In Hall, D. G., and Waxman, S. R. (Eds.) (in press).
Weaving a Lexicon. Cambridge, MA: MIT Press.

Learning to identify spoken words

Cynthia Fisher[*]

Barbara A. Church[†]

Kyle E. Chambers[*]

[*] University of Illinois

[†] S. U. N. Y. at Buffalo

Address correspondence to:

Cynthia Fisher, Department of Psychology, University of Illinois, Champaign, IL 61820, cfisher@s.psych.uiuc.edu.

Acknowledgements: The research described in this paper was supported by NICHD grant HD/OD34715-01 and NSF grant SBR 98-73450.

1. Introduction

Before children can learn the meaning or syntactic privileges of a single word, they must first learn to identify its sound pattern. This achievement sets the stage for other aspects of language acquisition. Once children can identify even a few words in connected speech, they can begin to observe how those words are distributed relative to other elements in sentences and relative to real-world contexts. On every theory, these observations constitute the primary data for word and syntax learning (e.g., Fisher, Hall, Rakowitz, & Gleitman, 1994; Pinker, 1984; Woodward & Markman, 1998). Language acquisition begins with the perceptual analysis that reveals words and other linguistic forms in the flow of world experience.

But word perception is no simple matter. Words are not static patterns to be matched against unvarying input. Instead, their sound patterns are systematically influenced by neighboring sounds, speech rate, the idiosyncrasies of the speaker's voice and regional dialect, the prosodic structure of the utterance, and the sentence's emotional tenor (e.g., Fisher & Tokura, 1996a, 1996b; Klatt, 1980; Lively, Pisoni, & Goldinger, 1994; Miller & Volaitis, 1989; Mullennix & Pisoni, 1990; Nusbaum & Goodman, 1994). Despite all this variability, listeners recognize words as such, readily abstracting across voices and pronunciations.

The traditional approach to the problem of variability, both in linguistics and in psychology, has been to assume that the mental lexicon contains only quite abstract information about the sounds of words. On an abstractionist view, the complexities of speech perception are kept out of the lexicon, via a process of normalization that reveals a context-independent underlying sound pattern for each word (see, e.g., Chomsky & Halle, 1964; Werker & Stager, 2000, and many others in between; see Lively et al. 1994, for a review).

However, this segregation of speech processing and lexical representations has been questioned in recent years. While it is clear that listeners must readily abstract over variability in the sounds of words, it is not so obvious that the lexicon must discard all details in the process of abstraction. Growing evidence suggests that knowledge of spoken words includes context-sensitive acoustic-phonetic details that are difficult to incorporate into an entirely abstract, categorical model of word recognition. This evidence comes from several sources—recent research in laboratory phonology, studies of language change over time, and research on implicit memory for speech.

In this chapter, we will review some of this evidence and present our own recent work on the learning mechanisms underlying spoken word recognition during acquisition. The picture that emerges from this evidence, though still speculative, is one in which there is no sharp division between phonological processing and the lexicon. Instead, basic properties of perceptual learning ensure that natural language phonology and the lexicon are inextricably linked. Highly detailed lexical representations might at first blush seem inconsistent with the fundamentally abstract nature of word identification: The listener’s goal, after all, is to determine what words were said, no matter who said them, or how. However, this is not the listener's only goal. We will argue that detailed and context-sensitive representations of spoken words are needed, both to learn the native-language phonetics and phonology, and to learn to identify and compensate for variations in dialect, accent, and social register.

2. Evidence for detailed and context-sensitive encoding in the lexicon

2.1. Small-scale cross-linguistic variation in sound systems

A recent surge of research in laboratory phonology has made it clear that languages differ in their phonetic details as well as in their phonological systems (e.g., Farnetani, 1997; Keating, 1985, 1990; Pierrehumbert, 1990). Speech sounds are coarticulated—the realization of each sound is influenced by the properties of nearby sounds. Listeners readily compensate for coarticulation effects, both taking context into account in identifying each speech sound, and using anticipatory coarticulation as evidence about the identity of upcoming sounds (e.g., Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Gow, 2001; Warren & Marslen-Wilson, 1987). Some context-dependent acoustic variability in speech is probably a natural result of vocal production constraints (and thus universal). Similarly, some compensation for coarticulation by listeners may follow from relatively low-level (and again universal) auditory contrast effects (e.g., Kluender, Diehl, & Wright, 1988). But languages also vary in how speech sounds are influenced by neighboring sounds (e.g., Farnetani, 1997; Keating, 1985, 1990; Pierrehumbert, 1990, 2000). Cross-linguistic variation in the vulnerability of segments to coarticulation suggests that speakers and listeners must learn how speech sounds are affected by various contexts in their language. Learning these facts would require quite detailed and context-sensitive representations of experience with language.

For example, speakers tend to lengthen vowels before voiced (e.g., /d, b, g/) rather than unvoiced (e.g., /t, p, k/) consonants (e.g., Chen, 1970; Crystal & House, 1988), and listeners use vowel length to identify both the vowel itself and the following consonant (Gordon, 1989; Klatt, 1976). The vowel in mad is longer than the vowel in mat, and this length difference is one of the cues listeners use to differentiate the final consonants /d/ and /t/. The usefulness of this cue is due in part to auditory contrast (Kluender et al., 1988), but voicing-dependent vowel lengthening is not produced or perceived uniformly across languages. English speakers produce much longer vowels before voiced consonants, but the analogous effect is much smaller in Czech or Polish (Keating, 1985). Accordingly, native speakers of different languages assign different weights to vowel length as one of the cues to the voicing of a final consonant (Crowther & Mann, 1992).

Pierrehumbert (2000) reviews evidence for many such differences between languages in the minutiae of how speech sounds are affected by their contexts. The existence of such fine-grained differences between languages tells us that native speakers must develop a quantitative estimate of how each speech sound is formed in their native language, and how it is influenced by the surrounding context. This kind of learning requires the listener to encode speech with very fine acoustic-phonetic detail and to link these detailed records of linguistic experience with information about the context in which each sound is heard.

2.2. Sound variability in the lexicon: Evidence from language change

These considerations make clear that to conquer the contextual variability of connected speech, listeners must readily encode (even while freely abstracting over) acoustic-phonetic details and contextual information. But what tells us these details are represented in the lexicon, rather than purely at sub-lexical levels of representation? It could be that listeners collect context-sensitive and detailed representations of speech as part of the process of acquiring the phonetics and phonology of their native language, but that only abstract representations make their way into the lexicon. Such a division of labor would enable the traditional, abstractionist view of the lexicon to account for fine-grained variation in phonological patterns across languages (e.g., Werker & Stager, 2000).

On the contrary, however, considerable research on phonological change over a language's history suggests that information about phonological variation is linked with particular lexical items. Languages' sound systems change over time, due at least in part to a tendency to save effort in articulation. Sounds come to be omitted or reduced in certain contexts, in turn altering the phonotactic regularities that define legal or likely sequences of phonemes. It has long been observed that these changes do not occur uniformly across words. Instead, phonological changes are diffused gradually across the lexicon (e.g., Chen & Wang, 1975). Bybee (2000) reviews data on sound change in languages, and finds that changes in progress are more advanced in highly frequent words. This frequency effect suggests that words undergo phonological change as they are used, and that each word retains a separate range of variability.

For example, various dialects of Spanish delete, or reduce to [h], the /s/ before a consonant. This occurs inside words (e.g., <u>estilo</u> "style" → <u>ehtilo</u>), but also happens at the ends of words. A word-final /s/ is sometimes deleted even when the next word starts with a vowel, and therefore the context would not license application of a phonological rule such as "delete /s/ before a consonant." Bybee argues that this happens because the word itself accrues information about the change in its final sound. The more often a word is used, the more opportunities it has to lose its /s/, and the more speakers learn to utter that particular word without it. Bybee (2000; see also Dell, 2000; Pierrehumbert, 2000) suggests, based in part on data like these, that lexical items represent more information about phonological variation than is traditionally supposed.

Evidence that lexical items are linked with information about phonological variability does not imply that there exist no abstract representations of phonological regularities. Speakers invent or borrow new words and apply the native-language phonology to those words. When we listen to foreign words we make perceptual errors that reflect the phonotactic regularities of our native language and that are not easily attributed to the influence of particular native-language words (e.g., Dupoux, Pallier, Kakehi, & Mehler, 2001). These phenomena require representations of phonological regularities that can easily be disentangled from existing words. In addition, entirely lexical representations of phonological variation would not account well for the language-change facts described above: Representations permitting no abstraction over lexical items would not explain why language change spreads across the lexicon at all, affecting multiple words and eventually approximating a general rule (e.g., Bybee, 2000; Pierrehumbert, 2000). The picture suggested by these data is one of multiple levels of abstraction over the speech data listeners encounter. We learn words and keep track of the details of how each word can vary, but also develop more abstract generalizations that allow us to apply native-language sound patterns to new words.

