Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition

HOLLY L. STORKEL
University of Kansas

ADDRESS FOR CORRESPONDENCE
Holly L. Storkel, University of Kansas, Department of Speech–Language–Hearing: Sciences and Disorders, 1000 Sunnyside Avenue, Room 3001, Lawrence, KS 66045-7555. E-mail: hstorkel@ku.edu

ABSTRACT
This study tests the claim that children acquire collections of phonologically similar word forms, namely, dense neighborhoods. Age of acquisition (AoA) norms were obtained from two databases: parent report of infant and toddler production and adult self-ratings of AoA. Neighborhood density, word frequency, word length, Density × Frequency and Density × Length were analyzed as potential predictors of AoA using linear regression. Early acquired words were higher in density, higher in word frequency, and shorter in length than late acquired words. Significant interactions provided evidence that the lexical factors predicting AoA varied, depending on the type of word being learned. The implication of these findings for lexical acquisition and language learning are discussed.

When a child encounters a novel word, he or she must store a representation of the referent, a representation of the phonological form, and an association between these two representations. The child's stored knowledge of the characteristics of the referent is referred to as the *semantic representation*, whereas his or her stored knowledge of form characteristics is referred to as the *lexical representation* (Levelt, 1989). The child must integrate these new representations with existing representations. Thus, it is assumed that associations are created between new and old representations that are similar along relevant semantic and lexical dimensions. For this reason, it is possible that the organization of existing representations may influence the creation and integration of new representations, specifically *lexical acquisition*.

The organization of lexical representations has been studied extensively. It has been proposed that lexical representations are organized into similarity neighborhoods based on phonological characteristics (Luce & Pisoni, 1998). One common operational definition of a *similarity neighborhood* is that it includes all the words differing by a one phoneme substitution, deletion, or addition in any word position (e.g., Luce, Goldinger, Auer, & Vitevitch, 2000; Luce & Pisoni, 1998). For

example, neighbors of /kɪt/ would include words such as /sɪt/, /koʊt/, /kɪŋ/, /ɪt/, and /skɪt/. Words that have many neighbors are said to reside in *dense neighborhoods*, whereas words with few neighbors are said to reside in *sparse neighborhoods*. Studies of spoken word recognition and speech production support the psychological reality of this hypothesized organization in adults and children.

In adults, words from dense neighborhoods are recognized more slowly and less accurately than words from sparse neighborhoods in a variety of tasks (e.g., Cluff & Luce, 1990; Goldinger, Luce, & Pisoni, 1989; Luce et al., 2000; Luce & Pisoni, 1998; Newman, Sawusch, & Luce, 1997; Sommers & Lewis, 1999; Vitevitch & Luce, 1998, 1999). Based on this evidence, it is assumed that words in dense neighborhoods compete with one another during word recognition (e.g., Luce & Pisoni, 1998). A different pattern is observed in speech production by adults. Dense neighborhoods seem to facilitate speech production with words from dense neighborhoods being more resistant to speech errors (Vitevitch, 1997) and being retrieved more rapidly than words from sparse neighborhoods (Vitevitch, 2002). In production, it is assumed that this facilitory effect of dense neighborhoods arises because neighbors provide additional activation for the target word, leading to faster access to dense neighborhoods relative to sparse neighborhoods (Vitevitch, 2002). Taken together, the findings from recognition and production suggest that the effect of neighborhood structure is robust, supporting the hypothesis that the lexicon is organized by similarity neighborhoods and that this structure influences spoken language processing.

Do children organize words into neighborhoods in the same manner as adults? Fewer studies have examined the effect of neighborhood density on language processing by children, and these have focused predominately on word recognition rather than speech production. Results show an inhibitory effect of dense neighborhoods on word recognition by children, paralleling the findings for adults. For example, children require more acoustic–phonetic information to recognize words from dense neighborhoods than words from sparse neighborhoods (Garlock, Walley, & Metsala, 2001; Metsala, 1997). Likewise, children are less accurate repeating words from dense neighborhoods than words from sparse neighborhoods (Garlock et al., 2001). This repetition task is thought to tap word recognition, rather than production, because the auditory word form must be recognized before repetition can occur. Moreover, repetition of an auditory stimulus does not necessarily entail retrieval of the lexical form from memory, which is inherent in other production tasks such as picture naming. In general, the effect of neighborhood density on word recognition by children is similar to that found for adults with dense neighborhoods inhibiting recognition. Given this similarity in word recognition, the organization of the child lexicon may be similar to that of the adult lexicon. This organization has the potential to influence lexical acquisition. In particular, dense neighborhoods may facilitate acquisition because the integration of a newly formed lexical representation with numerous existing lexical representations may serve to strengthen the new representation.

Empirical studies provide direct evidence that neighborhood density influences lexical acquisition. When children are exposed to an equal number of novel words from dense and sparse neighborhoods and learning is tracked over time, 17-month-old to 13-year-old children acquire novel words from dense neighborhoods more

rapidly than novel words from sparse neighborhoods (Hollich, Jusczyk, & Luce, 2002; Storkel, 2001; Storkel & Rogers, 2000). Note that in these studies neighborhood density was positively correlated with phonotactic probability (i.e., dense–common vs. sparse–rare), a measure of the likelihood of occurrence of a sound sequence. This correlation is a characteristic of the ambient language (Vitevitch, Luce, Pisoni, & Auer, 1999). These findings suggest that children learn words in dense neighborhoods earlier than those in sparse neighborhoods. This supports the hypothesis that integration with many other known words, as would occur in a dense neighborhood, may strengthen the newly formed lexical representation of a novel word (Jusczyk, Luce, & Charles–Luce, 1994). In this way, fewer exposures to the novel word would be required to form an accurate and detailed lexical representation, leading to more rapid lexical acquisition.

Past research examining the influence of neighborhood density on lexical acquisition has explicitly controlled numerous variables such as word frequency and word length (Hollich et al., 2002; Storkel, 2001; Storkel & Rogers, 2000). For example, in Storkel (2001) children received the same number of exposures to the novel words from dense versus sparse neighborhoods, thereby controlling word frequency. Likewise, the novel words from dense and sparse neighborhoods were the same length (i.e., three phonemes). This is important because neighborhood density is positively correlated with word frequency (Landauer & Streeter, 1973) and negatively correlated with word length (Bard & Shillcock, 1993; Pisoni, Nusbaum, Luce, & Slowiaczek, 1985). Specifically, words in dense neighborhoods tend to be higher in frequency and shorter in length than words in sparse neighborhoods. Moreover, word frequency and word length may influence lexical acquisition (e.g., Rice, Oetting, Marquis, Bode, & Pae, 1994). For this reason, if word frequency and word length were not controlled, it would be difficult to determine whether the influence of neighborhood density on lexical acquisition was attributable to neighborhood density alone or to these other correlated variables.

