

Problems with Phonemes

Notes for a lecture given to the Cambridge University Linguistics Society, 16th November 2006

John Coleman, University of Oxford Phonetics Laboratory,
with the assistance of graduate students Ekaterina Samoylova, Anastassia Loukina, Melissa Fox, Will Draper,
‘Ōiwi Parker Jones and Karen Park

1. Disagreement about definitions

1.1. Daniel Jones, *An Outline of English Phonetics*: “a family of sounds consisting of an important sound of the language (generally the most frequently used member of that family) together with other **related** sounds which ‘take its place’ in particular sound-sequences or under particular conditions of length or stress or intonation.” (Emphasis added.)

1.2. Cercle Linguistique de Prague (1931) Projet de terminologie phonologique standardisée. *Travaux du Cercle Linguistique de Prague* 4. 309-323.

Phonème — Unité phonologique non susceptible d’être dissociée en unités phonologiques plus petites et plus simples.

Unité phonologique — Terme d’une opposition phonologique quelconque.

Opposition phonologique — Différence phonique susceptible de servir, dans une langue donnée, à la différenciation des significations intellectuelles.

A more abstract view, allowing such realisations as Fang /u/ → [ɤ], Mandarin Chinese /i/ or /i/ → [ɹ]. (J. Kelly (1974) Close vowels in Fang, *BSOAS* 37 (1), 119-123; F. Dell (1994) Consonnes à prolongement syllabique en Chine, *Cahiers de Linguistique Asie-Orientale* 23: 87-94.)

1.3. Letter CXLIX from Trubetzkoy to Jakobson, 3-4 August 1935, concerning his impressions of the Second International Congress of Phonetic Sciences (University College, London, June 1935). In R. O. Jakobson, ed. (1975) *N. S. Trubetzkoy's Letters and Notes*. Mouton. English translations by Marvin Taylor and Anatoly Liberman in N. S. Trubetzkoy (2001) *Studies in General Linguistics and Language Structure*.

“After the farewell banquet, all kinds of diversions were organized; that is, some members of the congress made jocular speeches, sang songs, and the like. Every time the word *phoneme* turned up, it aroused an outburst of universal laughter. Horn composed a poem in Middle English on the themes of the congress. It ended in the following couplet:

wat is phonemes, wat is sunds?
twelf men haf twelf difinituins.

This couplet was later quoted by all and was rewarded with loud applause.”

2. Problem 1. Neutralization

2.1. N. S. Trubetzkoy (1935) *Anleitung zu phonologischen Beschreibungen*. Prague: Association Internationale pour les études phonologiques. [Translation: *Introduction to the Principles of Phonological Descriptions*, by L. A. Murray. The Hague: Martinus Nijhoff, 1968.]

Section 21. “For example, in many languages, the phonological contrast between voiced and voiceless consonants exists only before vowels and sonants while before obstruents voiced or voiceless quality is automatically regulated by the nature of the following sound and is therefore phonologically irrelevant.”

2.2. e.g. English tʃ, dʒ (but not tʒ, or dʃ) or (the mirror-image) cat[s] vs. dog[z]. Cf. [kɫab] *club* vs. [glav] *glove*.

2.3. Trubetzkoy (1933) *Charakter und Methode der systematischen phonologischen Darstellung einer gegebenen Sprache. Archives Néerlandaises de phonétique expérimentale* 8-9: 262-65. [Translation: 'The Systematic Phonological Representation of Languages', Taylor & Liberman pp. 12-13.]

“The dissolution (neutralization) of phonological oppositions is the most important kind of restriction of the use of phonemes. The resulting neutralized phonemes are treated by the linguistic consciousness as special phonemes and are represented in many writing systems by special characters. This perception of neutralized phonemes is sometimes expressed in their phonetic realization as well, so that they are pronounced as intermediate sounds between the two poles of a correlation ...

[Solution 1: rules/constraints on distribution] By the *target* of a sound rule I mean the position in which it operates (before consonants, between vowels, word initially, and so on). The same position can serve as the target for several rules. The more rules that share a given target point, the fewer phonemes that can occur in that position. In this way, positions of *minimal* and *maximum phoneme differentiation* can be identified for most languages. Indeed, *local phoneme inventories* can be drawn up for every position that serves as the target for at least one rule. The distribution of the rules may be such that the individual phoneme inventories do not correspond (e.g., those of word-initial and word-final position) and there is no position in which all phonemes of the language occur, **so it is sometimes advisable to dispense with a general phoneme inventory and work only with local ones.**” [= Solution 2]

2.4. Solution 2: polysystemicity – different systems of contrast in different places

Trubetzkoy to Jakobson, 17 May 1935 (*N. S. Trubetzkoy's Letters*, pp 334-5; Taylor & Liberman pp 246-8.)

