UNIVERSITY OF MURCIA

International Journal of English Studies

IJES

www.um.es/ijes

Is the Phoneme Usage-Based? – Some Issues[1]

GEOFFREY S. NATHAN[*]
Wayne State University

ABSTRACT

After a brief review of the history of the phoneme, from its origins in the nineteenth century to Optimality Theory, including some Cognitive Linguists' views of the concept. I argue that current 'usage-based' theorists views of the phoneme may not be able to explain some facts about how naïve speakers process language, both consciously and subconsciously. These facts include the invention of and worldwide preference for alphabetic writing systems, and language processing evidence provided by Spoonerisms, historical sound changes affecting all (or most) lexical items in a language and each other, and the fact that allophonic processes normally do not show lexical conditioning. I further suggest that storing speech in terms of a small number of production/perception units such as phonemes could be due to the fact that phonemes seem to optimize both efficiency and informativeness in much the same way as other basic-level categories.

KEYWORDS: history of phonology, phonemic processing, usage-based theories, natural phonology

[*] ***Address for correspondence***: Geoffrey S. Nathan, Department of English, Wayne State University, 5057 Woodward, Detroit, MI, 48202, USA. Tel. 1 313 577-8621, Fax: 1 313 577-0404, e-mail geoffnathan@wayne.edu

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

I. INTRODUCTION AND BRIEF HISTORY OF THE PHONEME

The phoneme has had a checkered history—it was first proposed as a name for something like a speech sound by Dufriche-Desgenettes (Anderson, 1985), but soon evolved into essentially its current sense of a minimal unit of sound in which lexical entries were 'spelled' by the time Baudouin de Courtenay explored the concept (Baudouin de Courtenay, 1972). Baudouin believed that phonemes were *intentions*, but that speakers often (perhaps normally) 'missed' their intentions due to *divergences*, some of which were automatic and unconscious (psychophonetic), and others of which were (we would now say) conventionalized (paleophonetic). Kramský (1972, 27) describes Baudouin as being interested in "the representational area of individuals' linguistic consciousness—the psychic equivalent of a sound" (Baudouin, 1972 [1894]: 152) (what we would now call an abstract mental image).

A countervailing conception, originally espoused by Saussure, argued that phonemes, like all other units in language, were defined by 'otherness', as existing solely in terms of their *value* in the system, as he defined for the future of many fields the notion of *structuralism* (Saussure, 1974 [1916]). Furthering this view was the view developed by the Prague School, as exemplified by Trubetzkoy (1939 [1969]), who introduced the notion that phonemes were defined by *features* of otherness, and that a phoneme was made up of a list of those features. Bloomfield, a proponent of much more concrete theorizing, nonetheless gave phonemes the same definition—a "bundle of distinctive features" (Bloomfield, 1933: 77).

Prague School linguists such as Trubetzkoy and Jakobson specifically attacked Baudouin's view of the phoneme because the Prague school believed that speculation about internal representations was inappropriate for linguists. Trubetzkoy argued for what we would now term an 'autonomous' definition of the phoneme, a structural one, because, under an appropriate division of labor, psychologists' job was to think about storage, but linguists' job was to understand systems in terms of the internally contrasting elements – "reference to psychology must be avoided in defining the phoneme since the latter is a linguistic and not a psychological concept" (Trubetzkoy 1939 [1969]: 38). Similarly, Jakobson (quoted in Kramský, 1972: 27) said Baudouin's theory called for "the disadvantageous transfer of phonological problems from the

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

firm ground of linguistic analysis to the hazy area of introspection and their being made on such unknowns as the psychic impulses of the speaker...".

A counterview, during the same period of the development of twentieth-century linguistic theory was that of Jones (Jones, 1967), who argued that Jones argued that any given phoneme is a 'family' of related sounds.

Edward Sapir, in his classic article on the subject (Sapir, 1972) made a strong case for a similar view to that of Baudouin, that phonemes were actual *mental* images, but that the production of the images diverged from their targets due to 'the application of absolutely mechanical phonetic laws' (p. 25). As evidence he cited such observations as the fact that his Southern Paiute consultant, whom he had taught to write Papago, insisted on writing what Sapir clearly heard as voiced consonants as voiceless, because the voicing was allophonic. Notice, incidentally, that defining phonemes through the tools of minimal pairs, and the concentration on 'otherness' does not occur within Sapir's conception of the phoneme. While minimal pairs may have been useful for the linguist attempting to break into a system s/he does not speak, Sapir did not appear to believe that native speakers use that methodology to learn the language in the first place.

Within the American Structuralist tradition the more 'rigorous' Bloomfieldians made no commitments to the psychologically real (as opposed to analytically useful) nature of phonemes, and Twaddell is famous for arguing that any attempt to speculate about the contents of the mind was akin to kindling 'a fire in a wooden stove' (Twaddell, 1935: 9) but Kenneth Pike, from a competing group of structuralists (and one of Sapir's students) argued against the classic Trager-Smith phonemicization of American English in part because he found it very difficult to get linguistics students to understand the vowel system Bloch/Trager/Smith proposed (Pike, 1947).

