122

PHONOLOGICA 1984

Proceedings of the Fifth International Phonology Meeting
Eisenstadt, 25 — 28 June 1984

Edited by

Wolfgang U. Dressler, Hans C. Luschützky,
Oskar E. Pfeiffer and John R. Rennison

Cambridge University Press
London New York New Rochelle
Melbourne Sydney
1987

Explanation in Phonology: Opinions and Examples

John J. Ohala

INTRODUCTION

There has recently been much commentary on the questions of whether it is possible to explain the sound patterns in language and, if so, what form those explanations will take — in particular, what facts and principles will be referred to from which the sound patterns follow; see Dinnsen (1980), Ladefoged (1980, 1983). Lass (1980), and Anderson (1981). I appreciate these writers' views and have learned much from them, but I find them much too pessimistic. In this paper I will offer what I believe to be a more hopeful alternative.

ARGUMENTS AGAINST EXPLANATION IN PHONOLOGY

Ladefoged questions our ability to account for certain language-specific phonetic facts by reference either to abstract psychological or physical phonetic entities. He rather endorses the notion that 'many interesting linguistic observations can be made in formal terms, independent of any other data' (1983:89).

Anderson allows that some aspects of phonology can be and have been explained by reference to phonetic and other facts, e.g., psychological, social, but doubts that all matters of interest to the phonologist, especially the very important ones, can be satisfactorily understood by reference to data which he regards as "external" to phonology.

Dinnsen accepts that phonetic explanations may be possible in principle but does not accept that any of those offered in the current literature meet the criteria for "true" explanations, which, according to him, must be deductive nomological (D-N) explanations, that is, law-like statements of cause and effect deduced from previously known facts and principles.

Lass goes farther than the others by arguing that D-N explanations, which every self-respecting scientific discipline is said to strive for, are not possible in *principle* when it comes to accounting for language change. This hits phonology 'right where it hurts' because sound change is one of the major areas in which explanations have been offered. Dinnsen and Lass present similar arguments for their positions: given a language exhibiting a certain sound change,[1] is it possible, they ask, that a D-N explanation can be constructed, according to which the sound change *had* to happen? That is, can all

[1] Or, for Dinnsen, a phonological rule, which amounts to the same thing since rules are generally necessitated by alternations created by sound change.

the necessary and sufficient conditions for the sound change be identified? The answer, of course, which no one would dispute, is 'no'. Under what seems to be identical circumstances, one language will undergo a certain sound change and another will not.

COUNTERARGUMENTS

I agree with Ladefoged that there are facts about speech behavior which are not yet explainable in any way and especially not by many of the currently fashionable phonological models. I also grant that we should not cease to make purely descriptive statements about any aspect of language structure even though we don't yet have an explanation for it. What strikes me as unfortunate about Ladefoged's views is that he seems to take the present primitive state of affairs in phonology as a reason to avoid attempts at explanation and instead to focus almost exclusively on pure description. I think that current research in phonology gives us reason to be optimistic that explanatory accounts for sound patterns not only are possible but are presently at hand.

Neither Ladefoged nor Anderson have demonstrated — à la Gödel and Heisenberg — any inherent block to our ultimate understanding of sound patterns in physical, psychological, and/or social terms. Rather, they simply present various examples — case studies — which they believe resist explanation by reference to things outside of phonology[2]. There is no proof that the cases they cite will resist such explanation forever. There is therefore no basis for debate here, except, perhaps, to show that some of their cases can be explained by principles they would consider 'outside' phonology and to point out certain weaknesses in argumentation.

Anderson, for example, cites the case of alternations in Fula, such as those in (1), as providing 'the crucial evidence... that phonological distinctions do not always correspond directly to phonetic observables' (1981:502).

(1) /war/ "kill" ~ /-bar/ ~ /mbar/
/war/ "come" ~ /-gar/ ~ /ŋgar/

He claims that these alternations justify setting up two underlying /w/'s, one [+anterior] and the other [−anterior]. He dismisses as irrelevant the account which claims that these alternations are the residue of historical processes which *did* have a phonetic basis at their origin and which are *now* maintained in the language as fossilized morphological variants (i.e. where original /b/ and /g/ in word initial position both shifted to /w/ due to their being 'grave' — having a low second formant; see Ohala & Lorentz, 1977; Ohala, 1979). In support of his contention he points to the 'internal coherence'

[2] I do not accept this division of the universe into 'inside' and 'outside' phonology but I will not make an issue of it here; see Ohala (1983b, in press a).

of these alternations and their apparent stability (although no independent evidence is cited for this point). The same, of course, could be said about virtually *any* alternation, e.g. those which resulted from Grimm's Law, Verner's Law, etc., as in English *father/paternal, foot/pedal, brother/fraternal, tooth/dental, heart/cardiac, horn/cornea,* etc. This is just what is meant by the term 'alternation', i.e. a regular correspondence of sound in different environments. Thus Anderson's argument boils down to this: the evidence for his analysis of the alternation is the fact of the alternation itself. In other words, he begs the question; this and similar cases he covers are therefore without any logical force.

The error in Dinnsen's and Lass' reasoning is quite fundamental: they believe that D-N explanations exist. Outside of mathematics and logic, however, D-N explanations do not exist. Physics and chemistry are often held up to us as the paradigm disciplines capable of D-N explanations. Such, at least, is the impression created by philosophically naive text-book writers and by the physics and chemistry public relations effort. It does not affect my point that many physicists and chemists themselves have fallen prey to this same propaganda. But it is not true; these physical scientists do not literally discover the 'laws' of nature. I do not begrudge physics and chemistry and related disciplines the credit they richly deserve for penetrating the mysteries of the physical world and therefore taming it. But that accomplishment by itself in no way justifies characterizing them as 'the exact sciences'. Anyone who has spent some time doing physics or chemistry in the laboratory (even in secondary school) or reading the primary literature in these disciplines (as opposed to others' 'digested' accounts of scientists' work) will realize that strictly law-like accounts are not possible. As a way of arguing this point let us imagine the following Gedankenkampf between a linguist and a physicist.

The linguist is presented with an isolated island populated with a few hundred monolingual speakers and is challenged to predict how their language will change within a given period of time. As suggested above, the linguist will not be able to do this with any accuracy. At best, she could specify a range of possibilities for the language's future and rule out an even larger range of impossibilities. If we score the Kämpferinnen with a 0 if they are inaccurate and 1 if they are accurate, then the linguist would get a score of 0. Now the physicist is challenged to predict the path of a billiard ball on a billiards table. The physicist would insist on knowing beforehand a) *whether* the ball will be hit, b) if so, with what force and at what angle, c) that after the ball was hit it would not be subject to any 'outside' forces, e.g., movements of the table, strong air currents, etc., d) that the ball will be perfectly round, the rim of the table of known and uniform elasticity, etc., etc. Naturally, none of this can be specified or guaranteed; this contest is to take place in the 'real world' not in the pages of a textbook. The physicist will also fail her challenges and will also get a score of 0. So the contest ends in a draw. The moral of the sto is this: neither one can insulate their problems from all the unpredictab and uncontrollable causal factors that impinge on their respective univers of observation.