2.3. Evidence from studies of implicit memory for speech

Recent research in the memory literature provides a quite distinct source of evidence for the same conclusion—that listeners routinely encode, and use in word identification, extremely detailed representations of particular words. Studies of adults' implicit memory for spoken words reveal a powerful learning mechanism that continually adapts adults' representations of words in response to listening experience (e.g., Church & Schacter, 1994; Goldinger, 1996; Schacter & Church, 1992). Each time we attend to a word, a lasting representation is created that aids later identification of the same word.

This facilitation, known as long-term auditory word priming, depends primarily on auditory, not semantic representations of experience, happens fast (on one trial), and has very long-lasting effects (e.g., Church & Fisher, 1998; Church & Schacter, 1994; Fisher, Hunt, Chambers, & Church, 2001; Schacter & Church, 1992). The facilitative effect of a single auditory presentation of a low-frequency word can be measured after a week's delay (Goldinger, 1996).

Auditory word priming functions abstractly, facilitating word identification despite ordinary variations in the sound of the word. These include changes in speech rate, fundamental frequency, speaker's voice, pronunciation, and adjacent context (e.g., Church, 1995; Church, Dell, & Kania, 1996; Church & Schacter, 1994; Goldinger, 1996; Poldrack & Church, 1997; Schacter & Church, 1992; Sheffert, 1998; Sommers, 1999).

On the other hand, many studies find that auditory word priming can be reduced by surprisingly small changes in the sounds of words. The facilitative effect of prior exposure to a word is greatest when the items used in testing match the learning items in nearly every respect. Priming is reduced when a test word is spoken in a different voice (Church & Schacter, 1994; Goldinger, 1996; Schacter & Church, 1992; Sheffert, 1998), at a different pitch or speaking rate (Church & Schacter, 1994), or when the details of the word's pronunciation change even slightly from the study episode (Church, Dell, & Kania, 1996; see also our own data from children reviewed below).

For example, many English words have multiple possible pronunciations. The word "retreat" can be spoken with a full vowel in its initial unstressed syllable (/ritrit/) or with the unstressed vowel reduced to a schwa (/rətrit/). These two possibilities exist quite generally in the English lexicon: An unstressed vowel can be reduced to a schwa, but under some circumstances will emerge as a full vowel. Church, Dell, and Kania (1996) presented listeners with a set of words subject to this kind of variation; some were produced with a reduced vowel and some with a full vowel. In a later test, listeners tried to identify low-pass filtered (therefore somewhat muffled and hard to hear) words. Some of the test words were repeated from the study phase and some were new. Of the studied words, half were presented with the same pronunciation at study and test (e.g., /ritrit/ → /ritrit/), while half changed pronunciation from study to test (e.g., /ritrit/ → /rətrit/). Adults more accurately identified studied words than new words, even if the studied words changed pronunciation from study to test. This is an example of an abstract benefit from repetition, spanning some of the ways words ordinarily vary. However, subjects most accurately identified items that retained the same pronunciation from study to test. This result implies that listeners encoded information about whether each word had been pronounced with a full or a reduced vowel. Both abstract and specific priming effects are also found when words undergo even smaller changes from study to test, including changes in vowel formant frequency that are difficult for listeners to discriminate (Church, 1995).

Data like these tell us that the rapidly-formed and long-lasting representations created whenever we attend to speech reflect acoustic-phonetic details specific to particular instantiations of spoken words. Listeners amass fine-grained and context-sensitive information about the sounds of each word, even while they readily abstract over these details to identify familiar words in new guises. Adult implicit memory for speech seems to have exactly the properties that would account for the language change effects reviewed above.

2.4. Distributional learning at multiple levels of analysis

Yet other lines of reasoning also lead to the conclusion that representations of words must be context-sensitive: In the process of identifying spoken words we must locate word boundaries in connected speech. Spoken words do not arrive at the ear pre-segmented by unambiguous acoustic cues to their boundaries. Languages recruit probabilistic cues to the location of word boundaries, including the typical stress patterns of words in the language (e.g., Cutler & Norris, 1988; Echols, Crowhurst, & Childers, 1997; Jusczyk, 1997), sequences of consonants that frequently occur at word boundaries (e.g., Mattys, Jusczyk, Luce, & Morgan, 1999; Mattys & Jusczyk, 2001), and allophonic (within-phonemic-category) differences between sounds at word edges and within words (Gow & Gordon, 1995; Jusczyk, 1997; Quené, 1992). These cues are language-specific, however, and even once learned, leave considerable ambiguity about the location of word boundaries. Theories of spoken word recognition therefore typically assume that adults find word boundaries in part by identifying the sound patterns of familiar words (e.g., Dahan & Brent, 1999; Klatt, 1980; Marslen-Wilson, 1987; McClelland & Elman, 1986; McQueen, Cutler, Briscoe, & Norris, 1995).

Word segmentation works in infancy in much the same way: By 8 months of age, infants carve words from the flow of speech by detecting locally predictable sequences of syllables

(Saffran, Aslin, & Newport, 1996; Aslin, Saffran, & Newport, 1998). After listening to four 3-syllable nonsense words randomly concatenated into a continuous stream, 8-month-olds listened longer to relatively novel 3-syllable "part words" that crossed the boundaries of the training words than to the now-familiar "words" from training. The stimuli were designed to offer no hints to word boundaries other than the highly predictable ordering of syllables within but not across word boundaries. Sequences of syllables that consistently repeat in connected speech become coherent, word-like perceptual units.

To find word boundaries via distributional analysis of sound sequences, listeners must represent information about sound patterns in context. These representations must be used abstractly to identify an old pattern in a new context, while still retaining enough specific context information to detect the repeating sequences of sounds that make up words.

Furthermore, distributional patterns in speech must be detected at multiple levels of analysis. Six-month-olds do not yet reliably recognize repeating sequences of phonemes or syllables, but they do use consistent metrical or rhythmic patterns to locate word-like units in continuous speech (Goodsitt, Kuhl, & Morgan, 1993; Morgan & Saffran, 1995; Morgan, 1996). Two-syllable nonwords with a consistent stress pattern quickly begin to cohere as perceptual units for young infants. Just as the 8-month-olds in Saffran et al.'s (1996) studies detected patterns of repeating syllables, younger infants pick up predictable metrical patterns in the speech stream.

By about 9 months, infants have begun to detect language-specific restrictions on the sequences of consonants and vowels that occur within words. These restrictions are known as phonotactic regularities. In English, for example, the /ŋ/ at the end of "sing" never occurs at the beginnings of words, the /h/ at the beginning of "hat" never occurs word-finally, and the sequence /tl/ never begins words. These are facts about English; other languages permit similar sounds to occur in different contexts. Nine-month-olds, like adults, use phonotactic restrictions and frequencies in speech processing: They listen longer to phonotactically legal or phonotactically frequent nonsense words (e.g., Friederici & Wessels, 1993; Jusczyk, Luce, & Luce, 1994), and use the probabilities of consonant sequences within words and across word boundaries to locate likely word boundaries in connected speech (Mattys et al., 1999; Mattys & Jusczyk, 2001).

For present purposes, notice that similar distributional evidence is needed to achieve word segmentation and to detect the phonotactic and prosodic sequencing regularities that characterize the native-language lexicon. In each case this learning requires a mechanism that both encodes specific information about context and permits enough abstraction to detect familiar patterns across a context change. For example, to learn that /h/ cannot occur syllable-finally, a listener must abstract over many syllables containing (and not containing) an /h/.

2.5. Summary

The evidence reviewed here suggests some conclusions about the nature of the phonological information encoded in the mental lexicon. To learn the sound system of the native language, we must encode detailed and context-sensitive representations of language experience, so that we can generate a quantitative estimate of how sounds are affected by different contexts in our language. Evidence from quite different sources suggests that these details are not kept out of the mental lexicon. The uneven, partly lexically-driven nature of sound change in languages suggests that speakers and listeners keep track of separate ranges of variability for each word (e.g., Bybee, 2000). Findings of both abstract and highly specific implicit memory for spoken words leads to the same conclusion (e.g., Church & Schacter, 1994; Goldinger, 1996). In learning to identify spoken words, listeners (a) collect detailed acoustic-phonetic information about how sound patterns are realized in different contexts, even as they (b) readily abstract across contexts to identify words as such despite contextual variability. Similar information is compiled for units of varying scope, yielding phonotactic regularities, words, and prosodic templates.

