HHS Public Access
Author manuscript
J Child Lang. Author manuscript; available in PMC 2017 July 24.

Published in final edited form as:
J Child Lang. 2003 May ; 30(2): 441–469.

Phonological neighbourhoods in the developing lexicon*

JEFFRY A. COADY and
University of Wisconsin-Madison

RICHARD N. ASLIN
University of Rochester

Abstract

Structural analyses of developing lexicons have provided evidence for both children's holistic lexical representations and sensitivity to phonetic segments. In the present investigation, neighbourhood analyses of two children's (age 3;6) expressive lexicons, maternal input, and an adult lexicon were conducted. In addition to raw counts and frequency-weighted counts, neighbourhood size was calculated as the proportion of the lexicon to which each target word is similar, to normalize for vocabulary size differences. These analyses revealed that children's lexicons contain more similar sounding words than previous analyses indicated. Further, neighbourhoods appear denser earlier in development relative to vocabulary size, presumably because children first learn words with more frequent sounds and sound combinations. Neighbourhood density as a proportion of the size of the lexicon then decreases over development as children acquire words with less frequent sounds and sound combinations. These findings suggest that positing fundamentally different lexical representations for children may be premature.

INTRODUCTION

During the first postnatal year, infants acquire a remarkable amount of information about the sound structure of their native language, even though they have virtually no productive use of that language. By 0;6, infants have the ability to discriminate virtually any phonetic contrast that could be relevant in any of the world's languages (see Aslin, Jusczyk & Pisoni, 1998 for a review). By 1;0, they have honed their discriminative capacities to concentrate on those speech sounds used phonemically in their native language (Kuhl, 1979, 1983; Hillenbrand, 1984; Werker & Tees, 1984; Jusczyk, Pisoni & Mullennix, 1992), how often they occur (Jusczyk, Luce & Charles-Luce, 1994), how they are combined to form the words of their language (Jusczyk, Friederici, Wessels, Svenkerud & Jusczyk, 1993), how they differ depending on their position in a word (Jusczyk, Hohne & Bauman, 1999), and how those differences signal syllable and word boundaries (Mattys, Jusczyk, Luce & Morgan, 1999).

*This research was conducted as part of the first author's dissertation at the University of Rochester. It was supported by a grant from the NIH (HD-37082) to the second author. We are grateful to Daniel Swingley, James Magnuson and Jason Zevin for assistance conducting the corpus analyses, and to Keith Kluender and three anonymous reviewers for helpful comments and suggestions.

Address for correspondence: Jeffry Coady, Department of Psychology, University of Wisconsin-Madison, 1202 W. Johnson St., Madison, WI 53706.

These sophisticated perceptual skills and distributional sensitivities contrast markedly with the fact that in the second postnatal year, as infants begin to speak, they often make errors in both speech perception and speech production (Eilers & Oller, 1976). Consequently, some researchers have suggested that once children begin to form lexical representations – the association between acoustic patterns and word meanings – they rely on holistic, syllable-based units rather than sequences of phonetic segments (Ferguson & Farwell, 1975; Menyuk & Menn, 1979). According to this view, children differentiate their earliest words based on overall acoustic or phonetic shape rather than on an analysis of individual phonetic segments. When the lexicon eventually grows to a size that prevents holistic lexical representations from adequately representing every word candidate, children must begin to pay attention to fine-grained acoustic-phonetic detail to differentiate among similar sounding words.

In spite of their sensitivity to phonetic detail and their increasing representational capacity, Treiman & Baron (1981) noted that even after the onset of the naming explosion, preliterate children do not appear to recognize that spoken words are composed of smaller units, which is a necessary precursor to the ability to read. Because adults can break words into their constituent phonemes, and because young children lack either the capacity or the metalinguistic skill required to do so, Treiman & Baron hypothesized that the period of holistic lexical representation extends well beyond the onset of the naming explosion. Specifically, they claimed that children store words in their lexicons holistically, and that these holistic representations gradually become segmentally based over the course of early childhood.

Support for this extended period of holistic lexical representation comes from evidence that children group spoken syllables on the basis of overall similarity, while adults group the same syllables on the basis of shared phonetic segments (Treiman & Baron, 1981; Walley, Smith & Jusczyk, 1986). Furthermore, when perceiving and producing fricative-vowel syllables, young children appear to rely more on the vocalic formant transitions than on the fricative noise spectra, using information that spans the syllables rather than information within individual phonetic segments (Nittrouer & Studdert-Kennedy, 1987; Nittrouer, Studdert-Kennedy & McGowan, 1989). Based on these and other findings, a number of researchers (e.g. Jusczyk, 1986; Walley, 1993) have concluded that children's sensitivity to individual phonemes is less than adults', and therefore that words in the developing lexicon must be stored holistically.

In contrast, more recent evidence using other methods of assessment suggests that very young children have access to more segmental (less holistic) representations, at least after the naming explosion. Gerken, Murphy & Aslin (1995) reported that children aged 3;0 to 4;0 use phonetic features and segments in a lexical access task. They found that children more often confused a target word with a nonsense word that differed by a single feature than they did with a non-word that differed by two features. Further, they found that children more often confused a target word with a nonsense word that differed by two features in a single position than they did with a nonword that differed by two features, but in two separate positions. If children were perceiving words holistically, these two conditions should have resulted in the same level of performance.

Swingley, Pinto & Fernald (1999) also provided evidence for a segmental level of analysis in the lexicons of very young children. Using an eye-tracking paradigm, they found that the timing of word identification was linked to the timing of the unfolding speech stream. That is, the latency to differentiate two words among children aged 2;0 depended on the degree of phonetic overlap at word onset. For example, *dog* and *doll* were differentiated later (after word onset) than *tree* and *truck* because the /da-/ portion of *dog* and *doll* has a longer duration than the /tr-/ portion of *tree* and *truck*. Furthermore, they found that infants' ability to differentiate two words was delayed when the words overlapped phonetically at onset (*doll* and *dog*), but not when the words rhymed (*doll* and *ball*). Thus, children appear to pay more attention to the phonetic segments (specifically to the word onsets) than to the overall acoustic shape of words. Furthermore, Swingley & Aslin (2000, 2002) have shown that infants as young as 1;2 are sensitive to the subtle mispronunciations that have been cited as evidence for children's holistic perceptual strategies, even when these changes are not at word onset. Using an eye-tracking paradigm, they found that children's fixations of an object's picture during auditory naming are briefly influenced by mispronunciations (e.g. substituting *cur* for *car*). These results suggest that more careful measurements and more sensitive experimental paradigms reveal that very young children do show evidence of sensitivity to segmental lexical information in referential speech perception tasks.

Computational analyses of lexical corpora

The foregoing summary suggests that the degree to which children's lexical representations are segmental versus holistic remains unclear. Another approach that could address this uncertainty is the use of simulation methods adapted from studies of spoken word recognition in adults. For example, Luce's Neighbourhood Activation Model posits that the speed and accuracy with which adults identify spoken words depend on target word frequency, neighbourhood density (the number of similar sounding words in the lexicon), and the frequency of these near phonological neighbours (Luce, Pisoni & Goldinger, 1990; Luce & Pisoni, 1998). One of these factors, neighbourhood density, may be particularly revealing of the nature of children's lexical representations. Consider two ways in which neighbourhood density could play a crucial role in the formation of the lexicon. On the one hand, as suggested by Schwartz (1988), children may be biased to add new words to their lexicon based on the ease with which these new words can be discriminated from words that are already in their vocabulary. If new words were preferentially added to these 'gaps' in the acoustic space of the lexicon, children's holistic perceptual strategies would reduce or eliminate the addition of similar sounding words (e.g. minimal pairs) to the lexicon, thereby maintaining a uniform sparseness to the phonological neighbourhoods. On the other hand, if children are sensitive to individual phonetic segments, then the expansion of the lexicon should not be fundamentally limited by the greater confusability of holistic lexical representations. As a result, the entry of new words into the lexicon will be facilitated for word candidates that contain common (more frequent) sounds and sound combinations, as suggested by Menn (1978) and Lindblom (1992). In this latter case, phonological neighbourhoods in the developing lexicon will actually be denser than those in the mature lexicon after scaling for vocabulary size.