Since D-N explanations (predictions) are impossible in *any* scientific d cipline, phonologists need not feel discouraged that they cannot produ them. What phonology *can* achieve — and in this it does not differ from oth scientific disciplines — is *deductive probabilistic* (D-P) explanations whi include appropriate statements as to the limited degree to which they ho Lack of perfect knowledge of the universe also means that we will not be al to explain everything, even in D-P terms. This may seem to be a triv observation, but part of Dinnsen's and Lass's dissatisfaction with some of t recently offered phonetic explanations in phonology is that these accou cannot explain everything. This is an unattainable requirement — by a science — and therefore in no way invalidates what has been done so f Rather than focusing on what cannot be done in a given discipline, attent should be given to what it *can* do, in particular, the 'fruitfulness' of its exp nations — in the sense suggested by Louis Pasteur:

> "...the characteristic of erroneous theories is the impossibility of ever forese new facts; whenever such a fact is discovered, those theories have to be gra with further hypotheses in order to account for them. True theories, on the contr are the expression of actual facts and are characterized by being able to pre new facts, a natural consequence of those already known. In a word, the cha teristic of a true theory is its fruitfulness." [From Vallery-Radot, 1911 : 24

This is precisely what has happened in the mature sciences, e.g. phy Since pre-historic times, people had a great many questions about the work of the physical world — the motion of bodies, the transfer of heat, the ca of weather, etc. But it was during the scientific revolution in the 16th cent that the physical sciences experienced their first significant success in get answers to *any* of these questions — in this case, those about the motio bodies. This was done by a mixture of rational and empirical means. T work was motivated by a hypothesis, i.e. an act of reason, but at the h of their method were the great pains taken to get high quality eviden support of their explanatory hypotheses; high quality in the sense of b minimally subject to the extraneous distortions which dilute its evide value. Careful, controlled, observations — that is what experiments are — the way to minimize such distortions.

The same methods and, as it turned out, some of the same physical con showed their fruitfulness by eventually leading to the answers to ques on the nature and behavior of heat. But by no means have all of the ques about the workings of the physical universe been answered. We still d fully understand the origin of weather and everyone knows that it be predicted very accurately. Are there critics within physics who say the whole discipline is bankrupt and not a respectable science because it hasn't solved the problem of weather? Obviously not. Yet this is the essence of some of Dinnsen's and Lass' criticisms of currently offered explanations in phonology. Whether in physics or phonology, the successful (partial) solution of even a few problems and the *fruitfulness* of these solutions is enough to show that appropriate logical and empirical methods are being used.

CANDIDATE EXPLANATIONS

I would now like to consider in some detail some candidate explanations for sound patterns. I say 'candidate' to emphasize that there is no claim that these explanations are 'correct' because, as outlined above, absolute certitude turns out to be an elusive goal. We know from the history of science that theories once thought to be 'true' have been replaced or revised by subsequent theories. There is no reason to expect that whatever theories we propose today will escape the same fate. It is, to repeat, the empirical support and the fruitfulness of the theories, not their 'truth', which mark valid explanations.

Voiceless laterals revisited. I would first like to reply to Lass' criticism of an explanation proposed in Ohala (1974). I tried to explain there why Norwegian /sl/ became /ʃl/ whereas /sn/ remained unchanged. I suggested the reason was that [l̥] is acoustically similar to [ʃ] (this was demonstrated by comparing their spectra) such that the following stages in the change were plausible: /sl/>[sl̥l]>[sʃl]>/ʃl/. If the /n/ had become voiceless in the same environment, the voiceless nasal that would result would have had very low intensity and be quite unlike any oral fricative.

Lass faults this explanation (39–42) by pointing out that the voiceless lateral in Welsh is regularly interpreted by non-Welsh speakers in Britain as [f], [θ], or even [x] but rarely as [ʃ]. In Ohala (1974) my purpose was to cite phonetic reasons for the different behavior of [l̥] and [n̥] (and on the basis of this, the different behavior of /s/ before these segments). Lass raises a completely different question in asking why the preferred re-interpretations of Welsh [l̥] should usually be the weak fricatives [f] and [θ]. There are a number of possible hypotheses one might plausibly entertain (and which could be systematically investigated if anyone deems the issue to be of sufficient importance). First it should be obvious that the measure of 'acoustic similarity' or 'distance' between sounds is a continuous one; [f], [θ], and [ʃ] are all more or less close to [l̥]: for that matter [s] is certainly closer to [l̥] than, say, [r] or [ʕ], although more distant than [ʃ]. Also, which sound is confused with another depends on which of the many phonetic features which characterize the sounds the listeners take to be most important. [ʃ] and [l̥] have similar spectral structure, [l̥] and [θ] or [f] are similar in having relatively low intensity. Listeners ᵛ different language backgrounds, e.g., Norwegian and English, may att to different cues. There may therefore be no incompatibility between facts of the Norwegian case and those cited by Lass. Second, the in the two languages may be phonetically different by virtue of t appearing in different environments; it is not unreasonable to expect that Norwegian [l̥] appearing after an [s] may have more intensity than the w initial [l̥] in Welsh. Moreover, they may just be phonetically different in any environment; see Ladefoged, 1980 for evidence of crosslanguage phon differences). Third, there may be English-specific phonotactic constra which would bias English ears to interpret /l̥/ (phonetically [l̥͡l]) as the missible cluster [fl] and not [ʃl]; different constraints may have applied in case of Norwegian. There are many avenues to explore if one has the pati and resources.

Lass also offers as counterevidence the fact that /sl/ clusters show no dence of changing to /ʃl/ in English. As indicated above, this type of argu is based on the false premise that an empirical discipline can possess perfect knowledge of the universe as to be able to predict (or post-dict course of events.

It is worth mentioning that the interaction of [ʃ] and [l̥] or (what is pho cally much the same thing) [ɬ], is not limited to Norwegian; it also show in Chadic (Newman, 1977).

Given the limits of our current knowledge, Lass' data do not contr and therefore refute the account I presented 10 years ago (which was a repetition of the account given by Haugen, 1942).

Spontaneous nasalization. Normally, distinctively nasalized v derive from sequences of vowel+nasal consonant, as given in (2).

(2) Latin *ventus* > French [vã]; Sanskrit *danta* 'tooth' > Hindi

On occasion, however, nasal vowels (or sometimes nasal consonants) app words which never had a nasal consonant at any point in their histor gives a couple of examples.

(3) Sanskrit *sarpa* 'snake' > Hindi [sãp]; Prakrit *pahuccai* 'attai Hindi [pəhə̃tʃ]

These are cases of so-called spontaneous nasalization. As it happens, such cases appear adjacent to consonants characterized by heavy airflow voiceless fricatives, aspirated stops, affricates.

My colleague, Mariscela Amador, and I attempted to test a hypo that I had made earlier, that vowels produced with a slightly open g might mimic the acoustic and perceptual effects of nasalization (Ohala

1980). A slightly open glottis creates a *branch* off the main oral resonating cavity in much the same way that the nasal cavity constitutes such a branch. They create similar acoustic effects, namely lowering of the relative amplitude of the first formant and increasing its bandwidth (Fant, 1973 : 8; Fujimura & Lindqvist, 1971). Several previous instrumental studies had shown that high air flow consonants like [s] or aspirated stops have greater-than-normal glottal opening and that this condition is assimilated to some extent by neighboring vowels (Slis, 1970).

To see whether physiologically oral vowels might sound nasalized on those portions immediately adjacent to voiceless fricatives, we used digital means to create a series of steady-state vowels by repeating single periods from the ends of vowels before an /s/ and, for the control cases, before the lateral /l/ and the nasal /n/ (to make sure that this method of creating vowels from single periods did not by itself introduce spurious nasalization or remove nasalization actually present). We presented these vowels to listeners to judge the degree of perceived nasalization on a 7-point scale, where 1 meant 'completely oral' and 7 'heavily nasalized'. We also recorded velic elevation to make sure that the vowels near /s/ and /l/ were not physiologically nasalized, and we sampled oral air flow, a rough indicator of glottal area, to verify that only the vowel before /s/ had significant glottal opening.

