CAMBRIDGE
UNIVERSITY PRESS

BRIEF RESEARCH REPORT

Phonological neighborhood measures and multisyllabic word acquisition in children

Melissa RAJARAM

The University of Texas at Dallas
Address for correspondence: Melissa Rajaram, The University of Texas at Dallas, 1966 Inwood Rd, Dallas, TX, 75235 USA. E-mail: rajaram.melissa@gmail.com

Abstract

Multisyllabic words constitute a large portion of children's vocabulary. However, the relationship between phonological neighborhood density and English multisyllabic word learning is poorly understood. We examine this link in three, four and six year old children using a corpus-based approach. While we were able to replicate the well-accepted positive association between CVC word acquisition and neighborhood density, no similar relationship was found for multisyllabic words, despite testing multiple novel neighborhood measures. This finding raises the intriguing possibility that phonological organization of the mental lexicon may play a fundamentally different role in the acquisition of more complex words.

Introduction

The organization of known words in a child's mental lexicon is believed to influence the acquisition of new words. Specifically, when a novel sound sequence is heard, it is thought to activate phonological representations of similar-sounding "neighbor" words, which in turn supports the creation of a new lexical representation (Storkel, Bontempo, Aschenbrenner, Maekawa & Lee, 2013). This model is supported by reports that words with more phonological neighbors are easier to acquire than those with few phonological neighbors in the mental lexicon (Hoover, Storkel & Hogan, 2010; Storkel, 2004b, 2009; Storkel et al., 2013; Storkel & Lee, 2011). Conversely, by identifying aspects of phonological similarity that predict ease of acquisition, we can gain a better understanding of the phonological organization of the mental lexicon.

The vast majority of the work in this area comes from studying the impact of neighborhood density on the acquisition of CVC words using behavioral experiments. For CVC words, the neighborhood density is typically defined as the number of words differing by the substitution, omission, or addition of a single phoneme (Luce & Pisoni, 1998). In behavioral experiments, children are taught

CrossMark

tightly controlled groups of CVC nonwords with either many neighbors (i.e., from dense neighborhoods) or few neighbors (i.e., from sparse neighborhoods). Children generally learn more words from dense neighborhoods than from sparse neighborhoods (Hoover et al., 2010; Storkel et al., 2013; Storkel & Lee, 2011), which is attributed to increased activation, spreading from a larger number of neighbors, easing the acquisition of new words (Storkel & Lee, 2011). This "spreading activation" hypothesis is also supported by corpus-based studies, where words with greater neighborhood density are learned at younger ages (Storkel, 2004a, 2009). Similarly, children with small vocabularies acquired a greater proportion of high density words, which is consistent with increased activation making these words more salient to children (Stokes, Kern & Dos Santos, 2012). We note that these trends are most pronounced for long-term word learning, e.g., when word acquisition is measured after a 1 week delay, in contrast to measuring word acquisition immediately after training, i.e., short-term learning (Storkel & Lee, 2011). While neighborhood density is not the primary determinant of long-term word acquisition (constituting a 5% main effect in CVC words; Storkel & Lee (2011)), the relationship is well-established and underpins our understanding of the organization of the mental lexicon.

However, a blind-spot of the field is multisyllabic words, which we define as words containing two or more syllables. Despite constituting a large portion of children's vocabulary (Kuperman, Stadthagen-Gonzalez & Brysbaert, 2012) and being well represented in parent-reported vocabulary checklists for young children (e.g., Fenson et al., 1994; Rescorla, 1989), there are virtually no studies of the effect of phonological neighbors in the acquisition of multisyllabic English words. Thus, it is an open question whether the effect of neighborhood density, and by extension the theory of spreading activation, generalizes to long-term multisyllabic word acquisition.

The closest related work on multisyllabic neighborhoods comes from studies of children's word acquisition in non-English inflectional languages, and word recognition in both English and inflectional languages. In inflectional languages, the effect of neighborhood density in multisyllabic acquisition was found to largely mirror that for CVC words. Children were more likely to correctly inflect multisyllabic words with greater neighborhood density during past tense verb inflection (Kirjavainen, Nikolaev & Kidd, 2012; Ragnarsdóttir, Simonsen & Plunkett, 1999), noun inflection (Granlund et al., 2019; Savičiute, Ambridge & Pine, 2018), and case inflection (Dąbrowska & Szczerbiński, 2006). However, these studies examine the process of acquiring word inflections through a different theoretical lens, e.g., rule-based or analogy-based approaches to inflectional morphology (Granlund et al., 2019). It is unclear if trends in this inflectional approach to word acquisition would generalize to English, which lacks a complex inflectional system.

The effect of neighbors on word recognition shows a more complicated dependence on the inflectional status of the language than word acquisition. In inflectional languages, multisyllabic words from dense neighborhoods were recognized more quickly than words from sparse neighborhoods in both children and adults (Arutiunian & Lopukhina, 2020; Vitevitch & Stamer, 2006). In contrast, English words from dense neighborhoods are recognized more slowly and less accurately than words from sparse neighborhoods, both in adults and children (e.g., Garlock, Walley & Metsala, 2001), thus inverting the trend in acquisition. Given the obvious difference in underlying processes, and the mismatch in ages analyzed, we use the English studies with multisyllabic words (Suárez, Tan, Yap & Goh, 2011; Vitevitch, Stamer & Sereno, 2008) primarily as guidance on the quantification of neighborhoods. Thus, there is a

lack of understanding as to how an English multisyllabic word's neighborhood density impacts its ease of acquisition.