3. Word recognition in children

Do young children represent words with anything like this level of detail? It may at first seem very unlikely that they do, simply because children are so error-prone in recognizing spoken words. Adult levels of skilled word recognition take a long time to develop. Even young schoolaged children make many more errors of word recognition than adults do (see Nittrouer & Boothroyd, 1990, Gerken, Murphy, & Aslin, 1995, and Swingley, Pinto, & Fernald, 1999, for reviews). Preschoolers are much less able to judge the difference between real words and minimally different non-word foils (e.g., Gerken et al., 1995). In comprehension and word-learning tasks young children often fail to discriminate newly-taught pairs of words that differ in only one speech segment, though they do better with familiar words (e.g., 'bear' versus 'pear'; Barton, 1980). Infants and children are also slower to identify familiar words than adults are: Fernald, Pinto, Swingley, Weinberg, and McRoberts (1998) found that 24-month-olds took longer to shift their eyes to a picture of a named object than adults did.

Even worse difficulties with word recognition are found for infants at the very start of word learning. Stager and Werker (1997) habituated 14-month-olds to a single nonsense syllable paired with a display of a moving object, and found that infants did not recover visual attention to the moving object display when the sound changed to a different syllable (e.g., from "bih" to "dih"). Younger infants tested in the same task discriminate these two sounds with ease. Stager and Werker argued that infants failed to detect the sound change in what was—for 14-month-olds—a word-learning setting. In similar studies, 14-month-olds succeeded if the words were less confusable (e.g., "neem" and "liff"; see Werker & Stager, 2000). Decades of research on speech perception have established that 14-month-olds and much younger infants discriminate syllables differing only in their initial consonant. Stager and Werker's findings suggest that using these speech discrimination abilities in the service of word learning presents a more difficult task.

Some have suggested that infants' and toddlers' difficulties with word recognition reveal a fundamental lack of detail in their lexical representations (e.g., Hallé & de Boysson-Bardies, 1996; Stager & Werker, 1997; Werker & Stager, 2000). This claim is derived from the traditional view of the lexicon as containing only enough information about the sounds of words to differentiate existing words. As Charles-Luce and Luce (1995) pointed out, each word is more distinct from its neighbors in the young child's tiny vocabulary than it would be in the more densely-packed lexicon of an adult. Given a sparser lexicon, less phonetic information would be required to differentiate each of the child's words from its neighbors. If the level of phonetic detail in the lexicon is driven by the existence of minimally different pairs of words, we should expect less detailed representations of the sounds of words in young children's lexicons.

However, findings of word recognition errors in young children do not force us to conclude that children's word representations are qualitatively different from adults' (see Fisher & Church, 2001). The task of word recognition requires processing at multiple levels of analysis; each offers opportunities for error. Most word recognition theories assume a number of distinct steps in accounting for word identification (e.g., see Jusczyk, 1997). A listener must (a) accomplish an acoustic-phonetic analysis of the speech wave, (b) generate a phonological representation of an incoming candidate word, (c) compare that candidate word with words in the existing lexicon, and (d) select the best match, retrieving knowledge about the selected word. Errors could creep in at any point in this process. For example, children might tend to make more errors in initial acoustic-phonetic analysis of speech input, introducing noise into the recognition process right at the start. Children might also err in the selection of the best match in the long-term lexicon or in the inhibition of similar-sounding words. Given all these sources of error, the finding that infants and children make more errors in identifying words need not mean that children's lexical representations differ in kind from those of adults. There might be more continuity in word representation and identification across development than meets the eye.

Consistent with this possibility, recent evidence suggests that very young listeners use phonetic information to identify words much as adults do. As adults, we match incoming sounds with word candidates in the mental lexicon *incrementally* and *in parallel* (e.g., Marslen-Wilson, 1987). For example, in a referential task, listeners took longer to select a word's referent when visible competitor objects had names that sounded similar at the start (e.g., *beaker* vs. *beetle*) than when all competitors' names differed from the target from the onset (e.g., *beaker* vs. *parrot*; Allopenna, Magnuson, & Tanenhaus, 1998). This and many other findings tell us that adults make use of partial information to rule out inconsistent word candidates; under many circumstances we can identify a word before its offset.

Swingley et al. (1999) found similar evidence of incremental use of phonetic information by 24-month-olds: In a preferential-looking comprehension task, children were quicker to move their eyes to a named target picture when the target was paired with a distractor with a dissimilar name (e.g., dog vs. truck) than when the distractor and the target words began with the same sounds (e.g., dog vs. doll). Swingley et al. found no evidence that children were slower to identify words in the context of distracters with rhyming names (e.g., duck vs. truck); thus it was not overall similarity in sound that slowed word recognition, but similarity at the beginning of the word. Apparently 24-month-olds, like adults, use sound information incrementally to identify words (see also Fernald, Swingley, & Pinto, 2001).

Subsequent studies using the same technique suggest that young children exploit considerable phonetic detail in identifying familiar words (Swingley & Aslin, 2000). Children 18 to 23 months old were instructed to look at one of two pictures; on some trials they heard correct pronunciations of the pictures' names (e.g., baby), and on other trials they heard mispronounced versions of them (e.g., vaby). In both cases children recognized the intended word, and looked at the relevant picture, but they were slower and less accurate when the words were mispronounced. Finally, Jusczyk and Aslin (1995) found that 7.5-month-olds who had been familiarized with two monosyllabic words (e.g., feet and bike) later listened longer to passages in which those words were embedded than to passages not containing the familiarized words. The same effect was not found when the words were replaced in the passages with foils differing only in a single consonant (e.g., zeet and gike); infants were not fooled by these sound-alike nonwords. Although infants and children make many more errors in identifying spoken words than adults do, these findings suggest that children's long-term memories of words, and the processes by which they identify words, do not differ in kind from those of adults.

4. Children's memory for spoken words

Thus far we have argued that learners need detailed representations of experience with words in order to learn the phonological patterns of their language, and that very fine levels of detail are associated with particular words in the adult lexicon. Furthermore, as discussed in

Section 3, although children's word recognition is error-prone, recent findings suggest considerable continuity in the representations and processes underlying word recognition from late infancy to adulthood. In the remainder of this chapter we describe recent work on memory for spoken words in young children. The phenomenon of long-term auditory priming, as studied in adults, suggests that fundamental mechanisms of implicit perceptual learning create appropriately detailed and flexible representations of the sounds of words. Here we present a line of research which explores this kind of learning in toddlers and preschoolers. Our findings point toward a startling sensitivity to acoustic-phonetic detail in very young listeners, and suggest more developmental continuity in implicit memory for speech than meets the eye.

4.1. Effects of experience on word recognition

To ask whether young children's learning about speech bears any resemblance to the adult pattern described above, the first step is simply to discover whether the same rapid and long-lasting facilitation of word identification can be found in young children at all. If a child hears a word just once, does this have any effect on later identification of that word? Several studies tell us that it does, and strongly support the developmental continuity of implicit memory for speech.

Church and Fisher (1998) found patterns of long-term auditory priming in 2-, 2.5- and 3-year-olds which were very similar to those found in adults. In an elicited imitation task, children more accurately identified and repeated mildly low-pass filtered words that they had heard presented once several minutes before, than words that were not played previously. Thus 2-, 2.5-, and 3-year-olds, like adults, gained a significant perceptual boost for word identification from hearing a word just once. When children's and adults' baseline performance in the task was roughly equated through differential levels of low-pass filtering, we found no significant change in the magnitude of the repetition priming effect from age 2 to college age, despite the enormous change in lexical knowledge across this age range.

Auditory word priming is also qualitatively similar in preschoolers and adults. First, as found for adults (Church & Schacter, 1994; Schacter & Church, 1992), the advantage shown by 3-year-olds in identifying previously-heard words did not depend on whether task used in the study phase required them to retrieve the meaning of each word (choosing its referent from a set of two objects) or not (judging whether a robot said each word “really well”). This difference in encoding task did, however, affect children’s ability to explicitly judge which of the test words they had heard in the study phase (Church & Fisher, 1998). These results, as well as findings with non-words presented in the next section, tell us that the facilitation in word identification due to simple repetition is largely mediated by perceptual, rather than semantic, representations of spoken words.

More recently, we have asked whether even infants show similar effects. To address this question, we adapted an object-choice and a preferential looking task to investigate the effects of repetition on 18- and 21-month-olds’ identification of familiar words.

In a task developed with Caroline Hunt, 18-month-olds simply listened to a set of eight familiar object words (e.g., bunny, candy, keys), each presented twice in the carrier phrase "Where's the X? X!") as an experimenter entertained the infant with an unrelated stuffed toy. No referent was presented for the study words (i.e., at this point there was no bunny, candy, or keys). Following this study phase, there was a 2-minute delay during which the infant played with another set of toys. Finally the test phase began: On each trial, two objects were displayed, equidistant from the child and out of reach, and the infant heard one of the objects named twice in the same carrier phrase heard at study. Half of the 16 test items had been heard in the earlier study phase, and half were new. For studied items, the same recorded tokens of the words in their carrier phrase had been presented in the study phase. After the test words were played, the experimenter moved the two objects forward to within the infant's reach, and said "Can you get it?" All choices were applauded. The measure of interest was how often the infant chose (touched first) the named object; choices were coded off-line from silent video. Infants were significantly more likely to select the named object when the test word had been studied. Hearing a familiar word just twice in a non-referential context made infants more likely to identify that word successfully several minutes later in a comprehension task.

