ELSEVIER

New Ideas in Psychology 25 (2007) 143–170

NEW IDEAS IN PSYCHOLOGY

www.elsevier.com/locate/newideapsych

How are words stored in memory? Beyond phones and phonemes

Robert Port

Department of Linguistics, Department of Cognitive Science, Indiana University, USA

Available online 25 April 2007

Abstract

A series of arguments is presented showing that words are not stored in memory in a way that resembles the abstract, phonological code used by alphabetical orthographies or by linguistic analysis. Words are stored in a very concrete, detailed auditory code that includes nonlinguistic information including speaker's voice properties and other details. Thus, memory for language resembles an exemplar memory and abstract descriptions (using letter-like units and speaker-invariant features) are probably computed on the fly whenever needed. One consequence of this hypothesis is that the study of phonology should be the study of generalizations across the speech of a community and that such a description will employ units (segments, syllable types, prosodic patterns, etc.) that are not necessarily employed as units in speakers' memory for language. That is, the psychological units of language are not useful for description of linguistic generalizations and linguistic generalizations across a community are not useful for storing the language for speaker use.

1. Introduction

How are words and other linguistic patterns stored in memory? The traditional view is that words are stored using segmental units like consonants and vowels. It seems intuitively obvious that speech presents itself to our consciousness in the form of letter-like symbolic units. When we hear someone say a word like *tomato*, we seem to hear it as a sequence of consonant and vowel sound units, which can be represented by the following notation:

tʰə ˈmeɪ ro

E-mail address: port@indiana.edu.

0732-118X/$ - see front matter © 2007 Published by Elsevier Ltd.
doi:10.1016/j.newideapsych.2007.02.001

This transcription in the alphabet of the International Phonetic Association (IPA, 1999) indicates the word has six consonant and vowel segments and is stressed on the syllable beginning with **m**. The notation indicates that two of the segments are complex: the [t] has an aspiration property and the vowel [e] includes an upward glide toward [i]. It is implied that this notation describes a pronunciation that is invariant across speakers, speaking rates, intonation contours and so forth. This paper will argue, however, that this abstract description does *not* resemble the form of words in memory. It will be shown that there is strong evidence from many areas of cognitive science supporting concrete, detailed representations that incorporate speaker properties, tempo, etc. Furthermore, as noted by Coleman (2002), there is virtually no evidence that supports the traditional view of linguistic representation. Segments are the basis of our intuitions about speech, but apparently *only* our intuitions, since our intuitions are strongly biased by the literacy education we have all received. The consequences of this conclusion are major and require reconsidering the goals of phonology and linguistics. Further once spoken words are shown not to be true symbols, our understanding of symbols in general must shift. This understanding also provides insight into what symbols really are and what their role has been in human cognition and the development of civilization.

1.1. Clarifying the traditional assumptions

The traditional assumption about the basic structure of language is that we speak in words that are symbols spelled from a small inventory of contrastive sound units called *phonemes* or *phonological segments* (Bloomfield, 1933; Chomsky & Halle, 1968; Saussure, 1916). These units are abstract symbol tokens that are invariant across context (so [d] is the same in the syllables we spell as [di, de, do], etc.) and the same across speakers. Thus the form of a word like *tomato* in memory is a phonetic transcription that is the same across speakers' voices, intonation contours, speaking rates and so forth. The letters used in the transcription represent sound units that do not overlap in time but are serially ordered just as orthographic letters are. Of course, these graphic letter-symbols can be elaborated with some additional diacritics that represent sound properties that may overlap one or more segments (e.g., nasalization of vowels, breathiness of voice, aspiration, etc.). Most modern linguists express the contrastiveness of the alphabet of tokens by differentiating the segments using distinctive features. This general description of speech—as a series of segments each of which is taken to be a vector of values on a small number of features—is nearly universal in linguistics and has been so for about a century. The way most language scientists conceptualize the form of words in memory is in this form (Chomsky & Halle, 1968; Liberman, Delattre, Gerstman, & Cooper, 1968; McCarthy, 2001; McClelland & Elman, 1986). Even within phonetics (where concern with acoustic and articulatory detail is strongest) phonetic transcriptions are typically postulated in this segmental form as well (Abercrombie, 1967; Jones, 1918a, p. 1; Pike, 1943; Ladefoged, 1972) although phoneticians do not typically assume that the list of segments or features must be small and fixed in size.

The assumption that a segmental description of speech is the only possible description is especially important in linguistics. As a premise, this is related to the assumption that learning a language places great strain on human memory so language must be encoded in some highly efficient way. It was thought that languages employ very restricted segment and feature inventories in order for language to serve as a reliable tool for communication

(Halle, 1985). Reliable reading and writing of some fixed symbol list also explains the claim that words must always be phonetically identical or else must differ from each other in at least one phonetic feature (Bloomfield, 1926; Chomsky & Halle, 1968). Finally, it justifies the assumption that the phonetic features in one language should partially overlap the phonetic features in any other language, since they are drawn from the same restricted universal inventory. The latter property makes direct comparisons across languages possible and supports the search for universal constraints on phonological systems since the common universal features of, for example, [± Voice] on obstruents, [± High] for vowels, place of articulation, etc. are found in many languages. It is typically assumed that these vaguely defined features are identical between languages without seeking evidence from close examination of speakers' behavior (Port & Leary, 2005).

The most important implication of the assumption of a universal phonetic alphabet is that the alphabet is the foundation for all the apparatus of formal linguistics: the notion of many error-free formal operations that require negligible time for their execution (Haugeland, 1985). No formal linguistics is possible without some apriori alphabet of discrete tokens. For the Chomsky and Halle model, the formal operations were ordered rules that create a surface form from underlying forms spelled in the phonetic alphabet. These days in 'optimality theory' (Prince & Smolensky, 1993; McCarthy, 2001) the operations include 'Gen' which generates a very large list of possible discrete transcriptions and 'Eval' to evaluate all these in terms of thousands of universal constraints (each stated in terms of a universal phonetic alphabet) so as to select the correct form. All formal theories must have an apriori inventory of tokens from which all else is constructed.

Another implication of the hypothesis of innate, abstract phonetic features is that this kind of representation is apparently what speakers must use for their lexicon, that is, for the repository of words they know. Since recognition of a word implies making contact with its representation in lexical memory, it seems clear that this must also be the representational code used for remembering specific utterances for a short period of time. That is, if I hear someone say *That's a tomato*, and remember later that I heard someone mention a *tomato*, the traditional theory claims that it is this phonological code—not very different from the transcription of *tomato* above and not very different in this case from its orthographic form—that is used to remember the sentence that I heard. After all, a linguistic representation of a word is the only representation there can be, if formal linguistics has anything to contribute to understanding the psychology of language.

2. Memory research: long term and short term

Does the experimental literature support the intuitions of linguists and others about the existence of such an abstract linguistic representation? There has been a great deal of research on memory for words over the past 50 years looking into the nature of the code used for storage of language. But it is difficult to find any evidence at all supporting a role for an abstract, phonological, segmented form of words, the kind linguists and many other scientists assume. Instead the data strongly suggest that listeners employ a rich and detailed description of words that combines linguistic and nonlinguistic properties of recently heard speech signals (Brooks, 1978; Coleman, 2002; Goldinger, 1996; Pisoni, 1997). This phenomenon, replicated a number of times, seems to provide the most powerful argument against the traditional view. Such a memory has been called *exemplar memory* and has been the target of investigation by a few phoneticians (Johnson, 1997;

Pierrehumbert, 2001) and some speech psychologists (Goldinger, 1996, 1998; Pisoni, 1997). It is worth looking closely at one of the most relevant of these studies. Palmeri, Goldinger and Pisoni (1993) explored a 'continuous word-recognition task,' where subjects listen to a continuous list of recorded words, some tokens of which are repeated, and indicate whether each word is new or a repetition. Words were repeated in the lists after various lags of from (1, 2, 4,..., 64) intervening words.

They also varied the number of talkers used to read the lists in steps from two speakers to 20 speakers, although subjects were told to ignore the variability in the talker voice and were never asked anything about the talkers. As shown in Fig. 1, accuracy was very high for recently presented words (over 90% correct for lags of less than four intervening words) and declined as the number of intervening words increased (falling to about 70% after 64 intervening words). However, there were several results that are completely unexpected by the traditional view of word representation. One is that the subjects did 8% better if the second presentation of a word had the same voice (in fact, the same recording) as the first presentation (as in Fig. 1). This difference in performance shows that listeners were somehow able to use idiosyncratic features of the speaker's voice to help recognize the repeated words. Even more surprisingly, the difference between same voice and different voice was unaffected by the number of voices, as shown in Fig. 2. The improvement for hearing the repetitions in the same (vs. different) voice did not differ from 2 to 20 different voices. This result suggests that the improvement in word memory for hearing the exact same recording is not likely to be due to a strategy to remember the voice and associate it with the word identity, since in this case the job should be harder when there are more voices to correctly associate with specific words than when there are fewer. One would expect greater voice uncertainty should reduce performance. The fact that there was no hint of an effect of the number of voices implies that speakers just automatically remember whatever voice characteristics there may be. It suggests use of something like an auditory (not linguistic) code for speech memory, and suggests they cannot help storing all that information perhaps because their episodic memory retains a detailed auditory description of each utterance heard in the recent past. Other experiments have shown that some

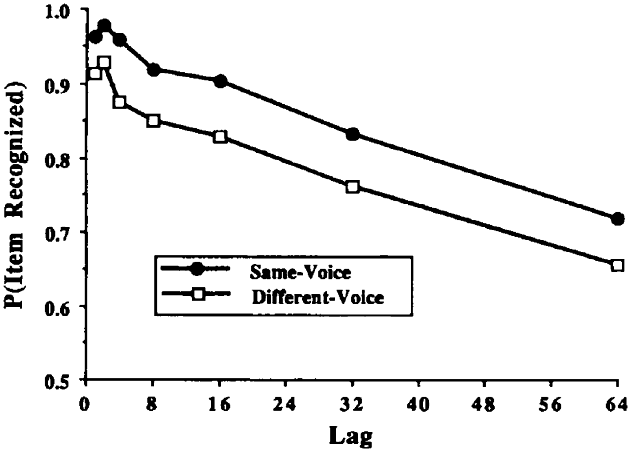

Fig. 1. Probability of correct recognition as a function of the number of intervening words in the continuously presented list. Reprinted with permission from Palmeri et al. (1993).

