No role for phonology in speech processing: Taking a strong position

Robert Port
Departments of Linguistics and Cognitive Science
330 Memorial Hall, Indiana University, 47405
Feb 8/2009

ABSTRACT

It will be claimed here that experimental evidence about human speech processing and memory for linguistic material shows that words are not spelled in memory from letter-like units, whether phones or phonemes. Linguists, like others with a Western education and lifetime literacy, identify any speech quite reflexively as a sequence of letter-sized units. Consonants and vowels seem like directly observable units of language but they are not. The language data that are available to learners are the rich auditory patterns of speech plus visual, somatosensory and motor patterns. The evidence is strong that people actually employ high-dimensional, spectro-temporal, auditory patterns to support speech production, perception and memory in real time. Abstract phonology (with its phonemes, distinctive features, syllable types, etc.) needs to be reconceived as a social institution – an inventory of patterns that evolves over historical time in some community as a structure of symmetries and regularities in the community's speech corpus Linguistics should study (and actually *is* studying) the phonological (and grammatical) patterns of various communities of speakers. But linguists should not expect to find the descriptions they make to be explicitly represented in the content of individual speakers' minds.

[To appear in *Proceedings of Meetings on Acoustics* (Acoustical Society of America). Online.]

This paper presents a novel and radical view of the relationship between speech processing by the individual speaker and the language of a community. The story presented here has been anticipated in other papers of mine in the past few years (Port, 2005, 2006, 2008) and has also been anticipated in some other recent work on phonetics (e.g., Faber, 1992; Pisoni, 1997; Hawkins, 2003; Linell, 2005; Johnson, 2006; Pierrehumbert, 2001, 2002, etc.). Still it may strike some readers as radical or even farfetched. The implications for theories in all areas of speech science and linguistics are far-reaching. Indeed, similar reasoning will apply to syntax just as much as to phonology, however this essay will focus just on phonology.

1. OVERVIEW OF THE ARGUMENT

For at least a century, linguists and psychologists have trusted their intuition that phonetic segments – i.e., consonants and vowels – are directly observable in speech. We assumed that C and V segments can be taken as valid raw data for the empirical study of language (Saussure, 1916; Jones, 1918, p. 1; Fant 1973; Ladefoged, 1972; Chomsky & Halle, 1968; Liberman, et al, 1968; IPA, 1999, Introduction). It was assumed Cs and Vs provide the appropriate psychological spelling system for every language. But my claim is that these powerful intuitions are at least partly (but probably mostly) a result of the lifelong literacy training to which all readers of this paragraph have been subjected (Faber, 1992; Linell, 2005; Olson, 1994; Port, 2006). The mass of experimental evidence over the past half century, it seems to me, actually supports a dramatic change in our thinking: the segment-based, ``economical,'' common-sense, low-bitrate view of linguistic memory is probably illusory. It is not the kind of memory we have. At the very least, any linguistic memory must be supplemented by a rich memory for speech material, resembling episodic memory for everyday events and activities, memory that is detailed and context-specific along with category assignments or identifications. Humans are certainly capable of abstract generalizations, but memory does not depend on abstract generalizations. What the relative contribution is of sensory abstraction versus concrete sensory images to record membership in various linguistic categories is still an open question. However memory research, much of it done on vision, offers strong evidence that people typically retain rich, detailed memory traces of specific sensory events lasting for at least a few days (e.g., Nosofsky, 1986; Shiffrin and Steyvers, 1997; Shiffrin, 2003; Tulving, 2002). It is possible (indeed, likely) that similar detailed, real-time memory representations for speech and language (as well as the coupled body movement, etc.) are also generated. This possibility of rich linguistic memory has been addressed by only a few psychologists and linguists (e.g., Klatt's LAFS speech recognition system, 1979; Jusczyk, 1993; K. Johnson, 1997, 2006; Pisoni, 1997; Pierrehumbert, 2001, 2002; Bybee, 2001; Coleman, 2002). I will try to show below there is surprisingly weak

evidence from real-time processing behavior that demands the hypothesis of a memory system for language that is restricted to abstract, speaker-independent units arrayed in serial order (the way phonemes and letters are arrayed).

But there is some evidence of abstract patterns in speech. It has been shown that speakers do generalize in perception across some phonological contexts. Thus, when a participant is familiarized with idiosyncratic pronunciations of some speech sound by a particular speaker, the subject will not only make a correction for training items spoken by the familiar voice (when presented with ambiguous stimuli), but also generalize the compensation to syllabic positions that have not been heard (Norris et al, 2003; Cutler et al, 2006; Nygaard & Pisoni, 1998). This is evidence that speakers can generalize a segment-sized unit and are thus not condemned to rely only on previously heard exemplars. An abstraction-only account (the traditional view) fails, but it appears that both concrete memories as well as some abstractions may be stored.

Of course, at the conscious level, if something appears in the visual field, then we might imagine that after the perceptual system has settled to an attractor (one that identifies the visual target as, say, a ``bird''), ordinary literate speakers like us will typically also arrive at some idealized alphabetic description of that word in terms of letters (orthographic or phonetic). But the intuition that all humans must automatically generate an alphabetical representation of words when they name something is probably not true. Nevertheless, the traditional view prevails very strongly among scientists of language despite the lack of evidence (cf. Ladefoged 1972; Hockett, 1955; Chomsky & Halle, 1968; Lyons, 1968; Liberman, et al., 1957, 1968; Kent and Minifie, 1977; Prince and Smolensky, 1993, Stevens, 1998).

The first goal of this paper is to attempt to shake the reader's confidence in segmental descriptions of words and speech by marshalling some of the empirical literature (much of it very well known) that repeatedly fails at critical points to support expectations of the hypothesis of a realtime psychological role for phones, phonemes or vectors of distinctive features. The second goal is then to reshape the landscape of phonology, phonetics and linguistics as a whole.

