Understanding Word Learning by Preschool Children: Insights from Multiple Tasks, Stimulus Characteristics, and Error Analysis

Jill R. Hoover and Holly L. Storkel

University of Kansas

*This project was supported by NIH grant DC004781, DC006545, DC000052.

Address for Correspondence: Holly L. Storkel, University of Kansas, Speech-Language-Hearing Department, Dole Human Development Center, 1000 Sunnyside Ave., Room 3001, Lawrence, KS 66045-7555, hstorkel@ku.edu, www.ku.edu/~wrdlrng.

Understanding Word Learning by Preschool Children: Insights from Multiple Tasks, Stimulus Characteristics, and Error Analysis

Lexical acquisition occurs rapidly and with relative ease in typically developing children. Children can form an initial representation of a novel word following only a single exposure (e.g., Dollaghan, 1985). The number of root words known increases from approximately 3,500 at the beginning of kindergarten to about 6,000 at the end of second grade (Biemiller, 2001; Biemiller & Slonin, 2001). While there is clear evidence that children learn words rapidly, what remains to be discovered is how children are able to accomplish this complex task so effortlessly. To learn a word, a child must store in memory two types of representations, semantic and word form, as well as the links between them (see Gupta this issue). The *semantic representation* refers to the meaningful information regarding a referent (e.g., "a toothed strip of rigid material for arranging the hair" for "comb"). A *word form representation* refers to the combination of sounds as a whole sequence and constitutes the word form (e.g., /koʊm/ for "comb;" Vitevitch & Luce, 1999). Moreover, children may capitalize on known phonological representations to support word learning. A *phonological representation* refers to the individual sounds of the language (e.g., /k/ /oʊ/ /m/ in "comb"). Gupta (this issue) formalizes these levels of word learning in his functional framework. Here we extend the framework to children, examining groups of typical learners and learners with phonological delays. In addition, we illustrate how using multiple tasks, manipulating the stimulus characteristics of the words to be learned, and analyzing errors can provide a window into the formation of each type of representation.

Multiple Tasks

A series of tasks can be used to examine the formation of phonological, word form, and semantic representations during word learning by preschool children (Storkel, 2001; 2003; 2004;

Storkel & Maekawa, in press). For example, children are exposed to nonwords paired with novel objects in a story narrative. Learning of the novel words is repeatedly tested as the story unfolds through three tasks: referent identification, form identification, and picture naming. In *referent identification*, children are shown three novel objects from the story. A nonword is presented auditorially and the child is asked to select the referent of the nonword by pointing to one of the three pictures. The presentation of the nonword may activate the child's word form representation. The word form representation then activates the appropriate semantic representation via the receptive link, allowing the child to select the correct picture. This task taps the receptive link and the semantic representation itself.

In *form identification*, children are shown a picture of one novel object followed by the sequential presentation of three auditory nonwords. In some studies, nonwords are paired with a colored square, whereas in other studies nonwords are paired with a picture of the target object (i.e., three pictures of the same object). The child's task is to select the correct nonword name of the picture. Presentation of the picture presumably activates the child's semantic representation which activates the word form representation via the expressive link, allowing the child to correctly select the target nonword. This task provides a window into the status of the expressive link and the word form representation.

Finally, *picture naming* requires children to spontaneously produce a trained nonword upon presentation of the corresponding referent. This task, like form identification, is thought to activate the child's semantic representation, in turn activating the word form representation via the expressive link. However, picture naming provides more detailed information about the child's word form representation than does form identification because a complete and accurate representation of the word form is needed to correctly produce it, whereas an incomplete representation could be sufficient to support correct selection of the appropriate nonword during form identification.

Stimulus Characteristics

In addition to the use of multiple tasks, the phonological and word form characteristics of nonword stimuli, namely phonotactic probability and neighborhood density, can be manipulated (Storkel, 2001; 2003; 2004; Storkel & Maekawa, in press). *Phonotactic probability* is a phonological characteristic and refers to the likelihood of occurrence of a sound sequence in a language (Vitevitch & Luce, 2004). Words can be composed of either common sound sequences or rare sound sequences. Common sound sequences (e.g., "sit") contain individual sounds that occur in many other words in the same position and sound sequences that co-occur in many other words of the language. On the other hand, words composed of rare sound sequences (e.g., "these") contain combinations of sounds that co-occur infrequently.