What defines a multisyllabic word's neighborhood? This is the key question that must be answered to extend studies of neighborhood density to multisyllabic words. Multisyllabic words exhibit types of variation fundamentally missing in CVC words, be it differing lengths of phonemes, multiple options for syllables to be compared, and (in English) lexical stress. In English, the traditional neighborhood measure often behaves unintuitively, e.g., parrot (/pɛrət/) and pear (/pɛr/) can never be in the same neighborhood. Such problems become especially acute as words get longer, a pattern that is also apparent in studies of word recognition in adults. For RECOGNIZING shorter bisyllabic words (three to five phonemes), the original neighborhood density (Luce & Pisoni, 1998) appears to capture the expected behavior in dense/sparse neighborhoods (Vitevitch et al., 2008). However, for English multisyllabic words with greater than 6 phonemes, this definition of neighborhoods yields virtually no neighbors (Storkel, 2004b), suggesting that either phonological neighbors have little role in acquisition of these words, or that we need a different measure of neighborhood density. Indeed, in word recognition for longer multisyllabic words with no traditional neighbors, a new measure based on the Phonological Levenshtein Distance (PLD20) was required to capture the expected behavior (Suárez et al., 2011), suggesting that novel neighborhood measures are needed for English multisyllabic words.

In our study, we explore neighborhood density measures that differ from the original (Luce & Pisoni, 1998) in two broad directions by 1) relaxing how similar two words need to be deemed neighbors (i.e., a difference of one phoneme is perhaps overly stringent), and/or 2) generalizing the concept of phonological similarity by incorporating phonological units other than phonemes (E. Bates et al., 1994). In addition to the original neighborhood density measure, we study three different measures of phonological neighborhoods (see Methods for exact definitions). The first neighborhood measure we considered was the PLD20 used in word recognition (Suárez et al., 2011), which incorporates the "edit distance" between word's phonemes rather than using a harsh cutoff. Second, we created a novel neighborhood measure by modifying the PLD20 above to use a sub-phonemic (Bailey & Hahn, 2005) edit distance instead of one based on phonemes. Third, inspired by the strong effect that the syllable stress has in children's word retrieval (Cutler, 2005) and recognition (Cooper et al., 2002), we designed a neighborhood measure using the suprasegmental phonological unit of the onset and nucleus (Marslen-Wilson & Zwitserlood, 1989) of the stressed syllable.

We investigated the association between these neighborhood measures and multisyllabic word acquisition in three, four and six year old children. Given the exploratory nature of this study, it was impractical to devote the extensive resources needed to perform behavioral experiments across multiple neighborhood measures (which would presumably need different sets of children). We therefore chose to adopt a corpus-based approach analyzing conversational transcripts from children at three, four and six-year-old of age (Paradise et al., 2005, 2003; Paradise et al., 2001). We incorporated two innovations to mitigate the indirectness of a corpus-based approach. First, analogous to the way in which nonwords control for ambient word frequency in behavioral experiments, we designed a novel acquisition measure, the Proxy for Acquisition from Conversational Transcripts (PACT), which statistically controlled for ambient frequency in children's word use. Second, to ensure that our results were not purely an outcome of our novel analysis methodology, we performed

a parallel analysis on CVC words, in effect a "known standard" (Baker & Dunbar, 2000) against which to interpret our multisyllabic results. We hypothesize that, like CVC words and multisyllabic words in inflectional languages, multisyllabic words from dense neighborhoods will be easier to acquire than those from sparse neighborhoods, for at least one of our neighborhood measures.

Method
Participants
Analyses were based on orthographic transcripts of child-caregiver conversations that were audio-recorded as a part of a longitudinal study of child development and middle ear effusion (Paradise et al., 2005, 2003; Paradise et al., 2001). Children in the cohort ($N = 752$) were demographically representative of the greater Pittsburgh, Pennsylvania area, were singleton births free of medical conditions and risk factors, and were from monolingual American English homes. Conversations were approximately 15 minutes long and occurred within two months of each child's third, fourth, and sixth birthdays (Paradise et al., 2005, 2003; Paradise et al., 2001).

Procedure
Children and adult caregivers played with a consistent set of toys; caregivers were instructed to "play and talk with your child as you would at home". Recordings were transcribed orthographically and coded for children's use of inflectional morphemes (i.e., plural -s, regular past tense -ed, and progressive aspect -ing) by trained research assistants using the Systematic Analysis of Language Transcripts (SALT) software (Miller & Iglesias, 2012). The number of digital transcript files available at each age varied due to sample attrition, equipment failure, or examiner error. We analyzed all available digitized transcripts for a total of 747 transcripts at age three, 683 transcripts at age four, and 696 transcripts at age six.

Analysis
Phonological word forms
To derive phonological word forms, we first compiled orthographic words used in the transcripts by children and adults with SALT's "Root Word List" function (Miller & Iglesias, 2012). We then removed orthographic words that were closed class words (e.g., Goodman, Dale & Li, 2008) or that had no corresponding entry in the CMU Pronunciation Dictionary (Weide, 2014). Trained raters then excluded words that either did not correspond to a standard American English form, were apparently misspelled, or were definite descriptions, such as proper names (Weizman & Snow, 2001). We next reduced words with English inflectional suffixes to their root forms to eliminate words that were multisyllabic only because of morphological operations (e.g., *biggest*). The remaining orthographic words were translated into Klattese phonological forms (Luce & Pisoni, 1998) based on the CMU Pronunciation Dictionary (Weide, 2014), yielding the unique phonological forms at each age (Table 1).

Neighborhood measures
The neighborhood measures were calculated with respect to the phonological forms spoken by children (Table 1), i.e., the expressive lexicons. We chose to use the

Table 1. Number of transcripts, total unique phonological forms spoken by children, and the subsets of phonological forms spoken by children and adults: all forms, multisyllabic and CVC, by age. The multisyllabic words ranged between 2 and 5 syllables.

