

Children make use of relationships across meanings in word learning

Sammy Floyd
Princeton University

Adele E. Goldberg
Princeton University

Abstract

Many words are associated with more than a single meaning. Words are sometimes “ambiguous,” applying to unrelated meanings, but the majority of frequent words are “polysemous” in that they apply to multiple *related* meanings. In a preregistered design that included two tasks, we tested adults’ and 4½-7-year-old children’s ability to learn 4 novel polysemous words or 4 novel ambiguous words. Both children and adults demonstrated a polysemy over ambiguity learning advantage on each task after exposure, showing better learning of novel words with multiple related meanings than novel words with unrelated meanings. Stimuli in the polysemy condition were designed and then normed to guard against learners relying on a simple definition in order to distinguish the multiple target meanings for each word from foils. We retested available participants after a week-long delay without providing additional exposure and found that adults’ performance remained strong in the polysemy condition in one task, and children’s performance remained strong in the polysemy condition in both tasks. We conclude that participants are adept at learning polysemous words that vary along multiple dimensions. Current results are consistent with the idea that ambiguous meanings of a word compete, but polysemous meanings instead reinforce one another.

Keywords: polysemy, word learning, semantics, development, cognitive linguistics

Correspondence address: Sammy Floyd, Department of Psychology, Princeton University, South Dr., Princeton, NJ, 08544. Email: sfloyd@princeton.edu. Acknowledgements: We thank the Princeton BabyLab, Dillon Day Camp, Hopewell Elementary School, and Charles Street School for allowing us to recruit participants, to RoseMarie Stevenson for significant help in subject recruitment, to Charlotte Jeppsen and Sarah Reid for help with data collection, and to Karina Tachihara for helpful discussions on design.

© 2020, APA. *Journal of Experimental Psychology: Learning, Memory and Cognition*. This paper is not the copy of record and may not exactly replicate the final, authoritative version of the article. Please do not copy or cite without authors' permission. The final article will be available, upon publication, via its DOI: 10.1037/xlm0000821

Introduction

Learning the meaning of words is essential to successful communication. An underappreciated fact that complicates word learning is the ubiquity of **polysemous words**, which are words with multiple conventional meanings (or “senses”) that are *related* to one another (Durkin & Manning, 1989; McCarthy, 1997; Fellbaum, 1998; Rodd, Gaskell & Marslen-Wilson, 2002). It has been estimated that between 40% to 84% of English words have polysemous senses (Durkin & Manning, 1989; McCarthy, 1997; Fellbaum, 1998; Rodd, Gaskell & Marslen-Wilson, 2002; Zipf, 1945). Extended senses can be evoked spontaneously in certain predictable ways (Nunberg, 1979; Pustejovsky, 1991; Rabagliati, Marcus, & Pykkänen, 2010). For example, we can refer to a work of art by the artist’s name (*a Picasso*), and as long as the context makes it clear that the name refers to an artist, we can do this without having witnessed the name used this way previously (e.g., *A Chella Man sold for a thousand dollars*). Recent work has demonstrated that children are able to take advantage of this type of productive or “regular” polysemy as they prefer a regular extension of a word’s meaning over a semantically unrelated extension (Rabagliati et al., 2010; Srinivasan, Berner, & Rabagliati, 2018; Srinivasan & Snedeker, 2011; Srinivasan & Snedeker 2014; Srinivasan, Al-Mughairy, Foushee, & Barner, 2017). That is, when a novel word is illustrated with one sense, children may expect the word to apply in a way that is systematically related. For instance, a word that refers to a certain material can be extended to apply to an object made of that material (as exists for the English words, *glass* and *tin*; Srinivasan et al., 2018); a word that refers to an animal can be extended to apply to the meat of that animal (as exists for *chicken*; Srinivasan & Snedeker, 2014); the word for a container can be used to refer to the container’s contents (e.g., wash vs. pour a *bowl*; Rabagliati et al., 2010); a word for an object can be used to refer to the object’s abstract content (e.g., a heavy vs. interesting *book*; Srinivasan & Snedeker, 2011); and a word for a tool can be used to refer to an action performed with that tool (*a hammer, to hammer*; Srinivasan et al., 2017). While these studies involved exposing children to novel words, each of the semantic relationships between familiar and new meanings already exists in English and is likely known by children, as exemplified by the instances provided. These same extension patterns appear in other languages as well, suggesting that they may be predictable (Srinivasan & Rabagliati, 2015).

Importantly, the multiple meanings of many words are often *not* predictable on the basis of any productive generalization or regular pattern (Fillmore & Langendoen, 1971; Rabagliati & Snedeker, 2013). For example, while pen caps, baseball caps, and bottle caps are all called *caps*, the relationships among these meanings do not recur in the same way in other words. The apparent similarity among the meanings of *cap* is not sufficient to predict that the same word will be used; a *baseball cap* is called *gorra* in Spanish, while a pen cap is called *tapa (de lapicero)*. Cross-linguistic differences such as these suggest that speakers often must learn the range of meanings individual words conventionally allow in their language (Bracken, Degani, Eddington & Tokowicz, 2017; Lehrer, 1990; Malt, 2010; Murphy, 2004; Pinker, 2007; Pye, 2017). Indeed, even the patterns which we might consider regular, such as extending a label based on a shared shape across meanings, are known to develop through exposure (Smith, Jones, Landau, Gershkoff-Stowe & Samuelson, 2002). Work on “co-lexification” across languages reveals recurring patterns of shared labels for concepts that are semantically related, with recurring polysemy being more common among languages that are diachronically related or in contact (Jackson et al. 2019; San Roque, Kendrick, Norcliffe, & Majid, 2018). Work on historical semantics has emphasized that the range of meanings associated with a given word typically changes over time, which provides further evidence that conventional meanings of words are often not predictable on the basis of productive mechanisms (Brinton & Traugott, 2005; Sweetser, 1990; Murphy, 2004).

While modeling work shows that the emergence of new senses is best explained by semantic closeness to a preceding meaning (Ramiro et al., 2018), it is uncontroversial that new meanings become conventional through exposure. For example, developmental work has documented that children are influenced by differences in the ways their language conventionally encodes polysemy from early stages of production. For example, English use the same preposition, *off*, to mean “removal from a flat surface” and “removal of enveloping clothing” (*he took the plate off the table, he took off his coat*), while Dutch children learn distinct terms for those meanings (*af* and *uit*, respectively) (Bowerman, 1996; Bowerman & Choi, 2001; Choi, McDonough, Bowerman & Mandler, 1999). Here, we consider the question of how

challenging it is for children to learn and retain multiple novel, irregular conventions which may vary along more than a single dimension.

The learners' task would be easier if the meaning of each word in any language at any given point of time could be captured by a single definition or set of necessary and sufficient features. However, as Wittgenstein (1953) established, the range of conventional meanings that common words allow is typically *not* captured by a single overarching definition. For example, we might attempt to define the English word *cap* in a general way as "something which tightly covers the top of something." But this potential definition does not accurately predict when the word *cap* is appropriate and when it is not; it would incorrectly predict corks, lids, and roofs could be called *caps*, and it would not predict mushroom *caps* or *cap and trade*, as they do not tightly cover anything. In fact, words are routinely extended along multiple dimensions (see Figure 1), only loosely clustering together according to *family resemblance* structures (Wittgenstein, 1953; Klein & Murphy, 2001, 2002), *radial categories* in which a central meaning is extended along more than one different dimension (Lakoff, 1987), or *chaining* from one sense to another (Heine, 1992; Ramiro et al., 2018). The fact that a word's range of conventional senses is not predictable from a definition or single "summary" representation entails that children must learn, rather than simply predict, additional senses of many words from witnessing the conventions of their language. To date, it has not been established that children are able to learn the kinds of multidimensional and idiosyncratic relationships that commonly underlie the structure of conventionally polysemous meanings (e.g., *cap* in English). The present work therefore extends work by Srinivasan and others by investigating whether children are able to take advantage of relationships among *conventionally* polysemous meanings, which vary along more than one dimension simultaneously. We refer to words that are associated with multiple related meanings that are not predictable by a productive generalization and which cannot be subsumed by a single definition as simply POLYSEMOUS in what follows.

In order to learn multiple meanings, it is possible that word learners track meanings in a way that allows relationships among them to be represented through shared attributes or generalization across items. Clustered or overlapping meaning representations may activate and ultimately strengthen nearby related meanings through spreading activation, making it easier to integrate and retain related meanings as compared to unrelated meanings. This type of learning mechanism would predict a learning advantage for conventional polysemous meanings over words with unrelated or ambiguous meanings, for example, the word *bat* is ambiguous insofar as baseball bats and flying bats do not share distinguishing features or relationships (Figure 1). Work on children's early semantic networks has found that children tend to produce words from more dense semantic and phonological neighborhoods earlier and that even infants are sensitive to semantic relations among different words and concepts (Arias-Trejo & Plunket, 2009; Borovsky, Ellis, Evan & Elman, 2016; Steyvers & Tenenbaum, 2005; Willits, Wojcik, Seidenberg, & Saffran, 2013). This work has generally assumed that each word corresponds to a single node in a network, but it suggests that semantic relationships and domain knowledge facilitate word learning. The current work asks whether children take advantage of multiple distinct relationships among the meanings of a *single word* during learning to enhance the learning and retention of a word's meanings.

Ambiguous words are far less common than polysemous words, accounting for roughly 4 to 7% of word meanings (Dautriche, 2015; Navigli & Ponzetto, 2012; Rodd et al., 2002). This difference in how prevalent polysemy is compared to ambiguity, raises another important motivation for the current research. Recent work has offered a speaker-based functional explanation for both polysemy and ambiguity (Harmon & Kapatsinski, 2017; Piantadosi, Tily & Gibson, 2012). This work argues that speakers tend to add new meanings to existing words because repeating familiar words is easier. Evidence comes from the fact that shorter and more frequent words tend to have more senses than longer and less frequent words (Piantadosi, Tily, and Gibson, 2012), and a more frequent newly learned word is more likely to be extended than a less frequent novel word (Harmon & Kapatsinski, 2017). This form-based motivation treats ambiguity and polysemy as a unified phenomenon, and therefore does not directly address why polysemy is so much more common than ambiguity. In the current work, we directly test whether ease of learning and retention serve to advantage conventional polysemy over ambiguity from the learner's perspective, over and above the

CHILDREN USE RELATIONSHIPS ACROSS MEANINGS IN WORD LEARNING

advantages that accrue to the speaker.¹

In the following experiments, we introduce participants to novel words and novel (unfamiliar) meanings in order to control for prior knowledge, including possible metaphorical or functional interpretations. The meanings are represented by images of novel objects in order to provide well-controlled and brief exposure, as is common in word learning experiments (Gentner, 1978; Golinkoff, Hirsh-Pasek, Bailey, & Wenger, 1992; Horst & Hout, 2016; Landau, Smith & Jones, 1998; Zettersten, Wojcik, Benitez & Saffran, 2018). By using novel words and novel meanings, we avoid potential transfer effects from the language children already know, and systematically avoid the type of productive extensions that are common cross-linguistically, which have been investigated by others (Srinivasan et al., 2017; Srinivasan, Berner, & Rabagliati, 2018; Srinivasan & Snedeker, 2011).