Neighborhood density is a word form characteristic and refers to the number of words that are phonologically similar to a given word, differing by only one phoneme (Luce & Pisoni, 1998). Words can reside in dense neighborhoods, having many neighbors (e.g., neighbors of *sit* include *bit, hit, sat, seat, sip*, and many more), or sparse neighborhoods, having few neighbors (e.g., neighbors of *these* include *cheese, peas, fees, those, ease*, and a few others).

Common sound sequences tend to have many neighbors and rare sound sequences tend to have few neighbors. Thus, past studies have typically manipulated these variables in tandem. Manipulation of these characteristics provides evidence of how known phonological and word form representations influence the creation of new word form and semantic representations during word learning. That is, word learning depends on the other sounds and words that the child already knows, and how the new word relates to these existing words may influence the speed of learning.

Error Analysis

Another way to assess the formation of underling representations is by analyzing the errors made by children across word learning tasks (Storkel, 2001; 2003; 2004; Storkel & Maekawa, in press). Error analysis is based on the hypothesis that knowledge is gradient, rather than all-or-none (Capone & McGregor, in press). Therefore, when a child fails to select the correct answer in a particular task, s/he may have partial knowledge (i.e., an incomplete representation) or no knowledge (i.e., no representation) of the new word. For example, in the referent identification task, the pictures presented to the child aid in determining the status of the semantic representation. One picture is the target object (e.g., a medium-sized purple pet with big ears), one is a semantically related object (e.g., a small green pet with antennae), and one is a semantically unrelated object (e.g., a machine that dispenses red candy balls). If the child selects the semantically related object, then this suggests that the child has only learned the general semantic category of the object, indicating a holistic representation. In contrast, if the child selects the semantically unrelated object, then the child's semantic representation may be nonexistent, highly impoverished or inaccurate because the child has not yet stored even basic semantic information about the referent of the target nonword.

Turning to form identification and picture naming, errors reveal the status of both the target and the substituted nonword. For example, if the child is shown a picture of one pet (e.g., a medium-sized purple pet with big ears named /jeɪp/) but responds with the name of the other pet (e.g., a small green pet with antennae named / koʊf/), then this would be a semantic error. In this case, the child has a holistic representation of the target referent and has only associated one word form representation (e.g., /koʊf/) with it. Thus, the word form representation of the target word may be holistic or impoverished and/or lack an expressive link, whereas the word form representation of the substituted word would be intact and have an expressive link to the holistic semantic representation. In contrast, if the child sees one pet (e.g., a medium-sized purple pet with big ears named /jeɪp/) but responds with the name of an unrelated object (e.g., a machine that dispenses red candy balls named /mɔɪd/), then this would be an unrelated error. Here, the child may have an impoverished semantic or word form representation of the target. The substituted word would be viewed as having an impoverished semantic representation and/or expressive link, but an intact word form representation because the child is able to produce the nonword.

Evidence from Typically Developing Preschool Children (Storkel, 2001)

Storkel (2001) provides an example of how multiple tasks, manipulation of stimulus characteristics, and error analyses can reveal the complexities of word learning. Storkel (2001) examined noun learning by presenting preschool children with eight nonword-novel referent pairs embedded in the context of a children's story. Learning was tracked periodically using the three previously described tasks. In addition, the characteristics of the nonwords to be learned were manipulated with half of the nonwords being common/dense sound sequences and half being rare/sparse sound sequences. Errors were analyzed in each task to determine which representations were vulnerable to failure during learning.

Multiple tasks and stimulus characteristics. Analysis of correct responses in the referent identification task showed that referents of common/dense sound sequences were identified more accurately than those of rare/sparse sound sequences. Therefore, phonotactic probability/neighborhood density affected the formation of semantic representations and receptive links. Analysis of the form identification data showed that common/dense sound sequences were identified more accurately than rare/sparse sound sequences. Phonotactic probability/neighborhood density appeared to influence the formation of a word form representation and an expressive link. Finally, on the picture naming task, children named common/dense sound sequences more accurately than rare/sparse sound sequences, proving further support that phonotactic probability/neighborhood density influenced word form representations. Therefore, phonotactic probability/neighborhood density appeared to have a global effect on word learning. Manipulation of stimulus characteristics lead to the additional insight that known phonological and word form representations influence how children learn words. Specifically, novel words that are similar to many other known words (i.e., common/dense) are learned more readily than those similar to few other known words (i.e., rare/sparse).