The perceived nasalization from vowel stimuli made in this way from the syllables /bal/, /ban/, and /bas/ produced by one of the four speakers was 3.2, 5.2, and 5.4, respectively, thus confirming the hypothesis that portions of vowels abutting voiceless fricatives may sound nasalized. We concluded that sound changes showing spontaneous nasalization came about when vowels that 'sounded' nasalized, even though they were not, were re-interpreted and spoken with actual nasalization. (For further details, see Ohala, 1983b.)

Dissimilation: the listener's 'fault'. In a couple of previous papers I have presented a new account of dissimilation (Ohala, 1981, 1983a, in press b, c). Perhaps the most well known example of dissimilation is Grassmann's Law (so-called), as exemplified in (4), where the diacritic '..' marks breathy-voicing.

(4) Proto-Indo-European *b̤end̤- 'bind' > Sanskrit band̤-.

I suggest that *listeners* are responsible for dissimilation. Through experience they learn that certain speech sounds introduce extraneous distortions on adjacent sounds (i.e. due to assimilation). They therefore formulate corrective rules which factor out these distortions — so they can figure out the pronunciation intended by the speaker. Dissimilation occurs when a listener invokes such corrective rules inappropriately, that is, factors out

aspects of pronunciation that were intended features of the pronunciat
Dissimilation is thus a form of low-level hypercorrection. Given a form
as /bend/ I suggest that the breathiness spilled over onto the segments i
vening, yielding [bend], and that the listener misinterpreted the breath
at the beginning of the word as being a distortion caused by the segme
the end. The listener therefore factored it out.

That listeners have and use these corrective rules has been demonstr
experimentally. One such study that I conducted with some of my collea
(Ohala, Kawasaki, Riordan, & Caisse, forthcoming, as reported in O
1981) involved requiring listeners to identify synthetic steady-state vo
on a continuum between /i/ and /u/. Somewhere in the middle of the contin
listeners stop hearing /i/ and start hearing /u/. However, this crossover p
is further front when the vowels appear with flanking apical conson
(/s—t/) than with flanking labials (/f—p/). The listeners presumably k
that apicals (but not labials) cause the vowel /u/ to be fronted and there
accept as /u/ a more fronted vowel in the apical environment.

Other explanations of dissimilation often appeal to articulatory princi
e.g. Ladefoged (1983) (à propos of Grassmann's Law):

> Aspirated consonants ... are very costly in that they use considerable res
> tory energy. A word with two such sounds is ... an obvious candidate for pru
> in any attempt to reduce the overall effort required for an utterance.

This conception of dissimilation, however, fails to explain why in the majo
of cases it is the first of two similar sounds which suffers dissimilation, whe
I account for this by the fact that assimilation, which feeds dissimilatio
predominantly anticipatory (Javkin, 1979 : 74ff). Furthermore, Ladefog
account is difficult to reconcile with cases of dissimilation involving p
of articulation — see (5) —, vowel quality, and tone, where it does not se
possible to identify the 'costlier' of two sounds.

(5) Ancient Chinese *pjam 'diminish' > Cantonese pin

My account would predict that the only features which should be sub
to dissimilation would be those which are known to spread by assimilat
over adjacent segments. This would include the features in (6a), which incl
place of articulation, but not those in (6b).

(6) a.	b.
labialization	fricative-ness
pharyngealization	stopped-ness
retroflexion	affricate-ness
glottalization	
aspiration	
palatalization	
place of articulation	

By and large, these predictions are borne out (see Ohala, 1981, 1983a, in press b, c, for further data, discussion, and references).

A potentially troublesome counterexample is the apparent dissimilation of voicing among the stops in IE roots, since although voicing spreads to adjacent segments, it is not known to spread to such an extent as to camouflage the distinctiveness of voicing of a whole CVC syllable. However, Gamkrelidze and Ivanov (1973) and Hopper (1973) have suggested that this series of stops was really glottalized — i.e. having a feature well known to dissimilate, e.g. in Quechua and Lahu. These reanalyses — mine of the nature of dissimilation and Gamkrelidze's of the IE stop system — are thus mutually supporting and, one may hope, will have fruitful consequences.

CONCLUSION

If one accepts the pessimistic views discussed above regarding the necessity or the possibility — now or ever — of trying to formulate explanatory accounts of sound patterns, if one accepts the view that phonology is different from other disciplines and is therefore exempt from the usual standards of scientific evidence, then one also must be content with the endless stream of new labels and new notations which either masquerade as explanatory constructs or deliberately forestall the search for them. One will then also have to be content with arguments supported by the most ambiguous evidence — ambiguous in the sense that it is open to many other interpretations. I believe this would be a grave mistake and that this would seriously retard progress in our field.

Looking at the history of various sciences, it seems that one discipline after another eventually recognizes the necessity of adhering to the standards developed in the 16th century — first in the discipline of physics. Chemistry joined the fold around the end of the 18th century, thanks to the efforts of Lavoisier, Priestley, Berzelius, Mendeleev, and others. Physiology — in the face of strong objections from the vitalists, who asserted that the mechanisms of living things would never be reduced to physical processes — nevertheless succeeded in doing just that in the first half of the 19th century, due to the efforts of Claude Bernard, Pasteur, Helmholtz, and others. Ivan Pavlov and others helped to make psychology a rigorous discipline in relatively recent times.

It seems to me that the more difficult the problems studied by a discipline, the greater the delay in its achieving some initial success which finally convinces its practitioners that it should conform to the accepted rigorous standards of empirical science. Physics was the first because it studied (then) the simplest and most easily controlled of problems. Psychology is the newcomer because it studies the most difficult problems. Nevertheless, the *same* basic methods work on all problems, even phonological ones.

Phonology undoubtedly deals with very difficult problems but none so difficult or so different in character as to bar or exempt it from condu(its research as other sciences do.

REFERENCES

ANDERSON, STEPHEN R. 1981. Why phonology isn't "natural." Linguistic In 12. 493–539.

DINNSEN, DANIEL A. 1980. Phonological rules and phonetic explanation. Jo of Linguistics 16. 171–191.

FANT, GUNNAR. 1973. The acoustics of speech. Speech sounds and features, by G Fant, 3–16. Cambridge: MIT Press.

FUJIMURA, OSAMU and JAN LINDQVIST. 1971. Sweep-tone measureme vocal-tract characteristics. J. Acous. Soc. Am. 49. 541–558.

GAMKRELIDZE, THOMAS V. and V. V. IVANOV, 1973. Sprachtypologie u Rekonstruktion der gemeinindogermanischen Verschlüsse. Phonetica 27. 15(

HAUGEN, EINAR. 1942. Analysis of a sound group: *sl* and *tl* in Norwegian.] 57. 879–907.

HOPPER, PAUL J. 1973. Glottalized and murmured occlusives in Indo-Eur Glossa 7. 141–166.

JAVKIN, HECTOR R. 1979. Phonetic universals and phonological change. Rep the Phonology Laboratory, Berkeley, No. 4.

LADEFOGED, PETER. 1980. What are linguistic sounds made of? Langua 485–502.

———— 1983. 'Out of chaos comes order': Physical, biological, and structural p in phonetics. Proceedings of the Tenth International Congress of Phonetic Sc ed. by Marcel P. R. van den Broecke & Antonie Cohen, 83–95. Dordrecht:

LASS, ROGER. 1980. On explaining language change. Cambridge: Cambridg versity Press.

NEWMAN, PAUL. 1977. Lateral fricatives ("hlaterals") in Chadic. Papers in linguistics, ed. by Paul Newman and Roxanna Ma Newman, 107–119.] Afrika-Studiencentrum.