The modern 'generative' history of the phoneme is well-known, from Halle's classic attack (Halle, 1959), as well as the basic statement of generative phonology (Chomsky & Halle, 1968). Essentially, generative phonology continued the Saussurean/Trubetzkoyan contrastive definition.[2] There was a time during the late twentieth century development of phonemic theory (not what it was called, of course, but what it was, nevertheless) when the notion of *distinctive features* was pushed as the definition of the phoneme to an extreme logical conclusion. It was

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

suggested that there could be phonemes (one per language) which were essentially the 'default' phoneme for the language. The default phoneme was defined by no features at all (one version of this was *Underspecification Theory* (see, for example Archangeli, 1988 and Steriade, 1996).

Within contemporary generative phonology the actual psychological status of the phoneme is not settled. There are those working within Optimality Theory who are simply not worried about the psychological reality of phonemes (or of the model in general),[3] and others who attempt to ground all of their theoretical constructs in physiological or psychological concepts (see e.g. the readings in Hayes, Kirchner & Steriade, 2004)

A non-generative stream that was very influential in the seventies and eighties, but has since been largely eclipsed, was Natural Phonology (Donegan & Stampe, 1979; Stampe, 1987). Natural Phonology argued that phonemes were concrete mental images of sounds, but that, as Sapir and Baudouin had argued, actual speech diverged from these stored sounds, and that the process of speaking included the real-time calculation of the divergences (and, in fact, that the divergences were dependent on speech situation, degree of precision intended by the speaker and the frequency of the utterance).

More significant, Stampe argued that speech *perception* was mediated by the same process, as hearers (who were also speakers) went through a process of 'sympathetic reconstruction' of what they heard, going through a process that could be paraphrased something like this: 's/he just said [x], which sounds like what would have come out if I had intended to say /y/'. Such a model explains many well-known phenomena across a large number of fields, including, for example, the facts that second language acquisition theorists had subsumed under the rubric of 'interference', as well as the 'phoneme restoration' facts (Samuel, 1996), in which subjects report hearing sounds that have been 'surgically' removed from a waveform, even if they are listening fairly carefully. The model is also not crucially dependent on the discovery of minimal pairs, as phonemes are units of intended perception, not classes of existing sounds.

The notion of the perception of intentions, incidentally, is not foreign to psychologists—although Stampe and Donegan did not explicitly refer to his work, it is clear that their view is similar to the Gibsonian school of perception (Gibson, 1979), which argued that organisms are attuned to the nature of their environment such that perception is of higher-level entities, and that

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

the perceptual system is designed to automatically reconstruct the objects 'behind' the percepts that impinge on the sense organs. Thus, if appropriately configured sets of moving lights are projected in a dark room people (and spiders, as Gibson showed in an elegant experiment) actually perceive moving objects, not coordinated moving sets of lights. The fact that little glowing lights attached to a person moving around in a totally dark room is perceived as a person (another experiment developed by Gibson) further emphasizes the importance of interpretation of percepts as unified entities. Similarly, the Haskins school of speech perception argued (see, for example Liberman & Mattingly, 1985) that people can be argued to perceive vocal tract movements, not abstract sound patterns.

II. THE PHONEME IN COGNITIVE GRAMMAR

The Cognitive Grammar revolution, which began in 1986 with the twin works of Lakoff (1987) and Langacker (1987) changed the way (at least some) linguists looked at such fundamental linguistic concepts as rules, lists and representations. It argued strongly for an insistence on psychological grounding for all linguistic units, claiming that there are no linguistically-specific principles at all. Lakoff and also Johnson (1987) argued that linguistic knowledge represented *embodied* experience, which Nathan (1999) argued meant that phonemes were represented as articulatory and acoustic mental images. Furthermore, Langacker argued, from the beginning, that all linguistic theoretical entities must be based on properties of the linguistic data without reliance on a priori linguistic categories such as abstract linguistic features that generative grammar has been forced to assume are innate categories.

Nathan (1986, 1996, 1999) argued that the Stampean Natural Phonology view of the nature of phonemes (and the ontological status of natural processes) was compatible with the Cognitive Grammar commitments to explanation and non-autonomy of theoretical constructs, arguing that phonemes were basic-level prototype structure categories (in the Roschian sense—(Rosch 1975, 1978)) and that natural processes were image-schema transformations analogous to those explored in some depth in Lakoff (1987). A similar claim, although not made within the Cognitive Grammar framework, had been made by Jaeger (Jaeger 1980a, b)

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

Other Cognitive Grammar theorists have taken issue with the fundamental notion of representation of abstract sound images, arguing that such abstract categories were inconsistent with the fundamental insistence of Cognitive Grammar on concrete, experience-based grounding for all linguistic units. The development of what has come to be called 'usage-based' linguistic theories (Langacker, 1988, Bybee, 1999, Barlow & Kemmer, 2000, Bybee, 2001, Kristiansen, 2006) has argued that speakers construct phoneme categories on the basis of numerous instances of actual speech events stored, in some cases, 'on top of each other', so that commonalities arise naturally out of similarities in acoustic images, or articulatory events.

