Is it a bat or a thing?
Referential contrast in the learning of homophones and superordinate terms

Daoxin Li (daoxinli@sas.upenn.edu) and **Christine Soh Yue (csohyue@sas.upenn.edu)**
Department of Linguistics, University of Pennsylvania
3401-C Walnut Street, Philadelphia, PA 19104 USA

John Trueswell (trueswel@psych.upenn.edu)
Department of Psychology, University of Pennsylvania
425 S. University Ave., Philadelphia, PA 19104 USA

Abstract

An important aspect of word learning is semantic generalization: when a novel word (e.g., 'fami') refers to something in the world, does that word have a more specific meaning (e.g., *dog*) or more general meaning (e.g., *animal*)? Here we focus on the role of semantic contrast between referential alternatives, which is an unstudied component of the word-learning process. We do this in the context of learning novel words cross-situationally, asking when learners adopt more specific meanings across instances (resulting in homophonic words: e.g., 'fami' means both *dog* and *butterfly*) or adopt a single superordinate meaning (e.g., 'fami' means *animal*). We hypothesize that learners' decisions about specific vs. general meanings will be informed by the semantic contrast among candidate referents within the referential domain. Learners will be more likely to establish homophonous meanings when contrasting referents are from a neighboring category of the target, and more likely to establish a superordinate meaning when contrasting referents are from more distant categories. We also expect homophone learning to be more difficult because of its additional demands on learning and memory: put simply, confirming two meaning hypotheses requires more evidence than confirming one. Our predictions were borne out in a series of experiments and in a cross-situational model of word learning. Overall, this study offers new perspectives for how learners form hypotheses for novel word meanings.

Keywords: word learning; referential contrast; homophone; superordinate term; memory

Introduction

When learning the meanings of novel words from their use in language, learners must determine not only what is being referred to (e.g., this dog or that cat) but also what is the relevant semantic generalization intended by the speaker (e.g., is the referent being characterized as a Dalmatian, dog or animal?) (e.g., Chomsky, 1959; Quine, 1960). Here we explore the role of referential contrast in determining semantic generalization. Referential contrast has been found to have a potent effect in the domain of speaker's lexical choice in language production: given a group of animals from different basic-level categories, if speakers want to pick out the Dalmatian, it will be sufficient to use the basic-level term 'the dog'; by contrast, when there are different breeds of dogs in the referential domain, speakers need to use a more specific, subordinate term like 'the Dalmatian' (e.g., Brennan & Clark, 1996). Here we bring this idea into the domain of word learning. We will examine the role of referential contrast in the learning of words cross-situationally, focusing on whether learners will adopt two specific meanings for a word (i.e., a homophone) or one broader meaning (i.e., a superordinate) from multiple encounters with a word. Learning homophones and superordinate terms are highly relevant to the study of semantic generalization because for both types of words, a label permits a broad range of referents, making it difficult to determine if a novel word has two specific meanings or one superordinate meaning (e.g., Dautriche & Chemla, 2016; Dautriche, Chemla & Christophe, 2016). For instance, 'bat' is used to refer to both the animal-bat and the baseball-bat. In this case, how do learners decide the meaning of 'bat': whether to learn two specific meanings, or to generalize to a superordinate meaning that encompasses both referents, like *thing*?

We hypothesize that local referential contrast will have a significant effect on learning outcomes. Consider a scenario in which a learner observes two actors using an unfamiliar language to converse about, and act upon, co-present objects (Fig 1a). If one person instructs the other to pick up an object and the other does so, conversational principles (e.g., Grice, 1975) dictate that this speech act must have been sufficiently informative to distinguish it from competing referents. Across multiple exchanges of this sort, when a particular novel word ('fami') has been paired with a sequence of semantically diverse referents (e.g., dog, butterfly, butterfly, dog) we predict we can drive learners to adopt one (superordinate) meaning or two (homophonous) meanings of this word solely by manipulating the referential competitors present on each trial (Fig 1b): Learners will be more likely to learn homophonous meanings (e.g., *dog* and *butterfly*) in the Lower-level contrast condition where the semantically contrasting referents are from a neighboring category of the target referents (e.g., other animals than the dog and the butterfly), and more likely to learn a superordinate meaning (e.g., *animal*) in the Higher-level contrast condition where the semantically contrasting referents are from a distant category (e.g., *office supplies*).