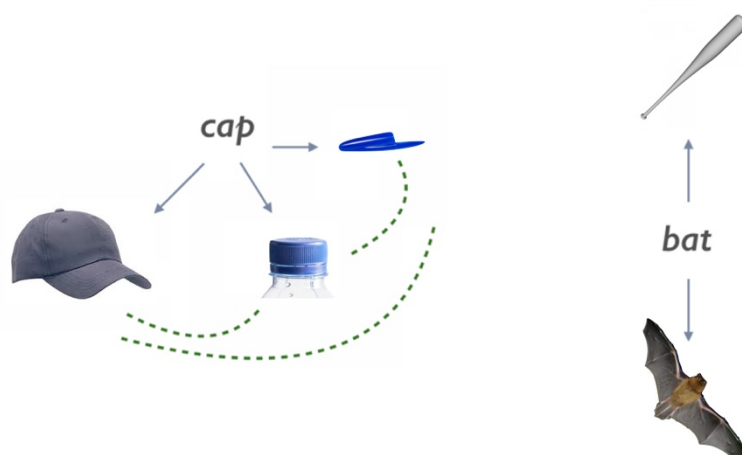


Figure 1: Polysemy: related senses associated with a single word form (e.g., *cap*) vs. Ambiguity/Homonymy: unrelated senses associated with a single word form (e.g., *bat*)².

¹ If we establish this independent motivation for polysemy, it affords a different or complementary interpretation of the correlation between ease of production and variability of meanings observed by Piantadosi, Tily and Gibson (2012). Rather than speakers' selecting words for additional meanings because they are frequent and therefore easy to produce, it may be, at least in part, that words become more frequent because they have been assigned additional meanings. That is, words that acquire additional meanings become relevant in a broader range of contexts. For example, the words *file*, *web*, *password*, increased in frequency since the 1980's because they took on meanings related to new technology (Google N-grams). An increase in frequency can be expected to lead to greater ease of production because frequently used words tend to become reduced (*family* is pronounced with 2 syllables; *needed* is pronounced more quickly than *kneaded*; *laboratory* becomes *lab*, Bybee, 1985). The causation may well go in both directions: we choose easily pronounceable (or *accessible*) words for new meanings *and* words with additional meanings become more accessible.

² *Public domain image sources*: Barracuda, 2006; Crisco, 2014; Delgado, 2012; Siedlecki, 2018; TexasRebel, 2007.

An independent line of work on non-linguistic categorization serves to highlight the challenges children face. When *adults* are asked to classify stimuli based on multiple features, they generally require dozens of exposures and explicit feedback regarding accuracy on each trial (Murphy, 2004). The learning becomes more challenging as the number of relevant dimensions increases (Shepard, Hovland & Jenkins, 1961; see also Kruschke, 1992, 2005; Nosofsky, 1984; Nosofsky et al., 1994), and as more item-level learning is required for classification (Smith & Minda, 1998). The challenge of the word-learning task children face is further highlighted by the fact that classification tasks generally ask adults to sort entities into only one or two categories at a time, and explicit feedback is provided. Word learners must navigate many more unfamiliar words, many of which are associated with multiple meanings. It is as yet unclear whether it is any easier to learn that a single label applies to a set of related meanings compared to a set of unrelated meanings.

There is reason to hypothesize that novel words with multiple meanings will be learned easier if the meanings are related to one another, stemming from two independent lines of work: work on lexical access in adults and work on non-linguistic categorization. Some of the earliest work on lexical access investigated whether all senses of ambiguous words were simultaneously activated (Swinney, 1979; Tabossi et al., 1987). To the extent that polysemy was recognized, it was suggested that listeners selectively activated an “underspecified” or vague meaning comprised of only features shared by all meanings, allowing listeners to avoid committing to a more specific interpretation until after disambiguation from context (Frazier & Rayner, 1990). In fact, words are recognized faster in lexical-decision tasks when they have multiple related meanings, whereas multiple unrelated meanings do not confer the same advantage (Armstrong & Plaut, 2008; Azuma & Van Orden, 1997; Brocher et al., 2016; Klepousniotau & Baum, 2007; Rodd et al., 2002; Rodd et al. 2012). Rodd et al (2004) has modeled the distinction between ambiguity and polysemy by positing a distributed representation of a word’s meanings in which unrelated meanings interfere with one another during retrieval in a way that related meanings do not. If, in parallel way, only ambiguous meanings interfere with one another in the initial stages of learning, we predict that it should be easier to learn and remember polysemous words when compared with ambiguous words.

Certain classic work on non-linguistic categorization also suggests that related meanings of a word should be easier to learn than unrelated meanings of a word. Rosch & Mervis (1975) found that exemplars that shared many features with other members of a category and few features with members of other categories (more *prototypical* exemplars) were easier to learn than less prototypical exemplars of the same category. At the same time, this work was performed on categories with structure that participants were already familiar with (e.g., furniture, birds, colors). Of current interest is the extent to which a new, latent complex structure will be useful to children during the initial learning of novel words. To ensure that perceptual similarity alone does not determine the requisite relationships required to learn the novel words in our paradigm, we report a Norming study below.

An additional issue raised in recent studies is that comprehension of newly learned novel words appears to degrade very quickly, particularly when participants are exposed to multiple candidate meanings at once (Aravind et al., 2018; Horst & Samuelson, 2008). This raises the possibility that participants simply rely on a strategy to satisfy experimental demands without retaining the meanings of new words (Trueswell et al., 2013), and is worrisome insofar as real-world vocabulary learning requires maintenance in long term memory. Therefore, to test whether multi-meaning words are retained by participants, we re-test all available learners after a delay of roughly a week.

To summarize, words in natural languages are commonly associated with a network of *related* conventional meanings. While polysemy is common, it is unclear whether children are influenced by relationships among meanings of a given word when those relationships must be gleaned during the word learning process itself. Current word learning models predict that all distinct potential meanings of a word should compete with one another and have no way to distinguish related from unrelated meanings. It is possible that children learn each word meaning independently, in which case learning polysemy would be essentially the same as learning ambiguity. On the other hand, related senses may reinforce each other to some extent during the learning process, predicting that polysemy should be easier to learn than ambiguity.

CHILDREN USE RELATIONSHIPS ACROSS MEANINGS IN WORD LEARNING

The present studies are the first that we know of to empirically compare the learning of polysemous and ambiguous words in a way that satisfies several desiderata. We present participants with stimuli that require them to learn word meanings that involve extensions across multiple dimensions, ensuring that no one single feature can be used to identify the meanings (this is confirmed by the additional norming task). Both novel labels and novel senses are used in order to mitigate the effects of previous experience and world knowledge, and no feedback is provided. To increase the ecological validity of the results, *multiple* novel words are learned, and we retest participants after a week delay and without re-exposure to investigate retention in long-term memory. We test 4 ½ - 7 children as well as adults, since children represent a more naïve group of participants who are less likely than adults to rely on metalinguistic strategies (Gombert, 1992) or metacognitive skills (Flavell et al., 1981), reducing the likelihood that their responses reflect learned strategies. Additionally, a number of findings in category and feature learning show young children are substantially more likely to attend widely to features, rather than categorize based on a single dimension as older children and adults tend to (Smith & Kemler 1977; Deng & Sloutsky, 2016; Plebanek & Sloutsky, 2017), indicating that young children may be well suited to learning the kinds of multidimensional word meanings that polysemous word learning requires. Finally, during this age range, children are known to learn 5-7 words per day (Cunningham, 2005), so the possibility of a polysemy over ambiguity advantage would have implications for our understanding of lexical networks as they are being formed.

We recognize that in reality, relatedness between meanings falls on a cline; while some senses are highly related, others share fewer features, and appear closer to ambiguity (Geeraerts, 1993; Tuggy, 1993). Therefore, in the two experiments described below, meanings for each word in a Polysemy condition were constructed to share similarities along two dimensions, while meanings in an Ambiguity condition were constructed by scrambling polysemy sets to reduce similarity between meanings. In a between-subjects design used in both experiments, adults and children were randomly assigned to either the polysemy condition or the ambiguity condition. During exposure, participants witnessed exposure videos which displayed a series of novel labels and pictures of novel objects (“meanings”), labeled with novel words, with pictures of foil objects interwoven.

In order to test whether participants successfully mapped the word forms to their intended targets, in a Label Matching task, Experiment 1 presented participants with four targets, each corresponding to a different word; participants were then asked to select the appropriate target for a particular word. To determine whether participants in fact construed multiple polysemous targets as targets of a single label and whether they were able to discriminate those target meanings from foils, a Sense Selection task (Experiment 2) asked participants to identify, for each of the 4 novel words, the 3 targets for each word, distinguishing them from 5 foils, which were witnessed but unlabeled during exposure. Thus, we determined whether participants were able to discriminate one target meaning from distractors which had been labeled differently (Exp. 1), and whether they could group commonly-labeled targets together in the presence of also-witnessed foils (Exp.2). All available participants were retested on the same two tasks after a week-long delay with no intervening exposure.

Experiment 1: Label Matching

As our interest is in vocabulary learning, Experiment 1 aims to determine whether adults and children were able to successfully match a novel label with its target meaning by selecting the target meaning from among 3 other meanings which had been labelled differently (distractors).

Method

Participants 84 children between ages 4½ - 7 were recruited to the lab or at local elementary schools ($M = 5:11$; $SD = 0.62$) and 84 adults were recruited from Amazon's Mechanical Turk through Turk Prime (Litman, Robinson & Abberbock, 2016). An additional six children's data was excluded from analysis for technical malfunction ($N = 2$), experimental error ($N = 2$) or missing date of birth ($N = 2$), and 3 adult participants' data was missing due to technical malfunction. Sample size and analyses were preregistered for both adults and children at the first timepoint (see supplementary material). Participants were randomly and equally assigned to the polysemy or ambiguity condition. Children's ages were matched across

CHILDREN USE RELATIONSHIPS ACROSS MEANINGS IN WORD LEARNING

conditions (Polysemy: $M = 5.88$, $SD = 0.67$; Ambiguity: $M = 5.90$, $SD = 0.57$; $\beta = -0.02$, $t = -0.375$, $p = 0.708$). Longitudinal analyses were not preregistered because we could not anticipate the rate of attrition, and thus sample size, prior to data collection. Thirty-seven children were available to be tested again after the 1 week delay (mean age = 6:2, $N = 18$ in polysemy condition, $N = 19$ in ambiguity). Fifty-nine adults were also re-recruited after the delay ($N = 31$ in the polysemy condition, $N = 28$ in the ambiguity condition). All procedures were approved by the Princeton University institutional review board. Before starting the study, adult participants and parents/guardians of child participants gave written consent.

Stimuli In the polysemy condition, one “prototypical” meaning shared a distinctive feature with a second target and a different distinctive feature with the third target, reducing the possibility that a single feature could distinguish all three target objects from foil objects. In Figure 2 with targets from the polysemy condition, the object in the center is the prototype; it shares a handle with the object on the left and a material with the object on the right, while the objects on the left and right share no distinguishing features with each other. For each of the 4 novel words, 5 foil “meanings” were also created (see supplemental materials). Thus, participants in the polysemy condition witnessed 4 novel words, each assigned three novel, *related* meanings (for a total of 12 target meanings); the videos also contained five novel foil meanings for each word (a total of 20 foils). Stimuli were normed to determine the extent to which visual similarity alone could predict results by exposing a separate group of participants to the same exposure videos without sound (see **Norming in Experiment 2** for details).



Figure 2: Example set of target objects in the polysemy condition (1 of 4 sets). The central object represents the prototype. No single feature can successfully discriminate all three targets from foils (see Figure 6 for foils and Exp. 2 Norming for confirmation).

Participants in the Ambiguity condition heard the same 4 novel labels and witnessed the same 12 target meanings and the same 20 foils, but in this condition the 12 target meanings were scrambled across the sets constructed in polysemy so that the 3 meanings of each word were relatively unrelated to one another (See Figure 3).

