

Who's afraid of similarity? Effects of phonological and semantic similarity on lexical acquisition.

Sarah Devi Sahni (sdsahni@gmail.com)

University of Wisconsin – Madison, Department of Psychology

Abstract

Children are sensitive to statistical regularities in speech and likely use these regularities when learning their native language. A central goal of current research is to understand which statistical regularities support different aspects of language acquisition and processing. In the current work we explore phonological and semantic similarity effects on early lexical acquisition. Using a computational model, behavioral findings from word learning studies are simulated and explored. With this model we demonstrate that acquisition can be facilitated by the distinctiveness of individual lexical mappings.

Introduction

Language acquisition research has robustly shown that children are sensitive to statistical regularities in speech, and utilize these regularities when learning their native language (for a review see Saffran & Sahni, in press). A central goal of current research in language acquisition is to understand which statistical regularities support different aspects of language acquisition and processing. Research on adult language processing has revealed that statistical regularities across words can affect lexical access and recognition (Dahan & Magnuson, 2006). Much of this work has examined effects of phonological similarity. Nevertheless, researchers have also examined the effects of semantic similarity along with phonological similarity (e.g. Mirman & Magnuson, 2008).

Phonological and semantic effects in lexical *acquisition* have also been examined. However, little of this work has simultaneously examined phonological and semantic effects in the same set of stimuli or set of studies. In the current work we used a computational model of word learning to investigate the influence of phonological similarity and semantic similarity on early word learning.

Phonological Similarity

Numerous researchers have shown that phonological similarity influences lexical recognition, recall, and access in adults (Dahan & Magnuson, 2006; Luce & Pisoni, 1998; Vitevitch & Luce, 1998; Vitevitch, Luce, Pisoni, & Auer, 1999). Luce's work demonstrates how lexical items that differ by a single phoneme (phonological neighbors) can be simultaneously activated and compete with spoken input (Luce & Pisoni, 1998). While this adult work suggests that phonological similarity impedes lexical processing, developmental work on phonological neighbors suggests that phonological similarity may aid typical lexical acquisition. Storkel (2004) examined whether phonological neighborhood density (together with word frequency and

word length) could predict the age of acquisition of early vocabulary items from the Macarthur-Bates Communicative Development Inventory (MCDI) lexical production norms (Dale & Fenson, 1996). She found that words with more phonological neighbors were acquired earlier than words with fewer phonological neighbors, even after accounting for effects of frequency and length. These results suggest that sound similarity (high phonological density) facilitates lexical acquisition.

In contrast with Storkel's work (2004), many nonce word learning studies suggest that infants struggle to learn words that are phonologically similar to one another or to words they already know. Using a habituation task, Stager and Werker (1997) found that 14-month-old infants were able to associate two novel labels with novel objects, but only when the labels were phonologically distinct, like *lif* and *neem*. Infants were unable to map phonologically similar labels *bih* and *dih* to separate objects. This result was quite surprising because using a similar task infants could discriminate the phonemic /b/-/d/ contrast at 8 months (Stager & Werker, 1997). Yet, it was not till 20-months that infants showed clear evidence of learning labels that differed on this contrast (Werker, Fennell, Corcoran, & Stager, 2002).

What can account for these disparate research findings? One important aspect of the child's environment that was not examined in this work is the referent or concept that labels map to. As similarity between labels affects lexical acquisition, it is likely that similarity between referents also affects acquisition. While there has been a significant amount of work investigating how young children will extend category labels based on referent properties', little work jointly examines the role of the label and the role of the referent in lexical acquisition.

Semantic Similarity

Some of the most interesting and revealing work on semantic development investigates label extension and categorization (Quinn & Johnson, 1997; Rakison & Oakes, 2003; Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976). Much of this work has emphasized how the structure of the environment enables infants to group objects and apply category labels to those groups. Their world is well structured; meaningful correlations occur and reoccur, while arbitrary correlations are rarely repeated. This experience allows children to tune into the meaningful and useful correlations in their world.