But critically, the learning of homophones is predicted to be challenging because it imposes additional, specific demands on learning and memory: If, across a series of learning instances for a word, a learner is entertaining that a word is a homophone, then the learner must necessarily

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

maintain two meanings in memory which are tested against the referents. If instead across that same sequence, learners are entertaining a single broader meaning, only one meaning needs to be retained and tested. Moreover, more confirmatory evidence is needed to learn homophones as the number of learning exposures are split in half when a word is treated by a learner as carrying two distinct meanings: In the four-exposure design of Figure 1b, for a homophone hypothesis, only two trials support each meaning; whereas for a superordinate hypothesis, all four trials support it. Below we briefly review prior work on learning word meaning generalizations, and the learning of homophones vs. superordinate terms before presenting our study.

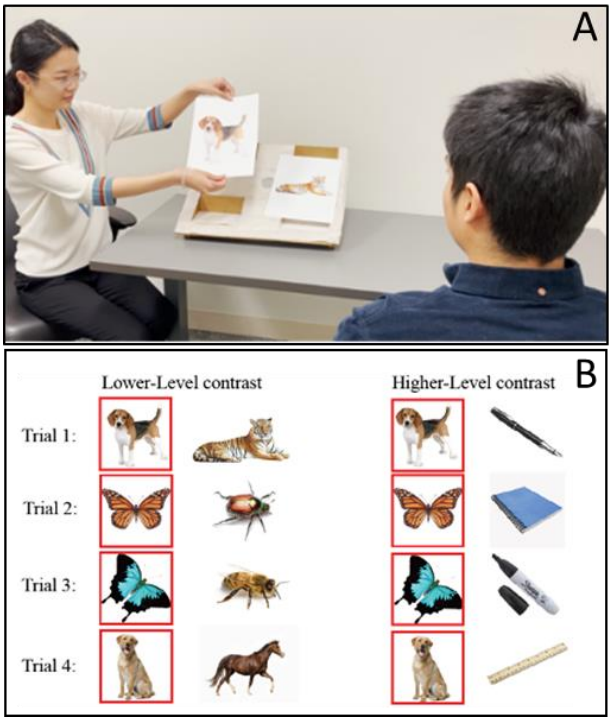

Figure 1: Design of the learning phase. 1a: Still frame example for a video in the learning phase. 1b: Objects shown in each learning trial for the word 'fami' in both conditions. The red square indicates the selected object.

Learning word meanings

There is a long history of studying how individuals make semantic generalizations during word learning (see e.g., Woodard & Markman, 1998; Bloom, 2002). Of particular interest to the present work is the observation that word meaning specificity is in part driven by semantic contrast among the words being used and learned within a local context. Most past work has focused on the role of explicit semantic contrast among known or to-be-learned lexical items (e.g., another novel word referring to exemplars that contrasted with those of the target word), rather than implicit referential contrast as studied here. In particular, although learners are known to have a "basic-level" bias (e.g., preferring to think that 'mipen' means *dog* rather than *animal* or *Dalmatian* in the presence of a spotted dog, Rosch & Mervis, 1975; Rosch, et al. 1976; Golinkoff, et al., 1995), non-basic-level meanings are adopted when explicit semantic comparison and contrast is provided. E.g., observational studies have found that adults will "anchor" novel non-basic-level words to a familiar basic-level term when introducing them to children (e.g., "See this dog? It's a poodle." "It's a kind of dog.", Blewitt, 1983; Shipley, Kuhn & Madden, 1983; Callanan, 1985; Clark & Wong, 2002). Experimental work has shown that contrastive information indeed has an important effect on children's hypotheses for word meaning regarding the level of generalization (e.g., children are more likely to learn non-basic-level meanings with anchoring such as "This is a wug. A wug is a kind of terval." than in ostensive labeling contexts, Callanan, 1989; Waxman, 1990; Waxman et al., 1991, 1997).

More recent experimental and modeling work has explored the extent to which the level of semantic generalization may be accomplished cross-situationally from a series of ostensive labeling events even in the absence of explicit semantic contrast. Xu and Tenenbaum (2007) offered evidence that learners can detect "suspicious coincidences" via Bayesian inference (e.g., several Dalmatians each labeled as a 'mipen' more likely support a subordinate over a basic-level meaning). However, debates exist as to the robustness of this effect (Spencer et al., 2011), and whether it is instead a product of semantic contrast: Wang and Trueswell (2019, 2022) found that so-called suspicious-coincidence effects only arise when learners are simultaneously taught another novel word whose exemplars contrasted with those of the target word at the subordinate level (e.g., different dogs other than Dalmatians; see also Choe & Papafragou, 2023, for similar findings).

