Journal of Experimental Child Psychology 104 (2009) 466–473

Contents lists available at ScienceDirect

Journal of Experimental Child Psychology

journal homepage: www.elsevier.com/locate/jecp

Brief Report

More is more: The relationship between vocabulary size and word extension

Emily E. Thom *, Catherine M. Sandhofer

Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90095, USA

ARTICLE INFO

Article history:
Received 5 January 2009
Revised 27 July 2009
Available online 3 September 2009

Keywords:
Language acquisition
Word learning
Vocabulary development
Category generalization
Color word learning
Fast mapping

ABSTRACT

This study experimentally tested the relationship between children's lexicon size and their ability to learn new words within the domain of color. We manipulated the size of 25 20-month-olds' color lexicons by training them with two, four, or six different color words over the course of eight training sessions. We subsequently tested children's ability to extend new color words to new instances. We found that training with a broader number of color words led to increased extension of new words. The results suggest that children's learning history predicts their ability to learn new words within domains.

Published by Elsevier Inc.

Introduction

Young children have been called "word-learning wizards" due to their impressive ability to map novel words to their intended referents on the basis of minimal exposure. Numerous experiments have documented this ability across multiple word types (e.g., Carey & Bartlett, 1978; Hirsh-Pasek & Golinkoff, 2006; Horst & Samuelson, 2008; O'Hanlon and Roberson, 2007). In these studies, children typically hear an unfamiliar object labeled with a novel label (e.g., "This is a dax one") and are asked to select other objects that share the same label (e.g., "Can you show me the other dax one?"). Word learning seems effortless and instantaneous in these word extension studies. For example, children seem to understand the meaning of novel object names after a single exposure (Heibeck & Markman, 1987). Furthermore, children appear to be so adept at word learning that they do not require objects to be explicitly labeled. In Carey and Bartlett's (1978) classic fast mapping study, 3- to 5-year-olds were asked to "bring me the chromium tray—not the blue one, the chromium one" in the presence of two trays: one blue and one olive-colored one. All of the children retrieved the intended tray.

* Corresponding author.
E-mail address: emilyt0623@ucla.edu (E.E. Thom).

0022-0965/$ - see front matter Published by Elsevier Inc.
doi:10.1016/j.jecp.2009.07.004

However, in other studies, the process of learning new words seems to be slow and difficult. For example, Rice (1980) trained 2- and 3-year-olds in three color words by showing the children multiple instances of each color, prompting them for the color labels, and providing corrective feedback when they provided incorrect labels. Rice found that most children required hundreds of trials to correctly and consistently label the three colors. Furthermore, children appear to require more experience to learn some word types than others. Heibeck and Markman (1987) found that children did not learn as much about color or texture words as they learned about shape words when given the same amount of exposure to each word type. Similarly, O'Hanlon and Roberson (2007) found that 3-year-olds learned shape terms more easily than they did color words under the same conditions. Furthermore, Gottfried and Tonks (1996) suggested that object labels are easier for children to learn than are color labels. Thus, children's word-learning wizardry does not appear to be uniform in all circumstances. Some words appear to be harder to learn than others (also see McMurray, 2007).

As a whole, these results are incongruous. Some studies suggest that children readily learn new words with minimal exposure, but other studies suggest that children require many trials to learn words. Moreover, across studies, children demonstrate markedly different rates of learning even within the same domain (color words). Why does learning words require only a few exposures in some studies but hundreds of exposures in other studies? One noteworthy difference between Carey and Bartlett's (1978) study and Rice's (1980) study was the size of children's color lexicon at the inception of each study. Children in Carey and Bartlett's study had an average of nine color words in their lexicon, whereas children in Rice's study had zero color words in their lexicon.

There are reasons to expect that within-domain lexicon size should affect learning. Training studies have demonstrated that extended practice in learning words within a particular domain facilitates learning of new words within the same domain (Gershkoff-Stowe & Hahn, 2007; Samuelson, 2002; Smith, Jones, Landau, Gershkoff-Stowe, & Samuelson, 2002). In these studies, the benefits of training generalize beyond the specific training set. For example, in Smith and colleagues' (2002) study, children who were trained in shape-based categories were more likely to categorize new, previously unseen objects by shape. In Samuelson's (2002) study, children easily named objects according to well-represented features in their training set but did not do so for less represented features.

In these studies, children received training in the same number of category instances. However, other studies have hinted that generalization may be influenced by the number of different instances experienced. Color production and comprehension appear to increase once children have accumulated four to five color words in their productive vocabularies (e.g., Soja, 1994). Similar findings have been reported for number words (e.g., Bloom & Wynn, 1997).