Research on the shape bias in categorization elegantly demonstrates how the structure of the environment can facilitate language learning. Many of the first words infants learn refer to categories of objects organized by shape.

Experience with these words seems to facilitate categorization abilities. Infants who know 150 words or more, can readily generalize names for newly learned objects to other objects with similar shapes while infants with less than 150 words cannot (Samuelson & Smith, 1999). Samuelson & Smith hypothesized that as children learn more words they extract organizing regularities and form generalizations. These generalizations may initially be restricted to a specific category (e.g. all spherical objects are balls). Then with increased exposure to labeled categories infants form second-order generalizations (e.g. things that are the same shape share a label). These generalizations allow infants to learn the category structure of the objects in their world. Crucially, it is only through sufficient exposure that infants' acquire higher-order generalizations and learn that objects that are the same shape are likely to share a label. If children acquire this bias due to the statistical regularities of the words they know, they must have significant experience with words organized by shape.

Storkel and Adolf (2009) assessed the effect of semantic set size on preschoolers' ability to learn new words. Semantic set size was defined as the number of objects that are meaningfully related to the target word. Subjects showed no difference in initial acquisition of items with large and small set sizes. However, one week after the initial test subjects showed better memory for objects with smaller set sizes. These results suggest that children can learn words more easily when they have a smaller semantic set size and the objects are more unique.

Rogers and McClelland's (2004) categorization model similarly predicts that it will be difficult to learn unique names for items that share many features with other items. Rogers and McClelland hypothesized that infants are sensitive to correlations among different types of directly observable features. These features, which co-occur in the exemplars of a single category, cannot individually define a category. Nor can a specific set of necessary and sufficient features define any category, there are always exceptions. However, the features that consistently co-occur, though not necessarily in every instance of a category, can define a category. For example, birds tend to fly, and have feathers, wings, and beaks. While these features do not always co-occur (penguins have wings but cannot fly) they frequently do and are said to coherently covary with one another. As infants interact in and explore their world they are naturally exposed to these correlations and regularities. Infants are sensitive to the coherent covariation and can use these constellations of features to identify new members of a category. Based on this work, two objects that share many properties will easily map to the same label. While this is beneficial when forming categories, it may be an impediment to children learning the names of similar objects, like "cup" and "glass".

In the current work we use a computational model of word learning to explore effects of phonological and semantic similarity on word learning. Research on phonological similarity is unresolved and suggests similarity

facilitates lexical acquisition in some situations but hinders acquisition in others. We propose that by using a computational model to explore effects of phonological and semantic similarity in a single task, we will be able to better understand this phenomenon.

Methods

The main goal of the model was to simulate behavioral experiments that tested infants' abilities to learn similar sounding labels (Werker & Fennell, 2004). In these studies, infants viewed novel objects on a video screen that were audibly labeled with a nonce word. Infants were repeatedly shown these stimuli until their interest had decreased and they were habituated. After habituation, infants received "same" and "switch" test trials. The same trials were the same as habituation trials. In switch trials the objects paired with each label were switched. That is, in switch trials *dih* was paired with the *bih* object, and *bih* was paired with the *dih* object. Longer looking times to switch trials were interpreted as dishabituation and evidence that children learned the mappings.

Architecture

The architecture of the model is presented in Figure 1. The model was composed of three layers: semantic, hidden and phonological. The phonological layer was the input layer and had 192 units (16 units coding phonetic features for each of 12 possible phonemes), the hidden layer had 200 units, and the semantic layer was the output layer and had 135 units. The semantic and phonological layers had recursive units as well as lateral connections between units within the layers. The semantic layer was the output layer over which targets were set and error was calculated.

