Phonemes as Mental Categories
Author(s): Geoffrey S. Nathan

Please see “How to cite” in the online sidebar for full citation information.

Please contact BLS regarding any further use of this work. BLS retains copyright for both print and screen forms of the publication. BLS may be contacted via http://linguistics.berkeley.edu/bls/.

The Annual Proceedings of the Berkeley Linguistics Society is published online via eLanguage, the Linguistic Society of America's digital publishing platform.
Phonemes as Mental Categories
Geoffrey S. Nathan
Southern Illinois University

This paper is a first attempt to explore the nature of a theory of phonology within the general framework of cognitive grammar (Langacker in press, Lakoff in press). It will be found that a number of basic tenets of natural phonology will fit elegantly into this theory, although some of them will require some reinterpretation.

Cognitive grammar holds that language is not different in kind from other human cognitive activities—seeing, hearing, categorizing, thinking, and that it follows the same organizing principles that these other cognitive activities display. In particular, cognitive grammar has adapted the prototype/basic-level notions of Eleanor Rosch (1973, 1977, 1978) as a basis for the organization of units in language, as in all other human activities.

According to Langacker, language is an inventory of structured units at varying levels and degrees of abstraction. The units may be either phonological or semantic, but cognitive grammar rejects any notion of an autonomous level of grammar that is not directly relatable either to the semantic or the phonetic substance of language. A great deal has been written within this framework, exploring the consequences of these assumptions. Representative studies can be found in Lakoff and Johnson 1980, Langacker in press, Brugman 1982, Lindner 1981, Yeagle 1983, Rubba 1986, Winters 1985 and Langacker ms.

This research suggests that human perception, organization and encoding of experience is mediated through our interaction with the world. Direct, perceptual and physiological experience underlies our understanding of more abstract domains, and is used to structure our knowledge about subjects that do not have direct physical existence.

At this point, aside from a few brief comments in Lakoff (in press) and Langacker (in press and MS), no serious attempt has been made to investigate the sound systems of language within this general framework. In order to do so, it will be necessary to identify some of the crucial concepts that cognitive grammar takes as central to the nature and functioning of language and ask how they could be relevant to phonological theory. Two essential ideas that this paper will attempt to integrate into phonological theory are those of the embodiment of mental entities (that is, their expression in perceptual terms) and the notion of prototype.

Prototype effects were first discovered by Rosch, and have since been shown to be crucial not only in general cognition, but also in linguistic categorization (see for example the work by Ross (1972) and Berlin et al. (1974) and the discussion in Lakoff (1982 and in press)). The notion of prototype expresses the sense that objects being categorized are not defined by the
presence or absence of criterial distinctive features, but rather in terms of degree of similarity to what are perceived as prototypical members. To give the prototypical example, a robin is perceived as a prototypical bird, while an ostrich and a kiwi are less good examples of 'birddom', although they are still birds. Similarly, following an example discussed at length in Lakoff (in press), a woman who has produced a child as the product of sexual union with her husband, and whose time is entirely taken up by looking after that child, cleaning her house and preparing meals for her cohabiting but working husband, is a prototypical 'mother'. Any person who deviates from this specification is, both sociologically and linguistically, non-prototypical. Thus we have 'working mothers', 'surrogate mothers', 'single mothers' and so on.

Cognitive non-prototypicality can lead to linguistic markedness. Prototype effects are a result of the kind of categorization that Rosch suggests human beings used in dealing with the world. Categories, according to Rosch (1973, 1977, 1978), are groups of objects in the world assembled because of their similarities to one another, organized around a central, prototypical image of the member of the category that is most representative of all the members. Rosch argues that 'categories appear to be coded in the mind neither by means of lists of each individual member of the category nor by means of a list of formal criteria necessary and sufficient for category membership but, rather, in terms of a prototype of a typical category member. The most cognitively economical code for a category is, in fact, a concrete image of an average category member'. (1977: 30)

In addition to the assumption that mental categories are structured around representative examples, it is assumed that language is made entirely of sets of categories of varying levels of complexity and abstraction, but that every unit, at every level is related directly either to semantics, that is, to understanding and meaning, or to phonetics—that is, to the phonetic substance of language. Cognitive grammar explicitly rejects the prevailing notion that it is productive to investigate the structure of grammar independently either of meaning or what means are being used to convey it. Consequently, cognitive grammar is a totally non-autonomous theory in which syntax is purely symbolic of meaning, and in which there are no units that are not either meaning or sound.

