

Phonemes: Can't Live With 'em, Can't Live Without 'em; Mostly the former

Why were phonemes invented?

It all starts with the simple question how to objectively represent data from languages, e.g. in a grammar. For millenia, people would simply use whatever writing system was used for the language. Thus a grammar of Hebrew written in 1528 would use the Hebrew alphabet which looks something like [תתרכ ל סבסמ] and a grammar of Norwegian written in 1821 would look something like [har bøndene kjøpte pær?]. This makes it impossible to report facts about an unwritten language. Being clever people, linguists simply used the spelling system of whatever languages they knew to write down a new language that they ran into, so if you're Danish or Norwegian and you hear a language with a vowel that sounds like the one in *rød* then you write [ø], and if you're Finnish, Swedish, German, Turkish, or Hungarian, you write [ö].

Some people, perhaps rightly so, objected that this amounted to defining unwritten languages in terms of the major European languages, not in their own terms. People used to be strongly influenced by grammatical categories for describing Latin, so a grammar in the old days would say how the genitive case is expressed, even if there is no such morphological category in the language. Especially under the influence of Boas, American field workers took to using systems of analysis that they took to be defined “in terms of the language itself”, rather than imposing a priori concepts from the outside onto the language. And since this era coincided with the frenzied birth of logical positivism and behaviorism, great emphasis was placed on developing objective, mechanical criteria for analysis, eliminating bias imposed by the analyst.

A huge stumbling block to this program is the objective rendering of speech. Recording technology in those days was nearly non-existent, and is also useless for communicating information about a language to others. There are a few archived (phonographic) recordings of language materials from the era — most are of such poor quality as to be uninterpretable, and in their day had no scientific use, because they could not be published or (for economic reasons) copied and distributed to other scientists. The only practical solution was to describe sounds via approximation, based on inferred physiological states, for example “[φ] is a bilabial voiceless frictionless spirant: just put your lips together and blow”. In principle, the symbols used to indicate these sounds could be anything, but by convention the symbols are given specific albeit general meanings.

The technology for studying the physiology of speech has improved a little over the past century, but still remains tricky enough that it is only rarely used in studying an unknown language, generally in limited doses by professional phoneticians. Lots of information can be gotten about the action of the articulators via introspection about tactile sensations in the vocal tract, knowledge of anatomy, and a few cute tricks. So for example one can use certain kinds of introspection about what's moving around the lips, coupled with visual information, to deduce that [o] involves rounding the lips. The implicit (and reasonable) assumption is that one can deduce a vocal tract configuration from the sound, if you have suitable training in the technique. Serious study of the acoustic properties of speech was technologically not possible for virtually the entire duration of the fieldworking era, and has been available in useful form to people doing language description for less than a decade (useful form being “cheap or free, easily useable software”: even the cheapest dedicated machines are horking and costs thousands of dollars). Fieldworkers know from experience that auditory impressions are necessarily subjective (they involve intro-

spective comparison of auditory experience with other auditory experiences acquired in life). A phonetic transcription is a supremely subjective entity, and quite at odds with the desideratum of total objectivity. It is unknown how actual individual field workers of the period resolved this philosophical conflict — most probably, the guy in the field did not care that much — but the general approach was to take the narrow phonetic transcription to be a convenient fiction not to be talked about too much in polite society. Apart from the fact that the details of the physical output were beyond serious scientific (mechanical) study, any comparison between languages where one declares “this entity in Kwakwala is identical to that entity in Dutch” introduces the ugly specter of universality and mentalism, and it is assumed as a matter of principle that phonetic properties across languages are at most similar, and can vary without limit.