“Twaddell's book [Twaddell 1935] is interesting but rather sterile. He attempts to give the phoneme a definition that would be free from metaphysics and ends up with the denial of phonological systems. In principle, he operates with our definition of the phoneme as *terme d'opposition* but draws an absurd conclusion: since in a given position only some, rather than all, phonemes are opposed, a special system has to be set up for each position (my *Teilsysteme!* ['partial systems']), but these systems are allegedly incompatible (it is not clear why); consequently, a unified system of phonemes does not exist in a language.”

2.5. J. R. Firth (1935) Phonological features of some Indian languages. *Proceedings of the Second International Congress of Phonetic Sciences*. Reprinted in J. R. Firth, *Papers in Linguistics 1934-1951*, 47-53.

“It may be practically convenient to think of a language as having a sound system, or phonetic structure as a whole, but this is little more than a list of all the possible alternances of sound in all contexts. This is an 'all-over' list, and not to be confused, for example, with a specific context of maximum alternance for a certain class of sound, e.g. the context in which the maximum number of plosives may alternate in English. ...

Let us take Marathi, for instance. In initial position only two nasal consonants can be used, **n** and **m**. In final position there is a three-term nasal alternance, but immediately preceding another consonant, especially stops, only one is possible, the nasal homorganic with the following consonant. Of these I have noted at least eight: alveolar **n** before **ts**, dental **n** before dental **t**, **d**, and similarly retroflex **ɳ**, palato-alveolar nasal, velar nasal,

bilabial, labio-dental, and a sort of nasal $\tilde{w}$. In transcription I should use the letter **n** to symbolize the initial **n** in contradistinction from **m** only, also for the final **n** which functions in a three-term alternance and again for the specific unique homorganic nasal on-glides before **ts**, **t**, **d**, and perhaps even all the rest. **But I should not thereby identify all these n sounds as linguistically and functionally the same element or unit. Indeed, they cannot be so identified.** The homorganic nasal on-glides to the stops could be represented by a separate nasal symbols such as $\sim$, but that is not really necessary. Surely we are free to use the same letter without being compelled to concoct a rationalized 'derivation' from the letter in the shape of a phoneme theory. Similarity of sound is no safe guide to functional identity, though it may serve as the basis of practical transcription symbols. There are not eight 'nasal' phonemes in Marathi. We might possibly say there are three, though I would prefer to say that the sounds we symbolize by **n** and **m**, for example, constitute the total nasal alternance in initial position, but only two out of three terms in final position, whilst they both serve also as unique homorganic on-glides.

The actual mechanism and act of utterance of **n**, for example, in each of the three cases would be different. They would actually be slightly different sounds, and their minor function [i.e. function in differentiating lexical items] would be different. Though writing them with the same symbol on practical phonetic grounds, I should not identify them in any other way. **That they are the same 'phoneme' is the very last thing I should say."**

Marathi is not particularly exotic: in English, too, we have initially only /m/ and /n/, final /m/, /n/ and /ŋ/, and pre-consonantly, a more extensive range of places of articulation: /mp/, /mʃ/, /nθ/, /ns/, /ntʃ/, /ŋk/.

2.6. Problems for Firth: (i) resyllabification, e.g. *rhythm* vs. *rhythmic*; (ii) ambisyllabicity, e.g. *whiten* vs. *widen*.

2.7. Solution 3 (almost the same as polysystemicity): Features and underspecification.

For example, plural /s~z/: [+consonantal, -vocalic, -sonorant, +continuant, -labial, +coronal, -distributed], without any mention of [±voice] – so this feature matrix does not represent a phoneme. Such underspecification leads us inevitably to the idea of a cluster of just a few features (e.g. place features such as [labial, coronal] spreading as a group):

$$\begin{array}{c} [+nasal] \quad k \\ \quad \backslash \quad | \\ [-labial, +coronal] \end{array}$$

Or even single features functioning relatively autonomously (e.g. [+nasal], above, or [±voice] in plurals).

