verbs and nouns than TTs, which leads LTs to start adding verbs to production sooner after they start producing nouns.

We also asked whether different results can be achieved by utilizing an artificially impoverished language input (the rich and poor environments described above). We found that the computational models took longer to learn the entire vocabulary in the poor environment, but the same relative results were observed for TT and LT simulations with respect to noun and verb contextual diversity. We evaluated this in two ways: First, TT simulations with a rich environment (i.e., the original contextual diversity values) compared to LT simulations with a poor environment (i.e., contextual diversity values calculated from an impoverished matrix of co-occurrences); and second, TT simulations with a poor environment compared to LT simulations with a rich environment. Using the best parameters from Table 3, the results generated from the PPA model show that the LT simulations produced lower average contextually diversity than the TT simulations for nouns (LTs-poor vs. TTs-rich: adjusted $R^2 = .81$, $F(1, 180,989.1) = 2530$, $p < .001$, 95% CI [1.96, 2.12]; LTs-rich vs. TTs-poor: adjusted $R^2 = .81$, $F(1, 180,989) = 2950$, $p < .001$, 95% CI [2.15, 2.31]), and higher average contextual diversity than the TT simulations for verbs (LTs-poor vs. TTs-rich: adjusted $R^2 = .33$, $F(1, 60,990.99) = 1467$, $p < .001$, 95% CI [−3.07, −2.77]; LTs-rich vs. TTs-poor: adjusted $R^2 = .30$, $F(1, 60,991.63) = 1431$, $p < .001$, 95% CI [−2.97, −2.67]). Thus, the relative difference in noun and verb learning between TTs and LTs appears to be robust to modest changes in the contextual diversity of the environment.

We simulated the vocabulary growth of the four base models and visualized the results in Figure 4. To do this, we produced 500 networks for each model using the best parameters in Table 3 for each talker type. Then, we averaged the contextual diversity of each network at each vocabulary size, separately for nouns and verbs. As can be seen in Figure 4, the PPA model (orange lines) is the only one that exhibits a similar growth curve to what was observed in children's vocabularies in both nouns and verbs (black lines). It also captures the qualitative shift in contextual diversity between nouns and verbs found in Study 1.

Why do LTs produce nouns with lower contextual diversity but verbs with higher contextual diversity compared to TTs?

One of the peculiarities of the above results is that the semantic maturation required for a word to enter the productive vocabulary seems to show a nonlinear relation with contextual diversity: TTs have a higher τ_n than LTs and also have higher contextual diversity for nouns; TTs have a higher τ_v than LTs, but have a lower contextual diversity for verbs. To disentangle this relation, we present here a series of simulations that reveal the underlying nonlinear relation, especially as it pertains to the relation between the semantic maturation for nouns, τ_n, and the semantic maturation for verbs, τ_v.

First, we generated PPA simulations from a noun-only environment using 10 different increasing τ_n values. The results confirmed that higher taus always result in higher average contextual diversity in noun vocabularies. Similar results were achieved when generating verb vocabularies in a verb-only environment. This suggests that the differences we observe in our data, which sample from both nouns and verbs simultaneously, may be a result of the gap in τ_n and τ_v values in the PPA model, which is larger for TTs than for the LTs (see Table 3). To examine this, we tested whether the size of this gap determines the average contextual diversity of the resulting simulated vocabulary. Simulating the vocabulary growth of 1000 vocabularies of size 50 using the PPA model with different $\tau_n - \tau_v$ size gaps, Figure 5 shows that when the size of the $\tau_n - \tau_v$ gap is narrow, the model produces vocabularies with higher average contextual diversity for verbs. As the gap increases, the contextual diversity for verbs falls.

This nonlinearity highlights the importance of different preverbal and verbal learning strategies. Note that

TABLE 2 Posterior probabilities of the base models and their extensions

Talker type	Model							
	PA	LA	PPA^0	PPA^1	PPA^2	PLA^0	PLA^1	PLA^2
TT	0.0000	0.0000	0.5006	0.0310	0.3776	0.0439	0.0088	0.0381
LT	0.0000	0.0000	0.6648	0.0017	0.3204	0.0062	0.0001	0.0068

Note: ϵ is .10; similar results are found with $\epsilon = .20$ and $\epsilon = .30$. ^{0}: $\beta_p \neq \beta_v \tau_n \neq \tau_v$. ^{1}: $\beta_p = \beta_v \tau_n \neq \tau_v$. ^{2}: $\beta_p \neq \beta_v \tau_n = \tau_v$.

