Superordinate Word Knowledge Predicts Longitudinal Vocabulary Growth

Molly Lewis
mollyllewis@gmail.com
Department of Psychology
Carnegie Mellon University

Eliana Colunga
colunga@colorado.edu
Department of Psychology and Neuroscience
University of Colorado-Boulder

Gary Lupyan
lupyan@wisc.edu
Department of Psychology
University of Wisconsin-Madison

Abstract

Does knowing certain words help children learn other words? We hypothesized that knowledge of more general (more superordinate) words at $time_1$ would lead to faster vocabulary growth as measured through vocabulary checklists administered at later timepoints. We find that this is indeed the case. Children who have similar vocabularies at $time_1$, but differ in their productive knowledge of more general words such as "animal," "picture," and "get" go on to have different rates of word learning. Knowledge of more general words is associated with faster vocabulary growth, particularly of words semantically related to the superordinate terms they are reported to produce. This positive relationship between knowledge of more general words and word learning remains even when controlling for measures of verbal and nonverbal intelligence.

Keywords: word learning; vocabulary; semantics; generality; hypernymy

Introduction

Learning a language requires learning its core vocabulary. During the first few years of life, children transition from a world of largely meaningless word forms to a world of meaningful, richly structured language. One way to understand how words transition from meaningless to meaningful is studying the relationship between word-forms and a child's external environment (e.g., Markman, 1990). For example, a child may observe the word "cup" being used in the presence of certain objects, and thereby narrow her hypothesis-space to the objects that are cups (Smith & Yu, 2008). Researchers have also studied social factors such as where a caretaker is looking during naming (e.g., Baldwin, 1993). These two strands of research have in common a focus on relationships between the words and the world. Another, less prominent strand of research has focused on language-internal factors such as the linguistic context in which a word appears (e.g., syntactic bootstrapping; Gleitman, 1990), and the relationship between the word being learned and a child's existing vocabulary (Beckage, Mozer, & Colunga, 2019; Beckage, Smith, & Hills, 2011; Hills, Maouene, Maouene, Sheya, & Smith, 2009b, 2009a). For example, Hills, Maouene, Maouene, Sheya, & Smith (2009b) showed that a good predictor of what nouns a child is likely to learn next are the semantic and phonological connections between the to-be-learned nouns, rather than connections to words the child already knows. Here, we take a novel approach to the study of language-internal factors by examining whether earlier knowledge of certain "seed" words at $time_1$ predicts a larger vocabulary at $time_2$.

Imagine two children with productive vocabularies of the same size; one child knows the more general (more superordinate) words "color," "animal," and "move" while the other knows words like "yellow," "fox," and "run." What might we predict about these children's subsequent vocabulary development knowing only this fact? We hypothesized that knowing the more general words would facilitate faster vocabulary growth as measured by the increase in productive vocabulary from $time_1$ to $time_2$. We test this hypothesis by quantifying the generality of children's productive vocabulary (as assessed through word checklist) and using it to predict how many (and which) words they go on to produce in the coming months.

Why would knowing more general words be expected to lead to faster development? One possibility is that learning such general words motivates children to seek out (or more effectively learn) their subordinates. Knowing that colors are something people talk about may motivate a toddler to learn more of the basic color terms. Another possibility is that knowing these words provides children with more powerful inductive biases (Gelman & Davidson, 2013). Learning that ants, pigeons, and dogs despite their many perceptual differences can be encompassed by a single general term— "animal"— may help in forming a more robust general semantic category that can be recruited in making inferences about new objects (it is an animal, therefore it___), as well as learning the names of those more specific categories. Yet another possibility is that knowing the category label may make it easier for children to ask adults what something is called: "What color is this?"; "What animal is this?" (a source of word knowledge that is not very well studied or understood).

The present studies were designed to look for an overall relationship between knowledge of more general words and vocabulary growth, but not to distinguish between the reasons for why this relationship would exist. Any correlational study like this one has to contend with the possibility that children who know more general words are, for one reason or another, simply better word learners more generally, and it is a difference in this more general ability that is responsible for faster word learning. We address this point in more detail below.