Others, working in frameworks separate from those of Cognitive Grammar, but compatible with it, such as Pierrehumbert (2002), have argued that prototype theory is not an appropriate model for storage of linguistic (or other) kinds of experience, proposing instead versions of *exemplar* theory, which, similarly to usage-based models argues that speakers store virtually all instances of everything they have heard, extracting commonalities from instances that are categorized as, in some sense, 'the same', with those commonalities often corresponding to units roughly the size of phonemes, but the phonemes themselves being secondary to the individual instances of individual lexical items (and, of course, larger, and perhaps smaller units). Recent work by those working explicitly within the framework of Cognitive Grammar, such as Bybee (2000, 2001), have argued that phonological theory has erred in assuming the existence of abstract phonological categories at the level of the phoneme.

Bybee has also argued that words are actually stored as individual instances, and that speakers evolve generalizations from similarities among pieces of the words, but without ever recoding the existing words in all their phonetic detail. That is, according to Bybee, phonemes are generalizations built upon existing stored entities, but do not, in any sense, change the way that words are stored. Bybee goes so far as to suggest that, for example, the allophones of phonemes in syllable-onset position may not be stored as in any sense 'the same' as those for what are traditionally thought of as the same phoneme in syllable-final position, suggesting, for example, that clear and dark [l] in English might not be categorized the same (Bybee, 2001: 88).

Although I will argue against this general view below, I simply point out here that not only do children not appear to have any difficulty spelling *leap* and *peel* with the same

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

consonants in different orders, but that it is extraordinarily difficult to get naive phonetics students even to *hear* the difference between the two sounds. Special training (such as a phonetics course) is required to hear this difference, which makes it quite different from the contrast, say between /s/ and /z/. Although these two sounds are frequently written the same (*loose* vs. *lose*, for example), children have no trouble learning the letter <z>, or using to represent /z/, and notice the jocular spelling of the plural with a <z> in, for example, the illegal file-sharing world of computer enthusiasts (illegal files are referred to as 'warez', Urbandictionary.com 'warez').

It is certainly not the case that the primary school teachers and parents who teach their children to spell are aware of these differences, but if the differences are as great as any other lexical difference we would predict clear and dark [l] would be easy to hear, not to mention other differences, such as aspiration. Children sometimes incorrectly categorize voiceless stops after /s/ as voiced (although very seldom, because that would violate very natural phonotactic restrictions which are exceptionless in English), but it is never difficult to teach children to spell 'stop' with a <t> (see Treiman, 1985, 1993 for some discussion), but, on the other hand, it is *very* difficult to get phonetics students to hear final devoicing in American English, so that they can hear that *bread is* actually pronounced with a voiceless unaspirated stop by most speakers most of the time.

III. WHY PHONEMES ARE OPERATIONAL MENTAL CATEGORIES

III.1. Phonemes and writing systems

As a counter to the view that phonemes, especially defined as recurring identical (even though they are not) sound entities, are an invention of linguists, it should be pointed out that the notion of a small number of segment-sized recurring units is one of the oldest ongoing concepts in human culture. Extensive discussion of writing systems can be found in Daniels and Bright (1996).

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

Recall that the Phoenecians began writing symbols that stood for consonants approximately 1800 BCE (there is some evidence that they got the practice from an earlier Egyptian variant on hieroglyphics). Their innovation was to use pre-existing symbols to represent the onset of the first (normally stressed) syllable of the word, so, for example, the symbol for 'ox' /ʔalep/ represented glottal stop, and the symbol for 'water' /mem/ represented /m/ (and the modern Roman uppercase letter has not changed much from that early symbol, amusingly enough). The alphabet (technically an 'abjad', because there were only symbols for consonants when first adopted) spread wildly around the Middle East in the second millennium BCE, becoming the basis for Hebrew, Arabic, Greek and finally Latin alphabets. Since that time, of course, it has spread throughout the world, becoming the basis for the majority of writing systems currently in use.

Of those not using Phoenician-based alphabets Korean uses a phoneme-based alphabet,[4] Japanese uses syllabaries (in addition to the Chinese-based logographic system), and, of course, the Chinese logographic system is still widely used by all Chinese speakers as well as still being used by older people in Korea, Vietnam, and, as mentioned above, is still an essential part of the overall writing system in Japan. All other living writing systems are variants of either alphabets or syllabaries (Mongolian, Coptic, Cree/Inuit). Thus, the most popular kind of writing systems are based on the assumption that there is a small number of repeatable, meaningless sound units, although opinions differ as to the exact size. Languages with complex syllable structure (essentially, any language with codas of any kind) normally choose an alphabet rather than a syllabary.[5]

The fact that, in culture after culture, language after language the writing system that survives is alphabetic (or, occasionally, syllabic) tells us that the psychological reality of understanding speech as made up of segment-sized, meaningless and recombinable units is very strong. It is true that the acquisition of literacy is a non-trivial task, but the fact that the vast majority of young children across many cultures learn to write an alphabet within less than a year suggests that the phoneme, a linguistic concept, has considerable psychological validity.

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

III.2. Phonemes and speech errors

There are a number of facts about how people process language sounds that any phonological theory will need to account for. Most of these are well known, but it seems useful to rehearse them here, as it will be necessary for any model based simply on generalizations of individual usage events to account for their occurrence as well as the more traditional representational models (especially the phonetically realistic ones, such as Natural Phonology).