Several studies have attempted to resolve this question about the development of phonological neighbourhoods by examining the structural properties of the lexicon in young children. Charles-Luce & Luce (1990) proposed that neighbourhood density may play an important role in the expansion of the developing lexicon. They hypothesized that words in the developing lexicon should be maximally distinct from one another because children are not sensitive to fine-grained phonetic detail. They examined this hypothesis by calculating phonological neighbourhoods for a 679-word lexicon typical of children aged 5;0, for a 943-word lexicon typical of age 7;0, and for a 20,000-word lexicon typical of adults. A word was considered a target word's neighbour if it differed from that target word by the addition, substitution, or deletion of a single phoneme in any word-position. For the analyses of the two child lexicons, Charles-Luce & Luce compared all words in each lexicon to the other words within that particular lexicon. For the analysis of the adult lexicon, on the other hand, they compared those words in each of the child lexicons to all of the 20,000 words in the adult lexicon; a complete analysis of neighbourhoods in the adult lexicon was not computed. Results were presented separately for three-, four-, and five-phoneme words, and showed that neighbourhood density increases over development for words of all lengths. Furthermore, Charles-Luce & Luce argued that this developmental increase in neighbourhood density is not due to differences in vocabulary size. They argued that if increasing vocabulary size alone accounted for the developmental increase in neighbourhood density, then neighbourhood density should be proportional to vocabulary size. They therefore examined the lexicons in terms of word length, and found that in the adult lexicon, roughly 9 percent of the words are three phonemes in length, while roughly 15 percent of all words are four phonemes, and 16 percent are five phonemes. However, neighbourhood density was higher for three-phoneme words than for four- and five-phoneme words. That is, three-phoneme words reside in denser neighbourhoods, even though four- and five-phoneme words out-number them. Thus, Charles-Luce & Luce argued that phonological neighbourhoods become more densely packed over the course of development, and that this increasing density is not primarily an effect of increasing vocabulary size. The increasing size of the lexicon and its associated demand for more differentiated phonological neighbourhoods were hypothesized as the motivating forces for the emergence of mature segmental representations from rudimentary holistic representations.

A similar analysis of the structural properties of the child's lexicon was provided by Logan (1992). He limited his sample to 4000 word tokens per child per sampling period, and found a significant increase in neighbourhood density between the ages of 1;6 and 2; 0, coincident with the naming explosion, with smaller increases thereafter. Interestingly, he coded words as sequences of phonemes, but he also coded the phonemes either by just their place of articulation or by just their manner of articulation. This coding by place or manner was an attempt to explore two candidate mechanisms for holistic lexical representations; namely, loosely specified bundles of articulatory features. While coding by place of articulation resulted in denser neighbourhoods than either coding by manner of articulation or the full phonemic coding, coding by manner of articulation was comparable to the full phonemic coding. Thus, in the developing lexicon, having words stored just in terms of their manner of articulation is a viable alternative to specifying all of a word's phonetic features. It is interesting to note that although Logan drew his lexicon from a very different source, his results for children aged 5;0 closely matched those reported by Charles-Luce & Luce (1990).

Dollaghan (1994) criticized both Charles-Luce & Luce's (1990) analysis strategy and their conclusions about lexical development. First, she argued that the developmental differences in neighbourhood density reported by Charles-Luce & Luce must be due to the huge differences in the vocabulary sizes analysed for children and adults. Dollaghan pointed out that Charles-Luce & Luce underestimated the size of children's vocabularies. Because Carey (1981) reported that a typical six-year-old has approximately 8,000 unique lexical entries, the five- and seven-year-olds' vocabularies of 679 and 943 words, respectively, used by Charles-Luce & Luce were unrealistically small. Furthermore, they calculated phonological neighbourhoods in the developing lexicon by comparing each word to several hundred other words, while those in the adult lexicon were calculated by comparing those very same words from the child lexicon to several thousand words in the adult lexicon – an order of magnitude difference. Dollaghan pointed out that there are a finite number of speech sounds used in any language, and those sounds can only be combined in limited ways to form the words of the language. When new words enter the lexicon, they must fit into this finite acoustic-phonetic space. Adult phonological neighbourhoods are more densely packed simply by virtue of the fact that adults have more words in their lexicons. Sparser neighbourhoods in the developing lexicon are an inevitable fact, and therefore are uninformative about the nature of developing phonological representations.

Furthermore, Dollaghan (1994) criticized the interpretations and conclusions reached by Charles-Luce & Luce (1990). This criticism focused primarily on the contention that considerable phonetic detail is necessary to differentiate between two lexical items that differ only by a single phoneme (i.e. that are neighbours, such as *cat* vs. *hat*, *cut*, or *can*). While these three neighbours of *cat* differ by only a single phoneme, they appear to adults to be rather easily discriminable from *cat*. However, in the context of a non-neighbour such as *fish*, which differs from *cat* in all three phonemes, these neighbours demand a fairly sophisticated discrimination mechanism and representational system. Dollaghan noted that 94 percent of the three-phoneme words in the Charles-Luce & Luce analysis had at least one phonological neighbour. Thus, she concluded that while children's sparser neighbourhoods are more likely due to children simply having smaller vocabularies rather than to organizational differences, this sparseness belies children's perceptual abilities. Walley (1993) reached a similar conclusion by noting that four- and five-year-olds (younger than the children considered in the Charles-Luce & Luce analysis) showed evidence of more segmental lexical representations.

Dollaghan also expressed concern that Charles-Luce & Luce ignored other factors known to affect word recognition, such as word familiarity and word frequency. She also pointed out that children's receptive vocabularies are probably a better measure of their actual lexicon at a young age than their expressive vocabularies; however, a comprehensive input lexicon was unavailable. She therefore conducted her own neighbourhood analysis by examining the expressive vocabularies typical of children aged 1;0 to 3;0. She based her analysis on 407 monosyllabic words, and found that approximately 84 percent of the words in the children's lexicon have at least one phonological neighbour, while 67 percent have two or more neighbours. Dollaghan concluded that, from the earliest stages of lexical acquisition, children must distinguish among phonologically similar entries in the developing lexicon, rendering global or holistic perceptual strategies ineffective.

In a subsequent analysis, Charles-Luce & Luce (1995) acknowledged Dollaghan's criticisms (1994), but also pointed out that her own analysis failed to correct for the noted shortcomings. To answer these criticisms, Charles-Luce & Luce presented the results from another neighbourhood analysis that examined the receptive vocabularies of children aged 1;1–1;9, which included 743 words. They found that the phonological neighbourhoods in children's receptive vocabularies were denser than those in their original study, but they again showed that words in the developing lexicon are less densely packed than those same words in the adult lexicon. They concluded that their initial claims (1990) had not been undermined, and that words in the developing lexicon are maximally differentiable. Therefore, Charles-Luce & Luce argued that holistic perceptual strategies are sufficient for young children to acquire much of their early lexicon.

To summarize, Charles-Luce & Luce (1990, 1995) and Logan (1992) found that words in the developing lexicon are, in principle, more discriminable than words in the adult lexicon, as evidenced by sparser phonological neighbourhoods. This structure would make it possible for children to use global, holistic perceptual strategies, and suggests that the addition of new words to the lexicon will be constrained by the presence of similar sounding words. In contrast, Dollaghan (1994) found that most words in the developing lexicon are, in fact, confusable with other words. This structure should preclude the use of global, holistic perceptual strategies, and suggests that the addition of new words to the lexicon is not constrained by the sound structure of the incoming words, but rather drives the lexicon to employ fine-grained segmental representations. The question of neighbourhood structure in the developing lexicon, therefore, remains unresolved.

While the foregoing corpus analyses of the structure of the developing lexicon have been useful in addressing specific claims concerning segmental versus holistic representations, they have several limitations: (a) they have only considered either children's expressive or receptive vocabulary, (b) they have underestimated the size of children's lexicons, and (c) they have relied on simple counts of whether a given word token was used, which do not appropriately normalize neighbourhoods by the sizes of the vocabularies considered. For the present study, a more comprehensive analysis of phonological neighbourhoods was conducted. First, children's expressive lexicons and their maternal input were drawn from samples of extended mother–child interactions. Similarity neighbourhoods were computed for both children's expressive vocabularies and for their maternal input, so that these can be considered together and language input can be directly compared to language output. Moreover, the sheer amount of speech included in the corpora should provide a more complete and representative sample of children's vocabularies, reducing any concerns about underestimating vocabulary size. Second, each child's corpus was weighted by the frequency with which each word token was used, thereby providing a frequency-weighted estimate of neighbourhood density. Third, because adults' lexicons are so much larger than children's lexicons, at least part of the differences in neighbourhood density must result from differences in vocabulary size. We therefore further examined neighbourhood density as a ratio of the number of neighbours to the number of words in a given lexicon size, in an effort to determine how neighbourhood density changes during the expansion of a ‘normalized’ lexicon.

STUDY 1. PHONOLOGICAL NEIGHBOURHOODS IN CHILDREN AGED 3; 6 AND YOUNGER

Method

Database—Phonological neighbourhoods were calculated for Adam and Sarah, and for their mothers, from the Brown corpus in the CHILDES database (Brown, 1973; MacWhinney, 1991). For both Adam and Sarah, the recording sessions began at approximately the age of 2;3.[1] All sessions from CHILDES were included until the child had reached the age of 3;6, which coincides approximately with the consistent use of consonant clusters (Templin, 1953). In addition, phonological neighbourhoods were calculated for an adult lexicon. The adult lexicon was taken from a computer-readable dictionary of 77,581 words that includes an orthographic representation, a phonetic transcription, a word frequency count based on the Kučera & Francis (1967) corpus, and word familiarity ratings (Nusbaum, Pisoni & Davis, 1984) for each entry.