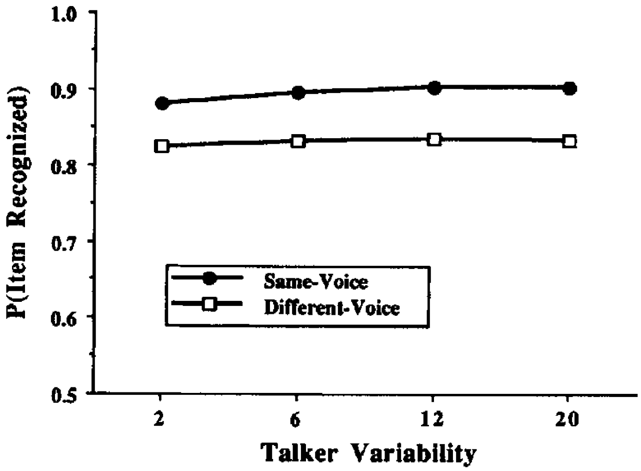

Fig. 2. Probability of correct recognition as a function of the number of voices used in each list. Reprinted with permission from Palmeri et al. (1993).

improvement due to identity of the voice lasts for up to a week (Goldinger, 1996) and that the speaker-identity information and phonetic information are integral (not separable) dimensions (Mullenix & Pisoni, 1990; Garner, 1974). Thus given this result, it should be expected that not only the speaker's voice, but also speaking rate, voice quality, etc. will be retained since the memory description is quite rich and detailed.

These results indicating detailed sensory memory representations are not unique to speech but are completely in line with what has been found repeatedly in visual memory experiments as well. Indeed, most current mathematical models of human memory assume that memory stores much information in the form of specific events rather than being forced to abstract some generalized 'prototype' for stimulus categories (Hintzman, 1986; Nosofsky, 1986; Raaijmakers & Shiffrin, 1992; Shiffrin & Steyvers, 1997). Relatively long-term representations of language seem to be coded in terms of specific episodes incorporating speaker details and nonlinguistic contextual features, rather than in an abstract phonological form. But are abstract properties not stored as well, one might ask? Of course, they probably are, especially for literate speakers, but they need not be. Concrete memories can be used to compute generalizations and abstractions in real time whenever needed (as shown by Hintzman, 1986). To see why, we need only imagine that the exemplars are coded into a large number of features, each with an activation value. When a set of overlapping exemplars are activated by a similar probe item, the other feature values that are shared by many exemplars will also receive more activation than the others. For example, if one is asked what the typical color of a tomato is, one can activate many specific episodes of '*tomato*' and is likely to find that the most active color feature is 'red.' If abstract generalizations can be computed directly from detailed exemplars, then the abstract prototypes appear redundant. The evidence so far suggests that the cognitive processing of language resembles perceptual processing in general, and that these processes are not necessarily dependent on a small number of abstract, general features but also employ a far larger number of concrete sensory features. To be specific, the traditional linguistic representation (whether phonetic or phonological) proposes 20–40 binary features (e.g., 4–6 bits of information) per segment. So at 10–15 segments per second, that implies only 40–60 bits/s for representing spoken language. Instead, we should be thinking of bit rates several orders of magnitude greater than this.

But we have been looking so far at long-term memory effects since the lexicon would seem to be a long-term memory structure. Perhaps we should also look at shorter-term representations of words. Conceivably short-term or working memory is where abstract linguistic symbol tokens may be found to play a cognitive role. Working memory is thought to be used by listeners for analyzing complex utterances. In fact, the term "phonological loop" has been used to describe short-term memory for words (Baddeley, 1986) because it appears to be a language-related store for words that can be rehearsed or cycled by some kind of subvocal speech. But do the data support interpreting Baddeley's term 'phonological' here as linguists understand the term? This would mean a code that is speaker independent, employing segments that are serially ordered (thus rate invariant) and specifiable with a small number of features. There has been a great deal of research over the past half century on short-term memory for words both heard and read (see classic reviews in Baddeley, 1986; Neisser, 1967). The basic task employed in the short-term memory tradition is 'immediate serial recall,' where the subject is presented a sequence of, typically, 7–10 words, either auditorily or visually, and is then asked to immediately recall the list of items in the correct order from first to last. It has long been known that lists with words that are phonologically similar to each other (e.g., *pay, weigh, grey,* or the English letters **b**, **t**, **c**, **e**)[1] are more difficult to recall correctly (more confusable in memory) than words that do not rhyme (e.g., *pay, sack, tie* or **b**, **j**, **o**, **x**) (known as the 'phonological similarity effect'). This is equally true whether the words are presented visually or auditorily (Baddeley, 1986), as would be expected if the same verbal (motor) store is relied on in both cases. But note that words that take longer to pronounce (measured in milliseconds) are recalled less well than words that are shorter (Baddeley, Thomson, & Buchanan, 1975), possibly because rehearsal time is proportional to actual pronunciation time. This suggests that the short-term store retains continuous time properties, but this is definitely not what the linguistic idea of a short-term phonological store would predict.

In a classic experiment, kindergarteners who neutralized the pronunciation of pairs like *wok* and *rock* (into approximately [wɑk]) were first shown to correctly perceive the distinction between [w] and [r] in the speech of others. But they themselves made more short-term memory confusions between words differing in these sounds (using lists like *red, red, wed, red, wed*) than did children who do not mispronounce them (Locke & Kutz, 1975). Their responses were made by pointing to a picture, so responding did not require saying the words aloud. These observations also suggest strongly that verbal working memory, or "phonological short-term store," uses a representation for words and a rehearsal process that is primarily articulatory and is not at all what linguists mean when they use the word 'phonological.'

So the evidence, then, consistently implicates an articulation-based representation for the "phonological loop" (Baddeley, 1990; Wilson, 2001). It appears that subjects try to prevent words from decaying by rehearsing them using some level of motor code. Even at a conscious level, we may notice ourselves rehearsing a telephone number by repeating it silently to ourselves as we move from the phonebook to the telephone. But this process

[1] Letters used to represent orthographic letters will be printed in bold with underline.

does not resemble the activation of a static, abstract spelling for words, as we should have expected on the traditional view.

The psychological data suggest that the nonlinguistic information and linguistic information are richly represented and sufficiently interlocked in their representation that listeners cannot just ignore or strip off nonlinguistic information.[2] For real-time processing of language—that is, for memory and for perceiving and interpreting utterances—it appears that very concrete representations of words are used, employing codes that are close to either sensory patterns (auditory, visual, etc.) or to parameters of motor control of the body. Of course, these recognition memory results imply a memory system capable of storing massive information about the properties of heard speech that is sufficient to record the speaker's voice, the speaker's emotional state and semantic intentions along with categorical information such as the orthographic spelling (if known), the speaker's name, the social context of the utterance, and so on. Presumably such episodic detail cannot be stored forever and patterns must be gradually converted to long-term memory representations. But there is no evidence, apparently, that this information is normally converted to an abstract and speaker-independent form, even though it is converted into that form when we write.

3. Other evidence

If this radical story has any merit, there ought to be more evidence than simply recognition memory and serial recall. As a matter of fact, there is a great deal of other evidence—much of which has been around for decades—but we linguists and psychologists have been unable to see the implications of these familiar results due to our conviction in the psychological reality of letter-based descriptions of speech.

3.1. Richness in dialect variation and language change

There is plenty of evidence that dialect change takes place gradually through small changes in target pronunciations (e.g., Labov, 1963; Labov, Ash, & Boberg, 2006). Idiosyncratic, social and regional dialect pronunciation targets seem to move smoothly through immediately adjacent regions of the vowel space and along other continua like voice-onset time, degree of constriction, place of articulation as well as the rate of motion of various gestures. In order for speakers to modify their pronunciations in tiny steps, speaker–hearers must be able to detect, remember and control such phonetic details. Most phonetic dimensions (e.g., place of articulation, voicing, lip rounding, temporal features, etc.) are continuous and appear to be learned and imitated with no difficulty. How could gradual sound changes, or dialect variability or subtle stylistic variation occur without speakers having detailed forms of storage for speech? Speech patterns are simply not binned into gross phonetic categories in memory as the traditional story based on lexical contrast would have it (Chomsky & Halle, 1968; see Port & Leary, 2005). Of course, this argument could have been raised at any time in the past 100 years, especially by phoneticians, but this argument against discrete phonetics seems not to have been proposed until recently (see Bybee, 2001; Coleman, 2002; Foulkes & Docherty, 2006).

[2] This is interesting because it suggests that simply 'detecting invariant cues' (Gibson, 1966; Port, 1986) while ignoring irrelevant variation in other properties of the signal (e.g., speaker identity, rate, etc.) may not be a strategy that can always be followed by listeners.

3.2. Frequency of occurrence

Frequency is well-known to have a major influence in speech perception. For example, when listening to words in noise, frequent words can be recognized more accurately than infrequent words (Savin, 1963). One way to model this is to postulate a resting activation level for each word pattern so that frequent words have higher resting activation which would make them easier to recognize (since they would already be closer to the recognition threshold) given incompletely analyzed auditory information. It has become increasingly clear as well that the frequency of words and phrases can have a major influence on speech production in most languages (Bybee, 2001; Phillips, 1984). Typically frequent words suffer greater lenition, that is, reduction in articulatory and auditory distinctness, than infrequent words (Bybee, 2001; Lieberman, 1963; Phillips, 1984). An example in my own speech can be seen in the difference between sentences 2a and 2b.

2a. *I'll give you a tomato or I'll give y'a tomato*
2b. *I'll see you tomorrow*

Example 2a illustrates a standard pattern in my speech where the initial **t̲** in *tomato* is pronounced with an aspirated [t] and the second **t̲**, since it is between vowels and the second vowel is unstressed, is flapped (almost invariably in my speech). The first **t̲** is not flapped even though it too falls between vowels and the following vowel is unstressed. The reason is that this **t̲** is word initial, and I will normally not flap a word-initial **t̲**. Example 2b looks the same but differs slightly. Although the unstressed initial syllable of *tomorrow* has the same context as the initial **t̲** of *tomato* and should be aspirated, in a casual speech situation I would most often pronounce the word with an initial flap. The critical difference is that the word *tomorrow* (like *today* and *to*) is more frequent than *tomato* and the sentence as a whole is a frequent expression in my speech. Thus, 2b is a relatively high-probability sentence compared to 2a. It seems that this higher probability pattern tolerates greater lenition than the less frequent pattern.