From a modern perspective, the problem is clear. It is impossible to study the grammar of a language and learn anything about the language, just looking at an acoustic waveform and the associated semantic interpretation. Even the act of unifying 100 tokens of a given word into a generalization “this is the word [lintu]” is massively difficult. There is a massive amount of variation in speech, and no two utterances of a word will ever be the same, even if said by the same person. The only objective and accurate way to represent a given utterance is to represent it with a sequence of universal symbols such as “12”, “-329”, “4510” and so on, noted every twentieth of a millisecond or so (i.e. through a digital recording). Given enough compute power, disk space, and work on algorithms it may in the distant future be possible to train a machine to reduce speech from arbitrary languages to fewer discrete symbols (a narrow phonetic transcription) without knowing the analysis of the language in advance, but by that time everybody will be speaking either English or Chinese so we won’t really care.

The practical problem for phoneticians is that in order to look at something like voice onset time and whether there is a difference in the beginning of words in the timing of the onset of vocal fold vibration in English *p,t,k* versus *b,d,g*, you must know in advance that there are such things, and which words illustrate them. The identification of things to be studied phonetically therefore relies on a prior analysis, and now we have to ask what that analysis was based on (surely not just tradition or orthography, is it?). Well, it’s based on identification of the phonemes of a language, which is based on analysing the distribution of phonetic variants and looking for contrast and oh, is that my tail up ahead that I’m chasing?

Saved by the Mechanical Phoneme

The procedure-happy taxonomic structuralists came up with a way to make life possible albeit uncomfortable, which is to basically not care (not a totally bad solution). It was decided that the only things that could be studied scientifically were the entities that form the basis for making word distinctions, and, according to them, the thing that distinguishes words is not sounds, but *phonemes*. Their analytic school is based on multiple “levels of analysis”, which are strictly separated and each has its own alphabet. For things resembling phonology (they never got very far with anything else, though they made some starts on word structure), the three levels of representation are:

Phonetic level, containing phones	[p ^h lɪŋks]
Phonemic level, containing phonemes	/plɪŋks/
Morph(ophon)emic level, containing morphophonemes	pliŋk-Z

Each level maps to the adjacent level by unordered correspondence rules, so /p/ → [p^h] / . ___ meaning “a /p/ phoneme corresponds to a [p^h] phone in syllable-initial position”. The alphabet of each level is autonomous, and thus the letter <p> at the phonemic level is not the same entity as the letter <p> at the phonetic level.

The symbols at the phonetic level, the phones (“sounds”), are traditionally assigned the standard IPA or other phonetic meanings. Sounds are mystical items, not to be discussed too much, unless you’re into physical stuff. We won’t say anything more about actual sounds, because it is too messy. Although, of course, people would have to say something about the phonetic level just so that the reader would have a basic understanding, and maybe not sound too goofy if they try to pronounce the language being described. As far as how they actually describe sounds, they typically do what impressionistic phoneticians do, which is describe as much stable detail as they can by repeatedly listening, and they write it down with discrete symbols that have widely agreed-upon interpretations. The major mysterious leap in relating a taxonomic grammar to actual speech is mediated by the phonetic transcription. Once you have the sounds on paper, in some fashion, everything is excruciatingly rigorous and scientific.

The elements of the phonemic level, the phonemes, have no phonetic value. “Phonetic value” is a property of the phonetic representation, not the phonemic representation. The phoneme is an analytic grouping of phones, which allows one to state the distribution of phones more compactly and generally. A phoneme is a label standing for a set of phones, and the specific member of the set which the phoneme maps to is determined by the context in which it occurs. The collection of phones which realise a particular phoneme are called the *allophones*. The relation between a phone and its phoneme must be *biunique*, meaning that there is exactly one phone in a given physical environment which represents any phoneme, and exactly one phoneme which is the cover symbol for a given phone.¹ In other words, /p/ can stand for [p^h] syllable initially, and no other phoneme can also stand for (become) [p^h] syllable initially. If [p^h] is an allophone of /p/, then it may only “come from” /p/, and only in the specified context(s). Though, the context would be a disjunction of factors. In short, when you see [p^h] you immediately know it comes from /p/, and when you see /p/ then (because of the context) you immediately know that it becomes [p^h]. These relations are expressed as context-sensitive *phonological rules*.