3. Problem 2: Never-contrasting phonemes, e.g. /h/ and /ŋ/ in English. The standard 'solution' to this problem makes reference to /h/ and /ŋ/ being "too dissimilar", without explaining how or why. Cf. Germanic *brāhta > brāhta [brāhta] > OE [brōhte], brōhte > Modern brought. Likewise wring/wrought.

If we are not bothered about phonemes in complementary distribution, why worry about allophones in complementary distribution? 'Penguins vs. polar bears'.

4. Problem 3: Non-uniqueness of phonemic analyses

Simple examples: (i) /t/ vs. /d/ in English: why not /th/ vs. /t/? (ii) /tʃ/ — one phoneme or two?

4.1. Yuen-Ren Chao (1934) The non-uniqueness of phonemic solutions of phonetic systems. *Bulletin of the Institute of History and Philology, Academia Sinica* Vol. IV, Part 4, 363-397. Reprinted in E. P. Hamp, M. Joos, F. W. Householder and R. Austerlitz, eds. *Readings in Linguistics I & II*.

Standard Mandarin	i	y	u	ɨ	ʉ	a
	Front		Back			
Velars			ku, k ^h u, xu		kʉ, k ^h ʉ, xʉ	ka, k ^h a, xa
Prepalatals	tɕi, tɕ ^h i, ɕi tɕia, tɕ ^h ia, ɕia	tɕey, tɕ ^h ey, ɕey				
Retroflex			ʈsu, ʈ ^h u, ʂu	ʈʂz, ʈ ^h ʂz, ʂʂ	ʈʂʉ, ʈ ^h ʂʉ, ʂʂ	ʈsa, ʈ ^h a, ʂa
Alveolars			tsu, ts ^h u, su	tsz, ts ^h z, sz	tsʉ, ts ^h ʉ, sʉ	tʂa, tʂ ^h a, sa

5. Problem 4: Marginal/loan phonemes (or are they?) e.g. [x] in Komi, Cheremis, English *loch*, *Bach*, sometimes. [g] and [ʃ] in Finnish, etc.

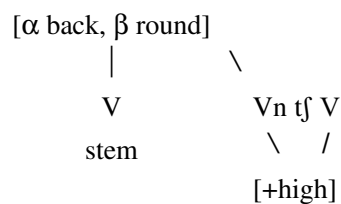
Charles C. Fries and Kenneth L. Pike (1949) Coexistent phonemic systems. *Language* **25**, 29-50. /t/ and /d/ are in complementary distribution in Mazateco (/d/ before nasals), but distinct in loanwords e.g. *siento*.

6. Problem 5: Non-locality/punctuality of certain phonological contrasts

6.1. Kirgiz vowel harmony. The challenge for phonemic theory is to explain the cooccurrence restrictions on vowels in separate syllables, the variations in the suffix, and the dependence of those variations on the stem vowels.

	[-back]		[+back]	
[-round]	bir + intʃi	'first'	atti + ntʃi	'sixth'
	bef + intʃi	'fifth'	dʒijirma + ntʃi	'twentieth'
[+round]	ytʃ + yntʃy	'third'	on + untʃu	'tenth'
	tørt + yntʃy	'fourth'	toguz + untʃu	'ninth'

6.2. Solutions: features/long components/prosodies/autosegments. (Zellig S. Harris (1944) Simultaneous components in phonology. *Language* **20**, 181-205.)



7. Problem 6: Elimination of the phonemic level in generative phonology

7.1. Phonemes $\Rightarrow$ allophones; morphophonemes $\Rightarrow$ phonemes: *tax*, *tacks* {^htæks, t^hæk+s} $\Rightarrow$ /t^hæks/

These two types of derivation define three levels: morphophonemic $\Rightarrow$ phonemic $\Rightarrow$ allophonic.

7.2. Morris Halle (1959) *The Sound Pattern of Russian* section 1.3. p. 22: “In Russian, voicing is distinctive for all obstruents except /ts/, /tʃ/ and /x/, which do not possess voiced cognates. These three obstruents are voiceless unless followed by a voiced obstruent, in which case they are voiced. At the end of the word, however, this is true of all Russian obstruents: they are voiceless, unless the following word begins with a voiced obstruent, in which case they are voiced. E.g. [mɔk li] “was (he) getting wet?” but [mɔg bi] “were (he) getting wet”; [zɛtʃ li] “should one burn?” but [zɛdʒ bi] “were one to burn”. Implies distinct morphophonemic and phonemic assimilation rules that are otherwise identical.