Error analysis by task and stimulus characteristic. In addition to the analysis of the correct responses across tasks, the errors made by children in each task were analyzed as another method of determining how phonotactic probability/neighborhood density influenced the formation of word form and semantic representations. On the referent identification task, there were more semantic errors made for common/dense than rare/sparse sound sequences. This suggests that when the target referent was similar to many other words, children were able to glean the general semantic features of the referent. In complement, children made more unrelated errors for rare/sparse than common/dense sound sequences during referent identification. Thus, when nonwords were similar to few other words, children were unable to retain even the most general semantic features. Taken together, nonwords composed of rare/sparse sound sequences were more vulnerable to the formation of weak semantic representations than nonwords composed of common/dense sound sequences.

The results of the error analysis for the form identification task mirrored those of the picture naming task, therefore only the picture naming errors are described here. When a child saw the referent of a rare/sparse sound sequence s/he tended to respond with the common sound sequence name of the semantically related referent. This shows that the child was able to create a holistic semantic representation of a rare/sparse nonword by retaining the feature of semantic category membership but was more apt to create an expressive link to the word form representation of the common/dense nonword rather than the target rare/sparse nonword. Analysis of the unrelated errors on the picture naming task suggested that regardless of the phonotactic probability of the target nonword, children tended to produce a word that was composed of a rare/sparse rather than a common/dense sound sequence. This suggests that children may have an accurate and detailed word form representation of rare/sparse nonwords but lack a correct expressive link. Therefore, the creation of an expressive link for rare/sparse sound sequences appeared especially vulnerable to failure.

Evidence from Clinical Populations (Storkel, 2004)

This approach from Storkel (2001) also has been applied to clinical populations of children, namely children with phonological delays (PD). Given the nature of their impairment, manipulation of phonological and word form characteristics may not lead to the same effects on word learning in children with PD as compared to their typical peers. Children with PD may not rely on phonological or word form representations in the same way as their typically developing peers when learning new words.

Storkel (2004) manipulated the phonotactic probability/neighborhood density of the nonwords and analyzed both correct and error responses. Multiple tasks were used but the picture naming results were the most revealing. The results showed that children with PD learned a similar number of words as their typical peers but different types. Specifically, children with PD showed a rare/sparse advantage when learning new words. This is the opposite pattern from typically developing children. Moreover, error analyses revealed that when children with PD were shown the picture referent of a common/dense sound sequence, they primarily responded with an unrelated rare/sparse sound sequences. Likewise, children with PD infrequently produced common/dense sound sequences. Thus, children with PD learned different words than typically developing children because of difficulty creating an expressive link and a word form representation of common/dense sound sequences. Because of their weak phonological representations, children with PD may not have been able to differentiate common/dense novel words from other phonologically similar known words. This may have lead to confusion between the novel word and other known words, impeding acquisition.

Clinical Implications

The results of Storkel (2001; 2004) highlight a variety of methods that can be used to gain knowledge of the formation of word form and semantic representations in children with typical language and in those with PD. These results suggest that word learning abilities may differ between these two populations. Importantly, children with PD scored within normal limits on standardized vocabulary tests, suggesting that these tests may be insensitive to differences in word learning (Dollaghan & Campbell, 1998). Current clinical diagnostic methods only allow one to determine the quantity of known words, rather than the representations that are vulnerable to failure during word learning. This is particularly important when planning treatment so that intervention can be directed towards the child's critical weakness. Assessment may be enhanced by incorporating additional diagnostic methods previously used solely in research settings. For example, Perdue & Storkel (2004) have examined the stimulus characteristics of several standardized vocabulary tests and have used this information to create a specially designed vocabulary probe that might allow identification of the types of words that are difficult for children to learn. Furthermore, tasks that are more sensitive to the process of word learning are being examined. Specifically, we are manipulating stimulus characteristics of items in a complex working memory task (Hoover & Storkel, 2005). The use of working memory measures has been found to be beneficial over standardized vocabulary tests, given their ability to detect differences in the process of word learning (Dollaghan & Campbell, 1998) and to reduce cultural bias (Rodekohr & Haynes, 2001). Clinicians may need to consider supplementing standard diagnostic practices with some of these newer techniques. To begin, clinicians may explore the manipulation of stimulus characteristics at the following websites: (1) phonotactic probability http://www.people.ku.edu/%7Emvitevit/PhonoProbHome.html; (2) neighborhood density http://128.252.27.56/neighborhood/Home.asp.