Study 1: Measuring word generality via hypernymy and hyponymy

Method

We quantified word generality using hypernymy and hyponymy of words in the MacArthur-Bates Communicative Development Inventory: Words and Sentences (MCDI; *N* = 680; Fenson et al., 1994) using WordNet (Miller, 1998), a structured database of English concepts organized by their semantic relations. Hypernymy was defined as the number of superordinate links above a given word (i.e., the word's 'parents'). Having few links indicates the word is close to the top of its semantic hierarchy. For example, "animal" has lower hypernymy score (6) than "dog" (14). Hyponymy was defined as the number of links below the word (i.e., the word's 'children'). High hyponymy scores are a sign that the word is part of a dense semantic neighborhood. For example, although "game" (noun) and "drink (noun, beverage)" and "game" are both listed as having 6 hypernyms; "game" has 193 hyponyms while "drink" has 23. Because the hyponymy distribution is highly skewed (some words have 100s of hyponyms), we log-transform it in all the analyses below. We carefully matched WordNet senses to those tested by the MCDI and sought to make reasonable inferences as to the likely sense parents are likely to be reporting on. Hypernymy scores could be calculated for 449 MCDI entries; hyponymy scores for 472 entries. We omit adjectives from this analysis because they are not represented in a hierarchical structure in WordNet.[1]

We validated the hyponymy and hypernymy measures by asking 71 adults (English speakers recruited on Amazon Mechanical Turk) to rate the 680 MCDI words on a 1-5 "generality" scale from "very specific" to "very general." Each word was rated by 9-11 participants. Each participant rated approximately 100 words. To make ratings more comparable across participants, we normalized each participant's response distribution to have a mean of zero and standard deviation of one.

Results

Table 1 shows examples of nouns and verbs rated as low and high in generality. Hyponymy (*M* = 2; *SD* = 1.6) and hypernymy scores (*M* = 7.3; *SD* = 3.4) calculated from WordNet were not correlated with each other ($r(447) = -0.07$, $p = 0.16$). This is somewhat surprising given that a word higher up in a semantic hierarchy (having a low hypernymy score) ought to have more words below it (and thereby a higher hyponymy score). In theory, this is true. In practice, however, words from different categories have very varied semantic neighborhoods (corresponding to very different hyponymy scores) and these differences swamp the theoretical tradeoff between hypernymy and hyponymy. Interestingly, human judgments of word generality (unscaled: *M* = 3; *SD* = 1; scaled: *M* = 0; *SD* = 1) were moderately correlated with *both* hyponymy ($r(470) = 0.43$, $p < .0001$) and hypernymy scores ($r(447) = -0.503$, $p < .0001$), i.e., words judged by human participants as more general tended to have more subordinate relations and fewer superordinate relations (Fig. 1). In an additive linear model predicting human generality judgments, both word hyponymy ($\beta = 0.32$, $SE = 0.03$, $Z = 10.09$, $p < .0001$) and hypernymy ($\beta = -0.4$, $SE = 0.03$, $Z = -12.93$, $p < .0001$) predicted independent variance in judgments ($R^2 = 0.39$).

We next asked whether the relationship between human judgments and hypernymy/hyponymy scores differed by part of speech. There was a reliable relationship between human judgments and WordNet hyponymy scores for both nouns ($r(364) = 0.46$, $p < .0001$) and verbs ($r(104) = 0.39$, $p < .0001$). The relationship between human judgments and WordNet hypernymy scores was stronger for nouns ($r(362) = -0.41$, $p < .0001$) than for verbs ($r(83) = -0.2$, $p = 0.07$).

Table 2 shows the pairwise correlation between human generality ratings, hypernyms, hyponyms and three lexical norms: Concreteness (Brysbaert, Warriner, & Kuperman, 2014), word frequency (Brysbaert & New, 2009), and age of acquisition (AoA; Frank, Braginsky, Yurovsky, & Marchman, 2017). These first-order correlations confirm some previously made claims about superordinate words (those that correspond to lower hypernymy values have later AoAs), as

Table 1: Five nouns (left) and verbs (right) rated as lowest and highest in generality in Study 1.

noun		verb	
low	high	low	high
rocking chair	animal	stop	think
rooster	people	tickle	get
tissue/kleenex	toy (object)	meow	have
carrots	picture	lick	show
toothbrush	friend	hug	do

Table 2: Pairwise correlations (Pearson's *r*) between measures of word generality and other related word measures. Correlations are for words with complete data for both measures. Hyper./Hypo. = WordNet hypernymy and hyponymy values; Gen. = average human generality rating; Conc. = concreteness; Freq. = log word frequency; AoA = Age of acquisition estimated from a MCDI database; * = $p < .01$.

