The acquisition of subordinate nouns as pragmatic inference: Semantic alternatives modulate subordinate meanings

June Choe (yjchoe@sas.upenn.edu)
Department of Linguistics, University of Pennsylvania
Philadelphia, PA 19104 USA

Anna Papafragou (anna4@sas.upenn.edu)
Department of Linguistics, University of Pennsylvania
Philadelphia, PA 19104 USA

Abstract

Word learning is characterized by a bias for mapping meanings at the "basic"-level such as apple, as opposed to a subordinate-level like red apple (Markman, 1990). The fact that learners nevertheless acquire subordinate nouns has been attributed to properties of the referential world that co-occur with the word (e.g., Xu and Tanenbaum, 2007b; Spencer et al., 2011). However, learners may also make inferences about the informativity of labels as intentional linguistic acts. We investigated whether learners exploit information about semantic contrast to generalize word meanings beyond the basic level. Experiment 1 found that the introduction of a labelled alternative at the subordinate level (green apple) eliminated the basic-level bias. Experiment 2 found that the presence of the alternative exemplar without a label merely suppressed the bias. We propose that the acquisition of subordinate-level meanings is facilitated by expectations of informativity which allow learners to enter the relevant alternatives into consideration.

Keywords: language; acquisition; word learning; pragmatics

Introduction

A major aspect of word learning involves identifying the level of specificity encoded by word meanings. This is especially challenging for meanings that enter into a subset-superset relationship, such as 'dog' vs. 'poodle' (Quine, 1960). Evidence suggests that learners show a bias for mapping nouns to the so-called "basic"-level meaning (dog), as opposed to a narrower, subordinate-level meaning (e.g., poodle; Markman, 1990). While it has been reported that children as young as two-to-three years old can generalize word meanings beyond the basic level with carefully designed linguistic support that makes category information salient (Callanan, 1985; Gelman and Markman 1986, 1987; Waxman et al., 1991, 1997), others have argued that it is neither reasonable nor necessary to expect such richness of input in child-directed speech to be the primary mechanism for the acquisition of subordinate nouns (e.g., Xu and Tenenbaum, 2007b).

Thus, recent efforts have focused on learners' ability to generalize to subordinate-level meanings from seeing just a few instances of positive evidence for that conjecture. In particular, word learning models such as those based on Bayesian inference have been shown to capitalize on the sheer number of positive examples for acquiring subordinate-level meanings (e.g., Xu and Tenenbaum 2007a, 2007b; Lewis and Frank 2018). For a Bayesian learner, a bigger category is more likely to be used for any referential expression due to its larger size, so it is assigned a higher prior than a narrower category which is consistent with a smaller subset of referential expressions. But for the same reason, the likelihood of repeatedly seeing subordinate-level exemplars is much greater when assuming the subordinate-level meaning rather than the basic-level meaning. Thus, when the learner sees multiple exemplars consistent with a subordinate-level meaning, they capitalize on the "suspicious coincidence" of that arrangement to infer that the label most likely encodes the subordinate-level category as opposed to the basic-level category.

However, it has also been reported that the basic-level bias is largely modulated by the presence of other exemplars in the task. In a series of cross-situational word learning experiments, Wang and Trueswell (2019, 2022) found that adults and 3-to-5-year-old children generalized the meaning of a novel label to the basic-level even when the label exclusively co-occurred with exemplars consistent with a narrower subordinate-level meaning (e.g., dalmatians). In fact, the crucial determinant for generalizations to subordinate-level meanings was the simultaneous learning of a second label when that label was paired with other members from the same basic-level category (e.g., non-dalmatian dogs). Critically, this effect disappeared when the second label was paired with members from a different basic-level category (e.g., birds), suggesting that learners generate task-specific inferences about which category levels are being highlighted in an ostensive labelling event, independently of the information about the distribution of the target label.

In this study, we begin with the assumption that learners make pragmatically driven inferences about the hypothesis space over which possible word meanings are proposed and evaluated. Unlike accounts that frame the acquisition of subordinate-level nouns as a question of how various cues apparent in the physical world interact and converge on a specific concept, we ask under what discourse contexts learners *expect* to hear a label with a narrower meaning. Under this framing, the meaningful difference between basic- and subordinate-level categories is not in the inherent size of

the area that they carve up in the conceptual space. Rather, the distribution of basic- and subordinate-level labels is primarily governed by speaker intent, which makes it first and foremost a linguistic act.