CHILDREN USE RELATIONSHIPS ACROSS MEANINGS IN WORD LEARNING

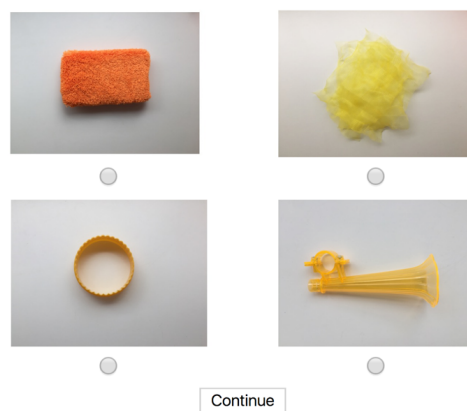


Figure 3: Example set of meanings in the Ambiguity condition (1 of 4 sets). Sets were constructed to reduce similarity between meanings. The same “prototype” images were used in both conditions.

Exposure Each of two exposure videos witnessed by each participant contained a stream of novel objects, presented one at a time for 2.5 seconds each. All target objects were named once by a human voice (“This is a *kaisee*”, labels counterbalanced), except the prototypical target object, which was shown and named twice. Ten foil images in each video were accompanied by the sound of a bell tone. The ambiguity sets were not constructed to share similarity, making it impossible to select a prototypical meaning, so one meaning of each ambiguous set was chosen at random to appear twice in the video to match exposure across conditions. Each of the 2 videos exposed participants to two novel words, each one of which labeled 3 novel target objects, offering participants in each condition the opportunity to learn 4 novel words with 12 meanings. Each video was approximately 55 seconds long, with targets and foils presented in a fixed, pseudo-randomized order, and the two videos in each condition were presented once each, in counterbalanced order.

Procedure Web scripts prevented participants from re-watching any exposure videos, and participants could not rewind to extend exposure or skip through. Children were tested in a quiet area, using noise-canceling headphones. The experiment was formatted identically for children and adults, but was shown to children on an external monitor while the experimenter sat to the child’s left, also facing the screen. Children selected answers using a wireless mouse, except if unable to operate it, in which case the experimenter selected the options which the child indicated by pointing.

Test On each of six trials, participants were prompted with one of the words they had been exposed to, and were shown four meanings, one from each of the 4 different novel words they had witnessed. See Figure 4. The meanings that had been witnessed twice in the videos (the “prototypes”) appeared on the same trials, to control for the amount of exposure. Each participant was tested on all senses of two of the four words they had witnessed. Word assignment was randomized across participants.



CHILDREN USE RELATIONSHIPS ACROSS MEANINGS IN WORD LEARNING

Figure 4: Sample test stimuli in the Label Matching task. Each of the four objects displayed had been labeled by a different word during exposure.

Results

We entered adult and child results from each trial into separate mixed-effect models with condition (polysemy vs. ambiguity) as the fixed effect using RStudio and the lmerTest library and fit the maximal random terms that convergence would allow (Barr et al., 2014; Kuznetsova, Brockhoff & Christensen, 2017; RStudio Team, 2017). For all models, we preregistered the maximal random effect structure that convergence would allow.

Adults Figure 5 shows performance in the two conditions for adults immediately after exposure in the top right panel. We analyzed results with a multilevel model which included condition as the fixed effect with random intercepts for subjects and items and a random slope and intercept for order. This model revealed significantly better accuracy in the polysemy condition in comparison to the ambiguity condition ($\beta = 0.31$, $t = 4.14$, $p = 0.002$).

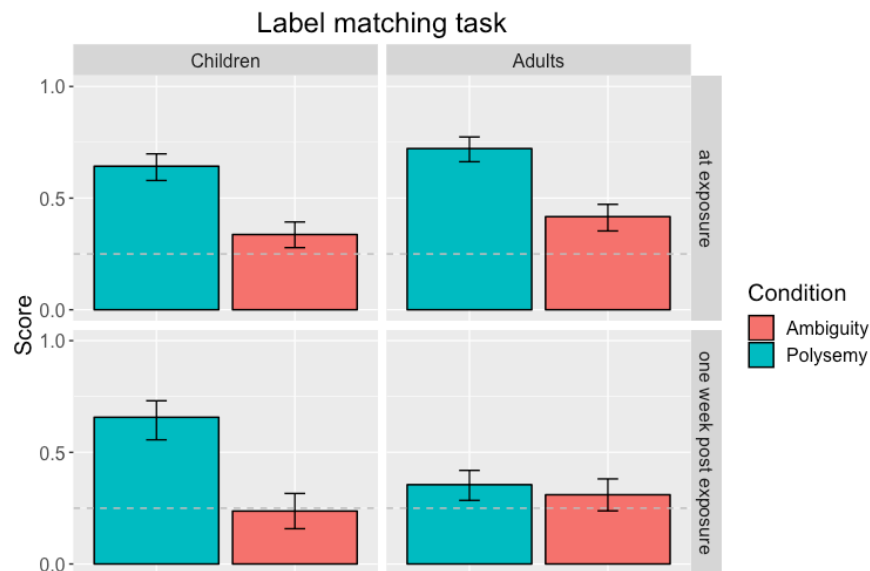


Figure 5: Average accuracy in the Label Matching task for the ambiguity and polysemy groups. Child performance (4½ -7 years old) at exposure in upper left panel ($n = 42$ per condition), adults' performance at exposure in upper right panel ($n = 42$ per condition); 4½ -7-year-old children's performance one week later in bottom left panel ($n = 18$ in polysemy, $n = 19$ in ambiguity); adults' performance one week later in bottom right panel ($n = 31$ in polysemy, $n = 28$ in ambiguity). Error bars represent bootstrapped 95% Confidence Intervals.

Children Figure 5 shows children's performance at the time of exposure in the upper left panel. As with adults, the preregistered multilevel model included condition as the fixed effect and random slopes and intercepts for order and subjects, and random intercepts for item. The model revealed significantly better accuracy in the Polysemy condition compared to the Ambiguity condition ($\beta = 0.31$, $t = 5.10$, $p = 0.002$).

Longitudinal results

The longitudinal analyses were not preregistered because we could not anticipate available sample sizes. The bottom panel in Figure 5 shows participants' accuracy averaged in each condition at the second timepoint. The children who returned for the retest did not differ in age across conditions, (polysemy: $M = 6.09$, $SD = 0.38$; ambiguity: $M = 6.17$, $SD = 0.36$; $\beta = -0.07$, $t = -0.61$, $p = 0.54$). Re-tested children returned

CHILDREN USE RELATIONSHIPS ACROSS MEANINGS IN WORD LEARNING

after an average of 7.19 days (range = 7-8 days, $SD = 0.4$ days), and this length was not significantly different across the conditions (polysemy: $M = 7.17$ days, $SD = 0.38$; ambiguity: $M = 7.21$ days, $SD = 0.42$; $\beta = -0.04$, $t = -0.33$, $p = 0.74$). Adults were re-recruited using Turk Prime (Litman et al., 2016) exactly 7 days later.

Children. To first determine whether the returning child participants were representative of the full child sample, we entered first timepoint results into a multilevel model to compare performance at exposure between the whole group with the returning subgroup, using Group, condition, and their interaction as fixed effects, and random intercepts and slopes for order, item, and subjects. This revealed no main effect of Group (full group) ($\beta = -0.03$, $t = -0.45$, $p = 0.66$) and no interaction of condition (polysemy) and group ($\beta = -0.02315$, $t = -0.30$, $p = 0.76$), but again revealed a main effect of the polysemy condition ($\beta = 0.30$, $t = 4.2$, $p = 0.001$).

The maximal converging model at the second timepoint included random intercepts and slopes for participants, items, and order, as well as condition as the fixed effect. Children again performed significantly better in the polysemy condition at the second time point ($\beta = 0.42$, $t = 6.63$, $p = 2.08e-08$). To determine if there were changes in accuracy across timepoints, we analyzed just those children who participated at both timepoints, using a model with timepoint, condition, and their interaction as the fixed effects and subjects, order, and items with random intercepts and slopes. This revealed a main effect of condition ($\beta = 0.33$, $t = 3.987$, $p = 0.007$), and a main effect of timepoint (T2) ($\beta = -0.15$, $t = -2.38$, $p = 0.02$) but no significant interaction of condition (polysemy) and timepoint (T2) ($\beta = 0.09$, $t = 0.77$, $p = 0.47$).

Adults. To determine whether the returning adult participants were representative of the full adult sample, we entered group (full vs. returning subgroup), condition, and their interaction as main effects (and random intercepts and slopes for order, item, and subjects). This revealed no main effect of group (returning subgroup) ($\beta = 0.01$, $t = 0.24$, $p = 0.81$) and no interaction of condition (polysemy) and group ($\beta = 0.00$, $t = 0.00$, $p = 0.99$), but again revealed a main effect of the polysemy condition ($\beta = 0.29$, $t = 3.379$, $p = 0.003$).

We also fit the maximal converging multilevel model to predict adult performance using condition as fixed effect, with random intercepts and slopes for subjects, items, and order. Unlike both the adult data at the first timepoint, and the child data at the first and second timepoints, adults' polysemy advantage at the second timepoint did not approach significance ($\beta = 0.05$, $t = 0.6$, $p = 0.56$).

Finally, we analyzed across timepoints, including just those participants who participated at both timepoints ($n = 59$), again using a maximal converging model with timepoint, condition, and their interaction as fixed effects (random intercepts and slopes for subjects, item, and order). This analysis revealed a significant main effect of the polysemy condition ($\beta = 0.31$, $t = 4.21$, $p = 0.0002$), but only a marginal effect for timepoint ($\beta = -0.14$, $t = -1.91$, $p = 0.06$), and a significant interaction between time and polysemy ($\beta = -0.26$, $t = -2.76$, $p = 0.007$). The interaction may suggest that though performance was lower overall for adults at the second timepoint, adults forgot significantly more in the polysemy condition. Alternatively, it may be due to the floor effect for ambiguity at the second timepoint: i.e. for performance in the ambiguity condition to decay at the same rate as the polysemy condition, it would have had to be reliably below chance at timepoint two (see Comparisons to chance).

Age comparisons. Our preregistered analyses reveal that the polysemy over ambiguity advantage was present both in children and in adults at the first timepoint, and an exploratory comparison between age groups allowed us to investigate whether it was significantly different at either age. We first compared performance at time of exposure, fitting a model with age group, condition, and their interaction as fixed effects and once again, the maximal converging structure including a random intercept for subjects and random slopes and intercepts for items and order. The model revealed a significant effect of condition (polysemy) ($\beta = 0.30$, $t = 5.62$, $p = 2.12e-07$), but no significant effect for age group (children) ($\beta = -0.08$, $t = -1.52$, $p = 0.13$), and no significant interaction ($\beta = 0.004$, $t = 0.05$, $p = 0.96$), suggesting that both adults and children exhibit the polysemy advantage to roughly the same degree at the first timepoint. We

also compared group performance across conditions at the second timepoint, and found a significant interaction between condition and age group ($\beta = 0.38, t = 3.78, p = 0.001$), but no main effect for condition ($\beta = 0.05, t = 0.56, p = 0.59$) or age group ($\beta = -0.07, t = -1.24, p = 0.23$). This suggests that, unlike timepoint 1, in which both adults and children showed a significant polysemy advantage, children at the second timepoint performed significantly better in the polysemy condition than adults, who showed little evidence of a polysemy advantage.

