Ecological Psychology, 22:304–326, 2010

ISSN: 1040-7413 print/1532-6969 online
DOI: 10.1080/10407413.2010.517122

Routledge
Taylor & Francis Group

Language as a Social Institution: Why Phonemes and Words Do Not Live in the Brain

Robert F. Port
Department of Linguistics and Department of Cognitive Science
Indiana University

It is proposed that a language, in a rich, high-dimensional form, is part of the cultural environment of the child learner. A language is the product of a community of speakers who develop its phonological, lexical, and phrasal patterns over many generations. The language emerges from the joint behavior of many agents in the community acting as a complex adaptive system. Its form only roughly approximates the low-dimensional structures that our traditional phonology highlights. Those who study spoken language have attempted to approach it as an internal knowledge structure rather than as a communal institution or set of conventions for coordination of activity. We also find it very difficult to avoid being deceived into seeing language in the form employed by our writing system as letters, words, and sentences. But our writing system is a further set of conventions that approximate the high-dimensional spoken language in a consistent and regularized graphical form.

Language seems to pose a problem for ecological psychology. Language is a domain where it seems very difficult to argue that linguistic structures like words are only in the environment because our intuitions are strong that words are internal, mental symbols of some kind. The claim that words and phonemes are internal symbols with an arbitrary link to meaning is very persuasive to most scientists—except perhaps those few who have a theoretical commitment

Correspondence should be addressed to Robert F. Port, Department of Linguistics, Indiana University, Memorial Hall 322, Bloomington, IN 47405. E-mail: port@indiana.edu

to studying information in the environment. I propose a new way to think about language that I believe resolves these difficulties. For a newborn, language is clearly just part of its environment—something it hears and may have a special interest in. The child must learn to use the language but does not need to represent it explicitly. The low-dimensional patterns of language (such as the ones we represent in our orthography) belong to the community of speakers, not to any individual speaker. Given the massive amount of variation, actual language use requires the rich dimensionality of speech spectra over time for perception to be successful. There is apparently no way to separate linguistic from nonlinguistic information, so speakers must store and deploy linguistic material using a relatively rich descriptive vocabulary (Port & Leary, 2005). For speech production, as well, speakers require subtle control to express specific interpretations, attitudes, and feelings.

In recent years, the Distributed Language Group (DLG) has blossomed, endorsing ideas that seem completely compatible with the characterization of language presented here. The term *distributed* emphasizes that linguistic structures are not located in each individual but are distributed across a population. The DLG community (see other articles in this issue) emphasizes that language cannot be separated from the rest of human behavior without severe distortion. The minimal case of language use is not a transcription of a sentence out of context but two or more people in conversation on some topic. Language is just one aspect of the intense interpersonal coordination exhibited by humans. In fact, I claim no fundamental distinction can be made between linguistic conventions and other cultural conventions such as culture-specific gestures and facial expressions (Harris, 1981). Our professional linguistic definition of *language* is narrowed to describe largely the aspects of human vocal interactions that happen to be preserved by typical alphabetical orthographies. Thus hand and body gestures and facial expressions, and so forth, are outside linguistics—even though many linguistic expressions are tightly coupled with specific facial and manual gestures. One thinks of American expressions and gestures such as "*I* don't know" (spoken with a shrug, raised eyebrows, and the lip corners turned down); "Just a teensy-weensy bit more cake, please" (with finger and thumb demonstrating how little); "You said *what*?!" (with brows furrowed and head pushed forward); "Stop! Stop! Stop!" (with hands raised, palms out, fingers splayed). Intonation is an unavoidable accompaniment of consonants and vowels. But even intonation seems like an awkward borderline case that is studied by only a few linguists. The conventions in all these domains are acquired by very similar kinds of imitation learning: of linguistic expressions, tone of voice, gestures, idioms, constructions (Goldberg, 2006; Wray & Perkins, 2000), pronunciations (Labov, Ash, & Boberg, 2006), intonation patterns, patterns of semantic interpretation (Bolinger, 1976), gestures of the hand, rhetorical constructions, and gestures like shrugs. Children each discover some way to

communicate appropriately and to recognize and use more and more linguistic patterns throughout their lives. Individual speakers will surely find idiosyncratic solutions to the problem of using a language. The idiosyncrasies of personal linguistic conventions, due partly to different patterns of exposure to the corpus, is one of the main reasons linguists should focus their attention on language as distributed across a community and abandon the notion of an "idealized speaker-hearer" whose cognitive representations we traditionally seek to model (Chomsky, 1965). One will not generally find linguistic generalizations represented physically within any speaker, although they can be found in the corpus of linguistic behaviors. If one wanted to study the detailed representational mechanisms of some single speaker, it might be possible, but then one is not likely to find the kind of generalizations that linguists are interested in: generalizations like the inventory of sound patterns used by most speakers, the word constructions, the syntactic generalizations—that is, the ways that people usually talk.

Every speaker is a language-production and perception machine different from every other speaker. Nevertheless, it still seems intuitively persuasive to us that the code for each linguistic item should be the same from speaker to speaker. But in fact every speaker's "code," that is, stored speech chunks with associated interpretations, will surely be different. The abstract linguistic units—phones, phonemes, words, phrases, and so on—can be defined only at the group level and, even then, only approximately. If this sounds like nonsense to the reader, consider that complex adaptive systems are now beginning to be understood (Holland, 1995). Each human community is able to develop a culture over time that tends to continually adapt to its environment by slowly changing and improving its technologies for cooperative behavior, sustenance, defense, training of the young, and so on (Richerson & Boyd, 2005). Of course, language is one of the most important parts of a culture. From the perspective of young people, the language they hear is just one aspect of the cultural environment they were born into.

I first review the traditional view of speech and language, criticizing it as I go, then try to encourage the reader to consider how literacy might influence our linguistic intuitions. Finally, I sketch my proposal that a language is essentially a kind of social institution, something created by a community of speakers over generations, and is not separately represented in the memory of each speaker.

LANGUAGE AS A MENTAL CODE AND ITS PREDICTIONS

The standard idea about language for at least the past century (and perhaps for a millennium) is that it consists of discrete sound units composed into discrete words, which are, in turn, composed into sentences. All mainstream linguistic

theories of the 20th century, from Saussure (1916) to Chomsky (1965), proposed variants of these ideas. Linguists have differed in how the set of phones or phonemes was to be determined and what their specific formal properties are, but practically all agreed that a small inventory (less than 50 or so) of consonant and vowel tokens are the basic units of psychological representation for language and that they are discrete, nonoverlapping (i.e., serially ordered), abstract "speech sounds" that are independent of their neighboring context. Words (or morphemes) are the next larger psychological units, which, in turn, are combined somehow into real-time utterances whose basic structure consists of phrases and sentences. If valid, all such theories should have many implications for the psychological form of language. They imply many testable predictions about the details of speech production and perception. Unfortunately (for proponents), almost none of these predictions are supported by experimental data (Lisker & Abramson, 1971; Port & Leary, 2005; Sampson, 1977).