One area that requires serious attention is the existence of phonological processing errors that are incomprehensible unless we can assume that speakers are dealing, online, in the process of speaking, with real units at the level of abstraction the analyses by Sapir/Baudouin/Stampe suggest. Consider two examples the author gathered while participating in recent conversations, one in a meeting about computer programs, the other in a personal conversation.

The first example illustrates that syllable units such as the rhyme participate in ongoing speech production. A native speaker of American English, while aiming at

1 [bæ̃nɚ pʰɑɚʔnɚ] 'Banner partner'

produced the following:

2 ['bɑɚʔnɚ ˌpʰæn…]

To explain this we note the syllable boundaries. Syllabification of the two words is as follows:

3 /b**æn**.ɚ/ and /p**ɑɚt**.nɚ/

What appears to have 'switched places' are the corresponding *rhymes* of the stressed syllables, leaving the onsets in place. Unless rhymes are actual linguistic units involved in the planning and production of speech we cannot explain why we find 2.

It goes without saying, of course, that rhymes are also crucial in the construction of poetry and song, and that both activities do not require literacy. While twentieth century song writers no doubt are generally able to read and write, there are many non-literate cultures (both contemporary and historical) in which poetry and song based on rhyme schemes, or on alliteration—the identity of *onsets*—is a standard part of the culture. And it is also certainly the case that preliterate children have no trouble appreciating nursery rhymes long before they can spell.

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

A more complex example, captured in the week this paper was being written is the following:

4 Zeno's Paradox -->['zi˩naksəz...]

Here what has exchanged is again the rhyme, but in this case, the rhyme of the secondarily stressed second syllables (or perhaps feet). That will account for the output [zinaks]. However, we also find evidence for the online, real time selection of the appropriate allomorph, [əz], rather than what would have come out without the speech error, namely [z]. This is exactly the same kind of "absolutely mechanical phonetic laws" that (Sapir, 1972: 25) referred to.

A final example, discussed in earlier publications, was the error produced when a speaker aimed at 'whatever was...'. Surely such a phrase, containing the word 'whatever' is as good a candidate for a stored lexical unit as anything one could imagine. Note that the word 'whatever' is always said, in American English, with a tap. However, the actual utterance was:

5 [waʔwəzɛvəɾ]

We cannot understand this utterance without admitting that the flap in 'whatever' is not stored as such. It cannot be, else where would the glottal stop have arisen? The glottal stop is, of course, the appropriate instantiation of the /t/ phoneme in 'what',[6] when it occurs preconsonantally.

Current models such as those proposed by Bybee and others, in which there is little online construction of speech from constituent units, cannot account for speech that appears to have been constructed from more abstract units in real time.

III.3. Evidence from vowel shifts

Another well-known fact about phonological behavior that requires some explanation comes from historical change. The widely studied sound change known as the Northern Cities Vowel Shift involves a chain of changes (whether it is a push or pull chain is not important for our purposes). The change, which applies in the American cities surrounding the Great Lakes (Buffalo, Cleveland, Detroit, Chicago, and, to some extent, St. Louis) applies to the lax vowels, and has the following character:

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

5 ɔ > ɑ > a, æ > ɛə, ɛ > ʌ

Labov (1996) describes this change as follows:

> The shift begins when /æ/, the vowel of *cad*, moves to the position of the vowel of idea /iə/ (1). The vowel /o/[7] in *cod* then shifts forward so that it sounds like cad to speakers of other dialects (2). /oh/ in *cawed* moves down to the position formerly occupied by *cod* (3), /e/ in *Ked* moves down and back to sound like the vowel of cud (4), *cud* moves back to the position formerly occupied by *cawed* (5), and /i/ in *kid* moves back in parallel to the movement of /e/ (in section entitled 'Chain Shifts'—no page numbers provided)

This change applies to every word containing the relevant sound, although other sound changes appear to work their way through the lexicon via one route or another. What is important about such sound changes is that they apply to every sound, and the same sound does the same thing in every word (with the exception, for lexical diffusion, that some words simply 'do nothing' and do not participate in the change). Incidentally, all sound changes (see e.g. Labov, in press) represent the changes that establish classic dialect and language families and cause linguists to set up genealogical trees).

How can this be accounted for in theories in which phonemes are abstractions generated from repeated individual instances of words containing the sound? If, for example, /æ/ is an abstraction discoverable by, in some sense, examining every instance of every word a speaker has ever heard that contains that sound, how can we explain the fact that the same speaker (always a child in the case of true 'transmission', not 'diffusion') changes every single word. If the categorization of the sounds is secondary to their storage we need to assume that each word is somehow 'indexed', and that speakers need to do a 'find and change' (analogous to what we do in a word processor). If, on the other hand, phonemes are the actual units in which words are 'spelled', perhaps as some kind of unification of a set of articulatory instructions and a set of (rough) formant patterns, then a single change in the phoneme itself, however it is represented, will lead to its being activated everywhere, since there is only one sound to change.

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

A second corollary to this view (namely that when children learn words they actually store them as sequences of this small set of basic units (as Taylor, 2006 also argues), is that we can understand how chain shifts can occur. Note that the Northern Cities Vowel Shift is such a chain shift. What this means is that in some sense the /ɑ/ vowel is interacting with the /æ/ vowel, which is interacting with the /ɛ/ vowel (I use the neutral term 'interacting' to avoid a commitment to 'push' versus 'drag' chains, a point that is unimportant to my argument).