We also performed an exploratory analysis to determine if the polysemy advantage was present throughout the age range within the child population (4½ to 7 years). We calculated a median split on age (median = 6.01 years) and entered each groups' data into the model with condition as the fixed effect (and random intercepts for subjects and items and a random slope and intercept for order). This model revealed a significant effect of condition (polysemy) ($\beta = 0.25, t = 3.11, p = 0.01$) in the younger half of participants. Using the same model used to predict performance in the older half, the model again revealed a significant effect of the polysemy condition ($\beta = 0.38, t = 5.89, p = 1.51e-05$), and a model predicting performance across the two groups using condition and age group found no significant interaction ($\beta = -0.13262, t = -1.405, p = 0.166745$), suggesting that the polysemy over ambiguity advantage is present throughout our child sample to a similar degree (random terms included an intercept for subjects and slopes and intercepts for items and block number).

Comparisons with chance

Final exploratory analyses include comparisons to chance. Interestingly, bootstrapped 95% confidence intervals show that children were above chance (0.25) in both conditions at timepoint 1 (ambiguity: $M = 0.337$, 95% C.I. = 0.278 - 0.393; polysemy: $M = 0.643$, 95% C.I. = 0.579 - 0.698), suggesting that children were to some extent also able to learn ambiguous words, consistent with past findings showing knowledge of ambiguous words at young ages (Backscheider & Gelman, 1995). Yet ambiguity was challenging for children, given how close to chance their performance was in the ambiguity condition. Further, their retention of ambiguous words overlapped with chance after the week delay (ambiguity: $M = 0.237$, 95% CI [0.158 - 0.316]; polysemy: $M = 0.657$, 95% CI [0.556 - 0.731]), indicating that they required more input than our exposure provided in order to retain multiple, unrelated meanings. In adults, we found performance parallel to children at the first timepoint (ambiguity: $M = 0.417$, 95% CI = [0.353 - 0.476], polysemy: $M = 0.722$, 95% C.I. [0.659 - 0.774]). However, their performance deteriorated after 7 days, resulting in a significant reduction in the polysemy condition across the timepoints, and worse performance overall at the second timepoint, where no polysemy over ambiguity learning advantage was evident. However, in a non-preregistered comparison to chance, we did find consistent above-chance performance in polysemy at the second timepoint for adults: $M = 0.355$, 95% CI [0.285 - 0.419], while their performance was not reliably above chance for ambiguity, as CIs overlapped with 0.25: $M = 0.310$, 95% CI [0.238 - 0.375].

Discussion. As predicted, both children and adults showed a significant polysemy over ambiguity advantage immediately following exposure, demonstrating that words with related meanings are easier to learn than words with unrelated meanings. The effect remained strong in children after a 1-week delay without re-exposure. Adult performance decayed after the week delay, and no polysemy over ambiguity advantage was evident (although performance only in polysemy condition remained above chance). There are two possible explanations for the difference between adults and children after delay without assuming that word learning is a critically different process in adults and children at this age. First, children were retested in person with the same experimenter and in the same location, while adults were re-recruited online. Because of this, children had additional cues to retrieve their past experience with the task, and thus the exposure; we know contextual cues such as physical surroundings can have significant effects on lexical memory (Baddeley, Eysenck & Anderson, 2009; Egstrom, et al., 1972; Godden & Baddeley, 1975). Because adults were recruited online, it was impossible to ensure they were in a similar context at the second exposure. Another, nonexclusive possibility is that adults' more densely-packed lexical and phonological networks impeded their ability to retrieve the novel word labels they had learned a week before. Recall that the task required participants to determine which of four meanings corresponded to the

particular label they were prompted with. Accurate performance was possible by a process of elimination: if two novel meanings were recalled with different labels, the witnessed label must belong to one of the remaining options. This strategy would require specific memory of word labels for the other items each trial, however adults in particular have been shown to struggle in name- and word-retrieval tasks (Dell & Gordon, 2003; Ramsar et al., 2014). If their denser lexicons made the task more difficult, then a different task that primarily depends on memory of meanings rather than memory of label-to-meaning mappings may reveal lasting evidence for stronger performance on polysemous word meanings. In Experiment 2, we provided participants with a task that evaluated this possibility.

Experiment 2: Sense selection

Experiment 1 demonstrated that both children and adults have a stronger memory for polysemous word-to-meaning mappings when asked to discriminate them from differently-labeled distractors immediately after exposure, and for children, the polysemy over ambiguity advantage remained after a week delay. One concern about the task in Experiment 1 is that participants were only asked to assign a single meaning to each label on each trial, so that they may have relied on memory of one or two novel words to perform well. Another issue is that Experiment 1 required participants to construe differently-labeled targets as dissimilar, but it did not directly test whether participants were able to group together commonly-labeled polysemous targets as sharing the same label. Therefore, in Experiment 2, we ask participants to discriminate the targets from a set of foils that had been witnessed during the same exposure video.

Specifically, on each trial participants were presented with three items from a single meaning set along with five foils, and were prompted with a label. Because this task presents participants with all three targets simultaneously, we report a separate norming study with a different group of participants in order to rule out two possible alternative explanations for participants' success in learning multiple polysemous meanings. First, it is possible that *a priori* visual similarity between the polysemous items is sufficient to explain participants' successful identification of targets at test. It is also possible that the higher relative exposure to the prototype in the videos may have facilitated learning of the prototype in each set of meanings, but that participants simply used this representation to infer the other two targets on-the-fly, without learning the two additional meanings. In our norming task, we rule out both of these possibilities by giving participants the same exposure as in the main task without label information, and then asked them to select the three items which they think are likely to share a label. Since participants only saw items from differently-labeled sets in the prior experiment, examining these explanations was not possible in Experiment 1.

The task of Experiment 2 did not require memory for specific labels. It instead probed whether participants' memory for which items had been labeled to test whether identification of the 3 target meanings of each word benefited from similarities between exemplars in the polysemy condition. This task then also allows us to address the possibility that adults' low performance at the second timepoint in Experiment 1 was due to the challenge of retaining new lexical items. If this led to a decay over time in adults' performance, we should see more robust retention at follow-up in Experiment 2.

Method

Participants The same 84 children and 84 adults from Experiment 1 were tested. As in Experiment 1, the same 37 children and 59 adults were tested again after a week, with no intervening exposure (again, children: $N = 18$ in polysemy, $N = 19$ in ambiguity; adults: $N = 31$ in polysemy, $N = 28$ in ambiguity).

In a **norming study**, a separate group of 84 adult participants were recruited from Amazon's Mechanical Turk (preregistered). This group witnessed the identical visual exposure but without any verbal labels, ensuring that they would see the prototypical objects twice as in the main task, and were then presented with the identical test items used in the main analyses (see Test) with the following instructions: "Three of these objects share a label. Your job is to guess which ones." (see SI for additional norming).

Stimuli & Exposure Adults and children took part in Experiment 2 immediately after Experiment 1 without any additional exposure. This order was used, rather than the reverse or a counterbalanced order, to minimize possible learning during the first task that would be relevant to the second task. Recall that Experiment 1 provided an indication that the targets on each trial had been labelled differently from one another since participants were asked to select *the* kaisee/nona/gazer/veebo from among a set of objects that had been witnessed with other labels. But it did not provide evidence of which three meanings belonged together with a label, which was required for accuracy in Experiment 2. Experiment 2, on the other hand, provided information that could be used in Experiment 1, since in Experiment 2, only targets which matched the label in the prompt were displayed to participants, along with a set of foils that had not been labeled. Remaining concerns about Experiment 2 being unduly influenced by Experiment 1 were addressed by a separate sample of 92 adult participants who were tested only on Exp. 2. Those results confirmed the same pattern of data reported here (see SI for results). It remains possible that the testing at timepoint 2 was affected by the testing at timepoint 1, but this was unavoidable, and importantly, whatever advantage was gleaned at timepoint 1 for the polysemy condition was also available for the ambiguity condition.

Test Participants performed a Sense Selection task in which they were shown three meanings of a word along with 5 of the foils from the corresponding exposure video, and were asked to “Pick three *kaisees*,” (the 4 nonce labels counterbalanced across conditions, see Figure 6). The same subset of children described in Experiment 1 were re-tested again after the week-long delay, without additional exposure; they were simply shown the same test trials in the same order as they had been a week earlier.



Figure 6: Polysemy trial (1 of 4). Participants were prompted with a label and asked to select 3 target meanings.

Results

Adults Figure 7 shows performance in the two conditions for both age groups, with chance at 1.125³. The preregistered maximal converging multilevel model included condition as the fixed effect, and subject,

³ Each trial offered participants to earn a score of 0, 1, 2, or 3 for selecting up to 3 correct targets out of 8, so, chance performance is calculated to be 1.125 out of 3, on the basis of the following:

$$\sum_{i=1}^3 i \cdot p(i) \text{ where } p(i) = \frac{C(5, (3-i)) \cdot C(3, i)}{C(8, 3)}$$

CHILDREN USE RELATIONSHIPS ACROSS MEANINGS IN WORD LEARNING

order, and items as random intercepts with random slopes for item and subject. It revealed significantly better accuracy in the polysemy condition in comparison to the ambiguity condition ($\beta = 0.65$, $t = 4.26$, $p = 0.0005$).

Children The same model was preregistered for children, and the same maximal random structure converged for the child data. Children also performed significantly more accurately in the polysemy condition compared to the ambiguity condition ($\beta = 0.84$, $t = 8.76$, $p = 2.64\text{e-}13$).

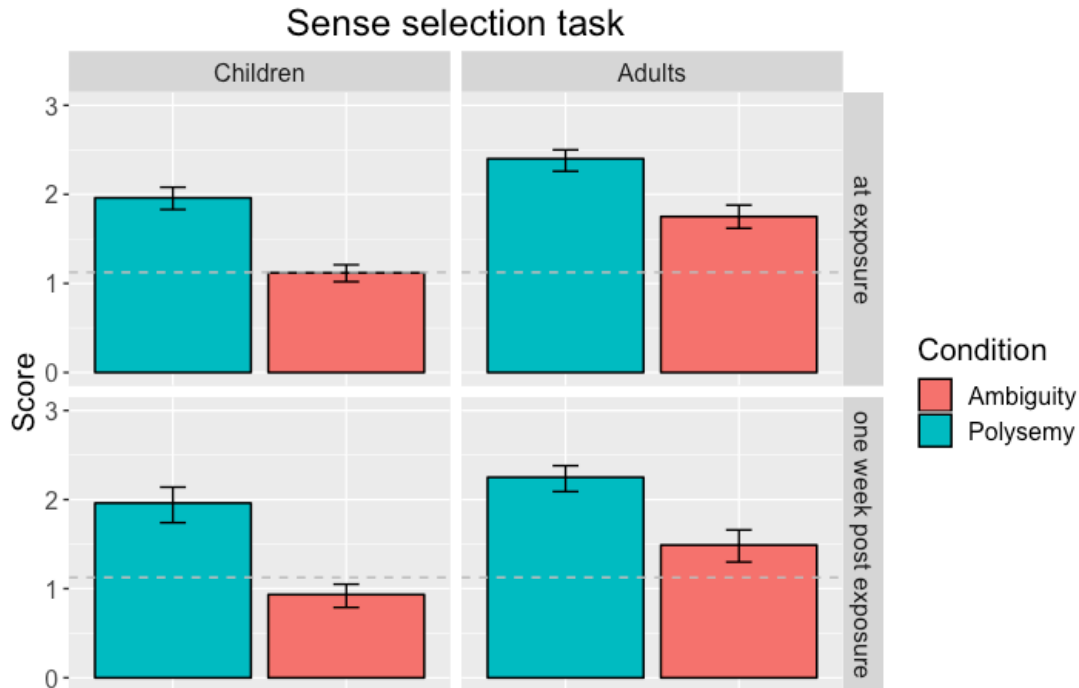


Figure 7: Average accuracy in the Sense selection task for the ambiguity and polysemy groups. Child performance (4½ -7 years old) at exposure in upper left panel ($n = 42$ per condition), adults' performance at exposure in upper right panel ($n = 42$ per condition); 4½ -7-year-old children's performance one week later in bottom left panel ($n = 18$ in polysemy, $n = 19$ in ambiguity); adults' performance one week later in bottom right panel ($n = 31$ in polysemy, $n = 28$ in ambiguity). Error bars represent bootstrapped 95% Confidence Intervals.