Because potentially confounding lexical factors were controlled, past empirical studies provide strong support that neighborhood density influences lexical acquisition in these well controlled conditions. At the same time, this level of experimental control limits the external validity of the neighborhood density effect in lexical acquisition. That is, it is possible that children may only acquire dense neighborhoods in well controlled settings. Alternatively, this influence of neighborhood density on lexical acquisition may be robust across both controlled and naturalistic settings. To address this issue, two naturalistic databases of words known by children were identified, and age of acquisition (AoA), neighborhood density, word frequency, and word length were computed for each word in each database. The lexical characteristics were examined as possible predictors of AoA in a linear regression analysis. If a null effect of neighborhood density is obtained, then this would suggest that other factors are stronger determinants of lexical acquisition in naturalistic contexts. In contrast, if a significant effect of neighborhood density emerges, then this would indicate that children acquire dense neighborhoods across a variety of learning contexts. This acquisition of dense neighborhoods, in turn, may have implications for language acquisition more generally.

METHOD

Databases

MacArthur communicative development inventories (CDI). The CDI consists of an extensive list of the words potentially known by young children (Fenson et al., 1993). Previous studies indicate that the CDI is a valid and reliable measure of early lexical acquisition with data available from approximately 1,800 children (see http://www.sci.sdsu.edu/cdi/#lexnorms, Dale & Fenson, 1996; Fenson et al., 1993, 1994). There are two versions of the inventories: infant and toddler. The infant inventory consists of 396 words and is intended for use with children age 8–16 months. Parents are asked to report both the words their infant comprehends and the words produced. The toddler inventory consists of 680 words and is intended for children age 16–30 months. Parents report the words their toddler produces. The infant and toddler production databases were combined for this study. In addition, only nouns were examined because of the previously reported discrepancies between noun and verb learning (e.g., Goldin–Meadow, Seligman, & Gelman, 1976; Leonard et al., 1982; Rice & Woodsmall, 1988) and to afford comparison with previous empirical work that has focused on noun learning (Hollich et al., 2002; Storkel, 2001; Storkel & Rogers, 2000; but see Storkel, 2003). This yielded 380 nouns.

Adult self-ratings of AoA. Adult self-ratings of AoA are collected by having adults estimate the age when they think they first learned a word and its meaning. Previous studies indicate that adult self-ratings are a valid and reliable measure of lexical development (e.g., Carroll & White, 1973a, 1973b; Gilhooly & Gilhooly, 1980; Gilhooly & Logie, 1980; Jorm, 1991; Winters, Winter, & Burger, 1978). Multiple databases of adult self-ratings are available (e.g., Bird, Franklin, & Howard, 2001; Carroll & White, 1973a; Gilhooly & Logie, 1980; Morrison, Chappell, & Ellis, 1997; Snodgrass & Yuditsky, 1996). Two databases were selected (Carroll & White, 1973a; Snodgrass & Yuditsky, 1996) because (a) they contained ratings for more than 200 words to provide a large enough sample for analysis and (b) the ratings were made by American English speakers to afford comparison with the previous studies of neighborhood density in lexical acquisition (Hollich et al., 2002; Storkel, 2001; Storkel & Rogers, 2000). These two databases were combined to yield 382 nouns. Eighty-six words appeared in both databases. These duplicate words were combined by averaging the AoA from each database to yield one rating. Self-ratings were made using a 9-point scale: 1 = age 2 years, 2 = 3 years, 3 = 4 years, 4 = 5 years, 5 = 6 years, 6 = 7–8 years, 7 = 9–10 years, 8 = 11–12 years, and 9 = 13 + years.

Variables

AoA. For each word on the CDI, the proportion of children who reportedly knew the word at each age in 1-month intervals is available. These data were converted into a single AoA defined as the earliest age when 50% or more of the children were reported to know the word (see also Fenson et al., 1994). Words that did not meet this criterion were eliminated, yielding 354 nouns. For the adult self-ratings,

the mean AoA rating across participants as reported in the original studies were used (Carroll & White, 1973a; Snodgrass & Yuditsky, 1996). The AoA then was standardized by creating z scores for each word, that is, z score $=$ (obtained value – M)/SD.

Neighborhood density.   Neighborhood density was computed for each word, using a 20,000 word dictionary containing familiarity ratings (Nusbaum, Pisoni, & Davis, 1984). Neighborhood density was operationally defined as all the words having a familiarity rating of 6 or more on a 7-point scale and differing from a given target word by a one phoneme substitution, deletion, or addition. The restriction in familiarity was intended to more closely approximate a child lexicon. Although this dictionary is based on an adult lexicon, previous studies have shown that it yields measures that accord well with those based on a child database (e.g., Jusczyk et al., 1994). In addition, use of this dictionary and operational definition supports comparisons across studies because most previous work on neighborhood density in American English has relied on these same procedures (Storkel, 2001; Storkel & Rogers, 2000). Furthermore, use of adult-based density counts should be viewed as a conservative approach because measurement error attributable to differences between adult versus child counts would be more likely to yield a null result than a significant result. Raw neighborhood density values were transformed to z scores.

Word frequency.   Word frequency was included as a variable of interest because of its potential influence on lexical acquisition and positive correlation with neighborhood density (Landauer & Streeter, 1973). Given this correlation, it is important to determine that any effect of neighborhood density is independent of word frequency. Likewise, previous work suggested that word frequency might interact with neighborhood density, potentially altering the effect of neighborhood density on lexical acquisition (Metsala, 1997). Word frequency counts were obtained from a study of the expressive vocabulary of first grade children (Moe, Hopkins, & Rush, 1982). This particular source was chosen because low frequency words were not eliminated from the corpus as in other counts (Kolson, 1960; Murphy, 1957). Raw frequency counts in Moe et al. (1982) indicate the number of occurrences of a given word out of 285,732 tokens. When a word was not found in the frequency count, the frequency was assumed to be zero to avoid missing data. For the CDI, 38 words were not found in the frequency count and were assumed to have a frequency of zero. For the adult self-ratings, 116 words were assigned a frequency of zero. Word frequency values were standardized by computing z scores.

Word length.   Word length was included as a variable of interest because of its potential influence on lexical acquisition and its negative correlation with neighborhood density (Bard & Shillcock, 1993; Pisoni et al., 1985). Word length was computed for each word by counting the number of phonemes in the transcription and then converting these raw values to z scores.

Table 1. *Density, word frequency, and word length characteristics of each database*

	CDI	Adult Self-Ratings
N for analysis	354	382
AoA (months)		
M	23	68
SD	4	26
Range	12–30	36–156
Neighborhood density		
M	7	5
SD	7	7
Range	0–33	0–29
Word frequency		
M	52	23
SD	110	66
Range	0–889	0–736
Word length		
M	4.40	5.11
SD	1.56	1.97
Range	1–10	1–11

Note: N, number of words; *M*, mean; *SD*, standard deviation.