For speech scientists outside linguistics (cognitive psychology, engineering, speech science, neuroscience, etc.), the issue is 'what is the relevance of linguistic theory and linguistic terminology for my research?' The surprising answer (given vividness of our intuitions) turns out to be not nearly as much as we thought. It will be argued that the ability of speakers to perceive and produce speech does not depend on a discrete symbolic code that is abstract and low dimensional. The low dimensional description that linguists call "phonological structure" actually exists only as a set of statistical generalizations across some corpus of speech. Linguists and psychologists who deal with language often act as though they believe that their phonetic alphabet (whether using IPA or Chomsky-Halle phonetic features) will eventually be operationalized in concrete physical terms. But physical definitions for segments and features have never been found. To succeed would require some acoustic property be found everywhere that a [d] or an [i] or a [-voice] feature occurs. It has been clear for half a century that such invariants will never be found with anything approaching the degree of generality that is demanded by the context-independence that the segments themselves are assumed to have (Jakobson, Fant & Halle, 1952; Liberman, et al. 1968; Stevens & Blumstein, 1978; Stevens, 1998; see Huckvale, 1997).

The statistical patterns that constitute a phonology live in a fairly high-dimensional space in continuous time, and were shaped by many generations of speakers. For various reasons (see Abler, 1989; Studdert-Kennedy, 2003), the phonological systems of languages tend toward a low-dimensional format of distributions that are roughly discrete. Thus approximately the same set of vowels typically occurs with various preceding and following consonants*: beat, bit, bet, bat; mean, min, men, man*; etc. (although the vowels differ in duration, nasality, etc.) This creates many symmetries and regularities across words and across segments (e.g., *b, d, g; p, t, k; m, n,ŋ*). But speakers do not know their language using a low-dimensional phonological code. They get along very well with a far richer and more concrete representation of speech for storing concrete or abstract linguistic fragments and chunks (Grossberg & Myers, 2000). This is why linguistic data support only approximate discreteness, symmetry and mutual independence for phonetic features. Of course, some individual speakers may have an organized, abstract knowledge of their own phonology (such as amateur or professional linguists and others whose profession involves speaking or singing skills), but these individuals will most often be found to have had much experience with an orthographic alphabet as well. Experience and skill with these discrete graphic tokens, the letters of the alphabet, provide a "cognitive scaffold" (Clark, 1997) that encourages phonological understanding in terms of letter-like tokens (Port, 2006). We do not come to the discipline of linguistics with only the biases of skilled speakers of, say,

English, but we also come with biases that follow from being skilled readers and writers. The effect of literacy on our "linguistic intuitions" has been generally acknowledged (but see Faber, 1992; Linell, 2005).

That is an outline of the new theory that will be defended below. The implications of this new view are broad since it releases linguists from concern about data from psycholinguistic experiments, and it releases psychologists and neuroscientists from responsibility to find the kind of things linguists claim they should be looking for. The new theory also helps us to see that all physical symbols as well as idealized symbol processing (including computing) are based on a specific technology – the Small Alphabet of letters and digits. Math, logic and formal linguistics (not to mention, of course, literacy itself) involve cognitive skills that depend on the small alphabet. The fact is that ordinary speech production and perception are acquired without an alphabet or any other technology. But the skill of transcribing speech with an alphabet usually requires at least a year of training (starting at age 5) when the orthography is a regular and consistent (like Finnish) but requires about 3 years with a less transparent orthography (like English, see Rayner et al., 2001; Ziegler & Goswami, 2005; Goswami et al., 2005). This training came so early in our lives, it is very difficult for us to imagine ourselves without skill at relating speech to letters and vice versa. It is inherently very difficult for us to think about speech without an alphabet, yet it is essential to do so if we want to understand linguistic behavior (see Port & Leary, 2005).

2. PREDICTIONS AND EVIDENCE

The next section will justify the theoretical ideas presented above by reviewing data from a range of different areas of research that are appropriate: phonetics, speech perception, language development, second-language learning and so forth. If the traditional theory of phonology is to be taken seriously as involving psychological claims, then we should review whether its inherent predictions are supported. Thus if words are actually stored using a small set of abstract discrete tokens like vowels and consonants, as assumed by most linguists, then many simple predictions should follow. For example,

1. the first prediction is that both synchronic variation and diachronic sound change should always exhibit discrete phonetic jumps whenever a feature changes or one segment type is replaced with a different segment type.

2. Second, each distinctive feature should have a single invariant physical definition across all contexts. This is what Chomsky and Miller (1963) called "the invariance condition" on phonetic features – that they each have an invariant physical correlate, since otherwise how would one know which features are present? But in the last half century there has been virtually no success at finding such physical invariants for distinctive features or segments.

3. A third prediction is for the absence of temporal effects that cannot be described using segments to do so (e.g., by inserting, deleting or replacing the segments or features) since the segment model allows only serially ordered tokens for representing events in time. Finally,

4. our memory for specific utterances should show evidence that the stored descriptions are invariant across contexts, across speakers and across speaking rates. This means the memories do not differentiate between contexts, speakers, rates, etc. So one kind of supporting evidence would be a tendency to remember words without remembering who spoke them. This could happen if the hypothetical associative link between the linguistic representation and the indexical features of the speaker were to be lost.

These are all simple consequences of assuming that the acoustic signal of speech somehow contains units directly analogous to letters, i.e., phones or phonemes. But it is well-known, and in some cases it has been known for a half century, that not one of the expectations above is fulfilled, as is shown below.

Rich Variation of Language. The traditional theory of language, as described by linguists, predicts that the abstract, canonical representation of each word (analogous to an orthographic spelling) is used by speakers both for recognizing and remembering what someone said. After all no alternative representation exists according to linguistics. But researchers on linguistic variation (e.g., Labov, 1994; Bybee, 2001) along with generations of literature from experimental phonetics (e.g., Peterson & Barney, 1952; Lisker and Abramson, 1964, 1967; Local, 2003; Hawkins, 2003; K. Johnson, 2001, 2006; Hay & Drager, 2007) have shown that the variety of actual pronunciations for any linguistic chunk that speakers may hear is seemingly unlimited and may vary along many

continuous parameters (e.g., frequency of vocal pitch, voice-onset time, formant values, etc.). Furthermore, speakers are perceptually sensitive to many aspects of many subtle variations. Most adult speakers are familiar with several regional or social accents, foreign-accented pronunciations and the idiosyncrasies of a large number of individual speakers. Some speakers can even imitate several dialects or speaker voices. It seems that all speakers vary their own pronunciations along a huge number of phonetic continua depending on subtleties of the social and pragmatic context and the cultivated linguistic skills of the speaker.