	Hyper.	Hypo.	Gen.	Conc.	Freq.
Hypo.	-0.07				
Gen.	-0.5*	0.43*			
Conc.	0.61*	-0.08	-0.49*		
Freq.	-0.42*	0.43*	0.56*	-0.52*	
AoA	-0.15*	-0.15*	0.18*	-0.21*	-0.1

[1] Our measures of hypernymy and generality closely map onto the distinction between subordinate, basic, and superordinate words. However, while the latter distinction is discrete, our measures are continuous, which we believe better reflects the psychological reality of the semantic hierarchy.

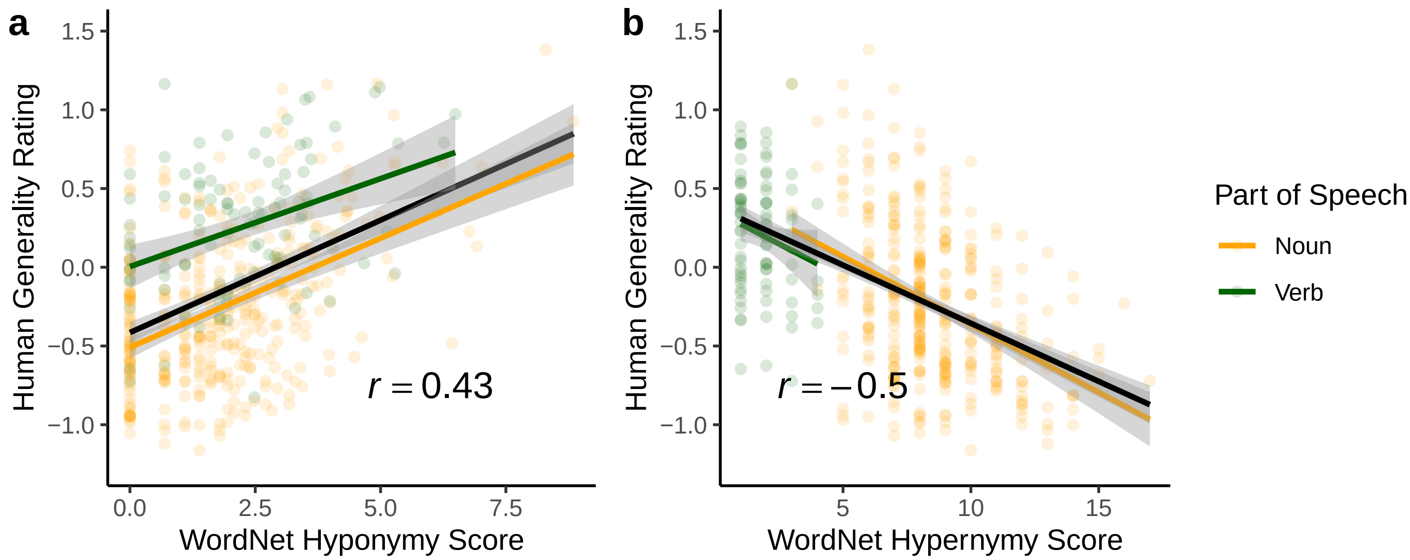

Figure 1: Normalized human generality ratings versus (a) WordNet hyponymy scores and (b) hypernymy scores. Each point corresponds to a word. Point color indicates part of speech. Lines show linear fit, with black showing the model fit across both parts of speech. Error bands indicate standard error.

well as more novel findings that greater hyponymy is associated with lower AoAs, perhaps because words in denser semantic neighborhoods are encountered in more informative learning contexts (Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976).

Taken together, our analyses suggest that word generality is a psychologically relevant property and that lexical database measures provide reasonably good estimates. This is especially noteworthy considering that WordNet contains many words and senses that our participants are unlikely to know.

Study 2: Word generality predicts vocabulary growth

In Study 2, we examined whether knowing more general words earlier on predicts faster subsequent vocabulary growth, as measured through repeated administrations of the MCDI.