Comparison across databases

The words in each database were compared because overlap or differences between the databases might affect the analyses of interest. There were 123 words that appeared in both the CDI and the adult self-ratings. Thus, approximately 65–68% of the words in each database were unique. Table 1 shows the descriptive statistics for each raw variable across each database. To examine differences in AoA, neighborhood density, word frequency, and word length across the databases, *t* tests were used. Significant differences were found across the databases for each dependent variable (all *t*s > 2.78; all *p*s < .01). The words from the CDI were earlier acquired, higher in neighborhood density, higher in word frequency, and shorter in word length than the words from the adult self-ratings.

RESULTS

To examine the effect of neighborhood density on lexical acquisition, a separate linear regression analysis was performed for each database: CDI and adult self-ratings. Words were treated as a random factor. The variables neighborhood density, word frequency, and word length were entered as possible predictors of AoA. The interaction terms of Density × Frequency and Density × Length also were of interest as potential predictors of AoA. These interaction terms were orthogonalized to avoid problems related to collinearity. Collinearity concerns correlations among predictor variables that then lead to inflation of the variance. Collinearity problems often arise when modeling main effects and interactions because the main effects are highly correlated with the interaction term (e.g.,

Table 2. *Results of the linear regression analysis for each database*

	B Estimate	Standard Error	t	p	Partial Correlation
CDI ($n = 354$)					
Neighborhood density	−0.191	0.070	−2.72	.007	−.14
Word frequency	−0.101	0.052	−1.94	.053	−.10
Word length	0.126	0.072	1.76	.080	.09
Density × Frequency	0.096	0.052	1.87	.063	.10
Density × Length	0.135	0.050	2.71	.007	.14
Adult self-ratings ($n = 382$)					
Neighborhood density	−0.208	0.062	−3.34	.001	−.17
Word frequency	−0.193	0.046	−4.19	<.001	−.21
Word length	0.241	0.063	3.82	<.001	.19
Density × Frequency	0.094	0.045	2.09	.037	.11
Density × Length	−0.040	0.044	−0.90	.367	−.05

Note: The partial correlation is between the given variable and AoA when all the other variables listed are controlled.

Kleinbaum, Kupper, & Muller, 1988). Interaction terms can be orthogonalized by regressing the interaction term on the main effects and using the residuals from this analysis in the analysis of interest (Little, Hoffman, Bovaird, & Widaman, 2002). These residuals represent the variance in the interaction that is not accounted for by the main effects, thereby eliminating the correlation between the interaction and the main effects. Thus, neighborhood density and word frequency were entered as predictors of the Density × Frequency interaction, and the residuals then were used as a predictor in the AoA analysis. Likewise, neighborhood density and word length were entered as predictors of the Density × Frequency interaction, and the residuals from this analysis then were entered as predictors of AoA.

All five predictor variables were used in each AoA regression analysis so that the independent effect of each variable could be determined. That is, the *t* statistics and corresponding *p* value for each predictor variable represent the significance of the effect of that particular variable when the other three predictor variables are controlled (e.g., Kleinbaum et al., 1988). In this way, the interpretation of the regression analyses is similar to the interpretation of a partial correlation (i.e., a measure of the linear relationship between two variables after controlling for the effects of the other variables in the analysis). It was critical that the analyses be performed in this way, rather than by only including significant predictor variables, to avoid the confounding effects of correlated predictor variables. The results are displayed in Table 2.

CDI

The regression model for the CDI was statistically significant, $F(5, 348) = 11.759$, $p < 0.001$, $r^2 = 0.15$. In terms of main effects, neighborhood density was a significant predictor of AoA. The slope for this variable was negative, indicating that as neighborhood density increased, AoA decreased. The main effect of word

Table 3. *Results of the linear regression analysis for subgroups of words within the CDI*

	B Estimate	Standard Error	t	p	Partial Correlation
Low frequency words ($n = 279$)					
Neighborhood density	−0.179	0.080	−2.23	.027	−.13
Word length	0.114	0.072	1.58	.116	.10
Density × Length	0.139	0.053	2.61	.010	.16
High frequency words ($n = 75$)					
Neighborhood density	0.014	0.156	0.09	.927	.01
Word length	0.714	0.378	1.89	.063	.22
Density × Length	0.033	0.156	0.21	.834	.03
Short words ($n = 222$)					
Neighborhood density	−0.249	0.069	−3.58	<.001	−.24
Word frequency	−0.147	0.056	−2.60	.010	−.17
Density × Frequency	0.148	0.058	2.56	.011	.17
Long words ($n = 132$)					
Neighborhood density	−0.419	0.515	−0.81	.418	−.07
Word frequency	−0.572	1.758	−0.33	.745	−.03
Density × Frequency	−0.532	1.256	−0.42	.673	−.04

Note: The partial correlation is between the given variable and AoA when all the other variables listed are controlled.

frequency and word length approached significance. Here, as frequency increased, AoA decreased. In contrast, as length increased, AoA increased. These main effects were qualified by the near significant interaction of Density × Frequency and the significant interaction of Density × Length. To explore these interactions, words were coded as low or high frequency based on z scores (i.e., negative = low; positive = high) or were coded as short or long based on z scores. Note that Table 1 displays the means for each variable and this coding of low versus high corresponds to scores below versus above the mean. To determine the Density × Frequency interaction, neighborhood density, word length, and Density × Length were entered in a regression analysis as predictors of AoA for each subgroup of words (low vs. high frequency). To determine the Density × Length interaction, neighborhood density, word frequency, and Density × Frequency were the predictors for the AoA regression analysis for each subgroup (short vs. long words). The results of these analyses are shown in Table 3.

Figure 1 shows the neighborhood density of the words produced at each age for low frequency (Figure 1a) versus high frequency words (Figure 1b). In the regression analysis for low frequency words, neighborhood density and Density × Length were significant predictors of AoA. High density words were acquired earlier than low density words. The Density × Length interaction will be considered below. In the regression analysis for high frequency words, only the effect of word length on AoA approached significance. Short words tended to be acquired earlier than long words.