The basic problem, the unavoidable question, is: how could these minute phonetic differences be employed in perception and controlled in production if linguistic memory could store only a canonical, abstract representation based on a minimal number of supposedly ``distinctive'' phonetic features in serial order? Since speakers recognize such a range of minute phonetic and temporal differences and can control many of them in their own speech, it seems obvious they must have a way of remembering them. The only conclusion is that speakers employ richly detailed phonetic memory representations for speech, and apparently not abstract supposedly "economical" ones using a short list of letter-like feature vectors (as claimed by Jakobson et al., 1952; Stevens & Blumstein, 1978; Stevens, 1998). Linguists insist there must be some ``economical'', uniform representation for each word. Words may have a fixed spelling in our orthography (that is enforced by the educational establishment of the country), but in the spoken language there is no evidence of an alphabet at all.

Recognition Memory. Another prediction of the hypothesis of alphabetical representation is this: When we remember what words someone said, we should rely on a memory that stores the linguistic information without speaker idiosyncrasies. Thus, we should predict difficulties on tasks that require remembering who said what. (Of course, subjects might also employ a completely different memory that stores indexical information about a speaker's voice, but it should be distinct from linguistic memory.) This prediction of an abstract code for language can be tested using "recognition memory'' experiments where, for each word in a spoken list, the participant responds whether it has been presented before. Linguistic theory claims that words are remembered in terms of abstract, serially ordered spellings using a small number of phonological (or phonetic) units. Thus, if I hear someone say *tomato*, in some dialect and speaking style, then what should be stored and available to support later cognitive operations should be a canonical phonological spelling, something like

[təme^{j}to^{U}]

which is approximated in our orthography as *tomato*. Indexical details about the specific utterance, such as the identity and sex of the speaker, the timing details of the pronunciation or subtle dialect variations, etc., are not part of the linguistic representation *per se* and should not be stored with the words themselves (Chomsky and Halle, 1968; Kent & Minifie, 1977; Pisoni, 1997). This abstract representation is often assumed to be somehow more efficient than one storing large amounts of auditory information much of which is, as we say, ``linguistically irrelevant'' (Jakobson, et al, 1952).

But speaker identity and timing patterns do, in fact, influence performance in recognition memory tasks (Goldinger, 1996; 1998; Pisoni, 1997). For example, if a subject hears a continuous list of spoken words and is asked to indicate when a word is repeated in the list, accuracy declines the greater the amount of time between the first presentation and the second (of course). But if the list is pronounced by many voices that change from word to word (and participants are told to ignore the voice), then one can compare performance on a word repeated with the same voice vs. repeated by a different voice. The traditional view of speech memory would predict no difference in performance since the code storing the word in memory will be the same if the word is the same. But the data show that words repeated in the same voice are recognized almost 10% more accurately than words repeated in a different voice (Palmeri, et al., 1993). This seems to imply that speakers store a representation that includes details about the speaker's voice. Even more surprisingly, the improvement is exactly the same whether just 2 voices read the list or 20 voices read the list. If the subjects were remembering the word in linguistic (i.e., speaker independent) form, and associating the abstract word with the identity of the speaker, then more voices should lead to more confusion than fewer voices. Yet there is no difference. This suggests the representations of the words are rich enough that the voices are very distinct from each other. Furthermore, some improvement for the same voice over a different voice can be detected up to a week later (Goldinger, 1998). The unavoidable inference from results like these is, again,+ that speakers automatically store much richer and more detailed representations than linguists and many other cognitive scientists ever imagined. Of course, speakers might store abstract representations as well, but evidently they are not limited to these. At the very least, any arguments claimed to support a realtime role for abstract

segments will need to be much more critically evaluated in future than they have been in the past.

Speech Perception. Many well-known phenomena of speech perception are quite incompatible with abstract, segmental representations but fit very well with a view of language storage that is concrete and detailed as proposed in exemplar models of memory (Hintzman, 1986; Goldinger, 1996; Pierrehumbert, 2001; K. Johnson, 2006). For example, a major problem in speech perception is the difficulty of accounting for the fact that we seem to perceive speech as segmented into context-free phones or phonemes, despite the fact that the acoustic signal is continuous, variable from context to context, and the ordered speech gestures overlap each other. In saying that we perceive speech as segmented, it is meant that our intuitions about the structure of speech are that it consists of letter-like units arranged serially (Fant, 1973; Liberman, et al., 1968; Kent and Minifie, 1977). These phones seem to be the direct output of our speech perception process. The discreteness and serial order of segments create the problem of "coarticulation": how is the linkage made between serial, discrete segments and the continuous-time speech signal? Acoustic information and articulatory gestures for neighboring segments overlap greatly. But the seeming directness of segmental speech perception probably comes about only when we learn to read, not when we learn to speak. Research on the "phonological awareness" of various subject groups has shown that the intuitions that allow us, for example, to add or delete a single segment from a word are found only in those who have had alphabet training (Morais, et al, 1979; Rayner, et al, 2001). So the entire coarticulation issue disappears as a problem for understanding speech perception and production once we realize that speakers employ rich and detailed speech memories. A preliterate child may be able to produce and correctly categorize productions of *Dee* and *dew* but not learn that they begin with ``the same consonant'' until he learns to write. Of course, the intuition of the coarticulation problem remains, but its explanation will come from better understanding of literacy skills. But in the case of reading, there are the discrete graphic letters themselves to help account for the vividness of our discrete letter-like intuitions.