But the question is how could frequency of occurrence be accounted for under the traditional theory of phonological representation? This kind of observation violates the Neogrammarian idea that the entire vocabulary should be subject to the same set of constraints specifiable only in phonological terms. It is very awkward in the traditional theory for each word to have its own specific pattern (Pierrehumbert, 2002). But to force frequency information into a traditional model, each word would have to have a feature called, say, 'Estimated frequency per million words' and then that numeral would be stored as part of the lexical representation of each word and influence the application of phonological rules. But for a rich exemplar memory system where details of actual utterances are stored for an extended time, frequency could be just a measure of how many tokens of a word or phrase there are in the database at the present time. Then if the word is probed (activated) because, say, a speaker is considering saying the word, then its total activation will be relatively high due to multiple instances in memory. The pronunciation can then be adapted to the estimated activation level in the listener's brain. It seems a rich exemplar memory would make frequency information available automatically.

3.3. Memory for auditory trajectories

Since the earliest spectrographic research on speech, it has been clear that, despite our strong segment-based perceptual experience of speech, consonant and vowel segments do not appear as consistent units of any kind in the acoustic stream of speech (Joos, 1948; Fant, 1960, 1973; Liberman et al., 1968). Chomsky & Miller (1963) pointed out that, in order to serve as the basis of a formal model of language, an adequate theory of phonetics must meet the 'linearity' and 'invariance' conditions on the relation between physical sound and the phonetic segments of linguistic description. That is, each abstract segment must have invariant acoustic cues in the same linear order as the segments themselves. Otherwise, the formal model is completely disconnected from physical reality. The generative school of phonology has always assumed that at some point these invariance conditions would be satisfied by the results of phonetics research. But, of course, these conditions have never been satisfied (Pisoni, 1997; Port and Leary, 2005). Acoustic correlates for segments or segmental features meeting these conditions have never been found for most segment types, even after all these years of trying (Stevens, 1980, 2000). Insisting nevertheless on a segmental description of speech forces an extremely context-sensitive and complex relation to the auditory or acoustic signal. There is, in general, nothing that preserves the linear order of the segments and nothing that is invariant to each segment type. One famous example is the acoustic distinction between [di] and [du].

In the utterance on the left in Fig. 3, at the release of the /d/, the second formant rises slightly toward the target F2 for the vowel /i/. There is a peak in the noise burst for the /d/ at about 2700 Hz. In the second syllable, the F2 falls toward a target value for the /u/ and the burst peak is at around 1700 Hz. Although the bursts are quite different and the formants move in opposite directions, we still hear the two syllables as beginning with the same stop consonant. These phenomena and many other similar effects convinced Liberman (Liberman, Harris, Hoffman, & Griffith, 1957; Liberman et al., 1968) that some special hardware must be employed by humans to transform these widely varying auditory patterns into the segmental description that has the two syllables in Fig. 3 beginning with a letter-like segment that is the same. Stevens and Blumstein (1978) attempted to deal with

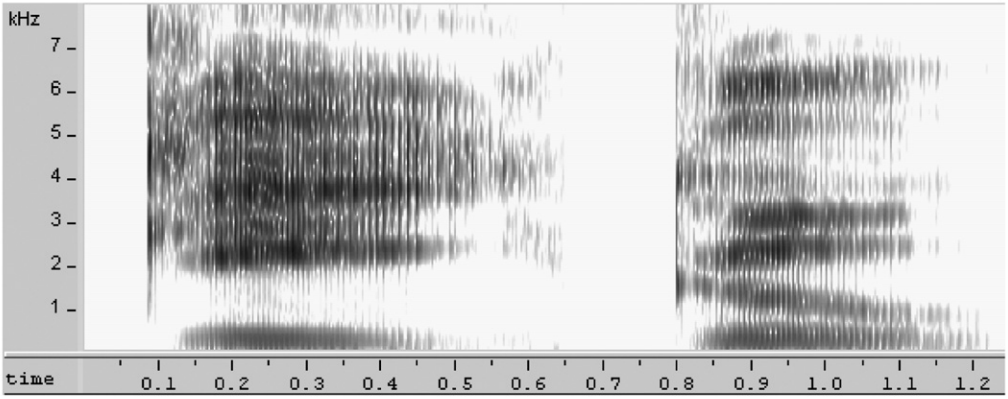

Fig. 3. Sound spectrogram of the syllables /di/ and /du/ spoken by a male speaker of American English (RP). Time in seconds on the x axis and frequency on y. Darkness shows intensity of roughly 300 Hz-wide bands.

this problem by proposing global spectral properties for the first 20–25 ms from onset of the burst (e.g, greater energy at higher frequencies, energy peak between 2200 and 3500 Hz, etc.) that they hoped would be invariant over following vowel and consonant contexts. Unfortunately, they had only very limited success. Research on speech perception over the past half century has led to an increasingly rich and varied set of "speech cues" shown to play a role in at least some contexts to shape listener's perceptions. Although some scientists have hoped that a distinction could be drawn between "major cues" (i.e., "linguistically relevant cues") and others that are minor enough to be ignored or backgrounded (Blumstein & Stevens, 1979; Chomsky & Halle, 1968; Stevens, 2000; Stevens, Keyser, & Kawasaki, 1986), most observers find no basis for treating some acoustic features as "distinctive" (that is, relevant to the linguistic spelling of words) and others as not distinctive. Essentially all of them can be shown to influence perceptual judgments under some conditions implying that such details are stored in memory and extracted for matching during the perceptual process.

Now if humans have a very rich memory capable of storing large amounts of detailed speech material including trajectories over time through a high-dimensional space of formants, stop bursts, fricative spectra shapes, etc., then what dimensions exactly do they use? My proposal is that each language learner, depending on the details of their auditory and linguistic experience, develops their own set of auditory features for speech in their native language. The details of speech cues are likely to differ in detail from speaker to speaker and, of course, they are likely to differ dramatically from language to language (see Hawkins and Smith, 2001; Kuhl & Iverson, 1995; Logan, Lively, & Pisoni, 1991; Strange, 1995; Werker and Tees, 1984).

Once the possibility of rich memory coding is entertained so that the supposed representational efficiency of phonemes no longer dominates our thinking, then many problems in phonological representation disappear. For example, "coarticulation" is said to be the influence of a segment on an adjacent or even nonadjacent segmental cues. Such phenomena challenge the view that segments are independent of each other. But the perceived invariance of a segment like [d] across all its contexts may be what we learned as we became skillfully literate. In realtime speech processing, coarticulation is invisible since learners will continue to remember lots of detailed trajectories in auditory-phonetic space anyway. So coarticulation disappears as a problem. We need not begin by assuming phonemes or phones as the descriptive units. Speakers simply store the auditory patterns they hear and recognize the word, morpheme or phrasal units in whatever sizes they were learned in. Coarticulation is only a problem if one assumes that the "real units" of language must be nonoverlapping and serially ordered. Of course, another problem appears in its place: why do /di/ and /du/ nevertheless seem so vividly to begin with the "same sound?" It is because of literacy training, as will be addressed in the next section.

3.4. 'Phoneme Awareness' comes primarily from literacy training

It has been known since the late 1970s that performance on the type of tasks called 'phonological awareness' tasks correlates very highly with reading skill (Anthony & Francis, 2005; Carroll, 2004; Liberman, Shankweiler, Fischer, & Carter, 1974; Ziegler & Goswami, 2005). Phonological awareness skills that are acquired early include counting the number of syllables in a word, identifying the stressed syllable of a word and recognition of rhyming word pairs. Others include identifying syllable onsets and rimes

(Bradley & Bryant, 1983). Many of the skills just mentioned may be learned before receiving reading instruction (Anthony & Francis, 2005; Liberman et al., 1974). But in recent years it has become clear that those tasks that require identifying segments and adding or removing them are beyond almost all people (whether children or adults) until they have received literacy training (Carroll, 2004; Perfetti, Francis, Bell, & Hughes, 1987; Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001; Ziegler & Goswami, 2005). Studies comparing Portuguese with and without literacy education (Morais, Cary, Alegria, & Bertelson, 1979) and literate Chinese with and without alphabet training showed that adults without alphabet training (Read, Zhang, Nie, & Ding, 1986) cannot do simple phonemic awareness tasks such as adding an initial [d] to [æb] (to yield [dæb]) or removing the final consonant from [bænd]. So apparently the intuition that linguists, many psychologists and most educated moderns have that speech presents itself directly to conscious experience in segmented form may be true of everyone reading this page, but the vividness of our intuitions about the segmental organization of speech is largely a consequence of training in reading and writing with an alphabet (as argued by at least Faber, 1992; Öhman, 2000, and implied by Firth, 1948; Olson, 1994). The clarity of our intuitions here apparently does not reveal anything at all about the form of linguistic memory required for spoken linguistic competence. The fact that the syllables we write as [di] and [du] share the same onset "speech sound" may be obvious us, but it is not obvious at all to someone who has had no alphabet training. It is important to recall that our families and communities invested great resources to educate us to have the skills required for competent use of alphabetical writing. These skills required years of daily practice—many hundreds of hours of training during youth and adulthood. Our intuitions about speech sounds cannot be assumed to be the same as the intuitions of those who have not experienced literacy training.

4. The alphabet as tool and trap

In his courses, Saussure (1916) pronounced forcefully that linguistics should study the spoken language and not the written language. Linguists have consistently emphasized this ever since. Certainly orthographic systems are not central issues in linguistics and neither are texts and manuscripts. But it turns out that, by employing phonetic transcriptions as the standard form of data for research in linguistics, the discipline never escaped from studying languages in written form (see Harris, 2000). It is easy to understand how we got into this situation since speech sounds are inherently very challenging for a perceptual system. Analysis of speech gestures themselves is very difficult for at least the following reasons:

(1) Speech gestures involve complex coordination of largely invisible articulators: e.g., the tongue, larynx, velum, etc. Only the lips and occasionally the tongue tip are visible to ourselves or others.
(2) Speech is articulated very quickly, with 10–15 segment-sized units per second. This is very fast compared to gross movements of the hand, arm or foot. If we try to speak artificially slowly, it sounds unnatural and you still cannot feel where your articulators are.
(3) There is an unlimited amount of variation. Words are pronounced slightly differently in detail in almost every utterance (e.g., Gracco & Abbs, 1989). Most English speakers have a large variety of ways to pronounce the orthographic word *and*—and few of

these versions sound much like the canonical pronunciation [ænd]. Every orthographic system embodies a combination of a linguistic analysis into roughly phoneme-like units, as well as an enormous number of arbitrary conventions that settle awkward and unclear analysis issues. For example, linguists have long puzzled over the issues like the following:

(a) Does *spit* have a **p** or **b**? Either spelling would have some justification and there is no contrast between these stops in this position. What phonological symbol should be used?
(b) Does *itch* have one consonant, as in the spelling [ɪč], or two, as in [ɪtš]?
(c) How many syllables are there in *hire* and *higher, mare* and *mayor* or *error* and *air?* Although they seem to differ in the number of syllables, they are usually identical in my pronunciation. Indeed, how many vowel and consonant segments are there in each of these words? It seems that it is our orthography that biases our intuitive reply to all such questions.
(d) How many words are in *White House?* Notice that this phrase, like *police academy*, has the same word-stress pattern as the single words *greenhouse, blackbird* and *shoehorn* and differs from *the white house, the green house* and *black bird.*
(e) Do *pail* or *fire* contain the glide [j]? In my speech, it sounds like they do but we do not spell them that way.