Because cognitive grammar is a non-modular, integrative approach to the nature of linguistic systems, all posited elements and principles require non-autonomous explanations. In particular, only the semantic and phonetic poles, and units coordinating those poles, are permitted to have a place in linguistic theory. As a consequence, we will expect the nature of the phonetic substance will determine the kind of structures that we find in language, and this turns out to be exactly the case.
The structure of phonological systems is determined by the substance out of which the systems are constructed—in particular, for spoken language, the nature of the human vocal apparatus and the nature of the constraints imposed by our perceptual apparatus on what sounds and combinations of sounds we can hear. We can predict, consequently, that non-spoken language will be subject to different constraints, because the nature of the producing mechanisms, the perceptual mechanisms and the conducting media are different. On the other hand, the general organizing principles will be the same—categorization in terms of prototypes and basic and higher level organization will be reflected in the organization of whatever units language uses.

What kinds of constraints will be placed on the phonological system by the fact that language is spoken? We will expect a number of general requirements will impose conditions on phonology, and will shape its possible nature: it must be audible, rhythmic, convey information, be efficient, and reflect pre-existing principles of human nature such as emotionality.

The fact that language must be audible guarantees that the primary building blocks of language will be vowels, since vowels are the loudest sounds that the vocal tract can produce. It is only when the vocal tract acts purely as a (tunable) amplifier of the vocal cords that truly loud sounds can be produced—any kind of closure will limit the amount of noise that can be produced. Consequently, language will be primarily made up of vowels—reflected in the fact that most of the actual time in speaking is made up of vowel production, and also in the fact that the prototypical speech sound is probably a vowel, although I do not believe this has yet been demonstrated experimentally.

Rhythmicity is also a general organizing principle of human behavior. In an interesting paper, Kelso, Tuller and Harris (1983) argue that rhythmic organization is a general consequence of energy-producing systems of all kinds by virtue of their self-regulating nature. Furthermore, Lindblom (1983) argues that the rhythmic nature adduced above, coupled with the conflicting requirements imposed by the nature of consonants versus vowels, will result in the syllabic organization of speech sounds that languages universally show. Consequently, the organization of speech sounds into rhythmic units or pulses (perhaps partially coordinating with chest pulses, as Stetson (1951) or jaw movement, as Kelso et al. suggest) is to be expected given the nature of the human motor system.

Speech must also convey information. This means that there must be enough differences to allow information to be communicated. The connection between difference and information has been known for a long time—in one sense since de Saussure, and in the mathematical sense since Shannon and Weaver. However, as we have seen and will see, information-
bearing capacity is not the only requirement that the use of language imposes on the structure of language. Information is just one of the prerequisites to a speech system. It will guarantee, however, that language will be made up of a reasonable number of contrasting units—no language seems to be able to make do with less than about thirteen, while similarly no language seems to need more than around a hundred different sounds. In addition, limits on the human perceptual system will set limits to the minimum distance between contrasting units. For example, no language uses more than three degrees of voice onset time (Nathan, MS), a fact that seems to be determined by mammalian perceptual systems.

Speech must also be efficient, in the physiological sense. This guarantees that, where possible, conflicting physiological instructions will be resolved so as to reduce the magnitude and number of movements that the vocal apparatus is required to undergo. As we will see, this requirement will lead directly to an explanation for why languages have allophones, and why we speak of phonological rules.

Lastly, we can expect language to be organized in accord with the nature of the human expressive system in a wider sense. For example, general emotional arousal leads to exaggerated behavior of all kinds, presumably because of a direct connection between arousal and the motor system. As a consequence, we would expect that languages would have at their disposal the means for expressing emotional arousal which would involve exaggerated motor activity. In general, languages do express intense emotional involvement through increased respiratory activity, (Ladefoged 1982) leading to louder, longer sounds that exhibit greater pitch excursions. Even in languages without contrastive stress in the English sense, like French, we find such devices as the accent d'insistence used to call attention to particular parts of the speech stream. Not all languages make use of stress, as is well known, but I know of no language which does make use of stress, but in which the stressed morphemes are old or unsurprising information.

Stress, as a resource, is available to language, but not every language 'chooses' to use it. On the other hand, having adopted a system of stress, stress is constrained in the way it operates to mark specialness in some manner, either as an emotional signal, an informational one, or, as Troubetzkoy put it, a culminating (boundary) signal.