If words are simply stored as individual items, perhaps with some cross-classifying 'index', then it would make no sense to say that the /æ/ vowel is 'pushing on' the /ɛ/ vowel, since there *is* no category of /æ/ vowels to do any pushing. It is certainly meaningless to say that the word 'bag' is pushing on the word 'ready'. We need to have some sense of active cognitive processing, involving categorization and production in real time, not 'recollected in tranquility' if we are to understand how phonemes could move in a coordinated manner, as we know they have done numerous times in the past, and continue to do as we study ongoing living speech. It is only a within a model in which phonemes have some independent, real existence, but are also instantiated in each word that contains that we can understand how real regular sound change, either historical or ongoing could take place.[8] That is, we need the notion of an inventory of actual phonemes and a lexicon of words 'spelled' in those phonemes.

III.4. Evidence from speech processing studies

Yet another way in which phonemes appear to be real, mentally coherent categories in which speakers operate in real time as they speak and hear their language comes from the fact, discussed by Cutler (2002), that allophonic processes normally do not show lexical conditioning:

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

> Studies of coarticulation reveal regularities which are determined by phonemic environment—the gestures which correspond to /k/ are different if the following vowel is high front /i/ rather than low back /ɔ/, for instance. Such studies have not revealed a role for the word itself as a determiner of regularity—high frequency words such as *key* and *cause* and low frequency words such as *kiwi* and *caucus* show the same patterns of variation. Without some expansion of the episodic modeling framework beyond word-specific phonetics, such regularities must presumably be ascribed to chance (p. 287).

A recent study of the perception of stress in contrasting language types emphasizes the importance of recognizing that some facts about words are simply not stored, no matter how many times the words may be repeated, and no matter how many 'obvious' the fact is to an outside observer. It is well known that French stress is completely automatic, falling on the final full vowel (needed to exempt schwas) of a breath group (Schane, 1968b). As such, French speakers regularly produce stressed syllables, but apparently do not store stress at all.

In order to test this assumption, Peperkamp and Dupoux (2002) attempted to teach French speakers (as well as Finnish, Hungarian, Polish and Spanish speakers) to repeat non-words that contrasted in stress placement alone. They found that French and Finnish speakers made significantly more errors in stress placement than did Hungarian speakers, while Polish and Spanish speakers did best. These languages can be rated along a scale of stress predictability—Hungarian, although having initial stress, has exceptional borrowed words, while Polish and Spanish have a stress 'window' of the well-known kind (somewhere in the last three syllables, depending on phonology, morphology and lexical effects). Speakers of French and Finnish simply can not get the stress right at all—they are, in the words of the authors, 'stress deaf'. But surely speakers of these languages *hear* stress all the time, and, if usage-based theories are correct, must be storing it. Nonetheless, they seem incapable of hearing it in the 'wrong' place, where their language forbids it, even if the task is simply to repeat a word heard immediately before.

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

Taylor (2006) suggests that phonemes are basic-level categories (following a discussion in Nathan 1986 and elsewhere). Why speakers should feel the need to store speech in terms of a small number of production/perception units remains a question that many continue to ask, and many now contest. I think a suggestion for the reason might be found in the original notion of basic level category, as proposed by Rosch, and elegantly summarized in a recent work on how basic musical motifs (such as the opening phrase of Beethoven's *Fifth Symphony*) constitute basic level categories. Zbikowski (2002) suggests:

> Rosch and her associates suggested that two contrasting principles influence the taxonomic level at which people prefer to categorize. The first is the efficiency principle, according to which people prefer to minimize the number of categories they must consider in making a categorization....The second principle is the informativeness principle, according to which people tend to maximize the informativeness of their categorizations. Since the most information about any entity is found at the most specific level of a taxonomy, you would use grand piano to categorize the thing sitting in the living room. Rosch and her associates proposed that the intermediate level of a taxonomy (in this case piano) optimizes both efficiency and informativeness and is thus the preferred level for basic categorization. A number of empirical operations converge at the basic level. The basic level is the highest level whose members have similar and recognizable shapes; it is also the most abstract level for which a single mental image can be formed for the category. The basic level is also the highest level at which a person uses similar motor actions for interacting with category members. The basic level is psychologically basic: it is the level at which subjects are fastest at identifying category members, the level with the most commonly used labels for category members, the first level named and understood by children, the first level to enter the lexicon of a language, and the level with the shortest primary lexemes' (pp. 32-33).

Similarly, Mompeán (2006) shows that categories at the level of abstraction roughly equivalent to the phonemic level are most easily perceived and stored.

III.5. Evidence on the recoding of stored examples

It is indeed the case that our ability to store things we have heard is much greater than those who worry about the 'poverty of the stimulus' would credit. Pierrehumbert (2002) cites evidence that

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

speakers recognize new words more easily the next time they hear them if they hear them in the voices of the speakers who introduced them to those words. She argues on that basis that phonemes are simply generalizations over numerous stored instances, without the need for much in the way of recoding (although she allows for the possibility that some repeated instances are similar enough to others that the new instances may not be stored separately.