Abbreviations: LA, lure of the associates model; LT, late talkers; PA, preferential acquisition model; PLA, progressive lure of the associates model; PPA, progressive preferential acquisition model; TT, typical talkers.

TABLE 3 Best parameter values for each of the best fitting base models

| Parameter | Models | | | | | | | |
| | PA | | LA | | PPA | | PLA | |
	LT	TT	LT	TT	LT	TT	LT	TT
β_p					0.75	0.75	1.3	1.3
β_v	2.02	1.69	1.99	1.72	0	0	0.1	0.2
τ_n					6.5	7	2.5	1.5
τ_v					10	18	3	8.5
MSE	492.1	355.5	477.0	346.5	364.9	229.5	447.6	329.02
Model probability	.0000	.0007	.0000	.0000	.9901	.9161	.0099	.0832

Note: The best parameter values for each base model were calculated after minimizing MSE. Approximate Bayesian Computation was used to compute models' posterior probabilities. ϵ is .10; similar results are found with $\epsilon = .05$, $\epsilon = .20$ and $\epsilon = .30$.

Abbreviations: LA, lure of the associates model; LT, late talkers; MSE, mean squared error; PA, preferential acquisition model; PLA, progressive lure of the associates model; PPA, progressive preferential acquisition model; TT, typical talkers.

because $\beta_p = .75$ changes to $\beta_v = 0$ once children learn their first word and become verbal, a larger $\tau_n - \tau_v$ gap means verbs spend more time accumulating information unrelated to contextual diversity ($\beta_v = 0$). This drives down the contextual diversity of verbs more for TTs than LTs, as TTs have a larger $\tau_n - \tau_v$ gap. To demonstrate this, we simulated an artificial learning environment (1000 times) with 18 nouns and 18 verbs, with each consecutive noun and verb assigned a matching but increasing contextual diversity value (e.g., 1, 10, 10, 20, 50, 100, …). We then simulated learning using the PPA model, $\beta_p = .75$ and $\tau_n = 7$, until the first word was learned and the learner transitioned to the verbal phase, $\beta_v = 0$. Then, to simulate the $\tau_n - \tau_v$ gap, we allowed the comprehension vectors to semantically mature until either 4 or 10 nouns were moved into the productive vocabulary, representing a small-gap or large-gap version, respectively. For each version, we repeated this procedure 1000 times and then sampled eight verbs in proportion to their average comprehension values over the 1000 simulations. In short, we created 1000 small or large gap verb vocabularies, with each based on 1000 periods of semantic maturation that differed, respectively, in the length of verbal period ($\beta_v = 0$). As expected, the average contextual diversity of the verb vocabularies generated using the large-gap comprehension values were lower than the verb vocabularies generated using the small-gap comprehension values (large-gap: $Mdn = 140.13$, $M = 140.16$; small-gap: $Mdn = 151.37$, $M = 151.53$; $W = 378,406$, $p < .001$). In sum, the $\tau_n - \tau_v$ gap drives down the contextual diversity of verbs for TTs relative to LTs, because TTs sample more information about verbs that is unrelated to contextual diversity ($\beta_v = 0$).

DISCUSSION

The present results demonstrate that comparison of the comprehension-expression gap between LT and TT populations provides a useful entry point for examining the potential role of semantic maturation in word development. Our best performing model, the PPA model, suggests that words that enter the receptive vocabulary experience a period of semantic maturation before moving into production. Nouns and verbs also mature at different rates, with verbs requiring more semantic maturation before production. In addition, contextual diversity appears to be most important during the pre-verbal phase, with the sensitivity to contextual diversity falling to zero during the verbal phase. Though contextual diversity has been found to predict productive vocabulary growth in prior work (e.g., Hills et al., 2010), the present research extends this work by suggesting that the role of contextual diversity in driving productive vocabulary growth may be a holdover from its influence on comprehension.