For example, if words are spelled from a discrete alphabet of modest size, then we should expect at least these five properties:

1. Phonetic jumps. Speech variation and historical sound changes should both exhibit jumps from one sound type to a neighboring sound type accompanied by some noisy blurring of boundaries. Thus, for example, a [t] might change to a [d] or an [i] to an [ɪ] (or in smaller steps if a larger phonetic alphabet were employed). But it has long been known that nothing resembling universal phonetic discreteness is found and that almost all of the parameters of speech (e.g., vowel quality, place of articulation, voice-onset time, vocal pitch, etc.) exhibit continuously varying target articulations, both within individual speakers (depending on context) and as differences between related dialects.

2. Segment-shaped physical correlates. There should be directly observable physical correlates of the boundaries between the segmental units (the consonants and vowels) like the ones we seem to hear phenomenally when listening to speech. But it has been known since the first speech waveform displays were made in early 20th century that letter-size units of speech cannot be identified in the speech pressure wave, even when frequency is displayed against time as in a sound spectrogram. Generally, it is very difficult, when looking at a spectrogram, even to determine how many letters it would take to transcribe a stretch of speech much less determine the location of any boundaries between the phones. The so-called segments we hear phenomenally (i.e., the consonants and vowels) do not at all resemble the intervals of homogenous wave type that are clearly visible on a sound spectrogram. The natural segmentation of the speech wave corresponds to abrupt changes in articulation, for example, to vocal tract closure, changes in glottal aperture, tongue or lip gestures, and so on (Fant, 1973). But nothing has the properties we expect at a "boundary between segments." In fact, it seems that the acoustic or auditory transitions between neighboring steady-states are the most informative part of the auditory information.

3. Physical invariance. It has generally been assumed that there must be acoustic invariants corresponding to the letters of phonemic or phonetic transcription, but research has failed to find acoustic patterns with those properties in most cases. If the letters of the phonetic alphabet comprise a valid psychological model, then each phone should exhibit the same acoustic pattern each time the same phone is used in a transcription (the invariance property) and the physical acoustic correlates of each symbol should have the same serial position as the symbols themselves (the linearity property; Chomsky & Miller, 1963). These constraints assure that the symbols can be reliably identified, ordered, and produced. If phones are to count as psychological "symbol tokens," then they must have invariance and linearity. Of course, these properties accurately describe letters (because letters have sufficient invariance of shape that we nearly always correctly identify them) and they describe bit strings, but do they describe speech?

Chomsky and Miller (1963) acknowledged that the phonemic level units of languages do not satisfy the invariance and linearity conditions. For example, stop consonants differing in place of articulation (e.g., /b, d, g/) are differentiated from each other by acoustic cues whose form tends to vary depending on both the preceding and following vowel as well as its position in the syllable (e.g., "dew" vs. "add"; Fant, 1973; Liberman, Delattre, Gerstman, & Cooper, 1968; Liberman, Harris, Hoffman, & Griffith, 1957; thus violating invariance). And the timing cues for [+voice] and [− voice] in, for example, "edger-etcher" or "lumber-lumper," are known to extend over most of the duration of a syllable (Hawkins and Nguyen, 2004; Klatt, 1976; Port & Leary, 2005; thus violating linearity). Chomsky and Miller apparently believed that the psychological phonetic units would surely eventually find satisfactory physical definitions that would ground their theory (see Pisoni, 1997). But data have continued to show the impossibility of satisfying linearity or invariance of phonetic symbols for describing speech—in English or any other language.

Our perceptual experience of speech resembles articulation much more closely than acoustics (Liberman et al., 1968), but the context-sensitive spectrotemporal cues must still be employed for pattern identity. The physical sound alone is quite sufficient to support perception of consonants and vowels that are invariant across contexts, so, again, rich and detailed aspects of all the contexts must be available for the task of stop categorization.

4. Speech timing based on letter-size segments. If words are fully specified by their transcription, then phones and phonemes (i.e., letterlike representations of speech) should support timing relations that are expressible simply in terms of the number and serial order of segments. (Thus a language might use both /ata/ and /atta/ to employ a "length" distinction.) Yet, many other subtleties of speech timing are exploited in languages of the world (e.g., Hawkins & Nguyen, 2004; Klatt, 1976; Lisker & Abramson, 1971; Port & Leary, 2005). Studies

of speech acoustics and speech perception converge on the view that speakers produce consistent subtle variations in the timing of speech gestures and listeners make use of much temporal detail for speech perception (Hawkins & Nguyen, 2004; Klatt, 1976; van Gelder & Port, 1995). So, from the viewpoint of what speakers listen to and what they control in speech production, any alphabetical representation will be so impoverished that it cannot support either real-life speech perception or expressive and fluent speech production.

5. Abstract linguistic memory. The hypothesis that words are represented in memory using an abstract, segmented alphabet also predicts that memories for linguistic material should be speaker independent and lack any timing detail. Thus, when a speaker, even an illiterate one, listens to an auditorily presented word list, the speech should be remembered in an abstract linguistic code using something that is roughly isomorphic with a transcription in some phonetic alphabet. Of course, speaker-specific information, regarding voice quality or the speaker's age and sex, might be stored as well, but speaker idiosyncrasies could not be part of the word representation itself. Thus, speaker properties could only have an association with the linguistic representation of the words. But in recognition memory experiments using lists of spoken words, it has been shown that listeners are better at recognizing that a word is repeated when the voice is identical in both occurrences than if the voice is different (e.g., Goldinger, 1996; Palmeri, Goldinger, & Pisoni, 1993). One possible interpretation is that a fairly detailed auditory representation is routinely stored, so with more features that match between details of the new utterance and the detailed record of the earlier one, the same-voice repetition is more likely to be noticed.

But also, on the traditional view, we should be able to make it more difficult to associate the voice with the word if the associations were confused by the use of many different voices instead of just 2. If participants tried to remember the voice to improve recognition of the word, then they should be able to do that much better with only 2 voices than with 20. Indeed, to whatever extent the participants might be guessing their answer, the 2-voice condition should exhibit higher scores than the 20-voice condition. Yet the performance benefit for the same voice repetition is the same no matter how many voices were used. This surprising result implies that speakers must routinely store auditory speech material in some rich and detailed form that includes speaker-specific properties. This result is compatible with something along the lines of a rich "exemplar" memory for which many experiments on vision have found support (Nosofsky, 1986).

