Phonology is not psychological and speech processing is not linguistic*

Robert Port

Indiana University

2/28/07

ABSTRACT

Experimental evidence about human speech processing and linguistic memory shows that words are not spelled from letter-like units, whether thought of as phones or phonemes. Linguists, like others with a Western education and a lifetime of literacy, identify speech quite automatically with a sequence of letter-sized units. Consonants and vowels seem like directly observable units in speech. However, the evidence is clear that we actually use high-dimensional spectro-temporal (i.e., auditory) patterns to support speech production, perception and memory in real time. Thus abstract phonology (with its phonemes, syllable types, distinctive features, etc.) needs to be re-conceived as a social institution – a system of patterns that evolves over historical time in some community playing almost no role in practical speech processing. There are really two sciences of language: First there is real-time **Speech-Language Processing**, the rich code used to store and manipulate information about linguistic patterns. Second is a proposed new **Phonology** that characterizes patterns in the corpus of a linguistic community (but makes no claims about psychological processing). This distinction cuts orthogonally across Chomsky's Competence vs. Performance since each has some properties of competence and some properties of performance. Experimental results force us to differentiate the language as a social institution (thus not necessarily `present' in each speaker's brain) from the actual psychological processing as we talk to each other. Chomsky thought speakers must ``know'' their language, but they do not because ``the language'' is an abstract, speaker-independent system that exists only at the level of the community. Speakers exhibit or perform a language, but do not need to ``know'' it.

Keywords: phonology, phonetics, exemplar memory, competence, performance

1. INTRODUCTION

For at least a century, linguists have trusted their intuitions that phonetic segments – consonants and vowels – are directly observable in speech. We assumed that segments can be taken as valid raw data for the empirical study of language (Saussure, 1916; Jones, 1918, p. 1; Ladefoged, 1972; Chomsky & Halle, 1968; IPA, 1999) and assumed they provide the psychological spelling system for every language. But these powerful intuitions are largely a result of the lifelong literacy training to which all readers of this paragraph have been subjected (Faber, 1992). The mass of experimental evidence over the past half century actually supports a very rich memory for language, resembling our memory for everyday events and activities which is detailed and context-specific. Humans are certainly capable of abstract generalizations, but memory does not depend on them. (See also Nosofsky, 1986 and Shiffrin and Steyvers, 1997 for exemplar-based models of list memory and categorization.) Almost no evidence supports a memory for language that uses abstract, speaker-independent tokens arrayed in serial order. But the intuition that we process language using abstract phones and phonemes strongly prevails in the field (Bloomfield, Chomsky, Liberman) despite the lack of evidence for them.

The goal of this paper is to marshal some of the empirical literature that is incompatible with any realtime psychological role for phones, phonemes or segmental distinctive features. 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 we linguists call ``phonological structure'' actually exists only as a set of generalizations across a speech corpus. These patterns are shaped over many generations of speakers. For various reasons, the phonological systems of language tend toward a

* Manuscript submitted to the *Society for Philosophy and Psychology* meeting in Toronto, Canada, June 2007.

low-dimensional (i.e., symbol-like) format. But speakers do not `know' their language using a code like this. They employ a far richer and more concrete representation of speech for storing linguistic fragments and chunks.

After making this case, the paper will argue that this conclusion about speech processing implies a division of labor between linguistics and psychology that is orthogonal to Chomsky's distinction of Competence (the mental and the linguistic aspects of speech data) and Performance (the physiological and physical aspects of speech). The revision proposed separates real-time Speech-Language Processing from a new Phonology that seeks description of the cross-speaker generalizations – the information that an adult learner of a second language needs to be taught. The implications of this new view are very broad since it suggests that all symbols and symbolic processing may be cultural technologies that are based on experience and skill using alphabetical orthographies and arithmetic symbols on paper.