Age in Years	Transcripts analyzed	Spoken by children	Spoken by *both* children and adults		
		All Forms	All Forms	Multisyllabic forms	CVC forms
Three	747	1,557	1,397	644	386
Four	683	1,687	1,495	717	393
Six	696	2,260	1,939	983	483

expressive lexicons because we did not have access to the children's entire lexicons, and neighborhood density values from children's EXPRESSIVE lexicons (e.g., Storkel & Hoover, 2010) have been shown to be associated with word acquisition (Hoover et al., 2010; Storkel & Lee, 2011). The neighborhood measures described below were calculated separately on the forms used at three, four, and six years (additional descriptive statistics in Supplementary Materials, Appendix A).

Original Neighborhood Density (ND).

The original ND was calculated by counting the number of words that differ by the additional, deletion or substitution of one phoneme (Luce & Pisoni, 1998). The original ND values for CVC words (Figure 1A) were used in the "known standard" analysis, and the original ND values for multisyllabic words (Figure 1 B1) were used in the multisyllabic analysis.

Phonological Levenshtein Distance Neighborhood (PLD20).

The PHONOLOGICAL LEVENSHTEIN DISTANCE (PLD20) was calculated by first determining the number of phoneme edits (i.e., phoneme substitutions, insertions, or deletions) to transform a multisyllabic word into all other words at a given age. For example, one substitution and two insertion edits are needed to transform *bear* (/bɛr/) into *parrot* (/pɛrət/). Next, to be consistent with prior work (i.e., Suárez et al., 2011), the mean of the 20 closest Levenshtein neighbors was calculated for each multisyllabic word to produce PLD20 values. Multisyllabic forms with lower PLD20 values were more similar to their 20 closest Levenshtein neighbors than multisyllabic forms with greater PLD20 (Figure 1B2).

Phoneme Feature Distance (P-FEAT20).

The PHONEME FEATURE DISTANCE (P-FEAT20) was calculated by first determining the average position-specific phoneme feature distances between a multisyllabic word and all other words at a given age. To find the phoneme distance between either consonants or vowels (i.e., not between a consonant and a vowel), words were aligned based on their stressed, or only, syllable, which was obtained from the CMU Pronunciation Dictionary (Weide, 2014). Consonants and vowels were each represented by four subsegmental features (see Supplementary Materials, Appendix B). Consonant features included place, manner, voicing, and sonority-obstruent (Bailey & Hahn, 2005). Vowel features included height, front-back, roundness, and
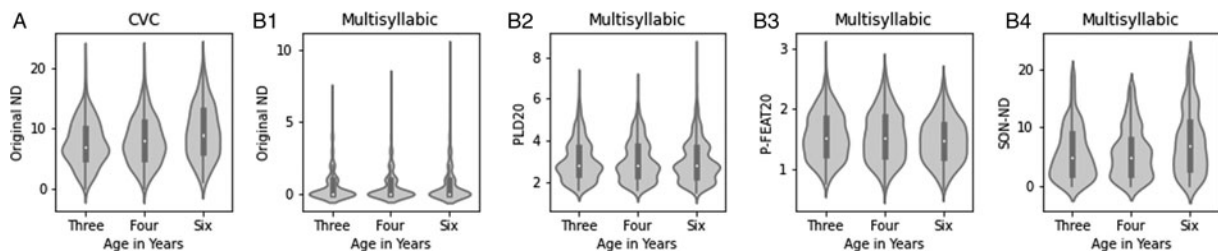
Figure 1. Violin plots of neighborhood measures for CVC (A) and multisyllabic (B1-B4) words used by children and adults at ages three, four and six years. Original ND (B1) produces few neighbors for multisyllabic words (no neighbors for 69.1% at age three, 68.9% at age four and 67.4% at age six years), while the alternative neighborhood measures (B2-B4) produce more varied distribution of values.

tenseness (International Phonetic Association, 1999). When a phoneme in one position could not be compared to a corresponding phoneme in the same position, e.g, in the case of finding the PFEAT distance between a multisyllabic and monosyllabic word, we considered all four phoneme features to be different.

In contrast to the PLD20 neighborhood measure, the inclusion of phoneme features in the PFEAT20 measure allows us to set a "weight" for phoneme substitutions in words based on their featural distance, i.e., the number of differing features. For example, the words *tin* (/tɪn/) and *pin* (/pɪn/) would have a smaller pairwise PFEAT distance than the words *tin* (/tɪn/) and *win* (/wɪn/) because /t/ and /p/ differ by one feature, while /t/ and /w/ differ by four features (see Supplementary Materials, Appendix B). After calculating pairwise PFEAT distance between words, to be consistent with the PLD20 measure (Suárez et al., 2011), the mean of the 20 closest Phoneme Feature Difference neighbors was calculated. Multisyllabic forms with a P-FEAT20 value closer to zero were more similar to their 20 closest phoneme feature distance neighbors, while multisyllabic forms with a P-FEAT20 values closer to four were less similar to their 20 closest neighbors (Figure 1 B3).

Stress: Onset Nucleus Neighborhood Density (SON-ND).

THE STRESS: ONSET NUCLEUS NEIGHBORHOOD DENSITY (SON-ND) was calculated for multisyllabic words by counting the number of words that contained identical onsets and nuclei of their stressed, or only, syllable. The CMU Pronunciation Dictionary (Weide, 2014) provided the vowel containing primary lexical stress, and syllable boundaries were placed with a computational implementation of the Maximal Onset Principle (see Supplementary Materials, Appendix A for github repository). Multisyllabic forms with higher SON-ND values had common onset-nucleus sequences, while multisyllabic forms with lower SON-ND values had less common onset-nucleus sequences (Figure 1 B4).