However, there is evidence against this view, and in favor of a view in which learners store speech 'as they would say it', rather than behaving as Taylor says, 'like tape recorders' (Taylor, 2006: 45). It is certainly the case that we can hear, and remember how a specific speaker sounds while saying a particular word, in some detail. Research on the perception and storage of music also argues that we have stored information about the absolute pitch of songs we have learned (say by hearing them repeatedly on the radio). Levitin (2006: 149) found that people who are asked to sing songs they have learned in that way normally start 'at or very near the absolute pitches of their chosen songs'. This would appear to argue that children (and presumably adults) learn new words in exactly the same way. But when children learn to speak there is no evidence that they pronounce words they learn from their fathers on a lower pitch than the words they learn from their mothers, and, of course, in general they pronounce all words at a higher pitch than either (in part, of course, because their vocal cords are smaller). But a moment's introspection will tell us also that when we are introduced to someone who has an unusual name, we do not repeat it on a different pitch if the person is female than if they are male. Clearly some instantaneous recoding has gone on. And, of course, the same recoding is quite possible. Although Levitin did not deal with this issue, it is very unlikely that, if someone with a bass voice is asked to sing a song he learned initially from a soprano that he will attempt to match the soprano range. Instantaneous transposition is something anyone can do (I am speaking here, of course, of actual singing—*writing transpositions*, in the sense of musical notation is a difficult

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

skill that takes several years of formal musical training). And, of course, we instantaneously repeat someone's name in our own dialect when we are introduced: when someone with the Northern Cities Dialect shift introduces themselves to me as [mɛət] I reply 'glad to meet you [mæt]'. Not only would it be rude to imitate the Detroit vowel, I simply cannot do so with the authentic vowel required, and my students find it amusing to hear me try.

III.5. Evidence from child language acquisition

Although much has been written on this topic it seems important to mention once again that there are numerous reports of children systematically substituting one phoneme for another in every word containing that phoneme. Smith (1973) is a classic study, for example, but a bilingual child that my colleagues and I have been studying systematically replaced all mid vowels (/e/, /o/) in his Spanish with high vowels for several months. Similarly, a child of my acquaintance currently learning English regularly replaces words containing voiceless dental fricatives (/θ/) with labiodentals (/f/). This kind of replacement, virtually universally reported, poses two problems for any usage-based theory. First, where do children learn to replace what are almost always *more marked* segments with *less marked* segments. They certainly are not learning it from their surroundings, because these are not only non-standard replacements, they are non-existent in the ambient language. And secondly, unless children are storing the words they attempt *with the phonemes as separate representations,* how else can they know which words contain the relevant sounds. If we assume that phonemes are, in some sense, 'calls to motor routines', rather than simply linguists' classification schemes, we can explain why children alter the routines, leading to replacements in every relevant word.

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

IV. CONCLUDING THOUGHTS

How can we sum up where the phoneme stands early in the twenty-first century among functionalist and other not strictly structuralist phonologists? It is certainly reeling under the combined assault of a large group of Cognitive Grammarians (Bybee, Kemmer, Langacker and others) and a smaller set of phonetically-oriented experimental phonologists (Pierrehumbert, Beckman (2000), Port (Port & Leary, 2005) and others). They provide evidence that speakers can hear fine phonetic detail, and appear to store it long-term. On the other hand, speakers have no metalinguistic access to that information. Unless the individual (or even dialectal) details cross phoneme boundaries (specifically the phoneme boundaries of the speaker/hearer), speakers have no way to talk about the differences. Without extensive training in phonetics speakers are unable to *describe* the clear [l]'s of Irish English or the distinctive way George W. Bush pronounces 'decider'.[9] Furthermore, no language (other than the specialized tool of the IPA) provides a way of recording such details, leading one to assume that speakers feel no need to identify it. Furthermore, people have been learning to 'spell' for almost four thousand years, and, however complex that task may be, it is a task that requires no scientific equipment, and can apparently be learned 'spontaneously' at least by some children (Read, 1986).

I believe that there remains reason to believe that people *do* hear a small number of distinct sound units *in* the words that they acquire as children. Furthermore, their storage of those sound units appears to be active not only in perception but also in production not only of speech they have heard before, but also of novel utterances, and even of the non-words produced during whatever processing errors are reflected in spoonerisms and other speech errors. There are very few psychological theories that remain quite as robust four thousand or so years after they were first invented, but So, although many are aiming at it, this four thousand year old target—the perception and production of speech sounds by naïve speakers in terms of recurring, identical

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

segments, which differ apparently without notice in different contexts, in both language-specific and universal ways—in short, the phoneme, appears to be resisting calls for its demise.

NOTES

1. I would like to thank Margaret Winters and José A. Mompeán for their comments on earlier drafts of this paper.

2. It should be pointed out, however, that even in the midst of the more orthodox generative view there were generative linguists, such as Schane (1968a, 1971) and Kiparsky (1982) who 'reverted' to a more traditional 'autonomous' view of the phoneme, not so much in the sense of autonomy from mental reality (as Trubetzkoy and Jakobson preferred) but in the Chomskyan polemic sense of not necessarily related to morphophonemic alternations, and derivable solely by distributional facts. It is important to note, however, that Schane and Kiparsky were arguing for the *psychological reality* of the 'autonomous' phoneme.