OHALA, JOHN J. 1974. Phonetic explanation in phonology. Papers from the session on Natural Phonology, ed. by A. Bruck, R. A. Fox, & M. W. LaGal 274. Chicago: Chicago Linguistic Society.

———— 1975. Phonetic explanations for nasal sound patterns. Na Papers from a symposium on nasals and nasalization, ed. by Charles A. Fe Larry M. Hyman, & John J. Ohala, 289–316. Stanford: Language Uni Project.

———— 1979. Universals of labial velars and de Saussure's chess a Proceedings of the Ninth International Congress of Phonetic Sciences, 41–47. Copenhagen: Institute of Phonetics.

———— 1980. The application of phonological universals in speech pat Speech and language. Advances in basic research and practice, Vol. 3, ed. b Lass, 75–97. New York: Academic Press.

———— 1981. The listener as a source of sound change. Papers from the session on Language and Behavior, ed. by C. S. Masek, R. A. Hendrick, & Miller, 178–203. Chicago: Chicago Linguistic Society.

———— 1983a. The direction of sound change. Abstracts of the Tenth International Congress of Phonetic Sciences, ed. by Marcel P. R. van den Broecke & Antonie Cohen, 253–258. Dordrecht: Foris.

———— 1983b. The phonological end justifies any means. Proceedings of the XIIIth International Congress of Linguists, August 29—September 4, 1982, Tokyo, ed. by Shirô Hattori and Kazuko Inoue, 232–243. Tokyo [Distributed by Sanseido Shoten].

———— In press a. Consumer's guide to evidence in phonology. Phonology Yearbook.

———— In press b. Phonological evidence for top-down processing in speech perception. Invariance and variability of speech processes, ed. by Joseph S. Perkell et al. Hillsdale, NJ: Lawrence Erlbaum Associates.

———— In press c. Linguistics and automatic speech processing. New systems and architectures for automatic speech recognition and synthesis (NATO Advanced Studies Institute, Bonas, France, July 1984), ed. by Renato De Mori & Ching Y. Suen. Berlin: Springer-Verlag.

OHALA, JOHN J., HARUKO KAWASAKI, CAROL RIORDAN, and MICHELLE CAISSE. Forthcoming. The influence of consonant environment upon the perception of vowel quality.

OHALA, JOHN J. and JAMES LORENTZ. 1977. The story of [w]: an exercise in the phonetic explanation for sound patterns. Berkeley Linguistics Society, Proceedings, Annual Meeting 3.577–599.

SLIS, IMAM H. 1970. Articulatory measurements on voiced, voiceless and nasal consonants. Phonetica 21. 193–210.

VALLERY-RADOT, RENÉ. 1911. The life of Pasteur. [Transl. by R. L. Devonshire.] London: Constable.

EXPERIMENTAL HISTORICAL PHONOLOGY

JOHN J. OHALA
University of California, Berkeley

Introduction

I use the word "experimental" in the title of this paper in a special way. I can explain this special use with the following example. Casual observation tells us that many of the developed nations have both an extensive television broadcast system and a declining birth rate while many of the developing nations have no extensive television and also have a steady or rising birth rate. We might reasonably hypothesize that a country's birth rate varies inversely--and we would imagine, causally--with the extent to which television broadcast is available to the general population. But this apparent causal relation may be spurious--there are numerous other factors besides television which might account for it. How can we test this hypothesis? We have to rely on experiment, which is simply controlled, careful observation--nothing fancier than that. That is, we have to make observations under circumstances where the predicted relation can appear or not appear, free from extraneous and confounding influences. We can either create these circumstances ourselves or we can let nature do it for us. In the first case we could introduce or remove television from a community and see if the birth rate changed as expected, or, in the second case we could wait until some natural event did the same and then observe the results. The "Great New York Blackout" of a few years back was just such an event. (As it happened, nine months after that night without television the number of births in the New York area increased dramatically but momentarily.)

I offer this example to illustrate first, that all hypotheses require experimental verification and second that there are these two types of experiment: that of the man-made controls and that of the nature-made controls.

In historical phonology nature has provided numerous instances of manipulation of some of the relevant variables and we can use these nature-made experimental circumstances to test many of our hypotheses regarding the mechanisms of sound change. But when the hypotheses become sufficiently rich and detailed, as they have in our field, we may have a more difficult time finding naturally-occurring experimental opportunities. In this case we had best have recourse to man-made experiments. It is this type of experiment that I will discuss and describe in this paper.

Experiment in this sense has not been used very much in historical phonology, although there are some interesting exceptions to this generalization, e.g., Rousselot (1891), Esper (1925), Haden (1938), Grammont (1939) and others. To be fair it must be pointed out that it was thought to be impossible to test hypotheses on sound change by using this kind of experiment since it was believed that sounds that are changing now are changing too slowly to be observed in one generation (Hockett 1965) and since past sound changes had happened too long ago to do anything about now. But even if these objections are true, they are rather beside the point because similar arguments, equally invalid, could be used to show that astrophysics could never be an experimental discipline. Most of the objects of study of astrophysics, e.g., the stars, galaxies, interstellar gases, etc., are too far away to be examined in detail and the span of time over which many astronomical events occur is far too long to be observed in hundreds of generations, let alone a single one. But since it is assumed--at least since the time of Galileo and Newton--that the stuff the stars are made of is essentially similar to the stuff earthly material is made of, and that they therefore must behave similarly, it is possible for astrophysicists to verify their hypotheses about the composition and behavior of stars by reference, in part, at least, to the results of experiments run in their laboratories here on earth in a tolerably short period of time. Thus Newton's great insight is suggested to have been the realization that whatever it was that made the apple fall to the ground was the same thing making the moon circle the earth.

Similarly, in historical phonology it is safe to assume that in most essential points the factors which caused sound change over the past few millenia, the normal domain of philological investigation, are present today and thus can be studied experimentally today (Haden 1938). Moreover, the assumption that sound change occurs too slowly to be observed has recently been challenged by Weinreich, Labov, and Herzog (1968).

Of course, present-day phonological and phonetic studies cannot tell us why or how or when, exactly, a given sound change may have occurred several hundreds of years ago, but, as Matthew Chen noted, they can usefully restrict the range of possible hypotheses we need entertain as explanations for the sound changes. Although these notions are not new, I don't think their consequences have been adequately exploited so far. I hope that some of the work I describe in this paper, in addition to other works referenced here, will be sufficient demonstration of the enormous value of experimental work in aiding in the solution to classical philological problems.

Theoretical Framework

I can most easily indicate my view of the place within historical phonology that the work I will describe occupies, by referring to the simple flowchart in figure 1.

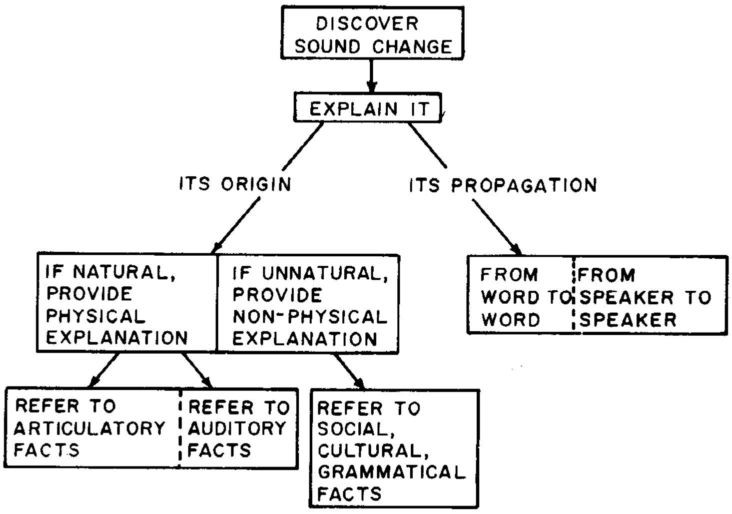


Figure 1. The task of the historical phonologist presented in pseudo-flow chart form.