2. PREDICTION AND EVIDENCE

If words were stored using abstract discrete tokens like vowels and consonants, as claimed by Chomsky and most phonologists, then many predictions should follow. For example, both synchronic variation and diachronic sound change should always exhibit discrete phonetic jumps as a feature changes or one segment type is replaced with another. And the context for any contextual variant would always be specifiable with just these discrete units. Each distinctive feature should have a single invariant definition for all contexts. And there should be no temporal effects observed that cannot be described in segmental terms since segments allow only serial order for representing time. Furthermore, our memory for specific utterances should exhibit evidence of descriptions that are invariant across context, across speakers and across speaking rates. But *none* of these expectations hold true.

Linguistic Variation. The traditional theory surely predicts that abstract, canonical representations of each word are used by speakers for remembering what someone said (after all there is no other representation offered by linguistics). But Labov, Bybee and other variationists, along with generations of literature from experimental phonetics (e.g., Peterson & Lehiste, 1960; Lisker and Abramson, 1964, 1967; Local, 2001), have shown that the variety of pronunciations for any word that speakers are sensitive to is staggering. Most of us are familiar with many subtle regional accents and foreign-accented pronunciations in our native language. It seems all speakers vary their own pronunciations along many phonetic continua depending on various subtleties of social and pragmatic context. How could these minute phonetic differences be employed in production and perception if we store language using a very low-dimensional phonological representation based on a small number of distinctive phonetic features? There is no end to the variety of pronunciations speakers must deal with. Since speakers recognize many such minute phonetic and temporal differences and can control many of them in their own speech, they must have a way of representing them, and thus clearly require richly detailed phonetic memories for speech, not abstract, supposedly ``economical'' ones using a narrowly restricted list of letter-like units.

Recognition Memory. A more direct source of evidence is found in recognition memory experiments. Linguistic theory claims that words are remembered in terms of abstract, serially ordered spellings using a small number of phonological units. Thus, if we hear someone say *tomato,* then what is supposedly stored and available to support later cognition should be a canonical phonological spelling of *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 are not part of this representation and should not be stored (Chomsky and Halle, 1968; Halle, 1985; Pisoni, 1997). The argument from modern linguistic theory is indeed that listeners do not need to store that information. The abstract representation is assumed to be more efficient than one using large amounts of information that is often linguistically irrelevant. But speaker identity and timing patterns do influence performance in recognition memory tasks (Goldinger, 1996; Palmeri, et al, 1993; Pisoni, 1997). For example, if a subject hears a long list of spoken words and is asked to indicate when a word is repeated in the list, accuracy declines (of course) the greater the amount of time between the first presentation and the second. But if the list is pronounced by many voices that change randomly, then words repeated in the same voice are recognized almost 10% more accurately than when the

repetition is in a different voice. Even more surprisingly, the improvement is exactly the same whether 2 voices read the list or 20 voices read the list. And some improvement lasts for up to a week. The unavoidable implication of results like these is that speakers automatically store much richer and more detailed representations than linguists ever imagined. Of course, speakers might store abstract representations as well, but evidently do not rely on them.

Speech Perception. Many well-known phenomena of speech perception are quite incompatible with abstract, segmental representations but fit well with a view of word storage that is highly concrete and detailed as proposed in ``exemplar models of memory'' (Hintzman, 1986; Pierrehumbert, 2001; K. Johnson, 2007). For example, ``coarticulation'' is the formidable problem of how people hear speech as consisting of nonoverlapping, context-free segments when the auditory stimulus exhibits a great deal of temporal overlap and context-sensitive variation that seems to be inaudible (Liberman, et al., 1968; Kent and Minifie, 1977). But the problem exists only for our conscious experience of language. Research on the phonological awareness of illiterates has shown that our segmental intuitions are found only in those who have had alphabet training (Morais, et al, 1979; Rayner, et al, 2001). So the entire coarticulation problem disappears if speakers employ a rich and detailed auditory memory. Thus [di] and [du] do not share any unit in memory, *contra* Liberman, et al., 1968. Linguistic memory does not extract an abstract, context-free invariant for each nonoverlapping consonant and vowel until one has had literacy training. Alphabet-literate people consciously describe speech to themselves using a phonological or orthographic code, but for realtime tasks they rely on a rich auditory memory.