Norming. We calculated “item predictability” scores for each novel word's set of three meanings in each condition by averaging participants' performance in the Norming (no exposure) study on each trial, and used these scores as a measure of predictability on the basis of similarity and on the basis of inference from higher frequency of the prototype (sample size and analysis preregistered). Item predictability was included as fixed effects with random slopes and intercepts for subjects and order in both populations. The models for both adults and children still revealed a significant advantage for adults and children in the polysemy condition, even in the presence of item predictability (adults: $\beta = 0.59$, $t = 2.82$, $p = 0.02$), children: $\beta = 0.78$, $t = 4.83$, $p = 0.0006$). In fact, visual predictability was not a significant predictor for either group: adults ($\beta = 0.08$, $t = 0.18$, $p = 0.87$), or children ($\beta = 0.15$, $t = 0.56$, $p = 0.59$). Thus, our prediction was confirmed that the polysemy over ambiguity advantage reflects a process of word learning and is not explained by the predictability of the items.

Longitudinal results. Once again, the exploratory analyses for timepoint 2 included timepoint and its interaction with condition as a fixed effect. The bottom panel of Figure 7 shows their accuracy averaged in each condition at the second timepoint in each group.

Children. The bottom left quadrant of Figure 7 shows children's scores on the second task one week after exposure. The model predicting performance of children at follow-up revealed a significant effect of the polysemy condition ($\beta = 1.02, t = 7.88, p = 1.63\text{e-}10$) including random intercepts and slopes for participants and order, and random intercepts for item (this analysis required model comparison using between two models with equally complex random structure). To determine if there were changes in accuracy across timepoints, we again analyzed just those children who participated at both timepoints using a model with timepoint, condition, and their interaction as the fixed effects, and random intercepts and slopes for subjects, items, and order. The polysemy over ambiguity advantage held; in fact, there was no evidence of memory decay after the week-long delay. That is, there was a main effect of condition (polysemy) ($\beta = 0.96, t = 7.80, p = 2.05\text{e-}09$), but no main effect of timepoint (T2) ($\beta = -0.17, t = -1.44, p = 0.16$) and no significant interaction of condition (polysemy) and timepoint (T2) ($\beta = 0.06, t = 0.35, p = 0.73$).

Again, to ensure that these participants were representative of our main sample on this task as well as the preceding task, we entered results into a multilevel model to compare performance on this task at the first timepoint between the whole group and the subgroup which returned for the follow up, using group, condition, and their interaction as main effects with maximal random structure (random intercepts for subjects random slopes and intercepts for item and order). Reassuringly, this revealed no main effect of group (full group) ($\beta = -0.01, t = -0.11, p = 0.91$) and no interaction of condition and group ($\beta = -0.07, t = -0.48, p = 0.64$), but again revealed a main effect of the polysemy condition ($\beta = 0.84, t = 47, p = 1.91\text{e-}10$).

Adults. The bottom right quadrant of Figure 7 shows adults' scores on the second task one week after exposure. Consistent with both the adult data at the first timepoint and the child data at the first and second timepoints, the model (including random intercepts and slopes for subjects, items, and order) found a significant polysemy advantage at follow up ($\beta = 0.76, t = 3.17, p = 0.01$).

We also analyzed across timepoints, taking just participants from both timepoints ($n = 59$), and again using a model with timepoint, condition, and their interaction as fixed effects and maximal converging random structure which included random intercepts and slopes for subject, order, and item. This analysis revealed a significant main effect of the polysemy condition ($\beta = 0.77, t = 3.91, p = 0.0009$), but no effect of timepoint ($\beta = -0.24, t = -1.24, p = 0.23$) or interaction between timepoint and polysemy ($\beta = -0.01, t = -0.04, p = 0.97$), suggesting that in this task, the polysemy advantage was present to the same extent across both timepoints.

To confirm that the returning adult participants were representative of the full adult sample in this task as well as the task reported in the first experiment, we entered results into a multilevel model to compare performance at exposure between the whole group with the returning subgroup, using group, condition, and their interaction as main effects (random intercepts and slopes for order, item, and subjects). This revealed no main effect of group (returning subgroup) ($\beta = -0.004, t = -0.05, p = 0.96$) and no interaction of condition (polysemy) and group ($\beta = 0.03, t = 0.24, p = 0.81$), but a main effect of the polysemy condition ($\beta = 0.65, t = 4.04, p = 0.0004$).

Age comparisons

To determine whether there was an effect of age on performance at the first timepoint on the Sense Selection task, we fit a model with age group (adults vs. kids) and condition as fixed effects (random slopes and intercepts for subjects, item, and order). This exploratory model revealed a significant effect of condition (polysemy) ($\beta = 0.65, t = 4.67, p = 0.0004$) and age group (children) ($\beta = -0.63, t = -4.52, p = 0.002141$), and no significant interaction of condition (polysemy) and age group (children) ($\beta = 0.19, t = 1.05, p = 0.31$). Adults' better performance suggests that unlike what was found for the Label Matching task of Experiment 1, the benefit of polysemy over ambiguity in learner's ability to discriminate targets from foils in the Sense Selection task may continue to develop with age. At the second timepoint, we also compared children and adults using a model with condition, age group, and their interaction as fixed effects (random effects and intercepts for subjects, items, and order), and found a significant effect of polysemy ($\beta = 0.76, t = 4.37, p = 0.001$) and of age group (children) ($\beta = -0.56, t = -3.07, p = 0.006$), but no significant interaction

($\beta = 0.27$, $t = 1.07$, $p = 0.30$). So, while the polysemy advantage was present across both groups, and children didn't do as well as adults overall, there was no evidence that the polysemy advantage was attenuated in children in particular.

In order to investigate possible age effects further, in another non-preregistered analysis, we entered child data with a median split by age (median = 6.01 years) to determine if the polysemy advantage was present throughout the age range in the sample. First, we entered each half of the data into a multilevel model with condition as the fixed effect (random intercepts for subjects and items and a random slope and intercept for order). This model once again revealed a significant effect of condition (polysemy) ($\beta = 0.73$, $t = 2.95$, $p = 0.01$) in the younger half of participants, and the same model found that the polysemy condition again predicted better performance ($\beta = 0.97$, $t = 6.68$, $p = 1.38\text{e-}06$). A model predicting performance across the two groups using condition and age group (random intercepts and slopes for items and order and random intercepts for subjects) found no significant interaction of age group within the child sample (older half) and condition (polysemy) ($\beta = -0.24$, $t = -0.86$, $p = 0.40$), again finding a main effect of condition (polysemy) ($\beta = 0.97$, $t = 6.40$, $p = 1.23\text{e-}07$), and no main effect of age group ($\beta = -0.14$, $t = -0.72$, $p = 0.48$).

Comparisons to chance. Bootstrapped 95% confidence intervals were used to compare to chance performance of selecting between 0 and 3 correct targets out of a possible 8 items (see Results, footnote 3). These exploratory analyses showed that children were reliably above chance in the polysemy condition at timepoint 1 (polysemy: $M = 1.96$, 95% C.I. [1.83 - 2.08], but not in the ambiguity condition as the bootstrapped 95% CI overlap with chance [1.03 - 1.21]. Furthermore, while children's performance after a week delay remained above chance in polysemy ($M = 1.96$), 95% C.I. [1.74 - 2.15], performance in ambiguity fell reliably below ($M = 0.93$), 95% C.I. [0.79 - 1.07]. Such low performance in children at follow-up in Experiment 2 may be considered surprising, as adults seemed to do comparatively better in Experiment 2 at follow-up as compared to Experiment 1. At timepoint 1 in Experiment 2, adults were also reliably above chance (ambiguity: $M = 1.75$, 95% C.I. [1.62 - 1.88]; polysemy: $M = 2.40$, 95% C.I. [2.26 - 2.50]), but unlike in Experiment 1, adult performance was reliably above chance in both conditions by the second timepoint (ambiguity: $M = 1.49$, 95% CI = 1.30 - 1.66, polysemy: $M = 2.25$, 95% CI = 2.09 - 2.38).

Discussion. Children and adults both showed a significant polysemy over ambiguity advantage and the younger half of children (4 ½ - 6) performed as well and showed as strong of a polysemy advantage as the older children. Both children and adults retained the polysemy advantage after a 1-week delay without re-exposure. Norming results confirm that this robust polysemy advantage cannot be attributed to visual similarity among items, as including visual similarity in the model had no significant effect for either children or adults. Adults outperformed children at both timepoints, unlike in Exp. 1, which is consistent with the idea that adults' poorer performance after the delay in Exp. 1 was likely due to the challenge of retrieving new (low-frequency) words from their denser lexicons.

General Discussion

The present experiments investigated children and adults' ability to learn 4 novel words which were assigned three distinct meanings apiece. We manipulated whether the three meanings of each word were related to one another (polysemy condition) or not (ambiguity condition). The meanings of each word in the polysemy condition were designed so that they shared no single distinguishing feature but instead varied along two different dimensions, reducing the possibility that participants could rely on a simple definition or a single extension rule in order to distinguish the three target meanings for each word from foils. Recall that novel labels and novel meanings were used in order to mitigate the effects of previous experience. No feedback was provided, and available participants were tested immediately after exposure and again after a one-week delay.

In each of two tasks, we consistently found a **polysemy over ambiguity advantage** in learning the novel words immediately after minimal exposure. That is, children and adults who were exposed to 4 novel

polysemous words outperformed those who were exposed to 4 ambiguous words in their ability to map novel labels to the appropriate target meaning, distinguishing that meaning from otherwise-labeled distractors (Experiment 1). Participants who had witnessed four polysemous words also performed better at correctly identifying and distinguishing multiple target meanings for each novel word from foils (Experiment 2). Since our interest is in vocabulary learning, which is consolidated in long term memory, we retested participants after a week-long delay without providing additional exposure.

After confirming that the participants who took part in the testing after a week did not differ significantly in age or performance from the full group, we performed the identical analyses for the data at the second time point. Remarkably, children's performance in the polysemy condition after a week remained well above their performance in the ambiguity condition in the label mapping task (Exp. 1) and in fact showed no decay after a week in the sense selection task (Exp. 2). Adults' performance fell to chance in the label mapping task after the weeklong delay, possibly because adults' larger lexicons made the retrieval of target words from long-term memory more demanding (Dell & Gordon, 2003; Ramscar et al., 2014). In the second task, which required participants to discriminate between commonly-grouped meanings and foils and did not require lexical access, adults showed the same polysemy > ambiguity advantage evident in children (Experiment 2).

The present findings are consistent with work on non-linguistic category learning insofar as meanings that shared more overlapping features were easier to categorize together than those that shared fewer distinguishing features. This suggests, perhaps not surprisingly, that word-learning may well be a special type of category learning, as long as "category" is not assumed to require a single vague or summary representation, but is instead understood to allow a range of subtypes with their own unique properties. Word learning is special insofar as learners need to use the labels as a cue to multidimensional category membership, and must learn the arbitrary mappings between exemplars and labels without receiving explicit instruction nor corrective feedback. The current findings are also consistent with prior work which suggests that ambiguous meanings of a word compete with one another during comprehension but polysemous meanings do not. If polysemous meanings are represented as partially overlapping, distributed representations (Rodd et al., 2012), polysemous meanings of a word may reinforce one another rather than compete with one another as is assumed by current models of word learning.