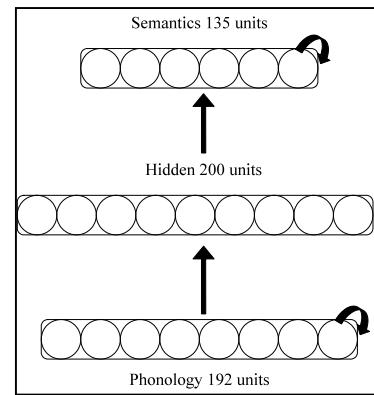


Figure 1: Network Architecture

Training

Three networks initialized with different small random weights, were trained on 332 nouns from the MCDI production checklist (Fenson, et al., 1994). The networks' task was to learn the mapping from phonological labels to semantic referents. Networks were presented with a label on the phonological layer and were to activate the correct set of semantic features describing the referent on the semantic

layer. For example, networks that had learned the word *dog* would activate the 44 semantic feature units that describe a dog (i.e., eats, has tail, is fun, is lovable, etc.) when presented with the phonological representation of *dog* across the phonological input layer.

Networks were trained using standard backpropagation (Rumelhart, Hinton, & Williams, 1986), with cross-entropy error calculated across output units. The learning rate was set to .005 with no momentum. Networks were trained in batches of 20 words. Output activations and weight matrices were saved every 500 training trials to evaluate the course of learning. Training for each word continued until the activation of each semantic output unit was within 0.2 of its target value or training was manually halted for testing.

Testing

To simulate the behavioral experiments, an analog of habituation and the same-switch procedure was used to test the networks. The networks were trained to differing levels of vocabulary size to simulate the different ages at which infants succeed and fail at the task. At these different stages of training the habituation and same-switch test procedures were simulated in the models.

In the behavioral work by Werker and colleagues (Werker & Fennell, 2004), infants were initially habituated to the stimuli. That is, they were repeatedly exposed to label-object pairs until their looking time decreased by 50%. They were next shown “same” and “switch” test trials, in which the label-object pairing from habituation was either preserved or switched. An increased looking time to the switch trials indicated dishabituation and acquisition of the label-object pairings.

As with infant participants the networks were habituated to the stimuli. Error across the output layer served as the model analog to looking time (Schafer & Mareschal, 2001). To establish the baseline error rate for the habituation phase, models were presented with correct label-object pairings for either *bih-dih* or *lif-neem*. After the first presentation of the novel words, activation on the semantic layer was recorded and compared to the semantic representation of the appropriate referent. This error value provided the baseline error rate for the habituation phase. Models were trained on the pair of novel words until error on the output layer reduced by 50% of baseline. Models were next tested with same and switch trials. On both same and switch test trials error across the semantic output layer was recorded. This error represented the mismatch between a model’s expectations and the semantic target of the nonce label. As with infant looking times, larger error indicates surprise and dishabituation from training (Schafer & Mareschal, 2001).

Phonological Representations

Phonological representations of the MCDI nouns and nonce words were based on representations from Joanisse and Seidenberg (1999). See the appendix for a list of features used to represent the phonemes of each word. These representations were slot-based and centered on the first

vowel such that when words were compared, phonemes in the same slot position were compared with one another. For example, the words /sta:r/ and /ka:r/ were aligned in vowel-centered slots such that the /a:r/s were aligned even though /sta:r/ has two initial consonants while /ka:r/ only has one.

Slot-based representations have known limitations and can cause delays in training (Plaut, McClelland, Seidenberg, & Patterson, 1996). In these representations phonemes across slots are independent from one another, and cannot facilitate learning across slots. Therefore though knowing the word *pencil* may facilitate acquisition of *penguin* because of the word-initial overlap; knowledge of neither *penguin* nor *pencil* can facilitate learning *playpen*, which has a word-final *pen*. Despite these limitations, vowel-centering has been shown to minimize this problem (Harm & Seidenberg, 1999).

Semantic Representations

Semantic representations of the MCDI nouns were taken from Howell, Jankowicz & Becker (2005). Howell et al. used a set of 97 perceptually grounded features to code each word in the MCDI (see the appendix for a list of all features). These features were a subset of the McRae, de Sa, and Seidenberg (1997) empirically derived feature set. Howell et al. chose to use only features that were directly observable by children 8 to 28 months old. They then gathered ratings on these 97 features from human raters for each concept on the MCDI. The final vector for each concept was created by averaging raters’ scores.