Thus, as shown in Figure 1, the syllables [di] (*Dee*) and [du] (*dew*) probably do not share any unit in actual memory, *contra* Liberman, et al., 1968 (and most of the literature of linguistics). Linguistic memory does not extract an abstract, context-free, nonoverlapping invariant for each consonant and vowel. At least, it normally does not until one has had training for alphabet literacy. Literate people can consciously describe speech to themselves using an orthographic or phonological (that is, idealized alphabetical) code, but for realtime tasks, apparently, they rely on their richly detailed auditory memory. Apparently, human speakers have the ability to hear a novel linguistic stimulus and to find the appropriate linguistic categories (e.g., word and phrase identities, the identities of various speech sound types, etc.) by searching a large, personal utterance memory for all the many kinds of closest matches (see Johnson, 2006). Of course, there are many closest matches since any given utterance fragment shares categories with many other fragments in memory based on phonological, grammatical and lexical similarities of this utterance to other remembered utterances.

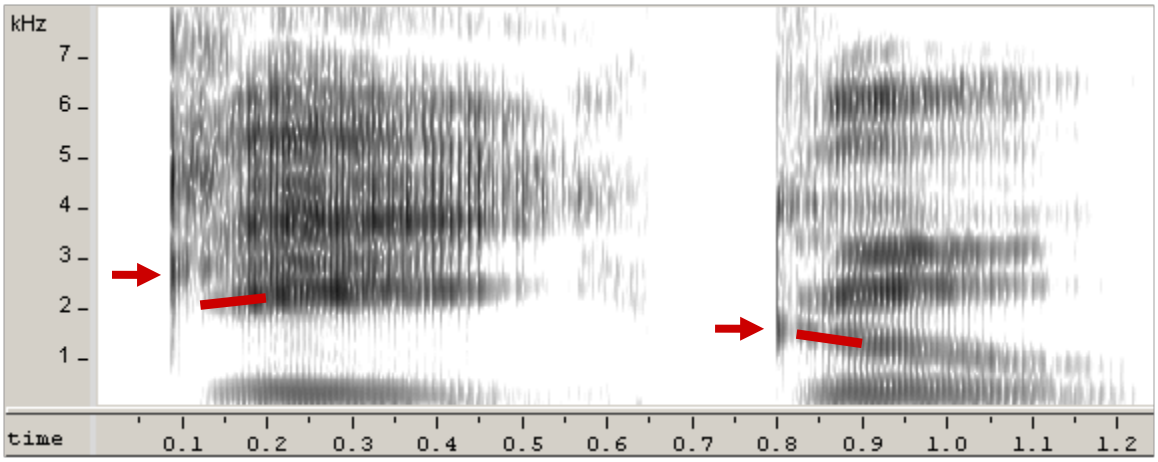

Figure 1. Spectrograms of the syllables *Dee* [di] on the left and *dew* [du] (by an American male) illustrate Liberman's point that the second formant goes in opposite directions from the burst into the vowel for *Dee* relative to *dew*. F2, an important cue, rises in [di] and falls in [du]. Plus the first resonance peak of the burst is higher in *Dee* than in *dew*. They sound to us like they have the same initial ``sound,'' but do not

have anything obvious that is a physical invariant. This situation is not rare, but is actually the norm wherever you look in speech. It is probably the case that learning to read with an alphabet helps us appreciate what is the same in the onset of *Dee* and *dew* (the origin of the articulatory gesture, of course).

All these phenomena make it difficult to imagine how the realtime psychological processing of language could possibly be limited to dealing with segmental units. Any representations used must not be dependent on the abstract, low-dimensional, non-overlapping descriptions proposed by traditional linguistics and powerfully supported by our intuitions. The data from many areas of speech research imply a speech memory that routinely stores large amounts of highly redundant and rich speech material (rather than minimal representations with a single form for each lexical item). These representations are coded by very rich auditory, articulatory and somatosensory codes (Guenther & Perkell, 2004). All these memory representations are linked to their categorical assignments as well (e.g., lexical categories, phonological categories, etc.). Things that belong to the same category need not share any specific property other than that people think of them as belonging to the same category (or use the same word or letter for them). A representation based on alphabetical principles is one that is obviously useful for a linguistic representation using ordered graphic tokens that can be written on paper, but it is not useful for someone talking in real time. Furthermore, each individual speaker's detailed auditory and linguistic code itself is sure to differ in detail from all other speakers, due to differing developmental histories.

It seems that people remember as much as they possibly can about details of specific utterances. Abstractions and generalizations can probably be extracted as needed from a memory system containing many detailed concrete instances and a large set of category labels. One very primitive model for each episode or exemplar (or utterance fragment) is simply a long vector of all co-occurring features of an event, and a model for long-term memory could be a large matrix of such exemplar vectors (Hintzman, 1986). Retrieval from memory would depend on the similarity of a probe vector (measured by comparing certain numbers in the vectors) to all features in all the stored vectors and a nonlinearity could strongly favor the closest matches over all others. Each category looks for its expected features in the stored vectors seeking a resonance between the input (probe) vector and a category. The result of such a process is a category decision (and perhaps also a subjective, conscious experience of the linguistic category) (Grossberg, 1995, 2003; K. Johnson, 2006). Evidence that favors exemplar-based rich memories and categories has been known for a long time in experimental psychology (Posner & Keele, 1968; Nosofsky, 1986; Shiffrin and Steyvers, 1997) and the human ability to remember randomly collocated events on a single exposure (such as your own memory for events that happened to you earlier today) is familiar to us all (Tulving, 2002; Goldinger, 1996). The precise role of exemplars vs. abstract generalization in human memory remains a major topic of research, but it is apparent, nonetheless, that traditional linguistic theory has been predicated on assumptions about human memory for linguistic material that are far too restrictive and are basically implausible. Low-dimensional serially ordered descriptors simply will not do the job of modeling the basic facts of linguistic behavior.