The CHILDES database consists of tape recorded sessions that were transcribed into English orthography. These transcriptions attempted to represent the actual utterances that the children and their mothers produced, including speech errors. Phonetic transcriptions for each individual word were then taken from the CMU Pronouncing Dictionary, an online machine-readable dictionary of North American English. These transcriptions were then hand checked, and the pronunciation of words not in the dictionary were based on the orthographic representation (e.g. *scuse* as a form of *excuse*). The only rhotic vowels included were /ɜ/ and /ɚ/; all other post-vocalic rhotics were treated as consonants. All unstressed syllables contained either /ə/ or /ɚ/. The following monosyllabic words, all of them function words, were transcribed with an unstressed vowel: *a*, *an*, *the*, *de*, *can*, *gon*, *was*, *were*, *to*, *of*, *or*, *for*, *her*, *yer*. Adam produced 2629 types/81,183 tokens; his mother, 2559 types/55,057 tokens. Sarah produced 1912 types/35,803 tokens; her mother, 2596 types/58,344 tokens. Because the original recordings were not analysed directly, there may have been transcription errors of which we are unaware. Further, because the transcriptions are in English orthography, we cannot assess the accuracy of the children’s articulations. Also, because the phonetic transcriptions were based on the words in isolation rather than in context, any suprasegmental factors that might have affected articulation were lost.

Procedure—As per Charles-Luce & Luce (1990), neighbours were defined as words that differ from a target word by the addition, substitution, or deletion of a single phoneme in any position. Similarity neighbourhoods were calculated based on phonetic transcriptions for each speech corpus individually. That is, the words in Adam’s corpus were compared to all other words appearing in that same corpus, without reference to the other corpora. The

[1]Eve’s sessions were excluded from the analysis because her family moved away when she was 2; 3, leaving an incomplete data set for the present purpose.

analysis included all monosyllabic words spoken by the children and their mothers, along with the monosyllabic words in the adult lexicon. Adam produced 1257 monosyllabic types while his mother produced 1175 monosyllabic types. Sarah produced 1001 monosyllabic types, while her mother produced 1214 monosyllabic types. Proper names, homonyms, and contractions were removed from the analyses. Inflected forms whose root forms also appeared in the corpus were also removed from the analyses. However, inflected forms whose root forms were not attested in the corpora remained in the analyses. For example, Adam produced the word *smashed*, although the word *smash* was unattested. The word *smashed* was therefore included in the analysis. Also, inflected forms that contained a vowel change were counted as separate entries. For example, *seen*, an inflected form of *see*, was removed from the analysis, but *been*, an inflected form of *be* was included as a separate entry. After removing proper names, homonyms, contractions, and inflected forms, phonological neighbourhoods were calculated for Adam's 923 remaining unique types, his mother's 843 remaining unique types, Sarah's 760 remaining unique types, and her mother's 870 remaining unique types. It is important to note that these analyses are based not on citation forms or intended targets, but rather on words that the children and their mothers actually produced. Table 1 summarizes the number of words of each length included in the neighbourhood analyses.

Of the 77,581 words in the adult lexicon, 23,182 had received a familiarity rating of 4 or higher, corresponding to 'know it's a word; don't know its meaning' (Nusbaum *et al.*, 1984). Of these, 7588 were monosyllabic. Proper names, homonyms, contractions, and inflected forms were then removed from the corpus, leaving 2561 unique monosyllabic forms. However, some high frequency words from the children's corpora, such as *a*, *is*, *the*, *it*, and *and* were excluded based on these criteria. Therefore, the MRC Psycholinguistic Database (Coltheart, 1981) was also consulted for familiarity ratings. Accordingly, another 34 words (cumulative frequency: 254,868, or one quarter of all word instances) were included in the analysis. Phonological neighbourhoods were calculated for these 2595 unique monosyllabic types.

After phonological neighbourhoods were calculated, the results were examined in two different ways. In one analysis, raw neighbourhood counts for each of the five corpora were examined and compared. In a second analysis, words from the children's corpora were directly compared to those same words in the adult corpus. In order to accomplish this, words were omitted from both the children's and adult corpora if they were not included in both. Charles-Luce & Luce (1990, 1995) originally used this method in order to directly compare neighbourhood density in the adult and child lexicons. There were 828 words in Adam's expressive vocabulary and 656 words in Sarah's expressive vocabulary that appeared in the adult lexicon. The excluded words were child forms of adult words, inflected forms that were removed from the adult analysis, or words that were not included in the adult analysis. Comparisons between the child lexicons and the adult lexicon were based on these words that appeared in both the child and adult lexicons.

Results

In order to directly compare the current results with previous findings, results are first shown as a function of word length, then for all words regardless of length. Figure 1 shows the results of the statistics computed for similarity neighbourhoods for Adam's and Sarah's expressive vocabularies and maternal input for three-phoneme words. Figure 2 presents the results for four-phoneme words, and Figure 3 shows the results for five-phoneme words. The corresponding results from the analysis of the adult corpus are also included in all of the graphs for comparison. In all of the corpora analysed, as word length increases, neighbourhood density decreases. Figure 4 shows the results of the statistics computed for all words regardless of length in the expressive vocabularies and maternal input of Adam and Sarah, respectively. All of these results are summarized in Table 2.

When we limit the comparison between child and adult lexicons to only those words common to both the child lexicons and the adult lexicon, which was the method used by Charles-Luce & Luce (1990, 1995), we see a similar, but more informative pattern of results. Words in Adam's expressive lexicon average 6.46 neighbours, while those exact words have an average of 13.91 neighbours in the adult lexicon. The words in Sarah's expressive vocabulary have an average of 6.37 neighbours, and an average of 14.76 neighbours in the adult lexicon. Two distinct patterns emerge from these analyses. First, neighbourhoods are denser in the adult lexicon than those for the same words in the developing lexicon. Second, for words that appear in the children's lexicons, the average number of phonological neighbours in the adult lexicon is higher than the average for all of the words in the adult lexicon. Recall that the average number of neighbours in the adult lexicon is 11.57 when all words are included, and 13.91 and 14.76 when the analyses were limited to words that appeared in Adam's and Sarah's lexicons, respectively. That is, while the results of the analyses of the children's lexicons do not significantly change when words are omitted from the analyses, limiting the adult analysis to words in the children's lexicons reveals that children's words are drawn from the denser than average neighbourhoods. This finding is not apparent from the initial analysis, and suggests that children are acquiring words from denser portions of the adult lexical neighbourhoods.

Discussion

The results from the current analyses support both sets of previous findings. Like Charles-Luce & Luce (1990, 1995) and Logan (1992), we found that neighbourhoods are sparser in the developing lexicon. As explained above, this is not surprising given that all words are contained within a finite acoustic-phonetic space. However, the current analyses also reveal that by the age of 3;6, words in children's lexicons are confusable with more words than previous analyses have indicated. Words in Adam's and Sarah's expressive vocabularies had an average of 6.5 neighbours, while words in the corpora analysed by Dollaghan (1994), and Logan (1992) had between 2.25–3.32 neighbours. Furthermore, even though Charles-Luce & Luce (1990, 1995) presented neighbourhoods as a function of word length, Logan (1992) showed that the results from their initial study (1990) were comparable to his own. Therefore, the current analysis revealed that words in the developing lexicon have roughly twice as many neighbours than did previous analyses. While finding that lexical neighbourhoods are this dense certainly does not preclude the use of global or holistic perceptual strategies, it does suggest that young children must possess considerable acoustic-phonetic skill to be able to differentiate among these neighbours, exactly the point raised by Dollaghan (1994).

STUDY 2. FREQUENCY-WEIGHTED NEIGHBOURHOOD DENSITY

Dollaghan (1994) pointed out that phonological similarity is only a part of what influences auditory word recognition. Target word familiarity, target word frequency, and neighbourhood frequency also contribute to this process. The previous analyses controlled for word familiarity by only including words that children actually said or heard from their mothers. However, only the number of similar words, but not their frequency, was considered. Therefore, in the following analysis, the effects of neighbourhood frequency were examined by weighting each of the neighbourhoods by their frequency.

Method

For each corpus in the previous analysis, neighbourhood frequency for each word was calculated by summing the frequency of the target word and the frequencies of all of its neighbours. Because words from higher frequency neighbourhoods are assumed to be harder to identify than those from lower frequency neighbourhoods (Luce *et al.*, 1990; Luce & Pisoni, 1998), neighbourhood density for each target word was weighted by log neighbourhood frequency. For example, consider the words *slab*, with ten neighbours, and *watch*, with five neighbours. The former has twice as many neighbours as the latter. However, the ten neighbours of *slab* rarely occur, while the five neighbours of *watch* are more frequent. After both neighbourhoods are weighted by their log frequency, however, their respective densities are almost identical.