Howell et al.’s patterns were composed of graded values that varied between 0 and 1, but the majority of features in the set were binary in nature (e.g., “is solid”, “is young” etc.). Therefore, all of the conceptually binary features were re-coded as 1’s and 0’s, with values above .5 becoming 1 and the remaining becoming 0. There were an additional 19 features that coded continuous dimensions (e.g., size, speed, colorfulness, etc.). These features were split into three units representing low, medium and high values of the feature. If a concept had a 0 on one of these continuous dimensions, the high, medium, and low units for that feature were all set to 0. This transformation resulted in semantic patterns using 135 units.

In addition to referents of words from the MCDI, representations for novel referents were created. To create these semantic representations an adult coder, blind to the hypotheses of the studies, looked at pictures and read descriptions of stimuli from published papers. Based on these pictures and/or descriptions each semantic feature was coded as 1 or 0, present or not present, for each novel object.

Results

Word learning experiments conducted by Werker and colleagues (2004) tested children between 14 and 20 months of age. To simulate results over this age range, we used the MCDI norms (Fenson, et al., 1994) to calculate the average number of words children at 14 months can understand. The

norms indicate that the majority of 14-month-olds know at least 64 words. The models reached this level of comprehension at 2500 weight updates. The MCDI comprehension norms do not have data on children older than 16 months, therefore a point later in training that corresponded to a larger vocabulary, 6500 weight updates and 306 known words, was used to simulate the 20-month data point.

We began by simulating the 14-month old studies. Weight matrices produced after 2500 training updates with the full MCDI vocabulary were loaded onto the models. Representations of the two nonce objects were paired with one label from each pair. As with the behavioral studies, the same nonce objects were used for *bih-dih* and *lif-neem*. Networks were habituated and tested with the same-switch procedure as described in the methods sections. All three models showed a larger switch preference when learning *lif* and *neem*, compared to *bih* and *dih* (see Figure 2). This was consistent with 14-month behavioral data (Werker & Fennell, 2004). This indicates that similar to children, the models found the switch trials to be a greater mismatch from what was expected when learning *lif* and *neem*, than when learning *bih* and *dih*.

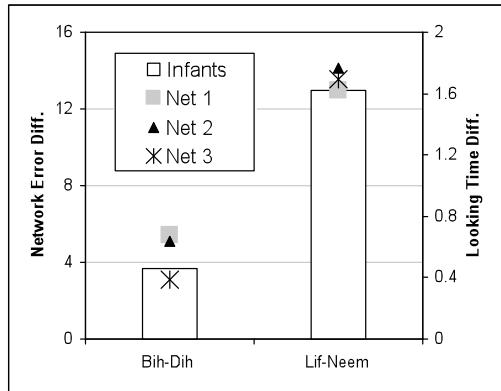


Figure 2: Switch preference for the three networks and infants from Stager and Werker (1997). Difference in error for the networks is labeled on the left y-axis and difference in looking time in seconds for behavioral data is labeled on the right y-axis.

A repeated measures 2 (trial type: same, switch) x 2 (nonce pair: *bih-dih*, *lif-neem*) ANOVA was run on output error from test trials. The main effect of trial type [$F(1,4)=529.571$, $p<.001$] was significant, showing increased error on switch trials for both pairs. There was also a significant interaction between trial type and nonce pair [$F(1,4)=132.42$, $p<.001$]. This result revealed that the switch preference for *lif-neem* was significantly greater than that for *bih-dih*. This replicates the crucial finding that dishabituation is significantly greater for labels that are distinct. The interaction between nonce pairs and test item type is a crucial replication of the Stager & Werker (1997) data.

This computational model of word learning maps phonological representations of labels to semantic feature representations of referents through a 200 unit hidden layer. Weights coming in and out of the hidden layer are adjusted

via the backpropagation algorithm. As the model is trained the hidden layer magnifies differences from the input that map to the correct set of semantic features. The activation across the hidden layer can be thought of as an internal representation of the input that maps to the correct features in the output. If two phonological labels produce similar patterns across the hidden layer, the model will more readily map these to similar referents.