The first task of the historical phonologist is to discover and perhaps do a preliminary classification of sound changes, but if he is a scientist he must progress beyond this taxonomic task and go on to explain them. "Explaining" sound changes can mean at least two separate things: explaining their origin or explaining their mode of propagation. Their origin may be due to physical or non-physical causes. Those sound changes, the so-called "natural" sound changes which are observed in numerous languages distant from one another typologically, chronologically, and geographically, must have been caused by factors universally present in all speakers and all societies throughout time, namely the inherent constraints of the speech production and perception system. Explanation of the origin of these sound changes will require reference to the facts of articulation and/or audition, as will be demonstrated below. Odd or "unnatural" sound changes--I intend nothing pejorative in using this term--, that is, those sound changes which are essentially unique to a particular language at a particular time, must be due to language-specific and culture-specific factors. Hyper-correction, spelling pronunciation, "fashionable" pronunciations, analogical extension of a grammatical paradigm, etc. are all well-recognized examples of this type of sound change (cf. Malkiel 1966, 1973).

Explaining how a sound change is propagated once it is started can be covered in two ways--partially overlapping. First one can consider how a change spreads from speaker-to-speaker, which is a matter of borrowing or learning (cf. Weinreich, Labov, and Herzog 1968). Second, one can treat the spread of a sound change from word-to-word within the lexicon of a single speaker--although this is related to its spread from speaker-to-speaker.[1] It is the second type of spreading I will treat in the second half of this paper.

Part I. Physically-actuated Sound Changes

I find it convenient to use the simple diagram in figure 2 to illustrate how physical factors initiate sound change.

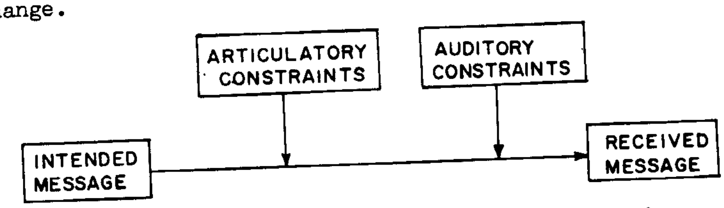


Figure 2. Schematic diagram of speech production process showing how the message (pronunciation) intended by the speaker can become contaminated by two types of "noise", that from the speaker's own articulatory constraints and th from the hearer's auditory constraints. In this way the pronunciation received by the hearer may differ from that intended by the speaker.

This is, of course, a very familiar type of diagram to communication engineers but it nevertheless incorporates notions of sound change advanced by linguists such as Sweet, Jakobson, Henning Andersen and others.[2] Although sound changes involve differences in the pronunciation of two speec communities separated in time, ultimately they must have thei origins in what passes between a single speaker and a single hearer. Implicit in this diagram is the suggestion that the origin of these differences in pronunciation lies in a misinterpretation by the hearer of the pronunciation intended by the speaker. This means some "noise" (in a technical sense) was added to the message (the pronunciation intended). The noise is that due to the constraints of the human articulatory or auditory mechanisms. Certain details of pronunciation not intended nor programmed by the speaker are nevertheless added to his pronunciation because of constraints of the speech production system. Likewise, certain details of the pronunciation as perceived by the hearer may be contributed by his own auditory apparatus. The hearer, of course, generally has no independent access to the mind of the speaker in order to discover what aspects of the pronunciation were intended and which were not, and so learns and repeats the pronunciation as it was perceived, not always as it was intended by the speaker. This noise is always and universally present. Thus "mini" sound changes, those involving only a single speaker and a single hearer, happen all the time. Casual observation supports this. The mechanism whereby one of these mini sound changes becomes "institutionalized" and encompasses a large segment of a speech community is certainly non-physical; sociolinguistic investigations ought to shed light on this aspect of the problem.

In the first half of this paper I will provide a few examples of how the constraints of the speech production and perception system can cause the hearer to perceive a pronunciation which differs from that intended by the speaker.

Epenthetic Stops

A well-known and well-documented example of articulatory constraints causing sound change is that of the so-called epenthetic stops that may appear between nasals and following obstruents, e.g., the [p] in something [sʌmpθɪŋ] or the [k] in length [lɛŋkθ] (Grandgent, 1896) and the stop between [l] and following sibilants, e.g., the [t] in false [fɔlts] or else [ɛlts]. The causes of these stops are reasonably well-known. The first case can be made clear by examination of figure 3. This figure shows some of the physiological parameters relevant to this case during two pronunciations (by an adult male speaker of American English) of the word Samson, on the left with no epenthetic stop and on the right with an epenthetic [p]. For each word are shown from top to bottom: air pressure sampled just behind the lips, the microphone signal, and the output of the nasograph (Ohala, 1971, 1972b) which is a rough indication the degree of velic opening. The two broken vertical lines indicate, approximately, the onset and offset, respectively,

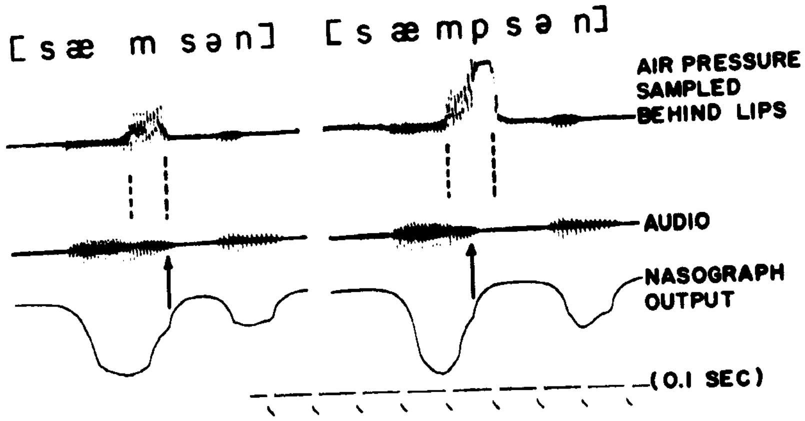


Figure 3. The origin of the epenthetic stop between nasal and following obstruent. Parameters from top to bottom: air pressure sampled behind the lips, microphone signal, nasograph signal (a rough indication of velic elevation, velic closure being indicated when the signal is at the top of its range; see Ohala, 1971, 1972b), and a 0.1 sec timing pulse. Broken vertical lines delimit the labial closure; arrows mark the approximate moments of velic closure. When the velic closure is made simultaneous with the release of the labial closure, as it is in the utterance on the left, n epenthetic stop results. But when the velic closure leads the labial release, as in the utterance on the right, an epenthetic [p] results, as is revealed by the rise in oral pressure.

of the labial closure. The arrow marks the approximate moment of velic closure. Ideally the [m] followed by the [s] requires that the velic closure should begin simultaneously with the release of the labial closure. This occurs in the utterance represented on the left. However, in the utterance on the right the velic closure is done prematurely with respect to the labial release and thus a complete stoppage of the air occurs, as is revealed by the rise in pressure behind the lips. This air pressure is then released when the labial release occurs and thus a labial stop is heard. This process is actually a partial denasalization of the nasal consonant in the environment of the following oral obstruent and is parallel to (but the reverse of) assimilatory nasalization. Like most such assimilations it is anticipator

The case of the dental stops that appear between [l] and following sibilants is aerodynamically similar to the precedin case (Ohala, 1972a). Figure 4 shows schematically the areas of tongue-palate contact for [l], [t], and [s].