We found the same rapid and long-lasting facilitation for spoken word identification in a visual preference comprehension task with 21-month-olds. The children sat on a parent's lap in front of two side-by-side video screens. All sound stimuli were played from a speaker centered between the two screens. During the initial study phase a set of words was presented twice, in the same carrier phrase used in the object-choice task described above. No referents for these words were shown; an aquarium video played on both screens as the study words were presented. A distractor phase followed, during which both video screens showed an unrelated children's video. On each trial of the final test phase, children saw different objects pictured on the two screens; the soundtrack instructed children to look at one of the pictures (e.g., "Find the truck. Truck!").

Early in each 6-second test trial the child heard an attention-getting phrase such as "Hey, look!", followed by the target words in their carrier phrase. The onset of the first presentation of the target word occurred when the pictures had been visible for 3 seconds; the onset of the second presentation occurred 1.5 seconds later. The target word was repeated to ensure that all children would eventually identify every word. The 1.5-second period between the onset of the first and second repetitions of the target word provided a test window within which to measure word identification speed and accuracy. We coded, frame by frame from videotape, the children's visual fixations to the two screens during the test period.

Children were faster and more accurate in identifying the studied items. We analyzed the mean proportion of time spent looking at the matching picture, divided into three 500 msec segments. Children were more likely to look at the target for studied than for new items; this effect was most striking early in the test period. Further analyses revealed a significant advantage for studied items during the first 100 msec following the onset of the target word. This very early advantage for studied words is almost certainly not due to faster identification of the sound pattern of the target word itself. Since it takes time to plan and initiate an eye movement (e.g., 150-175 msec for adults with no uncertainty about target location; see Rayner, 1998, for a review), the slight advantage for studied targets within 100 msec of target word onset appears too soon to be plausibly interpreted as a response to the target word itself.

There are two likely sources for this apparent precognition on the part of the children: First, in connected speech some information about upcoming sounds can be gained from coarticulation cues (e.g., Dahan et al. 2001; Gow, 2001; Warren & Marslen-Wilson, 1987). The target-distractor pairs used in this study had names that were dissimilar at the onset (e.g., candy and bug), making it possible that some phonological information about the target could be detected before the word itself began. Second, 17-month-olds spend more time looking at pictured objects whose names they know well (Schafer, Plunkett, & Harris, 1999). Children may have been better able to retrieve the names of target pictures on their own if they heard their names in the study phase. If so, then a slight early preference for the target might appear for studied items, even before any phonological information about the target word could be detected.

To isolate the effect of the study episode on perceptual identification of the test words, we examined shifts in fixation that occurred after the onset of the target word. Following Swingley et al. (1999), we examined trials in which the child happened to be fixating the distractor at the onset of the target word. We then measured the latency of shifts in fixation from the distractor to the target picture. For this analysis, we included only shifts that occurred within the 1.5-second test window, and discarded shifts that occurred within 200 msec of target word onset, on the assumption that these were too fast to be responses to the word. The 21-month-olds were reliably quicker to move their eyes from the distractor to the target picture for studied than for new targets. This result suggests that children were quicker to identify the sound patterns of familiar words if they had heard those words just twice in the study phase.

All of these studies show that toddlers' and preschoolers' spoken word identification benefits incrementally from repetition. Eighteen-month-olds were more accurate in choosing a named object, 21-month-olds were quicker to look at a named object, and 2- and 3-year-olds more accurately repeated a test word, if the word had been played once or twice several minutes earlier. Just as for adults, each experience with a word adds perceptual information to the system used to identify spoken words. This information makes the system better suited to identifying the same words in the future.

4.2. Encoding the details of pronunciation

The evidence reviewed above makes it clear that early word identification benefits from each encounter with the sounds of words. In all cases so far, however, we have examined the effect of hearing the same recorded token of the word at study and test; these data therefore tell us nothing about the nature of the long-term memories added to the word identification system each time young children listen to speech. In this section we focus on the content of these representations: How much phonetic detail is encoded in the long-term memories used for spoken word identification, and how readily do children abstract across these details?

In two experiments, we focused on words that can be pronounced in more than one way. For example, any standard dictionary gives the following pronunciation for the word potato: /pəteto/, with two /t/ consonants. But every speaker of standard American English knows that this word is often pronounced with only one /t/, and with the second /t/ reduced to a tap or flap. Both pronunciations are recognizable as tokens of the same word. Many English words permit the same variation (e.g., turtle, vegetable, daughter). In another series of experiments, we asked whether specific information about variation in a word's pronunciation is retained in the memory representations used to identify spoken words.

Church, Fisher, and Chambers (in preparation) used an elicited imitation task to study memory for word variability in 2.5- and 3-year-olds. As in the experiments described in the previous section, children first simply listened to a set of words presented while they watched an unrelated children's video. After working on a puzzle for at least 2 minutes, they heard and were asked to repeat test words. The test words were presented quietly, making them hard to hear; we measured the children's accuracy in repeating the words.

In the first experiment in this series, we tested for abstract priming spanning a change in pronunciation. Children heard 16 study words and were tested on 32 words. Half of the test words had been heard in the study phase and half were new. All of the studied words changed pronunciation from study to test. For example, a child who heard the /t/-version of a word in the study phase (e.g., /pəteto/) would now hear a version of the same word with a more /d/-like flap at test (/pəteɾo/). Both 2.5- and 3-year-olds identified and repeated the studied items more accurately than the new items. Since the studied words always changed pronunciation from study to test, this result shows that children's rapidly-formed perceptual representations of words contain components abstract enough to support word identification across acceptable variations in pronunciation. Children profited from simply hearing a word repeated, even if that word sounded a little different the next time it was heard.

Did children also retain information about how each word was pronounced, in addition to abstracting over the change? In a second experiment using the same materials, we asked whether children's memories of spoken words included information about whether each word had been pronounced with a /t/ or a /d/-like flap. The design and materials were just as described above; the only difference was that in the second experiment, all of the test items had been heard in the study phase; thus all were studied words. In this case, however, half of the test words appeared in the same pronunciation at study and test (e.g., /pəteto/ → /pəteto/), while half changed pronunciation from study to test (e.g., /pəteɾo/ → /pəteto/). Both 2.5- and 3-year-olds more accurately identified and repeated words that were pronounced the same way at study and test, than those that changed pronunciation. Children benefited most from past experience with a word if that word was pronounced in just the same way, rather than with a legal change in pronunciation. Taken together, these two studies show that on minimal exposure, children rapidly create long-term representations of spoken words that both abstract across legal variation in word pronunciation and retain quite specific information about how each word was pronounced.

Crucially, this information had to be linked with particular words heard in the experiment: Each child heard an equal number of words pronounced with a /t/ and with a flap, both at study and test. The pronunciation-specific priming effect we measured could only have been due to encoding a separate history of pronunciations for each word. This finding is strikingly reminiscent of the language-change phenomena reviewed earlier. Even 2.5- and 3-year-olds encode detailed and lexically-specific information about the sound patterns of words. These lexically-specific memories, as part of the system for identifying spoken words, could help explain why sound changes are not uniform across the lexicon.

4.3. Encoding sub-phonemic details

All of the word priming studies described so far have examined memory representations of existing English words. Children in these studies had probably already established a robust representation for each word before they came to be in our experiments; they also had probably learned that English words can vary in whether a /t/ is released, unreleased, or produced as a flap. It might be that the patterns of abstraction and specificity we find, or even the rapid facilitation for later word identification conferred by one or two repetitions of a word's sound pattern, depend on the existence of these prior lexical representations.

Fisher, Hunt, Chambers, and Church (2001) asked whether young children would show priming at all for entirely unknown words (nonwords), and if so, whether their rapidly-created representations of the new words exhibited the same flexible combination of abstraction and specificity found in our studies with real words. We again used an elicited imitation task: Children simply listened to a set of nonwords, each presented twice in an initial study phase. After a 2-minute distractor task they listened to and tried to repeat nonwords. As before, the test items were presented quietly, to increase the perceptual demands of the elicited imitation test. In a first study, 2.5-year-olds more accurately identified and repeated consonant-vowel-consonant (CVC, e.g., "biss", "yeeg") nonwords that they had heard just twice in the initial study phase. This finding tells us that the kind of rapid and long-lasting facilitation we measured for toddlers and preschoolers with real words can be found with entirely novel words as well. No previously established lexical representation is required for rapid sound-pattern learning to occur.