To better understand the models' behavior, hidden layer activations of the nonce words were examined prior to habituation. These activations represent the model's ability to discriminate the nonce labels based on current vocabulary size and composition, but prior to training on the nonce items. Weight matrices produced after 2500 and 6500 training trials on the nouns from the MCDI were loaded onto the networks. The networks were then tested on the *bih-dih* and *lif-neem* mappings. Activations produced on the hidden layer were recorded and the distance between patterns for labels in each pair was calculated. That is, for each model we compared activation patterns produced across the hidden layer for the label *bih* with the activation pattern produced by *dih*. Similarly, the hidden layer activation pattern produced by *lif* was compared to the pattern produced by *neem*. Euclidean distance between the two patterns was calculated to assess the model's ability to represent the input as two separate items (see Table 1).

Weight Update	Label	Distance between labels		
		Net 1	Net 2	Net 3
2500	<i>bih-dih</i>	0.93	0.78	0.78
2500	<i>lif-neem</i>	2.07	2.025	2.13
6500	<i>bih-dih</i>	1.80	1.58	1.74

Table 1: Euclidean distance between hidden representations of yoked label pairs.

After 2500 weight updates, the distance between hidden layer representations of *lif* and *neem* was greater than the difference between *bih* and *dih*. This greater difference shows that the model is better able to represent *lif* and *neem* as distinct labels. Hidden layer representations were also compared at 6500 weight updates when the model successfully maps *bih* and *dih* to distinct referents. With a larger and more diverse vocabulary, the difference between hidden layer representations of *bih* and *dih* is much greater, indicating that the more experienced model is better able to represent them as separate labels. However, the difference is still not as large as between *lif* and *neem* after 2500 updates, indicating that learning *bih* and *dih* when more experienced is possibly still harder than learning *lif* and *neem* at younger ages. This analysis indicates that with more experience the model is better able to represent the important differences between *bih* and *dih*.

Mapping to Distinctive Referents

A major goal of the current work was to examine the role of semantic similarity on lexical acquisition. In addition to

phonological similarity affecting acquisition it is likely the similarity of referents also affects word learning. To test this hypothesis, we created two new semantic patterns that were completely unique. Both patterns had 36 active semantic units, none of which overlapped. The units were chosen pseudo-randomly, and so patterns do not represent any real-world object. Using the same/switch method, we tested lexical acquisition of *bih* and *dih* and *lif* and *neem* paired with the distinct objects after 2500 updates. If semantic distinctiveness does not affect lexical acquisition, the interaction between test item type and label pair (*bih-dih* vs. *lif-neem*) should persist. Alternatively, if semantic distinctiveness can help to differentiate the label-object pairs, there should be no difference in the acquisition of *bih-dih* and *lif-neem*.

As seen in Figure 3, changing only the distinctiveness of the referents allows the model to learn *bih* and *dih* just as well as *lif* and *neem*. By making the referents of the two labels more distinct, similar-sounding labels are acquired as easily as distinctive sounding labels. A repeated measures 2 (test item type) x 2 (label pair) ANOVA was conducted to examine whether the acquisition of *bih-dih* differed from the acquisition of *lif-neem*, when they were mapped to distinct referents. While the significant main effect of test item type [$F(1,4)=482.437$, $p<.001$] persists, the interaction between test item type and label pair is no longer significant [$F(1,4)=.158$, $p=.711$]. Additionally, as seen in Table 2, hidden layer representations are further differentiated after training with distinct objects. This is true for both *bih-dih* and *lif-neem*.