3. PHONOLOGY ANEW

If words are actually stored in a detailed and speaker-specific way, then what role is there for a linguistic description in a low-dimensional, phonological alphabet? The phonological patterns, such as the consonant and vowel types, the distinctive features, the restricted range of syllable types, etc., of each language are obvious in the corpus of any speech community. But where could such structures come from if they are not explicitly represented in memory? The answer is that they emerge from the collective behavior of a community of speakers (Port, 2007, 2008). Linguists draw these generalizations across speech contexts and across speakers that reflect how speech patterns tend to cluster across utterances. One might predict that if a random group of speakers of different languages began to behave and interact as a community, the language they would create over a generation or two would exhibit a low-dimensional, abstract description like the phonologies of typical languages. Natural experiments similar to this have occurred in the creation of pidgin languages (Hall, 1966). A variety of laboratory simulations of the development of new language-like systems have been done with human subjects either by creating a task where they must communicate with each other using a completely novel medium (Gallantucci, 2005; Gallantucci & Steels, 2008) or by doing simulations using artificial computational agents (e.g., Steels & Vogt, 1997; deBoer, 2000; Cangelosi & Parisi, 2002). The point of these experiments and simulations is to demonstrate that new communication systems can self-organize rather quickly when a need arises (Kirby & Hurford, 2001). These systems do not seem to begin by establishing clearly definable symbols or words. The participants and computational agents may have no idea what communicates what or what the ``signals'' and ``meanings'' are. But they evolve behaviors that serve the function of communication in that community. My hypothesis is that natural languages evolve the same way.

The linguist interested in the phonology of a speaker group should look across a large set of utterances (by at least several speakers) and use whatever descriptive tools can be found (such as the IPA phonetic alphabet, sound spectrograms, speech analysis software, speech recognition algorithms and possibly even Chomsky and Halle's phonetic features) in an attempt to describe the patterns found there. The linguist's corpus is, of course, just an approximation to the corpus of the ambient language that presents itself to a typical language learner. Some of these patterns (e.g., many traditional phonological phenomena involving phonemes, features, syllables, etc.) can be described using an abstract technical alphabet like the IPA alphabet, as long as it is kept in mind that the alphabetical description may be easy for the phonologist to interpret, but is not what the speaker actually relies upon. Many other language-specific patterns require instrumental measurements of frequency-by-time trajectories (e.g., intonation contours, voice-onset time, mora patterns, phrase-edge lengthening, formant transitions, spectrum shapes, vowel and consonant durational patterns correlating with the voicing feature, etc.). Both kinds of descriptive tools, the impressionistic and the laboratory-based, aim to capture the properties that are shared across the speech community and also represent distinctions that are probably exploited by many speakers of the language to differentiate various sets of lexical items. This is how phonology should define its goals. It simply cannot make claims about realtime psychological representations either in all speakers of a language or in any particular speaker. The phonological descriptions that result from such research will resemble traditional ones but without the various constraints on the formalism that come from traditional phonological theory based on erroneous assumptions about the limitations of linguistic memory. Although a native speaker never needs a phonological description, this description, as partially embodied in an orthography, provides an essential resource for teaching a language to speakers of another language and also provides a practical basis for development (or reform) of an orthography. But phonological descriptions should not be expected to play much of a role in realtime perception and production by skilled speakers.

If speech memory is so much richer than we thought, then why, one might ask, do small graphic alphabets (such as the Greek, Latin and English alphabets) work as well as they do for converting language to graphical form? First of all, what does it mean to "work well"? A small alphabet for spelling words ``works well'' when a learner has to memorize the alphabet before beginning to read words. But would an alphabet work well for one who is still learning to talk? It is not likely. We know that one who is learning to speak needs to produce and perceive phonetic trajectories in time (Jusczyk, 1997; Jusczyk & Derrah, 1993). But alphabets have serious limitations since they leave out, for example, almost all the temporal characteristics of speech and much more (see Port & Leary, 2005). But first-language learners need to get timing subtleties just right if they want to be native speakers. A phonetic transcription provides only a very rough and approximate description for anyone except those who already know how the transcribed language should sound and can fill in the missing features.

But the question of why alphabets seem so appropriate is still important to answer. On the view presented in this essay, a language is a social institution that is shaped by generations of users. That shaping process tends toward a lexicon that exploits a very restricted set of combinations of the various phonetic degrees of freedom, i.e., the phonetic features of a language. There tend to be lots of differences between words but also much reuse of pattern fragments in new contexts. The result is that it is possible to describe much about the corpus of a language using a small set of distinct units. This is the description we call alphabetical writing, which turns out to support literacy for skilled speakers pretty well. A trend toward using a restricted set of patterns in the phonetic space probably reveals an attractor state for human languages. When a particular system as a whole approaches a low-dimensional description, it will typically result in many specific articulatory or auditory attractors (e.g., the phonemes, distinctive features, limited syllable types, distinctive intonation contours, periodic patterns, etc.) in the Speech-Language Processing system of speakers. Very likely, a phonology that approximates a low-dimensional characterization is easier to learn and to understand and may facilitate speakers in the creation of new words (Abler, 1989; Studdert-Kennedy, 2003). But this does not imply that these apparent speech categories are discrete cognitive units or ``symbols'' in our representation of language in memory. The phonological patterns of a community and the units of a realtime memory in a speaker exist on very different descriptive levels (communal vs. personal) and evolve on very different time scales.

Comparison with Competence vs. Performance. It may help the reader to appreciate the distinction developed here between Phonology and Speech-Language Processing by contrasting this distinction with Chomsky's (1965) distinction of Linguistic Competence and Linguistic Performance. There are fundamental differences and I will try to show they are orthogonal. For Chomsky, Competence is the formal core of language. It is purely discrete and

mental and the system that exhibits linguistic creativity but is surrounded on input and output (to the external world) by the Performance system. The processes that actually create the phonetic transcription belong to Performance.

Competence	Words in abstract, low-dimensional (letter-like) code (the phonology). Rules and operations for the generation of sentences Invariant across all speakers of a language Stable within a speaker over time
Performance	Speech production apparatus - implementation of phonological representations; muscular control Speech perception apparatus: hearing, phonetic perception (categorization into phonological units) Long-term memory

On this view when a listener hears speech, it is automatically converted by the Performance system into a string of discrete phonetic segments each of which is a short vector (not more than 40 or so binary values) of phonetic features. In speech production, the phonetic alphabet is the output of the Competence system which is implemented on something like a Performance keyboard. The mental Competence ``plays'' the physical Performance system in discrete time. Linguistics is concerned only with the formal aspects of language lying between the symbolic phonetic code as input and the same code as output. Performance is everything else outside the formal description of language. Thus, all continuous-time aspects of speech, any phonetic details lying below the level of discrete phonetic features and any constraints or errors due to memory limitations are all aspects of Performance and, thus, from the standpoint of the linguist, are irrelevant since they serve merely as noise that tends to obscure the true formal structure of language. Of course, I argue that this view is deeply mistaken.