Method

The MCDI was administered to a sample of 201 children (*N* females = 102) longitudinally as part of a larger study. All children were monolingual and not reported to have hearing difficulties. On average, each child completed the MCDI at 7.6 different timepoints (min = 2; max = 13). The first and last sessions occurred around 19 (*M* = 18.9; *SD* = 2) and 26 months of age (*M* = 26.4; *SD* = 3.3), respectively. Each timepoint was on average 1.3 months apart (*SD* = 0.5). A subset of the children (*N* = 18) completed several subtests of the Leiter-R IQ intelligence test during the last administration of the MCDI. We report children's performance on the fluid reasoning subtests with the caveat that although the Leiter-R has been claimed to measure children's general intelligence (Bay, 1996), there is good reason to doubt measurements of general intelligence in toddlers.

Results

Children on average learned 314.6 (*SD* = 201.9) words between the first and last administration of the MCDI (Fig. 2; 42.5 words per month; *SD* = 20.7; 54.1 words between administrations; *SD* = 37).

We fit additive mixed effect models predicting the number of words a child learned since the last administration of the MCDI as a function of a child's degree of word generality at the previous timepoint. Word generality was estimated as the mean hypernymy score of all words in the child's vocabulary. We controlled for change in age since the previous timepoint (in months), the vocabulary size at the previous timepoint, and the mean word frequency at the previous timepoint. Word frequency was estimated as the log frequency of words in adult speech in the English CHILDES corpus (North American and UK; MacWhinney, 2000). Session number was included as random slopes and child as a random intercepts.

Word hypernymy was a strong predictor of vocabulary growth: Children who knew less specific (more general) words at the previous timepoint tended to learn more words by the next timepoint ($\beta = -0.18$, *SE* = 0.04, *Z* = -5.05), controlling for the number and frequency of words they produced at the previous timepoint (as indicated on the MCDI) and the time elapsed since previous timepoint (Table 3). The same effect was observed in a model estimating vocabulary generality from the human ratings collected in Study 1, rather than WordNet hypernymy scores ($\beta = 0.23$, *SE* = 0.04, *Z* = 5.95). In contrast, hyponymy scores were not predictive of vocabulary growth ($\beta = 0.01$, *SE* = 0.03, *Z* = 0.4).

As evident in Figure 1, nouns on average have much lower

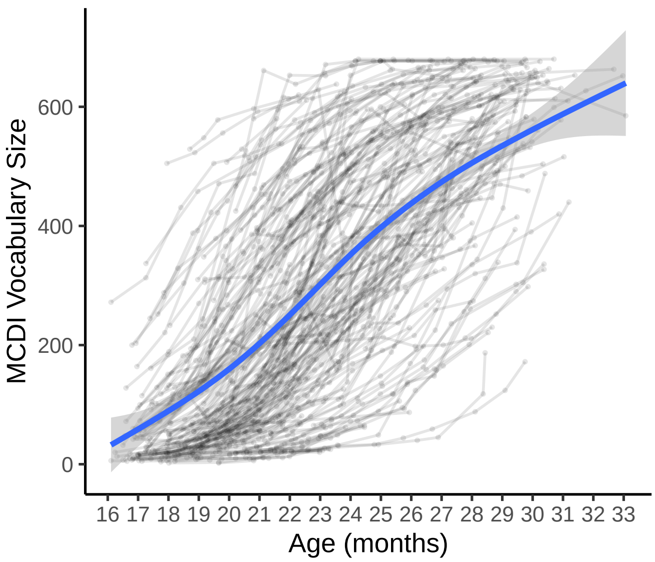

Figure 2: Longitidunal vocabulary growth. The x-axis indicates age in months and the y-axis indicates production vocabulary size as measured by the MCDI. The grey lines correspond to individual children. The blue line indicates the best fit line (LOESS fit) across children. Error bands show standard error. For visibility, the x-range is restricted to show ages older than 16 months.

hypernymy scores than verbs. This is largely a consequence of how WordNet is organized: nouns have much deeper hierarchies than verbs. To check whether the hypernymy effect we are seeing is driven largely by differences in children's vocabularies by lexical class, we scaled hypernymy by part of speech such that 0 indicates that word has average hypernymy for its lexical class. Word hypernymy remained a strong predictor of vocabulary growth (β = -0.12, *SE* = 0.03, *Z* = -4.31). Word hypernymy was also a strong predictor of vocabulary growth when hypernymy was scaled by semantic category (e.g., animals; β = -0.19, *SE* = 0.03, *Z* = -6.54).