6. Speech errors. Another prediction is that when speech errors occur, they should show evidence that the segment (or a vector of segmental distinctive features) is the unit that leads to the error. In fact, it seems the speech error distributions should resemble the distribution of typing errors because typing employs discrete context-invariant serial tokens as well. Thus, we should ex-

pect errors like (a) reversals of immediately adjacent segments (e.g., /tɛtn/ for intended /tɛnt/) rather than switches between distant phones or phonemes (as in "glean and pright" for "clean and bright," but the latter is a far more common type of error than switching adjacent segments in speech). We should find (b) the complete replacement of one phone by another phone, (c) no evidence that phonotactic constraints of the language influence the errors (because the segments are supposedly the confusable units), and (d) no evidence of attempts by speakers to simultaneously produce several competing gesture components of the phones. In fact, order reversals of adjacent phones are extremely rare and there is recent evidence from speech kinematics that speakers often attempt to produce two different gestures at the same time suggesting that the independent motoric units of speech are time-extended gestures rather than segments (Browman & Goldstein, 1992; Goldstein, Pouplier, Chen, Saltzman, & Byrd, 2007; Shattuck-Hufnagel & Klatt, 1979). Because the traditional speech error databases relied on auditory (or "impressionistic") phonetic transcription, many details of articulation were overlooked. Thus, altogether speech errors do not appear to provide much support for segments being the units of speech production as opposed to continuous-time phonetic gestures produced in parallel.

Notice that evidence relevant to these six predictions consistently supports greater richness (relative to an alphabet-based notation) of whatever representational methods speakers employ to remember language. Speech is not stored in memory in an abstract discrete code; it is stored using as much detail and richness as the speaker can achieve. Rich temporal and spectral detail are apparently what is required to achieve humanlike speech perception. Analogously, for talkers to achieve expressive natural performance, fine-grained continuous control of the vocal apparatus is clearly what is demanded.

Many of these arguments have been elaborated elsewhere (Port, 2007, 2009; Port, Cummins, & McAuley, 1995; Port & Leary 2005). But in all six of these expectations, the data fail to support the traditional view. Indeed, one fails to find any cognitive role for abstract, segment-size units that are invariant across speakers, across speaking rates, and across variation in neighboring contexts. The strong likelihood that speech relies on some relatively rich and detailed representational scheme pushes all conventional linguistic representations based on letterlike units out of any possible role in real-time speech performance. Consonant and vowel-size units appear to be completely irrelevant to both speech production and perception. But, of course, they are highly relevant for representing aspects of spoken language in an efficient graphical form.

Where Did We Go Wrong?

Yet, of course, this conclusion is very difficult to accept for modern psychologists and linguists. We have all resisted drawing this inference for well over a half

century now. Why is the inference that there is no real-time psychological role for phones or phonemes so difficult to entertain? One reason that has been suggested is why would speakers ignore the availability of an efficient, low-bit-rate code like the alphabet when it seems so obvious and readily available? The answer to this is, of course, that an alphabet-like representation is not, in fact, readily available—not until literacy skills are developed (Rayner, Foorman, Perfetti, Pesetsky, & Seidenberg, 2001; Ziegler & Goswami, 2005). It may seem quite transparent and obvious to us that "tap" has three "speech sounds" and that "trap" has four, or that the "same vowel sound" occurs in "bat" as in "ban," or that "spot" is pronounced with /p/, not a /b/. However, this way of listening to and thinking about speech sounds is not obvious at all to most 6-year-olds or to those with no alphabet training (Morais, Cary, Alegria, & Bertelson, 1979; Read, Zhang, Nie, & Ding, 1986). The most important reason we have not drawn the inference that segments are irrelevant is simply that our intuitions about language overwhelmingly testify that a word consists of letterlike parts. But word performances are not composed from letterlike units—neither acoustically, articulatorily, nor in memory.

Let us examine a transcription more carefully. A written word like "tomato" is usually pronounced in my dialect using the International Phonetic Association alphabet as approximately

$$[t^{h}əme^{I}ɾo^{U}].$$

That is, the initial [t] has an aspiration interval after the /t/ closure (suggested by the [h] superscript), the second and third vowel are diphthongs (defined by their movement in the vowel space as suggested by the superscript vowel symbols), and the second orthographic T is almost invariably pronounced as a tap or flap before an unstressed vowel (spelled phonetically as [ɾ] here) in my dialect. Notice that this transcription using superscripts preserves the same six segments as the orthographic spelling, but it expands the symbol inventory with additional letters so that single letters (with diacritic superscripts) can represent articulatory motions and multiple states. This way it suggests some of the dynamical gestures speakers make when pronouncing the word. Although many different pronunciations are possible for this orthographic word, they can all be approximated (we hope) by using some configuration of an expanded phonetic alphabet. Both the orthographic representation and the technical phonetic alphabet representation seem highly intuitive and at least approximately correct. But because the evidence in the previous section strongly suggests speakers do not use such a representation, could there be something that biases us to insist on using letters as the basic components of words? The answer is yes.

It may be difficult for us to recall, but every person reading this page spent several hours a week for many years learning to read and to refine their reading

skills. It is surely naïve to imagine that all this mental effort focused on use of an alphabet over several decades would have no consequences for our intuitions about the structure of language. Yet we linguists have never paid much attention to the possibility of biased intuitions in our interpretations of speech. It seems likely we tended to overlook this potential problem due to our lifelong training using the alphabet in our orthography. Only a few linguists have considered that our orthographic alphabet skills might, in fact, tend to dominate our linguistic understanding of speech (see Faber, 1992; Linell, 2005; Öhman, 2002). Chomsky (1965), on the other hand, insisted that we should trust our intuitions completely on matters of grammar and phonology and that doing linguistics is primarily a matter of explicating our intuitions about linguistic structure and finding formal notations for them. It is very easy for us to think about language using an alphabet and it is very difficult to think about the sounds of speech in any terms other than alphabetical ones.