With this result in hand, we went on to explore the contents of young children's representations of novel words. Three-year-olds heard two-syllable nonwords (CVC.CVC, e.g., "deece.gibe" /dis.gaɪb/, "tull.yave" /tʌl.jev/). Each syllable was recorded in two different contexts. For example, the component syllables of /dis.gaɪb/ were re-paired with other syllables to form /tʌl.gaɪb/ ("tull.gibe") and /dis.jev/ ("deece.yave"). All of the nonsense syllable pairs were recorded as coarticulated disyllables, with greater stress on the first syllable. This was done to encourage children to perceive them as possible two-syllable words. The resulting items had the phonological structure of English words like "council" or "bathtub." We arranged these items into various study and test lists to examine how abstract, and how specific, children's representations of the nonwords were.

First, we compared test items whose component syllables had been heard in an initial study phase to test items that were entirely new.   All of the studied items had changed recording context (and therefore recorded syllable token) from study to test. For example, a child who heard /tʌl.gaɪb/ and /dis.jev/ in the study phase would hear /dis.gaɪb/ and /tʌl.jev/ at test. Three-year-olds showed abstract priming for these recombined nonsense words. Children more accurately identified and repeated syllables that they had heard in the study phase then syllables that they had not heard, even though different tokens of the studied syllables were presented in a new word-like context at study and test. The rapidly-formed perceptual representations that aid word identification permit abstraction across acoustic and context variability, even for brand-new items.

Second, we looked for specificity effects, asking whether children encoded details relevant to the context in which they originally heard each syllable. In this study, 3-year-olds were again tested on two-syllable nonwords; all of the component syllables of the test items had been heard in the study phase. Half of the test items were presented in the same disyllabic context (and the same recorded token) at study and test, while half changed context (and recorded token) from study to test. The children more accurately identified and repeated syllables that were heard in the same context at study and test. This finding shows striking specificity in 3-year-olds' representations of brand-new spoken words. Based on just two exposures to a possible two-syllable word, children later more accurately identified test items if the same tokens of those syllables were heard in the same context.

What specific features of the disyllabic nonwords did the children encode?   There are two classes of possibilities. Syllables that changed context from study to test differed in both the adjacent context and the recorded token. The context change itself, or changes in syllable token, or both, could have caused the reduction in priming for changed-context items. Syllables recorded in different contexts will differ in systematic coarticulatory cues at the boundary between the two syllables. We argued in the opening sections of this chapter that both context information and detailed acoustic representations are needed to support phonological learning and spoken word identification; therefore we predicted that both context and token information should be encoded in memories of syllables, and that a change in either would reduce the magnitude of priming effects.

Fisher et al. (2001) reported a final study showing that the change in syllable token from study to test was responsible for a significant part of the context-specificity effect described above. In order to disentangle the effects of context from changes in the syllable token itself, we spliced the second syllables out of each of the disyllabic nonwords used in the previous experiments. Children heard the original disyllabic nonwords in an initial study phase (e.g., /dis.gaɪb/), participated in the same distractor task, and then listened to and tried to repeat the final syllables spliced out of their disyllabic contexts and presented in isolation (e.g., /gaɪb/). All of the test syllables were syllables the child had heard in the study phase. Half of the test syllables were the same syllable tokens, spliced out of the same-context disyllabic items (e.g., /dis.gaɪb/ → /(dis)gaɪb/), while half were different syllable tokens, spliced out of the changed-context disyllabic items (e.g., /dis.gaɪb/ → /(tʌl)gaɪb/). Children more accurately identified the nonsense syllables at test if they heard the same syllable token that had been presented at study, than if they heard a different token of the same syllable.

When children attend to speech, they encode token-specific detail about brand-new syllables. In our experiments, children benefited a little less from repeated exposure to a syllable if the syllable was presented in a different recorded context, or even if a different recorded token of the same syllable, with different coarticulation cues, was presented. At the same time, however, we measured a more abstract facilitation—syllables that had been heard before were repeated more accurately than those that had not, even though the studied syllables changed both recorded token and context from study to test. Taken together, these findings tell us that young children encode even brand-new words in excruciating detail, yet readily abstract over those details to identify a new rendition of the same syllable.

These token-specificity effects make it clear that the level of detail relevant for memory representations of spoken words does not depend on the existence of minimal pairs in the lexicon: There is no minimal pair of words that would force children to encode token-specific information in the mental lexicon. These findings strongly suggest that young children encode highly detailed representations of speech, and link these detailed representations with particular lexical items. Such representations would permit children to learn the details of how sounds are pronounced and how individual lexical items vary in their pronunciation.

4.4. Context-sensitive encoding

We have argued that very detailed and context-sensitive encoding of sound patterns is needed for children to learn the ways in which various sounds are affected by context in their language. Did context play any role in the context-change effect shown in Fisher et al. (2001)? In order to isolate the influence of context from syllable-internal token changes, Chambers, Fisher and Church (in preparation) tested 3-year-olds and adults with spliced-together versions of disyllabic nonword stimuli like those described in the previous section.

As in the previous context-change study, children were tested on disyllabic items all of whose component syllables had been heard in the study phase. Again, half of the test items were presented in the same disyllabic context at study and test, while half changed context from study to test. The crucial difference from the prior studies was that the same recorded syllable tokens were always presented at study and test; context-change items were simply the same syllable tokens re-spliced into a new disyllabic nonword. Children were tested in the elicited imitation task, and we measured the accuracy with which they repeated the novel words. Adults heard test items over headphones and tried to repeat them as quickly as possible; we measured reaction time rather than repetition accuracy.

In two studies we found small but consistent effects of a change in adjacent context for both children and adults, even though the recorded syllable token remained the same from study to test. Adults were reliably quicker, and children more accurate, in repeating same- than changed-context items. These results suggest that listeners encoded information about adjacent context, and received slightly less facilitation from repetition if the context changed.

Interestingly, we also found variability among the items in the magnitude of the context change effect. Using the adult reaction time data, we examined the possibility that some consonant sequences at the syllable boundary in our CVC.CVC items might sound more like they belonged within a single word than others. Phonotactic regularities govern sound sequences within words; consonant sequences vary in how frequently they appear within versus across the boundaries of English words. For example, the sequence /ft/ is common in English words (e.g., after, gift, lift), while /vt/ is rare within English words but occur across word boundaries fairly often (e.g., love to). Mattys and Jusczyk (2001; see also Mattys et al., 1999) found that 9-month-olds were better able to detect repeated words in connected speech if those words were edged with consonant sequences like /vt/, which are very unlikely to occur within words (and therefore good for word segmentation). Following Mattys et al. (1999), we computed both within-word and between-word transitional probabilities for the consonant sequences at the syllable boundaries of our disyllabic nonsense words, based on a phonetically transcribed on-line dictionary of about 20,000 English words (see Nusbaum, Pisoni, & Davis, 1984)[1]. In the adult reaction time data, we found that the within-word phonotactic frequency of the consonant sequence at the syllable boundary was significantly correlated with the magnitude of the context-change reduction in priming. Adults were slower to repeat studied syllables that had been spliced into a new disyllabic context at test; this context-change effect was larger for items that had more word-like consonant transitions at the syllable boundary.

[1] We estimated between-word transitional probabilities for the consonant sequences based on the product of the probability that each first consonant appears at the ends of words in the same dictionary, and the probability that each second consonant appears at the beginnings of words in the dictionary.

These findings suggest that both children and adults can encode information about the adjacent context when they listen to speech, and link that context information with their representation of a particular syllable. We also found intriguing evidence that phonotactic probabilities affect the likelihood of encoding context information across a syllable boundary. Future studies will examine the possibility that language-specific cues relevant to word segmentation, including stress pattern, syllable structure, and phonotactic regularities, partly determine the effective units of memory storage in word perception.

4.5 Context-sensitive encoding at multiple levels of analysis

The perceptual priming data from toddlers, preschoolers, and adults summarized above creates a picture of a very flexible lexicon full of highly-detailed and context-sensitive representations of spoken words. Contrary to the traditional view of lexical-phonological representations as abstract, a considerable amount of detail, including information about the context in which each sound appears, makes its way into the mental lexicon. Preschoolers and adults readily abstract over this wealth of detail and context information, to identify words or syllables as such. We also argued in earlier sections, however, that similarly context-sensitive and flexible learning must take place at a variety of levels of analysis over the same speech data.