A potential problem is that frequency counts in the children's lexicons were based on number of occurrences per several thousand words spoken, while those for the adult lexicon were based on number of occurrences per million. Therefore, frequency counts in the children's corpora had to be scaled to be comparable to adult frequency counts. A simple scaling would be inappropriate, because children's speech typically lacks many of the grammatical functor words that characterize fluent adult speech. For example, when Adam was asked 'Are those your checkers?', he responded 'Adam checker', meaning 'Those are Adam's checkers.' This response lacks four grammatical morphemes – the pronominal subject, the copula, the possessive, and the plural. While children may produce the content words at rates comparable to adults, they surely do not produce the function words at similar rates. Therefore, the scaling factors were based on the frequencies of the most common content words in the children's lexicons. For each of the child corpora, the frequency counts of the ten most frequent words were compared to the frequency counts of those same words in the adult corpus. The ratio of these words' frequencies in the adult lexicon to their frequencies in the child lexicons was then used as a scaling factor. This assumes that the relative frequencies of these most common words remain constant over development, which may or may not be a tenable assumption. However, in this context, this measure provides a more conservative estimate for scaling up the children's lexicons than simply multiplying the frequency counts from the children to achieve the number of occurrences per million.

The frequency counts in Adam's expressive lexicon were therefore scaled by a factor of 10.92, and those in his maternal input were scaled by a factor of 21.97. The counts in Sarah's expressive lexicon were scaled by a factor of 15.76, and those in her maternal input were scaled by a factor of 16.57. If a simple scaling had been done, these factors would be 12.32, 18.16, 27.93, and 17.14, respectively. Thus, for three of the four corpora considered, the corrected scaling factors more conservatively estimate frequency. The counterexample is Adam's maternal input, which is not assumed to be missing the grammatical function words.

Results and discussion

Figure 5 shows the results of the statistics computed for log frequency-weighted neighbourhood density for Adam's and Sarah's expressive vocabulary and maternal input for all words regardless of length. The corresponding results from the analysis of the adult corpus are also included for comparison. After scaling to make the frequency counts comparable, the average log frequency-weighted neighbourhood density for Adam's expressive lexicon was 28.07, while that for his maternal input was 24.08. In Sarah's lexicon, average weighted density was 27.71 in her expressive lexicon and 26.10 in her maternal input. In the adult lexicon, the average frequency-weighted density was 40.83. When we consider only those words common to the children's and adult corpora, we again see a similar, yet subtly different pattern of results. Average log frequency-weighted neighbourhood density for Adam's expressive lexicon was 27.70, while that for the same words in the adult lexicon was 54.62. Average weighted density in his maternal input was 24.65, while that for the same words in the adult lexicon was 51.12. In Sarah's lexicon, average weighted density was 26.87 in her expressive lexicon, 55.71 for those words in the adult lexicon, 26.17 in her maternal input, and 52.52 for those words in the adult lexicon.

As in the previous, unweighted analysis, neighbourhood density is greater in the adult lexicon than in the developing lexicon. However, in the current analysis the density measures from the previous analysis were simply weighted by neighbourhood frequency. In the unweighted analysis, words in the developing lexicon had roughly half as many neighbours as those same words in the adult lexicon (range 0.45–0.50). In this frequency-weighted analysis, the results are strikingly similar. After normalizing for frequency differences, log frequency-weighted neighbourhood density in the developing lexicon is roughly half that of the same words in the adult lexicon (range 0.48–0.54). That is, frequency weighting did not alter the relative size of lexical neighbourhoods in the developing and adult lexicons.

The results of this analysis further support the findings of Charles-Luce & Luce (1990, 1995) and Logan (1992). Words in the developing lexicon are less confusable with other words than those same words in the adult lexicon. Based on the way lexical neighbourhoods are calculated, these findings are incontrovertible. However, as Dollaghan (1994) pointed out, this is quite probably due to the disparate sizes of the vocabularies considered in the analyses. Recall that neighbourhoods in the adult lexicon were based on 2595 words, while those in the children's lexicons were based on roughly 800–900 words. Therefore, the influence of vocabulary size on neighbourhood density bears investigation.

STUDY 3. VOCABULARY SIZE, WORD LENGTH, AND NEIGHBOURHOOD DENSITY

Charles-Luce & Luce (1990) claimed that the differences in vocabulary size between children and adults could not account for differences in neighbourhood density between the developing and the mature lexicon. They offered as proof the finding that adults have more four- and five-phoneme words in their lexicons than three-phoneme words, but that three-phoneme words have more neighbours than four- or five-phoneme words. By this logic, if neighbourhood density were proportional to vocabulary size, then one would expect the less frequent three-phoneme words to have lower neighbourhood densities. Indeed, in the current analysis, the 492 three-phoneme words in Adam's expressive lexicon have an average of 7.56 neighbours, while the 1010 four-phoneme words in the adult lexicon have an average of 7.28 neighbours. Thus, neighbourhood size does seem to be unrelated to neighbourhood density. However, it is also the case in the current analysis as well as Charles-Luce & Luce's (1990, 1995) and Logan's (1992) previous analyses that shorter words reside in denser phonological neighbourhoods. Because this negative correlation between word length and neighbourhood density confounds the relationship between vocabulary size and neighbourhood density, examining the lexicon in terms of word length alone cannot fully explain this relationship. Therefore, the relationship between vocabulary size and neighbourhood density was examined by first calculating the average number of neighbours as a function of word length. Then, word length in phonemes in the children's lexicons and the average number of phonological neighbours in the children's lexicons were plotted as proportions of the adult lexicon.

Method

The results from the analyses of the children's and adult lexicons from Study 1 were re-plotted in terms of the average number of neighbours as a function of word length. Then, for the child lexicons, the average number of words of each length and the average number of neighbours were plotted as proportions of word length and number of neighbours in the adult lexicon.

Results and discussion

Figure 6 shows the average number of phonological neighbours as a function of word length in phonemes. It is clearly the case that neighbourhood size decreases as word length increases. The only exception to this finding is the number of neighbours for one-phoneme words in the adult lexicon, which have fewer neighbours than the two-phoneme words. This is the result of the limitations placed on the adult analysis, however, as many of the one-phoneme words attested in the children's speech were not included in the adult lexicon. Recall that all one-phoneme words are by definition neighbours of each other because they differ by the substitution of a single phoneme. Therefore, in this particular case, having fewer words translates directly to having fewer neighbours.

Figure 7 plots the average number of words of each length along with the average number of neighbours for words of each length as a proportion of the adult lexicon. This figure reveals that children's lexicons contain a disproportionate number of shorter words, with progressively fewer words as word length increases. Children's lexicons contain more than the seven one-phoneme words in the adult lexicon, and 69 percent of the two-phoneme words. This function continues to decline, until only 17 percent of the five-phoneme words and 16 percent of the six-phoneme words from the adult lexicon appear in the child lexicon. This is not surprising since the longer words contain consonant clusters, and the age of 3;6 was chosen for the current analysis because it is the age at which children are just beginning to consistently produce consonant clusters (Templin, 1953). The average number of neighbours for words of each length reveals a similar function. As word length increases, children's lexicons contain progressively fewer phonological neighbours. That is, the two-phoneme words in the child lexicon reside in neighbourhoods that are only 68 percent of the size of those same neighbourhoods in the adult lexicon. For five-phoneme words, the neighbourhoods in the child lexicon are only 16 percent of their size in the adult lexicon. This pattern of results suggests that children learn the shorter words from the adult lexicon first, and fill in the longer words. Because children have a greater proportion of the shorter words that reside in denser phonological neighbourhoods, neighbourhood density should actually decrease over development as children later acquire the longer words that reside in sparser phonological neighbourhoods.

While all results show that words in the developing lexicon have fewer neighbours than words in the mature lexicon, the different vocabulary sizes analysed remain a confounding factor. To correct for this, neighbourhood size was calculated as a proportion of the entire lexicon, which provides a measure of neighbourhood density independent of vocabulary size. That is, rather than calculating neighbours by counting the number of similar words, neighbours were determined by calculating the percentage of the child's or adult lexicon that served as a neighbour to a target word. Instead of using raw counts, the current analysis considers ratios.

STUDY 4. NEIGHBOURHOOD DENSITY INDEPENDENT OF VOCABULARY SIZE

Consider three ways in which children's perceptual strategies might influence lexical organization. First, there might be no effect. If this were the case, children's lexical acquisition would be uninfluenced by a word's sound form, and children's lexical neighbourhoods would simply be smaller versions of adult lexical neighbourhoods. Over development, then, neighbourhood density would scale uniformly as the size of the lexicon increased. The second possibility is that children utilize more global perceptual strategies, and as Charles-Luce & Luce (1990) originally hypothesized, they maintain a maximum sparseness in their lexical neighbourhoods. In this case, neighbourhoods in children's lexicons would be sparser even after correcting for differences in vocabulary size. Neighbourhood density would then increase over development as new words enter the lexicon. The final possibility is that children are sensitive to segmental information, and prefer to learn words with the more common sounds and sound combinations of their input. In this case, lexical neighbourhoods would actually be denser early in development and become sparser as children learn words with less common sounds and sound combinations.