Thus, broadly construed, acquisition appears to accelerate as the number of words that children know increases. Furthermore, it appears that this relationship exists within specific categorical domains. However, these previous findings are correlational. The causality of the relationship between lexicon size and subsequent learning and extension has not been tested. Determining the nature of the relationship between lexicon size and extension is integral to our understanding of the major factors that influence word learning and extension as well as the nature of rapid vocabulary growth. Thus, in this study, we sought to build on and strengthen the findings of previous research by *experimentally* testing the relationship between within-domain lexicon size and extension. We did so by manipulating the size of children's color lexicon by teaching them two, four, or six color words and then testing their ability to extend new, previously untrained color words. If the number of words in children's productive vocabularies matters for learning new words, then we should expect extension to increase as the number of trained words increases.

Method

Participants

Participants were 12 male and 13 female children, approximately 20 months old (M = 20 months 3 days, range = 16 months 0 days to 26 months 14 days) at the beginning of the study. Participants were randomly assigned to one of three color word conditions; of the 25 children, 8 were in the two-word condition, 9 were in the four-word condition, and 8 were in the six-word condition. All children were learning English as their primary language. Age and vocabulary size were equated across conditions. No children in this study were known to be color-blind or had immediate family members who were color-blind. In addition, 3 children were excluded due to unwillingness to participate and 3 other children moved away before completing the study.

Children were prescreened for the number of color words in their productive vocabularies. Parents completed the MacArthur–Bates Communicative Developmental Inventory: Words and Sentences and a 21-item color word questionnaire that included the 12 colors used in this study. Families who marked one or zero color words were invited to participate.

Colors

There were 12 colors used in this experiment: black, blue, brown, green, gray, orange, pink, purple, red, teal, white, and yellow. All training, test, and distractor object colors were randomly selected from this pool. Thus, children in each condition received training in a different random training set.

Stimuli

Objects were small and three-dimensional (approximately 3 cubic inches in size) and made from a variety of materials. Half of the objects were familiar to participants (e.g., hats), and half of the objects were novel (e.g., clay hooks) (see Appendix).

Procedure

There were 11 sessions. Sessions were conducted in a quiet corner of each participant's classroom and lasted approximately 6–12 min.

Session 1: Pretests to screen for color comprehension and production

Comprehension. The comprehension pretest was always administered first. Three familiar objects matching in shape, but differing in color, were placed in front of the child. The child was asked to select the target color (e.g., "Can you give me the blue one?"). The experimenter provided neutral feedback (e.g., "Thank you!"). Children who selected the correct color received two additional trials with that color. If children selected the target color in two of the three trials (above what would be predicted by chance), they were credited as comprehending the word. The order of the color trials was randomized for each child.

Production. The child was asked to label the color of a single familiar object (e.g., "What color is this?"). Neutral feedback was provided if the child answered the prompt (e.g., "Thank you!"). The order of trials was randomized for each child.

Exclusion/inclusion criteria. Children who comprehended or produced more than one color word during pretest were excluded from the study. In the various conditions, 4 children in the two-word condition, 3 children in the four-word condition, and 2 children in the six-word condition comprehended one word at pretest. For these children, the comprehended word was included in the training set to preserve the lexicon size across the three conditions. None of the 25 children produced any color words during pretest.

Sessions 2–9: Training

Training took place over eight sessions. The procedure closely followed the methodology used by Smith and colleagues (2002). First, the experimenter introduced a target-colored object. While playing with it, the experimenter labeled its color five times with varied syntax (e.g., "This is a red one," "The color of this is red"). Next, the experimenter introduced another target-colored object of the same color. The dyad then played with both objects as the experimenter labeled the color of the objects five more times. Finally, the experimenter introduced a third object that differed in color. The experimenter drew attention to the fact that this object was not the target color and asked the child to put it away (e.g., "Oh! This one is not red. Let's put it away"). The color of this third object was never labeled. This procedure was repeated for each color in the child's training set. Beginning in the fourth session, children were encouraged to repeat the color words (e.g., "Can you say red?"). These prompts were counted as part of the 10 naming instances of the color trial. The order of the color trials was randomized across sessions. Each trial lasted approximately 2 min. Table 1 lists the approximate session length for each condition and training type.