But alphabetical writing is a technology that achieved roughly its modern form only about 3,000 years ago. The earliest Greek alphabet might have been created by a Greek with a Phoenician education who wanted to apply Phoenician writing techniques to Greek (or perhaps by an educated Phoenician who also spoke Greek). The Phoenician alphabet itself was the culmination of 4–5 thousand years of earlier record-keeping technologies in the Middle East that were gradually getting easy enough to learn and convenient enough to use (Fischer, 2005). One major consequence of the development of literacy in the Middle East was the growth of the institution of school for teaching literacy to children. Schools have been a gradually increasing part of life in literate nations ever since (Olson, 1994). Parents in some communities begin teaching literacy to children as young as 2 by playing with alphabet blocks.

Alphabetical writing is certainly very useful, but letters are artifacts. It is difficult for children to learn to interpret letter sequences as syllables and to write down syllables as letter sequences, so we start teaching our children as young as possible. Of course, all of us who have had years of reading experience find the interpretation of continuous speech as a sequence of letterlike units to be trivially easy and completely natural. One might even ask, how could we expect to hear speech in any other way given all the practice we have had with letters?

Does this mean that phones and phonemes are illusions? Not at all. It just means they are not necessary participants in the real-time processing of language, either on input or output. What is illusory here is our conviction that phones and phonemes must play a functional role in any real-time cognition involving language simply because our conscious thoughts about language are vividly letter based. Phoneme-like sound patterns do, at least, play a functional role for the community and probably for individual speakers. They are regularities or symmetries in the speech corpus of a community that speakers learn to reproduce in their own speech (Newport & Aslin, 2004). How these approximations to

a low-bit-rate code provide benefit to the community and how they benefit individual speakers are important questions requiring further research (although there are various attempts in print to spell out the benefits of phonology and grammar, e.g., Studdert-Kennedy, 2003).

LANGUAGE AS CULTURE

My proposal is that we need to revise our thinking about language from the bottom up. Our mistake has been to assume that to study language is to study a form of "knowledge," to study the abstract representations in memory that make speech possible. Instead, a language is a kind of social institution, that is, a partially structured system of conventions created by a community of speakers and refined over generations. It is a technology developed by a community for coordination of behavior. Thus it is inherently distributed over space and time and represented differently in its real-time behavioral details in the brain of each speaker. Other institutional technologies include religious practices, an orthographic system, an educational system, and a system of community governance plus food-producing technologies like hunting, farming, fishing, and so on, with all of their accumulated knowledge, tools, manual skills, and social conventions. All of these systems and more comprise the culture of a community. All exhibit ratchetlike accumulation of knowledge and skills over generations that facilitate survival and reproduction in the ecological niche of the community (Richerson & Boyd, 2005; Tomasello, 1999).

A human community can be viewed as a complex adaptive system (Holland, 1995). Such a system supports the emergence of patterns and structures (i.e., symmetries of pattern) of many kinds on many spatial and temporal scales. We should think of a language, then, as one part of the culture of a community. It includes a socially maintained inventory of continuous-time speech fragments that facilitate coordinated action, as proposed by Arbib (2005). Some of the fragments of English include [bə, də, gə] (as in "between, deposit, garage"), [baɪ] ("buy"), [babaɪ sijə ɾəmarə] ("Bye-bye, see you tomorrow"), and so on. Note that minimal segments must have some extension in continuous time. The concept of discrete-time tokens with no temporal extent (exhibited by letters and supposedly exhibited by phones and phonemes) is a technical invention, a cognitive blend (Fauconnier & Turner, 2003) merging aspects of speech sound with aspects of letters of the alphabet. Learning to read is what drilled this blend into us. Although I have described the English fragments here using two different alphabets, no alphabetical description is adequate. The relevant fragments are actually learned as continuous-time patterns of sound, articulation, and somatosensation by speakers. These patterns tend to exhibit many symmetries. That is, some articulatory and auditory properties of one fragment resemble

properties of other fragments (i.e., "park" resembles "ark" and "Parr," which resemble "are" and "Pa," etc.). These overlaps often create the impression (if one were motivated to develop the shortest possible list) that each linguistic usage is composed from a set of nested units that linguists loosely call phones or phonemes. But, in fact, the set of components can never be fixed and can never be timeless. Speakers have much richer memories for language than are implied by any alphabetical descriptions (whether orthographic or phonetic). And speakers have production skills that allow control over an almost unlimited set of subtle articulatory properties. Indeed, most speakers can produce some utterances in other dialects or imitate some pronunciations in foreign languages. Because speakers remember considerable rich detail about speech, alphabets are clearly inadequate for just about any purpose (except for writing, of course, which is very useful if you can get the necessary training).

Homo sapiens seems to have found ways to allow human communities to specialize and to develop complex cultures that create a vast range of different technologies appropriate to the environmental niche of each community. Clearly, intense human sociability contributed to the development of language (Tomasello, 1999) along with the ability to learn statistical regularities after presentation of unfamiliar complex patterns (Newport & Aslin, 2004). Human communities are complex systems capable of adapting their culture over generations (e.g., Beckner et al., 2009; Hruschka et al., 2009; Port, 2007; Richerson & Boyd, 2005; Schoenemann, 1999; K. Smith, Brighton, & Kirby, 2003; K. Smith & Kirby, 2008). These social systems emerge in communities of many agents who interact with each other in complex ways.

Does this story imply then that, contrary to the standard view in linguistics (Chomsky, 1965; Pinker, 1994), humans have no specialized physiological or neural adaptations for language? After all, I am claiming that actual languages are emergent structures created by human communities, not by individual human brains. But surely there are many evolutionary adaptations that were necessary for the rapid acquisition and fluent use of language. However, those specializations should not be expected to look specifically linguistic. That is, they will not include any specific phonemes or distinctive phonetic features; nor will they include grammatical parts of speech or specifications for what constitutes a sentence. Instead, we should expect the specializations to relate to the perceptual and motor skills necessary for speech and to our proclivity to assign almost any aspect of our experience to one or another category depending on the conventions of our community. They are probably also related to the social skills of human infants and adults, such as the ability of human infants to direct attention to some object or event also attended to by their caregiver—what Tomasello (1999) calls "joint intentionality." These behavioral, neural, physiological, and anatomical specializations were presumably selected for over the past half million years or so and give us brains and bodies suitable for the emergence of practical

and effective human languages. Surely, the languages that emerge among Homo sapiens can only be those that are compatible with the neural, physical, and physiological properties exhibited by modern humans (as argued by Christiansen & Chater, 2008).