3. Consider, for example, the fact that the basic reference to OT, Prince and Smolensky (2004 [1994]) does not discuss the issue at all. Much of the literature on learnability of OT, in addition, is based on theoretical discussions of learnability as algorithms, rather than dealing with actual language input.

4. While the symbols in Korean are arranged into syllabic units, each component symbol represents a (somewhat abstract) phoneme. See, for example, King (1996).

5. Hittite attempted to use a syllabary, with mixed results. There was great inconsistency on how CVC structures were written—sometimes CV-V-VC, other times CV-VC, and yet others CV-V-V-VC. See Gragg (1996).

6. This is a fact about American English, but the speaker (the author) is a native speaker of that dialect for whom /t/ is always pronounced as glottal stop preconsonantally.

7. Labov uses the Trager-Smith transcription to describe American English. Thus he uses /o/ to represent what was historically /a/, and /oh/ to represent what is, in the dialects that have this sound, approximately /ɔ/

8. Not only does Labov assume in the articles mentioned above that there are real, categorically regular sound changes, but he is convinced that they exist, and that they serve as a counterexample to at least some part of the totally usage-based models that Cognitive Grammar has recently taken to heart (Labov, p.c.).

9. Note that the stigmatized pronunciation [nukjələ˞] is *phonemically* very distinct from the standard pronunciation [nuklijə˞], making comment on Bush's habitual pronunciation fodder for his critics.

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

REFERENCES

Anderson, S. R. (1985). *Phonology in the twentieth century: Theories of rules and representations*. Chicago: University of Chicago Press.

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology*, 5(2), 183–208.

Barlow, M., & Kemmer, S. (Eds.) (2000). *Usage-based models of language.* Cambridge: CUP.

Baudouin de Courtenay, J. (1972). An attempt at a theory of phonetic alternations. In Edward Stankiewicz (Ed. & Trans.), *A Baudouin de Courtenay anthology: The beginnings of structural linguistics*. Bloomington/London: Indiana University Press, pp. 144-213.

Beckman, M. (2000). The ontogeny of phonological categories and the primacy of lexical learning in linguistic development. *Child Development*, 71(1), 240-249.

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

Bybee, J. L. (1999). Usage-based phonology. In M. Darnell, E.-A. *et alia* (Eds.), *Functionalism and formalism in linguistics, I: General papers; II: Case studies.* Amsterdam: Benjamins, pp. 211-242.

Bybee, J. L. (2000). The phonology of the lexicon: Evidence from lexical diffusion. In M. Barlow & S. Kemmer (Eds.), *Usage-based models of Language*. Stanford: Center for the Study of Language and Information, pp. 65-86.

Bybee, J. L. (2001). *Phonology and language use.* Cambridge & New York: CUP.

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

Cutler, A. (2002). Phonological processing: Comments on Pierrehumbert, Moates et al., Kubozono, Peperkamp & Dupoux and Bradlow. In C. Gussenhoven & N. Warner (Eds), *Laboratory Phonology 7*. The Hague: Mouton de Gruyter, pp. 275-296.

Daniels, P. T. & Bright, W. (Eds.). (1996). *The world's writing systems*. New York: OU.

Donegan, P. J., & Stampe, D. (1979). The study of Natural Phonology. In D. Dinnsen (Ed.), *Current approaches to phonological theory.* Bloomington: Indiana University Press, pp. 126-173.

Gibson, J. J. (1979). *The ecological approach to visual perception.* Boston: Houghton Mifflin.

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

Gragg, G. B. (1996). Other languages. In Daniels, P. T. & Bright, W. (Eds.). (1996). *The world's writing systems*. New York: Oxford University Press, pp. 65-70.

Halle, M. (1959). *The sound pattern of Russian.* The Hague: Mouton.

Hayes, B., Kirchner, R. M., & Steriade, D. (Eds.). (2004). *Phonetically-based phonology.* New York: CUP.

Jaeger, J. J. (1980a). *Categorization in phonology: An experimental approach.* Ph.D. dissertation. University of California, Berkeley.

Jaeger, J. J. (1980b). Testing the psychological reality of phonemes. *Language and Speech*, 23, 233-253.

Johnson, M. (1987). *The body in the mind*. Chicago: University of Chicago Press.

Jones, D. (1967). *The phoneme: Its nature and use.* Cambridge: Heffer.

King, R. (1996). Korean Writing. In *The World's Writing Systems*. In P. T. Daniels & W. Bright (Eds.), Oxford: Oxford University Press, pp. 218-227.

Kiparsky, P. (1982) How abstract is phonology?. In Paul Kiparsky (Ed.), *Explanation in Phonology*. Dordrecht: Foris, pp. 119-164

Krámský, J. (1974). *The phoneme: Introduction to the history and theories of a concept.* München: W. Fink.

Kristiansen, G. (2006). Towards a usage-based phonology. *International Journal of English Studies*, 6(2), 107-140.