An alternative explanation for the relationship between prior vocabulary generality and vocabulary growth is that both variables are influenced by a third variable. One plausible candidate is intelligence. It may be that children with higher general intelligence tend to both know more general words *and* be better/faster word learners. A strong test of this possibility would require a word training study. Here, we attempted to examine this possibility statistically by controlling for children's IQ within the sample of children for whom this measure was available. We fit a mixed effect linear model predicting number of words a child learned since the last administration of the MCDI as a function of a child's vocabulary hypernymy at the previous timepoint, change in age since the previous timepoint (in months), the vocabulary size at the previous timepoint, the mean word frequency at the previous timepoint and child's IQ score. Controlling for the fluid reasoning subtests of the Leiter-R, vocabulary hypernymy remained a reliable predictor vocabulary growth (β = -0.51, *SE* = 0.12, *Z* = -4.21). Fluid reasoning did not strongly predict word growth (β = 0.16, *SE* = 0.11, *Z* = 1.41).

Table 3: Parameters of mixed model predicting vocabulary change as a function of mean vocabulary hypernymy at previous timepoint, mean vocabulary size at previous timepoint, mean vocabulary frequency at previous timepoint, and change in age from previous to current timepoint.

term	Beta	SE	t-value
(Intercept)	-0.01	0.02	-0.49
Vocabulary hypernymy at t-1	-0.18	0.04	-5.05
Vocabulary size at t-1	-0.40	0.04	-9.01
Vocabulary frequency at t-1	-0.29	0.03	-9.50
Change in age	0.42	0.02	16.92

Study 3: Knowing more general words predicts learning which words?

In Study 3, we ask whether the particular words that children know at a prior timepoint predict the particular words they learn at the subsequent timepoint. Specifically, we test the hypothesis that children who know a particular more superordinate word at a prior timepoint (e.g., "animal") will learn a more specific, semantically-related word at a subsequent timepoint (e.g., "cat"). To test this hypothesis, we leverage word embedding models trained on a large corpus of English text. These word embedding models allow us to quantify the semantic distance between prior known words and newly learned words.

Method

We identified the semantic coordinates of each of the words on the MCDI for which we had hypernymy scores (*N* = 449) in a word embedding model trained on English Wikipedia (Fig. 3a; Bojanowski, Grave, Joulin, & Mikolov, 2017). We then used these coordinates to analyze the vocabularies for the sample of longitudinal MCDIs collected in Study 2.

For each child at each timepoint, we identified the semantic coordinates of the known words at a target level of hypernymy (see Fig. 3b analysis schema). We then identified the newly learned words at the subsequent timepoint that were more specific that the target level of hypernymy. To quantify the semantic distance between the newly learned words and the prior-known, more-general words, we calculated the pairwise cosine distance between each prior-known word and each newly-learned word, and averaged across pairs. This quantity ("Target Comparison") summarizes the semantic similarity of newly-learned, specific words to prior-known, general words.

We compared this quantity to a control quantity estimated by randomly sampling words known by the child at the prior timepoint from all levels of hypernymy ("Control Comparison"). We selected the same number of random words as the