Particularly because the polysemy we investigate is irregular, we do not assume that children necessarily *expect* or predict that the word label will extend in any particular way. The current exposure provided positive evidence that the distinct meanings were labelled by the same word, which appears to provide an advantage during integration and long-term consolidation. Earlier work had emphasized that unrelated meanings of a word label were easier to learn when each meaning was used as a distinct syntactic category or was from a wholly different semantic field (such as labeling a novel animal *a glass*). For example, consistent with the current findings, Casenhiser (2005) found that children were reluctant to assign a familiar noun like *dog*, the meaning, "monster." At the same time, children were willing to reuse a label from a different grammatical category (*a did* could be a monster) (see also Dautriche, Fibla, Fievet, & Christophe, 2018). That is, ambiguous meanings are less difficult to learn when they apply in easily distinguishable contexts rather than in contexts that have the potential to overlap. In the current work, all novel words referred to artifacts and were used in nominal contexts, so the contexts provided were underdetermined and overlapping.

One reviewer suggested that participants in the polysemy condition may not have learned multiple meanings but instead only learned the prototypical meaning and then generalized that label to the other two meanings on the fly during the tasks. We agree with the suggestion that if participants somehow only learned the prototype, they could be expected to perform better in the polysemy condition than in the homonymy condition, since by design, the additional polysemous meanings were related to the prototype while the homonymous meanings were not. Yet an explanation of why and how learners might only learn the prototype during exposure is required to make this route to better performance in the polysemy condition feasible. In natural languages, learning only one prototype per form is not sufficient, since extensions of the same prototypical meaning vary across different languages (see Introduction). Moreover, the results of the norming study show that exposing learners to two instances of the prototype instead of one does not

predict performance on the task in either children or adults. We suggest that implicit recognition and comparison of the 3 labelled meanings results in their overlapping features being strengthened. That is, implicit recognition and comparison of the 3 labelled meanings would allow a prototype to emerge, but critically, this perspective involves the recognition and comparison of *multiple* meanings.

If polysemy were rare in natural language, the question of how words with multiple meanings are learned might be reasonably put aside until the learning of words with single senses were better understood. However, polysemy is ubiquitous, particularly in the case of highly frequent and early learned words, and it commonly includes distinct meanings from the same grammatical category and semantic field (as is the case for the various senses of *cap* and other examples in the introduction). The relative lack of previous work on children's learning of conventional, distinct but related meanings of words is surprising insofar as there has been a great deal of work emphasizing a finding that is the inverse of the multiple-meaning phenomenon reported here. In particular, learners generally disprefer assigning a second label to a concept (e.g., Clark, 1987; Markman & Wachtel, 1988; Merriman, Bowman & MacWhinney, 1989). That is, learning synonymous labels is challenging and true synonymy is rare in languages. Near synonyms are almost always distinguishable in one way or another: e.g., by dialect (*pop* vs. *soda*), register (*buy* vs. *purchase*), or attitude (*thrifty* vs. *stingy*). Rather than assigning multiple labels to the same concept (synonymy), both polysemy and ambiguity involve assigning multiple concepts to the same label.

Production-based accounts have offered an explanation as to why polysemy *and* ambiguity are common across the world's languages by emphasizing the advantage of reusing words that are easy to access and produce (Harmon & Kapatsinski, 2017; Piantadosi, Tily & Gibson, 2012). But it is important to distinguish polysemy from ambiguity, even though the distinction is undoubtedly gradient, because polysemy is so much more common than ambiguity (Dautriche, 2015; Navigli & Ponzetto 2012; Rodd et al 2002). The current work provides the first evidence that we know of demonstrating that conventional ("irregular") polysemous senses are easier to comprehend and retain than ambiguous senses.

We found little evidence of a developmental shift in the ability to learn novel polysemous words. Older children were not significantly more accurate in either task than younger children. Children in the label matching task performed as well as adults initially, and outperformed adults after the delay. Only in Experiment 2, which required participants to recall which meaning(s) were related, given a label, did adults significantly outperform children at both timepoints. Future work should test younger children to determine whether a polysemy over ambiguity advantage is evident in children younger than 4½.

Future work is also needed to determine exactly *why* polysemy is markedly easier to learn than ambiguity, but we can speculate about several possible (nonexclusive) factors. Witnessing a word in a new context is likely to direct learners' attention to features that were associated with the word's meaning(s) in previous contexts. If some of those features are shared by a possible new meaning, the label would then provide an attentional bias toward the new referent. As each exemplar is to some extent unique, the same process may be operative at a lower level, for instance, even when we recognize a new dog as a *dog*, or a new door as a *door*. As already mentioned, there may be less of a distinction between the relationships among multiple exemplars required for categorization tasks and the relationships among distinct meanings of polysemous words (see also Tuggy, 1993; Geerearts, 1993). If learners' attention is directed to features that have been associated with the word in other contexts, we might expect this to occur both during word learning and during comprehension. Given that the present experiments found a surprisingly robust memory for polysemous meanings after a week-long delay, it may also be possible that distinct but related meanings reinforce one another, or reduce decay in memory, with learners either representing related senses as partially overlapping in memory or distinctly but with the relationship itself being encoded.

Implications for models of human word learning

We know of no current models of human word learning that predict differences in children's learning of polysemy vs. ambiguity. The present work suggests that it is essential to expand our models of human word learning so that they not only include the rich internal structure of word meanings, but also reflect relationships among multiple meanings of words. The majority of work on vocabulary learning has focused on how children learn a single meaning for a given novel label rather than how children learn multiple

meanings of a word, as the task of assigning even a single meaning is recognized to be difficult (Quine, 1960; Gleitman, 1990). Current models of how children learn words have explicitly ignored the issues raised by words with multiple meanings. According to one recent model, “Propose but Verify” (PbV), the learner only tracks a single hypothesized word sense at any given point in time (Medina, Snedeker, Trueswell & Gleitman, 2011; Trueswell, Medina, Hafri, & Gleitman, 2013; Aravind et al., 2018; Woodard et al., 2016). This model aims to address the problem of referential ambiguity, in which learners must identify which referent in a scene is intended by a new word, but it inadvertently legislates both polysemy and ambiguity out of learners’ grasp. PbV proposes that a learner would hypothesize a meaning for the new word, but would then jettison that hypothesis and begin anew if the hypothesized meaning was not appropriate in next context. This problem with PbV has been recognized and a revised model, Pursuit, allows more than one hypothesized meaning to exist simultaneously (Stevens, Gleitman, Trueswell & Yang, 2016). However, Pursuit represents each possible sense as atomic, with no internal structure (Stevens et al., 2016). Thus, it is unable to capture the distinction between learning ambiguity and polysemy, as shared features or relationships between meanings are not represented. Most cross-situational word learning models, which rely on experience across many learning instances, also represent meaning atomically, with correct meanings emerging from competition between consistent and inconsistent referents (e.g., Yu & Smith, 2007). In particular, each possible meaning of a word is equally in competition with all other potential senses. In the case of polysemy, this counterintuitively predicts that evidence for the “baseball cap” meaning of *cap* would be evidence **against** the “bottle cap” meaning of *cap*.

These models of word learning surely involve simplifications that were not intended to capture the true nature of word meanings, but there has been precious little work investigating how novel words with multiple conventionalized senses are initially learned, particularly by children. The current results make clear that theories of human word learning should be developed in two main ways. First, competition between meanings cannot undermine the possibility of multiple, distinct word meanings for a single form. While competition can be recruited during online disambiguation, when a unique word or meaning must be selected from among alternatives (Hinton & Shallice, 1991; Joordens & Besner, 1994; Kawamoto, Farrar, & Kello, 1994; Luce & Pisoni, 1998; McClelland & Elman, 1986; Rodd, Gaskell, & Marslen-Wilson, 2004; McMurray, Horst & Samuelson, 2012; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), the robust learning evident in the present polysemy conditions suggests that encountering a related meaning of a word need not weaken—let alone eliminate—a distinct meaning of the same word in memory. We propose that multiple interpretations of a single word do not require competition between the meanings. Instead, we suggest new interpretations of a given word should be *added* to long term memory, with incorrect interpretations becoming less accessible over time from lack of reinforcement and from exposure to the meaning being labeled by the correct, conventional form (e.g., Goldberg, 2019).

Second, our results suggest the importance of distributed representations in learning. We propose that, rather than encoding or representing atomic and unanalyzed word meanings, that word learning requires representations with internal structure, without which it would be impossible to capture relatedness or similarity between attributes of a word’s meanings. For example, a representation such as *baseball cap* includes multiple attributes such as “covering” and “tightly fitting”, allowing it to share those attributes with a *bottle cap*, rendering the two meanings semantically related. Outside the field of word learning, distributed representations are common. For example, feature-based representations are widely used in models of word recognition (Hinton & Shallice, 1991; Joordens & Besner, 1994; Kawamoto, 1993; Kawamoto, Farrar, & Kello, 1994; Rodd, Gaskell, & Marslen-Wilson, 2004), although this work has tended to emphasize phonological and orthographic features as they are needed to recognize cases that require disambiguation. While models of word learning have occasionally included hand-coded semantic features, to date, they have not yet addressed the question of how polysemous or ambiguous words are learned (Fazly, Alishahi & Stevenson, 2010). It is possible that learning of unrelated meanings is more difficult due to interference during encoding or retrieval, rather than at the level of the meaning representation itself. However, we propose that the representations themselves are distributed, as some of our novel word meanings extend on a limited number of features, indicating that learners are guided by similarities between parts of meanings from their meaning representations, and not just the meanings as a whole.

CHILDREN USE RELATIONSHIPS ACROSS MEANINGS IN WORD LEARNING

Unifying both of these requirements, we propose that it is important to combine these two components (additive meanings, distributed representations) in order to account for the polysemy > ambiguity learning advantage found in the present data. Distributed representations provide the foundation for meanings to relate to one another, and the additive process allows additional meanings that overlap or relate to existing meanings to *strengthen* additional, related meanings, rather than compete with them.

Conclusion

Many words are associated with more than a single meaning, and quite often, meanings of a word are semantically related to one another. This is especially true in the case of frequent words, many of which are learned by young children. While certain meaning extensions can be predicted by very general rules and therefore may be created on-the-fly, other meaning extensions need to be learned on the basis of experience because they are idiosyncratic to individual words. We predicted that the existence of semantic relationships among meanings of a word (polysemy) would facilitate the learning of that word when compared with learning a word that was associated with the same number of unrelated senses (ambiguity). Experiments exposed adults and 4½-7-year-old children to 4 novel words, with 3 meanings each, during 2 minutes of exposure videos in which 20 foils were interspersed. The novel words were created using novel objects and novel labels in order to reduce interference from familiar labels or meanings. Two preregistered experiments confirmed our hypothesis, finding better accuracy for both age groups in the polysemy condition on both tasks: they were better able to identify which meaning was assigned a given label, a task that required distinguishing one of the label's meanings from meanings associated with other labels (Exp. 1); and they were more accurate at the selection of multiple meanings for each word from among also-witnessed but unlabeled foils (Exp. 2).