Counterbalancing exposure. To control for differences in exposure, two training types were counterbalanced within each condition. Half of the children in each condition received training that equated exposure to individual colors. Over the course of training, these participants received 8 trials for each color in their training set (1 trial per session). Thus, children in the two-word condition received 16 trials (2 colors $\times$ 8 sessions), children in the four-word condition received 32 trials, and children in the six-word condition received 48 trials. The other half of the children received training that equated the total number of trials. All participants received 48 trials during the study (6 trials $\times$ 8 sessions). In the two-word condition, participants received 3 trials for both of their training colors at each session. In the four-word condition, participants received 1 trial for each of their training colors plus 1 additional trial for two of their colors at each session. The repeated colors were systematically varied across the 8 sessions to preserve the even distribution of trials for each color. Children in the six-word condition received 1 trial for each of their training colors at each session.

Testing

Session 10: Color posttests. The production posttest was identical to the production pretest. The comprehension posttest was identical to the comprehension pretest with the exception that words from the child's training set were tested twice.

Session 11: Extension test. The extension test examined children's ability to map a new, previously untrained color word to its referent within a single naming session. Each child was presented with four new color words. An object was placed in front of the child, and the experimenter drew attention to the object and labeled the color five times using varied syntax (e.g., "This is a purple one," "The color of this is purple"). Three new objects were then placed between the target object and the child. The experimenter then asked the child to select the target-colored object (e.g., "Can you show me the other purple one?").

Results

Participant/condition characteristics

Analyses were conducted to ensure that pretest scores were similar across the three conditions. A one-way analysis of variance (ANOVA) found no statistically significant differences in comprehension pretest scores across conditions, $F(2, 22) = 0.931$, *ns.* No child correctly produced color words during the pretest; thus, there were no significant differences in production pretest scores among conditions.

Table 1
Means of training span and session length across conditions and training types.

	Two-word condition	Four-word condition	Six-word condition
Mean training span (days)	29.5	23.0	32.0
Approximate session length (trials per color) (min)	4	8	12
Approximate session length (total trials) (min)	12	12	12

The span of the experiment varied from child to child. However, all children participated in at least two sessions per week over the course of a few weeks. Table 1 shows the mean span for each condition. A one-way ANOVA found no statistically significant differences in training span among conditions, $F(2, 22) = 1.037$, *ns.*

Training

Tallies of types and tokens were made for the total of number color words that children produced throughout training and testing. Tallies included labels that children (a) uttered spontaneously and (b) repeated when prompted by the experimenter. The *type* count reflected the number of different color word types children produced, and the *token* count reflected the overall number of instances in which children produced color words.

Types

The mean type counts were 2.38 ($SE = .68$) for the two-word condition, 3.22 ($SE = .62$) for four-word condition, and 5.50 ($SE = 1.21$) for the six-word condition. A one-way ANOVA confirmed a main effect of condition on type count, $F(2, 22) = 3.41$, $p = .05$. Post hoc analyses determined that children in the six-word condition had a greater type count than children in the two-word condition, Tukey's HSD, $p < .05$. It is not surprising that children trained in more words had a greater type count.

Tokens

The mean token counts were 13 ($SE = 5.76$) for the two-word condition, 14 ($SE = 4.03$) for the four-word condition, and 20 ($SE = 6.47$) for the six-word condition. A one-way ANOVA revealed that these differences were not statistically significant, $F(2, 22) = 0.477$, *ns.* Together, the token and type counts suggest that color words did enter into children's productive vocabularies during the course of the study.

Retention

To compare performance on the posttests across conditions, raw scores were transformed into proportions correct. On the comprehension posttest, the mean proportions correct were .31 ($SE = .11$) for children in the two-word condition, .33 ($SE = .07$) for children in the four-word condition, and .31 ($SE = .08$) for children in the six-word condition. A one-way ANOVA conducted on pretest to posttest difference scores found no statistically significant differences, $F(2, 22) = 1.71$, *ns.* In addition, a one-way ANOVA conducted on only posttest scores found no statistically significant differences among conditions, $F(2, 22) = 1.06$, *ns.* Furthermore, these scores did not differ from chance (.33) for the two-word condition, $t(7) = -0.22$, *ns,* for the four-word condition, $t(8) = 0.077$, *ns,* or for the six-word condition, $t(7) = 1.40$, *ns.* On the production posttest, the mean proportions correct were .13 ($SE = .07$) for the two-word condition, .25 ($SE = .06$) for the four-word condition, and .19 ($SE = .13$) for the six-word condition. A one-way ANOVA conducted only on posttest scores found no statistically significant differences among conditions, $F(2, 22) = 0.61$, *ns.* Pretest to posttest training scores were also computed for untrained color words. Two one-way ANOVAs found no statistically significant differences in comprehension or production posttest scores among conditions, suggesting that learning of untrained color words did not differ among conditions for comprehension, $F(2, 22) = 0.12$, *ns,* or for production, $F(2, 22) = 1.60$, *ns.*

Extension test: Color word conditions

The question of interest was whether the number of words in which children were trained affected their ability to extend words to new instances. Raw scores on the extension test were transformed into proportions correct. Children in the six-word condition had a mean score of .50

($SE = .07$), children in the four-word condition had a mean score of .33 ($SE = .06$), and children in the two-word condition had a mean score of .25, ($SE = .07$). A one-way ANOVA found that these scores were significantly different, $F(2, 22) = 3.81$, $p < .05$. Post hoc analyses revealed that children performed significantly better in the six-word condition than in the two-word condition, Tukey's HSD, $p < .05$.