Thus, semantically contrasting lexical items play an important role in identifying novel word meanings. This suggests the more general role within referential contrast explored here.

Learning homophones and superordinate terms

Early studies of homophone learning have found that preschool children generally have difficulty with learning two distinct meanings for the same word (Casenhiser, 2005; Doherty, 2004; Mazzocco, 1997). It is more difficult for them to assign a new meaning to a familiar word (e.g. learning 'cup' can also label an unfamiliar object) than to learn a new meaning for a novel word (e.g. learning 'zud' labels an unfamiliar object), especially when the familiar meaning and the new meaning are syntactically or semantically close (Dautriche et al., 2018). Given the context of cross-situational word learning, one theory for homophone learning is that learners can learn homophones by precisely tracking the co-occurrence frequencies between all words and referents across learning instances: When there are two referents that frequently co-occur with one label across situations, learners would learn the word as a homophone. Yurovsky and Yu (2008) conducted a cross-situational word learning experiment with adults in support of this proposal. However,

because the experiment used unfamiliar objects, it is not entirely clear whether the participants associated the words with two distinct meanings as expected, or with one generalized meaning for the category that includes both referents. Indeed, Dautriche, Chemla and Christophe (2016) found both adults and 5-year-old French children were more likely to learn a superordinate meaning when the exemplars for a novel word were in a uniform distribution according to taxonomic hierarchies (e.g., a snake, a bird, a monkey, and a squirrel), and more likely to learn distinct, homophonous meanings when the exemplars were in a bimodal distribution (e.g., two different kinds of snakes and two different kinds of monkeys). Dautriche and Chemla (2016) observed similar patterns but also observed patterns reminiscent of lexical semantic contrast: Participants were more likely to learn homophones than superordinate terms when there was an intervention of other lexical items in the gap between the two meanings of the homophone.

In summary, learning homophones has been observed to be challenging, where people tend to avoid associating one word with different meanings. Factors that have been found to influence the learning of homophony vs. superordinate terms include organization of learning exemplars in the conceptual space and linguistic information. Here we examined another factor, i.e., referential-semantic contrast in the local domain.

The current study

In this study we have two hypotheses. First, local referential contrast informs learners' hypothesis for word meanings and thus influences the learning of homophones vs. superordinate terms. A local referential contrast at the superordinate-level will support the learning of a superordinate meaning, while a referential contrast from a neighboring category will support the learning of homophones. As discussed earlier, referential contrast plays an important role in the choice of definite expressions during language production (e.g., Brennan & Clark, 1996). Here we predict that this mechanism can also be used to determine word meaning generalization in cross-situational word learning. Our second hypothesis is that homophone learning will be more challenging and will benefit more from a greater number of learning trials compared to superordinate terms. We test our hypotheses in two experiments below.

Experiment 1: Four exposures per word

Experiment 1 exposed participants to four learning trials per novel word. They were instructed to learn a new language by watching videos where two people communicated in the language (Fig 1a). The novel words were systematically paired with semantically diverse referents in a particular order (e.g., 'fami' with a dog, a butterfly, another butterfly and another dog), and we manipulated the referential contrast across conditions to examine its effect (Fig 1b). In the test after exposure, we asked participants whether the nouns can refer to other objects to examine what word meaning they learned.

Methods

Participants Ninety-three English speaking adults were recruited from either the undergraduate subject pool at the University of Pennsylvania or Prolific. Fifty-one were assigned to the Lower-level contrast condition, and forty-two to the Higher-level contrast condition. The experiment was conducted online via PCIbex (Zehr & Schwarz, 2018).

Materials We created an artificial language using pseudo-words consisting of eight nouns, two verbs, and one exclamation. Words conformed to English phonotactics. Nouns were 'fami' 'kefu' 'lepa' 'nunu' 'poru' 'sati' 'supo' and 'tufa'. Verbs were 'dax' for 'pick something up' and 'zep' for 'turn something around'. An exclamation 'timo!' meant 'great job'. The meanings of the verbs and the exclamation word were taught at the beginning of the experiment.