Some Cultural Categories: Toward a New Linguistics

One important aspect of language as an inventory of conventions for behavior coordination is categorization. It is important not to confuse categories with formal tokens as linguists and others have done for long time. Formal tokens, like the letters [a-z,A-Z] and digits [0-9], are discretely different from each other and can be assumed to be recognized and produced with near-perfect accuracy (Haugeland, 1985) although they are a technology invented only a few millennia ago. The assumption that spoken language is constructed from such tokens has led to the speculation that spoken language might also be a formal system (Chomsky & Miller, 1963). A category, on the other hand, is a psychological grouping that is learned from one's community—a set of things considered the same by the community—no matter what the reason may be for calling them the same. (I avoid using the term *symbol* in this discussion because it is used in so many different ways as to be very confusing.) Cultural conventions chop up the world into categorical parts, many of which have names or verbal labels. A big part of learning one's culture and language is learning the things and events recognized by the society along with their simple or complex verbal labels (Evans & Levinson, 2009; Heft, 2007; Hodges & Baron, 2007; Sloutsky, 2003).

Apparently there are many cultural categories shaping our behavior that do not have words assigned to them by the popular culture but may have been explicated by some modern academic subculture. Examples of these are the various categories that linguists have discovered and found to be relevant to linguistic behavior in specific languages (e.g., phoneme, ±voice, noun, verb, conjunction, mora, etc.). Academic linguists have sometimes proposed (universal) formal-token status for these linguistic categories (Chomsky, 1965; Chomsky & Halle, 1968), but it is far more plausible that each culture uses somewhat different grammatical and phonological categories (Croft, 2000; Evans & Levinson, 2009). It is important to keep in mind that, as socially supported categories, the specifications will vary from language to language, and, in fine detail, from speaker to speaker (Bybee, 2007). The speech products of people may fall into various linguistic categories, but the speaker usually does not have explicit knowledge of what any of the categories are (unless the speaker has the benefit of literacy). The categories of everyday life, for example, the names of animals, plants, cultural artifacts, social roles, and so on, are conventions whose transmission depends on skilled language use and on social interaction.

Categories may be defined by a rule (e.g., definition of a square); by physical or functional similarity (e.g., wheel, chair, stove, etc.); or by any other means, including a simple list of partially arbitrary members (Glushko, Maglio, Matlock, & Barselou, 2008; Murphy, 2002; E. Smith & Medin, 1981). The regularities of a language are exhibited in the corpus of speech that each child is exposed to (Tomasello, 1999). The speaker adapts to the language regularities in some idiosyncratic way depending on their personal history of exposure to languages, dialects, and various speakers.

Linguistic Categories

Thus far I have tried to show that the representational abilities of speakers should be thought of as including distributions of idiosyncratic multidimensional descriptions of a large set of heard utterances along with many abstracted multidimensional templates. These representations do not resemble their abstract, context-free, static, written form. Fragments of these concrete memories, however, are assigned to categories in many ways by speakers of the language related to what we call semantics (e.g., singular/plural, etc.), phonology (e.g., ±voice, /b, d, g/), and syntax (e.g., grammatical categories). The traditional assumption that phonological patterns could be adequately represented in memory merely in terms of letterlike physical tokens, discretely ordered in time, was in part a description of some gross properties of the language, but, at the same time, it involved a projection of the properties of alphabet technology onto our understanding of human psychological processes.

A special trait of human communities is the tendency of cultures to categorize their world. One of the most important technologies created by human communities is spoken language. A language is shaped over generations with respect to the categorization provided by its "lexicon" and "grammar" as well as to the range of speech gestures and sound categories the speakers of a community employ (Hock & Joseph, 1996). But there will always be inherent uncertainty about exactly what the patterns are. The uncertainty stems from variation in the corpus to which each speaker is exposed and from such basic issues as exactly what community is under study and whose speech is representative of it. Of course, there are speech chunks at a large range of sizes, from syllable pronunciations to speech formulas to entire spoken "texts" (such as prayers or epics; Wray & Perkins, 2000). But there are no universal criteria for how to distinguish word-size pieces from common phrases, from idioms, or from "styles of talking" about things. Our orthography has made partly arbitrary decisions about what counts as a word (thus deserving separation from other words by a space) versus a phrase or sentence, but each speaker must find his or her own way to store memories of linguistic fragments and generate new utterances.

What are some of the structures of language that are created by communities of speakers? Although it is clearly an illustration of "literacy bias," we tend to think, first, of a "lexicon," a list of wordlike chunks. But there are also "grammatical regularities," illustrated by the structure of simple sentences or case markings on nouns and tense markings on verbs. The sentence is a fundamental concept for written language but probably plays a role in the spoken language primarily for speakers with extensive literacy training, such as professors, lawyers, newscasters, and so on. Languages always seem to employ a restricted inventory of gestures and sounds. A problem for linguistics is that our literacy bias leads us to insist that syllables are composed from so-called speech sounds, that is, letter-size units, the consonants and vowels. These hypothetical "sounds" (although they are not really physical sounds at all) are supposed to be just like letters except they occur in real time as sound patterns. But it has been known for 50 years that speech sound is not divisible into countable units that line up with letters in any consistent way (Fant, 1973; Joos, 1948). The phonology of a language provides a continuous series of dynamic gestures that are partly serial and partly simultaneous (Browman & Goldstein, 1992; Goldstein & Fowler, 2003).

The importance of the notion of "restricted inventory" is that words in any specific language or dialect tend to resemble each other quite a bit, such that, for most speech fragments, other fragments can be found in the corpus that are very similar. To illustrate this, notice that sometimes pairs of words differ from each other in very similar ways. For example, the distinctions between the six English words in Set 1 shown here are fairly similar to the differences between the six words of Set 2. (An asterisk marks the forms that are not actual words in my dialect.) The orthographic notation here should be interpreted as standing for or pointing out various continuous speech gestures or their continuous-time acoustic description. As noted earlier, some acoustic differences will be the same between "ban-pan" and "Bill-pill," although the formant transitions from the stops into the vowel (known to be essential bearers of information about stop place of articulation) will be different in the two lists. But the third set, where the stops occur in a different position in the syllable, presents more of a problem. The voicing distinction between [b] and [p], and so on, is manifested quite differently between "ban-pan" (in initial position) as opposed to "lab-lap" (in syllable final position) because the vowels in the words ending in [b,d,g] in Set 3 are quite a bit longer than the vowels in the words ending in [p,t,k] in Set 3 and the aspiration that distinguished "ban" from "pan" occurs nowhere.

ban	pan	Bill	pill	lab	lap
Dan	tan	dill	till	lad	*lat
*gan	can	gill	kill	lag	lack
Set 1		Set 2		Set 3	

But English orthography implicitly claims the same consonantal "speech sounds" are used in all three sets of words, even though research on the physical speech signals (as argued earlier) shows that speakers of the language actually store and make use of some detailed auditory form for all the words. Presumably languages gain some benefit from favoring word specifications involving this kind of partial similarity between syllable-initial and syllable-final contexts because many languages have similar sets of consonants in syllable-initial and syllable-final position.