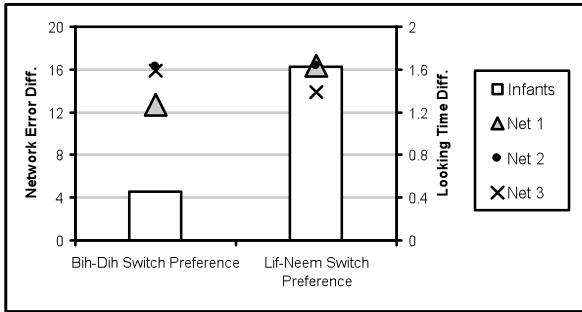


Figure 3: Switch preference for three networks mapping to distinct objects and infants from Stager and Werker (1997). Difference in error for the networks is labeled on the left y-axis and difference in looking time in seconds for behavioral data is labeled on the right y-axis.

Conclusions

The natural world provides infants with strong correlations between linguistic structure and object properties. This structure supports the young child's difficult task of mapping labels to concepts and referents in their world.

In the present work we examined how structure among word forms and words referents can influence word learning. Word learning studies by Werker and colleagues (2004) suggested that high phonological density inhibits acquisition, while Storkel (2004) suggests that in some

contexts, phonological density should facilitate acquisition. Using a computational model of word learning, we explored the role that semantic referents of novel words may play in these findings.

Label	Training	Net 1	Net 2	Net 3
<i>bih-dih</i>	Prior to habituation	.832	.783	.78
<i>lif-neem</i>	Prior to habituation	2.01	2.02	2.11
<i>bih-dih</i>	Post habituation	2.43	2.57	2.53
<i>lif-neem</i>	Post habituation	3.8	3.93	3.36

Table 2: Euclidean distance between hidden representations of yoked label pairs when mapping to **distinct** referents.

The computational model examined effects of semantic and phonological similarity on the *process* of word learning. Using model analogs to habituation, we simulated the basic finding that it is difficult to learn similar sounding labels like *bih* and *dih*. By examining the hidden layer representations of these items we found that the surface similarity of the labels affected the model's ability to treat them as separate items. However, models were able to successfully map *bih* and *dih* to separate objects when the objects were completely distinct. Training with these distinct objects allowed the models to pull apart representations of words that had similar labels, as shown in Table 2.

Importantly, this simulation showed that the referents of labels, and their relationship to other items in the input, can affect word learning. This finding brings to light the need to consider the effects of semantic structure when studying word learning.

Acknowledgments

This work has been generously supported, in part, by the 2008 APA dissertation award, and pre-doctoral grant F31DC008737 from the National Institutes of Health. The author wishes to thank four anonymous Cognitive Science reviewers for helpful comments and suggestions.

Appendix: Sound & Semantic Feature Sets

Sound features: voiced, consonantal, vocalic, sonorant, lateral, continuant, noncontinuant, advanced tongue root, nasal, labial, coronal, anterior, high, distributed, dorsal, radical.

Semantic features: size, weight, strength, speed, temperature, cleanliness, tidiness, brightness, noise, intelligence, goodness, beauty, width, hardness, roughness, height, length, scariness, colorfulness, is black, is blue, is brown, is gold, is green, is grey, is orange, is pink, is purple, is red, is silver, is white, is yellow, is conical, is crooked, is curved, is cylindrical, is flat, is liquid, is rectangular, is round, is solid, is square, is straight, is triangular, has feather, has scales, has fur, is prickly, is sharp, is breakable, made of china, made of cloth, made of leather, made of metal, made of plastic, made of stone, made of wood, climbs, crawls, flies, leaps, runs, swims, breathes, drinks, eats, makes animal noise, singles, talks, has four legs, has beak, has door, has shell, has eyes, has face, has fins, has handle, has leaves, has legs, has paws, has tail, has teeth, has wheels, has whiskers, has wings, is

annoying, is comfortable, is fun, is musical, is scary, is strong smelling, is young, is old, is comforting, is lovable, is edible, is delicious.

References

Dahan, D., & Magnuson, J. S. (2006). Spoken-word recognition. In M. J. Traxler & M. A. Gernsbacher (Eds.), *Handbook of Psycholinguistics* (pp. 249-283). Amsterdam: Academic Press.