Noun-learning videos were created. In each, an actress was seated at a table that had two photographs. An actor faced the actress (Fig 1a). After a two second pause, the actor produced an utterance (e.g., "Dax fami") and the actress then acted on an image (i.e., picking up the dog). The actor then exclaimed "timo!" and nodded his head. We used this artificial language paradigm instead of just teaching participants the novel nouns using English to better approximate the natural circumstances of early word learning.

Each target noun had four learning videos. The actress's selections across these videos (Fig 1b) permitted two basic-level meanings. E.g., for 'fami', the target was the dog in two videos and the butterfly in the other two. In principle the novel noun could have either two distinct meanings (e.g., *dog* and *butterfly*) or a superordinate meaning that encompasses both (e.g., *animal*). Our crucial manipulation was the unselected image on each trial. In the Lower-level contrast condition, they were from two neighboring categories from the selected images (i.e., other mammals for dog targets and other insects for butterfly targets); whereas in the Higher-level contrast condition, competitors were from a single distant category (e.g., office supplies). We counterbalanced across trials the verb ('dax' or 'zep') used in the instruction and the left/right position of selected image.

If participants can use referential contrast to inform word meaning, then they would be more likely to learn two distinct meanings, dog and *butterfly*, in the Lower-level contrast condition and a single superordinate meaning *animal* in the Higher-level contrast condition. To test this, test materials for each word were designed to probe semantic generalizations. For each word, we prepared two tests. Participants would see an attested referent from the exposure phase (e.g., a dog) and told "This was called a fami". They then saw four new exemplars, each time asked "Is this also a fami?" (Fig 2). We denote the four test items as Level-1 to Level-4 items. The Level-1 item was from the same basic-level category as the attested referent, e.g., another dog. The Level-2 item was from a slightly broader category that encompassed the attested referent (e.g., dog) but not the other attested referent of the same word (e.g., butterfly). For example, we picked a

pig from the *mammal* category, which encompasses the dog but not the butterfly. The Level-3 item was from an even broader category that encompassed both referents for the word, e.g., a swan, from the *animal* category. Finally, the Level-4 test items came from an outside category, e.g., a door, which was not from the *animal* category. Overall, these test items from different categories would allow us to determine participants' level of meaning generalization.

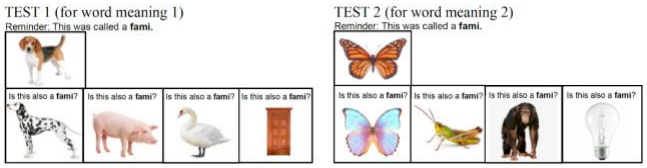

Figure 2: Reminder and test pictures for Referent 1 (left) and Referent 2 (right) of 'fami'.

Procedure The experiment consisted of a familiarization phase, a learning phase, and a test phase. During the familiarization, participants were told that they were going to learn a new language, which is called Tulka. They were then explicitly taught the meanings of three words: 'dax' for 'pick something up', 'zep' for 'turn something around' and 'timo' for 'great job'. Participants were tested on these words via multiple-choice to assure they learned them. They could not proceed until all questions were answered correctly.

During the learning phase, participants were told that they would now learn the names for objects in Tulka by watching two people who are experts in Tulka communicating in this language. Their job was to listen to the speech, watch the actress carry out the action, and then repeat aloud what they heard the actor say. The repetition requirement was to assure that participants paid attention to the stimuli. Responses were recorded by PCIbex's Media Recorder. Participants were first presented with a practice trial in English. Then the learning trials began, where the participants watched the videos for the eight novel nouns (32 trials total). The order of the eight nouns were counterbalanced across participants. For half of the participants exposure words were blocked such that all four exposures for a word occurred in a row. All other participants received interleaved exposure during learning.

The test phase began after watching all the learning videos. For each attested referent, participants were first reminded of the noun that co-occurred with it during the learning phase. For example, they would be presented with the picture of an attested dog, and reminded that "this was called a fami". At the same time, we played an audio of the pseudo-noun, which was recorded by the same actor from the learning videos. We included the reminder in our test phase because we did not intend to test participants' memory; instead, what we were interested in was to what extent participants would generalize the meaning given that they already learned the pairing between the word label and the referent. After seeing the reminder, participants would need to decide for each of the four test items whether it could also be referred to by the novel noun (Fig 2): "How about this? Is this also a fami?" Only one image was shown on the screen at a time, and the order of the four test items were randomized. The four test items for one referent were presented immediately next to one another, while the two referents for a single word were interleaved with other words. We used this test paradigm to make it natural for homophones: In the real world it is unlikely that there will be situations where the *animal-bat* and the *baseball-bat* appear at the same time and people need to determine whether they are both 'bats'. Instead, the *animal-bat* and the *baseball-bat* will occur in different situations, and people will make the decision whether it is a 'bat' separately for the two referents. Therefore, we decided to only test one referent at a time and include intervening materials between the tests for the two referents of one word, instead of putting them all in a grid.