To learn native-language phonotactic regularities, for example, listeners must encode information about the contexts in which each speech segment (i.e., each consonant or vowel) can occur, abstracting across syllables to detect sub-syllabic regularities. Even in abstracting these sub-syllabic regularities, however, not all context information can be lost: Phonotactic regularities are restrictions on the context in which each sound can occur, including syllabic position (e.g., "ng" can be syllable-final but not syllable-initial in English) and adjacent sounds (e.g., only a small subset of consonants can appear syllable-finally after a long vowel; Harris, 1994). Languages also have phonotactic regularities less absolute than the ban on initial "ng" in English: Vocabularies are characterized by probabilistic phonotactic patterns, and listeners (including 9-month-old infants) are sensitive to these probabilities (e.g., Jusczyk et al. 1994; Kessler & Treiman, 1997; Mattys et al. 1999; Mattys & Jusczyk, 2001).

The findings we have reviewed above suggest the operation of exactly the kind of implicit learning and memory mechanism required to learn these regularities: Information about sound patterns is encoded flexibly, both retaining fine acoustic detail and information about context, and permitting abstraction across these context-sensitive representations. If such flexible representations of speech are laid down with each listening experience, then phonotactic regularities should emerge from the accumulation of incremental changes derived from listening experience.

Dell, Reed, Adams, and Meyer (2000) found that adults could learn new phonotactic regularities in a speech production task. On each trial, adults saw a set of four nonsense syllables (e.g., mak ghin saf hing) and were asked to repeat them in order, in time to a metronome. The metronome encouraged a fast speech rate, increasing the likelihood of slips of the tongue. Participants repeated many such lists over four experimental sessions on different days. All stimulus syllables complied with the phonotactic restrictions of English (e.g., "ng" always syllable-final, "h" only initial), but the experimental stimuli also displayed new phonotactic regularities, a subset of the English possibilities. For example, a subject might see materials in which /f/ occurred only as a coda consonant (syllable-final, e.g., "saf"), and /s/ only as an onset consonant (syllable-initial). Dell et al. examined the speech errors the subjects made during this task. They found, as expected, that speakers' errors respected the phonotactic regularities of English essentially 100% of the time. In addition, however, participants' speech errors rapidly came to respect the experimentally-induced phonotactic regularities: Errors that were "legal" with respect to the experimental phonotactic constraints were much more likely than errors that violated those constraints. The experiment-wide phonotactic constraints were honored even in errors producing syllables that were never presented in the experiment; this extension to novel syllables suggests the subjects in these experiments learned relatively abstract, sub-syllabic regularities in sound sequences.

Dell et al. attributed their findings to implicit learning in the language production system: Over many trials, the production system was given more practice with (for example) an initial /s/, and even in error became more likely to produce this familiar pattern. Could such rapid phonotactic learning occur based on listening alone?

Onishi, Chambers and Fisher (2002) examined this question by asking adults simply to listen to sets of nonsense syllables that exhibited novel phonotactic regularities, again a subset of the possibilities of English. Subjects listened to sets of CVC syllables in which consonants were artificially restricted to syllable onset or coda position. Onishi et al. chose constrained syllable sets that could not be differentiated based on a single phonetic feature or set of features (group 1: /b, k, m, t/, group 2: /p, g, n, ch/). Subjects heard either group 1 consonants as onsets and group 2 consonants as codas (e.g., /bæp/), or the reverse assignment (e.g., /pæb/).   Approximately 25 syllables were presented during the initial study phase, presented 5 times each in a simple cover task in which subjects rated clarity of articulation. After a two-minute distractor task subjects heard more CVC syllables and tried to repeat them as quickly and accurately as possible. Onishi et al. measured the latency to begin speaking. The items in the test phase included two kinds of novel syllables—legal syllables, consistent with the experimental phonotactic constraints established in the study phase, and illegal syllables, violating those constraints.

The adult subjects were reliably quicker to repeat the legal than the illegal items, revealing that they had learned the novel phonotactic restrictions imposed in the study phase. Furthermore, Onishi et al. anticipated that the phonotactic learning found in this study would diminish during the test phase: The test phase included as many illegal as legal syllables, so throughout the test phase subjects were given evidence against the phonotactic restrictions established earlier in the experiment. This prediction was upheld: The advantage for legal items was significant in the first half of the test trials, and was gone in the second half.   Adult listeners readily acquired restrictions on the syllable positions in which consonants could occur, simply based on listening to a set of syllables.

In a second experiment, Onishi et al. asked whether listeners could acquire phonotactic constraints involving more complex interactions between sounds. Adults heard syllables exhibiting second-order phonotactic regularities in which consonant position depended on the adjacent vowel: For example, /b/'s might be onsets and /p/'s codas for syllables containing the vowel /æ/, but these positions would be reversed for syllables containing the vowel /I/. Thus each consonant appeared equally often in both onset and coda position in each study list; in order to pick up the phonotactic regularities embedded in these materials, subjects had to attend to the relationship between consonants and vowels. In the test phase, subjects were again faster to repeat new syllables that were legal than those that were illegal. As expected, Onishi et al. found that these second-order phonotactic restrictions waned during the test phase, as subjects heard as many illegal as legal test items.

Although the subjects in these experiments were adult native speakers of English, in only a few minutes of listening they picked up new regularities in the sequencing of segments within syllables. These findings testify further to the great flexibility of phonological representations: Listeners abstracted across syllables to learn that consonants were restricted to a subset of syllable positions. At the same time, listeners readily kept track of context information, thus learning that consonant position could depend on the adjacent vowel. These listening-based effects parallel the effects of speaking practice found in Dell et al.'s (2000) studies. Just as ongoing speaking practice makes it easier to produce phonotactically frequent patterns, listening practice makes it easier to identify frequently-occurring sound patterns, even in new syllables.

Ongoing studies ask how abstract the phonotactic learning uncovered in these experiments was. Consonant-vowel transitions contain considerable information about phoneme identity, and useful estimates of phonotactic probabilities rely on the frequencies of pairs of phonemes as well as on position-sensitive phoneme frequencies (e.g., Bailey & Hahn, 2001). Subjects in Onishi et al.'s experiments generalized the patterns they acquired in the study phase to new syllables, but may have done so by learning transitions between consonants and vowels rather than by establishing something more like the general rule "/b/'s are onsets." Preliminary data suggest, however, that adults can generalize newly-learned consonant position regularities to new vowels. Such findings would tell us that listeners establish abstract, vowel-independent knowledge of the syllable positions in which each consonant occurs, even as they detect patterns of co-occurrence between consonants and vowels.

5. What good are detailed lexical representations?

We have presented many kinds of evidence that call into question traditional assumptions about how speech perception and word identification work. A common theme among diverse views of word identification has been the assumption that the lexicon contains only abstract information about the sounds of words (see Lively, Pisoni, & Goldinger, 1994, for a review). This assumption is compelling in part because it seems to follow almost inevitably from the demands of the word recognition task—we do identify words as such, as abstract phonological patterns, across great variability in their acoustic-phonetic details. The assumption that word representations are abstract has held even greater sway in the developmental literature on word representations. As reviewed in Section 3, several researchers have invoked underspecified word representations in children to account for their relatively poor performance in spoken word identification (e.g., Werker & Stager, 2000; Hallé & de Boysson-Bardies, 1996).

The conclusion demanded by the data presented in this chapter, however, is that phonological representations in the mental lexicon are not so abstract, either for adults or for young children. On the contrary, exceedingly fine acoustic-phonetic details make contact with the lexicon. Our data testify to the continuity across development of the implicit learning and memory mechanisms relevant to speech processing. Preschoolers and infants, like adults, possess a learning mechanism that creates and updates long-term representations of the sounds of words to reflect each experience with language. One or a few exposures to a word facilitates later identification of the same sound pattern. The representations laid down when young children listen to speech include arbitrarily fine details about each token. This level of detail is far beyond that required to differentiate minimal pairs of words like bat and pat. We argue that the same perceptual learning mechanisms which permit adults to adapt continually to new words, new speakers, new dialects, and acoustic circumstances, play a role in the development of the auditory lexicon from the start (see also Nusbaum & Goodman, 1994).

The tendency for even infants to retain item-specific perceptual information has been uncovered in other domains as well, including memory for particular musical performances (Palmer, Jungers, & Jusczyk, 2001) and the identification of a colorful mobile as a cue to carry out a previously rewarded action (e.g., Hartshorn, Rovee-Collier, Gerhardstein, Bhatt, Klein, Aaron, Wondoloski, & Wurtzel, 1998).