Similarly, short-term verbal memory or the so-called ``phonological loop'' (Baddeley, 1986) is not phonological at all in the sense that linguists use this term. It is evidently a store based on a motor code, something very concrete and sensitive to context, and nothing like the abstract, static code that structural linguistics predicts (Wilson, 2001).

All these phenomena imply that the realtime psychological processing of language does not rely on the abstract, low-dimensional non-overlapping descriptions proposed by traditional linguistics (and as suggested by our intuitions). They imply a speech memory that stores large amounts of redundant speech material (rather than representations with a single form for each lexical item) coded by a very rich auditory or articulatory code (rather than one that is maximally efficient for representation with the kind of ordered graphic tokens we can write on paper). 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 people remember all they possibly can about details of specific utterances. Abstractions and generalizations can be easily extracted as needed from a memory containing only concrete instances. (See Hintzman, 1986, who models each episode as a long vector of all co-occurring features and models long-term memory as a matrix of such exemplar vectors. Retrieval from memory depends on the similarity of the probe to features in the stored vectors.) An exemplar-based memory system is also capable of creativity when needed despite its use of a richly detailed representational code. If this proposal for rich linguistic memory seems radical and implausible, consider that exemplar theories of memory are now widely accepted in experimental psychology (Nosofsky, 1986; Shiffrin and Steyvers, 1997) and the human ability to remember randomly collocated events on a single exposure (such as your `episodic memory' for events that happened to you earlier today) is well acknowledged (Goldinger, 1996; Reilly and Norman, 2005). It seems that linguistic theory has been predicated on unrealistically restrictive assumptions about the capabilities of human memory.

3. A NEW PHONOLOGY

If words are stored in this detailed and speaker-specific way, then do we still need linguistic descriptions that employ a low-dimensional alphabet? Yes, we do. Phonological patterns – the segment types, distinctive features, syllable types, etc., of each language – are obvious in the corpus of any speech community. The linguist looks across a large set of utterances and, using whatever descriptive tools are available (such as the IPA phonetic alphabet, sound spectrograms, etc.), attempts to describe the patterns found there. This corpus is, of course, an approximation to the ambient language as it presents itself to the language learner. Some of these patterns (e.g., many traditional phonological phenomena involving phonemes, features, syllables, etc.)

can be described using an alphabet of phonetic symbols. But many other patterns require more careful measurements of frequency-by-time trajectories (e.g., voice-onset time, mora patterns, formant transitions, spectrum shapes, intonation contours, etc.). Both kinds of descriptive tools aim to (a) capture the properties that are shared across the speech community, (b) represent distinctions that are exploited by most speakers of the language to differentiate various sets of lexical items, and (c) describe the properties that differentiate the phonology of one language from another. These descriptions provide essential resources for teaching a language to speakers of another language and they provide a practical basis for development or refinement of an orthography. But a phonological description should not be expected to play much of a role in the realtime perception and production by skilled speakers.

If speech memory is so much richer than we thought, then why, one might ask, do small alphabets work as well as they do for recording language? First of all, alphabets do not do as well as most people think since they leave out, for example, all the temporal characteristics and much more (see Port & Leary, 2005). They are only effective for people who already speak the language. But the question is still worth answering. Apparently, a language as a social institution is shaped by generations of users toward a system that restricts the phonetic degrees of freedom in many ways. It is possible to describe much about the corpus using a small set of distinct units – enough to support literacy for fluent speakers. Employing a restricted set of patterns in the phonetic space probably constitutes an attractor for a speech community. When the system as a whole approaches a low-dimensional description, it will typically exhibit many local attractors (e.g., the phonemes, distinctive features, limited syllable types, distinctive intonation contours, etc.). Presumably 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. But this does not imply that the apparent units, the `symboloids' of this corpus of speech, are discrete cognitive units in our representation of language in memory. The phonological patterns and the units of a memory code exist on very different descriptive levels and develop on very different time scales.