Acquisition measure

The PROXY FOR ACQUISITION FROM CONVERSATIONAL TRANSCRIPTS (PACT) served as the measure of word acquisition. Given that the amount of ambient adult word use can influence child word use (Ambridge, Kidd, Rowland & Theakston, 2015; Goodman et al., 2008), we accounted for this factor in our dependent measure to produce a more accurate measure of word difficulty. Briefly, we followed Goodman et al. (2008), who suggested that the residual remaining after partialing out the contribution of ambient frequency "may be conceptualized as a measure of the difficulty of a word." Thus, words that children use more often than expected based on their ambient frequency are considered relatively "easier" to acquire, while words that children use less often than expected based on ambient frequency are considered relatively "harder" to acquire (Table 2). We note that we are unable to control for when a word was learned relative to its usage. This distinction could be important since depending on the stage of learning (short vs long term), neighborhood density is thought to show a different impact on acquisition (Storkel & Lee, 2011). However, as children only learn a small proportion of their vocabulary on a given day (Bloom & Markson, 1998), it is unlikely that multiple children had just learnt exactly the same word in our relatively short sessions. So, the PACT score, which integrates over all the words and children, is likely to reflect the majority of words that were not learnt in the immediate past and to act as an ambient-language corrected measure of long term learning.

Table 2. Examples of CVC and multisyllabic words with low and high PACT values across ages. Low PACT words were less than 1.75 SD below the mean, while high PACT value words were greater than 1.75 SD above the mean.

Age	Low PACT ("harder" to acquire)	High PACT ("easier" to acquire)
Three Years	fawn, fig, buck excavator, finesse, aerial	ham, duck, pig carrot, welcome, ladder
Four Years	nerve, nag, league network, incline, disposal	walk, laugh, peach mother, spider, airplane
Six Years	pike, doubt, jig mignon, grimace, pesticide	coat, dig, bowl puppy, toilet, kitten

To create the PACT values, we determined the best-fit relationship between the child and adult frequency for all words that occurred in both sets of transcripts at each age (Table 1) and then calculated the residuals. We used the percentage of transcripts that contained a phonological form as the measure of frequency to make the PACT values insensitive to constant repetition of a word from one particular child or adult. The SALT Root Word List's "%Transcript" column provided the percentage of child or adult transcripts that contained an orthographic word form. For the majority of words, there was a one-to-one correspondence between the orthographic form and phonological word form. In cases where two or more orthographic forms corresponded to one phonological form (e.g., *right/write*), we retained the highest of the %Transcript values. Both child %Transcript and adult %Transcript values were highly skewed and transformed with a natural logarithm (Goodman et al., 2008). Visual inspection suggested that a quadratic fit better captured the relationship than a linear fit (Figure 2A), and residuals for each word were calculated to the best-fit curve, yielding the PACT value. Positive residuals suggested a word was "easier" to acquire, while negative residuals suggested a word was "harder" to acquire relative to other words at age.

The PACT is intended to quantify relative difficulty of acquisition at a particular age rather than absolute difficulty across ages. For example, even though words tend to be more easily acquired with age, the average PACT value across all ages is zero (Table 3), with the "main effect" of age effectively removed. Nonetheless, individual subsets of words can show interesting trends within and across ages. Within each age, the PACT values for CVC words were positive, while the multisyllabic words were negative, reflecting that the CVC words were consistently easier to acquire than the multisyllabic forms at all ages (Table 3.). Across ages, following a fixed set of multisyllabic words (Figure 2C), we saw a clear trend where the PACT values increased from age three (mean = −0.01) to age four (mean = 0.03) to age six (mean = 0.13). This suggests that this subset of words went from being harder than the typical word at age three to easier than the typical word at age six. Thus, while comparing PACT across ages can be meaningful, care must be taking in interpreting the observed results.

Statistical analyses

The CVC and multisyllabic data were analyzed in R (R Core Team, 2017) using linear mixed models (lme4; Bates, Mächler, Bolker & Walker, 2015). In both analyses, the dependent measure was the PACT values. Fixed effects included the children's age
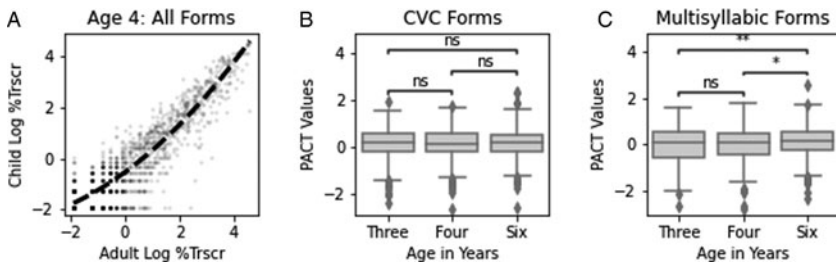
Figure 2. When creating Proxy for Acquisition from Conversational Transcripts (PACT) values, A) the best fit quadratic curve is found between the logarithm of the percent transcripts for children and adults at each age (shown only at age four years). When examining the distributions of PACT values from words in common across all ages using a Wilcoxon test (with Bonferroni correction), (B) the CVC PACT values ($n = 319$) did not differ significantly across ages, (C) but the multisyllabic PACT values ($n = 434$) differed between ages four and age six years, and between ages three and age six years. NOTE: ns = not significant, $*p < .05$, $**p < .01$.

Table 3. PACT values (and Standard Deviation) for different subsets of words.

Word Type	Age Three	Age Four	Age Six
All Words	0.00 (0.68)	0.00 (0.65)	0.00 (0.67)
CVC	0.10 (0.68)	0.04 (0.70)	0.08 (0.68)
Multisyllabic	−0.06 (0.68)	−0.05 (0.61)	−0.07 (0.67)

and the neighborhood measure(s). Age was entered as a categorical variable, while the neighborhood measure(s) were continuous variables. We included a random intercept for words to control for the fact that different target words could have intrinsic effects on acquisition (e.g., words vary based on semantics, number of syllables, phonemes, etc.). The interaction of age and neighborhood measure(s) was also included in both analyses. Statistical significance for each factor was determined using Satterthwaite's method (lmerTest; Kuznetsova, Brockhoff & Christensen, 2017). We note that our model was primarily designed to allow us to test the effect of different neighborhood measures on the PACT, and not the effect of age, whose interpretability is limited in the present construction. The output of the statistical analysis of our model produces three types of terms:

1. Estimates of the effect of the neighborhood measures. These are our primary focus and represent the overall "slope" (across all ages) for that measure. A significant effect implies that variation in the measure could explain differences in acquisition.
2. Interaction terms of the neighborhood measures with age: these represent how the "slopes" for a measure change with age.
3. Intercepts of age: These reflect the overall offset needed at each age after accounting for the fits above.