Although our observed data indicate differences in the role of contextual diversity between LTs and TTs depending on the type of word (especially for verbs, with the largest difference between talker groups), the PPA model allowed us to tease apart this relation. We found that the relation between the semantic maturation required for nouns and verbs makes the average contextual diversity of vocabularies as a whole an unreliable measure to determine the degree in which children exploit contextual diversity for word learning. Therefore, we compare maturation thresholds to measure word semantic maturation across populations (i.e., best parameter values for the PPA model), which is different from how much of this learned information contributes to word production due to the influence of unrelated learning during the verbal phase. Our results suggest that it is not that LTs pay more attention to contextual diversity for verbs, but rather that verbs require less semantic maturation for LTs than for TTs. Our findings are consistent with those of McGregor et al. (2002) who found that children with specific language impairment showed poorer semantic knowledge about words already produced. Also, these results are consistent with those of Horvath and Arunachalam (2021) who found that
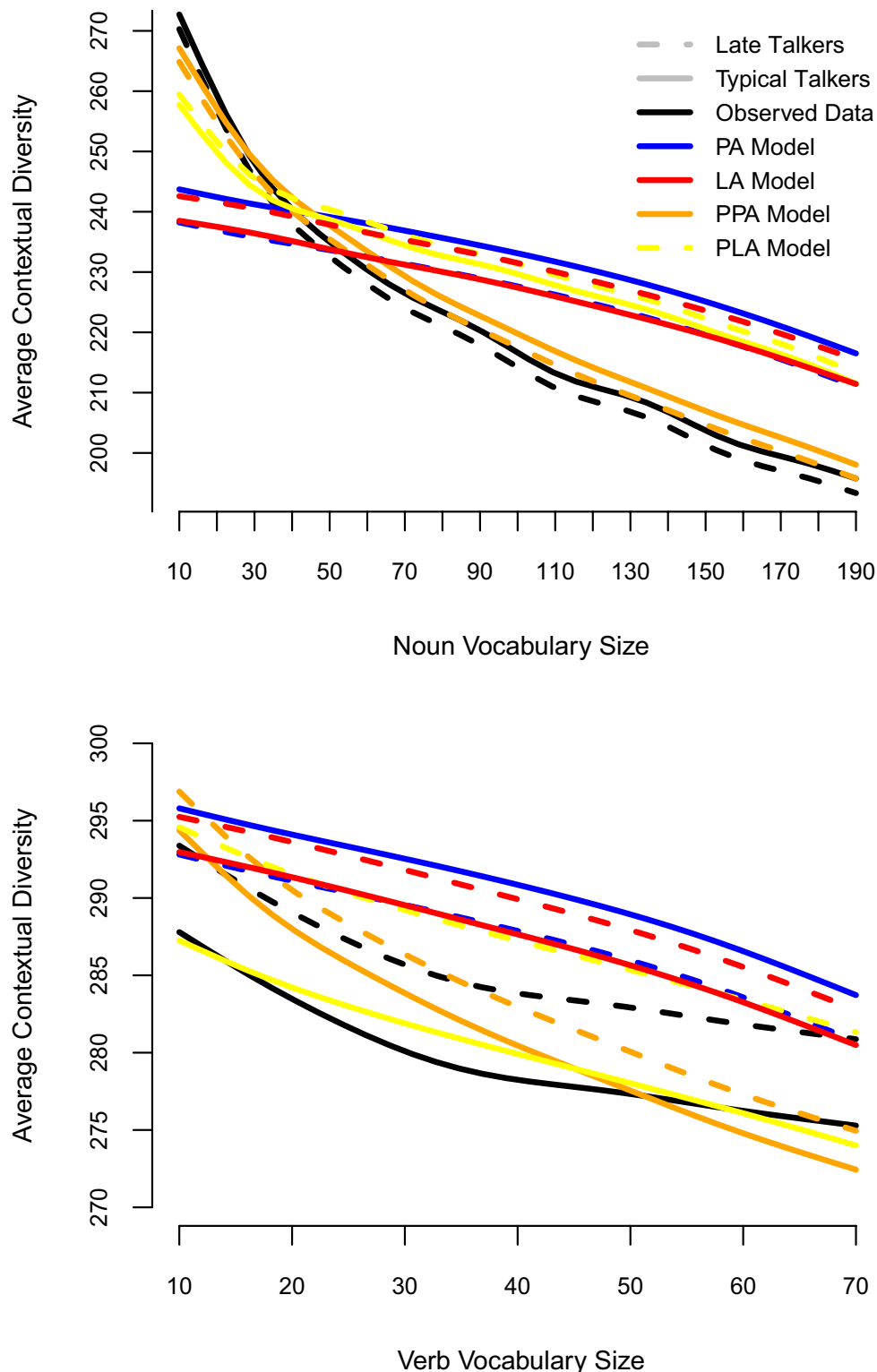
FIGURE 4 The predicted average contextual diversity of the noun and verb vocabularies of TT and LT populations as modeled by the four base computational models. *Note*: Preferential acquisition (PA), lure of associates (LA), progressive preferential acquisition (PPA), and progressive lure of associates (PLA). Best parameter values calculated for each base model were used for each talker type. All models are compared against the observed data, which is in black. Each model is represented by a color and each type of talker is represented by a type of line.