Given all these difficulties about how words should be spelled, what the founders of linguistics needed was some method for recording speech, stabilizing as much variation as possible and comparing languages with each other. The discipline of linguistics could not make much progress without the conceptual and theoretical tool (just as much as the practical tool) provided by a consistent *alphabet* that was small enough to be easy to teach and precise enough that additional variation could be ignored without obvious misrepresentation. In the early 20th century, the founders of the disciplines of phonetics and linguistics (such as Passy, Saussure, Daniel Jones, Troubetzkoy and others) needed to be able to think about speech using a model they could understand, remember and teach. The representation of language by strings of letters on paper is a powerful technology for the representation of language in a form that can be studied. Indeed, phonetic transcription was the only graphical representation available before the Second World War. A discipline of linguistics could never have gotten off the ground without the prior development of an IPA alphabet.

The alphabets developed over the past three millennia into the major European orthographies (e.g., Latin, French, Old Church Slavic, Russian, English, German, etc.) were, of course, based on the Greek alphabet which was itself the culmination of 3000–4000 years of middle-eastern experimentation with graphical technologies for recording linguistic information (see Hock & Joseph, 1996). The pre-Greek development of writing in the middle east exhibited a trend toward smaller and smaller inventories of tokens along with a corresponding increase in the phonological awareness needed to be skilled in using the alphabet. The developers of the scientific alphabet of the International Phonetic Association in the late 1880s sought an idealized consistent alphabet with one letter for each "speech sound" (IPA, 1999). But what exactly is a speech sound? No satisfactory definition was ever provided (see Twaddell, 1935). Of course, all linguists and phoneticians shared the intuition that a speech sound is whatever we are intuitively motivated to represent with a single letter.

It is now difficult to think about spoken language in any other way than in terms of its serially ordered alphabetical form (see Derwing, 1992). After all, it has only been since the 1950s that acoustic and computational technology (e.g., the sound spectrograph and its software successors) allowed us a good look at speech sounds in continuous time. The sound spectrogram (as shown above in Fig. 3) represents frequency and intensity against time along a continuous spatial axis rather than representing time with the discrete axis of an alphabet—a major improvement (Port, Cummins, & McAuley, 1995). Still, interpreting various kinds of graphic images is something we humans are good at. Understanding complex dynamical events with many degrees of freedom (e.g., an economic system, an ecological system or the speech production process) without visual or spatial scaffolding remains very difficult (Abraham & Shaw, 1983; Clark, 1997).

The IPA phonetic alphabet was a major technological achievement which provided a list of letters with relatively consistent interpretations that could distinguish very approximately the most obvious sounds in the languages most familiar to linguists. But what was not realized at the time is that these letters, just like the conventional orthographic ones, are really only an engineered method for language representation—a culturally transmitted technology whose constraints stem not just from the spoken language but also from human visual perception, ease of drawing, limitations on hand and arm-motion, etc. and the 3000-year history of this technology. The letters chosen from a small list remain there on a sheet of paper indefinitely and are designed to be sufficiently distinct visually that we can preserve a very low error rate in reading, copying and writing them. The number of symbols is small enough and their interpretation sufficiently transparent that it is feasible to train a willing 5–6 year-old child to successfully read and write in a year or two (or maybe three for an inconsistent orthography like English and French, see Ziegler and Goswami, 2005).

4.1. Mental representation of language

Since classical times, western Europeans wondered about human thinking and linguistic competence. What might account for our language skills? The speculation was natural enough that language might be remembered in a way that resembles its storage on paper. This was the idea of Polish linguist Baudouin de Courtenay in the 1870s that led to his postulation of the *phoneme,* a notion that was eventually picked up by Jones (1918b) and others. The phoneme is basically a variant of the concept of a *letter* that is generally hypothesized to be psychological—something in the mind of speaker/hearers. (Actually, Jones preferred to view a phoneme as a set of similar phones and was uncomfortable with a psychological interpretation, see Twaddell, 1935). Whatever the ontological claim, everyone knew what they were like. The phoneme was taken to be invariant over time, invariant between speakers, serially ordered, discrete, nonoverlapping, static and drawn from a small enough set to be reliably produced and perceived. The primary difference between a letter and a phoneme is that a letter is a visible shape on a piece of paper while a phoneme is an invisible token (probably within someone's mind). This idea made sense to many linguists and was quickly adopted by the linguistic community worldwide (see Twaddell, 1935). The phoneme is best understood as a blend of ideas from orthography and a few ideas from psychology (Fauconnier & Turner, 2002). It represents a projection onto the mind of something external that is familiar and easily understood—at least for those with an alphabet-based education.

Of course, in the early 20th century, linguistics had no choice but to rely on phonetic transcription for the description of languages. Transcription using the generalized and consistent cross-linguistic alphabet of the IPA was the best tool available for representation of an arbitrary language. But during the first half of the century new technologies for recording became available (e.g., wire recorders, tape recorders, oscillographs, etc.) which created a problem since consonants and vowels could not be straightforwardly identified in these records. In the late 1940s, Joos (1948) presented the sound spectrograph to the field of linguistics and within a decade or two, speech synthesizers under computer control came online (see Klatt, 1987 for a history). Research with these tools revealed the enormous discrepancy between, on one hand, actual speech gestures and speech acoustics over time and, on the other, the representation offered by a phonemic or phonetic transcription (Ladefoged, 1980). The evidence continued to accumulate that phonetic transcription lacked an enormous amount of information present in continuous-time representations (Hawkins & Smith, 2001; Joos, 1948; Lisker & Abramson, 1971; Sampson, 1977). Furthermore, the transcription process itself was not well understood and unreliable (Eisen & Tillman, 1992; Lieberman, 1965) and the perception of speech has been shown to be rather permanently shaped early in life (Logan et al., 1991; Werker & Tees, 1984). Halle (1954) admitted that speech technology shows that speech is ''not a sequence of clearly separated events, but rather a continuous flow of sound, an unbroken chain of movements'' nevertheless ''investigators of language ... have usually preferred to describe language as a sequence of discrete events.'' Even phonetician Abercrombie (1967, p. 42) acknowledged that ''we describe [segments] as if they were produced by postures of the organs'' even though ''speech is not really a succession of discrete postures.'' However, he declared, ''the only practicable way to describe [speech] is as if it were.''

Unfortunately, between 1920 and the 1960s, linguistics (at least in the United States) became ever more committed to a theoretical view of language predicated on the existence of a closed, universal inventory of segmental sound tokens—a ''universal phonetic alphabet'' presented in the form of a list of segmental distinctive features (Chomsky & Halle, 1968; Jakobson, Fant, & Halle, 1952; see Port & Leary, 2005). If phones and phonemes are discrete and words are spelled from them, then words must be discrete too. In fact, discreteness at all levels of language (phonemes, morphemes, words, sentences, etc.) is assured by discreteness at the phonetic level—just as it is guaranteed for all written alphabetical language by discrete typewriters and recently by the ASCII code for digital computers. Since the continuous-time representations of speech were quite incompatible with the theoretical assumption that a language is a system of formal symbols, linguists have tended to keep the tape recorder and spectrograph out of the phonology classroom. So the field has become trapped. For many linguists, the very notion of a theoretical approach to linguistics demands symbolic input and thus is dependent on the mysterious and poorly investigated process of phonetic transcription as the only gateway to the scientific study of language. In recent years there has been a movement toward ''laboratory phonology'' which has attempted to bridge the gap, but these efforts have yet to yield a coherent new framework that can encompass both phonetics and formal phonology (see the Cambridge University Press series, *Papers in Laboratory Phonology*, now moved to Mouton Press). Traditional generative phonology appears to have no theoretically consistent way to make contact with continuous-time representations of linguistic performance.

But it appears that language is not actually discrete or formal at all. It has some general resemblance to a formal symbol system (as will be shown below), but the evidence is overwhelming that it cannot actually *be* a formal system (Port & Leary, 2005). Only orthographic written language comes fairly close to a discrete symbol system.

5. Human symbol processing

We all understand that humans are symbol processors. Without formal symbol processing skills humans could never have developed orthographies, arithmetic, mathematics and statistics, formal logic, computer programming, our cell-phone controls and computer text editors. Ability to exhibit competence in all of these depends on our ability to do some amount of formal thinking—that is, thinking where we imagine the manipulation of discrete tokens 'in our minds' to achieve logical deduction or to predict future events. But the cultural and technological source of formal thinking has remained confusing and unclear to us in the 20th century (Newell & Simon, 1976; Fodor & Pylyshyn, 1988; see van Gelder & Port, 1995).

Where do formal symbols come from?[3] Chomsky and many others (e.g., Fodor, 1975; Newell & Simon, 1976) assert the Platonic and Cartesian idea that symbol systems are available apriori for humans. They are taken to be cognitive resources ready at hand (and probably at birth) for use during language acquisition, linguistic processing and many other kinds of reasoning. It is often assumed that formal thinking is inherently human—just as language capability is said to be inherently human. But the evidence for this is almost entirely intuitive. Evidence available today suggests that formal thinking is a skill dependent on cultural learning and the historical development of social institutions, like schools, to train our children in these symbolic cognitive skills (Donald, 1991; Olson, 1994; Tomasello, 1999). The amazing power of a small alphabet and a number system that is closed under basic operations is clearly a development of Middle Eastern and European culture. The development of algebra from Al-Khwarizmi to Descartes to Chomsky is an accomplishment of inestimable importance (cf. Lakoff & Núñez, 2000). But an essential component of any formal system is the inventory of tokens given apriori (e.g., the letters and numbers) that are normally graphically specified (except in computers, of course). These original concrete symbols are the archetypes from which, on my proposal, all other abstract symbols of the mind are analogical extensions. Conceptual symbols in various sciences, e.g., 'the gross national product' or 'the species gray squirrel' or 'noun phrase,' have no invariant physical form but we can think of each of these concepts *as if* it were a symbol token, and then do some sort of modeling using that imagined token.