More specifically, we propose that the distribution of basic- and subordinate-level labels co-occurring with an object follows naturally from the level of informativity intended by a speaker when discussing that object. Thus, in the case of basic vs. subordinate nouns, identifying the intended meaning would greatly benefit from the learner inferring the appropriate level of informativeness for a novel word from an ostensive labelling event. And when the situation presents many choices for labelling a referent, the label that corresponds to the narrowest possible category (subordinate-level) is often the most informative (Murphy and Brownell, 1985).

In the present study, we probe the nature of these pragmatic inferences by testing the role of semantic contrast in adult learners' basic- vs. subordinate-level generalization of novel words from single learning trials. We hypothesize that the rate of basic-level generalizations for an ostensive target label (e.g., 'mipen' paired with a red apple) would decrease if the target is immediately followed by a labelled semantic alternative at the subordinate level (e.g., 'kalmick' paired with a green apple), under the assumption that the presence of the alternative makes it clear that the more informative (subordinate-level) categories are relevant to the task (on the role of relevant alternatives, see Barner et al., 2011; Skordos and Papafragou, 2016). Additionally, we hypothesize that this effect of contrast is primarily linguistic, as opposed to merely conceptual, and should thus be stronger when the alternative is labelled rather than simply present but unlabelled (Clark, 1987, 1988, 1990). In two experiments, we test these hypotheses respectively.

Experiment 1

Participants Fifty-three adults participated in Experiment 1. Participants were recruited from the undergraduate subject pool at a university in the United States and on Prolific, a platform for online subject recruitment.

Materials and Procedure

Experiment 1 was hosted online on PCIbex (Zehr and Schwarz, 2018). There were ten trials: eight critical trials and two catch trials which tested for attention and color vision deficiency. Each trial was divided into the *learning phase* and the *testing phase*.

Images from eight distinct "semantic domains" (i.e., superordinate-level categories) corresponding to each critical trial were prepared. The domains were balanced in the number of natural (e.g., fruits) and artifact (e.g., cars) kinds. For each domain, a total of fourteen images of single exemplars were created. Twelve of these images (all set against a naturally occurring background) were used in the testing phase: two were from the target subordinate category (e.g., red apples), two from an alternative subordinate category (e.g., green apples), three other members from the same basic category (e.g., other apples), and five members from other basic-level categories within the same semantic domain (e.g., non-apple fruits). The remaining two images – one additional member from each of the two subordinate-level categories – showed the exemplar without a background and were used in the learning phase.

At the beginning of the experiment, a cartoon character, Sally, appeared on the center of the screen and introduced herself as a foreign language speaker. Before proceeding to the trials, Sally told participants that they would be learning words from her native language and that they should pay attention as she would ask questions about these words later. All communication from Sally was delivered in written form, appearing in a speech bubble.

The learning phase began with Sally on the screen by herself for one second, after which objects appeared at Sally's sides one-by-one. When the learning phase introduced two objects, one appeared to the left and then the other appeared to the right of Sally. When the learning phase only introduced one object, it appeared once to the left of Sally. Only one object was visible at any given time and each object stayed on screen for seven seconds while Sally labelled the object.

We manipulated the presence of a semantic alternative in two conditions in a within-participant design. In the *No Contrast* condition, only the target was shown and labelled (Figure 1a). In the *Contrast* condition, the target was followed by a semantic alternative at the subordinate-level with a different label (Figure 1a-b). After all object(s) had been presented, the learning phase concluded with Sally returning back to the upright position for one second.