References

Biemiller, A. (2001). Teaching vocabulary: Early, direct, and sequential. *American Educator, 25,* 25-28.

Biemiller, A. & Slonin, N. (2001). Estimating root word vocabulary growth in normative and advantaged populations: Evidence for a common sequence of vocabulary acquisition. *Journal of Educational Psychology, 93,* 498-520.

Capone, N.C. & McGregor, K.K. (in press). The effect of semantic representation on toddlers' word retrieval. *Journal of Speech, Language, and Hearing Research*.

Dollaghan, C.A. (1985). Child meets word: "fast mapping" in preschool children. *Journal of Speech and Hearing Research, 28,* 449-454.

Dollaghan, C.A & Campbell, T.F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research, 41,* 1136-1146.

Gupta, P. (this issue). What's in a word? A functional analysis of word learning. *Perspectives on Language Learning and Education.*

Hoover, J.R. & Storkel, H.L. (2005, June). The assessment of word learning in children with phonological disorders: Nonword language processing task. Poster session presented at the annual meeting of the Symposium for Research and Child Language Disorders, Madison, WI.

Luce, P.A. & Pisoni, D.B. (1998). Recognizing spoken words: The neighbourhood activation model. *Ear and Hearing, 19,* 1-36.

Perdue, A.N. & Storkel, H.L. (2004, June). Phonotactic probability and the Peabody Picture Vocabulary Test – 3. Poster session presented at the annual meeting of the Symposium for Research and Child Language Disorders, Madison, WI.

Rodekohr, R.K. & Haynes, W.O. (2001). Differentiating dialect from disorder: A comparison of two processing tasks and a standardized language test. *Journal of Communication Disorders, 34,* 255-272.

Storkel, H.L. (2001). Learning new words: Phonotactic probability in language development. *Journal of Speech, Language, and Hearing Research, 44,* 1321-1337.

Storkel, H.L. (2003). Learning new words II: Phonotactic probability in verb learning. *Journal of Speech, Language, and Hearing Research, 46,* 1312-1323.

Storkel, H.L. (2004). The emerging lexicon of children with phonological delays: Phonotactic constraints and probability in acquisition. *Journal of Speech, Language, and Hearing Research, 47,* 1194-1212

Storkel, H.L. & Maekawa, J. (in press). A comparison of homonym and novel word learning: The role of phonotactic probability and word frequency. *Journal of Child Language.*

Vitevitch, M.S. & Luce, P.A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language, 40,* 374-408.

Vitevitch, M.S. & Luce, P.A. (2004). A web-based interface to calculate words and nonwords in English. *Behavior-Research-Methods,-Instruments-and-Computers, 36,* 481-487.

Questions

1. What aspect of language does phonotactic probability represent?
   a) morphology
   b) syntax
   c) phonology
   d) semantics

2. Which of the following patterns is true regarding the effect of phonotactic probability/neighborhood density on word learning in children with typical language development?
   a) Common/dense sound sequences are learned more easily than rare/sparse sound sequences in all tasks
   b) Rare/sparse sound sequences are learned more easily than common/dense sound sequences in all tasks
   c) Common/dense sound sequences and rare/sparse sound sequences are learned at the same rate in all tasks
   d) The effect of phonotactic probability varies by task

3. Which of the following patterns is true regarding the effect of phonotactic probability/neighborhood density on word learning in children with phonological delays?
   a) Children with phonological delays learn new words the same way as typically developing children
   b) Common/dense sound sequences are learned more easily than rare/sparse sound sequences in all tasks
   c) Rare/sparse sound sequences are learned more easily than common/dense sound sequences in all tasks
   d) Common/dense sound sequences and rare/sparse sound sequences are learned at the same rate in all tasks

4. Analysis of unrelated errors in referent identification allows one to assess the formation of:
a) phonological representations
b) both lexical and semantic representations
c) lexical representations
d) semantic representations

5. Standardized vocabulary tests can be used to assess:
a) The quantity of words known by children
b) The types of words known by children
c) How children learn new words
d) Phonological, lexical, and semantic representations

Key

1. c
2. a
3. c
4. d
5. a