The key here is that the notion of a non-physical yet formal symbol is a cognitive achievement that I speculate would require at least several years of education. My hypothesis is that all mental symbols get their discreteness only analogically, by extension from physical ones like letters and numerals. So the important point here is that it is not the spoken language that provides the model for formal languages, but rather our conventional graphical representations that provide the model for our intuitions about spoken language (Olson, 1994, chapter 4). The creation of modern mathematical systems (such as groups, semigroups and, of course, string grammars) may only have been possible

[3] By a formal symbol I mean a discrete token (like a letter or digit) used for reasoning or cognitive manipulation. The term symbol is used in many other senses, of course.

for a mind that is skilled with orthographic and numerical symbol manipulation. Writing language with a small alphabet is a skill that very likely encourages the belief that language itself is actually symbolic. The symbolic representation of language exhibited by our orthography seems very concrete to our linguistic consciousness. These orthographic skills may suggest to many that a low-dimensional description of spoken language must be possible as well. It is this conviction that inspired Chomsky and most modern linguists to attempt formal linguistic analyses. But true symbol processing requires physical implementation, such as by written tokens on paper, as when we do a long division problem or any formal proof, or check the grammar of a sentence we just composed at the keyboard. Unfortunately, spoken language does not have the properties of written language.

To a limited degree, we can do formal reasoning mentally, as if we were using physically discrete tokens for words, numbers, etc. Most of us can mentally compose a line or two at time of computer code, do simple long division problems, etc., but these skills are derived from practice in reading, writing and doing arithmetic by actually manipulating tokens on paper. For any challenging symbol processing tasks, like doing calculations, writing computer programs with many lines of code or composing an essay or a letter, we depend on a written medium, on actual physical layout of the symbols to which we can visually refer. Letters and numerals are useful to us both in their familiar roles of storage and display, and also in their scientific and technical role as cognitive aids. But these scaffolded cognitive acts ultimately depend on external, physical representations—always graphical until the advent of computing machinery. We alphabet-literates often learn to bootstrap our reasoning by using graphic symbols to scaffold careful thinking (Clark, 2006; Olson, 1994; Ong, 1982). But, despite these uses of symbols and despite the gross similarities between the written language and a true symbol system, the evidence strongly indicates that low-dimensional formal symbols (such as those of modern linguistics) will never provide adequate scientific models of the form of language in human memory. The concept of a formal symbol system is a culturally transmitted technology that was inspired, on my hypothesis, by orthographic writing and arithmetic. One incidental consequence of our incorporation of symbols and symbolic modeling into our conscious cognitive processes happens to be that we experience vivid intuitions about the symbolic nature of language.

6. Implications for linguistics as a discipline

Naturally enough, this drastic rethinking of the nature of speech memory has consequences for linguistics as a whole. But, of course, the repudiation of our linguistic intuitions does not mean that the patterns of phonological and grammatical structure linguists have been studying do not demand both description and explanation.

Each language has a large inventory of meaningful word- and morpheme-like fragments[4] that are constructed almost entirely from smaller pieces that are reused in

[4] The distinction between 'morpheme' and 'word,' however, is surely also primarily derivative from orthographic convention. Some meaningful components of language are more freely commutable than others and some languages (like English) have phonological markers for 'compound words' (which are often ignored in our orthography, cf. *potato peeler* and *woodpecker* which have the same word-stress pattern for me despite the different number of orthographic words). It has proven impossible to justify any cross-linguistic definition of a word or defend any claim of a sharp distinction between word and morpheme in general. The term 'word' in the title and throughout this paper should be understood generically to mean any meaningful linguistic unit, including morphemes and familiar phrases, that is likely to be stored in memory.

other morphemes and words, and which only vaguely resemble similar pieces of other languages. It seems that a randomly chosen word rarely contains unique sounds or sound sequences.[5] Almost every syllable, or at least every syllable part, reappears in some other vocabulary items—as suggested by the table below. An orthographic alphabet is used here for convenience of communication, but within a speaker, each word here should be imagined to be a bundle of similar trajectories through auditory space in memory. The exemplar memory contains clusters of neighbors of various kinds in the speech-auditory space. For example, consider the various categories or neighboring groups of words that the word *slow* could be said to belong to:

Example: *slow* [slo]
1. [slo-] *slope*, *Sloan*, *Slovak*, …
2. [sl-] *sleeve*, *slid*, *sled*, *Slade*, *slack*, *slide*, *slaughter*, *sludge*, …
3. [s(l, m, n)-] *snow, smile, slit,* …
4. {-o] *row*, *low*, *stow*, *doe*, *toe*, *grow*, *sew*, *Shmoe*, *blow*, , …

We can think of each group of words in terms of specific similarities to define a category of words. We might then initially describe the phonology of a language as the set of interlocking and overlapping categories of partly similar words and phrases in the memory of speakers. Since these categories are generalizations across many speakers and because there are still noticeable differences within each group (e.g., the vowel in *slow* is different from *slope* and *Sloan*), the description of the categories cannot be completely precise. These categories could be said to be, in some sense, "employed for spelling words"—but that is misleading since it suggests the categories are manipulable symbol tokens when they are really only sets of utterance fragments with partial resemblance to each other. Of course, the phonological patterns differ dramatically from language to language. For most languages, one can graphically represent these patterns by using an alphabet, that is, by using a small set of vowel and consonant types that are sufficient to differentiate most vocabulary entries that seem to sound obviously different. (Of course, it always depends on who is listening. Those who use the IPA alphabet will apply it differently due to their native language.) The lexicon of many languages consists of items that are distinct from each other in such a way that they exhibit neat cases of a matrix-like structure of, for example, places vs. manners, (e.g., [p t k, b d g, m n η]). Minimally distinct tables of words are revealing. For example, in my own speech, I find the following series of similar front vowels combined with various initial and final consonants:

bead/Bede	*beat/beet*	*bean*	*beam*
bid	*bit*	*bin/been*	- - - -
bed	*bet*	*Ben*	- - –
bad	*bat*	*ban*	*bam*

The dashes mark "cells" that have no entry in my speech. Of course, they are cells only in the sense that, e.g., we can imagine the vowel of *bid* surrounded by the consonant pattern of *beam*. That is, one could say *bim* and most skilled English speakers would recognize it as belonging to categories of lexical entries that include both *beam* and *bid* (as well as to

[5] A rare example of a 'unique' is the final cluster in *kiln* in my version of English since there is currently no other word ending in [-ln] in my English.

ream-rim, bead-bin, etc.). So the dashed cells are ‘available’ in this odd sense as potential new words.

Such tables are easy to construct in any language although it must be noted that such tables can only be created if one is willing to overlook many noticeable variations (and many more that are less noticeable, Hawkins & Nguyen, 2004; Hawkins, 2003). For example, in my midwestern American speech the vowel in *bat* is noticeably lower (i.e., tongue lower and F2 and F1 closer together) than the vowel in the *bad,* and *bad* is usually lower than the vowel in the two nasal-final words where the so-called “short-A” (low front vowel) is very raised and sounds quite “tense” (Labov et al., 2006). And, of course, all the vowels are shorter in the **t**-final words than in the others and are strongly nasalized in *beam* and *bam*. But what advantage is there to ignoring these differences? The main advantage is that one can economize on the number of letters. For skilled readers, we can isolate a set of short-A-like vowels that seem to recur in a huge set of lexical items from *Sam* to *bachelor* to *Democrat*. And since many words fit into tables of similar patterns, it is practical to assign graphic symbols to unit-like patterns like the short-A. Of course, this is just what was discovered by the alphabet founders 3000 years ago. The point is that the reason to use one symbol for all these variant sounds is so that one can employ fewer letters (assuming you have learned the proper way to interpret them).

These patterns can also be studied at a more detailed level. For example, some generalizations can be drawn across a set of word pairs like *bad/bat, limber/limper, sender/center, ruby/rupee, buzz/buss, felled/felt*, etc. The traditional term for this distinction is the ‘voicing’ or ‘tensity’ feature, but whatever label is used, these pairs differ in a large number of correlated properties (Lisker, 1984; Port, 1981). There is a difference in the glottal gestures accompanying them and equally salient differences in the temporal detail of the words. In general, the ‘voiced’ member of each pair has a longer vowel and shorter consonant constrictions (whether the consonant is a stop, fricative, nasal or glide, singleton or cluster) than the corresponding ‘voiceless’ partner (Hawkins & Nguyen, 2004; Lisker, 1984; Port, 1981; Port & Leary, 2005). Although it may be tempting (for reasons of economy) to assert that these distinctions are simply consequences of a discrete feature assigned to a single segment, a more realistic way to describe this is to say that there are many pairs of words in English that exhibit a similar relationship between their syllable codas without assigning this distinction to any single segment and without attempting to abstract a discrete symbol from these pairs.

What is the explanation for all these regular patterns? Although the traditional view claimed that these patterns reflect the discrete alphabet-like and feature-like code used to represent words psychologically in real time, it seems likely that these relatively discrete pronunciation patterns have very little to do with any units of representation in memory but reflect social pressures applying over generations to shape the vocabulary used by a speaker community to approach a lower-dimensional description. The lower-dimensional description is a structure that is shaped over many generations and exists in the individual speaker only implicitly in the form of the clusters and categories of similar speech trajectories that we have been reviewing here.

There are probably many reasons for these long-term pressures toward a low-dimensional description. One important one is a fundamental property of language that has been noted by many observers over the years (e.g., Abler, 1989; Chomsky & Halle, 1964; Dietrich & Markman, 2003; Goldstein & Fowler, 2003; Hockett, 1968; Holland, 1995; Studdert-Kennedy, 2003; Von Humboldt, 1836/1972). If a language could employ a limited set of building blocks and combine them to make new structures in a way that does not completely merge or destroy the identity of the components themselves, an enormous variety of novel potential patterns becomes ‘‘available’’ in the sense that they become somewhat implicit in the data. The principle is to construct an expanding set of structures by ‘‘reusing’’ the relatively discrete patterns. The units we actually use are not formally discrete objects, that is, not really building blocks, but it is true that these patterns tend not to simply dissolve or merge when combined, but retain sufficient independent identity to permit their identification with categories of partly similar words. This approximate independence allows the reuse of pattern fragments in many other contexts (and, of course, also makes it convenient to employ the same letter or pattern of letters to represent them graphically). The principle seems rather similar to that found in the combination of fragments of genomic material in genetic reproduction (although genes may not be as much like an alphabet as was formerly thought either, Stotz & Bostanci, 2006). Apparently, human languages employ several levels of nested patterns such as a large set of meaningless sound categories, and meaningful words and morphemes that are combinable into phrases and utterances (Hockett & Altman, 1968). Speakers have reason to seek greater contrast and distinctiveness in some situations and, at other times, to seek greater ease of articulation. These factors led over time to a tendency toward maximal differences between categories (along psychophysical dimensions) and often the historical collapse of patterns that are ‘‘similar enough’’ into merged equivalence classes that we have traditionally called ‘‘allophones’’.