	Phonology	**Speech-Language Processing**
Competence-like	Words are abstract (over speakers, rates) Low dimensional (approximately) Stable over time	Words stored in a physically definable code – Speech perception Speech production: Utterance construction and creativity
Performance-like	Temporal patterns Auditory trajectories Inventory of variant pronunciation	Real-time speech perception Categorization Speech production

The distinction proposed here is orthogonal to Chomsky's distinction since parts of both the proposed Phonology and Speech-Language Processing can be characterized as Competence-like and parts as Performance-like. Thus, looking first at my proposed Phonology, the social institution, it allows abstract word descriptions that may be differentiated using a small letter-like code relying on a rather small number of dimensions to do much of the work of distinguishing lexical entries – just as phonologists have insisted since the days of the Prague school (Troubetzkoy, 1939). These phonological descriptions are mostly invariant across speakers and generally stable over time. On the other hand, these descriptions should also include properties that resemble Chomsky's Performance system since phonological patterns include various kinds of temporal properties and auditory trajectories in continuous time. They also exhibit considerable variability and are basically distributed clusters in a space that has enough dimensions to store rich detail.

Similarly, the proposed Speech-Language Processing system that characterizes the linguistic agent stores words in an auditorily definable code (which Chomsky and Halle claimed for their universal phonetic features but were never able to make good on, because the features also had to be segmental and intuitively accessible) (Pisoni, 1997; Port and Leary, 2005). This system also accomplishes the construction and generation of utterances - again resembling Chomsky's Competence. Still, Speech-Language Processing does such Performance-like tasks as speech perception and the assignment of speech sounds to categories. The distinction proposed here cuts across Chomsky's distinctions by separating properties of the speech corpus available to the language learner – the system of regularities that is shaped over historical time to be useful for a community of speakers – from the actual concrete skills the individual speaker employs for speaking and listening in accordance with those phonological patterns.

4. CONCLUSIONS

The story presented here makes a radical break with the past and corrects a widespread error in our thinking and theorizing about language. The mistake was to trust our intuitions when we should have been more skeptical. The powerful intuitions we have relied upon for centuries -- that language is always structured in terms of discrete letter-like tokens – is largely a side effect of our years of literacy education and extensive practice of literacy – an effect that has been overlooked or discounted for too long. We all taught ourselves (with initial help from a teacher) to listen to and think about language in letter-based terms. It is likely that this ability is important for the skillful use of our orthography (see McCandliss et al. 2003 for discussion of some neurological effects of years of reading practice). Consequently we phoneticians, linguists, psychologists and speech scientists were all quite sure that the "real" structure of language had to somehow make use of the discrete symbol strings that we use for reading – despite all the contrary evidence that has been in front of us at least since the appearance of the sound spectrograph 60 years ago (Joos, 1948).

Linguists, from Saussure to Chomsky to Prince & Smolensky, 1993, have hoped that phonology could provide both (a) a description of the psychological code for realtime linguistic processing and also (b) a description of language that is the same from speaker to speaker. Despite the frequent insistence that a language is basically a code (Saussure, Hockett, Chomsky), individual speakers solve all these representational problems independently of each other, and thus differently. Language cannot be a simple code with discrete signifier tokens and discrete meanings shared by all speakers (see Harris,1981; Love, 2004). What the language learner needs to remember about his language is vastly more concrete and detailed than we thought. But since speech gestures are distributed in continuous time and exhibit dynamics reflecting the human vocal tract (e.g., Browman & Goldstein, 1992, 1995), it makes sense that memory and perception would demand continuous time as well. Phonological patterns (e.g., phonological categories like ``phonemes'' and ``features'') are only implicitly present in the memory of speakers (unless they are literate) and are not explicitly "represented" as we thought. Phonological generalizations - the patterns that are shared by the speech of a community - comprise categories alright, but they exist as statistical regularities only at the level of the community. These categories do not become real symbols until we assign letters to them (thus the written language uses symbols, but it seems the spoken language does not).

It turns out that what we have been loosely calling Linguistic Cognition has two very different parts. The first is the social product, the Language, the ``grammar'' and ``phonology,'' the community's ways of talking. The second is the realtime Speech-Language Processing system for learning, producing and perceiving speech (and everything else about life, of course). The first is categorized in various ways and polished or shaped by the community from generation to generation, while the second is born and, as we know, flickers out after some number of years. Because of idiosyncratic experiences, individuals are sure to differ in detail in their analyses. Phones and phonemes, though they come readily to our conscious awareness of speech, are not valid empirical phenomena suitable as the data basis for linguistics since they are not physically definable. So far, only people with literacy training can transcribe speech into an alphabet – after enough training. Phones and phonemes are interpretations of speech that are most strongly accessible to those with experience using alphabetical writing. Although this essay has addressed only the issue of phonology, it should be obvious that if language and phonology are reinterpreted this way, the rest of language will also have to be reinterpreted, if only because the nondiscreteness of phonetics guarantees nondiscreteness in syntax . The possibility of rich memory greatly changes our understanding of all aspects of language.

5. REFERENCES

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

Browman, C., and Goldstein, L. (**1992**). "Articulatory phonology: An overview," Phonetica **49**, 155-180.

Browman, C., and Goldstein, L. (**1995**). "Dynamics and Articulatory Phonology," in *Mind as Motion: Explorations in the Dynamics of Cognition*, edited by R. Port, and T. v. Gelder (MIT Press, Cambridge, Mass), pp. 175-193.

Bybee, J. (**2001**). *Phonology and Language Use* (Cambridge University Press, Cambridge, UK).

Cangelosi, A., and Parisi, D. (**2002**). *Simulating the Evolution of Language* (Springer-Verlag, London).