Coding To determine whether participants learned one superordinate or two distinct meanings for a novel word, we need to consider their responses for both referents of the word at the same time. The responses were coded into one of three categories: 'Homophone', 'Superordinate', and 'Other'. They were coded as 'Homophone' if participants answered 'yes' to the Level-1 tests of both referents of a word, and answered 'no' to all the other tests. For example, consider the meaning for 'fami' in Figure 2. Participants were considered to learn two different meanings, *dog* and *butterfly*, if they decided that only the new dog and the new butterfly among all the test items could be called a 'fami'. Responses were coded as 'Superordinate' if participants answered 'no' to the Level-4 tests of both referents of a word and answered 'yes' to all the other tests. To illustrate, consider 'fami' again: If participants learned a superordinate meaning *animal*, then all the test items but the door and the bulb would be a 'fami'. All other response patterns were coded as 'Other'.

Results and discussion

The proportion of different response types (Homophone, Superordinate, Other) is shown in Figure 3. Blocking vs. interleaving words during exposure had no effect on responding, nor did this interact with condition. Thus, for simplicity, we collapsed across blocking when reporting results here. We analyzed the results through mixed-effects logistic regression using the lmerTest package in R. First, hierarchical modeling showed that Condition was a significant predictor of 'Homophone' responses ($X^2(1)=14.21$, $p<0.001$): There were significantly fewer 'Homophone' responses in the Higher-level contrast condition compared to the Lower-level contrast condition ($\beta=-4.47$, $SE=1.84$, $z=-2.43$, $p=0.02$). Similarly, for 'Superordinate' responses, we also found Condition to be a significant predictor ($X^2(1)=20.52$, $p<0.001$), with significantly more 'Superordinate' responses in the Higher-level contrast condition than the Lower-level contrast condition ($\beta=3.95$, $SE=0.62$, $z=6.42$, $p<0.001$).

Notably, in the Lower-level contrast condition, there was a substantial proportion of 'Other' responses (62.7%). In these responses, participants almost always selected 'yes' to the

two Level-1 test items and almost never selected 'yes' to the two Level-4 test items; that means, the 'Other' responses were typically somewhere between 'Homophone' responses and 'Superordinate' responses, where participants selected 'yes' to both Level-1 test items and also some but not all of the Level-2 and Level-3 test items. The proportions of selecting three, four, and five 'yes' answers among the eight test items for a word were similar, without further systematic patterns that could be identified. This was actually expected given the memory constraint: Without enough opportunities to test their hypotheses, participants were at chance.

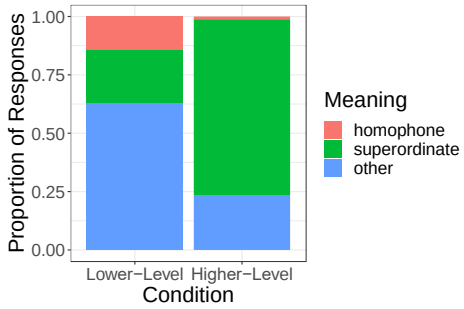

Figure 3: Coded responses at test for Experiment 1.

The results of this study were as predicted in that participants were more likely to assign two distinct meanings to words learned in the Lower-level contrast condition than words in the Higher-level contrast condition. This occurred even though the particular referents demonstrated for the words were identical across the condition (e.g., two dogs and two butterflies). The only difference across the conditions was the nature of the referential competitors.

Experiment 1 has also found an overall paucity of Homophone responses, although they were more frequent in the Lower-level contrast condition than in the Higher-level contrast condition. This is likely due to the uncertainty associated with maintaining two meanings rather than one: each meaning was only confirmed once, as opposed to the superordinate meaning that was confirmed three times. The closeness of the two meanings of our homophones (e.g., *dog/butterfly*) may have also contributed to difficulty (as observed in e.g., Dautriche et al., 2018).