As stated above, phonological neighbourhoods in the developing lexicon were calculated by comparing each word to several hundred other words, while those in the adult lexicon were calculated by comparing each word to several thousand words in the adult lexicon. To correct for this size discrepancy, the present analysis calculated neighbourhood size as a proportion of the lexicon rather than as an absolute number. This is based on the idea that a word's confusability is a function not of the absolute number of similar sounding words, but rather of the proportion of the lexicon with which the target word is confusable. While this conception of word confusability is entirely speculative and not based on any empirical findings, it could potentially be informative. Consider two different words, *more* and *noise*. In Adam's expressive lexicon, the word *more* has four phonological neighbours, and in the adult lexicon, the word *noise* has four neighbours. So the word *more* is confusable with 4 of 922 other words in Adam's expressive lexicon, or 0.43%, while the word *noise* is confusable with 4 of 2594 other words in the adult lexicon, or 0.15%. By just including raw counts that ignore vocabulary size, we lose sight of the possibility that a word with four neighbours may be more confusable in the smaller lexicon than a word with four neighbours in the adult lexicon, particularly if children are using more holistic acoustic representations.

Method

Neighbourhood density was recalculated for the children's and adult corpora, this time in terms of the proportion of the lexicon with which each word was confusable. The ratio of the number of neighbours to the number of all monosyllabic words considered in the neighbourhood analysis was calculated for each word in all of the corpora. For the children's lexicons, the ratios were calculated based on each lexicon individually. Ratios for each of the words in the child lexicons were then calculated based on the entire adult lexicon. As a result, the adult and child lexicons could again be directly compared, as in Study 1, but this time in terms of neighbourhood density RELATIVE TO VOCABULARY SIZE.

Results and discussion

Figure 8 shows neighbourhood density as the ratio of phonological neighbours to all monosyllabic words in the lexicon for Adam's expressive lexicon and maternal input. The results from those same words in the adult lexicon are overlaid. Figure 9 shows these same results for Sarah's expressive lexicon and maternal input. The mean ratio of phonological neighbours to all words in Adam's expressive vocabulary is 0.0070, while the ratio for those same words in the adult lexicon is 0.0054 (a difference of 29.6%). The ratio in his maternal input is 0.0066, while that for the same words in the adult lexicon is 0.0053 (a difference of 24.5%). The ratio of phonological neighbours to all words in Sarah's expressive vocabulary is 0.0084, while that for those words in the adult lexicon is 0.0057 (a difference of 47.4%). The ratio in her maternal input is 0.0070 while those same words have a ratio of 0.0054 in the adult lexicon (a difference of 29.6%). In all cases, the ratio is larger in the child lexicon than for the same words in the adult lexicon.

These ratio data show that in proportion to neighbourhood size, neighbourhood density actually decreases considerably between the age of 3;6 and adulthood. This suggests that children are more likely to learn new words that contain the more frequent sounds and sound combinations, as Menn (1978) and Lindblom (1992) have suggested. If children were not sensitive to segmental information and were maintaining maximal distinctiveness among their lexical entries, then we would expect that the ratio of number of neighbours to the number of all words in the lexicon would increase over the course of development as more words are acquired and packed into the increasingly dense lexicon. This was not the case. Further, these results suggest that differences in the absolute size of lexical neighbourhoods are the result of comparing lexicons of different sizes within a finite acoustic-phonetic space.

GENERAL DISCUSSION

As in all previous studies that compared children's lexicons to the adult lexicon, we found that neighbourhood density increases over development. However, a closer examination of the words in children's lexicons reveals that this apparent increase does not tell the whole story. By the time children reach the age of 3;6, words in their lexicons are confusable with an average of six other words in their lexicons. While finding that neighbourhoods are this dense does not rule out the use of holistic perceptual strategies, it does suggest that children must possess considerable sensitivity to acoustic-phonetic detail to be able to differentiate among these similar sounding words. Moreover, the pattern of results suggests that children actually acquire words with more frequent sounds and sound combinations before those that contain the less frequent sounds and sound combinations. Three findings from the present study support this claim. First, experiment 1 revealed that the words in the children's lexicons are actually those words in the adult lexicon that reside in denser than average phonological neighbourhoods. That is, children first learn those words that are more confusable than average, based on the adult lexicon. Second, experiment 3 revealed that children have more of the shorter words and fewer of the longer words from the adult lexicon, and that neighbourhood density is inversely related to word length. Thus, again, we found that children have earlier acquired words from the denser neighbourhoods in the adult lexicon. Finally, in the fourth analysis, neighbourhoods were calculated in terms of ratios rather than as simple counts. While the absolute number of neighbours increases over development, neighbourhood density expressed as a proportion of the lexicon is higher in young children than in adults. Together, these results provide evidence that children acquire those words that contain the more frequent sounds first, and later fill in words with less frequent sounds.

While this and previous neighbourhood analyses of the developing lexicon have been useful in explaining the development of lexical organization, the data do not speak to the acoustic nature of children's lexical forms, nor to their preferred perceptual strategies. While a sparsely packed lexical space would certainly benefit listeners, it is not necessary for holistic perceptual strategies. Similarly, although it is widely accepted that the developing lexicon eventually grows to such a size that holistic perceptual strategies become inefficient (e.g. Walley, 1993), a densely packed space does not preclude holistic forms or perceptual strategies. Results of this and other neighbourhood analyses are merely suggestive because they are not based on behavioral findings. Furthermore, these studies are based on a snapshot of the lexicon at a static point in time. The lexicon in early childhood is a dynamic entity. Based on Carey's (1981) estimate, children are acquiring approximately five new unique lexical entries per day. Given these limits, neighbourhood analyses do not directly address the question of how lexical entries are stored or how new lexical items are added. In spite of these limitations, the current results do clearly show that children's lexicons contain many confusable words, and that children's lexicons actually appear denser earlier in development. From this pattern of lexical organization, children's sensitivity to fine-grained acoustic-phonetic detail is likely to play an important role in vocabulary growth.

These findings add to a growing body of evidence that shows that children become increasingly sensitive to fine-grained acoustic-phonetic information in the speech stream at a young age (Walley, 1993; Gerken *et al*., 1995; Swingley *et al*., 1999; Swingley & Aslin, 2000, 2002). While previous studies have shown that children are not as sensitive as adults to individual phonetic segments (Treiman & Baron, 1981; Jusczyk, 1986; Walley *et al*., 1986; Nittrouer & Studdert-Kennedy, 1987; Nittrouer *et al*., 1989; Walley, 1993), they do show evidence of such sensitivity when the task demands are simplified. This raises the question of how words are stored in the developing lexicon. Ferguson & Farwell (1975) originally postulated that very young children contrast and identify words as wholes. Menyuk & Menn (1979) also suggested that speech perception in young children involves units larger than individual phonemes, partly because units larger than individual phonemes are more accessible to adult perceivers in recognition tasks. Because both children and adults show evidence of sensitivity to both individual segments and to units larger than the individual phoneme, the holistic/segmental dichotomy may be unwarranted.

References

Aslin, RN., Jusczyk, PW., Pisoni, DB. Speech and auditory processing during infancy: constraints on and precursors to language. In: Kuhn, D., Siegler, R., editors. Handbook of child psychology: cognition, perception, and language. Vol. 2. NY: Wiley; 1998.

Brown, R. A first language: the early stages. Cambridge, MA: Harvard University Press; 1973.

Carey, S. The child as word learner. In: Halle, M.Bresnan, J., Miller, G., editors. Linguistic theory and psychological reality. Cambridge, MA: MIT Press; 1981.

Charles-Luce J, Luce PA. Similarity neighbourhoods of words in young children's lexicons. Journal of Child Language. 1990; 17:205–15. [PubMed: 2312642]

Charles-Luce J, Luce PA. An examination of similarity neighbourhoods in young children's receptive vocabularies. Journal of Child Language. 1995; 22:727–35. [PubMed: 8789521]

CMU pronouncing dictionary [On-line]. Available: http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Coltheart M. The MRC psycholinguistic database. Quarterly Journal of Experimental Psychology: Human Experimental Psychology. 1981; 33A:497–505.

Dollaghan CA. Children's phonological neighbourhoods: half empty or half full? Journal of Child Language. 1994; 21:257–71. [PubMed: 7929681]

Eilers RE, Oller DK. The role of speech discrimination in developmental sound substitutions. Journal of Child Language. 1976; 3:319–29.

Ferguson CA, Farwell CB. Words and sounds in early language acquisition. Language. 1975; 51:419–39.

Gerken LA, Murphy WD, Aslin RN. Three- and four-year-olds' perceptual confusions for spoken words. Perception & Psychophysics. 1995; 57:475–86. [PubMed: 7596745]

Hillenbrand J. Speech perception by infants: categorization based on nasal consonant place of articulation. Journal of the Acoustical Society of America. 1984; 75:1613–22. [PubMed: 6736425]

Jusczyk, PW. Toward a model of the development of speech perception. In: Perkell, JS., Klatt, DH., editors. Invariance and variability in speech processes. Hillsdale, NJ: Erlbaum; 1986.

Jusczyk PW, Friederici AD, Wessels JMI, Svenkerud VY, Jusczyk AM. Infants' sensitivity to the sound patterns of native language words. Journal of Memory and Language. 1993; 32:402–20.

Jusczyk PW, Hohne EA, Bauman A. Infants' sensitivity to allophonic cues for word segmentation. Perception & Psychophysics. 1999; 61:1465–76. [PubMed: 10598463]

Jusczyk PW, Luce PA, Charles-Luce J. Infants' sensitivity to phonotactic patterns in the native language. Journal of Memory and Language. 1994; 33:630–45.