Figure 2 shows the neighborhood density of the words produced at each age for short (Figure 2a) versus long words (Figure 2b). Results of the regression analysis
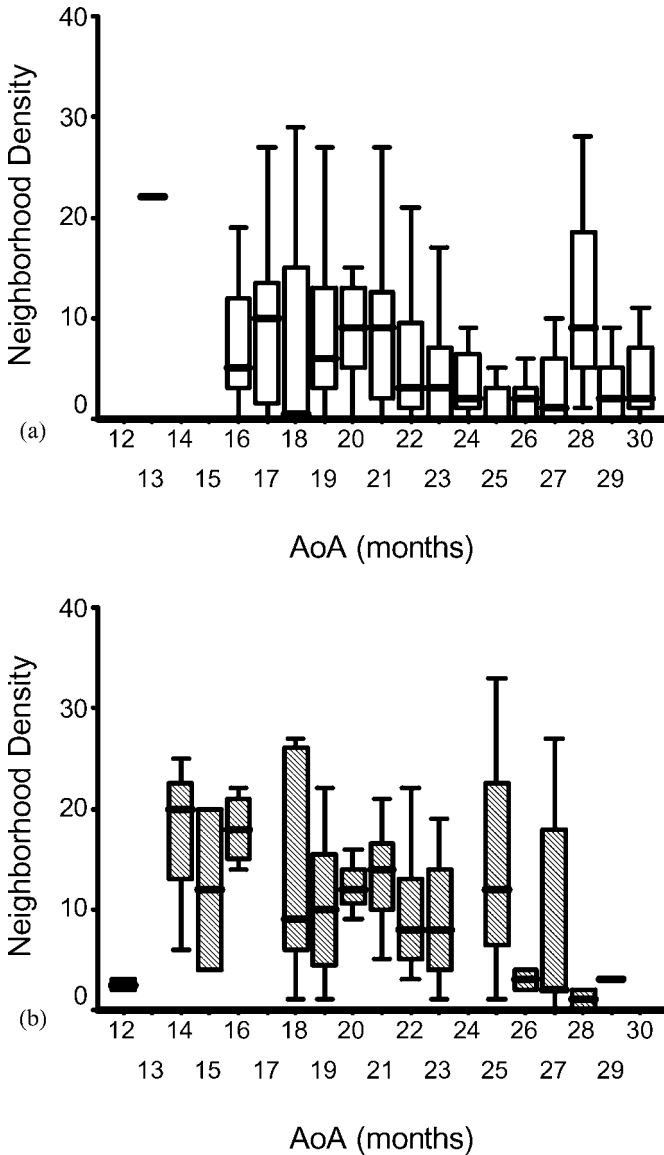
Figure 1. The neighborhood density of the words produced by 50% of infants/toddlers in the CDI at each age in months. Open bars represent (a) low frequency words and filled bars represent (b) high frequency words. The line within the bar represents the 50th percentile. The ends of the bars represent the 25th and 75th percentiles. The horizontal top and bottom bars indicate the lowest and highest values, excluding outliers.
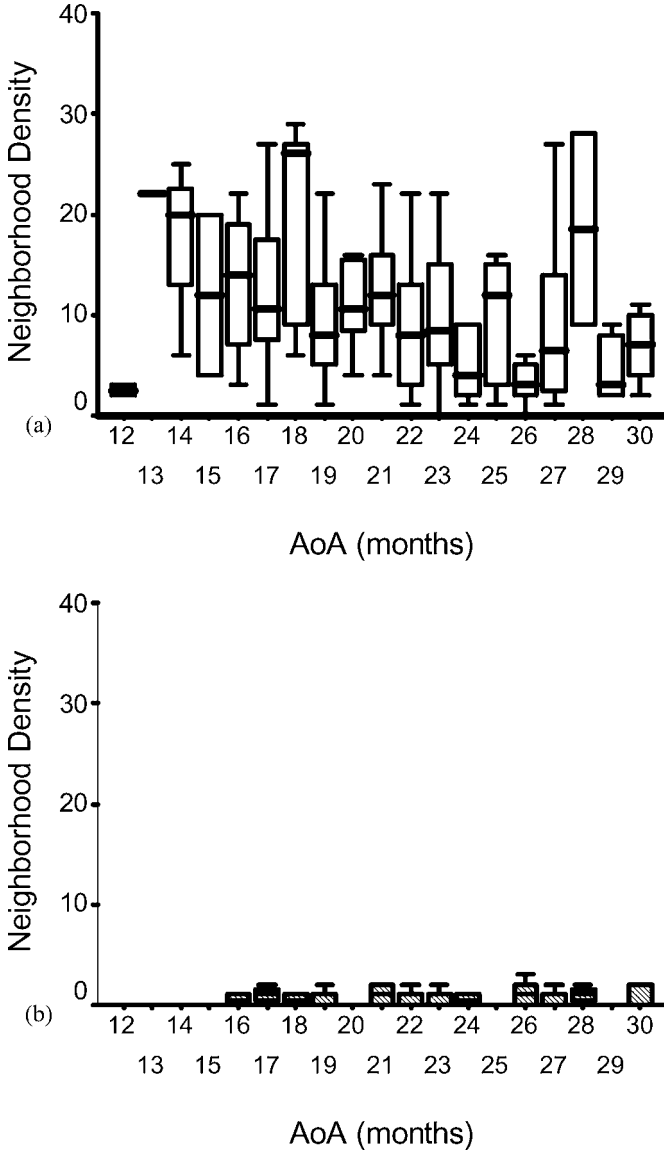
Figure 2. The neighborhood density of the words produced by 50% of infants/toddlers in the CDI at each age in months. Open bars represent (a) short words and filled bars represent (b) long words. The line within the bar represents the 50th percentile. The ends of the bars represent the 25th and 75th percentiles. The horizontal top and bottom bars indicate the lowest and highest values, excluding outliers.

for short words indicated that neighborhood density, word frequency, and
Density × Frequency were significant predictors of AoA. High density or high
frequency words were acquired earlier than low density or low frequency words.
Relative to the Density × Frequency interaction, the effect of neighborhood density
on AoA was more robust for low frequency words than for high frequency words,
as previously described. Results of the regression analysis for long words showed
that none of the variables were significant predictors of AoA. Figure 2 shows that
there was very little variability in neighborhood density for long words, potentially
accounting for this limited effect of density on AoA of long words. In summary,
the Density × Length interaction appeared to be attributable to a significant effect
of neighborhood density for short words, but not for long words.

Adult self-ratings

The regression model for the adult self-ratings was statistically significant,
$F(5, 376) = 28.365$, $p < 0.001$, $r^2 = 0.27$. In terms of main effects, neigh-
borhood density, word frequency, and word length were significant predictors
of AoA. As neighborhood density and word frequency increased, AoA decreased.
In contrast, as word length increased, AoA increased. This parallels the findings
from the CDI. These main effects were qualified by a significant interaction of
Density × Frequency. This significant interaction was explored by performing
separate regression analyses for low versus high frequency words, as previously
described for the CDI. Although the Density × Length interaction was not signif-
icant for this database, the effect of neighborhood density for short versus long
words was explored as a point of comparison to the CDI analysis. Table 4 displays
the results for the regression analysis.

Figure 3 shows the neighborhood density of the words produced at each age
for low frequency (Figure 3a) versus high frequency words (Figure 3b). Analy-
sis of low frequency words showed that neighborhood density, word length, and
Density × Length were significant predictors of AoA. Words higher in density
were acquired earlier than words lower in density, and shorter words were acquired
earlier than longer words. The interaction of density and length will be considered
below. For high frequency words, word length was the only significant predictor
of AoA with short words being acquired earlier than long words.