Thus, the effect of age is spread out over the second and third types of terms. For example, a significant age offset may arise from a uniform shift in PACT scores at

that age, but may also reflect changes in neighbor effects at that age (i.e., interaction term), which in turn, requires a change in offset to fit the data. Because age effects are not the primary goal of this study, not to mention the complexity of deconvolving these effects along with the subtleties of the PACT score (discussed above), we shall largely refrain from interpreting any significant effect in these terms.

Next, to examine if multicollinearity might affect the results of the multisyllabic mixed linear model, we calculated the Variance Inflation Factor (VIF) of the neighborhood measures (usdm; Naimi, Hamm, Groen, Skidmore & Toxopeus, 2014). We chose a conservative VIF threshold of four (O'Brien, 2007), and consider values less than this to suggest that the model coefficients were not poorly estimated due to multicollinearity. Finally, we completed two sets of univariate analyses for the CVC and multisyllabic words. At each age, we used ordinary least squares regression to determine the relationship between a single neighborhood measure and the PACT values.

Results

Table 4 presents the CVC mixed linear effects analysis. There were 539 unique CVC words and a total of 1,262 observations across the three ages. Not all words were present at every age. The fixed effect of original ND was statistically significant ($p = 0.032$), but neither the fixed effect of age, nor any interactions between age and original ND were found to be significant. See Supplementary Materials, Appendix C for additional CVC analyses with the alternative neighborhood measures.

Table 5 presents the multisyllabic mixed linear effects analysis. There were 1,254 unique multisyllabic words with a total of 2,344 observations across the three ages. Not all words were present at every age. None of the fixed effects of neighborhood measures were statistically significant, and no interactions between age and neighborhood measures were statistically significant. While the intercept at age three was statistically significant ($p = 0.036$), as discussed in the methods, this term is reflective of the interplay of age with the PACT scores and the neighborhood measures, rather than the effect of neighborhood measures with PACT scores which is the primary focus of this study. All VIF values for the neighborhood measures were below the threshold of 4 (i.e., 1.4 for ND, 1.0 for SOND, 2.4 for PLD20 and 2.1 for PFEAT20), suggesting that the coefficients were not poorly estimated due to

Table 4. Summary of the CVC linear mixed model analysis

Predictor	Estimate	95% CI	SE	t value	Pr(>|t|)
Intercept	−0.144	−0.284 to -0.004	0.072	−2.009	0.045 *
Age(6)	0.075	−0.068 to 0.218	0.073	1.031	0.303
Age(3)	0.004	−0.141 to 0.150	0.074	0.059	0.953
ND	0.017	0.002 to 0.033	0.008	2.151	0.032 *
Interactions					
Age(6):ND	−0.003	−0.018 to 0.012	0.008	−0.391	0.696
Age(3):ND	0.006	−0.011 to 0.023	0.009	0.697	0.486

NOTES: Formula used in R: PACT ~ age * ND + (1|word). The random effect for word accounted for 27% of the variance in the model. Reference level for Age = Four (4). *p < .05.

Table 5. Summary of the multisyllabic linear mixed model analysis.

Predictor	Estimate	95% CI	SE	t value	Pr(>|t|)
Intercept	−0.089	−0.310 to 0.132	0.113	−0.785	0.433
Age(6)	0.205	−0.042 to 0.452	0.126	1.624	0.105
Age(3)	0.290	0.020 to 0.56	0.138	2.098	0.036*
ND	0.024	−0.024 to 0.071	0.024	0.970	0.332
SOND	0.001	−0.010 to 0.011	0.005	0.104	0.917
PLD20	−0.008	−0.079 to 0.062	0.036	−0.232	0.817
PFEAT20	−0.004	−0.162 to 0.155	0.081	−0.043	0.965
Interactions					
Age(6):ND	−0.015	−0.065 to 0.034	0.025	−0.608	0.543
Age(3):ND	−0.049	−0.107 to 0.008	0.029	−1.684	0.092
Age(6):SOND	−0.001	−0.011 to 0.010	0.005	−0.171	0.864
Age(3):SOND	0.003	−0.010 to 0.015	0.006	0.427	0.669
Age(6):PLD20	−0.057	−0.136 to 0.021	0.040	−1.427	0.154
Age(3):PLD20	−0.022	−0.107 to 0.063	0.044	−0.507	0.612
Age(6):PFEAT20	0.010	−0.167 to 0.188	0.091	0.115	0.908
Age(3):PFEAT20	−0.148	−0.339 to 0.042	0.098	−1.521	0.129

NOTES: Formula used in R: PACT ∼ age * (ND + SOND + PLD20 + PFEAT20) + (1|word). The random effect for word accounted for 21% of the variance in the model. Reference level for Age = Four (4). *$p < .05$.

multicollinearity. See Supplementary Materials, Appendix E for an additional modeling approach to address multicollinearity.

The CVC and multisyllabic univariate analyses mirrored the results from the mixed model analyses. The CVC analyses showed a statistically significant relationship between the original neighborhood density and PACT values at all ages. In addition, the variance captured in these analyses was comparable to the 5% main effect reported by Storkel and Lee (2011): 2.1% at age three, 1.5% at age four, and 1.2% at age six years. The multisyllabic analyses did not display a consistent association between any of the novel neighborhood measures and word acquisition. See Supplementary Materials, Appendix D for more details.