Morphophonemes	{mɔk li}	{mɔg bi}	{zɛtʃ li}	{zɛtʃ bi}
	↓ Assimilation		{tʃ} does not have a voiced cognate	
Phonemes	/mɔk li/	/mɔg bi/	/zɛtʃ li/	/zɛtʃ bi/
	{l} is not an obstruent		↓ Assimilation	
Allophones	[mɔk li]	[mɔg bi]	[zɛtʃ li]	[zɛdʒ bi]

7.3. **Solution:** Halle's generative alternative:

Morphophonemes	{mɔk li}	{mɔg bi}	{zɛtʃ li}	{zɛtʃ bi}
	↓ Assimilation		↓ Assimilation	
	<i>(Indefinitely many) intermediate representations</i>			
	↓ other rules		↓ other rules	
Allophones	[mɔk li]	[mɔg bi]	[zɛtʃ li]	[zɛdʒ bi]

8. Defending the phoneme?

Sanford A. Schane (1971) The phoneme revisited. *Language* 47, 503-521: surface contrast implies phonemes?

“**The phoneme as a viable phonological unit for capturing relevant surface contrasts has been rejected within generative phonology.** However, certain diachronic and synchronic effects can only be explained in terms of the phoneme. The historical phenomena of ... are due to surface contrast rather than to storable morphophonemic or phonetic processes. From these observations I conclude that the phoneme must be recognized as a phonological entity. Synchronically, a phonemic representation is not to be discovered by applying a set of procedures to a phonetic representation, nor does it exist as an autonomous level within a generative phonology; rather, it is to be characterized as **a representation of the relevant surface contrasts** ... Thus the phoneme does in fact have a place in generative phonology ... So long as generativists do not generate a narrow phonetic representation, their rules will generate explicitly a broad phonetic representation, which, implicitly, is a representation of surface contrasts. It can be no mere coincidence, then, that the output of a generative phonology is so often almost amazingly identical to a classical representation.”

Or perhaps not so amazing, when we acknowledge that very many published generative phonological analyses are actually re-worked from earlier classical phonemic descriptions.

9. Problem 7: Phonemes and alphabetic writing.

9.1. Sometimes people point to the existence of alphabetic writing as evidence for the mental reality of phonemes. But we can turn the question round and ask: of all the times that writing systems have been created, why has alphabetic writing only been discovered once or twice – and then only imperfectly phonemic – whereas syllabic/moraic writing systems are far more frequent (= have been independently discovered many times)? And, after the invention of writing, why did it take *over a thousand years* for alphabetic writing to be “discovered”.

9.2. If phonemes are *real* units of lexical representations, why is it that evidence for phonemic awareness is far stronger in alphabetic literates than in illiterates, or non-alphabetic literates (such as literate Chinese)?

José Morais (1985) Literacy and awareness of the units of speech: implications for research on the units of perception. *Linguistics* **23**, 707-721.

10. Persistence

Pinker (1994) *The Language Instinct*: “morphemes, in turn, are built out of phonemes. ... a phoneme corresponds to an act of making a sound.”

Jurafsky and Martin (2000) *Speech and Language Processing*: “We generally capture ... pronunciation variation by positing an abstract class called the **phoneme**, ... realized as different **allophones**”

Harley (2001) *The Psychology of Language*: “A **phoneme** is a basic unit of sound in a particular language.”

Ashby and Maidment (2005) *Introducing Phonetic Science*: “a phoneme being one of the basic phonological units of which all words in the language are composed.”

Why are phonemes retained in elementary pedagogical treatments of phonological theory?

- 1) The idea is easy to understand (for alphabetic literates). It is a simple model ... *at first*.
- 2) They are useful for certain purposes (e.g. pronouncing dictionaries) – convenience of the fiction.
- 3) They were a historically important construct, so it is hard to *not* talk about them.
- 4) The alternatives (feature theory, autosegmental & prosodic structure) are harder for beginners to understand. Maybe.
- 5) Laziness?
- 6) Ignorance of the older literature. (Short-sighted textbooks?)

Trask (1999) *Key Concepts in Language and Linguistics*: “the distinctive features ... were increasingly regarded as the true fundamental units of phonology. Since the 1980's, this trend has continued to the point at which many phonologists now work exclusively with features, and ignore phonemes altogether – though an understanding of phonemes is still considered essential in most introductory linguistics courses.”