Figure 4 shows the neighborhood density of the words produced at each age
for short (Figure 4a) versus long words (Figure 4b). In the regression analysis
of short words, neighborhood density and word frequency emerged as significant
predictors of AoA. Early acquired words resided in dense neighborhoods, and
later acquired words resided in sparse neighborhoods. Likewise, early acquired
words tended to be higher in frequency than later acquired words. The interaction
Density × Frequency approached significance with the effect of neighborhood
density on AoA tending to be more robust for low frequency than high frequency
words. Turning to the regression analysis for long words, there were some difficul-
ties performing the analysis due to collinearities among the predictor variables. For
this reason, Density × Frequency could not be included in the regression analysis.
This modified analysis showed that word frequency was the only significant predic-
tor of AoA. High frequency words were acquired at earlier ages than low frequency

Table 4. *Results of the linear regression analysis for subgroups of words within adult self-ratings*

	B Estimate	Standard Error	*t*	*p*	Partial Correlation
Low frequency words (*n* = 309)					
Neighborhood density	−0.237	0.079	−2.99	.003	−.17
Word length	0.217	0.071	3.06	.002	.17
Density × Length	−0.028	0.054	−0.52	.604	−.03
High frequency words (*n* = 73)					
Neighborhood density	0.018	0.062	0.30	.767	.04
Word length	0.220	0.099	2.23	.029	.26
Density × Length	−0.022	0.048	−0.47	.642	−.06
Short words (*n* = 240)					
Neighborhood density	−0.264	0.048	−5.54	<.001	−.34
Word frequency	−0.180	0.041	−4.41	<.001	−.28
Density × Frequency	0.071	0.040	1.76	.080	.11
Long words (*n* = 142)[a]					
Neighborhood density	−0.140	0.200	−0.70	.484	−.06
Word frequency	−1.513	0.414	−3.66	<.001	−.30
Density × Frequency	Could not be estimated due to collinearities				

Note: The partial correlation is between the given variable and AoA when all the other variables listed are controlled.
[a]The constant had to be excluded from the model because of collinearities with neighborhood density.

words. Long words showed little variation in neighborhood density, attenuating the effect of this variable on AoA. As in the CDI analysis, the Density × Length interaction appeared to be attributable to significant effects of neighborhood density on the AoA of short words, but not long words where limited variation in density was observed.

DISCUSSION

Past work has shown that words from dense neighborhoods are acquired more quickly than words from sparse neighborhoods (Hollich et al., 2002; Storkel, 2001; Storkel & Rogers, 2000). The purpose of the current study was to establish the external validity of this claim by examining the effect of neighborhood density on naturalistic samples of lexical acquisition. To rule out the confounding effects of correlated variables, the effects of word frequency and word length on lexical acquisition were also examined. Table 5 summarizes the results across the two naturalistic databases of lexical acquisition. The effects observed generally were consistent across the two databases. Early acquired words resided in dense neighborhoods whereas later acquired words resided in sparse neighborhoods. Importantly, the effects obtained mirrored those of previously reported empirical studies (Hollich et al., 2002; Rice et al., 1994; Storkel, 2001; Storkel & Rogers, 2000). This provides further support that the organization of existing lexical representations
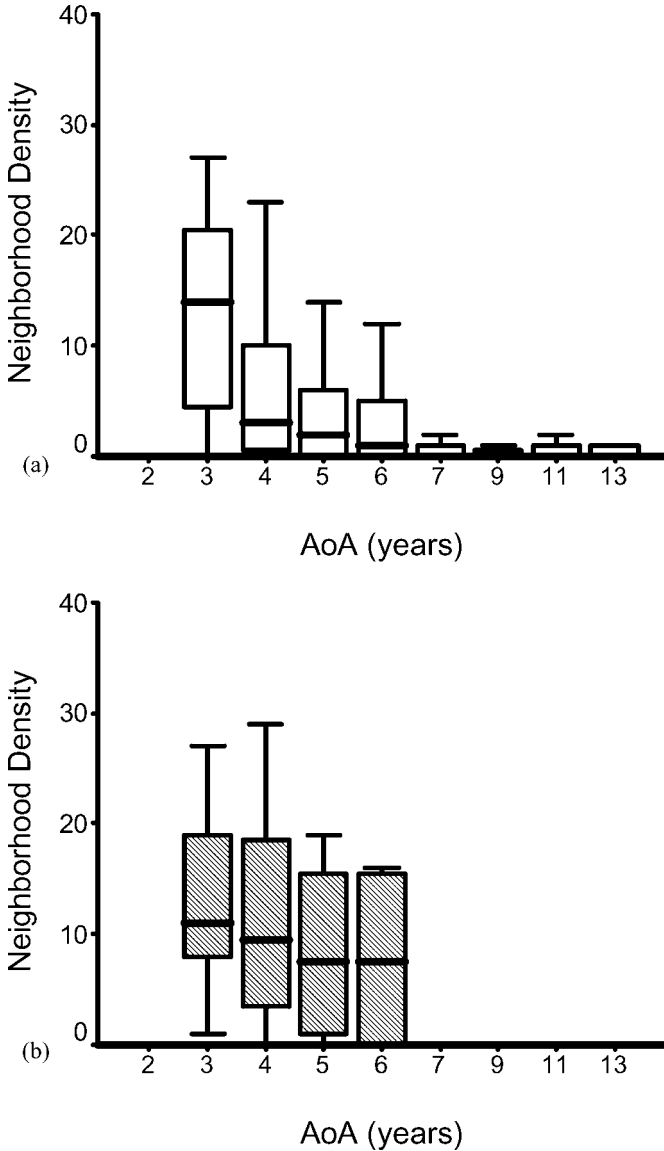
Figure 3. The neighborhood density as a function of adult self-rating of AoA converted to age in years. Open bars represent (a) low frequency words and filled bars represent (b) high frequency words. The line within the bar represents the 50th percentile. The ends of the bars represent the 25th and 75th percentiles. The horizontal top and bottom bars indicate the lowest and highest values, excluding outliers.
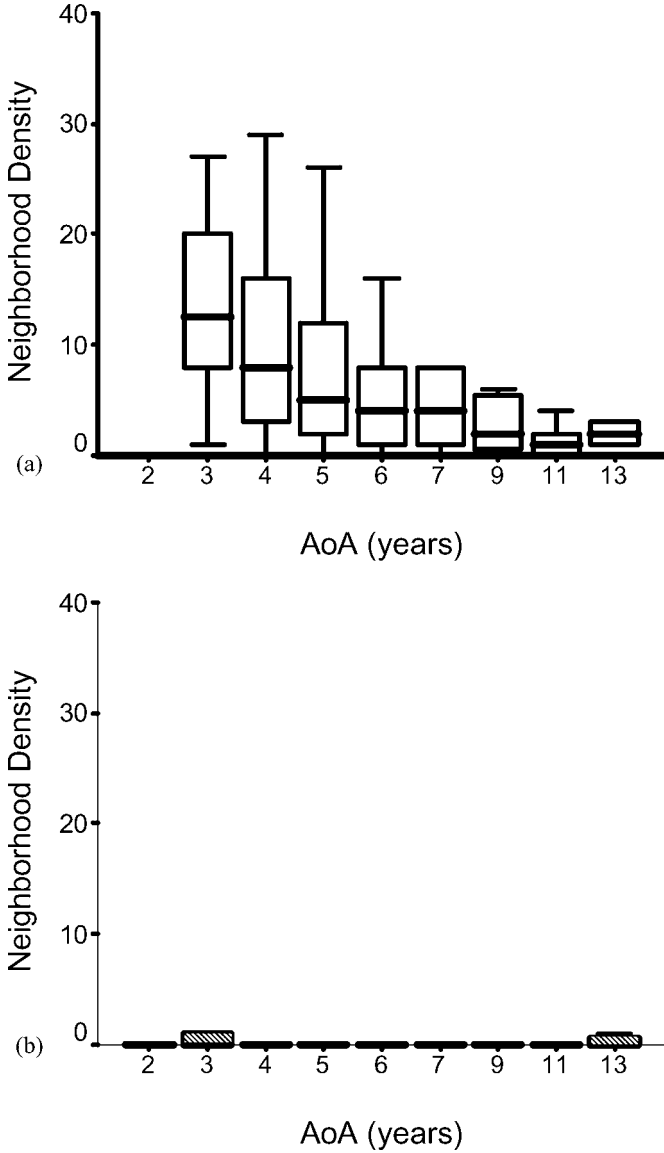
Figure 4. The neighborhood density as a function of adult self-rating of AoA converted to age in years. Open bars represent (a) short words and filled bars represent (b) long words. The line within the bar represents the 50th percentile. The ends of the bars represent the 25th and 75th percentiles. The horizontal top and bottom bars indicate the lowest and highest values, excluding outliers.