One-sample *t* tests were conducted to compare performance in each condition with chance. If children were responding randomly, then they would be expected to have an average proportion correct of .33. Results showed that children in the six-word condition performed better than expected by chance, $t(7) = 2.54$, $p < .05$. Children's performance in the two- and four-word conditions did not differ significantly from chance for the two-word condition $t(7) = -1.20$, *ns*, or for the four-word condition, $t(8) = 0.06$, *ns*.

Children's training span (i.e., the number of days they spent in training and testing) was sorted into two categories based on the median values. An independent samples *t* test did not find statistically significant differences between the two groups, suggesting that children who were in training for shorter or longer periods of time did not differ on the extension test, $t(23) = -.25$, *ns*. An independent samples *t* test found that, across conditions, performance did not differ significantly between training types, $t(23) = -1.06$, *ns*.

Finally, we asked whether children who were more successful in comprehending or producing the trained color words performed better on the extension test. Pearson product moment correlations were calculated between scores on the comprehension and production tests and extension test scores. Neither test was statistically significant, suggesting that retention of trained words on the posttest was unrelated to performance on the extension test. In addition, a one-way ANOVA with posttest production scores entered as a covariate was not significant, $F(3, 21) = 1.29$, *ns*. However, a one-way ANOVA with posttest comprehension scores entered as a covariate was marginally significant, $F(3, 21) = 2.55$, $p < .10$, $r^2 = .267$. This suggests that scores on the comprehension posttest may have accounted for some of the variability in performance among conditions.

Discussion

Our study sought to understand why word learning appears fast and effortless in some circumstances and slow and effortful in others. We hypothesized that vocabulary size within a domain was related to subsequent word learning within that same domain. To test this idea, we experimentally manipulated the number of labeled instances of color words to which children were exposed and subsequently examined their ability to rapidly learn new, previously untrained color words. We found that children who were trained in more words were better at extending new words than children who were trained in fewer words. Our results are consistent with the idea that experience in labeling objects within a domain leads to increased extension of novel words within the same domain. This study cannot disambiguate whether increased extension resulted from experience in labeling distinct category instances or from overall exposure to labeling particular objects or properties within a category. Nevertheless, our results provide an important first step in establishing the causal link between experience in labeling and word learning and extension.

Results of this study shed new light on previous findings demonstrating that vocabulary size is related to rapid word learning and extension within particular domains. For example, recent studies found that training in certain word types led to sharp increases in children's productive vocabulary outside of the laboratory (Samuelson, 2002; Smith et al., 2002). However, the vocabulary growth displayed different patterns; in Smith and colleagues' (2002) study children's vocabularies for object labels increased, whereas in Samuelson's (2002) study children's overall vocabulary increased. Our results suggest that this difference can be attributed to the nature of training; children in Smith and colleagues' study were trained exclusively with object labels, whereas children in Samuelson's were trained in words from multiple domains. The results of the current study suggest that rapid vocabulary growth occurs in domains where children have had experience in learning words and, by extension, that vocabulary growth may be slower in domains where children have had less experience. This may also account for the different rates of learning across domains (Braisby & Dockrell, 1999; Heibeck & Markman, 1987) and even within the same domain (Carey & Bartlett, 1978; Rice, 1980).

Interestingly, our results suggest that explicit retention might *not* be necessary for increased extension to occur. Recall that we found differences in extension of new words among conditions despite the fact that explicit retention of the trained words was somewhat limited.

This finding has two important implications. First, it suggests that experience in labeling, and not explicit retention of words, leads to increased extension. However, it does not suggest that memory for previously labeled category instances is unimportant for extension. The fact that comprehension was partially related to extension suggests that children had some implicit memory for the trained color words. A test of recognition more sensitive than our current comprehension test might have better tapped children's memory for the words and revealed a stronger relationship between comprehension and extension. Because color words tend to be acquired slowly, a longer period of training might have also led to increased performance on the comprehension posttest. Indeed, because comprehension often precedes production, a longer period of training might have also led to increased performance on the production posttest.