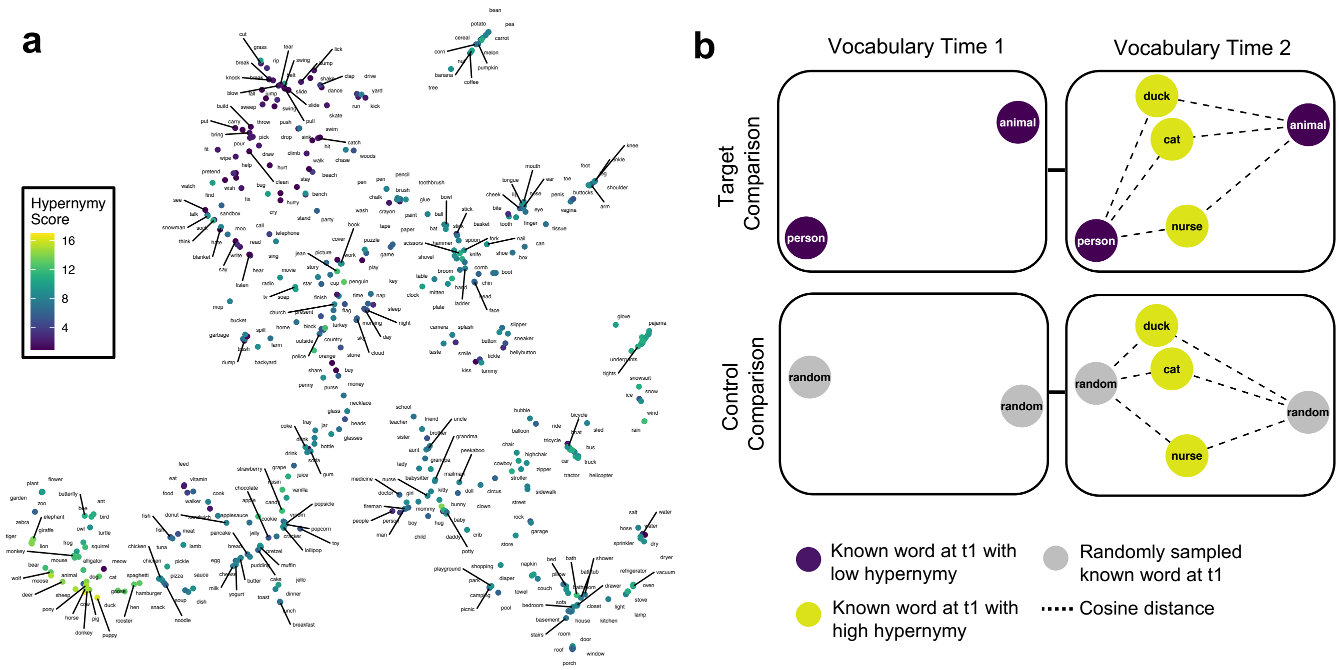

Figure 3: (a) Semantic coordinates of words on the MCDI words and sentences form, derived from a word embedding model trained in a English Wikipedia. Two dimensional coordinates are derived using the umap algorithm (McInnes, Healey & Melville, 2018). Point color corresponds to hypernymy score from WordNet, ranging from more subordinate (specific) in yellow, to more superordinate (general) in blue. Distance between points corresponds to semantic similarity. (b) Schema of Study 3 analysis.

number of known words at the target level of hypernymy. We then used these randomly selected words to calculate pairwise cosine distances in the same way as described above.

For each child, we averaged across timepoints and hypernymy levels to estimate an average cosine distance for the target and control comparisons. If newly-learned words are semantically close to prior-known, more-general words, we should expect the cosine distance in the Target Comparison to be smaller relative to the cosine distance in the Control Comparison.

Results

Relative to the control comparison (M = 0.783 [0.782, 0.784]), children tended to learn specific words that were semantically closer to the more general words they were indicated to know at the prior timepoint (M = 0.773 [0.771, 0.775]); *paired* t = -13.41; $p < .0001$; Fig. 4).

General Discussion

We show that which words a child knows predicts the rate at which they go on to learn new words. Specifically, knowing more general nouns and verbs–those that are higher up in a semantic hierarchy–predicts faster rate of word learning. This relationship holds when we control for size of the child's vocabulary at the previous timepoint and the mean frequency of the words they know. In Experiment 3, we show that these more general/more superordinate words are not associated with learning just any words, but specifically words related to the superordinate words the child knows at the previous timepoint.

Does knowing more general words help children learn words? That is, is the relationship between hypernymy and subsequent word learning a causal one? We sought to rule out the possibility that the relationship between hypernymy in early vocabulary and speed of word learning is due to a common cause such as general intelligence. The relationship between hypernymy and rate of word learning remains when controlling for fluid reasoning scores, and it is worth emphasizing that in all analyses we already control for children's initial vocabulary size which would control out the hypernymy effect if it were just that some children are better word learners in general. What our analysis cannot rule out is that children who know more general words at an early age may be in a more language-rich environment and it is this language-richness (however defined) that is responsible for their knowing these words and for their learning words at a faster rate. For example, it may be that some parents present their children with language such as "This is a chicken, it is a kind of bird. What other animals do you see?" which may lead both to greater knowledge of superordinates and faster vocabulary growth without the superordinates being the specific cause. It is also possible that differences in knowledge of more superordinate words reflects children's idiosyncratic interests. A child who learns "color" or "bird" earlier on may