The principle of complementary distribution is essential to partitioning phones into coherent sets (phonemes). The idea is to group the phones into subsets, such that no two members of a subset ever appear in exactly the same physical environment. If two phones ever do appear in the same environment, they cannot be allophones of the same phoneme. Instead, they illustrate the property of *contrast* (and, allophones of the same phoneme cannot contrast). The typical way to illustrate contrast is via a minimal pair such the distinct words [pʌn] vs. [bʌn]. The reason why such examples make a knock-down argument for contrast, hence different-phoneme status, is that if they did involve the same phoneme, there can (definitionally) only be one phone that appears in that context (they can’t be in free variation because selection of [p] versus [b] is not free, but instead indicates selection of a specific word).

¹ An unpleasant complication is introduced by free variation, where a given phoneme can be pronounced freely in any of two or more ways, in a given environment. When you have free variants, for example if [s] and [θ] are interchangeable, the principle is that you can always select either one phone or the other, and that the choice of one versus the other never signals a different word.

If minimal pairs are not available, a plausible but non-probative argument can be made based on near-minimal pairs. For example, if the closest you could get to minimal pairs for [p] vs. [b] in Gwambomambo is:

[palk] [apo] [kepsi] [gilap] [uport] [poit]
 [bark] [abe] [kebši] [gülāb] [ubork] [bout]

then the claim of phonemic status would not automatically be rejected, but allophonic status cannot be conclusively disproven because the subtle contextual differences could be included in the rule, i.e. “/p/ becomes [b] before __ark#”. Before you laugh this possibility out of the room, remember that the only concern of the taxonomist is stating objective and mechanical procedures of analysis, which could in principle be performed by a machine. Plausibility is an undefined concept for them (since it implies a judgement on the part of the analyst).

One further fact about phonemic analysis: it is not optional. That is, these are both the necessary and the sufficient conditions for partitioning phones into phonemes. Hence one could partition the set of phones of English {p,t,k,p^h,t^h,k^h} into the phonemes {/p/,/t/,/k/,/p^h/,/t^h/,/k^h/}. In addition, one can partition the phones into {/p/,/t/,/k/}. Only the latter solution is allowed, and this is guaranteed by the principle of economy of phonemes, which requires an analysis to minimize the number of phonemes.

The morphophonemic level is another level of analysis, which groups together phonemes. The requirement of biuniqueness is not imposed on this level, which is how phonological neutralization is dealt with. In the case of German final devoicing, where generative underlying *d* becomes [t], /t/ and /d/ are separate phonemes and therefore the mapping between /d/ (actually [d], using the typical symbol for a morphophoneme) and [t] cannot be done by a phonological rule, since phonological rules can only produce allophones from phonemes: a phonological rule cannot have a phoneme as an output, it can produce only an allophone. Since you cannot know from the physical environment whether [bunt] is from /bund/ or /bunt/, the rule covering devoicing in German cannot be phonological, therefore it is *morphophonemic*. A morphophonemic rule is one that maps between morphemic and phonemic representations. The adjective “federal” would be [bund] and the adjective “colorful” would be [bunt]. There is a morphophonemic rule [d] → /t/ / __# in German, also something like [Z] → /s/ / {p,t,k}__ and [Z] → /z/ / {b,d,g,m,n,l,r,a}__ in English. Morphophonemes, especially “process morphophonemes”, can be any symbol. Thus, [snV*-r-dek-a] in Klamath is phonemically /snedadka/ which is pronounced [snetət^ha]: “V*” mean “copy of the next vowel” and “r” means “copy the following syllable”. These symbols are purely arbitrary conventions created by the author of the book, and can be replaced with any symbol that you wish.