Several other explanatory principles have been proposed for why phonological patterns like these evolve and endure over generations of speakers and why different languages sometimes arrive independently at similar patterns. In addition to (a) supporting linguistic creativity through recombination of units, the proposed reasons include (b) reducing the number of degrees of freedom for articulatory control (e.g., Browman & Goldstein, 1993; Lindblom, MacNeilage, & Studdert-Kennedy, 1984; Martinet, 1960; Studdert-Kennedy, 2003) and (c) improving the reliability and speed of speech perception (e.g., Jakobson et al., 1952; Goldstein & Fowler, 2003). A lexicon employing a fairly small number of these near-symbolic units may be a historical attractor state by partially satisfying factors (a)–(c). As categories in the social lexicon, the items implicitly suggest a space parameterized by combinations of these units—even if the ‘‘units’’ themselves have fuzzy edges, imprecise or context-sensitive definitions and vary somewhat from context to context.

The important point here is that we should think of phonological patterns of various sizes as reflecting the social nature of language. Phonological structure is an emergent adaptive system of patterns that appears within the speech of a community and is always in flux. The phonology contains categories whose characteristics are defended by the slow (and weak) maintenance processes of the social institution as a whole. Thus, ‘‘violations’’ of these structures can be expected to occur frequently. Speakers can, in principle, control any aspect of their productions that they want and can sometimes imitate minute idiosyncrasies in the pronunciations of others. Thus language variation and change can result (Bybee, 2001; Labov, 1963) along with incomplete neutralization (Port & Crawford, 1989; Warner, Jongman, Sereno, & Kemps, 2004) and many other phenomena. The factors listed above gradually bias the distributions of the real-time phonology parameters produced by the community.

One key idea is here that there are at least two time scales over which to look at language, and two distinct sets of phenomena to explain. Beginning at the longer time scale, the structure of phonology serves the community of speakers and is manifested as behavior tendencies apparent in the statistics of the speech of community members. Phonological structure as a social institution makes available to the child language learner a large set of patterns of words, etc., in use plus many others that could be words or utterances but are not (or not yet). For these purposes, all the variant sounds that seem to be, for example, short-A variants can be treated as the same (and could be spelled the same) regardless of vowel quality detail, nasalization, duration differences, etc. Of course, since the set of components is not well specified, exactly what the set of potential words is must remain completely fuzzy. Potential words and phrases need to be similar enough to previous vocabulary items to be fairly easy for a listener to recognize, interpret or learn, and yet different enough that they will tend not to be confused with existing items. There is no clear boundary between 'possible word' (or phrase or sentence) and 'impossible word' (or phrase or sentence) as assumed by traditional linguistic theory. And there is no possibility of formally specifiable constraints. There are just some patterns that are common, some that are rare, and a huge number of combinations that will seem possible to speakers but have never occurred. Utterances are recognizable as a function of their degree of similarity to the huge set of previously heard utterances. Speaker sensitivity to the degree of similarity to familiar items has been demonstrated several times (Frisch, Large, & Pisoni, 2000; Pierrehumbert, 1994).

Speakers tolerate great amounts of variation. But over the long run, the lexicons of languages tend toward attractor states where it seems *almost as if* they employed a small number of sound categories that keep a discrete distance from each other for spelling the lexicon. Certainly, when there are longstanding orthographic conventions, this impression can be very compelling to literate speaker–hearers. The proposal here is that on a time scale of generations and looking over a community of speakers, a simple, componential phonological inventory is a state that the speech of a community only approaches. The unfortunate mistake in late 20th century linguistics was to assume, inspired by the practical success of orthographies and mathematical systems with small alphabets, that language actually *is* such an ideal symbol system and that speakers normally process language in such formal terms. There is no evidence suggesting that language is a formal system or that it can be succinctly described using a formal system.

Turning to a very short time scale, actual productions of words in real time respond to many contextual factors and occasionally result in deviations from the category generalizations that may be observed in the speech of the community. This is one reason why it is so difficult to use unedited audio recordings of real speech for traditional phonological analysis. The phonological objects are very hard to identify in actual recordings, so an alphabet-trained human is required to do the identifying. In fact, on the short time scale during syllable production, it seems reasonable to argue that phonology, as a pattern of social behavior, cannot apply. Since a single utterance has no regularities and no distributions, it can have no phonological structure. By analogy, a single atom has mass and energy but it is meaningless to ask about its pressure. Pressure is a property only of large aggregates of atoms, such as a volume of gas.

From this perspective then, a description of the phonology of a language should be an attempted snapshot, across some time window over some group of speakers, of the distributional pattern of utterances. These utterances can be accurately described only as trajectories through a sufficiently rich phonetic space over time—although letter-based transcriptions, for their convenience and intuitiveness, will still be useful as well. The real data of linguistics, however, are only speech gestures and phonetic trajectories in time. Research on phonology should study the distributions and the ways the distributions are pushed around by the talking habits of speakers of the community. If we insist on locating the precise form of "linguistic knowledge" in some single speaker, then probably the best that could be done is to point to a large set of centroids in clouds of data points in a high-dimensional space of utterances (see Pierrehumbert, 2001, 2002). When we look very closely at a single speaker, the knowledge of a language will have only remote similarities to a symbol list as represented by the alphabetic form of written language or a phonetic transcription. The individual speaker's knowledge is better thought of as a very large inventory of utterances in a high-dimensional space and a range of control parameters for speech production. At any moment in time in any specific speaker, the units of social phonology may be quite invisible and do not matter—unless the speaker has learned how to employ graphical transcription. For those who have advanced literacy skills, there is probably also an additional description, a low-dimensional, stable linguistic understanding of the utterance database resembling the linguistic analyses of modern linguists.

7. Discussion

7.1. Rich memory

At this point it may help to speculate as to what the proposed "rich memory" in various modalities is like. Although there are many open issues about human memory (Raaijmakers & Shiffrin, 1992; Whittlesea, 1987), it seems to be generally accepted that it is a system that attempts to store as much detail as possible about events in life, using auditory, visual and somatosensory information. The features used for this coding are whatever sensory and temporal patterns the individual has learned for describing and differentiating events in the environment, but these features generally differ in detail from person to person. This kind of storage is massively redundant since many similar events will occur multiple times. This memory includes so-called 'episodic memory,' linking co-occurring information from many modalities (Brooks, 1978; Gluck, Meeter, & Myers, 2003; Goldinger, 1998) and is rich enough in detail that it is natural to think of it as an 'exemplar memory,' that is, a memory that stores concrete examples (Hintzman, 1986; Nosofsky, 1986; Shiffrin & Steyvers, 1997; Smith & Medin, 1981). My proposal is that language is stored using some version of such a system. The dimensionality of this memory will vary from speaker to speaker but it is surely far richer than linguists have ever considered in the past. Of course, these memories may include many prototypes and abstractions as well and, for people with the appropriate education, the memory for language will include an orthographic and perhaps an approximate phonological or phonemic description as well.

This memory makes possible the perception of the identity of phonological fragments based on some similarity measure. Stored information includes the categories that each utterance fragment (e. g., word, morpheme, etc.) might belong to. This memory, because of its redundancy, can differentiate fragments based on their frequency of occurrence. Turning to the production problem, the speaker also uses frequency implicit in the memory to determine details of how to pronounce a fragment in any particular situation. Somehow apparently, the database of tokens of individual speech fragments (such as words) is able to influence a speaker's choice of pronunciation decisions, since speakers (especially younger ones) modify their pronunciations to be more similar to what they hear others say (Goldinger, 1998; Labov, 1963; Pierrehumbert, 2001).

Finally, this memory is also what a subject in a recognition memory experiment relies on to detect the repetition of a word. Evidence from recognition memory for speech was reviewed which suggests that the form of language in memory cannot resemble any traditional linguistic descriptions, whether they be more 'phonetic' (that is, more detailed but still segmented and invariant across speakers) or more 'phonological' (that is, focused on information that is relevant to lexical distinctness, rather like an idealized orthographic representation) (Hawkins & Smith, 2001). The traditional representation postulates some supposedly "minimal" and "maximally efficient" coding. But this coding seems fairly well suited only to a native speaker who wants to read words using a minimal graphical representation, but not very suited to a listener who wants not just to differentiate words, but also to extract information about the speaker, the speaker's state-of-mind and many aspects of the context.

Another reason to believe in a richer memory for speech rather than an abstract memory is experiments on speech perception showing that finding invariant patterns for consonant- and vowel-sized units has proven maddeningly difficult (e.g.; Cole et al., 1997). The most effective speech recognition systems make little use of segment-sized units (Huckvale, 1997; Jelinek, 1988). This suggests that our intuitions about segments must have some other source than the acoustic signal or articulatory gestures themselves. But the vivid conventional model of the relation between letter-segments and speech sounds and gestures is something we all share. The transcription of *tomato* in Example 1 seems right to us all. However, the model that is intuitive to most of us—writing words using letter strings—is something we were taught in school; not something we learned as we became competent speakers and hearers

A glance at western education practices shows us where our strong segmental intuitions come from—from literacy training. In order to write language with a very small token inventory, children in alphabet cultures are trained to interpret speech in terms of a small set of graphic tokens. One incidental result of this training is that literate people tend to automatically interpret speech in terms of letter-sized units. A much more important consequence is that over the past three millennia our culture has developed ways to exploit the power of symbol systems, using them, not just for alphabetic writing standards, but also in advanced symbol-based technologies like arithmetic, logic, mathematics and computer programming.

7.2. Linguistics

Such a drastic revision of the nature of phonetic representations necessarily forces a reconsideration of what linguistics can be. Without realizing what it was doing, linguistics made a gamble that there will be a low-dimensional description of each language that can be expressed using an alphabet. Linguists tended to take for granted that there is some basic alphabet capable of supporting their formal models. They assumed, in effect, that there exists some low-bitrate code for speech representation. But the data reviewed in this essay seem instead to support a much richer memory for speech that employs a large amount of information about instances of speech and the probabilities of occurrence of its fragments.