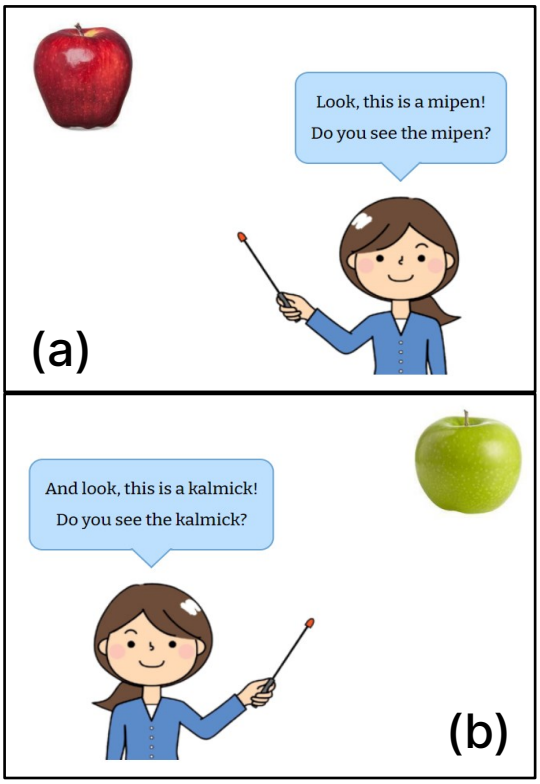

Figure 1: Presentation of the target (red apple, "mipen", top panel) and the alternative (green apple, "kalmick", bottom panel) in the learning phase of the *Contrast* condition in Experiment 1. In the *No Contrast* condition, only the target was shown and labelled, as in the top panel.

In the testing phase of each trial, eighteen images were presented in a 3-by-6 grid (Figure 2). Each image was placed inside a 150-pixel square cell with 15 pixels of row and column gaps. The grid included two matches to the target subordinate category, two matches to the alternative subordinate category, three other matches to the basic category, three matches to the superordinate category, and eight non-matches (i.e., members of other superordinate categories). The images were laid out in the grid in randomized order between design groups. Figure 2 shows an image grid from the test phase for the fruit domain, where participants are asked to find matches to the label paired with a red apple ("mipen") in the learning phase.

Figure 2: Images shown in the testing phase for a trial involving the fruit domain. Selections representing the basic-level generalization (all apples) are marked with a blue outer border and selections representing the target subordinate-level generalization (red apples) are marked with an additional inner yellow border.

Sally instructed the participants to select all matches for a novel label introduced in the learning phase. For example, if the target label was "mipen" paired with a red apple in the learning phase, Sally asked participants: "Do you see any other mipens below? Click on all mipens you see!". Participants interacted with the image grid by clicking one image at a time. The testing phase was not timed, and participants could freely select and unselect any images as many times as they wished, as long as at least one image was selected before proceeding. The final set of selections as well as the target and timestamp of individual clicks made in the testing phase were recorded for analysis.

Coding

The set of selections at test was coded into one of three categories: Subordinate, Basic, and Other. A response was coded as Subordinate if it reflected a narrow generalization to only the subordinate-level category (e.g., both red apples selected). A responses was coded as Basic if they included both subordinate target exemplars and also all other members from the basic-level category (e.g., all apples). Responses were coded as Other if the selections represented an incomplete subset of the Subordinate or Basic set, such as selecting only one of two subordinate targets (an incomplete Subordinate response) or both subordinate targets and only one or two out of three other basic-level exemplars (an incomplete Basic response). [1] Lastly, selections which included exemplars from other semantic domains (e.g., a planet after seeing a dog in the learning phase) were judged to reflect inattentiveness and excluded from analysis.

Results and Analysis

Data from three participants who answered incorrectly on the two catch trials were discarded. Additionally, 12 trial-level responses were excluded because the selections included exemplars from other semantic domains. 388 responses from the remaining 50 participants were entered into the analysis.

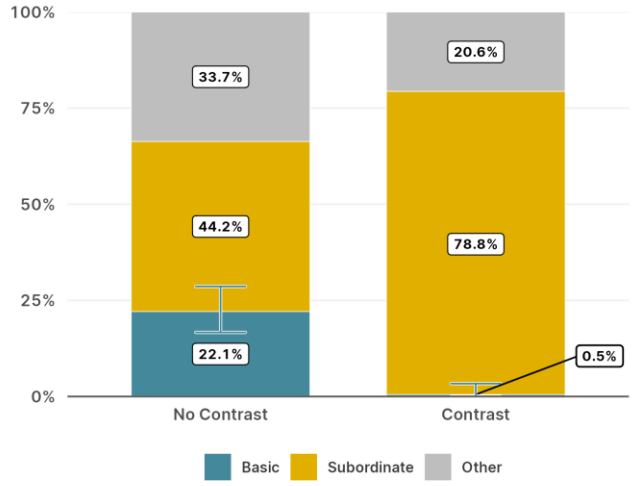

Figure 3: Responses at test for Experiment 1.