In sum, then, the nature of the human vocal and perceptual apparatus guarantees that language will be made up of strings of vowels and consonants, organized into syllables that (approximately) alternate open and closed positions in a rhythmic pattern. There will be enough contrasting units for information to be conveyed (although the number required varies by an order of magnitude), but each unit will permit enough variation to allow for efficient transition between motorically different activities. Finally, languages will have available
some device for calling attention to portions of the speech stream for purposes ranging from emotional coloring to signalling the beginning or ending of words.

Let us return at this point to the question of the nature of phonological organization, to see how this conception of phonological substance might interact with the concept of categorization that Rosch and others have been promoting.

The beginnings of this theory can be found in a suggestion by Jaeger (1980), who investigated the psychological reality of various putative phonological relations and entities. She found that, as one might predict, native speakers of English were able to perform various concept-formation tasks which involved associating various allophones of a single phoneme. With greater difficulty they were able to treat as single categories classes of sounds that were grouped by various proposed SPE-style features. Some features, such as voice and continuancy fared better than others such as anteriority. Finally, she investigated the status of the vowel-shift rule, finding that speakers apparently formed categories on the basis of the spelling system, rather than on the underlying forms proposed in SPE and the various revisions thereof. Of course, if one is investigating the psychological reality of phonological systems rather than taking the purely structuralist point of view that all that is being described is patterns, then such things as the influence of spelling are likely to become relevant to a correct description.

At the end of her dissertation, Jaeger suggests that Rosch's categories are the appropriate way to think about the relationship between phonemes and allophones, and it is this conception that I will explore here.

Let us suppose that a phoneme is a category of sounds. Each of the sounds that would be traditionally classified as an allophone will therefore be a member of that category. We can think therefore of a large network of sounds scattered like points on a (multi-dimensional) map. Each language will draw boundaries around some of the points (analogous to, say, state boundaries), enclosing sets of the sounds into categories that will constitute the phonemes. Other regions of the map will simply be unused, for each language. Thus a phoneme is a class of sounds that native speakers consider as being somehow 'the same', in much the same way as English speakers all perceive robins, bluejays and even chickens as 'birds'.

However, Rosch argues, at least for color categorization, that people are not free to group members of a category in any arbitrary manner. She argues that 'there are colors and forms which are more perceptually salient than other stimuli in their domains...Such colors and forms more readily attract attention than other stimuli and are more easily remembered than less salient stimuli' (1973: 114)

Analogously, for sounds, we could argue that some sounds within a category are more representative of the entire
category than others, and in fact that less prototypical members of the category will be heard as the more prototypical member. Furthermore, we could argue that languages are not free to group any arbitrary set of sounds into categories. Instead, I will argue that the sounds that are picked out as the most prototypical are selected because they are perceptually and articulatorily the most distinctive from the point of view of the phonetic substance discussed above. Furthermore, the sounds that are perceived as being members of the same category will be related to the prototypical sounds as natural, contextual extensions of the prototypical sound.

Put in another way, one set of principles will determine that some sounds are somehow more distinctive, while another set of principles will determine possible relationships between the most distinctive sounds and sounds that are similar to those, but are, because of their phonetic context, more appropriate to their surroundings than the prototypical sounds.

This conception of phonological organization is not unique to cognitive grammar, and in fact has been around for a long time. It is a version of the theory of natural phonology as proposed in Stampe 1973, Donegan and Stampe 1979, Donegan 1978 and 1985, Dressler 1985 and elsewhere. Donegan 1978 is a particularly elegant statement of the conception I am proposing here, and much of what I have to say at this point draws heavily from her work.

Donegan argues that there is a class of fortitions, which define the best examples of the class of vowels (and, analogously, consonants). These fortitions are a set of phonological processes which say, for example, that front vowels are likely to be unrounded, especially the lower they are, or that tense vowels are likely to be raised. She presents articulatory and acoustic arguments for the purpose of these processes, arguing that fortitions exist in language in order to make the segments more distinct, and consequently more hearable. Although the total set of fortitions defines maximum hearability for all possible sounds, any single language will select a subset of all possible fortitions, which will consequently define the set of phonemes for that language. Because the human perceptual and articulatory apparatus is the same for all people, we can expect that the total range of possible fortitions is limited, but that languages will differ in which and how many possible sound distinctions they choose.

From our point of view, then, we can say that the fortitions of a language define the center of each phoneme as the maximally contrastive member of the set, in much the same way as natural phonology has argued that the fortitions define the phoneme inventory of a language. Since the fortitions are determined themselves by articulatory and acoustic considerations, the sets of possible phoneme-centers will follow along the lines investigated by researchers such as Jakobson (1968) and Greenberg (1966), who found that there were
universal organizing principles governing the possible inventories of phonemes.