Table 5. *Summary of the findings from the current study*

	CDI	Adult Self-Ratings
Neighborhood density	Early acquired = dense	Early acquired = dense
Word frequency	Early acquired ∼ frequent	Early acquired = frequent
Word length	Early acquired ∼ short	Early acquired = short
Density × Frequency	∼Significant	Significant
Low frequency	Early acquired = dense	Early acquired = dense
		Early acquired = short
High frequency	Early acquired ∼ short	Early acquired = short
Density × Length	Significant	Not significant
Short words	Early acquired = dense	Early acquired = dense
	Early acquired = frequent	Early acquired = frequent
Long words	No significant effects	Early acquired = frequent

Note: ∼, trends that approached significance.

influences the creation and integration of new representations. In particular, similarity to many existing lexical representations facilitates lexical acquisition.

New observations also emerged. In particular, the effect of neighborhood density was only evident for low frequency words, not for high frequency words. In addition, neighborhood density predicted AoA for short words but not for long words. Previous research documenting an effect of neighborhood density on lexical acquisition has tended to provide few exposures to the words to be learned (i.e., low frequency exposure) and has focused on monosyllabic words containing few phonemes (i.e., short words). Thus, the findings of the current study are in line with previous empirical studies by showing robust effects of neighborhood density on AoA for low frequency and short words. The current findings suggest that neighborhood density may play a lesser role when learning high frequency or long words. This reduced effect of neighborhood density for high frequency words may be understood by considering the relationship between high frequency and exposure. That is, frequency is indicative of the number of times a child is likely to encounter a word. The current findings suggest that the number of exposures to a word may neutralize the effect of neighborhood density. Specifically, if given enough exposures to a word, a child will learn the word regardless of its neighborhood density. To illustrate, if a child were exposed to two new words, one from a dense neighborhood and one from a sparse neighborhood, then with only minimal exposure to both words (i.e., low frequency) an advantage for the dense word over the sparse word would be predicted. With greater exposure (i.e., high frequency), the child would likely learn both words and no dense neighborhood advantage would be observed. Note from this scenario that a dense neighborhood advantage would be expected for all words following minimal exposure (e.g., during fast mapping) and that this difference between dense and sparse neighborhoods would be reduced with greater exposure (e.g., during extended mapping). In contrast, the reduced effect of neighborhood density for long words is likely attributable to limited variability in density. In particular, density for long words ranged from 0 to 8 for the CDI and from 0 to 3 for the adult self-ratings versus 0 to 33 and 0 to 29

for short words for the CDI and adult self-ratings, respectively. In this way, only small changes in density are possible for long words, and these small changes in density may not be sufficient to produce a learning advantage.

In terms of the effect of word frequency, early acquired words were higher in frequency than later acquired words. This effect appeared to be more robust for short words than for long words. Significant frequency effects for long words were only observed for adult self-ratings. This inconsistent effect of word frequency on AoA for long words may relate to limited variability in word frequency for long words. The standard deviation of the frequency for long words was 42 (range = 0–326) for the CDI and 13 (range = 0–109) for adult self-ratings, as compared to 131 (range = 0–889) for CDI short words and 80 (range = 0–736) for adult self-ratings short words. Changes in frequency for long words may not have been sufficiently large to produce a learning advantage. Counter to this hypothesis, analysis of the adult self-ratings of long words, which showed a more limited frequency range than the CDI long words, did yield a significant effect of word frequency. Here it is important to remember that the Density × Frequency interaction term could not be included in the regression analysis of the AoA long words due to problems of collinearities. It could be that some of the variance predicted by word frequency may actually be attributable to the Density × Frequency interaction and, if the regression model could have been estimated with this interaction included, the effect of word frequency may have been reduced. Future investigation of acquisition of long words may shed light on this issue.

Turning to the main effect of word length, early acquired words were shorter than later acquired words, and this effect was consistent across databases for high frequency but not for low frequency words. Note that this is the opposite effect observed for neighborhood density where the density effect on AoA was more consistent for low frequency than high frequency words. This pattern of results may indicate that these two lexical characteristics influence different aspects of lexical acquisition. In particular, word length may be more critical to the formation of a lexical representation whereas neighborhood density may play a greater role in the retention of a lexical representation. Word length likely influences initial processing of the phonological form of the novel word and the ability to hold the novel sound sequence in working memory while an initial lexical representation is formed (see Gathercole & Baddeley, 1993, for review). Once a lexical representation is formed, neighborhood density may influence retention because integration with many existing lexical representations may strengthen the new representation. The importance of each of these components of lexical acquisition may vary by word frequency. It is possible that word frequency plays a role in retention of a lexical representation. With each exposure to a word, the lexical representation is accessed and modified or elaborated. Alternatively, in an episodic view of long-term memory, greater exposure to a word would increase the number of exemplars of that word in memory (e.g., Goldinger, 1998). For high frequency words, word length would influence the initial processing of the phonological form and the creation of the new lexical representation. Neighborhood density may be less important because frequency itself would facilitate retention of the new representation in memory, minimizing the facilitory contribution of dense neighborhoods. For low frequency words, word length may still influence initial

processing of the sound sequence and creation of a new representation, but in some cases this may be overshadowed by the influence of neighborhood density on the retention of the representation. Because the number of exposures to the word is relatively minimal for low frequency words, neighborhood density may be the main factor that promotes retention of the new lexical representation in memory. Investigation of the influence of these variables over the time course of lexical acquisition may support or refute this account.