Evidence of the great specificity of adults' memories for spoken words, as well as the historical sound-change data reviewed above, have persuaded some theorists to adopt episodic or instance-based theories of spoken word recognition (e.g., Bybee, 2000; Goldinger, 1998; Jusczyk, 1997; Pisoni, 1997), based on episodic theories of memory and categorization (e.g., Hintzman, 1986; Logan, 1992; Nosofsky, 1988). On an episodic view, each encounter with a word is retained in the memory stores used for word identification. Episodes are context-sensitive and finely detailed, and abstraction across tokens is achieved based on either the central tendency of the set of tokens activated during the recognition process, or direct similarity comparisons with individual episodes. Another class of models developed to account for the specificity of the memories that participate in word identification invokes both episodic and abstract representations (e.g., Church, 1995; Church & Schacter, 1994; Moscovitch, 1994; Schacter, 1994). Our data do not permit us to choose between such models, though we tend to prefer a hybrid model that includes both categorical and episodic representations. For now, however, we can conclude that traditional abstractionist models of the lexicon are inadequate to explain the data. Lexical representations of the sounds of words are linked with token-specific and context-sensitive details.

These details might seem to be mere obstacles to the main task of recognizing what word has been uttered. We have argued, however, that representations encoding very fine acoustic-phonetic detail and permitting easy abstraction over these details are required to learn the sound system of the native language. Languages vary in the fine points of how speech sounds are produced and how they are influenced by various contexts. Thus for example, English-learning listeners come to use vowel-lengthening as a powerful cue to final consonant voicing, while this cue is less relevant for Polish-learning listeners. In order to learn such sound-pattern regularities, learners must gather quite detailed information about speech sounds and the contexts in which they have occurred.

Finally, we should mention another set of phenomena that plainly demonstrate listeners' and speakers' ability to track the details of variability in language use. The sounds of words depend heavily on stylistic factors. Klatt (1980) gave a casual-speech transcription of the sentence "Would you hit it to Tom?" to dramatize the importance of adjacent context in ordinary word identification. As ordinarily produced, this sentence displays many contextual modifications, such as the palatalization that merges the consonants at the boundary of "would you," the deletion of one of the adjacent /t/s in "it to," and the demotion of the /t/ in the middle of "hit it" to a tap or flap. These are characteristics of <u>casual</u> American English speech. If we speak in a more formal setting, the same words can sound quite different.

More generally, language use is an inescapable part of the way that speakers create a social persona, and listeners diagnose the life history, education, and affiliations of others. Particular pronunciations of words can carry social stigma or prestige, and therefore be adopted or avoided depending on the formality of the setting and on the speaker's social aspirations and group identifications (e.g., Wardaugh, 1998).

For example, Labov (1972) tracked the appearance of post-vocalic /r/ in the speech of New Yorkers. Pronouncing words like <u>fourth</u> and <u>floor</u> with or without an /r/ has social meaning: Listeners rated speakers who included the /r/'s higher in job prospects, but rated speakers who omitted their /r/'s superior in their likely toughness in a fight. Labov collected utterances of "fourth floor" from salespeople in department stores varying in elegance, and found that salespeople in the fanciest stores were most likely to include the /r/'s. Men and women also tend to maintain separate ranges of variability in language use, and to enact different style changes as they move from one context to another (Labov, 1998). Girls and boys pick up on these differences, presumably much as they pick up on gender-typed behavior in non-linguistic domains. Though we are often unaware of it, we negotiate personal styles through the way we use language, including the details of how words are pronounced and how they are affected by their linguistic contexts.

In summary, language use is made up of more than words, characterized by their abstract phonological form. To function in a linguistic society, listeners must detect words, but also native-language phonological patterns at varying scales, changes in word pronunciation that do not apply uniformly across the lexicon, and styles of speech that are laden with social meaning. We have described evidence that young children's rapidly-formed representations of spoken words have the right combination of abstraction and specificity to learn about speech at all these levels. By studying the implicit learning and memory mechanisms that create representations of spoken words, we can explore how languages come to be structured at so many levels.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of Memory & Language, 38, 419-439.

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. Psychological Science, 9, 321-324.

Bailey, T. M., & Hahn, U. (2001). Determinants of wordlikeness: Phonotactics or lexical neighborhoods? Journal of Memory & Language, 44, 568-591.

Barton, D. (1980). Phonemic perception in children. In Child Phonology, Vol. 2: Perception (pp. 97-116). New York: Academic Press.

Bybee, J. (2000). Lexicalization of sound change and alternating environments. In M. B. Broe & J. B. Pierrehumbert (Ed.), Papers in Laboratory Phonology V: Acquisition and the lexicon (pp. 250-268). New York: Cambridge University Press.

Carroll, J. B., Davies, P., & Richman, B. (1971). The American Heritage word frequency book. New York: American Heritage Publishing Co.

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

Chen, M. (1970). Vowel length variation as a function of the voicing of the consonant environment. Phonetica, 22, 129-159.

Chen, M. Y., & Wang, W. S.-Y. (1975). Sound change: Actuation and implementation. Language, 51, 255-281.

Chomsky, N., & Halle, M. (1968). The sound pattern of English. New York: Harper and Row.

Church, B. A. (1995). Perceptual specificity of auditory priming: Implicit memory for acoustic information. Unpublished doctoral dissertation, Harvard University, Cambridge.

Church, B. A., & Fisher, C. (1998). Long-term auditory word priming in preschoolers: Implicit memory support for language acquisition. Journal of Memory & Language, 39, 523-542.

Church, B. A., & Schacter, D. L. (1994). Perceptual specificity of auditory priming: Implicit memory for voice intonation and fundamental frequency. Journal of Experimental Psychology: Learning, Memory, & Cognition, 20, 521-533.

Church, B. A., Dell, G., & Kania, E. (1996, November). Representing phonological information in memory: Evidence from auditory priming. Paper presented at the meeting of the Psychonomic Society, Chicago, IL.

Crystal , T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. Journal of the Acoustical Society of America, 83, 1553-1573.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. Journal of Experimental Psychology: Human Perception and Performance, 14, 113-121.

Dahan, D., & Brent, M. (1999). On the discovery of novel word-like units from utterances: An artificial-language study with implications for native-language acquisition. Journal of Experimental Psychology: General, 128, 165-185.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: Evidence for lexical competition. Language & Cognitive Processes, 16, 507-534.

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

Dell, G. S. (2000). Commentary: Counting, connectionism, and lexical representation. In M. B. Broe & J. B. Pierrehumbert (Ed.), Papers in Laboratory Phonology V: Acquisition and the lexicon (pp. 335-348). New York: Cambridge University Press.

Dell, G. S., Reed, K. D., Adams, D. R., & Meyer, A. S. (2000). Speech errors, phonotactic constraints, and implicit learning: A study of the role of experience in language production. Journal of Experimental Psychology: Learning, Memory & Cognition, 26, 1355-1367.

Dupoux, E., Pallier, C., Kakehi, K., & Mehler, J. (2001). New evidence for prelexical phonological processing in word recognition. Language and Cognitive Processes, 16, 491 -- 505.

Echols, C. H., Crowhurst, M. J., & Childers, J. B. (1997). The perception of rhythmic units in speech by infants and adults. Journal of Memory & Language, 36, 202-225.

Farnetani, E. (1997). Coarticulation and connected speech processes. In W. J. Hardcastle & J. Laver (Eds.), The Handbook of Phonetic Sciences, (pp. 371-404). Oxford: Blackwell Publishers.

Fernald, A., Pinto, J. P., Swingley, D., Weinberg, A., & McRoberts, G. W. (1998). Rapid gains in speed of verbal processing by infants in the 2nd year. Psychological Science, 9, 228-231.

Fernald, A., Swingley, D., and Pinto, J.P. (2001). When half a word is enough: infants can recognize spoken words using partial phonetic information. Child Development, 72, 1003-1015.

Fisher, C., & Church, B. A. (2001). Implicit memory support for language acquisition. In J. Weissenborn & B. Hoehle (Eds.), Approaches to bootstrapping: Phonological, syntactic, and neurophysiological aspects of early language acquisition (pp. 47-69). Amsterdam:

Benjamins.

Fisher, C., & Tokura, H. (1996a). Acoustic cues to grammatical structure in infant-directed speech: Cross-linguistic evidence. Child Development, 67, 3192-3218.

Fisher, C., & Tokura, H. (1996b). Prosody in speech to infants: Direct and indirect acoustic cues to syntactic structure. In J. L. Morgan & K. Demuth (Ed.), Signal to syntax: Bootstrapping from speech to grammar in early acquisition (pp. 343-363). Mahwah, NJ: Lawrence Erlbaum Associates.

Fisher, C., Hall, D. G., Rakowitz, S., & Gleitman, L. (1994). When it is better to receive than to give: Syntactic and conceptual constraints on vocabulary growth. Lingua, 92, 333-375.

Fisher, C., Hunt, C., Chambers, K., & Church, B. A. (2001). Abstraction and specificity in preschoolers' representations of novel spoken words. Journal of Memory & Language, 45, 665-687.

Friederici, A. D., & Wessels, J. M. (1993). Phonotactic knowledge of word boundaries and its use in infant speech perception. Perception & Psychophysics, 54, 287-295.