Any particular phoneme, however, will be made up not of a single sound, but of a class of sounds, all of which are classed as 'the same' by native speakers. What is the source of the variation in which sounds are actually produced? This brings us to the second fundamental concept of natural phonology--lenitions. Unless the language explicitly forbids it, the prototypical sounds are automatically replaced by less prototypical sounds if the phonetic surroundings require it. The principles governing these replacements are known as lenitions, and are what are usually represented as 'natural phonological processes' (although, within natural phonology they are only one half of the actual set).

Consider, for example, nasalized vowels. The addition of nasal coupling to vowels reduces the audibility of the resonances that render vowels distinct (i.e. nasality hides formants). Consequently, languages avoid phonemic vowel nasalization. On the other hand, due to the relative sluggishness of the velum as an articulator, it is easier to produce vowels with nasal resonances if a nasal consonant is in the same syllable. Thus we can say that there is a fortition excluding nasal vowels as prototypical vowels, but a lenition permitting nasal vowels as members of non-nasal vowel categories. A child acquiring English need do nothing special with respect to vowel nasalization--it is easier not to try to hear nasalized vowels, but under certain circumstances, it is easier to produce them, even if the target is non-nasalized.

What advantages does this conception of phonological systems give us? For one thing, several constraints on phonological representation that have been proposed by various theorists in various schools 'fall out as theorems' within this system. The constraint proposed by Postal (1968) which has come to be called the Naturalness Condition is automatically required by this theory. Postal argued that phonological representations must be just that--phonological. In this he was arguing against totally abstract representations in which non-phonetic features were posited. This theory goes even further, to argue that underlying and surface representations are all constrained by the requirement that they be sounds. In this case, the theory is even more strongly constrained, in that all representations must be sounds. This effectively excludes a great many possible alternatives, including archiphonemes and other incompletely specified sets of feature matrices, such as those proposed in Arcangeli (1984). In many ways it is similar to the kind of representations that Sapir apparently advocated (McCawley 1979, Sapir 1933). The theory imposes a kind of extreme naturalness constraint that disallows any representation at any level that is not pronounceable, because all representations are simply sounds.

There are, however, some additional complications which
need to be addressed at this point. In recent work Lakoff has suggested that the structure of categories is 'radial', in that less prototypical members may form chains of association with each other, leading further and further away from the central member. Any category may contain a number of chains, arranged like spokes on a wheel, in which members on adjacent spokes may have nothing in common, but are relatable only back through the prototypical member. A good example might be the somewhat overworked case of 'cup'. It is clear that various ceramic containers are all appropriately called cups, but in order to relate the extremely diverse entities of 'two sticks of butter' and 'hands held with palm concave' we cannot go directly from one to the other, but rather must travel through the intermediate of the ceramic entity, recognizing that each of the others is an extension, in different directions, of the prototypical, ceramic, handled, saucered cup.

Phonemes appear to be organized in exactly the same manner, with different extensions of the prototype being, perhaps, unlike each other, but relatable through a reasonable chain of lenitions to the prototype. For example, the /t/ in English may be pronounced as an ejective alveolar stop in certain circumstances (e.g. What?!) but as zero in others (as in a frequent pronunciation [sərdeɪ] for Saturday).

This conception of chains of related phones, successively further from the prototype, is the analog, within this theory, to sequential application of rules. Consider, for example, the system of English alveolar stops. We have at least the following:

```
1) t \_ tʰ \_ d \_ d
   \_ t \_ t
   \_ \_ t
   \_ \_ D
   \_ \_ D
```

The ejective and implosive sounds occur as instances of emphatic coda and onset respectively. The voiceless [t] occurs as an allophone of /t/ after /s/ and as an allophone of /d/ in initial position (see Nathan to appear for discussion). The voiceless flap occurs in syllable-final position in words such as 'hothouse' and even in 'plate'. The flap is voiced intervocally in standard examples like 'city'. /d/ of course may also be realized as both kinds of flap in cases like 'madder' and 'red-hot'. Here we have a complex case of overlapping and interlocking categories. A path along one of these chains constitutes a derivation. As Stampe and others have emphasized, there is extensive evidence that speakers travel these paths in the process of actively constructing speech, because when they produce Spoonerisms, the allophones
that are used are appropriate even if the actual phonemes have been rearranged.