Lexical representations and language learning

The acquisition of dense neighborhoods may have consequences for the structure of underlying lexical representations. It has been suggested that lexical representations initially may be holistic or less segmentally detailed (Charles–Luce & Luce, 1990, 1995; Logan, 1992; Metsala & Walley, 1998; Storkel, 2002). Given that young children know very few words, a holistic representation may be sufficient to distinguish each word from every other and avoid confusion among words. However, as more words are acquired it is likely that a holistic representation ultimately will not be sufficient to uniquely distinguish each word. It is assumed that at this point, lexical representations become segmentally detailed. Moreover, this change from holistic to detailed representations may occur differentially across the lexicon such that certain words may be more likely to become segmentally represented. In particular, words in dense neighborhoods may be more likely to have segmentally detailed representations earlier in acquisition because of the greater potential for overlap with other words in the language (Metsala & Walley, 1998). In support of this hypothesis, Storkel (2002) provided evidence that preschool children do organize words into neighborhoods based on similarity relationships but the relevant similarity relationship differed across neighborhoods. Specifically, words in dense neighborhoods were organized segmentally by phoneme similarity, but words in sparse neighborhoods seemed to be organized holistically by manner class similarity (see also Logan, 1992).

These proposed changes in underlying lexical representations may have consequences for language learning. As children learn language, sounds that were previously produced inaccurately become accurate, and this sound change is often implemented on a word-by-word basis, a process termed *lexical diffusion* (Ferguson & Farwell, 1975; Stoel–Gammon & Cooper, 1984). Thus, a sound may be produced accurately in some, but not all, target words of the language, and this is not predicted by context (i.e., word position; adjacent sounds). Recent evidence suggests that the words vulnerable to sound change can be distinguished from those that are resistant to sound change based on lexical characteristics. In particular, words from dense neighborhoods are more likely to undergo sound change than words from sparse neighborhoods (Gierut & Morrisette, 1998; Gierut & Storkel, 2002; Morrisette, 1999; Storkel & Gierut, 2002). Dense neighborhood structure appears to facilitate lexical diffusion of sounds. This facilitation of sound change may be related to the segmentally detailed nature of lexical representations of words residing in dense neighborhoods. As a result, acquiring dense neighborhoods may ultimately facilitate certain aspects of phonological acquisition (but see Gierut, Morrisette, & Champion, 1999; Morrisette & Gierut, 2002). Note that an

alternative interpretation of these findings is that these effects may be attributable to phonotactic probability, rather than neighborhood density, due to the previously described correlation between density and phonotactic probability. However, Vitevitch (2002) showed neighborhood density effects on speech production even when phonotactic probability was controlled. Therefore, it is equally possible that the effect of neighborhood density on sound change could be independent of phonotactic probability.

Written language acquisition may also be influenced by changes in lexical representations associated with dense neighborhood structure. In particular, phonological awareness, an important precursor to the acquisition of written language in alphabetic orthographies, may arise when lexical representations become more segmentally detailed (Metsala & Walley, 1998). In support of this hypothesis, children appear to show better phonological awareness for words from dense neighborhoods than for words from sparse neighborhoods (Metsala, 1999). Thus, acquisition of dense neighborhoods may help to foster phonological awareness.

Inconsistent findings across databases

Although the two databases showed a consistent effect of neighborhood density, word frequency, and word length on lexical acquisition, the steepness of the slopes of the regression equations and the age when a plateau in these effects occurred varied across databases. For example, the regression equation for the CDI showed a shallower slope for neighborhood density than did the equation for adult self-ratings. In addition, the effect of neighborhood density appeared to level off at about 24 months for the CDI and at about 4–5 years for adult self-ratings. These discrepancies across databases may relate to differences in measurement scales. Specifically, the scale range varied across databases with the CDI database using a scale from 12 to 30 months and the adult self-ratings database using a scale from 2 to 13 years. Likewise, the scale increments differed across databases with the CDI using 1-month intervals and the adult self-ratings using 1- and 2-year intervals. Measurement sensitivity may be affected by these scale differences. Future work is needed to further examine how the effect of lexical characteristics on acquisition may change over time to determine when these effects might begin to plateau. The point of plateau could be indicative of a transition in lexical acquisition from one set of lexical factors being influential to a different set of factors becoming influential. Longitudinal research might be particularly helpful in providing evidence of the location of this plateau.

CONCLUSION

This study extends the results from previous empirical work by examining naturalistic lexical acquisition. The results provided additional evidence that early acquired words reside in dense neighborhoods whereas later acquired words reside in sparse neighborhoods. This acquisition of dense neighborhoods appears to have consequences for phonological acquisition in terms of both productive phonology and phonological awareness. Furthermore, the predictors of AoA appeared to change, depending on the types of words being learned. That is, the AoA

of low frequency words was predicted by neighborhood density. In contrast, the AoA of high frequency words was predicted by word length. For short words, both neighborhood density and word frequency predicted AoA. Predicting the AoA of long words was difficult, with word frequency emerging as a significant predictor in only one database.

ACKNOWLEDGMENTS

REFERENCES

Bard, E., & Shillcock, R. (1993). Competitor effects during lexical access: Chasing Zipf's tail. In G. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing* (pp. 235–275). Hove, UK: Erlbaum.

Bird, H., Franklin, S., & Howard, D. (2001). Age of acquisition and imageability ratings for a large set of words, including verbs and function words. *Behavior Research Methods, Instruments, & Computers, 33*, 73–79.

Carroll, J. B., & White, M. N. (1973a). Age-of-acquisition norms for 220 picturable nouns. *Journal of Verbal Learning and Verbal Behavior, 12*, 563–576.

Carroll, J. B., & White, M. N. (1973b). Word frequency and age of acquisition as determiners of picture-naming latency. *Quarterly Journal of Experimental Psychology, 25*, 85–95.

Charles–Luce, J., & Luce, P. A. (1990). Similarity neighbourhoods of words in young children's lexicons. *Journal of Child Language, 17*, 205–215.

Charles–Luce, J., & Luce, P. A. (1995). An examination of similarity neighbourhoods in young children's receptive vocabularies. *Journal of Child Language, 22*, 727–735.

Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 551–563.

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

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

Fenson, L., Dale, P. S., Reznick, J. S., Thal, D., Bates, E., Hartung, J. P., Pethick, S., & Reilly, J. S. (1993). *The MacArthur Communicative Development Inventories: User's guide and technical manual*. San Diego, CA: Singular Publishing.

Ferguson, C. A., & Farwell, C. B. (1975). Words and sounds in early language acquisition. *Language, 51*, 419–439.

Garlock, V. M., Walley, A. C., & Metsala, J. L. (2001). Age-of-acquisition, word frequency, and neighborhood density effects on spoken word recognition by children and adults. *Journal of Memory and Language, 45*, 468–492.

Gathercole, S. E., & Baddeley, A. D. (1993). *Working memory and language*. Hove, UK: Erlbaum.

Gierut, J. A., & Morrisette, M. L. (1998). Lexical properties in implementation of sound change. In A. Greenhill, M. Hughes, H. Littlefield, & H. Walsh (Eds.), *Proceedings of the 22nd Annual Boston University Conference on Language Development* (pp. 257–268). Somerville, MA: Cascadilla Press.