This is an example of symmetrical patterns created by a community of speakers. These words show many symmetries because, for example, the "voicing property" seems to be the same for all the words in the same column in the aforementioned sets and the place of articulation feature is nearly identical for all stops in the same row. Of course, this is true only sometimes because (a) the contrast between /t/ and /d/ is largely neutralized before an unstressed vowel, as in "budding/butting" (although not completely neutralized in my dialect; cf. Fox & Terbeek, 1977; Port & Crawford, 1989), and (b) the difference between /b,d,g/ and /p,t,k/ is not found after an /s/ (cf. "spot, stow, Scot," but words like "sbot, sdow, sgot" do not occur in English) although the claim of "identity" across contexts may be approximately true in many cases. So the very simple and discrete regularities and the neat serial order suggested by the alphabetical notation in the aforementioned tables apply only to letters or other graphic tokens on paper and not to the sounds of speech for which they are gross oversimplifications. Actual phonologies are only approximately discrete and require far more degrees of freedom than letters do.

In this section, the arguments for viewing spoken language as a rough inventory of conventional speech patterns have been made. The apparent letterlike component parts of words are only approximately the same from context to context. Thus they cannot be the actual components speakers rely on. They are simply categories of different items that are treated as the same by the language even though the degree of identity is highly variable. It is our expectation as literate observers of language that spoken language must employ a small alphabet just as the written language does. This has biased linguists to imagine that all those little acoustic details (e.g., the formant transitions, the small changes in vowel quality, the small differences in timing, etc.) must not really matter for the psychological processing of language. We linguists thought, "Speakers can surely ignore all that messy detail and still produce and understand language. After all, the written language ignores that detail. So we should be able to treat spoken language just like other formal systems." But speakers cannot ignore all that small stuff. We have been deceiving ourselves into thinking we could for over a century now. It is time to look the data squarely in the eye and abandon all low-bit-rate formal models for spoken language if we seek to understand human speech as it is produced and perceived. It is at the communal level,

in the social institution, where the low-dimensional description works—at least approximately.

Literacy

Looking at human history in the long view, human communities probably first developed language roughly 100 thousand years ago (Mithen, 2006). The primary evidence is that this is when evidence of cultural flowering begins. It is when tools made of bone and human campsites began to show evidence of continuous cultural change and specializations for particular ecological niches. Improved language skills would have greatly facilitated the cumulative enrichment of cultures by a "ratchet effect" (Tomasello, 1999). Then much more recently, certain human communities developed a practical and easily taught alphabetic writing system (only 3,000 years ago). The approximately low-dimensional statistical patterns in human speech, as described in the previous section, can be mapped reasonably well onto a short list of alphabet tokens (Harris, 2000; Olson, 1994). Thus useful texts could be created requiring a minimum amount of learning of arbitrary relationships. Over the past 3 millennia, there has been some progress in making teaching literacy more effective, but the basic methods for teaching reading and writing have hardly changed: memorize the suggestive names for letters, then practice reading and writing short common words until the child "gets the hang of it." The alphabet literacy we all share trains us to experience speech as a sequence of letters from a very small, fixed alphabet and to experience discrete words and sentences as well. The fit between the alphabet patterns and the real phonology of a language is good enough that with sufficient training we have been content to overlook all the problems (because we learned to just "get over" the inconsistencies and the arbitrariness of spelling and writing while still very young). Just in the past century, formal mathematical grammars have taken the properties of orthographic letters (discreteness, serial order, etc.) and generalized them to create rich mathematical systems (Harrison, 1978) that led to the development of computer programs and to speculation that language might have a formal symbolic representation in the human brain (Chomsky, 1965; Chomsky & Miller, 1963; Saussure, 1916).

Any community provides its children with exposure to a fraction of the linguistic culture (somewhat different fractions for each speaker, of course, and for any speaker at any particular point in time). The linguistic culture of a community provides a massive set of conventional categorizations of the experience of that community. These include both unnamed categories and ones that have a lexical name or a description that requires only one or a few words to describe it. Young speakers learn "speech chunks" by rote example at first (see Grossberg, 2003; Grossberg & Myers, 2000, for an explicit model that chunks speech), but gradually they come to categorize speech the way their family and

neighbors do (see Werker & Tees, 1984, for results on the earliest stages of phonological acquisition) and learn to control their own body to produce speech that conforms to the conventions of their community. None of this requires that individuals produce or perceive speech using identical descriptors or that they have any conscious knowledge of the categories they implement.

How could it be that a speaker does not need to "know the language" in order to use it? The key idea is that it is the community that undergoes self-organization and creates the language in historical time, not individual agents. The argument is analogous to the famous case of termites that follow simple behavioral biases about where to deposit their little wads of dirt mixed with a pheromone. These biases happen to result in periodically spaced columns of a fixed height with a ceiling created above to support the next layer of columns (Kugler & Turvey, 1987). The termites themselves have no idea what they are doing. They are simply behaving instinctually under guidance by genes as to where to deposit their wads of dirt. Yet regular and periodic physical structures are created for the benefit of the colony by animals with no representations of either periodic columns or layers. For the case of language as well, the community has become the relevant complex adaptive system, not the individual. (Of course, an individual agent may also be a complex adaptive system, but no individual human normally creates a language or a grammar for a language.) The group creates structures (e.g., lexicons, phonologies, rituals, etc.) as conventions. The agents in the system are imitators as well as users of the community speech conventions. It is a serious mistake to insist that these community structures could only exist if they are part of the psychological constitution of the agents.

So it seems highly likely that agents (i.e., speakers) differ from each other in the details of their representations (because there is nothing that can enforce identity). Still speakers in the same community will behave in ways that generally accord with the conventions followed by others. What the speaker must learn are appropriate behaviors, both perceptual and motor. The behaviors are, no doubt, tied to the microcircuitry of their nervous system, their vocal-tract anatomy, and their auditory and linguistic experience. The "linguistic units" of the community language (such as the ones we formalize in English orthography and in linguists' grammars) are not something native speakers have any use for in order to speak correctly.

What Is Predicted if Language Is a Social Institution?

If this story is on the right track and linguistic structures are essentially communal not personal, then many predictions seem to follow. The theory predicts at least the following. I believe all of them are supported by evidence, but I point to just one or two references in support of each.