Dale, P. S., & Fenson, L. (1996). Lexical development norms for young children. *Behavior Research Methods, Instruments & Computers*(28), 125-127.

Fenson, L., Dale, P. S., Reznick, J. S., Bates, E., Thal, D. J., & Pethick, S. J. (1994). Variability in early communicative development. *Monographs of the Society for Research in Child Development*, 59(5), v-173.

Harm, M. W., & Seidenberg, M. S. (1999). Phonology, reading acquisition, and dyslexia: Insights from connectionist models. *Psychological Review*, 106(3), 491-528.

Howell, S. R., Jankowicz, D., & Becker, S. (2005). A model of grounded language acquisition: Sensorimotor features improve lexical and grammatical learning. *Journal of Memory and Language*, 53(2), 258-276.

Joanisse, M. F., & Seidenberg, M. S. (1999). Impairments in verb morphology following brain injury: A connectionist model. *Proceedings of the National Academy of Sciences of the United States of America*, 7592-7597.

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

McRae, K., de Sa, V. R., & Seidenberg, M. S. (1997). On the nature and scope of featural representations of word meaning. *Journal of Experimental Psychology: General*, 126(2), 99-130.

Mirman, D., & Magnuson, J. S. (2008). Attractor dynamics and semantic neighborhood density: Processing is slowed by near neighbors and speeded by distant neighbors. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 34(1), 65-79.

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal and impaired word reading: Computational principles in quasi-regular domains. *Psychological Review*, 103(1), 56-115.

Quinn, P. C., & Johnson, M. H. (1997). The emergence of perceptual category representations in young infants: A connectionist analysis. *Journal of Experimental Child Psychology*, 66(2), 236-263.

Rakison, D. H., & Oakes, L. M. (Eds.). (2003). *Early category and concept development: Making sense of the blooming, buzzing confusion*: Oxford University Press.

Rogers, T. T., & McClelland, J. L. (2004). *Semantic cognition: A parallel distributed processing approach*: MIT Press.

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.

Rumelhart, D. E., Hinton, G. E., & Williams, R. (1986). Learning internal representations by error propagation. In D. E. Rumelhart, J. L. McClelland & P. R. Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition*. (Vol. 1). Cambridge, MA: MIT Press.

Saffran, J. R., & Sahni, S. D. (in press). Learning the sounds of language. In M. Joanisse, M. Spivey & K. McCrae (Eds.), *Cambridge Handbook of Psycholinguistics*. Cambridge: Cambridge University Press.

Samuelson, L. K., & Smith, L. B. (1999). Early noun vocabularies: Do ontology, category structure and syntax correspond? *Cognition*, 73(1), 1-33.

Schafer, G., & Mareschal, D. (2001). Modeling infant speech sound discrimination using simple associative networks. *Infancy*, 2(1), 7-28.

Stager, C. L., & Werker, J. F. (1997). Infants listen for more phonetic detail in speech perception than in word-learning tasks. *Nature*, 388(6640), 381-382.

Storkel, H. L. (2004). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*(25), 201-221.

Storkel, H. L., & Adolf, S. M. (2009). The effect of semantic set size on word learning by preschool children. *Journal of Speech, Language, and Hearing Research*, 52(2), 289-305.

Vitevitch, M. S., & Luce, P. A. (1998). When words compete: Levels of processing perception of spoken words. *Psychological Science*, 9(4), 325-329.

Vitevitch, M. S., Luce, P. A., Pisoni, D. B., & Auer, E. T., Jr. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language*, 68(1), 306-311.

Werker, J. F., & Fennell, C. (2004). Listening to sounds versus listening to words: Early steps in word learning. In D. G. Hall & S. R. Waxman (Eds.), *Weaving a lexicon*. (pp. 79-109): MIT Press.

Werker, J. F., Fennell, C. T., Corcoran, K. M., & Stager, C. L. (2002). Infants' ability to learn phonetically similar words: Effects of age and vocabulary size. *Infancy*, 3(1), 1-30.