Gierut, J. A., Morrisette, M. L., & Champion, A. H. (1999). Lexical constraints in phonological acquisition. *Journal of Child Language, 26*, 261–294.

Gierut, J. A., & Storkel, H. L. (2002). Markedness and the grammar in lexical diffusion of fricatives. *Clinical Linguistics & Phonetics, 16*, 115–134.

Gilhooly, K. J., & Gilhooly, M. L. (1980). The validity of age-of-acquisition ratings. *British Journal of Psychology, 71*, 105–110.

Gilhooly, K. J., & Logie, R. H. (1980). Age-of-acquisition, imagery, concreteness, familiarity, and ambiguity measures for 1,944 words. *Behavior Research Methods, Instruments, & Computers, 12*, 395–427.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review, 105*, 251–279.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28*, 501–518.

Goldin–Meadow, S., Seligman, M., & Gelman, R. (1976). Language in the two-year-old. *Cognition, 4*, 189–202.

Hollich, G., Jusczyk, P. W., & Luce, P. A. (2002). Lexical neighborhood effects in 17-month-old word learning. In B. Skarabela, S. Fish, & A. H.-J. Do (Eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development* (Vol. 1). Sommerville, MA: Cascadilla Press.

Jorm, A. F. (1991). The validity of word age-of-acquisition ratings: A longitudinal study of a child's word knowledge. *British Journal of Developmental Psychology, 9*, 437–444.

Jusczyk, P. W., Luce, P. A., & Charles–Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*, 630–645.

Kleinbaum, D. G., Kupper, L. L., & Muller, K. E. (1988). *Applied regression analysis and other multivariable methods* (2nd ed.). Belmont, CA: Duxbury Press.

Kolson, C. J. (1960). *The vocabulary of kindergarten children.* Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh, PA.

Landauer, T. K., & Streeter, L. A. (1973). Structural differences between common and rare words: Failure of equivalence assumptions for theories of word recognition. *Journal of Verbal Learning and Verbal Behavior, 12*, 119–131.

Leonard, L. B., Schwartz, R. G., Chapman, K., Rowan, L. E., Prelock, P. A., Terrell, B., Weiss, A. L., & Messick, C. (1982). Early lexical acquisition in children with specific language impairment. *Journal of Speech and Hearing Research, 25*, 554–564.

Levelt, W. J. M. (1989). *Speaking: From intention to articulation.* Cambridge, MA: MIT Press.

Little, T. D., Hoffman, L., Bovaird, J., & Widaman, K. F. (2002). *On the merits of orthogonalizing powered and interaction terms: Implications for modeling interactions among latent variables.* Paper presented at the Society for Multivariate Experimental Psychology.

Logan, J. S. (1992). *A computational analysis of young children's lexicons* (Tech. Rep. No. 8). Bloomington, IN: Indiana University, Speech Research Laboratory.

Luce, P. A., Goldinger, S. D., Auer, E. T., & Vitevitch, M. S. (2000). Phonetic priming, neighborhood activation, and PARSYN. *Perception & Psychophysics, 62*, 615–625.

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

Metsala, J. L. (1997). An examination of word frequency and neighborhood density in the development of spoken-word recognition. *Memory and Cognition, 25*, 47–56.

Metsala, J. L. (1999). Young children's phonological awareness and nonword repetition as a function of vocabulary development. *Journal of Educational Psychology, 91*, 3–19.

Metsala, J. L., & Walley, A. C. (1998). Spoken vocabulary growth and the segmental restructuring of lexical representations: Precursors to phonemic awareness and early reading ability. In J. L. Metsala & L. C. Ehri (Eds.), *Word recognition in beginning literacy* (pp. 89–120). Mahwah, NJ: Erlbaum.

Moe, A. J., Hopkins, K. J., & Rush, R. T. (1982). *The vocabulary of first grade children*. Springfield, IL: Thomas.

Morrisette, M. L. (1999). Lexical characteristics of sound change. *Clinical Linguistics & Phonetics, 13*, 219–238.

Morrisette, M. L., & Gierut, J. A. (2002). Lexical organization and phonological change in treatment. *Journal of Speech, Language & Hearing Research, 45*, 143–159.

Morrison, C. M., Chappell, T. D., & Ellis, A. W. (1997). Age of acquisition norms for a large set of object names and their relation to adult estimates and other variables. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 3*, 528–559.

Murphy, H. A. (1957). The spontaneous speaking vocabulary of children in primary grades. *Journal of Education, 140*, 1–105.

Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 873–889.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). *Sizing up the Hoosier mental lexicon* (Research on Spoken Language Processing Rep. No. 10, pp. 357–376). Bloomington, IN: Indiana University, Speech Research Laboratory.

Pisoni, D. B., Nusbaum, H. C., Luce, P. A., & Slowiaczek, L. M. (1985). Speech perception, word recognition and the structure of the lexicon. *Speech Communication, 4*, 75–95.

Rice, M. L., Oetting, J. B., Marquis, J., Bode, J., & Pae, S. (1994). Frequency of input effects on word comprehension of children with specific language impairment. *Journal of Speech and Hearing Research, 37*, 106–122.

Rice, M. L., & Woodsmall, L. (1988). Lessons from television: Children's word learning when viewing. *Child Development, 59*, 420–429.

Snodgrass, J. G., & Yuditsky, T. (1996). Naming times for the Snodgrass and Vanderwart pictures. *Behavior Research Methods, Instruments, & Computers, 28*, 516–536.

Sommers, M. S., & Lewis, B. P. (1999). Who really lives next door: Creating false memories with phonological neighbors. *Journal of Memory and Language, 40*, 83–108.

Stoel–Gammon, C., & Cooper, J. A. (1984). Patterns of early lexical and phonological development. *Journal of Child Language, 11*, 247–271.

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

Storkel, H. L. (2002). Restructuring of similarity neighbourhoods in the developing mental lexicon. *Journal of Child Language, 29*, 251–274.

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

Storkel, H. L., & Gierut, J. A. (2002). Lexical influences on interword variation. In B. Skrabela, S. Fish, & A. H. J. Do (Eds.), *Proceedings of the 26th Annual Boston University Conference on Language Development* (Vol. 2, pp. 665–676). Sommerville, MA: Cascadilla Press.

Storkel, H. L., & Rogers, M. A. (2000). The effect of probabilistic phonotactics on lexical acquisition. *Clinical Linguistics & Phonetics, 14*, 407–425.

Vitevitch, M. S. (1997). The neighborhood characteristics of malapropisms. *Language and Speech, 40*, 211–228.

Vitevitch, M. S. (2002). The influence of phonological similarity neighborhoods on speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 735–747.

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

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., Pisoni, D. B., & Auer, E. T. (1999). Phonotactics, neighborhood activation, and lexical access for spoken words. *Brain and Language, 68*, 306–311.

Winters, J. J., Winter, L., & Burger, A. L. (1978). Confidence in age-of-acquisition estimates and its relationship to children's labeling performance. *Bulletin of the Psychonomic Society*, 361–364.