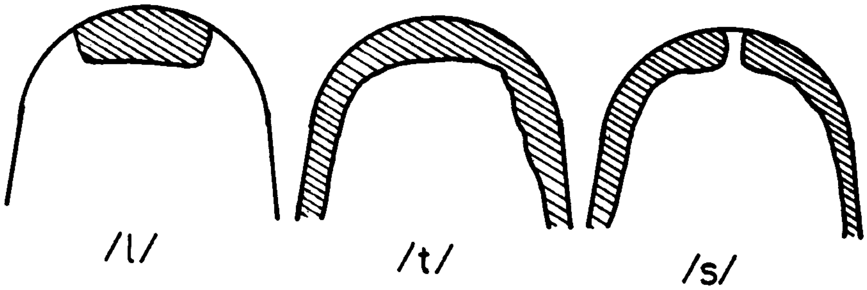


Figure 4. Schematic diagram of the hard palate showing the areas of tongue-palate contact for the consonants [l], [t], and [s]. The contact areas for [l] and [s] are partially complementary; when added together, as they may be in a transition from [l] to [s], their combined area of contact produces a [t]. This is the origin of the epenthetic [t] in such words as false [fɔlts] and else [ɛlts].

It can be seen that the contact areas for [l] and [s] are to a certain extent complementary: [l] has contact in the middle of the alveolar ridge but not along the sides, and [s] has contact along the sides, but not in the middle. In moving from an [l] to an [s], contact and release of contact must be made simultaneously in these complementary areas. To the extent that they are not simultaneous, complete contact all around the alveolar ridge may result and thus complete stoppage of the air, that is, a [t], will result.

This much is quite elementary and has been known for some time. But there is an interesting assymmetry in these phenomena: the epenthetic stop is apparently far less likely to appear if the following consonant is voiced than if it is voiceless (Grandgent, 1896). Hector Javkin and I (Javkin and Ohala, 1972) have investigated the epenthetic stops between [l] and following [s] or [z] by recording tongue-palate contact (using a dynamic palatograph) while simultaneously recording oral air pressure. We concluded after examination of the data obtained, that the sequence [l] + [z] is just as likely to yield a complete closure around the alveolar ridge as is [l] + [s], the difference, however, lies in the fact that the vibration of the vocal cords throughout the [l(d)z] sequence offers relatively high resistance to the air flowing into the mouth, such that when the complete stoppage does occur, the air pressure does not build up very much and consequently when released, does not create a very audible noise burst vis-a-vis the fricative noise for the following [z]. Thus, the epenthetic stop may be present but it is not heard. Although more study is needed we presume a similar explanation holds for the nasal + obstruent sequences.

Lowering of Nasalized Vowels

A case which involves reference primarily to auditory factors is that of the tendency for nasalized vowels to lower, exemplified in such French word pairs as fine, fin [fin], [fɛ̃] , or brune, brun [bʁyn] , [bʁœ̃] . Various authors have attempted to provide basically articulatory explanations for this phenomenon (Passy, 1891, Martinet, 1955, Delattre, 1970) but I think a simpler explanation can be given by reference to auditory-acoustic facts. House and Stevens (1956), using an electronic analogue of the oral and nasal tracts, studied the effect on various vowels' spectra of varying degrees of nasal and oral tract coupling (i.e., physiologically, varying amounts of velopharyngeal opening). Figure 5 shows their results for the high front vowel [i] .

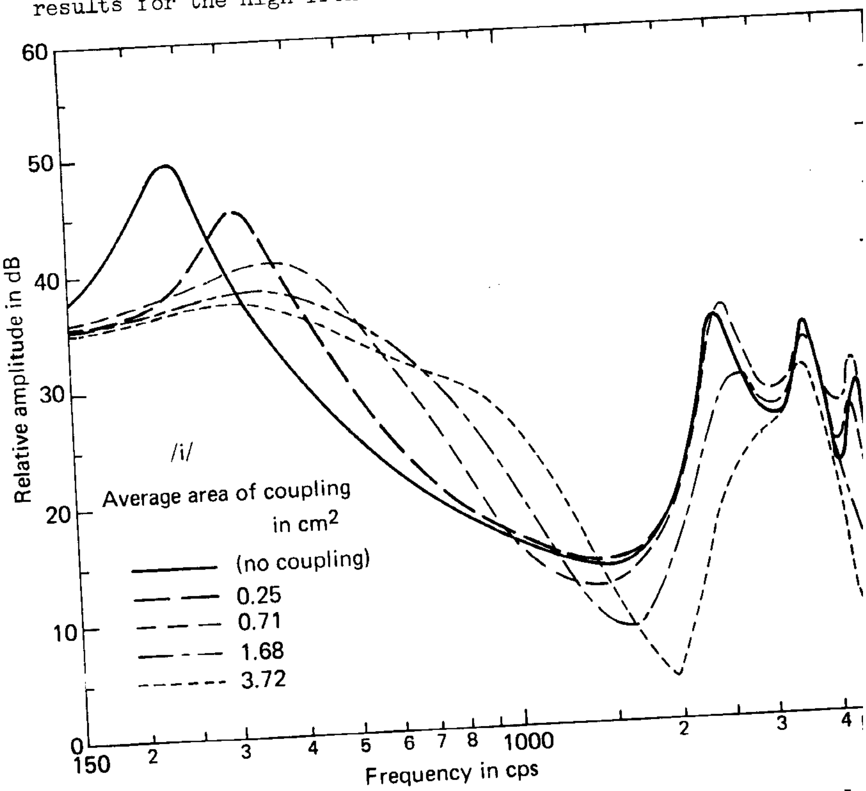


Figure 5. The effect of nasal coupling (= amount of velc pharyngeal opening) on the spectrum of the vowel [i] (fror House and Stevens, 1956). Solid line: spectrum of [i] ı no nasal coupling; broken lines: spectrum with varying amounts of nasal coupling. Among other changes, there i upward shift in the peak frequency of the first formant - from about 240 Hz to a maximum of about 360 Hz.

There are a number of differences to be noted but the most important for our purposes is the rise in the first formant (the changes in the other formants are relatively small). Fant (1960) arrived at essentially the same conclusions in an independent study. In figure 6 is shown graphically the change in frequencies of the 1st and 2nd formants of the vowels due to increasing nasalization as derived from the House and Stevens data.