Following previous experimental work on the acquisition of subordinate nouns (e.g., Xu and Tenenbaum, 2007a, 2007b; Spencer et al., 2011; Lewis and Frank, 2018; Wang and Trueswell, 2019, 2022), we first conducted an analysis of the rate of Basic responses. We found a significant decrease in the proportion of Basic responses between the *Contrast* condition and the *No Contrast* condition ($p < 0.0001$, Chi-squared test[2]), with Basic responses decreasing from 22% to 0.5% when a semantic alternative was present with a different label. This difference was also reflected in the proportion of Subordinate responses, which increased from 44% to 79%.

Additionally, we observe two somewhat unexpected patterns. One is the low rate of Basic responses, both on the absolute scale and compared to Subordinate responses.

[1] The coding scheme also included Superordinate responses (e.g., all animals), but none were observed across both our experiments.

[2] Given the lack of item- and participant-level variation across conditions, a mixed-effects model was not appropriate for the data.

Empirically, the magnitude of the basic-level bias has been reported to vary widely, putting 22% on the low end but still comparable to that of other studies (e.g., Jenkins et al. (2015) reports 26%). At the same time, we suspect that a few experimental artifacts may be at play, such as the overall difficulty of making a Basic response (five images had to be selected, as opposed to two-to-three in most other studies).

Another related observation is the relatively high rate of Other responses, whose distribution differed substantially between conditions. Of the 67 Other responses in the *No Contrast* condition, two-thirds were incomplete subsets of the Basic response (i.e., a selection of both subordinate targets and some but not all of the five basic-level exemplars), partially explaining the overall low rate of Basic responses.[3] Interestingly, these responses are entirely absent among the 39 Other responses in the *Contrast* condition. Instead, over half of these were "mutually exclusive" responses (Gelman et al., 1989), which includes all basic-level members except the two from the alternative subordinate-level category (e.g., all apples except for the two green apples). While both may be construed as errors in generalizing to the basic-level category, their source differs.[4] The "incomplete Basic" responses are likely driven by the failure to identify the appropriate basic-level category to generalize to. In the case of "mutually exclusive" responses, however, the basic-level category is correctly identified but participants nevertheless pursue a narrower interpretation which excludes the semantic alternative. This is further investigated in Experiment 2.

Experiment 2

Participants

Ninety adults who did not previously participate in Experiment 1 were recruited on Prolific for Experiment 2.

Materials and Procedure

Experiment 2 was similar but sought to disentangle the effect of labelling from the mere presence of the alternative referent. The labelling of the semantic alternative was manipulated in two conditions. The *Labelled Alternative* condition replicated the *Contrast* condition of Experiment 1. In the *Unlabelled Alternative* condition, the alternative was present but not labelled: Sally simply drew attention to the object by saying "(And) look over here! Do you see this?" To guard against possible presentation effects, the order in which the target referent appeared relative to the alternative referent was counterbalanced (*target first* vs. *target second*).

Coding

Because participants saw both the target and the alternative in the learning phase for all critical trials, the coding scheme followed that of the *Contrast* condition in Experiment 1.

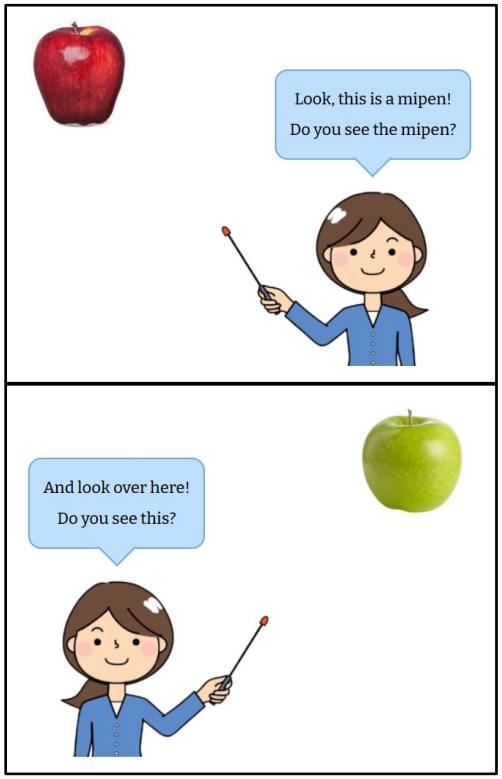

Figure 4: Presentation of the target (top panel) and the alternative (bottom panel) in the *Unlabelled Alternative* condition of Experiment 2. The *Labelled Alternative* condition is the same as Figure 1.