Labov, W. (1996). The organization of dialect diversity in North America http://www.ling.upenn.edu/phono_atlas/ICSLP4.html

Labov. W. (2007). Transmission and diffusion. To appear in *Language*, June 2007

Lakoff, G. (1987). *Women, fire and dangerous things. What categories reveal about the mind.* Chicago: University of Chicago Press.

Langacker, R. W. (1987). *Foundations of Cognitive Grammar*. Stanford: Stanford University Press.

Langacker, R. W. (1988). A usage-based model. In B. Rudzka-Ostyn (Ed.), *Topics in Cognitive Linguistics*. Amsterdam: Benjamins, pp. 127–161.

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

Levitin, D. J. (2006). *This Is your brain on music: The science of a human obsession*. New York: Dutton

Liberman, A., & Mattingly, I. (1985). The motor theory of speech perception revisited. *Cognition*, 21, 1–36.

Mompeán, J. A. (2006). The phoneme as a basic-level category: Experimental evidence from English. *International Journal of English Studies*, 6(2), 141-172.

Nathan, G. S. (1986). Phonemes as mental categories. In *Proceedings of the 12th Annual Meeting of the Berkeley Linguistics Society*, 12, 212–224.

Nathan, G. S. (1996). Towards a Cognitive Phonology. In B. Hurch & R. Rhodes (Eds), *Natural Phonology: The state of the art*. Berlin: Mouton/de Gruyter, pp. 107–120.

Nathan, G. S. (1999). What functionalists can learn from formalists in phonology. In *Proceedings of the Symposium on formalism and functionalism.* Amsterdam: Benjamins, pp. 305–327.

Peperkamp, S., & Dupoux, E. (2002) A typological study of stress 'deafness'. In C. Gussenhoven & N. Warner (Eds) *Laboratory Phonology 7*. The Hague: Mouton, pp. 203-236.

Pierrehumbert, J. (2002). Word specific phonetics. In C. Gussenhoven & N. Warner (Eds), *Laboratory Phonology 7*. The Hague: Mouton de Gruyter, pp. 101–139.

Pike, K. (1947). On the phonemic status of English diphthongs. In V. B. Makkai (Ed.), *Phonological theory: Evolution and current practice*. New York: Holt Rinehart and Winston, pp. 145–151.

Port, R.F. & Leary, A. P. (2005). Against formal phonology, *Language*, 81(4), 927-964.

Prince, A. & Smolensky, P. (2004 [1994]). *Optimality Theory: Constraint interaction in Generative Grammar*. Oxford, UK: Blackwell. Originally published as Technical Report Number 2 of the Rutgers Center for Cognitive Science.

Read, C. (1986). *Children's creative spelling*. London, England: Routledge & Kegan Paul.

Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104, 192–233.

Rosch, E. (1978). Principles of categorization. In E. Rosch & B.B. Lloyd. (Eds.) *Cognition and categorization*. Hillsdale, NJ: Lawrence Erlbaum, pp. 27–48.

Samuel, A. (1996). Phoneme restoration. *Language and Cognitive Processes*, 11, 647–653.

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.

Sapir, E. (1972). The psychological reality of phonemes [La réalité pschologique des phonèmes] (Translated and reprinted from *Journal de psychologie normale et pathologique* (1933), 30, 247–265). In V. B. Makkai (Ed.), *Phonological theory: Evolution and current practice*. New York: Holt Rinehart and Winston, pp. 22-31

Saussure, F. d. (1974). *Course de linguistique générale. Edition critique préparée par Tullio de Mauro*. Paris: Payot.

Schane, S. A. (1968a). On the non-uniqueness of phonological representations. *Language*, 44, 709-716.

Schane, S. A. (1968b). *French phonology and morphology*. Cambridge, MA: MIT Press.

Schane, S. A. (1971). The phoneme revisited. *Language*, 47, 503-521.

Smith, N. V. (1973). *The acquisition of phonology: A case study*. Cambridge: CUP.

Stampe, D. (1987). On phonological representation. In W. U. Dressler *et alia* (Eds), *Phonologica 1984*. London: Cambridge University Press, pp. 287–300.

Steriade, D. (1996). Underspecification and markedness. In J. A. Goldsmith (Ed.), *The handbook of phonological theory*. Cambridge, MA: Blackwell, pp. 114-174.

Taylor, J. R. (2006). Where do phonemes come from? A view from the bottom. *International Journal of English Studies*, 6(2), 19-54.

Treiman, R. (1985). Spelling of stop consonants after /s/ by children and adults. *Applied Psycholinguistics*, 6, 261-282.

Treiman, R. (1993). *Beginning to spell: A study of first-grade children*. New York: Oxford University Press.

Trubetzkoy, N. S. (1939). *Gründzüge der phonologie [Principles of Phonology]* (C. Baltaxe, Trans. and Ed.). Prague [Los Angeles]: *Travaux du cercle linguistique de Prague* [University of California Press].

Twaddell, W. F. (1935) *On defining the phoneme. Language*, Vol. 11, No. 1, Language Monograph No. 16.

'warez', Urban Dictionary http://www.urbandictionary.com/define.php?term=warez

Zbikowski, L. M. (2002) Conceptualizing music: Cognitive structure, theory and analysis. Oxford: Oxford University Press.

© Servicio de Publicaciones. Universidad de Murcia. All rights reserved.