Gerken, L., Murphy, W. D., & Aslin, R. N. (1995). Three- and four-year-olds' perceptual confusions for spoken words. Perception & Psychophysics, 57, 475-486.

Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. Journal of Experimental Psychology: Learning, Memory and Cognition, 22, 1166-1183.

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

Goodsitt, J. V., Morgan, J. L., & Kuhl, P. K. (1993). Perceptual strategies in prelingual speech segmentation. Journal of Child Language, 20, 229-252.

Gordon, P. C. (1989). Context effects in recognizing syllable-final /z/ and /s/ in different phrasal positions. Journal of the Acoustical Society of America, 86, 1698-1707.

Gow, D. W. J. (2001). Assimilation and anticipation in continuous spoken word recognition. Journal of Memory & Language, 45, 133-159.

Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. Journal of Experimental Psychology: Human Perception & Performance, 21, 344-359.

Hallé, P. A., & de Boysson-Bardies, B. (1996). The format of representation of recognized words in infants' early receptive lexicon. Infant Behavior and Development, 19, 463-481.

Harris, J. (1994). English sound structure. Oxford: Blackwell.

Hartshorn, K., Rovee-Collier, C., Gerhardstein, P., Bhatt, R. S., Klein, P. J., Aaron, F., Wondoloski, T. L., & Wurtzel, N. (1998). Developmental changes in the specificity of memory over the first year of life. Developmental Psychobiology, 33, 61-78.

Hintzman, D. L. (1986). "Schema abstraction" in a multiple-trace memory model. Psychological Review, 93, 411-428.

Jusczyk, P. W. (1997). The discovery of spoken language. Cambridge, MA: MIT Press.

Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. Cognitive Psychology, 29, 1-23.

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

Keating, P. (1985). Universal phonetics and the organization of grammars. In V. Fromkin (Ed.), Phonetic linguistics: Essays in honor of Peter Ladefoged (pp. 115-132). New York: Academic Press.

Keating, P. A. (1990). Phonetic representations in a generative grammar. Journal of Phonetics, 18, 321-334.

Kessler, B., & Treiman, R. (1997). Syllable structure and the distribution of phonemes in English syllables. Journal of Memory and Language, 37, 295-311.

Klatt, D. H. (1976). Linguistic uses of segmental durations in English: Acoustic and perceptual evidence. Journal of the Acoustical Society of America, 59, 1208-1221.

Klatt, D. H. (1980). Speech perception: A model of acoustic-phonetic analysis and lexical access. In R. A. Cole (Ed.), Perception and production of fluent speech (pp. 243-288). Hillsdale, NJ: Erlbaum.

Kluender, K. R., Diehl, R. L., & Wright, B. A. (1988). Vowel-length differences before voiced and voiceless consonants: An auditory explanation. Journal of Phonetics, 16, 153-169.

Labov, W. (1972). Sociolinguistic patterns. Philadelphia: University of Pennsylvania Press.

Labov, W. (1998). The intersection of sex and social class in the course of linguistic change. In J. Cheshire & P. Trudgill (Ed.), The sociolinguistics reader, volume 2: Gender and discourse (pp. 7-52). London: Arnold.

Lively, S. E., Pisoni, D. B., & Goldinger, S. D. (1994). Spoken word recognition. In M. A. Gernsbacher (Ed.), Handbook of Psycholinguistics (pp. 265-301). New York: Academic Press.

Logan, G. D. (1992). Shapes of reaction-time distributions and shapes of learning curves: A test of the instance theory of automaticity. Journal of Experimental Psychology: Learning, Memory, & Cognition, 18, 883-914.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word recognition. Cognition, 25, 71-102.

Mattys, S. L., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. Cognition, 78, 91-121.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. Cognitive Psychology, 38, 465-494.

McClelland, J. L., & Elman, J. L. (1986). Interactive processes in speech recognition: The TRACE model. In J. L. McClelland & D. E. Rumelhart (Ed.), Parallel distributed processing: Explorations in the microstructure of cognition (pp. 58-121). Cambridge, MA: MIT Press.

McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. Language and Cognitive Processes, 10, 309-331.

Miller, J. L., & Volaitis, L. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. Perception & Psychophysics, 46, 505-512.

Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. Journal of Memory & Language, 35, 666-688.

Morgan, J. L., & Saffran, J. R. (1995). Emerging integration of sequential and suprasegmental information in preverbal speech segmentation. Child Development, 66, 911-936.

Moscovitch, M. (1994). Memory and working with memory: Evaluation of a component process model and comparisons with other models. In D. L. Schacter & E. Tulving (Ed.), Memory systems 1994 (pp. 269-310). Cambridge, MA: MIT Press.

Mullennix, J. W., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. Perception & Psychophysics, 47, 379-390.

Nittrouer, S., & Boothroyd, A. (1990). Context effects in phoneme and word recognition by young children and older adults. Journal of the Acoustical Society of America, 87, 2705-2715.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. Journal of Experimental Psychology: Learning, Memory, & Cognition, 14, 700-708.

Nusbaum, H. C., & Goodman, J. C. (1994). Learning to hear speech as spoken language. In J. C. Goodman & H. C. Nusbaum (Ed.), The development of speech perception (pp. 299-338). Cambridge, MA: MIT Press.

Nusbaum, H. C., Pisoni, D. B., & Davis, C. K. (1984). Sizing up the Hoosier mental lexicon: Measuring the familiarity of 20,000 words. Research on Speech Perception, Progress report no. 10. Speech Research Laboratory, Psychology Department, Indiana University, Bloomington, Indiana.

Onishi, K., Chambers, K. E., & Fisher, C. (in press). Learning phonotactic constraints from brief auditory experience. Cognition.

Palmer, C., Jungers, M. K., & Jusczyk, P. W. (2001). Episodic memory for musical prosody. Journal of Memory & Language, 45, 526-545.

Pierrehumbert, J. (1990). Phonological and phonetic representation. Journal of Phonetics, 18, 375-394.

Pierrehumbert, J. (in press). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee & P. Hopper (Ed.), Frequency effects and the emergence of linguistic structure (pp. 137-157). Amsterdam: John Benjamins.

Pinker, S. (1984). Language learnability and language development. Cambridge MA: Harvard University Press.

Pisoni, D. B. (1997). Some thoughts on "normalization" in speech perception. In K. Johnson & J. W. Mullennix (Eds.), Talker variability in speech processing (pp. 9-32). New York: Academic Press.

Poldrack, R. A., & Church, B. (1997, November). Auditory priming of new associations. Paper presented at the meeting of the Psychonomic Society, Philadelphia, PA.

Quené, H. (1992). Durational cues for word segmentation in Dutch. Journal of Phonetics, 20, 331-350.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. Psychological Bulletin, 124, 372-422.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-old infants. Science, 274, 1926-1928.

Schacter, D. L. (1994). Priming and multiple memory systems: Perceptual mechanisms of implicit memory. In D. L. Schacter & E. Tulving (Ed.), Memory Systems 1994 (pp. 233-268). Cambridge, MA: MIT Press.

Schacter, D. L., & Church, B. A. (1992). Auditory priming: Implicit and explicit memory for words and voices. Journal of Experimental Psychology: Learning, Memory, and Cognition, 18, 915-930.

Schafer, G., Plunkett, K., & Harris, P. L. (1999). What's in a name? Lexical knowledge drives infants' visual preferences in the absence of referential input. Developmental Science, 2, 187-194.

Sheffert, S. M. (1998). Voice-specificity effects on auditory word priming. Memory & Cognition, 26, 591-598.

Sommers, M. S. (1999). Perceptual specificity and implicit auditory priming in older and younger adults. Journal of Experimental Psychology: Learning, Memory, & Cognition, 25, 1236-1255.

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

Swingley, D., & Aslin, R. N. (2000). Spoken word recognition and lexical representation in very young children. Cognition, 76.

Swingley, D., Pinto, J. P., & Fernald, A. (1999). Continuous processing in word recognition at 24 months. Cognition, 71, 73-108.

Wardaugh, R. (1998). An introduction to sociolinguistics. Oxford: Blackwell.

Warren, P., & Marslen-Wilson, W. (1987). Continuous uptake of acoustic cues in spoken word recognition. Perception & Psychophysics, 41, 262-275.

Werker, J. F., & Stager, C. L. (2000). Developmental changes in infant speech perception and early word learning: Is there a link? In M. B. Broe & J. B. Pierrehumbert (Ed.), Papers in Laboratory Phonology V: Acquisition and the lexicon (pp. 181-193). New York: Cambridge University Press.

Woodward, A. L., & Markman, E. M. (1998). Early word learning. In W. Damon (Series Ed.) & D. Kuhn & R. S. Siegler (Vol. Eds.), Handbook of child psychology: Vol . 2. Cognition, perception, and language (5th ed., pp. 371-420). New York: Wiley.