Jusczyk PW, Pisoni DB, Mullennix J. Some consequences of stimulus variability on speech processing by 2-month old infants. Cognition. 1992; 43:253–91. [PubMed: 1643815]

Kučera, F., Francis, W. Computational analysis of present day American English. Providence, RI: Brown University Press; 1967.

Kuhl PK. Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. Journal of the Acoustical Society of America. 1979; 66:1668–79. [PubMed: 521551]

Kuhl PK. Perception of auditory equivalence classes for speech in early infancy. Infant Behavior and Development. 1983; 6:263–85.

Lindblom, B. Phonological units as adaptive emergents of lexical development. In: Ferguson, CA.Menn, L., Stoel-Gammon, C., editors. Phonological development: models, research, implications. Timonium, MD: York Press; 1992.

Logan, JS. A computational analysis of young children's lexicons (Tech. Rep. No. 8). Bloomington, IN: Speech Research Laboratory, Dept. of Psychology, Indiana Univ; 1992.

Luce PA, Pisoni DB. Recognizing spoken words: the neighbourhood activation model. Ear & Hearing. 1998; 19:1–36. [PubMed: 9504270]

Luce, PA., Pisoni, DB., Goldinger, SD. Similarity neighbourhoods of spoken words. In: Altmann, G., editor. Cognitive models of speech processing: psycholinguistic and computational perspectives. Cambridge, MA: MIT Press; 1990.

MacWhinney, B. The CHILDES project: Tools for analyzing talk. Hillsdale, NJ: Erlbaum; 1991.

Mattys SL, Jusczyk PW, Luce PA, Morgan JL. Phonotactic and prosodic effects on word segmentation in infants. Cognitive Psychology. 1999; 10:465–94.

Menn, L. Phonological units in beginning speech. In: Bell, A., Hooper, JB., editors. Syllables and segments. Amsterdam: North-Holland; 1978.

Menyuk, P., Menn, L. Early strategies for the perception and production of words and sounds. In: Fletcher, P., Garman, M., editors. Language acquisition: studies in first language development. Cambridge: CUP; 1979.

Nittrouer S, Studdert-Kennedy M. The role of coarticulatory effects in the perception of fricatives by children and adults. Journal of Speech & Hearing Research. 1987; 30:319–29. [PubMed: 3669639]

Nittrouer S, Studdert-Kennedy M, McGowan RS. The emergence of phonetic segments: evidence from the spectral structure of fricative-vowel syllables spoken by children and adults. Journal of Speech & Hearing Research. 1989; 32:120–32. [PubMed: 2704187]

Nusbaum, HC., Pisoni, DB., Davis, CK. Research on speech perception progress report No. 10. Bloomington, IN: Speech Research Laboratory, Department of Psychology, Indiana University; 1984. Sizing up the Hoosier mental lexicon: measuring the familiarity of 20,000 words.

Schwartz, RG. Phonological factors in early lexical acquisition. In: Smith, MD., Locke, JL., editors. The emergent lexicon: the child's development of a linguistic vocabulary. NY: Academic Press; 1988.

Swingley D, Aslin RN. Spoken word recognition and lexical representation in very young children. Cognition. 2000; 75:1–20. [PubMed: 10815775]

Swingley D, Aslin RN. Lexical neighbourhoods and the word-form representations of 14-month-olds. Psychological Science. 2002; 13:480–4. [PubMed: 12219818]

Swingley D, Pinto JP, Fernald A. Continuous processing in word recognition at 24 months. Cognition. 1999; 71:73–108. [PubMed: 10444905]

Templin MC. Speech development in the young child. 3: The development of certain language skills in children. Journal of Speech and Hearing Disorders. 1953; 17:280–5.

Treiman, R., Baron, J. Segmental analysis ability: development and relation to reading ability. In: MacKinnon, GE., Walker, TG., editors. Reading research: advances in theory and practice. Vol. 3. NY: Academic Press; 1981.

Walley AC. The role of vocabulary development in children's spoken word recognition and segmentation ability. Developmental Review. 1993; 13:286–350.

Walley AC, Smith LB, Jusczyk PW. The role of phonemes and syllables in the perceived similarity of speech sounds for children. Memory & Cognition. 1986; 14:220–9. [PubMed: 3736395]

Werker JF, Tees RC. Cross-language speech perception: evidence for perceptual reorganization during the first year of life. Infant Behavior & Development. 1984; 7:49–63.

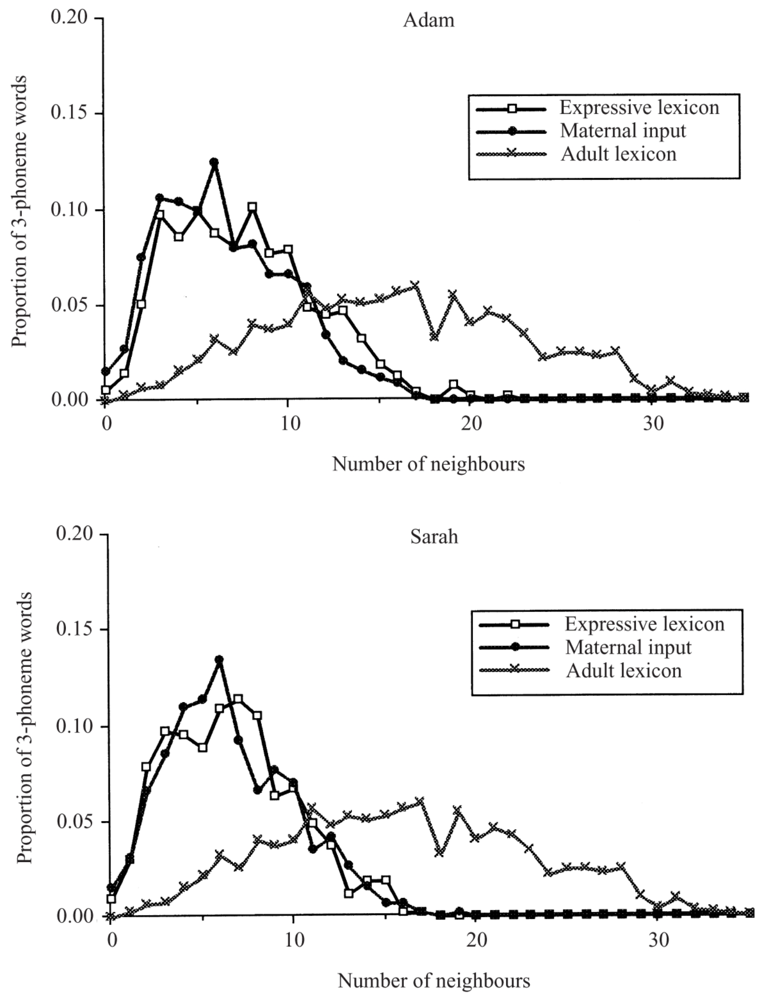

Fig. 1.

Proportion of three-phoneme words in Adam's (top panel) and Sarah's (bottom panel) expressive lexicons and maternal input as a function of the number of phonological neighbours. Adult results are overlaid.

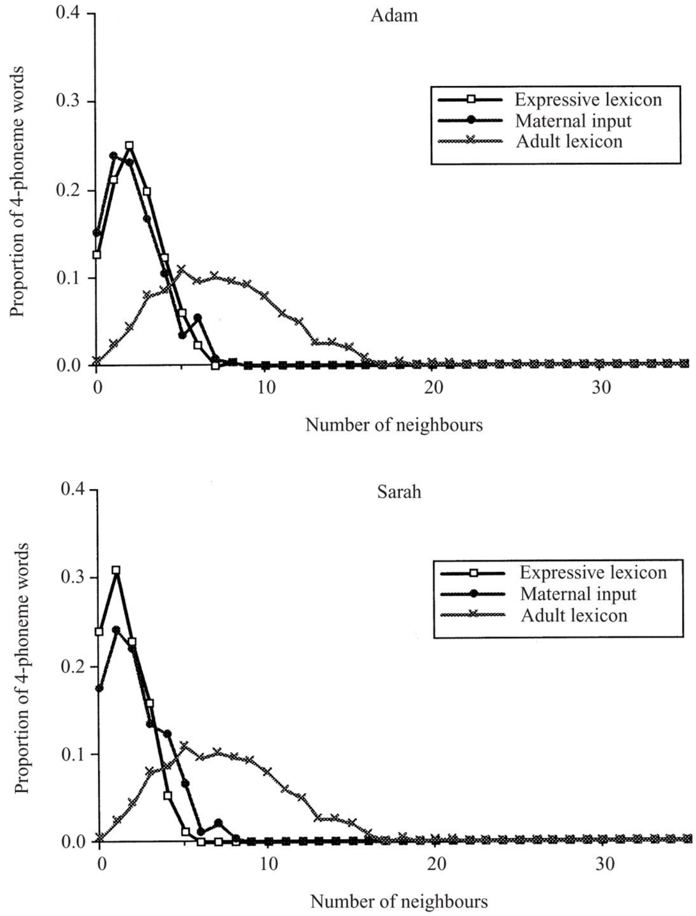

Fig. 2.