Results and Analysis

After applying the same filtering criteria as in Experiment 1, 669 responses from 86 participants entered the analysis.

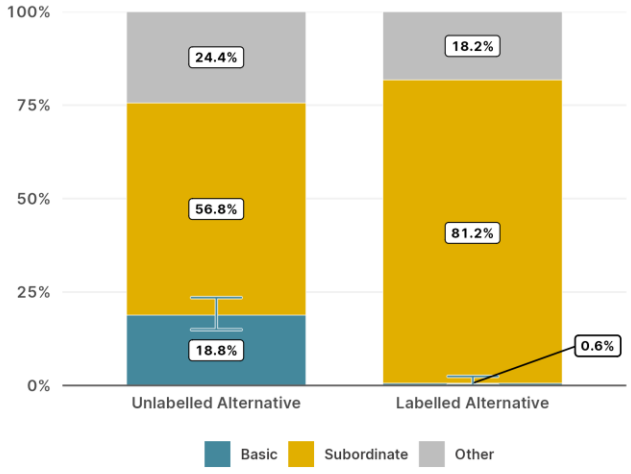

Figure 5: Responses at test for Experiment 2.

[3] In fact, under an alternative coding scheme which accounts for partial selections of the basic-level set (e.g., Lewis and Frank, 2018), the proportion of Basic response in the *No Contrast* condition increases to 40%, which is closer to the empirical average for the magnitude of the basic-level bias using this paradigm.

[4] Notably, none of the "incomplete Basic" selections in the *No Contrast* condition pattern like the "mutually exclusive" selections in the *Contrast* condition. For example, no participants selected all apples except for the two green apples after seeing a single red apple labelled in the learning phase.

We fitted a mixed-effects logistic regression model to the rate of Basic responses using the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages in R (R Core Team 2021). The model summary is shown in Table 1.[5]

Table 1: Mixed-effects logistic regression model fitted to Basic responses in Experiment 2.

	β (SE)	t	p
(Intercept)	-7.6 (1.7)	-4.5	<0.0001
Label	-2.7 (0.5)	-5.3	<0.0001
Order	2.5 (0.9)	2.9	0.0036

We find a significant main effect of Label (β = -2.7, SE = 0.5, $p < 0.0001$) such that there was an overall lower rate of Basic responses when the alternative was labelled compared to when it was simply present and unlabelled. The Order nuisance variable also reached significance ($\beta = 2.5$, $SE = 0.9$, $p < 0.01$), indicating that the basic-level interpretation of the target label was more likely when the target was shown first in the learning phase, before the alternative was introduced.

Similar to in Experiment 1, we find that this difference in Basic responses between the conditions is straightforwardly reflected in the rate of Subordinate responses, which increased from 57% to 81%. Compared to the *No Contrast* condition of Experiment 1, the *Unlabelled Alternative* condition shows lower Basic responses (19% vs. 22%) and higher Subordinate responses is higher (57% vs. 44%). We also find that the *Contrast* condition of Experiment 1 is successfully replicated by the *Labelled Alternative* condition, with similar rate of Basic (0.6% vs. 0.5%) and Subordinate (81% vs. 79%) responses.

Lastly, because the semantic alternative was always present across both conditions, we were able to compare the rate of "mutually exclusive" responses as observed in the *Contrast* condition of Experiment 1. These responses occurred in 9 out of 83 Other responses in the *Unlabelled Alternative* condition, and 34 out of 62 Other responses in the *Labelled Alternative* condition. While the mechanism driving the "mutually exclusive" responses is unclear, these responses pattern specifically with the labelling of the semantic alternative, not just its presence, suggesting that these "mutually exclusive" interpretations are specific to the processing of linguistic, as opposed to conceptual, contrast. We intend to further investigate the specificity of conjectures narrower than the basic-level category in future work.

General Discussion

This study investigated whether learners use semantic contrast to make subordinate-level generalizations. We proposed that the introduction of a semantic alternative at the subordinate level would make subordinate-level conjectures accessible to a learner who would otherwise default to a basic-level generalization for a novel label. Results show that the presence of a labelled alternative eliminated the basic-level bias during word learning with adult learners (Experiment 1); furthermore, the mere presence of the alternative referent without a label slightly suppressed the basic-level bias but failed to eliminate it (Experiment 2). These results are consistent with our hypothesis that learners disambiguate between subordinate- and basic-level meanings for a label by inferring the level of informativity expected from a labelling event (and not just from any type of contrast).