Flapping in English is a problem case. Because you get neutralization of the medial consonant in “rider” and “writer”, the rule cannot be phonological (biuniqueness fails). There must be a flap phoneme, /r/, and a morphophonemic rule which turns the morphophonemes [d] and [t] into /r/ in the relevant context. But everybody knows that /r/ is not a phoneme of English. Squirm.

The distribution of [h] and [ŋ] in English is a problem case. The fact is that [h] only appears syllable initially and not in the context [v__v], and [ŋ] only appears postvocally at the end of a syllable or in the context [v__v]. Therefore *h* and *ŋ* are in complementary distribution. Therefore they are allophones of one phoneme? Live by the sword, die by the sword.

Mandarin Chinese palatals are a problem case. Before non-high and back vowels there is a contrast between k , t^s and \check{c} , hence each is a phoneme. None of these can appear before front high vowels. Furthermore, the palatal affricate $[t^c]$ can only appear before front high vowels. In other words, $[t^c]$ is in complementary distribution with $/k/$, as well as with $/t^s/$ and $/\check{c}/$. It must therefore be reduced to being an allophone of one of these segments, and only one, but which segment it is an allophone of is completely arbitrary. Any theory which forces one into arbitrary and unjustified statements is undesirable, qua theory.

Russian voicing assimilation is a problem case. The Russian obstruent system has some interesting gaps in voicing: $/x/$, $*/\gamma/$; $/\check{c}/$, $*/\check{j}/$; $/t^s/$, $*/d^z/$. There is a rule which assimilates the voicing of obstruents to that of an immediately following obstruent, so $/kd/ \rightarrow [gd]$, $/bk/ \rightarrow [pk]$. By this rule $/x/$, $/\check{c}/$ and $/t^s/$ become $[\gamma]$, $[\check{j}]$ and $[d^z]$ before voiced obstruents. The former rule is neutralising (t becomes d and d becomes t) and therefore must be morphophonemic. The latter rule is non-neutralising (these sounds arise *only* because of this rule, and are in complementary distribution with their brethren), so must be phonological. Thus the theory must be rejected because it cannot express this process, which common sense tells us is one single process, as one rule. This is the “celebrated Halle argument against the taxonomic phoneme” (juice and cake will be served afterwards).

Phoneme as mental unit

The preceding explains the view of the American taxonomic structuralists. European structuralist tradition is another matter. The term “phoneme” was actually introduced by Baudouin de Courtenay (a Polish linguist with a French name teaching in Tataristan in the 19th century). I don’t know or care what his theory of the phoneme was, but it was certainly not the mechanist ~ taxonomist view. Another view of “phoneme” is that it is a mental unit (which was an impossibility in the anti-mentalist taxonomic structuralist view). Thus what unifies $[t]$ and $[t^h]$ in English is that speakers “think of them as being the same”. The main problem with this approach is that it has no general predictive basis for the language analyst, that is, you can’t tell if two phonetic objects are unified as a phoneme, except to perform a psychological test on speakers. The problem is, there is no independent test of the validity of these tests: we don’t know what speakers are basing the judgements on; we don’t even know how to conduct the experiment. Typically, to do an experiment, a standard taxonomic phonemic analysis is assumed and non-robust phonemes are filtered out. There is actually very little objective evidence to give credence to the claim that phonemes have a special cognitive status. While students coming into 201 are often unaware that English $[p]$ and $[p^h]$ are objectively different and that $/p/$ and $/b/$ are different, one might suspect orthographic influence a bit (or, a huge amount). This section is short: it amounts to saying “nice idea, where’s the beef?”.

The Generative Phoneme

This section is really short. There isn’t any. The closest you’re gonna get is “underlying segment”. The phoneme has no status in generative phonology. Except that people tend to forget history, so phonemes have a way of creeping into usage; but they have no status in the theory. They came back to life for a few years in some versions of Lexical Phonology; they pop up now and again in terms of “Contrast Preservation” in Optimality Theory. We’re still not sure what a “contrast” is.