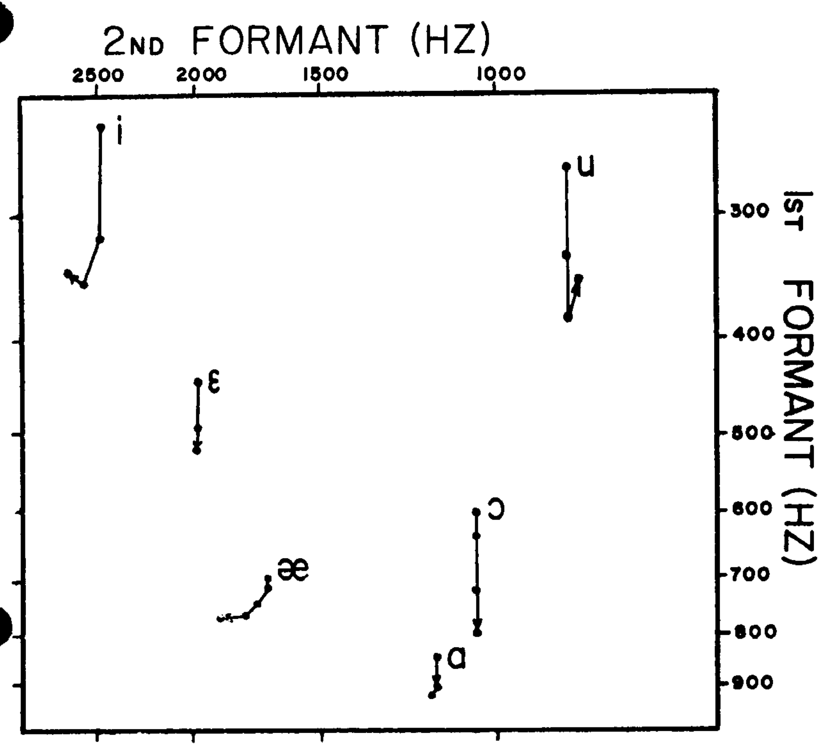


Figure 6. Effect of increasing nasal coupling (= velopharyngeal opening) on the formant frequencies of various English vowels (measured from the data of House and Stevens, 1956). This acoustic space, defined by formant one vs. formant two, closely matches the auditory vowel space, defined impressionistically by the labels high-low vs. front-back. Thus auditorily, increased nasalization lowers some vowels.

In this formant one vs. formant two space, upon which the traditional auditory vowel space is based (Joos, 1948, Ladefoged, 1967), the effect of increased nasal coupling is to lower the high vowels [i] and [u] and the mid back vowel [ɔ] . Thus we can say that nasalization makes these vowels appear to lower auditorily although there need be no actual change in the articulation of the vowel. Whether this auditory effect might not lead to a change in articulation subsequently, due to faulty imitation on the part of other speaker/hearers, is a separate question. It is not unlikely though.

Nasalization in the Environment of Glottal and Pharyngeal Consonants

Nasalization of vowels in the environment of nasal consonants is well known. Less well known and somewhat rare is the fact that vowels also occasionally become nasalized in the environment of glottal and pharyngeal consonants. This has been documented for Chinese by Chen (1973), for Lahu (a Lolo-Burmese language) by Matisoff (1972), and for East Gurage (a Semitic language of Ethiopia) by Hetzron (1969). Matisoff (personal communication) has also pointed out a probable instance of this in certain British English dialects, namely, those that render half as [hɑ̃f]. This is also related to the fact that the long stretches of perseveratory nasalization in such languages as Tereño and Sundanese (described by Bendor-Samuel (1966) and Robins (1957), respectively) can pass through glottal consonants.[3]

In seeking an explanation for this I used the nasograp to investigate soft palate activity during glottal and non-glottal consonants in American English (Ohala, 1971, 1972b) Some of the resulting records are shown in figures 7 and 8.

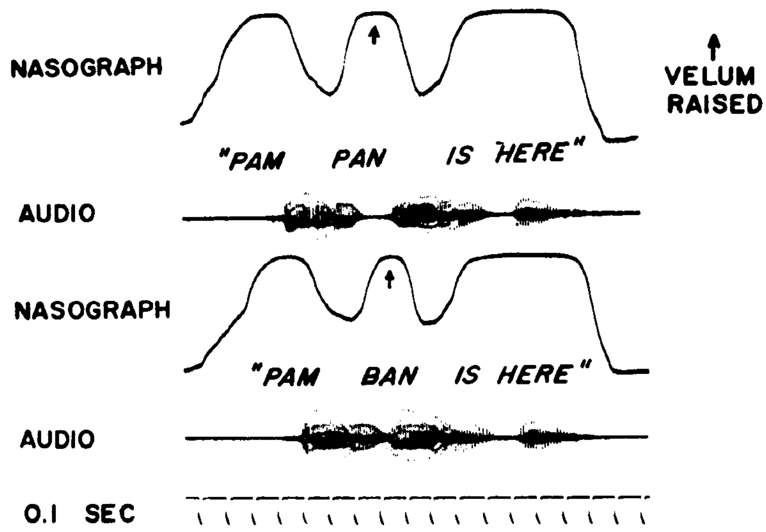


Figure 7. Nasograph and microphone signals for two utterances showing the velum is elevated and closed during the oral obstruents [p] and [b] (consonants marked by arrows). These traces should be compared with those in figure 8.

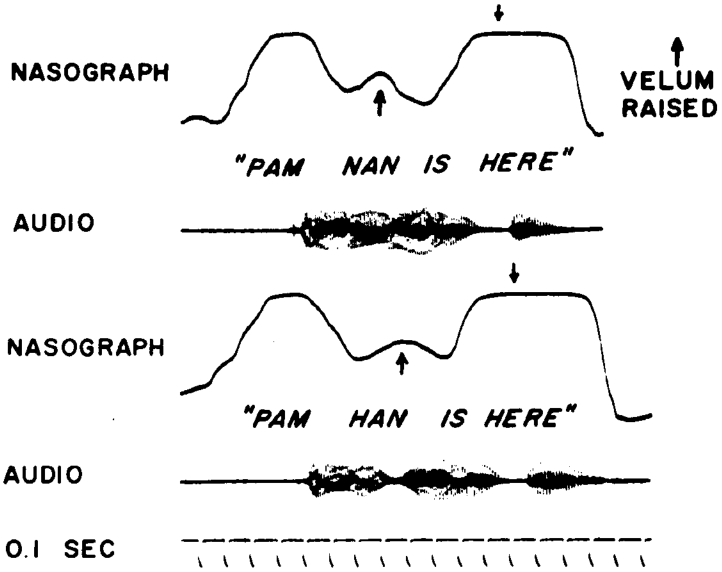


Figure 8. Parameters as in figure 7. The nasograph signals show the velum remains lowered and open during [n] (top utterance) and during the [h] in the heavily nasalized environment (consonants marked by upward arrows). However, during the [h] in <u>here</u> (marked by downward arrows), where it is flanked by an obstruent and a high vowel - both of which require an elevated velum -, there is velic closure.

The frame sentence "Pam _an is here", [pʰæ̃́m æ̃́nIzhIr], provided a heavily nasalized environment in which the various consonants could be spoken. Figure 7 shows that the oral obstruents [p] and [b] require velic closure. This is expected since the essential acoustic cue of stops is a noise burst at their release and this requires a build-up of air pressure in the mouth which in turn requires an air-tight oral cavity, including velic closure. Figure 8, however, reveals that the soft palate position for [h] is about the same as tha for [n] in this nasalized environment. Similar results were obtained for the glottal stop. Unlike the oral obstruents, glottal (and probably pharyngeal) consonants do not require soft palate elevation since they involve air pressure build-up further back in the vocal tract than the point where the nasal and oral cavities join (Schourup, 1973). Also a lowered velum will not greatly distort the acoustic quality of the broad-band noise produced by these consonants. This second factor is probably more important since in the case of high vowels and [l] the build-up of air pressure is also at the glottis and yet the soft palate must generally be raised during these sounds in order to avoid distorting their distinctive acoustic character. This is not to say that the velum must be lowered during these consonants because these figures also show that during the [h] in here, in these utterances, where it is adjacent to an oral obstruent and a high vowel, the velum is elevated. Thus the position of the velum during glottal and pharyngeal consonants must be largely contextually determined. Nevertheless, given the compatibility of these consonants and nasalization we can explain the tendency or likelihood of nasalization occurring on vowels adjacent to them.