1. Variability between speakers. Because linguistic categories have conventional specifications, individual agents should find alternative ways to implement the linguistic conventions in various situations. Small differences in pronunciation or intended meaning or in the use of grammatical conventions, because they occur in a rich, real-time communicative context, should be easily tolerable. So speaker-to-speaker variation should be large if we look at details (e.g., Bybee, 2007; Docherty & Foulkes, 2000; Labov et al., 2006). This is predicted because categories demand only approximate sameness whereas speakers have continuous control over many parameters of their productions.

2. Within-speaker variation. Any given speaker may implement the conventions differently depending on the context and the speaker's evaluation of the communicative needs of the moment, such as the prevailing signal-noise ratio at each moment of speech (e.g., Lindblom, 1990). They may also gradually change their target pronunciations along some parameters during their lifetime (Harrington, 2006; Sancier & Fowler, 1997). There is nothing to prevent small random variations and nothing to deter gradual shifts in pronunciation targets in ways that reflect the speaker's social situation at the moment of speaking.

3. Temporal patterns. Timing patterns should include many that cannot be well described using serial letterlike units. Durational ratios and periodic patterns, for example, should be possible because these require some form of measurement of continuous time and not mere counting of letterlike segments (e.g., Klatt, 1976; Port, Dalby, & O'Dell, 1987; Port & Leary, 2005). This is predicted because speech memory relies on memory representations that approximate continuous time (not discrete time).

4. Indexical information in speech memory. Linguistic memory should include the speaker's idiosyncratic voice information rather than be abstracted away from the specific utterance (Palmeri et al., 1993; Pisoni, 1997). This is predicted because auditory representations for speech chunks do not have abstract acoustic specifications but are hypothesized to be more concrete and rich in detail than linguists usually imagine. And the word-specifying information cannot be somehow factored from the rest of the auditory properties of speech.

5. Dialect variation and change over time. Social groups that are distributed geographically should exhibit slow change of pronunciation and other usages over time (e.g., Bybee, 2007; Hock & Joseph, 1996; Labov et al., 2006). This is predicted because linguistic features are simply conventional aspects of the culture of a community. Like all other cultural traits, they should exhibit gradual change over time. Variability is tolerated and small changes along various continua may be imitated by others depending on the details of each speaker's communication network and the relative prestige of talkers. People talk primarily with others in a restricted group, so consistency at the local scale is what is most important for efficient language use.

6. Computer speech recognition will not succeed by recovering letterlike segments. Speech recognition engineers should find it difficult to try to identify consonant and vowel segments as a step toward recognizing larger chunks of speech like words and phrases (e.g., Huckvale, 1997; Jelinek, 1988). Words are not composed from simple segments or segmental features. Instead, they are defined by overlapping continuous gestures and acoustic trajectories. Static letterlike segments should be of very limited use for speech recognition (even though native speakers can be trained to use letters for reading and writing).

There seems to be plenty of observational and experimental support for every one of these predictions. On the other hand, we saw earlier that the predictions of the abstract, segment-based notion of speech found almost no supporting evidence.

CONCLUSIONS

It seems that the radical story endorsed here is compatible with a great deal of the experimental data dealing with language performance. It is also compatible with the ideas of the DLG because they emphasize the tight social embedding of linguistic behavior as well as the vast differences between written and spoken language. Because we linguists and psychologists are all literate, we find it difficult to see that our thoughts and intuitions about language have been profoundly shaped by the reading and writing skills we worked so diligently to achieve in our youth. Almost all linguists have followed Saussure (1916) in claiming to be studying spoken language not written language. But the fact is that almost all modern linguists, like Saussure, never really escaped from letter-based characterizations of language. Audio (and video) recordings are rarely found in the linguistics classroom or in most linguistics research. When we think of "words," "speech sounds," and "sentences" in our descriptions of language, we are importing the conventions of our writing system and trying to use them uncritically as hypotheses about psychological representations. Confusion about the distinction between the written language and the spoken language creates difficulties for many types of research (e.g., see Love's 2004 criticisms of Clark, 2001).

What I have tried to show here is that the neat codelike units of our normal conscious thoughts about language—that is, the letterlike "speech sounds," the discrete words, morphemes, prepositional phrases, full sentences, and so on—play at most a very tiny, specialized role in real-time conversational performance. The kind of speech patterns linguists are interested in only exist at the level of the community and reflect a kind of social institution that appears as the cultural environment for each child. Linguistics and the psychology of language should abandon attempts to describe the representations of abstract linguistic

units stored in people's heads—even though this has been the goal of these disciplines for at least a century. The patterns of language that linguists are interested in simply do not exist in the form of abstract, formal, low-bit-rate structures that are common across all speakers. Instead, our every utterance is a creative behavioral response to our experience, potentially shaped by every possible facet of our lives, but also generally compatible with the conventions observable in the speech of others in our community. A language is, first of all, the set of conventions about speech shared by some community. For several millennia some communities have also had a system of orthographic conventions based partly on the phonetic value of letters. Of course, the letters, too, are part of our language, but letters differ in profound ways from speech and have seriously confused us in our attempts for the past century to understand human linguistic capabilities.

ACKNOWLEDGMENTS

I am grateful to Philip Carr, Stephen Cowley, Ken deJong, Carol Fowler, Bert Hodges, Diane Kewley-Port, Tom Schoenemann, and Mark van Dam for helpful comments on this article.

REFERENCES

Arbib, M. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences, 28,* 105–167.

Beckner, C., Blythe, R., Bybee, J., Christiansen, M., Croft, W., Ellis, N., ... Schoenemann, T. (2009). Language is a complex adaptive system: Position paper. *Language Learning, 59*(1), 1–27.

Bolinger, D. (1976). Meaning and memory. *Forum Linguisticum, 1,* 1–14.

Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica, 49,* 155–180.

Bybee, J. (2007). *Frequency of use and the organization of language.* New York, NY: Oxford University Press.

Chomsky, N. (1965). *Aspects of the theory of syntax.* Cambridge, MA: MIT Press.

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

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

Christiansen, M., & Chater, N. (2008). Language as shaped by the brain. *Behavioral and Brain Sciences, 31,* 489–558.

Clark, A. (2001). *Mindware: An introduction to the philosophy of cognitive science.* New York, NY: Oxford University Press.

Croft, W. (2000). Parts of speech as language universals and as language-particular categories. In P. Vogel & B. Comrie (Eds.), *Approaches to the typology of word classes* (pp. 65–102). Berlin, Germany: Mouton de Gruyter.