Experiment 2: Eight exposures per word

In Experiment 2, we doubled the number of learning trials while keeping all other aspects of the design the same. We predicted that homophone learning should be easier in Experiment 2 compared to Experiment 1. Now each homophonous meaning (e.g., *dog* and *butterfly*) is confirmed three times rather than once as was the case in Experiment 1.

Methods

Participants Sixty-one English-speaking adults who did not previously participate in Experiment 1 were recruited from the undergraduate subject pool at the University of Pennsylvania. Thirty-five of the participants were assigned to the Lower-level contrast condition, and twenty-six to the Higher-level contrast condition.

Materials Materials were the same as in Experiment 1, except that we created a new learning video for each selected image in existing videos. In each new video, the unselected image and the verb in the instruction were the same as the existing video, but we switched the left/right position of the selected image and the unselected image to prevent participants from learning any association between words and positions. Therefore, each word had eight learning trials.

Procedure The procedure was the same as Experiment 1, except that the number of exposure trials were doubled.

Coding The coding scheme was the same as Experiment 1.

Results and discussion

The proportion of different response types is shown in Figure 4. We analyzed the results using mixed effects regression. Condition was a significant predictor for both 'Homophone' responses ($X^2(1)=38.27$, $p<0.001$) and 'Superordinate' responses ($X^2(1)=45.64$, $p<0.001$): There were fewer 'Homophone' responses in the Higher-level contrast condition ($\beta=-5.12$, $SE=0.98$, $z=-5.23$, $p<0.001$) and more 'Superordinate' responses in the Higher-level contrast condition ($\beta=5.79$, $SE=1.00$, $z=5.77$, $p<0.001$). When compared to Experiment 1, post-hoc analyses revealed significantly more 'Homophone' responses in the Lower-level contrast condition in Experiment 2 than Experiment 1 (*odds ratio*=0.07, $SE=0.04$, $z=-4.76$, $p<0.001$), while the rates of 'Superordinate' responses in the Higher-level contrast condition showed no significant difference across experiments (*odds ratio*=2.33, $SE=1.36$, $z=1.45$, $p=0.15$). We also manipulated blocking across groups in Experiment 2 but we found no significant effect or interaction, so we collapsed the groups.

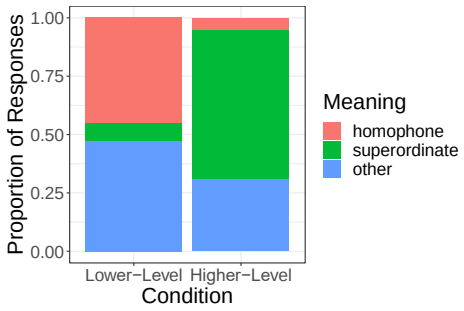

Figure 4: Coded responses at test for Experiment 2.

In summary, in Experiment 2, we successfully replicated the effect of referential contrast on learning homophones vs. superordinate terms. Moreover, given more learning trials, we found learners became much more successful with learning homophones in the Lower-level contrast condition.

General Discussion

In this work we examined whether learners can use referential contrast to learn the correct level of meaning generalization during cross-situational word learning. In particular, while most previous studies on the role of contrast in word learning presented the contrast in different situations and involved learning a label for the semantically contrasting object, e.g., learning the word for Dalmatians in one trial and learning the word for different dogs in another trial, in the current study we examined referential contrast in the local domain, where there were two objects in one trial and the semantically contrasting object was not named. It is established in the language production literature that speakers' lexical choice of referring expressions is influenced by the referential contrast in such situations, and here we set to investigate whether this mechanism can also be used in word learning.

In two experiments, we presented participants with exactly the same word-referent mappings but different referential contexts, so that the alternative referent on each trial contrasted with the target referent at different generalization levels. In the Higher-level contrast condition, the alternative referent contrasted with the target referent at the superordinate level, e.g., it was an inanimate object while both target referents were animals. In such a context, a superordinate term like 'animal' would be sufficient to pick out the target referents, so we predicted that learners would be more likely to learn a superordinate meaning for the novel word. By contrast, if the alternative referents were from a neighboring category as in the Lower-level contrast condition, the learners' hypothesis for the novel word would need to be more specific, e.g., *dog* and *butterfly* instead of *animal* in general; therefore, given enough learning trials, they should be more likely to learn two distinct meanings for a homophone. As predicted, we found that learners can use referential contrast to form hypotheses for novel word meanings. Although it is difficult to learn homophones with relatively close meanings (*dog/butterfly*), learners showed signs of doing so when the referential contrast supported it. Superordinate terms, which violate basic-level preferences, were also arrived at via semantic contrast. Overall, the results indicate that learners arrive at the correct level of meaning generalization not just by tracking all the word-referent pairs, which was kept the same across conditions in our experiments; rather, their hypothesis is dependent on inferences about the speaker's intended message based on the local situational and conversational context, where semantic contrast can play an important role in informing the intended level of meaning generalization.