Remarkably, after a full week's delay, the subset of children who were available to be retested again showed the same polysemy over ambiguity advantage, despite having had no additional exposure. Moreover, the younger half of children performed as well as the older half, so that no clear developmental changes in the polysemy over ambiguity advantage were evident. Adult performance after the delay was mixed: in Exp. 1, only the polysemy condition remained above chance, but direct evidence of a polysemy over ambiguity advantage was absent, due to lower performance overall. On the other hand, adults did display a lasting polysemy over ambiguity advantage in Exp. 2, which did not require them to access and compare the meanings of multiple word labels. The recognition that polysemous meanings are ubiquitous and relatively easy to learn implies that the working assumption, commonly made in modeling work on human word-learning that words map uniquely onto a single meaning or that multiple potential meanings of a word are always in competition with one another, should be retired.

The current results raise new questions. Given that the current polysemy over ambiguity advantage was equally evident in younger and older children, we can ask, do children demonstrate the advantage at the very outset of vocabulary learning or do they need to learn to learn polysemous meanings? Since novel words can be learned to apply to related meanings after such limited and potentially confusing exposure, how is it that learners constrain meanings to apply to just the range that is specific to that word in the given language? The results also raise questions about populations that may have more difficulty recognizing relatedness among meanings. For example, individuals on the Autism spectrum have been claimed to hyper-focus on distinctions at the expense of recognizing relationships. Does this imply that individuals on the Autism spectrum lack the same polysemy over ambiguity advantage evident in neurotypicals (Floyd & Goldberg, *in prep*)? These and many other questions come to the fore once we recognize that words with multiple related meanings are more common and easier for typical learners than words with unrelated meanings.

References

- Apresjan, J. D. (1974). Regular polysemy. *Linguistics*, 12(142), 5-32.
- Aravind, A. & de Villiers, J., Pace, A., Valentine, H., Golinkoff, R., Hirsch-Pasek, K., Iglesias, A., Sweig Wilson, M. (2018). Fast mapping word meanings across trials: Young children forget all but their first guess. *Cognition*, 177(10).
- Armstrong, B. C., & Plaut, D. C. (2008). Settling dynamics in distributed networks explain task differences in semantic ambiguity effects: Computational and behavioral evidence. In B. C. Love, K. McRae, V. M. Sloutsky (Eds.), *Proceedings of the 30th Annual Conference of the Cognitive Science Society* (pp. 273–278). Austin, TX: Cognitive Science Society.
- Azuma, T., & Van Orden, G. C. (1997). Why safe is better than fast: The relatedness of a word's meanings affects lexical decision times. *Journal of Memory and Language*, 36, 484– 504. doi:10.1006/jmla.1997.2502
- Backscheider, A.G., Gelman, S.A. (1995) Children's understanding of homonyms. *Journal of Child Language*, 22, pp. 107–127
- Baddeley, A., M. W. Eysenck, and M. C. Anderson. (2009). *Memory*. Psychology Press, New York, NY, US.
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, 68(3), 10.1016/j.jml.2012.11.001.
- Barracuda1983. (2006). *File:Pipistrellus flight2.jpg*. (2014, November 24). Wikimedia Commons, the free media repository. Retrieved 13:14, June 11, 2018 from https://commons.wikimedia.org/w/index.php?title=File:Pipistrellus_flight2.jpg&oldid=140434049.
- Bowerman, M. (1996). Learning how to structure space for language: A crosslinguistic perspective. In P. Bloom, M. A. Peterson, L. Nadel, & M. F. Garrett (Eds.), *Language and space* (pp. 385-436). Cambridge, MA: MIT press.
- Bowerman, M., & Choi, S. (2001). Shaping meanings for language: universal and language-specific in the acquisition of semantic categories. In *Language acquisition and conceptual development* (pp. 475-511). Cambridge University Press.
- Bracken, J., Degani, T., Eddington, C. M., & Tokowicz, N. (2017). Translation semantic variability: How semantic relatedness affects learning of translation-ambiguous words. *Bilingualism: Language and Cognition*, 20, 783-794.
- Brinton, L., & Traugott, E. (2005). *Lexicalization and Language Change (Research Surveys in Linguistics)*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511615962
- Brocher, A., Koenig, J.P., Mauner, G. & Foraker, S. (2017) About sharing and commitment: the retrieval of biased and balanced irregular polysemes. *Language, Cognition and Neuroscience*, 1-24.
- Brown, R. W. (1957). Linguistic determinism and the part of speech. *The Journal of Abnormal and Social Psychology*, 55(1), 1.
- Casenhiser, D.M. (2005). Children's resistance to homonymy: An experimental study of pseudohomonyms. *Journal of Child Language*, 32(2), 319-343.
- Choi, S., & Bowerman, M. (1991). Learning to express motion events in English and Korean: The influence of language-specific lexicalization patterns. *Cognition*, 41, 83-121.
- Choi, S., McDonough, L., Bowerman, M., & Mandler, J. M. (1999). Early sensitivity to language-specific spatial categories in English and Korean. *Cognitive Development*, 14, 241-268. doi:10.1016/S0885-2014(99)00004-0.
- Crisco 1492. (2014). *File:Cap of "Aqua"-brand water bottle.jpg*. Wikimedia Commons, the free media repository. Retrieved 13:09, June 11, 2018 from https://commons.wikimedia.org/w/index.php?title=File:Cap_of_%22Aqua%22-brand_water_bottle.jpg&oldid=231564425.
- Clark, E. V. (1987). The principle of contrast: A constraint on language acquisition. *Mechanisms of language acquisition*, 1, 33.

- Cunningham, A.E. (2005). Vocabulary growth through independent reading and reading aloud to children. In E.H. Hiebert, M.L. Kamil (eds.) *Teaching and learning vocabulary: Bringing research to practice*, Mahway, NJ: Lawrence Erlbaum. 45-68.
- Dautriche, I. (2015) *Weaving an Ambiguous Lexicon (PhD)*.
- Dautriche, I., Fibla, L., Fievet, A., & Christophe, A. (2018). Learning homophones in context: Easy cases are favored in the lexicon of natural languages. *Cognitive Psychology*, 104, 83-105. doi:10.1016/j.cogpsych.2018.04.001
- Delgado, C. (2012). *File:BIC blue pen cap.jpg*). Wikimedia Commons, the free media repository. Retrieved 13:11, June 11, 2018 from https://commons.wikimedia.org/w/index.php?title=File:BIC_blue_pen_cap.jpg&oldid=218255820.
- Dell, G. S., & Gordon, J. K. (2003). Neighbors in the lexicon: Friends or foes. *Phonetics and phonology in language comprehension and production: Differences and similarities*, 6, 9-37.
- Deng W., Sloutsky V. M. (2016). Selective attention, diffused attention, and the development of categorization. *Cognitive Psychology*, 91, 24–62.
- Durkin, K & Manning, J (1989). Polysemy and the Subjective Lexicon: Semantic Relatedness and the Salience of Intraword Meanings. *Journal of Psycholinguistic Research*, 18, 577-612.
- Egstrom, G.G., Weltman, G., Baddeley, A.D., Cuccaro, W.J., and Willis, M.A. (1972). *Underwater work performance and work tolerance*. Report no. 51, Bio-Technology Laboratory, University of California, Los Angeles.
- Fazly, A., Alishahi, A., & Stevenson, S. (2010). A probabilistic computational model of cross-situational word learning. *Cognitive Science*, 34(6), 1017–1063.
- Fellbaum, C. (Ed.) (1998) *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.
- Fillmore, C. J. (1992). Corpus linguistics or computer-aided armchair linguistics. *Directions in corpus linguistics. Proceedings of Nobel Symposium* (82), pp. 35-60.
- Fillmore, C. J., & Langendoen, D.T. (1971). *Studies in linguistic semantics*. (Edited by Charles J. Fillmore and D. Terence Langendoen.). New York, etc.: Holt, Rinehart & Winston
- Fillmore, C. J. (1977). Topics in lexical semantics. *Current issues in linguistic theory*, 76, 138.
- Flavell, J. H., Speer, J. R., Green, F. L., August, D. L., & Whitehurst, G. J. (1981). The development of comprehension monitoring and knowledge about communication. *Monographs of the Society for Research in Child Development*, 1-65.
- Floyd, S. & Goldberg, A.E. (in prep). *Children on the Autism spectrum are challenged by complex word meanings*.
- Foraker, S., & Murphy, G. L. (2012). Polysemy in sentence comprehension: Effects of meaning dominance. *Journal of Memory and Language*, 67(4), 407-425.
- François, A. (2008). Semantic maps and the typology of colexification. From polysemy to semantic change: *Towards a typology of lexical semantic associations*, 106, 163.
- Frank, M. C., Goodman, N. D., & Tenenbaum, J. B. (2007). A Bayesian framework for cross-situational word-learning. In *Advances in neural information processing systems*.
- Frazier, L., & Rayner, K. (1990). Taking on semantic commitments: Processing multiple meanings vs. multiple senses. *Journal of Memory and Language*, 29(2), 181–200.
- Gathercole, S. E., Willis, C. S., Emslie, H., Baddeley, A. D. (1992). "Phonological memory and vocabulary development during the early school years: A longitudinal study". *Developmental Psychology*. 28 (5): 887–898. doi:10.1037/0012-1649.28.5.887.
- Geeraerts, D. (1993) Vagueness's puzzles, polysemy's vagaries. *Cognitive Linguistics* 4, 223-272.
- Gentner, D. (1978). A study of early word meaning using artificial objects: What looks like a jiggy but acts like a zimbo. *Papers and Reports on Child Language Development*, 15, 1–6.
- Gentner, D. (1978). On relational meaning: The acquisition of verb meaning. *Child Development*, 49, 988-998.
- Gentner, D. (1981). Verb semantic structures in memory for sentences: Evidence for componential representation. *Cognitive Psychology*, 13, 56-83.