Proportion of four-phoneme words in Adam's (top panel) and Sarah's (bottom panel) expressive lexicons and maternal input as a function of the number of phonological neighbours. Adult results are overlaid.

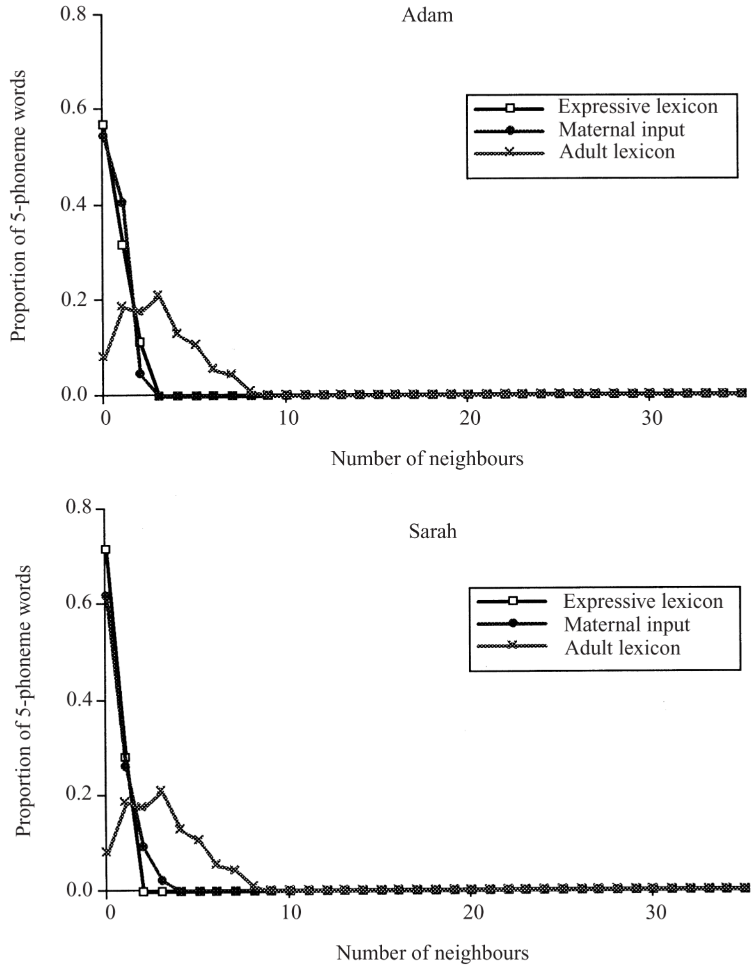

Fig. 3.

Proportion of five-phoneme words in Adam's (top panel) and Sarah's (bottom panel) expressive lexicons and maternal input as a function of the number of phonological neighbours. Adult results are overlaid.

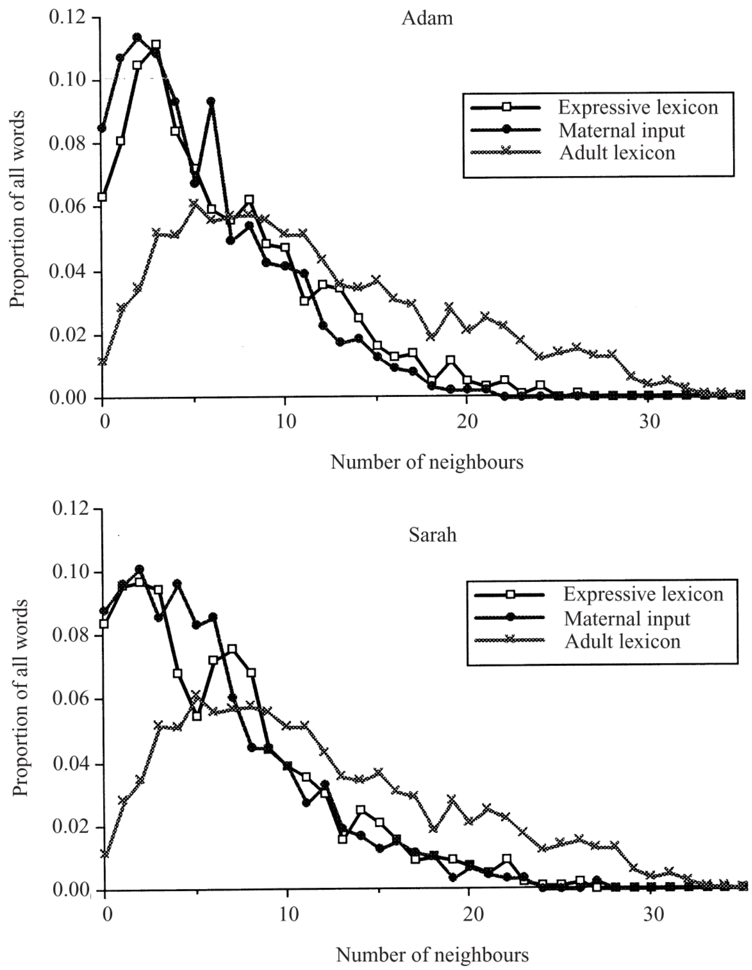

Fig. 4.

Proportion of all words in Adam's (top panel) and Sarah's (bottom panel) expressive lexicons and maternal input as a function of the number of phonological neighbours. Adult results are overlaid.

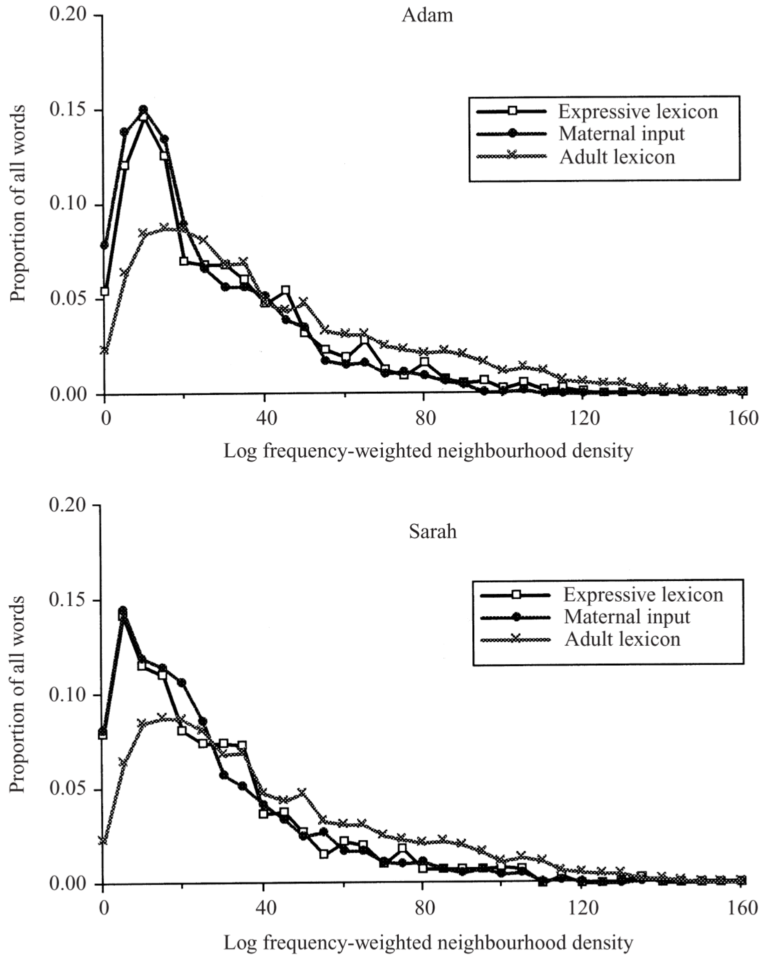

Fig. 5.
Frequency-weighted neighbourhood density for all words in Adam's (top panel) and Sarah's (bottom panel) expressive lexicons and maternal input. Adult results are overlaid.

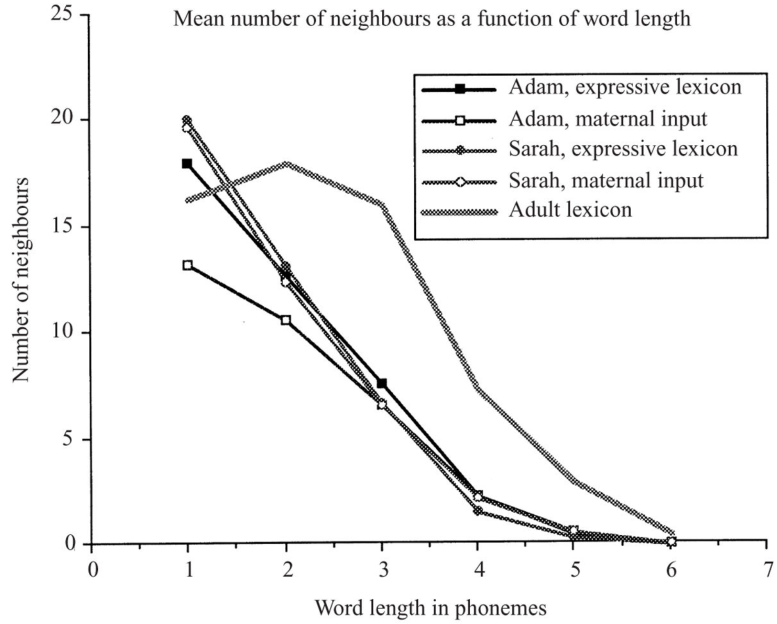

Fig. 6.