Allophone as Fiction

It is not at all clear that there is such a thing as an “allophonic rule”, as part of the phonology or “symbolic grammar”. Consider the word “school”. Phonetic descriptions may say that the *k* is rounded, so phonetically you have [sk^wuwl]. Even more fine-grained descriptions will tell you that the initial *s* is also somewhat rounded, not as much as the *k*. Another description of the facts is that lip protrusion required for the vowel *u* precedes the actual onset of the vowel *u*, beginning somewhere around where the *s* is. Now, what the heck is the difference between this, and saying that you have a semi-rounded *s* and a round *k*? It comes down to the theory of properties (features, forthcoming), and whether the concept “semi-round” is legitimately part of phonology. The acid-test of validity for distinctions in generative phonology, as mediated by the theory of distinctive features, is whether it is possible to represent such a contrast. Thus if [u] is [+round] and [k^w] is [+round], what can half-round *s* be, other than [-round]? Binary feature representation thereby limits the possible phonological segments. That means that it is actually impossible to describe this process phonologically.² Similarly, it has been noted that velars become fronted in English before front vowels, cf *keep*, *kill* etc. What is not widely commented on is that the actual degree of fronting varies as a function of the vowel, so that fronting is maximal with *keep* and minimal with *cap*. The best way to describe this situation is to simply say that the tongue body is anticipatorily placed in the actual position of the following vowel; this clearly is not a phonological description.

Numerous examples of this nature can be found, accounting for a huge portion of the set of assumed allophonic rules. In a few cases, experimental data has been brought out to show that a more accurate description of the facts is achieved by abandoning the assumption that there is a categorical shift from one state to another, and assuming instead that the property (such as the frontness of velars) is due to processes outside of phonology, namely phonetic implementation.

The question immediately becomes, what are the necessary and sufficient conditions for a process to be deemed phonological versus phonetic, meaning specifically “dealt with by categorical symbolic operation a.k.a. phonology” versus “dealt with in terms of whatever continuous physical functions define phonetic implementation”? Skirmishes over this continue, but certain territory can clearly be carved out — if you assume that there is a difference between phonetics and phonology in the first place (SPE denies it, rendering all phonetics part of the phonology, and certain phoneticians such as the Articulatory Phonologists deny it by subsuming all phonology under phonetics). Phonological rules operate in terms of a small set of properties which can have at most two states, i.e. [+voice] or [-voice]. There is essentially no way to represent time in phonology, except via precedence (and, as we will see later, domain of association and number of skeletal positions). The concept “halfway into the vowel” is beyond the reach of phonology, as is “40 msc”. So too is any reference to actual physical states, except as mediated by the features. There is no feature that informs us about the jaw, therefore any operation best understood in terms of movement of the jaw must be phonetic.³ On the other side, it is at least widely as-

² There is an escape clause, but not a good one. SPE feature theory actually anticipated this problem and introduced scalar feature values, not just binary ones. There are apparently no living adherents to this viewpoint.

³ One has to be careful not to engage in sophistry when arguing for jaw-sensitivity. You could construct an argument that the jaw *could* be the crucial factor, but this does not mean that the process *must* be described in terms of the jaw.

sumed that phonetic implementation is “blind” in the sense that it cannot refer to lexical or morphological properties, or anything other than the actual physical material around a given sound. It is also somewhat assumed that phonetic implementation cannot neutralise distinctions between sounds, and thus a rule neutralising /zap/ and /zab/ to [zap] could not be phonetic. However, this restriction is arbitrary, and we could say that it doesn’t follow from the nature of the theory of phonetic implementation (except, there is no such theory, though there may be many individuals who would agree with that restriction). So unfortunately, it is not clear that there is anything that *can’t* be handled by phonetic implementation, and if we can’t objectively partition the class of processes into relatively neat piles, we’re back where we started from.