The second implication of this finding is that vocabulary retention—lexicon size—may be a proxy measure of experience in labeling. For example, children with larger lexicons are likely to have had more experience in labeling and, as a result, are better at extending new labels. This also suggests that when category extension is the desired outcome, broad learning may be more beneficial than deep learning; that is, learning less about a greater number of different category instances may be more beneficial than learning more about a fewer number of instances.

Why should labeling and not just prior experience matter?

The act of labeling the category instances appears to be key to our findings (see Horst & Samuelson, 2008, for a discussion of factors that promote word referent mappings). From birth, children experience many examples of colors on a daily basis. If simply seeing instances of color without experiencing them as labeled properties mattered for extension, then all of our participants should have performed well on the extension test. But this was not the case in our study. Moreover, we would expect young children to learn and extend color words quite easily in the real world. However, previous research has shown that color word learning can be quite difficult for young children.

Learning the labels of multiple examples of a category may promote selective attention to features relevant to the higher order category and, conversely, may detract attention from irrelevant features. Moreover, the relevance of features should be highlighted and reinforced as the number of experienced instances increases. Increased attention to category-relevant features should narrow the pool of possible referents and streamline acquisition and extension of new labels. Indeed, research has shown that attention to category-relevant features is important for category generalization and word learning and that it increases as a result of experience with members of the category (O'Hanlon and Roberson, 2006, Experiment 3; Smith et al., 2002).

Thus, the more experience children have in learning the labels within a particular category, the more fine-tuned their attention is to category-relevant features and the more readily they should be able to learn and extend new labels. This would account for why rapid word learning appears to develop at different rates across domains. Because different features matter for different categories, children must tune their attention to the features that matter for each category. To do so, children must accumulate enough experienced instances within each category—instances of color to tune attention to color, instances of shape-based object categories to tune attention to shape, and so forth. Children may accumulate instances from each domain at different rates, causing rapid word learning to develop at different rates across domains. This suggests that children's proclivity for fast mapping within a domain should be predictable by the number of words they already know within that domain and that categories in which the instances themselves may be easier to accumulate should be learned earliest. In other words, the nature of children's previous experience in learning words dictates the nature of their vocabulary growth.

Appendix. Familiar and novel objects

Familiar objects	Novel objects
Baskets	“Hexagons”
Blocks	“Tops”
Boxes	“Figure 8s”
Eggs	“Pom-poms”
Bracelets	“Spirals”
Hats	“Clay squares”
Bean bags	“Hooks”
Feet	“Squares”

References

Bloom, P., & Wynn, K. (1997). Linguistic cues in the acquisition of number words. *Journal of Child Language, 24*, 511–533.

Braisby, N., & Dockrell, J. (1999). Why is colour naming difficult? *Journal of Child Language, 26*, 23–47.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Proceedings of the Stanford Child Language Conference, 15*, 17–29.

Gershkoff-Stowe, L., & Hahn, E. R. (2007). Fast mapping skills in the developing lexicon. *Journal of Speech, Language, and Hearing Research, 50*, 682–696.

Gottfried, G. M., & Tonks, S. J. M. (1996). Specifying the relation between novel and known: Input affects the acquisition of novel color terms. *Child Development, 67*, 850–866.

Heibeck, T. H., & Markman, E. M. (1987). Word learning in children: An examination of fast mapping. *Child Development, 58*, 1021–1034.

Hirsh-Pasek, K., & Golinkoff, R. M. (2006). *Action meets word*. New York: Oxford University Press.

Horst, J. S., & Samuelson, L. K. (2008). Fast mapping but poor retention in 24-month-old infants. *Infancy, 13*, 128–157.

McMurray, B. (2007). Defusing the childhood vocabulary explosion. *Science, 317*, 631.

O’Hanlon, C. G., & Roberson, D. (2006). Learning in context: Linguistic and attentional constraints on children’s color term learning. *Journal of Experimental Child Psychology, 94*, 275–300.

O’Hanlon, C. G., & Roberson, D. (2007). What constrains children’s learning of novel shape terms? *Journal of Experimental Child Psychology, 97*, 138–148.

Rice, N. (1980). *Cognition to language*. Baltimore, MD: University Park Press.

Samuelson, L. K. (2002). Statistical regularities in vocabulary guide language acquisition in connectionist models and 15–20-month-olds. *Developmental Psychology, 38*, 1016–1037.

Smith, L. B., Jones, S. S., Landau, B, Gershkoff-Stowe, L., & Samuelson, L. (2002). Object name learning provides on-the-job training for attention. *Psychological Science, 13*, 13–19.

Soja, N. N. (1994). Young children’s concept of color and its relation to acquisition of color words. *Child Development, 65*, 918–937.