Beyond these findings, the data are broadly consistent with previous findings in the literature on the acquisition of subordinate nouns. The classic basic-level bias measured in the *No Contrast* condition of Experiment 1 has a magnitude that is small but still within the range of reported values from previous studies (see Jenkins et al., 2015, for a review). Additionally, our finding that Basic responses diminish in the presence of a semantic alternative at the subordinate-level is also consistent with results from cross-situational paradigms (Wang and Trueswell, 2019, 2022). In fact, we observe a strong effect of a labelled alternative even after just a single instance of the target label, without needing to expose the learner to multiple exemplars cross-situationally. Lastly, the presentation order effect in Experiment 2 is consistent with prevailing hypothesis-testing models of word learning, whereby a learner could initially posit a basic-level interpretation of the target label, which may or may not be revised upon encountering the semantic alternative (e.g., Medina et al., 2013; Stevens et al., 2017), especially if the labelled target is understood to be an anchor that introduces a basic-level category with respect to which the following (alterative) referent is to be interpreted (Waxman et al., 1991, 1997).

The general idea that contrast facilitates conjectures at the subordinate-level is certainly not new, given the known role of contrast in language acquisition (Markman, 1984, 1990; Clark, 1987, 1988, 1990). For example, Clark's (1987) Principle of Contrast states that "any difference in form in a language marks a difference in meaning"; and has been argued to allow learners to discover new mappings between concept and form. In the context of acquiring subordinate nouns, however, learners may need to engage in more complex inferential work beyond a simple search for some non-basic-level meaning that also happens to be consistent with what is observed in the referential scene. Here we proposed that the crucial task for the learner is to discover the *intended* dimension of contrast, where some dimensions of contrast are easier to access than others depending on the level of informativity that is assumed of the labelling event. Building on the previous literature on contrast, we argued that the specific meaning of contrast is only recognizable when the learner considers the correct (i.e., intended) alternative set. For the task of disambiguating between labels that form

[5] Model formula: Basic ~ Label + Order (1 + Order | Participant) + (1 | Item). Both predictors were sum coded. The interaction term could not be included because no Basic response was observed in the condition crossing *Labelled Alternative* and *target second*.

subset-superset relationships, we predicted that this choice of the alternative set is guided by the level of informativeness that the learner expects from the labelling event.

The role of alternatives in facilitating inferences about meanings that are normally difficult to access has been extensively explored elsewhere in the pragmatics literature. For example, we find a striking parallel to the study of scalar implicatures, specifically on the issue of young children's difficulty with generating the pragmatic interpretation of "some" as meaning 'some but not all' (Noveck, 2001; Papafragou and Musolino, 2003; among others). This difficulty has sometimes been argued to reflect children's limitations in processing (Guasti et al., 2005; Tieu et al., 2015), similar to how the task of learning subordinate-level categories has been treated by some in the literature on word learning (e.g., Ross and Murphy, 1996; Sloutsky et al., 2007; Sloutsky, 2010). For present purposes, we note that, when the stronger alternative "all" was introduced in a prior context, children were more successful in generating the pragmatic 'not-all' interpretation of "some" (Skordos and Papafragou, 2016; Barner et al., 2011). Furthermore, children arrived at the 'not-all' interpretation of "some" even when previously exposed to "none" (instead of "all"; Skordos and Papafragou, 2016), suggesting that establishing the *relevance* of the appropriate scale, not necessarily the accessibility of a particular stronger alternative, is key to computing scalar implicatures. In other words, despite the fact that "none" itself does not participate in the pragmatic meaning of "some", it nevertheless constrains children's search for the relevant scalar relationship by making the issue of quantification salient to the discourse context.

Likewise in our study, the presence of the semantic alternative, despite not contributing to the semantic or distributional profile of the target label itself, nevertheless constrained the interpretation of the target label to a subordinate-level category. Here, the successful inference depended on the learner's ability to recognize which scale is being highlighted in the task – the default, lateral contrast between mutually exclusive categories at the same level, or the harder-to-access vertical contrast between basic- and subordinate-level categories (see also Frank and Goodman, 2012, 2014). We expect these vertical inferences to be especially crucial for early word learning when the referential world rarely offers direct negative evidence to rule out the basic-level interpretation (Jenkins et al., 2015). Furthermore, although we have primarily focused on cases where the learner is mapping existing concepts to new labels, such as in the task of associating the concept 'dog' with the sound /dɒg/ (Gleitman and Trueswell, 2020), an informativity account could also be useful in explaining how *newly formed* concepts are mapped to new labels (Waxman and Markow, 1995; Lupyan et al. 2007; Lupyan and Thompson-Schill, 2012; LaTourrette and Waxman, 2019; Caplan, 2022).