children, particularly LTs, formed stronger verb representations when the novel verb was presented in a consistent linguistic context.

The smaller semantic maturation gap between nouns and verbs in LTs makes them produce verbs sooner after nouns than TTs. This earlier verb production by LTs may
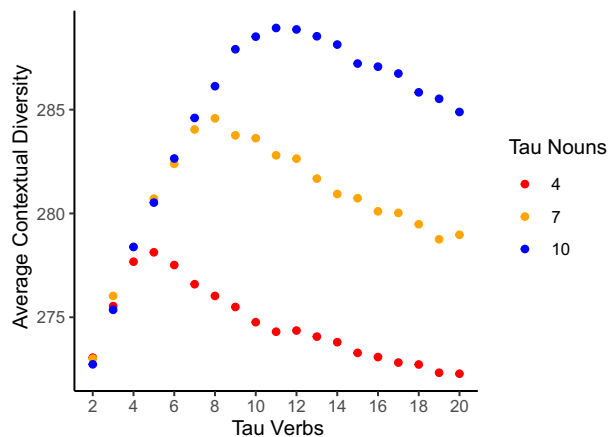
FIGURE 5 Average contextual diversity of verb vocabularies using different tau values. *Note*: The PPA model was utilized to generate the vocabularies, with $\beta_p = .75$ and $\beta_v = 0$, as in Table 3.

explain the reduced noun-verb production gap found in children with early language delay (i.e., proportion of verbs produced subtracted from proportion of nouns produced; Jiménez et al., 2020). We also found a small difference between the amount of semantic maturation needed for nouns between LTs and TTs. However, the influence of word concreteness and frequency may explain the noun differences between the talker groups. This suggests that previous findings, such as Beckage et al.'s which included different types of words in their vocabulary analysis, could have been driven by the influence of the verbs alone. In fact, contextual diversity of verbs is known to better correlate with age of acquisition than nouns (Hills, 2012). The inclusion of more than one type of word in the analysis could also be the reason why our network results do not corroborate Beckage et al.'s findings. In sum, LTs are delayed off the starting block, but once they start using contextual diversity to learn words, they require less information before moving words into production, giving rise to the signature pattern of noun and verb development we see here.

What factors influence these differences in semantic maturation is still an open question. These may be related to innate abilities or the learning environment, or a combination of both. The variation among families in the quantity and quality of child-directed speech is known to predict children's vocabulary growth (Hart & Risley, 1995; Hoff & Naigles, 2002). This is an important issue, as the differences between LTs and TTs might be related to differences in the degree of contextual diversity in parental speech, and not so much on children's innate abilities to exploit contextual diversity as a word learning strategy. However, our model simulations generated with an artificially impoverished language input demonstrated that the same differences between talker types are preserved. This is possible since the distribution of contextual diversity values is the same, just with lower values of contextual diversity than the original richer