Discussion

In the current study, we examined the relationship between phonological neighborhoods and multisyllabic word acquisition in English. We hypothesized that, like CVC words, multisyllabic words from dense phonological neighborhoods would be easier to acquire than multisyllabic words from sparse neighborhoods. A potential complication is that multisyllabic words show types of variation (e.g., number of phonemes and lexical stress) that are absent in CVC words, and it is thus unclear whether the measure of neighborhood density used for CVC words (Luce & Pisoni, 1998) would generalize to multisyllabic words. We therefore created three additional, multisyllabic-specific,

neighborhood measures and sought to test which (if any) of these are associated with multisyllabic acquisition.

As it would be impractical to test multiple neighborhood measures using behavioral experiments, we developed a corpus-based approach using conversational transcripts from children. Our analysis was designed to overcome the indirect nature of a corpus-based approach where, in contrast to behavioral experiments, confounding factors can only be controlled for post-data-collection. First, we created a new measure of word acquisition, i.e., the PACT, which corrects for the impact of ambient adult language on children's word use. Next, we used mixed linear models for statistical analysis, which allowed us to simultaneously test the contribution of neighborhood effects based on multiple measures at multiple ages. Finally, to ensure that our results were not purely an outcome of our novel analysis methodology, we performed a parallel analysis on CVC words – in effect, a "known standard" (Baker & Dunbar, 2000) against which to interpret our multisyllabic results.

These analyses were performed on conversational transcripts from children at three, four and six years. We found that our PACT measure captured some expected aspects of ease of acquisition: within each age CVC words are relatively "easier" to acquire than multisyllabic words (Table 3), and multisyllabic words became relatively easier to acquire as children develop (Figure 2). The mixed linear model analysis did not find a significant relationship between multisyllabic acquisition and any neighborhood measure, nor any significant interactions between the neighborhood measures and the children's age. In fact, the only significant effect was the intercept term at age three which, as discussed in the methods, is hard to interpret. It is difficult to deconvolve contributions from age on PACT scores versus age on neighborhood measures, and, in any case, neither is the primary interest in this study. In contrast to the multisyllabic analysis, the CVC "known standard" analysis revealed a statistically significant relationship between acquisition and neighborhood density, consistent with previous literature (Hoover et al., 2010; Storkel et al., 2013; Storkel & Lee, 2011). Additionally, to get a sense of the magnitude of the effect in our CVC analysis, we performed three separate univariate analyses (see Supplementary Materials, Appendix D), where we captured a comparable amount of variance as reported by Storkel and Lee (2011). Thus, despite testing multiple neighborhood measures and replicating the well-accepted relationship between neighborhood density and CVC acquisition, we found no support for a similar association in multisyllabic words.

One possible interpretation for the discrepancy between the CVC and multisyllabic results is that this is an artifact of the multisyllabic measures we chose to use. In other words, dense neighborhoods do in fact support multisyllabic acquisition in English, but the measures of neighborhood density we tested were simply unable to capture this relationship. Given that these neighborhood measures were chosen to encompass (what we considered to be) plausible phonological relationships, future discovery of a "true" neighborhood measure (i.e., predictive of multisyllabic acquisition) would likely expand the aspects of phonology thought to impact the organization of the mental lexicon. If future studies fail to find a significant effect of neighborhood density, we will have to consider the possibility that there is in fact no neighborhood measure for which dense neighborhood support multisyllabic acquisition. The model of lexical acquisition used here posits that a) words exhibit some relative organization in the mental lexicon, and b) words with more neighbors in this organization are more easily acquired due to shared activation between neighbors. Thus, if words in dense phonological neighborhoods are not easier to

acquire, it would suggest that phonology has a limited role to play in the organization of multisyllabic words. In sum, our study suggests that we must either expand our concept of neighborhoods for multisyllabic English words or that their process of acquisition differs from CVC words.

Another perplexing question is how to reconcile the limited effect of multisyllabic neighborhoods we find in the acquisition of English words with the more pronounced role found in acquisition of inflection languages and in word recognition. Differences in both the methodology and language inflection status relative to our study means the results are not directly comparable. For example, studies from inflectional languages that show a positive effect of neighborhood density use a distinct measure of acquisition that emphasizes learning the correct inflectional suffixes rather than whole words (e.g., Granlund et al., 2019). In word recognition, there is an even more complex pattern: English studies suggest a negative impact of neighborhood density, but in adults rather than children (Suárez et al., 2011; Vitevitch et al., 2008), while studies in inflectional languages show a positive impact of neighborhood density in both adults and children (Arutiunian & Lopukhina, 2020; Vitevitch & Stamer, 2006). It is unknown if these disparities represent a fundamental difference in mental processes between multisyllabic recognition and acquisition. Further studies are needed to resolve the complex interaction between the degree of language inflection, word length, word recognition and word acquisition, and the implications of such a finding.

We acknowledge our study has several limitations, in scope and methodology, which impact the interpretation of our results. First, ease of word acquisition is affected by many phenomena, and phonological neighbor density accounts for only a small proportion of the variance. Thus, rather than accurately predicting ease of acquisition, we aimed to identify neighborhood measures significantly correlated with acquisition as a means to study how the spreading activation model applies to multisyllabic word learning. One challenge common to all studies in the field is the impossibility of determining the entire vocabularies of children. Following past work (Storkel & Hoover, 2010), we defined our neighborhood measures based solely on words spoken by children, yet it is an open question whether using all known words would yield different conclusions. Another challenge in the field is the impossibility of directly observing or perturbing word activations to infer a causal effect on acquisition. Instead, behavioral experiments, the gold standard in the field, use nonword controls (for confounding effects like frequency) with sparse/dense neighborhood densities (a proxy for activation) to test for a statistical effect on acquisition. Even though we statistically controlled for the effect of word frequency, we acknowledge that our corpus-based-approach is even less direct. Our methodology is thus more susceptible to other confounding effects (e.g., phonotactics), which could potentially compete with the effect of neighborhoods density, which is itself small. Moreover, unlike behavioral experiments, we were unable to control for when a word was learned, and it is possible that a few recently learned words may be included. This may lead to further diluting our signal in light of the intriguing finding that neighborhood density affects short-term and long-term learning differently (Storkel & Lee, 2011). For these reasons, it is conceivable that our methodology contributed to the lack of a significant effect of multisyllabic neighborhood density and that future studies with more powerful approaches may yet uncover a significant effect with these same neighborhood measures. Nonetheless, our methodology was able to recover the effect of neighborhood density on CVC