Lastly, we return to a discussion of previous works that have reported children and adult's success in acquiring subordinate-level nouns through the sheer number of instances of positive evidence, such as via inferences about sampling (e.g., Xu and Tenenbaum, 2007b; Lewis and Frank, 2018) or attending to changes in various dimensions of presentation style (e.g., Spencer et al., 2011; Jenkins et al., 2015, 2021). How can our claims about the importance of informativity-based inferences be reconciled with these findings that seemingly jettison the need for the leaner to attend to the linguistic context beyond just the moment(s) when the target label is uttered?

We argue that these observations are not mutually exclusive. Inferences about speaker intent are pervasive, and arguably inevitable in word learning since labels do not simply describe the world (Grigoroglou and Papafragou, 2021). Rather, labels for objects invoke concepts and meanings that the speaker intends to communicate something about, and this information is often delivered to the listener by highlighting (in verbal and non-verbal ways) how the label contrasts with a *specific* set of *relevant* alternatives. In this sense, it is not the referential world that is suspicious or surprising in its presentation of multiple subordinate-level exemplars paired with a novel label. Instead, it may be that the only way for the learner to make sense of an interlocutor repeating the same label multiple times with what looks like members of the same kind of an object, is to interpret this act as an insistence to revise the current conjecture for the label, resulting in a move away from the initial basic-level generalization to a more conservative conjecture. In this sense, the effect of multiple instances of positive evidence may be indirect, which also explains its disappearance under more informative contexts that relieve the learner from computing speaker intent from such unreliable cues (Wang and Trueswell, 2019, 2022). In other words, the basic level is no doubt a conceptually privileged taxonomic category due to the salience of the kinds of perceptual features that define it, but it is also a linguistically privileged one because it lies at the default level of informativity assumed in conversations. In sum, properties of "the world" may indeed affect how learners generalize the meaning of a novel label, but only in so far as these cues contribute to the learner forming an expectation about what the speaker intends to convey.

Conclusion

In two word learning experiments, we showed that semantic alternatives facilitate mappings to subordinate-level meanings, and especially so when the alternative is labelled. Results suggest that learners can use linguistically marked contrast to reason about the level of informativity for a word's meaning that is expected from an ostensive labelling event. This sensitivity to the informativeness of an utterance offers a possible mechanism for the acquisition of subordinate nouns despite the apparent sparsity of evidence for subordinate-level meanings offered by the referential world.

References

Barner, D., Brooks, N., & Bale, A. (2011). Accessing the unsaid: The role of scalar alternatives in children's pragmatic inference. *Cognition*, *118*, 84-93.

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

Caplan, S. (2022). *The Immediacy of Linguistic Computation* [Doctoral dissertation, University of Pennsylvania].

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.

Clark, E. V. (1987). The Principle of Contrast: A constraint on language acquisition. In B. MacWhinney (Ed.), *Mechanisms of language acquisition.* Hillsdale, NJ: Lawrence Earlbaum Assoc.

Clark, E. V. (1988). On the logic of contrast. *Journal of Child Language*, *15(2),* 317–335.

Clark, E. V. (1990). On the pragmatics of contrast. *Journal of Child Language*, *17*, 417–431.

Frank, M. C., & Goodman, N. D. (2012). Predicting pragmatic reasoning in language games. *Science*, *336*, 998.

Frank, M. C., & Goodman, N. D. (2014). Inferring word meanings by assuming that speakers are informative. *Cognitive Psychology*, *75*, 80-96.

Gelman, S., A., & Markman, E. M. (1986). Categories and induction in young children. *Cognition, 23*, 183-209.

Gelman, S., A., & Markman, E. M. (1987). Young children's induction from natural kinds: the role of categories and appearances. *Child Development, 58*, 1432-1541.

Gelman, S. A., Wilcox, S. A., & Clark, E. V. (1989). Conceptual and lexical hierarchies in young children. *Cognitive Development*, *4*, 309-326.