Mean number of neighbours as a function of word length in phonemes for both Adam's and Sarah's expressive lexicons, maternal input, and the adult lexicon.

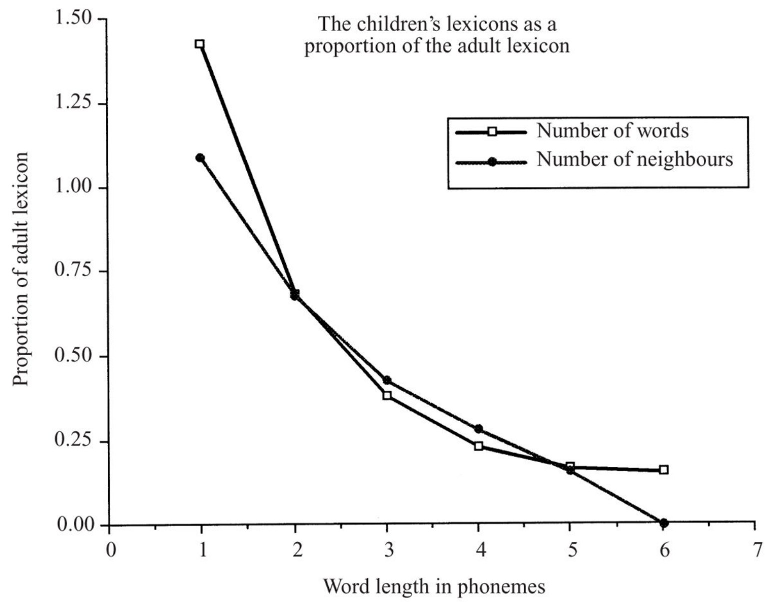

Fig. 7.

The number of words and the number of neighbours in the children's lexicons as a proportion of the number of words and the number of neighbours in the adult lexicon, for words of different lengths.

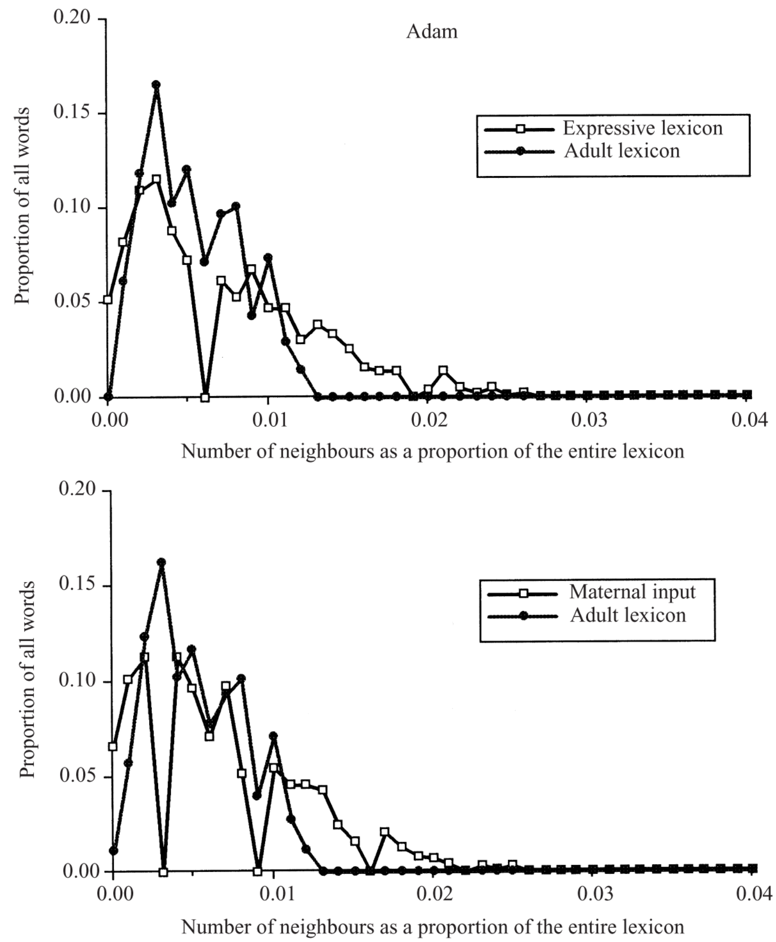

Fig. 8.

Proportion of all words in Adam's expressive lexicon (top panel) and maternal input (bottom panel) and those same words in the adult lexicon as a function of the ratio of phonological neighbours to all monosyllabic words.

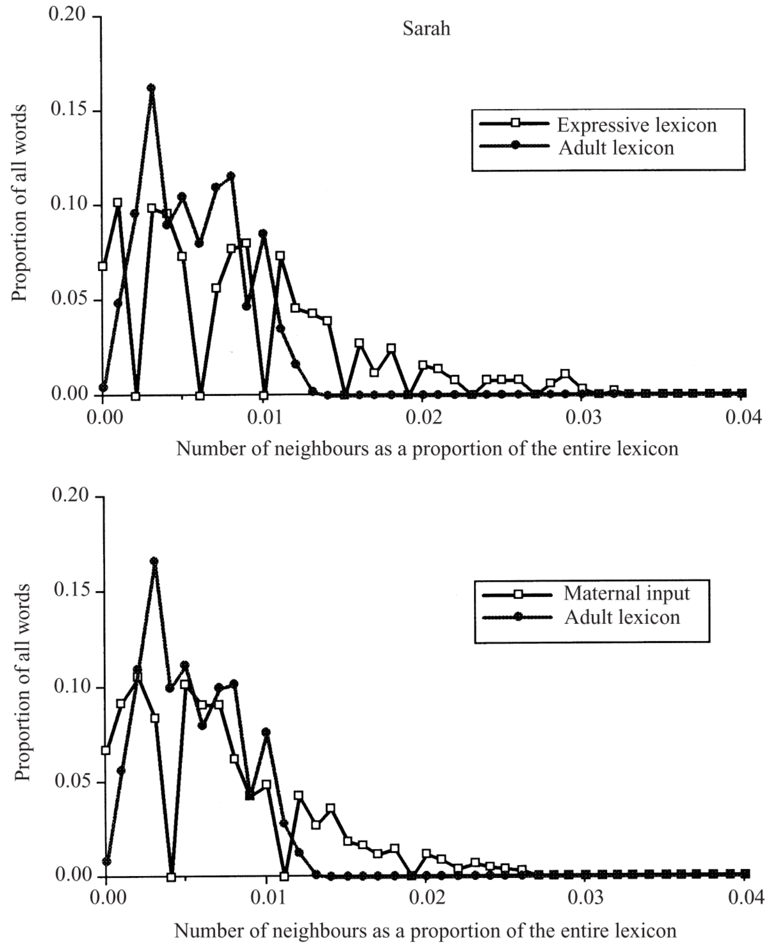

Fig. 9.

Proportion of all words in Sarah's expressive lexicon (top panel) and maternal input (bottom panel) and those same words in the adult lexicon as a function of the ratio of phonological neighbours to all monosyllabic words.

TABLE 1

Number of monosyllabic words of each length in the children's and adult lexicons

	Adam		Sarah		
Word length in phonemes	**Expressive lexicon**	**Maternal input**	**Expressive lexicon**	**Maternal input**	**Adult lexicon**
6	2	3	1	1	11
5	35	42	25	42	215
4	267	255	171	244	1010
3	492	440	429	454	1187
2	117	97	122	117	165
1	10	6	12	12	7
All words	923	843	760	870	2595

TABLE 2

Mean number of neighbours for words of each length in the children's and adult lexicons

	Adam		Sarah		
Word length in phonemes	Expressive lexicon	Maternal input	Expressive lexicon	Maternal input	Adult lexicon
6	0 (0)	0 (0)	0 –	0 –	0.45 (0.52)
5	0.54 (0.70)	0.50 (0.59)	0.28 (0.46)	0.52 (0.77)	2.93 (1.91)
4	2.27 (1.54)	2.22 (1.71)	1.51 (1.23)	2.16 (1.73)	7.28 (3.58)
3	7.56 (3.93)	6.53 (3.54)	6.60 (3.45)	6.55 (3.48)	15.96 (6.76)
2	12.63 (5.32)	10.57 (4.88)	13.15 (5.43)	12.39 (5.26)	17.96 (7.13)
1	18.00 (4.81)	13.17 (4.49)	20.00 (4.26)	19.67 (4.25)	16.29 (6.63)
All words	6.50 (5.11)	5.42 (4.34)	6.50 (5.38)	5.99 (5.00)	11.57 (7.41)