acquisition with effect sizes comparable to that published. Thus, barring any unforeseen confounds specific to multisyllabic words, our results suggest that neighborhood density (or at least measures considered here to quantify it) has a larger role in the acquisition of CVC words as compared to multisyllabic words.

In conclusion, we developed a corpus-based approach to examine the relationship between multiple neighborhood measures and multisyllabic acquisition in three, four and six year old children. While we were able to replicate the relationship between neighborhoods and CVC acquisition, we were unable to detect a relationship between neighborhood measures and multisyllabic acquisition. These results suggest that multisyllabic words might be organized based on as of yet undiscovered phonological relationships in the mental lexicon, or, alternatively, that a multisyllabic word's phonological characteristics have a limited role in the organization of words in the mental lexicon. Regardless, as multisyllabic words are a substantial portion of the words children acquire, this work highlights the need for specific studies of neighborhood measures and multisyllabic acquisition in English.

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S0305000920000811

References

Ambridge, B., Kidd, E., Rowland, C. F., & Theakston, A. L. (2015). The ubiquity of frequency effects in first language acquisition. *Journal of Child Language*, *42*(2), 239–273. https://doi.org/10.1017/S030500091400049X

Arutiunian, V., & Lopukhina, A. (2020). The effects of phonological neighborhood density in childhood word production and recognition in Russian are opposite to English. *Journal of Child Language*, (May). https://doi.org/10.1017/S0305000920000112

Bailey, T. M., & Hahn, U. (2005). Phoneme similarity and confusability. *Journal of Memory and Language*, *52*(3), 339–362. https://doi.org/10.1016/j.jml.2004.12.003

Baker, L. M., & Dunbar, K. (2000). Experimental design heuristics for scientific discovery: the use of "baseline" and "known standard" controls. *International Journal of Human Computer Studies*, *53*(3), 335–349. https://doi.org/10.1006/ijhc.2000.0393

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting Linear Mixed-Effects Models Using lme4, 67(1). https://doi.org/10.18637/jss.v067.i01

Bates, E., Hartung, J., Marchman, V., Thal, D., Fenson, L., Reilly, J., … Reznick, J. S. (1994). Developmental and Stylistic Variation in the Composition of Early Vocabulary. *Journal of Child Language*, *21*(1), 85–123. https://doi.org/10.1017/S0305000900008680

Bloom, P., & Markson, L. (1998). Capacities underlying word learning. *Trends in Cognitive Sciences*, *2*(2), 67–73. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/21227068

Cooper, N., Cutler, A., Wales, R., Cooper, N., Cutler, A., & Wales, R. (2002). Language and Speech Constraints of Lexical Stress on Lexical Access in English : and Non-native Listeners. *Language and Speech*, *45*(3), 207–228. https://doi.org/10.1177/00238309020450030101

Cutler, A. (2005). Lexical Stress. In *The Handbook of Speech Perception* (p. 144). https://doi.org/10.1063/1.3033202

Dąbrowska, E., & Szczerbiński, M. (2006). Polish children's productivity with case marking: The role of regularity, type frequency, and phonological diversity. *Journal of Child Language*, *33*(3), 559–597. https://doi.org/10.1017/S0305000906007471

Fenson, L., Dale, P. S., Reznik, 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*. Retrieved from http://psycnet.apa.org/psycinfo/1995-20882-001

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*(3), 468–492. https://doi.org/10.1006/jmla.2000.2784

Goodman, J. C., Dale, P. S., & Li, P. (2008). Does frequency count? Parental input and the acquisition of vocabulary. *Journal of Child Language*, *35*(3), 515–531. https://doi.org/10.1017/S0305000907008641

Granlund, S., Kolak, J., Vihman, V., Engelmann, F., Lieven, E. V. M., Pine, J. M., … Ambridge, B. (2019). Language-general and language-specific phenomena in the acquisition of inflectional noun morphology: A cross-linguistic elicited-production study of Polish, Finnish and Estonian. *Journal of Memory and Language*, *107*(May 2018), 169–194. https://doi.org/10.1016/j.jml.2019.04.004

Hoover, J. R., Storkel, H. L., & Hogan, T. P. (2010). A cross-sectional comparison of the effects of phonotactic probability and neighborhood density on word learning by preschool children. *Journal of Memory and Language*, *63*(1), 100–116. https://doi.org/10.1016/j.jml.2010.02.003.A

International Phonetic Association. (1999). Handbook of the International Phonetic Association: A guide to the use of the International Phonetic Alphabet. Cambridge University Press.

Kirjavainen, M., Nikolaev, A., & Kidd, E. (2012). The effect of frequency and phonological neighbourhood density on the acquisition of past tense verbs by Finnish children. *Cognitive Linguistics*, *23*(2), 273–315. https://doi.org/10.1515/cog-2012-0009

Kuperman, V., Stadthagen-Gonzalez, H., & Brysbaert, M. (2012). Age-of-acquisition ratings for 30,000 English words. *Behavior Research Methods*, *44*(4), 978–990. https://doi.org/10.3758/s13428-012-0210-4

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models . *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

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

Marslen-Wilson, W., & Zwitserlood, P. (1989). Accessing Spoken Words: The Importance of Word Onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(3), 576–585. https://doi.org/10.1037/0096-1523.15.3.576

Miller, J. F., & Iglesias, A. (2012). Systematic Analysis of Language Transcripts (SALT) Research Version. *Computer Program*.