Gleitman, L.R. & Trueswell, J.C. (2020). Easy Words: Reference Resolution in a Malevolent Referent World. *Topics in Cognitive Science*, *12*(1), 22-47.

Grigoroglou, M., & Papafragou, A. (2020). Contributions of pragmatics to word learning and interpretation. In A. Papafragou, J. Trueswell & L. Gleitman (Eds.), *The Oxford Handbook of the Mental Lexicon*. Oxford: Oxford University Press.

Guasti, M. T., Chierchia, G., Crain, S., Foppolo, F., Gualmini, A., & Meroni, L. (2005). Why children and adults sometimes (but not always) compute implicatures. *Language and Cognitive Processes*, *20*, 667, 696.

Jenkins, G. W., Samuelson, L. K., Smith, J. R., & Spencer, J. P. (2015). Non-Bayesian noun generalization in 3-to 5-year-old children: Probing the role of prior knowledge in the suspicious coincidence effect. *Cognitive Science*, *39*(2), 268–306.

Jenkins, G. W., Samuelson, L. K., Penny, W., & Spencer, J. P. (2021). Learning words in space and time: Contrasting models of the suspicious coincidence effect. *Cognition, 210*.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13), 1-26.

LaTourrette, A., & Waxman, S. R. (2019). A little labelling goes a long way: Semi-supervised learning in infancy. *Developmental Science*, *22*(1).

Lewis, M. L., & Frank, M. C. (2018). Still suspicious: the suspicious-coincidence effect revisited. *Psychological Science*, *29*(12), 2039–2047.

Lupyan, G., Rakison, D. H., & McClelland, J. L. (2007). Language is not just for talking: Redundant labels facilitate learning of novel categories. *Psychological Science*, *18*, 1077-1083.

Lupyan, G., & Thompson-Schill, S. L. (2012). The evocative power of words: activation of concepts by verbal and nonverbal means. *Journal of Experimental Psychology*, *141*, 170-186.

Markman, E. M. (1984). The acquisition and hierarchical organization of categories by children. In C. Sophian (Ed.), *Origins of cognitive skills: The 18th Annual Carnegie Symposium on Cognition*.

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

Medina, T.N., Snedeker, J., Trueswell, J.C., & Gleitman, L.R. (2011). How words can and cannot be learned by observation. *Proceedings of the National Academy of Sciences, 108*, 9014-9019.

Murphy, G. L., & Brownell, H. H. (1985). Category differentiation in object recognition: typicality constraints on the basic category advantage. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *11*(1), 70–84.

Noveck, I. (2001). When children are more logical than adults: experimental investigations of scalar implicatures. *Cognition*, *78*, 165-188.

Hillsdale , NJ: Lawrence Erlbaum Associates, pp. 376-406.

Papafragou, A. & Musolino, J. (2003). Scalar implicatures: experiments at the semantics-pragmatics interface. *Cognition*, *86*, 253-282.

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

R Core Team. (2021). R: A Language and Environment for Statistical Computing. https://www.R-project.org/

Ross, B. H., & Murphy, G. L. (1996). Category-based predictions: Influence of uncertainty and feature associations. *Journal of Experimental Psychology*, *22*, 736-753.

Skordos, D., & Papafragou, A. (2016). Children's derivation of scalar implicature: alternatives and relevance. *Cognition*, *153*, 6-18.

Sloutsky, V. M., Kloos, H., & Fisher, A. V. (2007). When looks are everything: Appearance similarity versus kind information in early induction. *Psychological Science*, *18*, 179-185.

Sloutsky, V. M. (2010). From perceptual categories to concepts: What develops? *Cognitive Science*, *34*, 1244-1286

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.

Stevens, J.S., Trueswell, J.C., Yang, C. & Gleitman, L.R. (2017). The pursuit of word meanings. *Cognition, 41*(4), 638-676.

Tieu, L., Romoli, J., Zhou, P., & Crain, S. (2015). Children's knowledge of free choice inferences and scalar implicatures. *Journal of Semantics*, 1-30.

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*.

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

Waxman, S. R., & Markow, D. B. (1995). Words as invitations to form categories: Evidence from 12- to 13-month-old infants. *Cognitive Psychology*, *29*(3), 257–302.

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.

Xu, F., & Tenenbaum, J. B. (2007a). Sensitivity to sampling in Bayesian word learning. *Developmental Science*, *10*(3), 288–297.

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

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