Docherty, G., & Foulkes, P. (2000). Speaker, speech and knowledge of sounds. In N. Burton-Roberts, P. Carr, & G. Docherty (Eds.), *Phonological knowledge: Conceptual and empirical issues* (pp. 105–127). Oxford, UK: Oxford University Press.

Evans, N., & Levinson, S. (2009). The myth of language universals: Language diversity and its importance for cognitive science. *Brain and Behavioral Sciences, 32,* 429–492.

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, The Netherlands: John Benjamins.

Fant, G. (1973). Descriptive analysis of the acoustic aspects of speech. In G. Fant (Ed.), *Speech sounds and features* (pp. 17–31). Cambridge, MA: MIT Press.

Fauconnier, G., & Turner, M. (2003). *The way we think: Conceptual blending and the mind's hidden complexities.* New York, NY: Basic Books.

Fischer, S. (2005). *A history of writing*. London, UK: Reaktion Books.

Fox, R., & Terbeek, D. (1977). Dental flaps, vowel duration and rule ordering in American English. *Journal of Phonetics, 5,* 27–34.

Glushko, R., Maglio, P., Matlock, T., & Barselou, L. (2008). Categorization in the wild. *Trends in Cognitive Science, 12*(4), 129–135.

Goldberg, A. (2006). *Constructions at work: The nature of generalization in language.* Oxford, UK: Oxford University Press.

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.

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, Germany: Mouton de Gruyter.

Goldstein, L., Pouplier, M., Chen, L., Saltzman, E., & Byrd, D. (2007). Dynamic action units slip in speech production errors. *Cognition, 103,* 386–412.

Grossberg, S. (2003). The resonant dynamics of speech perception. *Journal of Phonetics, 31,* 423–445.

Grossberg, S., & Myers, C. W. (2000). The resonant dynamics of speech perception: Interword integration and duration-dependent backward effects. *Psychological Review, 107,* 735–767.

Harrington, J. (2006). An acoustic analysis of 'happy-tensing' in the Queen's Christmas broadcasts. *Journal of Phonetics, 34,* 439–457.

Harris, R. (1981). *The language myth.* London, UK: Duckworth.

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

Harrison, M. (1978). *Introduction to formal language theory.* Reading, MA: Addison-Wesley.

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

Hawkins, S., & Nguyen, N. (2004). Influence of syllable-final voicing on the acoustic onset of syllable-onset /l/ in English. *Journal of Phonetics, 32,* 199–231.

Heft, H. (2007). The social constitution of perceiver-environment reciprocity. *Ecological Psychology, 19,* 85–105.

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

Hodges, B., & Baron, R. (2007). On making social psychology more ecological and ecological psychology more social. *Ecological Psychology, 19,* 79–84.

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

Hruschka, D., Christiansen, M., Blythe, R., Croft, W., Heggarty, P., Mufwene, S., ... Poplack. (2009). Building social cognitive models of language change. *Trends in Cognitive Science, 13,* 464–469.

Huckvale, M. (1997). *Ten things engineers have discovered about speech recognition.* Paper presented at the ASI Workshop on Speech Pattern Recognition, St. Helier, Jersey, UK, September 1997.

Jelinek, F. (1988). *Applying information theoretic methods: Evaluation of grammar quality*. Paper presented at the Workshop on Evaluation of Natural Language Processing Systems, Wayne, PA, December 1988.

Joos, M. (1948). Acoustic phonetics [Supplement]. *Language Monograph, 24,* 1–136.

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

Kugler, P. N., & Turvey, M. T. (1987). *Information, natural law and the self-assembly of rhythmic movement*. Hillsdale, NJ: Erlbaum.

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

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. *Journal of Experimental Psychology, 54,* 358–368.

Lindblom, B. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W. J. Hardcastle & A. Marchal (Eds.), *Speech production and speech modelling* (pp. 403–439). Amsterdam, The Netherlands: Kluwer.

Linell, P. (2005). *The written language bias in linguistics.* Oxford, UK: Routledge.

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

Love, N. (2004). Cognition and the language myth. *Language Sciences, 26,* 525–544.

Mithen, S. (2006). *The singing Neanderthal: The origins of music, language, mind and the body.* Cambridge, MA: Harvard University Press.

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

Murphy, G. (2002). *The big book of concepts.* Cambridge, MA: MIT Press.

Newport, E. L., & Aslin, R. (2004). Learning at a distance I: Statistical learning of non-adjacent dependencies. *Cognitive Psychology, 48,* 127–162.

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

Öhman, S. (2002). Phonetics in a literal sense: 1. Do the letters of the alphabet have a pronunciation? *Proceedings of Fonetik, TMH-QPSR Quarterly Progress Report, KTH, 44*(1), 125–128.

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

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.

Pinker, S. (1994). *The language instinct: How the mind creates language.* New York, NY: William Morrow.

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, CA: Academic Press.

Port, R. (2007). How words are stored in memory: Beyond phones and phonemes. *New Ideas in Psychology, 25,* 143–170.

Port, R. (2009). The dynamics of language. In R. Meyers (Ed.), *Encyclopedia of complexity and system science* (pp. 2310–2323). Heidelberg, Germany: Springer-Verlag.

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. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 339–371). Cambridge, MA: Bradford Books/MIT Press.

Port, R., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of Acoustical Society of America, 81,* 1574–1585.

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

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.

Richerson, P., & Boyd, R. (2005). *Not by genes alone: How culture transformed human evolution.* Chicago, IL: University of Chicago Press.

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

Sancier, M., & Fowler, C. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25,* 421–436.

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

Schoenemann, P. T. (1999). Syntax as an emergent characteristic of the evolution of complexity. *Minds and Machines, 9,* 309–346.

Shattuck-Hufnagel, S., & Klatt, D. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior, 18,* 41–55.

Sloutsky, V. (2003). The role of similarity in the development of categorization. *Trends in Cognitive Sciences, 7,* 246–251.

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

Smith, K., Brighton, H., & Kirby, S. (2003). Complex systems in language evolution: The cultural emergence of compositional structure. *Advances in Complex Systems, 6,* 537–558.

Smith, K., & Kirby, S. (2008). Cultural evolution: Implications for understanding the human language faculty and its evolution. *Philosophical Transactions of the Royal Society, B, 336,* 3591–3603.

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, MA: Harvard University Press.

van Gelder, T., & Port, R. (1995). It's about time. In R. Port & T. van Gelder (Eds.), *Mind as motion: Explorations in the dynamics of cognition* (pp. 1–44). Cambridge, MA: MIT Press.

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.

Wray, A., & Perkins, M. R. (2000). The functions of formulaic language: An integrated model. *Language and Communication, 20,* 1–28.

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