Naimi, B., Hamm, N. A. S., Groen, T. A., Skidmore, A. K., & Toxopeus, A. G. (2014). Where is positional uncertainty a problem for species distribution modelling? *Ecography*, *37*(2), 191–203. https://doi.org/10.1111/j.1600-0587.2013.00205.x

O'Brien, R. M. (2007). A caution regarding rules of thumb for variance inflation factors. *Quality and Quantity*, *41*(5), 673–690. https://doi.org/10.1007/s11135-006-9018-6

Paradise, J. L., Campbell, T. F., Dollaghan, C. A., Feldman, H. M., Bernard, B. S., Colborn, D. K., … Smith, C. G. (2005). Developmental Outcomes after Early or Delayed Insertion of Tympanostomy Tubes. *New England Journal of Medicine*, *353*(6), 576–586. https://doi.org/10.1056/NEJMoa050406

Paradise, J. L., Dollaghan, C. A., Campbell, T. F., Feldman, H. M., Bernard, B. S., Colborn, D. K., … Smith, C. G. (2003). Otitis Media and Tympanostomy Tube Insertion During the First Three Years of Life: Developmental Outcomes at the Age of Four Years. *Pediatrics*, *112*(2), 265–277. https://doi.org/10.1542/peds.112.2.265

Paradise, J. L., Feldman, H. M., Campbell, T. F., Dollaghan, C. A., Colborn, D. K., Bernard, B. S., … Smith, C. G. (2001). Effect of early or delayed insertion of tympanostomy tubes for persistent otitis media on developmental outcomes at the age of three years. *New England Journal of Medicine*, *344* (16), 1179–1187. https://doi.org/10.1056/NEJM200104193441601

Ragnarsdóttir, H., Simonsen, H. G., & Plunkett, K. (1999). The acquisition of past tense morphology in Icelandic and Norwegian children: An experimental study. *Journal of Child Language*, *26*(3), 577–618. https://doi.org/10.1017/S0305000999003918

Rescorla, L. (1989). The Language Development Survey: a screening tool for delayed language in toddlers. *The Journal of Speech and Hearing Disorders*, *54*(4), 587–599. https://doi.org/10.1044/jshd.5404.587

Savičiute, E., Ambridge, B., & Pine, J. M. (2018). The roles of word-form frequency and phonological neighbourhood density in the acquisition of Lithuanian noun morphology. *Journal of Child Language*, *45*(3), 641–672. https://doi.org/10.1017/S030500091700037X

Stokes, S. F., Kern, S., & Dos Santos, C. (2012). Extended Statistical Learning as an account for slow vocabulary growth. *Journal of Child Language*, *39*(1), 105–129. https://doi.org/10.1017/S0305000911000031

Storkel, H. L. (2004a). Do children acquire dense neighborhoods? An investigation of similarity neighborhoods in lexical acquisition. *Applied Psycholinguistics*, *25*(2), 201–221. https://doi.org/10.1017/S0142716404001109

Storkel, H. L. (2004b). Methods for minimizing the confounding effects of word length in the analysis of phonotactic probability and neighborhood density. *Journal of Speech, Language, and Hearing Research : JSLHR*, *47*(6), 1454–1468. https://doi.org/10.1044/1092-4388(2004/108)

Storkel, H. L. (2009). Developmental differences in the effects of phonological, lexical and semantic variables on word learning by infants. *Journal of Child Language*, *36*(2), 291–321. https://doi.org/10.1017/S030500090800891X

Storkel, H. L., Bontempo, D. E., Aschenbrenner, A. J., Maekawa, J., & Lee, S.-Y. (2013). The Effect of Incremental Changes in Phonotactic Probability and Neighborhood Density on Word Learning by Preschool Children. *Journal of Speech, Language, and Hearing Research*, *56*(5), 1689–1700. https://doi.org/10.1044/1092-4388(2013/12-0245)

Storkel, H. L., & Hoover, J. R. (2010). An online calculator to compute phonotactic probability and neighborhood density on the basis of child corpora of spoken American English. *Behavior Research Methods*, *42*(2), 497–506. https://doi.org/10.3758/BRM.42.2.497

Storkel, H. L., & Lee, S.-Y. (2011). The Independent Effects of Phonogactic Probability and Neighborhood Density on Lexical Acquisition by Preschool Children. *Language and Cognitive Processes*, *26*(2), 191–211. https://doi.org/10.1080/01690961003787609

Suárez, L., Tan, S. H., Yap, M. J., & Goh, W. D. (2011). Observing neighborhood effects without neighbors. *Psychonomic Bulletin and Review*, *18*(3), 605–611. https://doi.org/10.3758/s13423-011-0078-9

Team, R. C. (2017). R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.r-project.org/

Vitevitch, M. S., & Stamer, M. (2006). The curious case of competition in Spanish speech production. *Language and Cognitive Processes*, *21*(6), 760–770. https://doi.org/10.1080/01690960500287196

Vitevitch, M. S., Stamer, M. K., & Sereno, J. A. (2008). Word Length and Lexical Competition: Longer is the Same as Shorter. *Language and Speech*, *51*(4), 361–383. https://doi.org/10.1177/0023830908099070

Weide, R. (2014). The CMU pronunciation dictionary, release 0.7a. Retrieved from http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Weizman, Z. O., & Snow, C. E. (2001). Lexical input as related to children's vocabulary acquisition: effects of sophisticated exposure and support for meaning. *Developmental Psychology*, *37*(2), 265–279. https://doi.org/10.1037/0012-1649.37.2.265