environment. Therefore, to obtain different simulation results the distribution of contextual diversity must be different: caregivers of LTs would need to produce speech in which the relative distribution of contextual diversity differed from that of TTs. However, this may be constrained by linguistic limitations on word usage. On the other hand, impoverished environments, as the one that we artificially created, did lead to simulated learners acquiring words more slowly. This supports the idea that one of the factors that might slow down word learning in LTs is the lower semantic richness in the parental speech they experience at home. Nonetheless, LTs can use cross-situational evidence to learn words in a word learning intervention (Alt et al., 2014) and thus future work needs to look more closely at how the learning environment for individual children predicts word production. This is the focus of ongoing work in numerous laboratories including our own. The insight into semantic maturation from the PPA model provides an advance in computational approaches toward achieving that.

Our results found similar attention parameters between LTs and TTs, meaning that both populations are equally sensitive to contextual diversity in the environment. This corroborates Vuksanovic and Bjekic's findings (2013), which found that LTs have a comparable frequency of joint attention as TTs in word learning situations. However, given that the input provided to our best model is the same across talker groups and that there are no differences in attention, one might wonder what other child-level variables might be causing the lower semantic maturation. What our model is telling us is that, for TTs, the semantic information learned through contextual diversity carries less importance for verb production. This is because the TT simulations proportionally learned more information unrelated to contextual diversity than the LT simulations before producing verbs, affecting the weight of contextual diversity in production. This suggests that TTs might be relying more strongly on other types of cues to support word production as they are further along in the verbal phase. Though contextual diversity is one of the stronger predictors of early word learning, numerous other word features have been shown to be at play, such as phonological cues, perceptual features of the objects words refer to, concreteness, and other semantic features (e.g., Engelthaler & Hills, 2017; Stokes et al., 2019; Storkel & Lee, 2011).

The results of the present work suggest that the factors mentioned above may be especially important after the preverbal phase, where contextual diversity appears to play less of a role. In this sense, contextual diversity may set the stage for future development, but may not remain an active player in word learning once children become increasingly active in their language learning. In fact, we suggest in our analysis that the verbs produced by children with large noun-verb production gaps, that is, TTs, are less influenced by contextual diversity because

the relative semantic differences between words during the comprehension phase decreases due to the longer exposure to unrelated learning. This unrelated learning might well represent the impact that learning other word features has on language acquisition. Future modeling work on word maturation should aim to incorporate the potentially dynamic roles of attention to various cues in the learning environment, especially phonological influences on word promotion from comprehension to production (Sahni & Rogers, 2008; Stokes et al., 2019; and see Stella et al., 2017).

This work has numerous limitations. One is the use of cross-sectional data to make inferences about vocabulary growth of individual children. Although this approach is not unusual, the results presented in Study 1 should be contrasted with results obtained from longitudinal data to consider individual variability. Since no demographic information about the sample was available, it is not possible to confirm the generalization of the results across different sociocultural groups. Furthermore, we cannot confirm that these results can be cross-linguistically generalized as they are based only on the English language. The distribution of words differs across language types, which could influence how children perceive the semantic relatedness between words. Although we found that a much higher semantic maturation was needed for verbs than for nouns, this finding needs to be contrasted with results generated using verb-friendly languages, where the time gap between the acquisition of nouns and verbs is shorter (e.g., Imai et al., 2008). Also, in the current work, we have lemmatized and stemmed words in child-directed speech for our analysis, however, it is known that morphemes are also relevant universal cues for noun and verb acquisition across many languages (Moran et al., 2018). Further research could evaluate if children learn more semantic information through contextual diversity when considering root words with different morphemes (e.g., *jump* vs. *jumping*). In addition, though this research has used contextual diversity to evaluate the ability of semantic maturation to explain differences between LT and TT populations, other word properties such frequency, concreteness, phonology, and even object features have been shown to play a role (Engelthaler & Hills, 2017; Gendler-Shalev et al., 2021; Hills et al., 2009a; Stella et al., 2017; Storkel & Lee, 2011). Though measures of contextual diversity are often found to be more explanatory than frequency in direct comparisons (e.g., Adelman et al., 2006; Baayen, 2010; Hills et al., 2009a, 2010; Johns & Jones, 2022), future work will be needed to evaluate the potential role of frequency and other word properties during semantic maturation. Finally, our computational models treat production and comprehension as separate and discrete phases of development. This is unlikely to be the case (e.g., Pickering & Garrod, 2013). Future work will be needed to investigate how comprehension, production, and semantic maturation themselves "mature" over the course of early learning.