Our work has also confirmed the memory constraint on word learning. We found that when there were four learning trials for a word, learners struggled to learn homophones, but superordinate meanings were learned easier, which was expected since one needs to keep more hypotheses in mind. Moreover, by definition, two homophone meanings will receive fewer confirmations than one subordinate meaning from the same sequence resulting in less meaning certainty. Consistent with this, homophone learning benefited significantly from a doubling of learning trials, where learners had more opportunities to confirm each of the two hypotheses and learn the word meanings (Exp 2).

Modeling Findings The major findings can be captured by models that aim to resolve referential ambiguity through cross-situational comparisons. Memory-Bound Pursuit (MBP; Soh & Yang, 2021; Yue, LaTourette, Yang & Trueswell, 2023) is a current hypothesis-testing word learning model which incorporates a memory constraint: There is an unlimited lexicon where word-meaning pairs that are sufficiently confirmed are permanently stored, and a finite memory buffer that stores a limited number of hypothesized word-meaning pairs during learning. When the memory buffer reaches its capacity, words are forgotten probabilistically weighted by inverse frequency, i.e., less frequently encountered words are more likely to be removed from the memory buffer. For more details, see Soh & Yang (2021) and Yue et al. (2023). While the influence of referential contrast on resolving semantic ambiguity is not a current aspect of MBP, we can accommodate it by having the model pick a level of meaning generalization: We provide the model with basic-level hypotheses (e.g., *dog*, *butterfly*) in the Lower-level contrast condition and superordinate hypotheses (e.g., *animal*) in the Higher-level contrast condition. We then ran MBP simulations of our experiments and the results are shown in Figure 5. Without real modifications on the algorithm, MBP accurately predicted the effect of condition and number of learning trials.

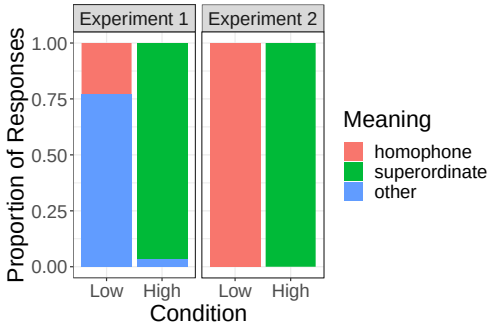

Figure 5: Coded responses at test for MBP simulations.

We are now testing whether children, who are tasked with the challenge of language acquisition with more limited cognitive abilities than adults, can also use referential contrast in the same way to learn homophones vs. superordinate terms. On one hand, previous studies have shown that children are sensitive to semantic contrast in word learning (e.g., Callanan, 1989; Waxman et al., 1991; Wang & Trueswell, 2019); on the other hand, the semantically contrasting object was always explicitly named, so it is an interesting question whether children can still make use of the semantic contrast based solely on the co-present referential domain, and whether they can integrate this information cross situationally.

Acknowledgements

This work is supported by the University of Pennsylvania Language and Communication Sciences Research Funds awarded to D. Li. Thank you to our participants for participating in the experiments; to members of the Language Learning Lab at the University of Pennsylvania for helpful comments; to Abimael Hernandez Jimenez for help with creating the stimuli; and to CogSci reviewers whose comments improved the paper.

References

Blewitt, P. (1983). Dog versus collie: Vocabulary in speech to young children. *Developmental Psychology, 19*(4), 602–609.

Bloom, P. (2002). *How children learn the meanings of words*. MIT press.

Brennan, S. E., & Clark, H. H. (1996). Conceptual pacts and lexical choice in conversation. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(6), 1482–1493.

Callanan, M. A. (1985). How parents label objects for young children: The role of input in the acquisition of category hierarchies. *Child Development, 56*(2), 508–523.