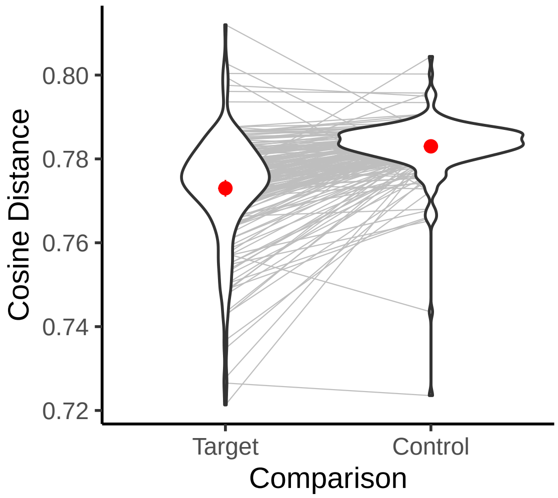

Figure 4: Study 3 results showing average cosine distance between newly learned specific words and prior known general words ("Target"), relative to a control analysis of random words ("Control"). Grey lines correspond to individual children, averaged across timepoints and hypernym levels. Red points and ranges show group means and bootstrapped 95% confidence intervals.

just be more interested in learning about colors or birds and it is this interest that is responsible for driving subsequent word growth in these semantic regions. Showing a true causal relationship between hypernymy and word growth requires a training study which we are in the process of conducting.

Our study has several limitations worth noting. Our measures of word knowledge are based on checklists completed by the child's parents or caretakers. These checklists have good validity and correlate with how well children do on more objective assessments such as Picture-Vocabulary tests. Still, these checklists are indirect measures of children's language. The use of checklists, however validated, also provides a very coarse estimate of a child's word knowledge. For example, checking off the verb "love" could mean the child uses it in a very narrow context, such as saying "I love you" or that they use it in a more adult way: to describe people, animals, objects, and actions. That we find effects of hypernymy even using such an indirect and coarse measure as a word checklist suggests that more precise language measures that, for example, incorporate word senses and contextual information, may reveal much richer connections between the words a child knows and the words they are likely and unlikely to learn next.

References

Baldwin, D. A. (1993). Infants' ability to consult the speaker for clues to word reference. *J. Of Child Language*, *20*(2), 395–418.

Bay, M. (1996). An exploratory factor analysis of the Leiter-R.

Beckage, N., Mozer, M. C., & Colunga, E. (2019). Quantifying the role of vocabulary knowledge in predicting future word learning. *IEEE Transactions on Cognitive and Developmental Systems*, *12*(2), 148–159.

Beckage, N., Smith, L., & Hills, T. (2011). Small worlds and semantic network growth in typical and late talkers. *PloS One*, *6*(5), e19348.

Bojanowski, P., Grave, E., Joulin, A., & Mikolov, T. (2017). Enriching word vectors with subword information. *Trans. Of the Assoc. For Computational Linguistics*, *5*, 135–146.

Brysbaert, M., & New, B. (2009). Moving beyond Kučera and Francis: A critical evaluation of current word frequency norms and the introduction of a new and improved word frequency measure for American English. *Behavior Research Methods*, *41*(4), 977–990.

Brysbaert, M., Warriner, A. B., & Kuperman, V. (2014). Concreteness ratings for 40 thousand generally known English word lemmas. *Behavior Research Methods*, *46*(3), 904–911.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., Pethick, S. J., ... Stiles, J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, i–185.

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *J. Of Child Language*, *44*(3), 677.

Gelman, S. A., & Davidson, N. S. (2013). Conceptual influences on category-based induction. *Cognitive Psychology*, *66*(3), 327–353.

Gleitman, L. (1990). The structural sources of verb meanings. *Language Acquisition*, *1*(1), 3–55.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009a). Categorical structure among shared features in networks of early-learned nouns. *Cognition*, *112*(3), 381–396.

Hills, T. T., Maouene, M., Maouene, J., Sheya, A., & Smith, L. (2009b). Longitudinal analysis of early semantic networks: Preferential attachment or preferential acquisition? *Psychological Science*, *20*(6), 729–739.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk. Transcription format and programs* (Vol. 1). Psychology Press.

Markman, E. M. (1990). Constraints children place on word meanings. *Cognitive Science*, *14*(1), 57–77.

McInnes, L., Healy, J., & Melville, J. (2018). Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv Preprint arXiv:1802.03426*.

Miller, G. A. (1998). *WordNet: An electronic lexical database*. MIT press.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *8*(3), 382–439.

Smith, L., & Yu, C. (2008). Infants rapidly learn word-referent mappings via cross-situational statistics. *Cognition*, *106*(3), 1558–1568.