In sum, our findings suggest that semantic maturation during the comprehension-expression gap is partly driven by contextual diversity, and differences in the threshold for semantic maturation help to explain the differences between TTs and LTs as well as the differences between nouns and verbs.

FUNDING INFORMATION

REFERENCES

Adelman, J. S., Brown, G. D., & Quesada, J. F. (2006). Contextual diversity, not word frequency, determines word-naming and lexical decision times. *Psychological Science*, *17*, 814–823. https://doi.org/10.1111/j.1467-9280.2006.01787.x

Alt, M., Meyers, C., Oglivie, T., Nicholas, K., & Arizmendi, G. (2014). Cross-situational statistically based word learning intervention for late-talking toddlers. *Journal of Communication Disorders*, *52*, 207–220. https://doi.org/10.1016/j.jcomdis.2014.07.002

Baayen, R. H. (2010). Demythologizing the word frequency effect: A discriminative learning perspective. *The Mental Lexicon*, *5*(3), 436–461.

Bates, E., Marchman, V., Thal, D., Fenson, L., Dale, P., Reznick, J. S., Reilly, J., & Hartung, J. (1994). Developmental and stylistic variation in the composition of early vocabulary. *Journal of Child Language*, *21*, 85–123. https://doi.org/10.1017/s0305000900008680

Beaumont, M. A., Zhang, W., & Balding, D. J. (2002). Approximate bayesian computation in population genetics. *Genetics*, *162*, 2025–2035.

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PLoS One*, *6*, e19348. https://doi.org/10.1371/journal.pone.0019348

Bilson, S., Yoshida, H., Tran, C. D., Woods, E. A., & Hills, T. T. (2015). Semantic facilitation in bilingual first language acquisition. *Cognition*, *140*, 122–134. https://doi.org/10.1016/j.cognition.2015.03.013

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*, 904–911. https://doi.org/10.3758/s13428-013-0403-5

Csárdi, G., & Nepusz, T. (2006). The Igraph software package for complex network research. *InterJournal Complex Systems*, *1695*, 1–9.

Csillery, K., Francois, O., & Blum, M. G. B. (2012). Abc: An R package for Approximate Bayesian Computation (ABC). *Methods in Ecology and Evolution*, *3*, 475–479. https://doi.org/10.1111/j.2041-210X.2011.00179.x

Engelthaler, T., & Hills, T. T. (2017). Feature biases in early word learning: Network distinctiveness predicts age of acquisition. *Cognitive Science*, *41*, 120–140. https://doi.org/10.1111/cogs.12350

Evert, S. (2008). Corpora and collocations. In *Corpus linguistics. An international handbook* (pp. 1–53). https://doi.org/10.1515/9783110213881.2.1212

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). *MacArthur Communicative Development Inventory: Users guide and technical manual.* Singular Publishing Company.

Fourtassi, A., Bian, Y., & Frank, M. C. (2019). The growth of children's semantic and phonological networks: Insight from 10

languages. *Cognitive Science*, *44*, e12847. https://doi.org/10.31234/osf.io/37npj

Fraïsse, C., Roux, C., Gagnaire, P. A., Romiguier, J., Faivre, N., Welch, J. J., & Bierne, N. (2018). The divergence history of European blue mussel species reconstructed from Approximate Bayesian Computation: The effects of sequencing techniques and sampling strategies. *PeerJ*, *6*, e5198. https://doi.org/10.7717/peerj.5198

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. https://doi.org/10.1017/s0305000916000209

Gendler-Shalev, H., Ben-David, A., & Novogrodsky, R. (2021). The effect of phonological complexity on the order in which words are acquired in early childhood. *First Language*, *41*, 779–793. https://doi.org/10.1177/01427237211042997

Gentner, D. (1982). *Why nouns are learned before verbs: Linguistic relativity versus natural partitioning*. Center for the Study of Reading technical report, no. 257.