Callanan, M. A. (1989). Development of object categories and inclusion relations: Preschoolers' hypotheses about word meanings. *Developmental Psychology, 25*(2), 207–216.

Casenhiser, D. M. (2005). Children's resistance to homonymy: An experimental study of pseudohomonyms. *Journal of Child Language, 32*(2), 319–343.

Choe, J., & Papafragou, A. (2023). The acquisition of subordinate nouns as pragmatic inference. *Journal of Memory and Language, 132*, 104432.

Chomsky, N. (1959). Review of B. F. Skinner's verbal behavior. *Language, 35*, 26–58.

Clark, E. V., & Wong, A. D. W. (2002). Pragmatic directions about language use: Offers of words and relations. *Language in Society, 31*(2), 181–212.

Dautriche, I., & Chemla, E. (2016). What homophones say about words. *PLoS ONE, 11*(9), e0162176.

Dautriche, I., Chemla, E., & Christophe, A. (2016). Word learning: Homophony and the distribution of learning exemplars. *Language Learning and Development, 12*(3), 231–251.

Dautriche, I., Fibla, L., Fievet, A-C., & Christophe, A. (2018). Learning homophones in context: Easy cases are favored in the lexicon of natural languages. *Cognitive Psychology, 104*, 83-105.

Doherty, M. J. (2004). Children's difficulty in learning homonyms. *Journal of Child Language, 31*(1), 203–214.

Golinkoff, R. M., Shuff-Bailey, M., Olguin, R., & Ruan, W. (1995). Young children extend novel words at the basic level: Evidence for the principle of categorical scope. *Developmental Psychology, 31*(3), 494–507.

Grice, P. (1975). Logic and conversation. In P. Cole, & J. Morgan (Eds.), *Syntax and Semantics* (Volume 3). New York: Academic Press.

Mazzocco, M. M. (1997). Children's interpretations of homonyms: A developmental study. *Journal of Child Language, 24*(2), 441–467.

Quine, W. V. O. (1960). *Word and object*. Cambridge, MA: MIT Press.

Rosch, E., & Mervis, C. B. (1975). Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology, 7*(4), 573–605.

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

Shipley, E. F., Kuhn, I. F., & Madden, E. C. (1983). Mothers' use of superordinate category terms. *Journal of Child Language, 10*(3), 571–588.

Soh, C., & Yang, C. (2021). Memory constraints on cross situational word learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 43).

Spencer, J. P., Perone, S., Smith, L. B., & Samuelson, L. K. (2011). Learning words in space and time: Probing the mechanisms behind the suspicious-coincidence effect. *Psychological Science, 22*(8), 1049–1057.

Wang, F. H., & Trueswell, J. C. (2019). Spotting dalmatians: Children's ability to discover subordinate-level word meanings cross-situationally. *Cognitive Psychology, 114*, 101226.

Wang, F. H., & Trueswell, J. C. (2022). Being suspicious of suspicious coincidences: The case of learning subordinate word meanings. *Cognition, 224*, 105028.

Waxman, S. R. (1990). Linguistic biases and the establishment of conceptual hierarchies: Evidence from preschool children. *Cognitive Development, 5*, 123–150.

Waxman, S. R., Lynch, E. B., Casey, K. L., & Baer, L. (1997). Setters and Samoyeds: The emergence of subordinate level categories as a basis for inductive inference in preschool-age children. *Developmental Psychology, 33*(6), 1074–1090.

Waxman, S. R., Shipley, S., & Shepperson, B. (1991). Establishing new subcategories: The role of category labels and existing knowledge. *Child Development, 62*, 127–138.

Woodward, A. L., & Markman, E. M. (1998). Early word learning. In W. Damon (Ed.), *Handbook of child psychology: Vol. 2. Cognition, perception, and language* (pp. 371–420). John Wiley & Sons, Inc.

Xu, F., & Tenenbaum, J. B. (2007). Word learning as Bayesian inference. *Psychological Review, 114*(2), 245–272.

Yue, C. S., LaTourrette, A. S, Yang, C., & Trueswell, J. (2023). Memory as a computational constraint in cross-situational word learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 45).

Yurovsky, D., & Yu, C. (2008). Mutual exclusivity in cross-situational statistical learning. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 30).

Zehr, J., & Schwarz, F. (2018). PennController for Internet Based Experiments (IBEX).