Chomsky, N. (**1965**). *Aspects of the Theory of Syntax* (MIT Press, Cambridge, Massachusetts).

Chomsky, N., and Halle, M. (**1968**). *The Sound Pattern of English* (Harper and Row, New York).

Chomsky, N., and Miller, G. (**1963**). "Introduction to the formal analysis of natural languages," in *Handbook of*

Mathematical Psychology edited by R. D. Luce, R. R. Bush, and E. Galanter (Wiley and Sons, New York), pp. 323-418.
Clark, A. (**1997**). *Being There: Putting Brain, Body, and World Together Again* (Bradford Books/MIT Press, Cambridge, Mass.).
Coleman, J. (**2002**). "Phonetic representations in the mental lexicon," in *Phonetics, Phonology and Cognition*, edited by J. Durand, and B. Laks (Oxford University Press, Oxford), pp. 96-130.
Cutler, A., Eisner, F., McQueen, J., and Norris, D. (**2006**). "Coping with speaker-related variation via abstract phonemic categories," in *10th Conference on Laboratory Phonetics* (Paris, France).
deBoer, Bart (2000) ``Self-organization in vowel systems.'' Journal of Phonetics **28**, 441-465.
Faber, A. (**1992**) "Phonemic segmentation as epiphenomenon: Evidence from the history of alphabetic writing," in *The Linguistics of Literacy*, edited by P. Downing, S. Lima, and M. Noonan (John Benjamins, Amsterdam), pp. 111-134.
Fant, G. (**1973**) *Speech Sounds and Features* (MIT Press, Cambridge, Mass.).
Galantucci, B. (**2005**). "An experimental study of the emergence of human communication systems," Cognitive Science **29**, 737-767.
Galantucci, B., and Steels, L. (**2008**). "The emergence of embodied communication in artificial agents and humans," in *Embodied Communication in Humans and Machines*, edited by I. Wachsmuth, M. Lenzen, and G. Knoblich (Oxford University Press, Oxford), pp. 229-256.
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.
Goswami, U., Ziegler, J., and Richardson, U. (**2005**) "The effects of spelling consistency on phonological awareness: A comparison of English and German," Journal of Experimental Child Psychology **92**, 345-365.
Grossberg, S. (**1995**). "Neural dynamics of motion perception, recognition learning and spatial attention," in *Mind as Motion: Explorations in the Dynamics of Cognition*, edited by R. Port, and T. v. Gelder (MIT Press, Cambridge, Mass), pp. 449-490.
Grossberg, S. (**2003**). "The resonant neural dynamics of speech perception," Journal of Phonetics **31**, 423-445.
Grossberg, S., and Myers, C. W. (**2000**). "The resonant dynamics of speech perception: Inter-word integration and duration-dependent backward effects," Psychological Review **107**, 735-776.
Guenther, F., and Perkell, J. (**2004**). "A neural model of speech production and its application to studies of the role of auditory feedback in speech," in *Speech Motor Control in Normal and Disordered Speech*, edited by B. Maasen, R. Kent, H. Peters, P. VanLieshout, and W. Hulstijn (Oxford University Press, Oxford, England), pp. 29-49.
Hall, R. (**1966**). *Pidgin and Creole Languages* (Ithaca, Cornell University Press).
Harris, R. (**1981**). *The Language Myth* (Duckworth, London).
Hawkins, S. (**2003**). "Roles and representations of systematic fine phonetic detail in speech understanding," Journal of Phonetics **31**, 373-405.
Hay, J., and Drager, K. (**2007**). "Sociophonetics," Annual Review of Anthropology **36**.
Hintzman, D. L. (**1986**). "`Schema abstraction' in a multiple-trace memory model," Psychological Review **93**, 411-428.
Hockett, C. (**1955**). "Manual of Phonology," (Linguistic Society of America, Baltimore, Md).
Huckvale, M. (**1997**). "10 things engineers have discovered about speech recognition," in *NATO ASI Speech Patterning Conference* (Jersey, England).
IPA (**1999**). *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet* (Cambridge University Press, Cambridge, England).
Jakobson, R., Fant, G., and Halle, M. (**1952**). *Preliminaries to Speech Analysis: The Distinctive Features* (MIT, Cambridge, Massachusetts).
Johnson, K. (**1997**). "Speech perception without speaker normalization: An exemplar model," in *Talker Variability in Speech Processing*, edited by K. Johnson, and J. Mullenix (Academic Press, London), pp. 145-166.
Johnson, K. (**2001**). "Spoken language variability: Implications for modeling speech perception.," in *Proceedings of the Workshop on Speech Recognition as Pattern Classification*, edited by R. Smits, Kingston, J, Nearey, T., Zondervan, R (Max Planck Institute for Psycholinguistics, Nijmegen).
Johnson, K. (**2006**). "Resonance in an exemplar-based lexicon: The emergence of social identity and phonology," Journal of Phonetics **34**, 485-499.
Jones, D. (**1918**). *An Outline of English Phonetics* (Teubner, Leipzig, Germany).