Gleitman, L. R. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*, 3–55. https://doi.org/10.1207/s15327817la0101_2

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends Cognitive Science*, *7*, 219–224. https://doi.org/10.1016/S1364-6613(03)00080-9

Goldin-Meadow, S., Seligman, M. E., & Gelman, R. (1976). Language in the two-year old. *Cognition*, *4*, 189–202. https://doi.org/10.1016/0010-0277(76)90004-4

Hart, B., & Risley, T. (1995). *Meaningful differences*. Brookes.

Hartig, F., Calabrese, J. M., Reineking, B., Wiegand, T., & Huth, A. (2011). Statistical inference for stochastic simulation models–theory and application. *Ecology Letters*, *14*, 816–827. https://doi.org/10.1111/j.1461-0248.2011.01640.x

Hills, T. (2012). The company that words keep: Comparing the statistical structure of child- versus adult-directed language. *Journal of Child Language*, *40*(3), 586–604. https://doi.org/10.1017/S0305000912000165

Hills, T. T., Maouene, J., Riordan, B., & Smith, L. B. (2010). The associative structure of language: Contextual diversity in early word learning. *Journal of Memory and Language*, *63*, 259–273. https://doi.org/10.1016/j.jml.2010.06.002

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009a). Longitudinal analysis of early semantic networks preferential attachment or preferential acquisition? *Psychological Science*, *20*, 729–739. https://doi.org/10.1111/j.1467-9280.2009.02365.x

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009b). Categorical structure among shared features in networks of early-learned nouns. *Cognition*, *112*, 381–396. https://doi.org/10.1016/j.cognition.2009.06.002

Hirsh-Pasek, K., & Golinkoff, R. M. (Eds.). (2010). *Action meets word: How children learn verbs*. Oxford University Press.

Hoff, E., & Naigles, L. (2002). How children use input to acquire a lexicon. *Child Development*, *73*, 418–433. https://doi.org/10.1111/1467-8624.00415

Hoffman, P., Ralph, M. A. L., & Rogers, T. T. (2013). Semantic diversity: A measure of semantic ambiguity based on variability in the contextual usage of words. *Behavior Research Methods*, *45*, 718–730. https://doi.org/10.3758/s13428-012-0278-x

Horvath, S., & Arunachalam, S. (2021). Repetition versus variability in verb learning: Sometimes less is more. *Journal of Speech, Language, and Hearing Research*, *64*(11), 4235–4249.

Hsiao, Y., & Nation, K. (2018). Semantic diversity, frequency and the development of lexical quality in children's word reading. *Journal of Memory and Language*, *103*, 114–126. https://doi.org/10.31234/osf.io/p9xk4

Imai, M., Li, L., Haryu, E., Okada, H., Hirsh-Pasek, K., Golinkoff, R. M., & Shigematsu, J. (2008). Novel noun and verb learning in Chinese-, English-, and Japanese-speaking children. *Child Development*, *79*, 979–1000. https://doi.org/10.1111/j.1467-8624.2008.01171.x

Jiménez, E., Haebig, E., & Hills, T. T. (2020). Identifying areas of overlap and distinction in early lexical profiles of children with autism spectrum disorder, late talkers, and typical talkers. *Journal of Autism and Developmental Disorders*, *51*, 3109–3125. https://doi.org/10.1007/s10803-020-04772-1

Johns, B. T., & Jones, M. N. (2022). Content matters: Measures of contextual diversity must consider semantic content. *Journal of Memory and Language*, *123*, 104313.

Jones, M. N., Johns, B. T., & Recchia, G. (2012). The role of semantic diversity in lexical organization. *Canadian Journal of Experimental Psychology/Revue Canadienne de Psychologie Expérimentale*, *66*, 115–124. https://doi.org/10.1037/a0026727

Küntay, A. C., & Slobin, D. I. (2002). Putting interaction back into child language: Examples from Turkish. *Psychology of Language and Communication*, *6*, 5–14.

MacWhinney, B. J. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.). Erlbaum.