It must be acknowledged that the approach to memory endorsed here would seem to be compatible with a language in which every word is simply different from every other word but with little or no reuse of components. After all, if memory is very detailed and rich, why would languages need to have phonologies? Why do they all appear to build words from a large, but still limited, number of component fragments (features, segments, onsets, codas, etc.) as was shown in Section 6. These facts require additional explanatory principles. But the explanation cannot be that such components are the manipulanda of some formal system for speech production and perception. It seems likely that phonological structures are instead the result of gradual shaping of the vocabulary by the speaker community. This realm of patterns is what the child language learner is exposed to and uses as the empirical basis for acquiring linguistic skills. The phonology should be viewed as a social institution that is polished and streamlined by its speakers over the generations so that it approaches a (possibly) more efficient componential structure that resembles a system of formal components.

If this turns out to be correct, then linguistics will need to abandon its goal of describing the form that language exhibits as it is used in realtime. Linguistics should observe and record the patterns in the speech of a community. The resulting descriptions will be of use to teachers of the language, dictionary writers, orthography designers and others. But they are not likely to be of much explanatory value to those seeking to understand realtime language processing.

8. Concluding points

1. Spoken language differs from written language in more fundamental ways than we thought. It is not a real symbol system although it does approach one to a very limited degree.
2. Instead, the child learns about the distributions of lexical patterns in a high-dimensional space of common sound patterns, that is, the phonology of the child's linguistic community. As children learn to speak, they store phrases and "words" as rich and complex high-dimensional patterns, learning eventually to categorize them into lexical and phonological categories (and may be supported in this by orthographic spellings). A phonetic transcription does not begin to capture all this essential richness. At a microlevel, each speaker must discover their own detailed auditory-phonetic code for identifying and storing linguistic chunks of their language. Different speakers of the same language will only have statistical similarities to each other. But speakers of different languages may have dramatically different and incommensurable codes for describing and storing words.
3. The discipline of linguistics, and phonology in particular, should really be concerned with regularities in the speech of a community of speakers—with the patterns that comprise the linguistic environment of the language learner. Such environments include the ones known as standard literate English, Spanish, French, etc. as well as the oral language environments of various subcommunities of speakers. These patterns should be studied and taught using whatever tools are useful (including, to be sure, both orthographic and phonetic alphabets). But it would be misleading to claim that an alphabetical description directly captures anything like "the memory code" for the language that is used by individual speakers. Their memory code is something very different and much richer. These generalizations and their degree of prevalence in the community are useful information for a second-language learner of any language.

4. Finally, since language is not truly symbolic, it becomes clear that the notion of a symbol and all that western culture has developed using symbol patterns is strongly dependent on the physical, that is, graphical, character of letters and numbers (see Clark, 2006). Because they have a consistent physical form, we can reason confidently with them. This was more difficult to see before when we did not differentiate between real symbols whose formal properties are supported by physical properties, and those symbol-like units, such as words and phonemes, that do not have physical tokens to define them concretely. It used to seem that both speech sounds and written words, as well as our thoughts about numbers and abstract quantities, employed real symbols. Abstract numbers seem to us to be just as real as concrete number tokens on a page. Actually, spoken words only approximate symbols, and in our formal thinking they "stand for" symbols. It is only written words or tokens with discrete graphical correlates that are guaranteed to support real formal operations.

Acknowledgments

This paper contains ideas inspired by many colleagues. Thanks especially to David Pisoni, Ken de Jong, Sarah Hawkins, Janet Pierrehumbert, Keith Johnson, John Coleman, Terrence Nearey, Diane Kewley-Port and Andy Clark. I am also grateful to Mark VanDam and Sarah Hawkins for helpful comments on various drafts.

References

Abercrombie, D. (1967). *Elements of general phonetics*. Chicago, Illinois: Aldine-Atherton.

Abler, W. (1989). On the particulate principle of self-diversifying systems. *Journal of Social and Biological Structures*, *12*, 1–13.

Abraham, R., & Shaw, C. (1983). *Dynamics: The geometry of behavior, part 1*. Santa Cruz, California: Aerial Press.

Anthony, J., & Francis, D. (2005). Development of phonological awareness. *Current Directions in Psychological Science*, *14*, 255–259.

Baddeley, A. D. (1986). *Working memory*. Oxford, UK: Oxford University Press.

Baddeley, A. D. (1990). The development of the concept of working memory: Implications and contributions of neuropsychology. In G. Vallar, & T. Shallice (Eds.), *Neuropsychological impairment of short-term memory* (pp. 54–73). New York: Cambridge University Press.

Baddeley, A. D., Thomson, N., & Buchanan, M. (1975). Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, *14*, 575–589.

Bloomfield, L. (1926). A set of postulates for a science of language. *Language*, *2*, 153–164.

Bloomfield, L. (1933). *Language*. New York, NY: Holt Reinhart Winston.

Blumstein, S., & Stevens, K. (1979). Acoustic invariance in speech perception: Evidence from measurements of the spectral characteristics of stop consonants. *Journal of Acoustical Society*, *66*, 1001–1017.

Bradley, L., & Bryant, P. E. (1983). Categorizing sounds and learning to read—a causal connection. *Nature*, *271*, 419–421.

Brooks, L. R. (1978). Nonanalytic concept formation and memory for instances. In E. Rosch, & B. B. Lloyd (Eds.), *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Bybee, J. (2001). *Phonology and language use*. Cambridge, UK: Cambridge University Press.

Carroll, J. M. (2004). Letter knowledge precipitates phoneme segmentation but not phoneme invariance. *Journal of Research in Reading*, *27*, 212–225.

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

Chomsky, N., & Miller, G. (1963). Introduction to the formal analysis of natural languages. In R. Luce, R. Bush, & E. Galanter (Eds.), *Handbook of mathematical psychology*, Vol. 2 (pp. 323–418). New York: Wiley.

Clark, A. (1997). *Being there: Putting brain, body, and world together again*. Cambridge, MA: Bradford Books/MIT Press.

Clark, A. (2006). Language embodiment and the cognitive niche. *Trends in Cognitive Science*, *10*, 370–374.

Cole, R., Mariani, J., Uszkoreit, H., Varile, G. B., Zaenen, A., Zampolli, A., et al. (1997). *Survey of the state of the art in human language technology*. Pittsburgh, Pennsylvania: Center for Spoken Language Understanding, Carnegie Mellon University.

Coleman, J. (2002). Phonetic representations in the mental lexicon. In J. Durand, & B. Laks (Eds.), *Phonetics, phonology and cognition* (pp. 96–130). Oxford: Oxford University Press.

Derwing, B. (1992). Orthographic aspects of phonological competence. In P. Downing, S. Lima, & M. Noonan (Eds.), *The linguistics of literacy* (pp. 193–210). Amsterdam: John Benjamins.

Dietrich, E., & Markman, A. (2003). Discrete thoughts: Why cognition must use discrete representations. *Mind and Language*, *18*, 95–119.

Donald, M. (1991). *Origin of the modern mind: Three stages in the evolution of culture and cognition*. Cambridge, MA: Harvard University Press.

Eisen, B., & Tillman, H. G. (1992). Consistency of judgments in manual labeling of phonetic segments: The distinction between clear and unclear cases. In *Paper presented at the international conference on spoken language processing*, 1992.

Faber, A. (1992). Phonemic segmentation as epiphenomenon: Evidence from the history of alphabetic writing. In P. Downing, S. Lima, & M. Noonan (Eds.), *The linguistics of literacy* (pp. 111–134). Amsterdam: John Benjamins.

Fant, G. (1960). *The acoustical theory of speech production*. The Hague: Mouton.

Fant, G. (1973). *Speech sounds and features*. Cambridge, MA: MIT Press.

Fauconnier, G., & Turner, M. (2002). *The way we think: Conceptual blending and the mind's hidden complexities*. New York, NY: Basic Books.

Firth, J. R. (1948). Sounds and prosodies. *Transactions of the Philological Society*, 127–152.

Fodor, J. (1975). *The language of thought*. Cambridge, MA: Harvard University Press.

Fodor, J., & Pylyshyn, Z. (1988). Connectionism and cognitive architecture. *Cognition*, *28*, 3–71.

Foulkes, P., & Docherty, G. (2006). The social life of phonetics and phonology. *Journal of Phonetics*, *34*, 409–438.

Frisch, S., Large, N., & Pisoni, D. (2000). Perception of wordlikeness: Effects of segment probability and length of the processing of nonwords. *Journal of Memory and Language*, *42*, 481–496.

Garner, W. R. (1974). *The processing of information and structure*. Potomac, Maryland: Erlbaum.

Gibson, J. (1966). *The senses considered as perceptual systems*. Boston: Houghton-Mifflin.

Gluck, M., Meeter, M., & Myers, C. (2003). Computational models of the hippocampal region: Linking incremental learning and episodic memory. *Trends in Cognitive Science*, *7*, 269–276.

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.

Goldstein, L., & Fowler, C. (2003). Articulatory phonology: A phonology for public language use. In N. O. Schiller, & A. S. Meyer (Eds.), *Phonetics and phonology in language comprehension and production* (pp. 159–207). Berlin: Mouton de Gruyter.

Gracco, V., & Abbs, J. (1989). Sensorimotor characteristics of speech motor sequences. *Experimental Brain Research*, *75*, 586–598.

Halle, M. (1954). The strategy of phonemics. *Word*, *10*, 197–209.

Halle, M. (1985). Speculations about the representation of words in memory. In V. Fromkin (Ed.), *Phonetic linguistics: Essays in honor of Peter Ladefoged* (pp. 101–114). Orlando, Florida: Academic Press.

Harris, R. (2000). *Rethinking writing*. London: Continuum.

Haugeland, J. (1985). *Artificial intelligence, the very idea*. Cambridge, MA: Bradford Books-MIT Press.

Hawkins, S. (2003). Roles and representations of systematic fine phonetic detail in speech understanding. *Journal of Phonetics*, *31*, 373–405.

Hawkins, S., & Nguyen, N. (2004). Influence of syllable-final voicing on the acoustic properties of syllable-initial /l/ in English. *Journal of Phonetics*, *31*, 199–231.

Hawkins, S., & Smith, R. (2001). Polysp: A polysystemic, phonetically rich approach to speech understanding. *Italian Journal of Linguistics-Rivista de Linguistica*, *13*, 99–188.

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

Hock, H. H., & Joseph, B. (1996). *Language history, language change and language relationship* (2nd ed.). Berlin: Mouton de Gruyter.

Hockett, C. (1968). *The state of the art*. The Hague: Mouton.