- Godden, D and Baddeley, A. (1975). Context dependent memory in two natural environments. *British Journal of Psychology*, 66 (3): 325–331.
- Goldberg, A.E. (2019). *Explain me this: Creativity, Competition and the Partial Productivity of Constructions*. Princeton NJ: Princeton University Press.
- Golinkoff, R. M., Hirsh-Pasek, K., Bailey, L. M., & Wenger, N. R. (1992). Young-children and adults use lexical principles to learn new nouns. *Developmental Psychology*, 28, 99–108. doi: 10.1037/0012-1649.28.1.99
- Gombert, J. E. (1992). *Metalinguistic development*. Chicago: University of Chicago Press.
- Harmon, Z., & Kapatsinski, V. (2017). Putting old tools to novel uses: The role of form accessibility in semantic extension. *Cognitive Psychology*, 98, 22-44. doi:10.1016/j.cogpsych.2017.08.002
- Heine, B., & Reh, M. (1984). *Grammaticalization and reanalysis in African languages*. Hamburg: Helmut Buske Verlag.
- Hinton, G. E., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review*, 98, 74-95. doi:10.1037/0033-295X.98.1.74
- Horst J.S. & Hout M.C. (2016). The Novel Object and Unusual Name (NOUN) Database: A collection of novel images for use in experimental research. *Behav. Res. Methods*, 48: 1393-1409
- Horst, J. S., & Samuelson, L. K. (2008). Fast Mapping but Poor Retention by 24-Month-Old Infants. *Infancy*, 13(2), 128-157. doi:10.1080/15250000701795598
- Imai, M. & Gentner, D. (1997). A crosslinguistic study on constraints on early word meaning: Linguistic influence vs. universal ontology. *Cognition* 62:169-200.
- Jackson, J.C., Watts, J., Henry, T.R., List, J.M., Forkel, R., Mucha, P.J., Greenhill, S.J., Gray, R.D. and Lindquist, K.A., 2019. Emotion semantics show both cultural variation and universal structure. *Science*, 366(6472), pp.1517-1522.
- Joordens, S., & Besner, D. (1994). When banking on meaning is not (yet) money in the bank: Explorations in connectionist modeling. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20(5), 1051-1062.
- Kawamoto, A. & Farrar, W. & Kello, C. (1994). When two meanings are better than one: Modeling the ambiguity advantage using a recurrent distributed network. *Journal of Experimental Psychology: Human Perception and Performance*. 20. 1233-1247. 10.1037/0096-1523.20.6.1233.
- Klein, D. E., & Murphy, G. L. (2001). The representation of polysemous words. *Journal of Memory and Language*, 45, 259–282.
- Klein, D. E., & Murphy, G. L. (2002). Chapter has been my ruin: Conceptual relations of polysemous senses. *Journal of Memory and Language*, 47, 548–570.
- Klepousniotou, E., & Baum, S. R. (2007). Disambiguating the ambiguity advantage effect in word recognition: An advantage for polysemous but not homonymous words. *Journal of Neurolinguistics*, 20(1), 1–24. doi:10.1016/j.jneuroling. 2006.02.001
- Kruschke, J. K. (1992). ALCOVE: An exemplar-based connectionist model of category learning. *Psychological Review*, 99(1), 22-44. (doi:10.1037/0033-295X.99.1.22)
- Kruschke, J. K. (2005). Learning involves attention. In: G. Houghton (Ed.), *Connectionist Models in Cognitive Psychology*, Ch. 4, pp. 113-140. Hove, East Sussex, UK: Psychology Press.
- Kuznetsova A, Brockhoff PB and Christensen RHB (2017). “lmerTest Package: Tests in Linear Mixed Effects Models.” *Journal of Statistical Software*, 82(13), pp. 1–26. doi: 10.18637/jss.v082.i13.
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package ‘lmerTest’. *R package version*, 2(0).
- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Landau, B., Smith, L. B., & Jones, S. S. (1988). The importance of shape in early lexical learning. *Cognitive Development*, 3(3), 299-321. DOI: 10.1016/0885-2014(88)90014-7.
- Landau, B., Smith, L. B., & Jones, S. (1998). Object shape, object function, and object name. *Journal of Memory and Language*, 38, 1–27.

- Lehrer, A. (1990). Polysemy, conventionality, and the structure of the lexicon. *Cognitive Linguistics*, 1, 207–246.
- Litman, L., Robinson, J., & Abberbock, T. (2016). TurkPrime. com: A versatile crowdsourcing data acquisition platform for the behavioral sciences. *Behavior Research Methods*, 1-10.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: the neighborhood activation model. *Ear and hearing*, 19(1), 1–36.
- Malt, B.C. (2010). Naming artifacts: Patterns and processes. In B. Ross (Ed.), *The psychology of learning and motivation: Advances in research and theory* (pp. 1-38). Amsterdam: Elsevier.
- Markman, E. M., & Wachtel, G. F. (1988). Children's use of mutual exclusivity to constrain the meanings of words. *Cognitive psychology*, 20(2), 121-157.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18(1), 1-86.
- McMurray, B., Horst, J. S., & Samuelson, L. K. (2012). Word learning emerges from the interaction of online referent selection and slow associative learning. *Psychological Review*, 119, 831–877. <http://dx.doi.org/10.1037/a0029872>
- Merriman, W. E., Bowman, L. L., & MacWhinney, B. (1989). The mutual exclusivity bias in children's word learning. *Monographs of the society for research in child development*, i-129.
- Meyer, D.E., & Schvaneveldt, R.W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90(2), 227-234.
- Murphy, G.L. (2004). *The big book of concepts*. Cambridge, MA: MIT Press.
- Navigli, R., & Ponzetto, S.P. (2012). BabelNet: The automatic construction, evaluation and application of a wide-coverage multilingual semantic network. *Artificial Intelligence*, 193, 217-250.
- Nosofsky, R.M. (1984). Choice, similarity, and the context theory of classification. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 10, 104-114.
- Nosofsky, R. M., Gluck, M. A., Palmeri, T. J., McKinley, S. C., & Glauthier, P. (1994). Comparing models of rule-based classification learning: A replication and extension of Shepard, Hovland, and Jenkins (1961). *Memory & Cognition*, 22, 352-369.
- Nunberg, G. (1979). The non-uniqueness of semantic solutions: polysemy. *Linguistics and Philosophy*, 3(1).
- Piantadosi, S.T. , Tily, H. & Gibson, E. (2011). The communicative function of ambiguity in language. *Cognition*, 122. 280—291
- Pinker, S. (2007). *The Stuff of Thought: Language as a Window into Human Nature*. New York, NY: Viking.
- Plebanek, D. J., & Sloutsky, V. M. (2017). Costs of Selective Attention: When Children Notice What Adults Miss. *Psychological Science*, 28(6), 723–732. <https://doi.org/10.1177/0956797617693005>
- Pustejovsky, J. (1991) The Generative Lexicon. *Computational Linguistics* 17 (4): 409-411.
- Pye, C. (2017). *The Comparative Method of Language Acquisition Research*. 10.7208/chicago/9780226481319.001.0001.
- Rabagliati H. & Snedeker J. (2013). The Truth About Chickens and Bats: Ambiguity Avoidance Distinguishes Types of Polysemy,” *Psychological Science*, 24(7), 1356–1360.
- Rabagliati, H., Marcus, G.F., & Pytkänen, L. (2010). Shifting senses in lexical semantic development. *Cognition*, 117(1), 17-37.
- Ramiro, C., Srinivasan, M., Malt, B.C., & Xu, Y. (2018). Algorithms in the Historical Emergence of Word Senses. *Proceedings of the National Academy of Sciences*, 115, (10).
- Ramscar, M., Hendrix, P., Shaoul, C., Milin, P., & Baayen, H. (2014). The myth of cognitive decline: Non-linear dynamics of lifelong learning. *Topics in cognitive science*, 6(1), 5-42.
- Rodd, J., Gaskell, G., & Marslen-Wilson, W. (2002). Making sense of semantic ambiguity: Semantic competition in lexical access. *Journal of Memory and Language*, 46(2), 245-266.
- Rodd, J. (2004). When do leotards get their spots? Semantic activation of lexical neighbors in word recognitions. *Psychonomic bulletin & review* (11). 434-9. 10.3758/BF03196591.

- Rodd, J. & Berriman, R & Landau, M & Lee, T & Ho, C & Gaskell, G & Davis, M. (2012). Learning new meanings for old words: Effects of semantic relatedness. *Memory & cognition*, 40, 1095-108. 10.3758/s13421-012-0209-1.
- Rosch, E., & Mervis, C.B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive psychology*, 7(4), 573-605.
- RStudio: Integrated Development for R. RStudio, Inc., Boston, MA URL <http://www.rstudio.com/>
- Samuelson, L. K., and Smith, L. B. (2005). They call it like they see it: spontaneous naming and attention to shape. *Dev. Sci.* 8, 182–198. doi: 10.1111/j.1467-7687.2005.00405.x
- San Roque, L., Kendrick, K. H., Norcliffe, E., & Majid, A. (2018). Universal meaning extensions of perception verbs are grounded in interaction. *Cognitive Linguistics*, 29(3), 371-406.
- Shepard, R. N., Hovland, C. I., & Jenkins, H. M. (1961). Learning and memorization of classifications. *Psychological Monographs*, 75 (13, Whole No. 517).
- Siedlecki, P. (2018). [Image of baseball bat] Retrieved from <https://www.publicdomainpictures.net/en/view-image.php?image=151442&picture=baseball-bat>
- Smith L. B., Kemler D. G. (1977). Developmental trends in free classification: Evidence for a new conceptualization of perceptual development. *Journal of Experimental Child Psychology*, 24, 279–298.
- Smith, J. D., & Minda, J. P. (1998). Prototypes in the mist: The early epochs of category learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 24(6), 1411-1436.
- Smith, L. B., Jones, S. S., Landau, B., Gershkoff-Stowe, L., and Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science* 13, 13–19. doi: 10.1111/1467-9280.00403
- Srinivasan, M., & Snedeker, J. (2011). Judging a book by its cover and its contents: The representation of polysemous and homophonous meanings in four-year-old children. *Cognitive Psychology*, 62(4), 245-272.
- Srinivasan, M. & Snedeker, J. (2014). Polysemy and the taxonomic constraint: Children’s representation of words that label multiple kinds, *Language Learning and Development*, 10(2), 97-128
- Srinivasan, M. & Rabagliati, H. (2015). How concepts and conventions structure the lexicon: Cross-linguistic evidence from polysemy. *Lingua* <http://dx.doi.org/10.1016/j.lingua.2014.12.004>
- Srinivasan, M., Al-Mughairy, S., Foushee, R., & Barner, D. (2017). Learning language from within: Children use semantic generalizations to infer word meanings. *Cognition*, 159, 11-24.
- Srinivasan, M., Berner, C., & Rabagliati, H. (2018). Children use polysemy to structure new word meanings. *Journal of Experimental Psychology: General*. doi:10.1037/xge0000454
- Stevens, J.S., Gleitman, L.R., Trueswell, J.C., & Yang, C. (2016). The Pursuit of Word Meanings, *Cognitive Science*, 1–39.
- Steyvers, M., & Tenenbaum, J. B. (2005). The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive science*, 29(1), 41-78.
- Sweetser, E. (1990). *From Etymology to Pragmatics: Metaphorical and Cultural Aspects of Semantic Structure (Cambridge Studies in Linguistics)*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511620904
- Swinney, D.A. (1979). Lexical access during sentence comprehension:(Re) consideration of context effects. *Journal of verbal learning and verbal behavior*, 18(6), 645-659.
- Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological research*, 49(2-3), 161-167.)
- Tanenhaus, M., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, 268(5217), 1632-1634. doi:10.1126/science.7777863
- Taylor, J. (1995). *Linguistic categorization. Prototypes in linguistic theory*. Oxford University Press.
- TexasRebel. (2007). *File:Baseball cap.jpg*. Wikimedia Commons, the free media repository. Retrieved 13:07, June 11, 2018 from https://commons.wikimedia.org/w/index.php?title=File:Baseball_cap.jpg&oldid=124032924

CHILDREN USE RELATIONSHIPS ACROSS MEANINGS IN WORD LEARNING

- Trueswell, J.C., Medina, T.N., Hafri, A., & Gleitman, L.R. (2013). Propose but verify: Fast mapping meets cross-situational word learning. *Cognitive Psychology*, 66(1), 126–156.
- Tuggy, D. (1993) Ambiguity, polysemy, and vagueness. *Cognitive Linguistics* 4, 273-290.
- Wittgenstein, L. (1953). *Philosophical investigations*. John Wiley & Sons.
- Yoshida, H., Smith, L.B. (2003). Shifting ontological boundaries: How Japanese- and English-speaking children generalize names for animals and artifacts. *Developmental Science*, 6, 1–34.
- Yu, C. (2008). A statistical associative account of vocabulary growth in early word learning. *Language learning and Development*, 4(1), 32-62.
- Yu, C., & Smith, L.B. (2007). Rapid word learning under uncertainty. *Psychological Science* 18(5), 414-420.
- Yurovsky, D., Yu, C., Smith, L. B. (2013). Competitive processes in cross-situational word learning. *Cognitive Science*, 37, 891–921.
- Zettersten, M. Wojcik, E., Benitez, V.L. & Saffran, J. (2018). The company objects keep: linking referents together during cross-situational word learning. *Journal of Memory and Language*, 99:62-73.
- Zipf, G. K. (1945). The meaning-frequency relationship of words. *Journal of General Psychology*, 33, 251-256. <http://dx.doi.org/10.1080/00221309.1945>.