McGregor, K. K., Newman, R. M., Reilly, R. M., & Capone, N. C. (2002). Semantic representation and naming in children with specific language impairment. *Journal of Speech, Language, and Hearing Research.*, *45*, 998–1014.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science*, *317*(5838), 631. https://doi.org/10.1126/science.1144073

Moore, C., Dailey, S., Garrison, H., Amatuni, A., & Bergelson, E. (2019). Point, walk, talk: Links between three early milestones, from observation and parental report. *Developmental Psychology*, *55*, 1579–1593. https://doi.org/10.1037/dev0000738.supp

Moran, S., Blasi, E. D., Schikowski, R., Küntay, A. C., Pfeiler, B., Allen, S., & Stoll, S. (2018). A universal cue for grammatical categories in the input to children: Frequent frames. *Cognition*, *175*, 131–140. https://doi.org/10.1016/j.cognition.2018.02.005

Nematzadeh, A., Fazly, A., & Stevenson, S. (2014). *Structural differences in the semantic networks of simulated word learners* (Vol. *36*, pp. 1072–1077). Proceedings of the Annual Meeting of the Cognitive Science Society.

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, *36*, 329–347. https://doi.org/10.1017/s0140525x12001495

Reilly, S., Wake, M., Bavin, E. L., Prior, M., Williams, J., Bretherton, L., Eadie, P., Barrett, Y., & Ukoumunne, O. C. (2007). Predicting language at 2 years of age: A prospective community study. *Pediatrics*, *120*(6), e1441–e1449. https://doi.org/10.1542/peds.2007-0045

Reilly, S., Wake, M., Ukoumunne, O. C., Bavin, E., Prior, M., Cini, E., Conway, L., Eadie, P., & Bretherton, L. (2010). Predicting language outcomes at 4 years of age: Findings from early language in Victoria study. *Pediatrics*, *126*, e1530–e1537. https://doi.org/10.1542/peds.2010-0254

Rosa, E., Tapia, J. L., & Perea, M. (2017). Contextual diversity facilitates learning new words in the classroom. *PLoS One*, *12*, e0179004. https://doi.org/10.1371/journal.pone.0179004

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 of the United States of America*, *112*, 12663–12668.

Sahni, S. D., & Rogers, T. T. (2008). *Sound versus meaning: What matters most in early word learning?* (Vol. *7*, pp. 280–285). IEEE International Conference on Development and Learning.

Stella, M., Beckage, N. M., & Brede, M. (2017). Multiplex lexical networks reveal patterns in early word acquisition in children. *Scientific Reports*, *7*, 1–10. https://doi.org/10.1038/srep46730

Stokes, S. F., de Bree, E., Kerkhoff, A., Momenian, M., & Zamuner, T. (2019). Phonology, semantics, and the comprehension–expression

gap in emerging lexicons. *Journal of Speech, Language, and Hearing Research*, 1–14. https://doi.org/10.1044/2019_jslhr -19-00177

Storkel, H. L., & Lee, S. Y. (2011). The independent effects of phonotactic probability and neighborhood density on lexical acquisition by preschool children. *Language and Cognitive Processes*, *26*, 191–211. https://doi.org/10.1080/01690961003787609

Thal, D., & Bates, E. (1988). Language and gesture in late talkers. *Journal of Speech, Language, and Hearing Research*, *31*, 115–123. https://doi.org/10.1044/jshr.3101.115

Vergara-Martínez, M., Comesaña, M., & Perea, M. (2017). The ERP signature of the contextual diversity effect in visual word recognition. *Cognitive, Affective, & Behavioral Neuroscience*, *17*, 461–474. https://doi.org/10.3758/s13415-016-0491-7

Vitevitch, M. S. (2019). *Network science in cognitive psychology.* Routledge. https://doi.org/10.4324/9780367853259

Vitevitch, M. S., Chan, K. Y., & Goldstein, R. (2014). Insights into failed lexical retrieval from network science. *Cognitive Psychology*, *68*, 1–32. https://doi.org/10.1016/j.cogpsych.2013.10.002

Vuksanovic, J., & Bjekic, J. (2013). Developmental relationship between language and joint attention in late talkers. *Research in Developmental Disabilities*, *34*, 2360–2368. https://doi.org/10.1016/j.ridd.2013.04.017

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*, 3–36.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.