The distinction promoted here between Phonology and Speech-Language Processing seems to have much in common with Chomsky's (1965) distinction of Linguistic Competence and Linguistic Performance. But there are important differences. For Chomsky, Competence is the formal core of language. It is purely discrete and mental, the system that exhibits linguistic creativity. So beginning at its periphery, Competence begins for the speaker-hearer with the phonetic alphabet. The listener hears speech which is automatically converted somehow 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 plays the role of a keyboard on which the mind (i.e., Linguistic Competence) plays the body in discrete time. Linguistics is concerned only with the formal aspects of language lying between the symbolic phonetic code as input and as output. Performance, however, 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, from the standpoint of the linguist, serve merely as noise obscuring the formal structure of language.

	(new) Phonology	**Speech-Language Processing**
Competence	- abstract letter-like representations - invariant across speakers, rates, etc. - approximate low-dimensional code - slow change over historical time	- memory representations of words - create new combinations
Performance	- high-dimensional phonetic description - temporal patterns	- speech perception - high-dimensional phonetics - real-time processing - motor limitations - changes rapidly in time

But, as shown in the table above, the distinction proposed here is orthogonal to Chomsky's distinction since parts of both my proposed Phonology and Speech-Language Processing would be characterized as Competence and parts as Performance. Thus, my social Phonology can be approximately

described using low-dimensional, abstract, speaker-invariant representations that can include phoneme-like units, but it will also include temporal patterns and high-dimensional phonetic details when appropriate. Similarly, my proposed Speech-Language Processing incorporates memory for chunks of linguistic material like words and phrases and is also the system that creates new words or fresh utterances appropriate for the speaker's communicative needs. That sounds like Competence, but Speech-Language Processing also does Performance-like tasks such as speech perception using very high-dimensional descriptions of speech and only functions in real, continuous time, not discrete time.

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 speakers as a social institution – from the actual skills the 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 reveals a serious error in our thinking and theorizing about language (and, more generally, our thinking about symbols). The mistake stems from trusting our intuitions when we should have been more skeptical. The powerful intuition we have relied upon for a century that language is structured in terms of discrete letter-like tokens is really just one side effect of our years of literacy education – one that has been overlooked. We taught ourselves (with initial help from a teacher) to listen to and think about language in segmental terms probably because this ability is important for the skillful use of our orthography. Consequently we phoneticians, linguists, psychologists and philosophers were all quite sure that the ``real'' structure of language had to be discrete and segmental, despite all the contrary evidence in front of us at least since the appearance of the sound spectrograph at mid-century (Joos, 1948). The discipline of phonology cannot (as linguists had hoped from Saussure to Chomsky & Halle to Prince & Smolensky, 1993) offer both a description of the psychological code for linguistic memory and also a description of our language that is suitable for writing it down and for teaching to adult learners. What the native speaker needs to know about a language is vastly more concrete and detailed than we thought. And phonological generalizations – the patterns that are shared by the speech of an entire community – are patterns that exist as statistical regularities only at the level of the corpus of that community. It turns out that what we have been loosely calling Linguistic Cognition depends both on the individual's psycholinguistic skills exhibited in real time and also on a linguistic social institution consisting in part of phonological regularities and patterns that are shared across the community. Phones and phonemes are not empirical phenomena, directly observable in the data of language. They are interpretations of speech made only by those with years of experience with alphabetical writing.

5. REFERENCES

[1] Baddeley, A. D. (1986). *Working Memory*. Oxford, U. K.: Oxford University Press.

[2] Bybee, Joan (2001). *Phonology and Language Use*. Cambridge, UK Cambridge University Press.