Joos, M. (**1948**). "Acoustic Phonetics," Language Monograph **23**.
Jusczyk, P. (**1993**). "From general to specific capacities: The WRAPSA model of how speech perception develops," Journal of Phonetics **21**, 3-28.
Jusczyk, P. (**1997**). *The Discovery of Spoken Language* (MIT Press, Cambridge, Mass.).
Jusczyk, P., and Derrah, C. (**1987**). "Representation of speech sounds by young infants," Child Development **64**, 675-687.
Kent, R., and Minifie, F. (**1977**). "Coarticulation in recent speech production models," Journal of Phonetics **5**, 115-135.
Kirby, S., and Hurford, J. (**2002**). "The emergence of linguistic structure: An overview of the iterated learning model.," in *Simulating the Evolution of Language*, edited by A. Cangelosi, and D. Parisi (Springer-Verlag, London), pp. 121-148.
Klatt, D. (**1979**). "Speech perception: A model of acoustic-phonetic analysis and lexical access," Journal of Phonetics **7**, 279-312.
Labov, W. (**1994**). *Principles of Linguistic Change, Internal Factors, Volume I* (Wiley-Blackwell).
Ladefoged, P. (**1972**). *A Course in Phonetics* (Harcourt Brace Jovanovich, Orlando, Florida).
Liberman, A. M., Delattre, P., Gerstman, L., and Cooper, F. (**1968**). "Perception of the speech code," Psychological Review **74**, 431-461.
Liberman, A. M., Harris, K. S., Hoffman, H., and Griffith, B. (**1957**). "The discrimination of speech sounds within and across phoneme boundaries," Journal of Experimental Psychology **54**, 358-368.
Linell, P. (**2005**). *The Written Language Bias in Linguistics: Its Nature, Origins and Transformations* (Routledge, Oxford).
Lisker, L., and Abramson, A. (**1964**). "A cross-language study of voicing in initial stops: acoustical measurements," Word **20**, 384-422.
Lisker, L., and Abramson, A. (**1967**). "Some effects of context on voice-onset time in English stops.," Language and Speech **10**, 1-28.
Local, J. (**2003**). "Variable domains and variable relevance: Interpreting phonetic exponents," Journal of Phonetics **31**, 321-339.
Love, N. (**2004**). "Cognition and the language myth," Language Sciences **26**, 525-544.
Lyons, J. (**1968**). *Introduction to Theoretical Linguistics* (Cambridge University Press, Cambridge).
McCandliss, B., Cohen, L., and Dehaene, S. (**2003**). "The visual word form area: Expertise for reading in the fusiform gyrus," Trends in Cognitive Sciences **7**, 293299.
Morais, J., Cary, L., Alegria, J., and Bertelson, P. (**1979**). "Does awareness of speech as a sequence of phones arise spontaneously?," Cognition **7**, 323-331.
Norris, D., McQueen, J., and Cutler, A. (**2003**). "Perceptual learning in speech," Cognitive Psychology **47**, 204-238.
Nosofsky, R. (**1986**). "Attention, similarity and the identification-categorization relationship," Journal of Experimental Psychology: General **115**, 39-57.
Nygaard, L., and Pisoni, D. (**1998**). "Talker-specific learning in speech perception," Perception & Psychophysics **60**, 355-376.
Olson, D. R. (**1994**). *The World on Paper: The Conceptual and Cognitive Implications of Writing and Reading* (Cambridge University Press, Cambridge).
Palmeri, T. J., Goldinger, S. D., and 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.
Peterson, G., and Barney, H. L. (**1952**). "Control methods used in a study of the vowels.," JASA **24**, 175-184.
Pierrehumbert, J. (**2001**). "Exemplar dynamics: Word frequency, lenition and contrast," in *Frequency Effects and the Emergency of Linguistic Structure*, edited by J. Bybee, and P. Hopper (John Benjamins, Amsterdam), pp. 137-157.
Pierrehumbert, J. (**2002**). "Word-specific phonetics," in *Laboratory Phonology 7*, edited by C. Gussenhoven, and N. Warner (Mouton deGruyter, Berlin), pp. 101-140.
Pisoni, D. B. (**1997**). "Some thoughts on `normalization' in speech perception," in *Talker variability in speech processing*, edited by K. Johnson, and J. Mullennix (Academic Press, San Diego), pp. 9-32.
Port, R. (**2006**). "The graphical basis of phones and phonemes," in *Second Language Speech Learning: The Role of Language Experience in Speech Production and Perception*, edited by M. Munro, and O.-S. Bohn (John Benjamins, Amsterdam, Holland), pp. 349-365.
Port, R. (**2007**). "What are words made of?: Beyond phones and phonemes," New Ideas in Psychology **25**, 143-170.
Port, R. (**in press, 2009**). "Dynamics of language," in *Encyclopedia of Complexity and Systems Science*, edited by R.

Myers (Springer Verlag, London).
Port, R. F., and Leary, A. (**2005**). "Against formal phonology," Language **81**, 927-964.
Posner, M., and Keele, S. W. (**1968**). "On the genesis of abstract ideas," Journal of Experimental Psychology, Learning, Memory and Cognition **77**, 353-363.
Prince, A., and Smolensky, P. (**1993**). *Optimality Theory: Constraint Interaction in Generative Grammar* (Rutgers University Center for Cognitive Science, New Brunswick, New Jersey).
Rayner, K., Foorman, B., Perfetti, C., Pesetsky, D., and Seidenberg, M. (**2001**). "How psychological science informs the teaching of reading," Psychological Science in the Public Interest **2**, 31-74.
Saussure, F. d. (**1916**). *Course in General Linguistics* (Philosophical Library, New York).
Shiffrin, R., and Steyvers, M. (**1997**). "The effectiveness of retrieval from memory," in *Rational Models of Cognition*, edited by M. Oaksford, and N. Chater (Oxford University Press, Oxford, United Kingdom), pp. 73-95.
Steels, L., and Vogt, P. (**1997**). "Grounding adaptive language games in robotic agents," in *European Conference on Artificial Life 1997*, edited by I. Harvey, and P. Husbands (MIT Press, Cambridge, Mass).
Stevens, K., and Blumstein, S. (**1978**). "Invariant cues for place of articulation in stop consonants," Journal of Acoustical Society of America **64**, 1358-1368.
Stevens, K. N. (**1998**). *Acoustic Phonetics* (MIT Press, Cambridge, Mass.).
Studdert-Kennedy, M. (**2003**). "Launching language: The gestural origin of discrete infinity," in *Language Evolution: The State of the Art*, edited by M. H. Christiansen, and S. Kirby (Oxford University Press, Oxford, UK).
Troubetzkoy, N. (**1939**). *Grundzüge der Phonologie* (Travaux du cercle linguistique de Prague).
Tulving, E. (**2002**). "Episodic memory: From mind to brain," Annual Review of Psychology **53**, 1-25.
Ziegler, J., and Goswami, U. (**2005**). "Reading acquisition, developmental dyslexia and skilled reading across languages: A psycholinguistic grain size theory" Psychological Bulletin **131**, 3-29.