Let us consider a case where the rule interaction becomes somewhat more problematic. In English liquids are devoiced following syllable initial voiceless stops. Thus [pliːz] and [kreit]. Consequently, we can say that a possible path exists between 1 and voiceless $\dagger$, sanctioned by the natural process of aspiration. In addition, English possesses a lenition permitting the deletion of schwa under certain circumstances. In our current terms we would say that zero is one of the alternative realizations of schwa (which may itself be an alternative realization of some other vowels). That is, included within the phoneme zone of schwa is nothing. This permits a potential 'rule-ordering' conflict between the two processes—what is to stop a chain whereby schwa chooses the zero allophone, and 'consequently' the liquid in a word like 'police' chooses the voiceless allophone, becoming virtually homophonous with 'please'. Donegan and Stampe (1979) argue that some processes are marked as non-iterative, while others are not so restricted. What we need here is some notion that a voiceless 1 can 'be an /l/' in the case where nothing is missing between the stop and the /l/, but cannot 'be an /l/' when there is a zero representing a vowel. The obvious answer here is that not devoicing the /l/ maintains the identity of the missing vowel—the missing vowel is 'recoverable', to use the terminology of the abstractness controversy. In that case, what we have is a restriction on permissible paths that permits two potentially distinct words from becoming merged. This of course is a semantic constraint on 'rule ordering'—a rule may not apply if it will conceal the existence of an otherwise invisible segment.

The conception of 'rule ordering' that is being presented here, is of course, global, in the now discredited sense that generative semantic theory proposed. Within a non-modular theory, however, there can be no a priori rejection of globality in rule application. And in a theory in which the only possible units are either phonological or semantic, any rule connecting the two levels will of necessity be global. Since cognitive grammar permits only phonological and semantic units, or units that are schematic of one or the other, we must expect that phonological units will be influenced either by other phonological units or by semantic ones—there are no other considerations. Given the notion of global constraints on whether a particular phoneme can 'be' another, the idea of any kind of sequential rule ordering can be eliminated. For any particular pair of possible extensions of the prototype sounds, a particular language can 'choose' whether the extension of the extension will deviate too far from the prototypical target. Although [pliːs] seems to me to be too 'sloppy' a pronunciation, I can certainly imagine other people who would not reject it, and as Donegan and Stampe show,
reorderings from counterfeeding to feeding with increasing tempo, or with extremely familiar words are frequent.

Donegan and Stampe argue that counterfeeding constraints of this kind are the only ones needed in phonological theory, and I have argued that they exist entirely for semantic purposes. The nature of language as the cognitive grammarians see it predicts that the only restrictions on possible rule types will be semantic ones, and that is exactly what we find.

One notion that I have not discussed at all is the distinction, proposed by natural phonology, between processes and rules. The distinction in our terms is between categorization (both \( t^n \) and D are instances of \(/t/\)) and higher level schemas (the morpheme divine, if it is recognized to be morphologically complex in the first place, may be pronounced either with an \(/ai/\) or with an \(/I/\). Processes deal exclusively with what sounds may count as other sounds, and with which sounds are the 'good' sounds for a particular language. Rules (in the technical sense) on the other hand, relate alternative pronunciations of morphemes, not phonemes, and consequently are semantically, rather than phonologically based. A similar conclusion is reached in lexical phonology, where a similar distinction is made between lexical and post-lexical rules (see Rubach 1984 for discussion), where lexical rules require information about morphemic identity, while post-lexical rules are purely rules of pronunciation.

Notes

* This paper has benefitted from the input of a large number of people. These include Margaret Winters, Jo Rubba, Ron Langacker, George Lakoff, Wolfgang Dressler and Lee Hartman. I alone am responsible, however, for the current state of this paper.

1 Study of languages that do not use the vocal tract, such as ASL, indicates the the basic organizing principles differ in just the ways that one would expect from the nature of the 'phonological' resources available. For instance, because visual space can be organized more complexly, ASL allows multiple third person pronouns, distinguished by location, whereas spoken language, with a single channel normally permits only one, or, occasionally, with an obviative system, two third person referents.

2 Some interesting relationships between the rhythmic nature of speech and that of music can be found in the work of many of the metrical phonologists although they are operating under a different framework.

Bibliography

Arcangeli, Diana. 1984. Underspecification in Yawelmani
Phonology and Morphology. Doctoral Dissertation. MIT.


Kelso, J.A. Scott, Betty Tuller and Katherine S. Harris. A "dynamic pattern" perspective on the control and coordination of movement. in MacNeilage.


MS. A Usage-based theory of language.