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

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

[5] Faber, A. (1992) Phonemic segmentation as epiphenomenon: Evidence from the history of alphabetic writing. In Downing, P., Lima, S. and Noonan, M. (Eds.) *The Linguistics of Literacy.* Amsterdam, John Benjamins.

[6] 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.

[7] Halle, M. (1985). Speculations about the representation of words in memory. In V. Fromkin (Ed.), *Phonetic Linguistics: Essays in Honor of Peter Ladefoged* (pp. 101-114). Orlando, Florida: Academic Press.

[8] Hintzman, D. L. (1986). `Schema abstraction' in a multiple-trace memory model. *Psychological Review, 93*, 411-428.

[9] IPA. (1999). *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*. Cambridge, England: Cambridge University Press.

[10] Johnson, Keith (2007) Speech perception without speaker normalization: An exemplar model. In Johnson, K. and J. Mullenix (Eds.) *Talker Variability in Speech Processing*. London, Academic Press.

[11] Jones, D. (1918). *An Outline of English Phonetics*. Leipzig, Germany: Teubner.

[12] Kent, R. and Minifie, F. (1977) Coarticulation in recent speech production models. *Journal of Phonetics 5*, 115-135.

[13] Labov, W. (1963). The social motivation of a sound change. *Word, 19*, 273-309.
[14] Ladefoged, P. (1972). *A Course in Phonetics*. Orlando, Florida: Harcourt Brace Jovanovich.
[15] Liberman, A. M., Delattre, P., Gerstman, L. & Cooper, F. (1968). Perception of the speech code. *Psychological Review 7*4, 431-461.
[16] Lisker, L. & Abramson, A. (1964) A cross-language study of voicing in initial stops: Acoustical measurements. *Word 20*, 384-422.
[17] Lisker, L. & Abramson, A. (1967) Some effects of context on voice-onset time in English stops. *Language and Speech 10*, 1-28.
[18] Local, J. (2003) Variable domains and variable relevance: Interpreting phonetic exponents. *Journal of Phonetics 31*, 321-339.
[19] Nosofsky, R. (1986). Attention, similarity and the identification-categorization relationship. *Journal of Experimental Psychology: General, 115*, 39-57.
[20] Morais, J., Cary, L., Alegria, J., & Bertelson, P. (1979). Does awareness of speech as a sequence of phones arise spontaneously? *Cognition, 7*, 323-331
[21] O'Reilly, Randall and Kenneth Norman (2005) Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in Cognitive Sciences 6*, 505-510.
[22] 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.
[23] Peterson, G. & Barney, H. (1952) Control methods used in a study of the vowels. *Journal of Acoustical Society of America. 24*, 175-184.
[24] Pierrehumbert, J. (2003) Phonetic diversity, statistical learning and acquisition of phonology. *Language and Speech 26*, 115-154.
[25] Pisoni, D. (1997) Some thoughts on `normalization' in speech perception. In Johnson, K. and Mullenix, J. (Eds.) *Talker Variability in Speech Processing*. London, Academic Press.
[26] Port, R. F., & Leary, A. (2005). Against formal phonology. *Language, 81*, 927-964.
[27] Port, R. (2007, in press) How words are represented in memory: Beyond phonetics and phonology. *New Ideas in Psychology* (Elsevier)
[28] Prince, A. & Smolensky, P. (1993) *Optimality Theory: Constraint Interaction in Generative Grammar.* New Brunswick, New Jersey, Rutgers University Center for Cognitive Science.
[29] 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
[30] Saussure, F. d. (1916). *Course in General Linguistics* (W. Baskin, Trans.). New York: Philosophical Library.
[31] Shiffrin, R., & Steyvers, M. (1997). A model for recognition memory: REM: Retrieving effectively from memory. *Psychonomic Bulletin and Review, 4*, 145-166.
[32] Wilson, M. (2001). The case for sensorimotor coding in working memory. *Psychonomic Bulletin and Review, 8*, 44-57.