Hockett, C., & Altmann, S. (1968). A note on design features. In T. A. Sebeok (Ed.), *Animal communication: Techniques of study and results of research* (pp. 61–72). Bloomington, IN: Indiana University Press.

Holland, J. (1995). *Hidden order: How adaptation builds complexity*. Cambridge, MA: Perseus Books.

Huckvale, M. (1997). Ten things engineers have discovered about speech recognition. In *Paper presented at NATO ASI workshop on speech pattern processing*, Jersey, UK.

IPA. (1999). *Handbook of the international phonetic association: A guide to the use of the international phonetic alphabet*. Cambridge, England: Cambridge University Press.

Jakobson, R., Fant, G., & Halle, M. (1952). *Preliminaries to speech analysis: The distinctive features*. Cambridge, MA: MIT.

Jelinek, F. (1988). Applying information theoretic methods: Evaluation of grammar quality. In *Paper presented at the workshop on evaluation of natural language processing systems*. Wayne, Pa.

Johnson, K. (1997). Speech perception without speaker normalization: An exemplar model. In K. Johnson, & J. Mullenix (Eds.), *Talker variability in speech processing* (pp. 145–166). London: Academic Press.

Jones, D. (1918a). *An outline of English phonetics*. Leipzig, Germany: Teubner.

Jones, D. (1918b). The phonetic structure of Sechuana. *Transactions of the Philological Society*, *1917–1922*, 99–106.

Joos, M., (1948). Acoustic phonetics. *Language monograph*, vol. *23*, Linguistic Society of America.

Klatt, D. (1987). Review of text-to-speech conversion for English. *Journal of Acoustical Society*, *82*, 737–793.

Kuhl, P., & Iverson, P. (1995). Linguistic experience and the perceptual magnet effect. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121–154). Timonium, Maryland: York Press.

Labov, W. (1963). The social motivation of a sound change. *Word*, *19*, 273–309.

Labov, W., Ash, S., & Boberg, C. (2006). *The atlas of North American English*. Berlin: Mouton de Gruyter.

Ladefoged, P. (1972). *A course in phonetics*. Orlando, Florida: Harcourt Brace Jovanovich.

Ladefoged, P. (1980). Phonetics and phonology in the last 50 years. *UCLA Working Papers in Phonetics*, *103*, 1–11.

Lakoff, G., & Núñez, R. (2000). *Where mathematics comes from: How the embodied mind brings mathematics into being*. New York, NY: Basic Books.

Liberman, A. M., Delattre, P., Gerstman, L., & Cooper, F. (1968). Perception of the speech code. *Psychological Review*, *74*, 431–461.

Liberman, A. M., Harris, K. S., Hoffman, H., & Griffith, B. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Experimental Psychology*, *54*, 358–368.

Liberman, I. Y., Shankweiler, D., Fischer, F. W., & Carter, B. (1974). Explicit syllable and phoneme segmentation of the young child. *Journal of Experimental Child Psychology*, *18*, 201–212.

Lieberman, P. (1965). On the acoustic basis of the perception of intonation by linguists. *Word*, *21*, 40–54.

Lindblom, B., MacNeilage, P., & Studdert-Kennedy, M. (1984). Self-organizing processes and the explanation of phonological universals. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Explanations of language universals* (pp. 181–203). Berlin: Mouton.

Lisker, L. (1984). Voicing in english: A catalogue of acoustic features signalling /b/ vs. /p/ in trochees. *Language and Speech*, *29*, 3–11.

Lisker, L., & Abramson, A. (1971). Distinctive features and laryngeal control. *Language*, *47*, 767–785.

Locke, J. L., & Kutz, K. J. (1975). Memory for speech and speech for memory. *Journal of Speech and Hearing Research*, *18*, 176–191.

Logan, J., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *Journal of the Acoustical Society of America*, *89*, 874–886.

Martinet, A. (1960). *Elements of general linguistics*. Paris: Armand Colin.

McCarthy, J. (2001). *A thematic guide to optimality theory*. Cambridge, England: Cambridge University Press.

McClelland, J., & Elman, J. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition*, *7*, 323–331.

Mullenix, J., & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception and Psychophysics*, *47*, 379–390.

Neisser, U. (1967). *Cognitive psychology*. Appleton-Crofts.

Newell, A., & Simon, H. (1976). Computer science as empirical inquiry: Symbols and search. *Communications of the Association for Computing Machinery*, *19*, 113–126.

Nosofsky, R. (1986). Attention, similarity and the identification–categorization relationship. *Journal of Experimental Psychology: General*, *115*, 39–57.

Öhman, S. E. G. (2000). Expression and content in linguistic theory. In M. Gustafsson, & L. Hertzberg (Eds.), *The practice of language* (pp. 99–107). Dordrecht, Holland: Kluwer Academic.

Olson, R. (1994). *The world on paper: The conceptual and cognitive implications of writing and reading*. Cambridge: Cambridge University Press.

Ong, F. J. (1982). *Orality and literacy: The technologizing of the word* (1st ed.). London: Routledge.

Palmeri, T. J., Goldinger, S. D., & Pisoni, D. B. (1993). Episodic encoding of voice attributes and recognition memory for spoken words. *Journal of Experimental Psychology, Learning Memory, and Cognition*, *19*, 309–328.

Perfetti, C., Francis, D., Bell, L. C., & Hughes, C. (1987). Phonemic knowledge and learning to read are reciprocal. *Merrill-Palmer Quarterly*, *33*, 283–319.

Phillips, B. (1984). Word frequency and the actuation of sound change. *Language*, *60*, 320–342.

Pierrehumbert, J. (1994). Syllable structure and word structure. In P. Keating (Ed.), *Papers in laboratory phonology III* (pp. 168–188). Cambridge, UK: Cambridge University Press.

Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.

Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology 7* (pp. 101–140). Berlin: Mouton de Gruyter.

Pike, K. L. (1943). *Phonetics: A critical analysis of phonetic theory and a technique for the practical description of sounds*. Ann Arbor: University of Michigan Press.

Pisoni, D. B. (1997). Some thoughts on normalization in speech perception. In K. Johnson, & J. Mullennix (Eds.), *Talker variability in speech processing* (pp. 9–32). San Diego: Academic Press.

Port, R. (1981). Linguistic timing factors in combination. *Journal of the Acoustical Society of America*, *69*, 262–274.

Port, R. (1986). Invariance in phonetics. In D. Klatt, & Perkell (Eds.), *Invariance and variability in the speech processes*. London: Lawrence Erlbaum Associates.

Port, R., & Crawford, P. (1989). Incomplete neutralization and pragmatics in German. *Journal of Phonetics*, *17*, 257–282.

Port, R., Cummins, F., & McAuley, D. (1995). Naive time, temporal patterns and human audition. In R. Port, & T. v. Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 339–371). Cambridge, MA: Bradford Books/MIT Press.

Port, R. F., & Leary, A. (2005). Against formal phonology. *Language*, *81*, 927–964.

Prince, A., & Smolensky, P. (1993). *Optimality theory: Constraint interaction in generative grammar*. New Brunswick, New Jersey: Rutgers University Center for Cognitive Science.

Raaijmakers, J., & Shiffrin, R. (1992). Models for recall and recognition. *Annual Review of Psychology*, *43*, 205–234.

Rayner, K., Foorman, B., Perfetti, C., Pesetsky, D., & Seidenberg, M. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, *2*, 31–74.

Read, C., Zhang, Y., Nie, H., & Ding, B. (1986). The ability to manipulate speech sounds depends on knowing alphabetic writing. *Cognition*, *24*, 31–44.

Sampson, G. (1977). Is there a universal phonetic alphabet? *Language*, *50*, 236–259.

Saussure, F. d. (1916). *Course in general linguistics (W. Baskin, Trans)*. New York: Philosophical Library.

Savin, H. (1963). Word-frequency effect and errors in speech perception. *Journal of Acoustical Society of America*, *35*, 200–206.

Shiffrin, R., & Steyvers, M. (1997). A model for recognition memory: REM: retrieving effectively from memory. *Psychonomic Bulletin and Review*, *4*, 145–166.

Smith, E., & Medin, D. (1981). *Categories and concepts*. Cambridge, MA: Harvard University Press.

Stevens, K. N. (1980). Acoustic correlates of some phonetic categories. *Journal of Acoustical Society of America*, *68*, 836–842.

Stevens, K. N. (2000). Toward a model for lexical access based on acoustic landmarks and distinctive features. *Journal of Acoustical Society*, *111*, 1872–1891.

Stevens, K. N., & Blumstein, S. (1978). Invariant cues for stop place of articulation. *Journal of the Acoustical Society of America*, *64*, 1358–1368.

Stevens, K. N., Keyser, S. J., & Kawasaki, H. (1986). Toward a phonetic and phonological investigation of redundant features. In J. Perkell, & D. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 426–463). Hillsdale, NJ: Lawrence Erlbaum.

Stotz, K., & Bostanci, A. (2006). The representing genes project: Tracking the shift to post-genomics. *Community Genetics*, *9*, 190–196.

Strange, W. (1995). Cross-language studies of speech perception: A historical review. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 3–45). Timonium, MD: York Press.

Studdert-Kennedy, M. (2003). Launching language: The gestural origin of discrete infinity. In M. H. Christiansen, & S. Kirby (Eds.), *Language evolution: The state of the art*. Oxford, UK: Oxford University Press.

Tomasello, M. (1999). *The cultural origins of human cognition*. Cambridge, Mass: Harvard University Press.

Twaddell, W. F. (1935). On defining the phoneme. *Language monograph,* vol. 16, Linguistic Society of America.

van Gelder, T., & Port, R. (1995). Its about time. In R. Port, & T. v. Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 1–44). Cambridge, MA: MIT Press.

Von Humboldt, W. (1836/1972). *Linguistic variability and intellectual development* (G. C. Buck & F. A. Raven, Trans.). Philadelphia, PA: University of Pennsylvania Press.

Warner, N., Jongman, A., Sereno, J., & Kemps, R. R. (2004). Incomplete neutralization and other sub-phonemic durational differences in production and perception: Evidence from Dutch. *Journal of Phonetics*, *32*, 251–276.

Werker, J., & Tees, R. C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant Behavior and Development*, *7*, 49–63.

Whittlesea, B. (1987). Preservation of specific experiences in representation of general knowledge. *Journal of Experimental Psychology: Learning Memory and Cognition*, *13*, 3–17.

Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin and Review*, *8*, 44–57.

Ziegler, J., & Goswami, U. (2005). Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory. *Psychological Bulletin*, *131*, 3–29.