Tonal Development

The development of distinctive tone on vowels due to the loss of some voicing distinction on preceding consonants is well documented for Punjabi, Chinese, and many Southeast Asian language (see Hyman, 1973). Most of the cases are those in which a voicing distinction is lost leaving a high tone on the vowel that followed a previously voiceless, -- especially a voiceless aspirated -- obstruent, and a low tone on the vowel following a previously voiced obstruent. In addition there is the case of Punjabi in which the breathy-voiced consonants (the so-called "voiced aspirates") were merged with the voiceless unaspirated consonants leaving a low tone on the vowel that followed. The reasons for this are rather clear--more or less. It is well known that there are slight pitch changes on the vowels following these different consonant types. This is illustrated in figure 9 which shows the fundamental frequency and rectified, integrated audio curve for four Hindi nonsense words containing four bilabial stop types as spoken by an adult female speaker of Standard Hindi.

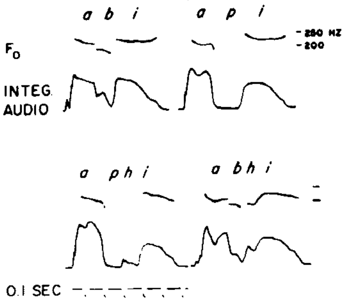


Figure 9. Fundamental frequency (F_0) and the rectified, integrated audio signal (= voice amplitude) during four Hindi nonsense words. The fundamental frequency is slightly higher after the release of the voiceless unaspirated stop [p] and the voiceless aspirated stop [p^h] than it is after the release of the voiced stop [b]. Also, the fundamental frequency after the release of the breathy-voiced stop [b^h] is considerably lower than it is after the other three. These small fortuitous pitch perturbations, which occur due to the physical constraints of the speech production apparatus, are instrumental in the development of tone.

Comparing the fundamental frequency after the release of the [b], [p], [p^h], it is obvious that the pitch is a little higher for the voiceless stops [p] and [p^h], than it is for the voiced stop [b]. And the fundamental frequency immediately after the release of the breathy-voiced stop [b^h], is considerably lower than it is after any of the other stop types. Hindi is not usually considered to be a tonal language but we can see here the "seeds" of the sound change which made Punjabi a tone language as well as the effects which caused tonal development in Chinese and other tone languages. (These slight pitch perturbations are known to occur in other non-tonal languages as well; see Lea, 1973 and Ohala, 1973.)

There is some dispute as to the exact causes of these pitch perturbations - one theory attributing them to aerodynamic effects and one to differences in vocal cord tension. I will not go into details of the dispute here (see Ohala, 1973 for a discussion), even though it is relevant for analyses of the histories of tone languages. In any case it is sufficient for our purposes to know that even without intending to, a speaker will produce these small pitch changes as a fortuitous consequence of the normal articulation of voiced, voiceless, and breathy-voiced consonants. Moreover these small pitch differences have been shown to serve as perceptual cues for the differentiation of stop types in languages not considered to be tonal, e.g., Russian and English (Chistovich, 1969; Haggard, Ambler, and Callow, 1970; Fujimura, 1971). Of particular interest in this regard is Fujimura's study in which he asked American English-speaking subjects to judge as "k" or "g" the initial consonant in some synthesized CV sequences which differed both in voice onset time and in the pitch contour following the stop release. The results are shown in figure 10, where percentage of "k" judgements is plotted against voice onset time of the stimulus.

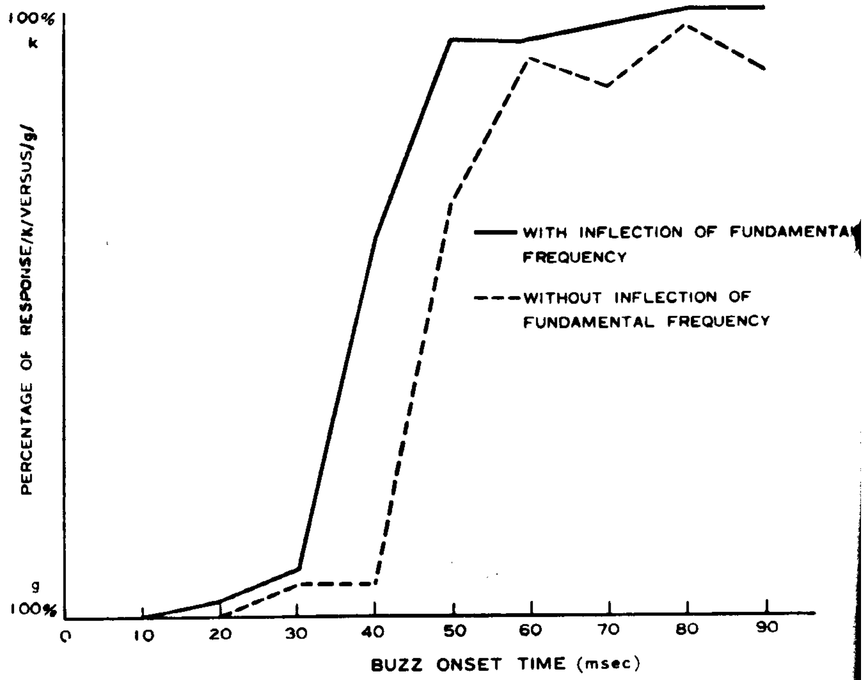


Figure 10. American English listeners' response to synthetic CV stimuli (from Fujimura, 1971). Abscissa: percentage of "k" judgements; ordinate: voice onset time stimulus. Broken line: subjects' response to series of stimuli with no pitch change; solid line: subjects' respo to series of stimuli with small upward pitch contour. The addition of the small pitch rise makes the stimuli more "g"-like.

The broken line shows the response to all those stimuli having no pitch contour. For this series the "boundary" between the "k" and "g", in regard to voice onset time, is about 45 ms. The solid line shows the response to the stimuli which had a slight pitch rise. Because of this pitch rise the stimuli became a bit more "g"-like since the boundary is now about 35 ms. It appears, then, that the English hearer uses not just a single acoustic cue to differentiate these stop types but actually a hierarchy of cues. In this case voice onset time is clearly the most important cue and the pitch contour is of much less, but still appreciable, importance. No doubt the intensity and frequency of the noise burst have some intermediate perceptual importance. There is, in fact, abundant evidence that listeners use a multitude of perceptual cues to distinguish what the phonologist may prefer to conside a "minimal" contrast (Lisker, 1957; Fry, 1960; Delattre, 1967

Ainsworth, 1972, to mention just a small fraction of the relevant literature). A sound change could occur then, when one speaker/hearer uses a different ranking of the relative importance of the various cues than did some other speaker/hearer, and yet both could be using the same acoustic representation of a given word in speaking. One could imagine, for example, that at some early stage in Punjabi when breathy-voiced stops still were pronounced, some speakers began to treat the pitch rise as more important than the breathy-voice itself. As the breathy-voice cue became of lesser and ultimately of no importance perceptually - which need not be a very gradual process - it would no longer be maintained in the production of those words which once possessed it.

Nasalization of Low Vowels

The physical explanations I am putting forth for sound change - unlike some other purportedly explanatory theories of sound change - are empirical claims in the true sense of the word: they can potentially be falsified. Various types of evidence could be marshalled to disprove the models presented here. I can illustrate this by considering a case which seems to have a simple explanation at first but which upon examination of further data is shown to be more complicated and unexplained by present models.

Chen and Wang (1973) have documented quite well the generalization that distinctive nasalization sits more comfortably on low than on high vowels. They have shown that nasalization invades the vowels of a language generally by hitting the low vowels first and subsequently, if at all, moving to the high vowels. Also, if distinctive nasalization is lost on vowels, it will depart from the high vowels first, lingering for a longer time on low vowels. It is interesting therefore to find that the phonetic literature for over 100 years (Czermak, 1857; Moll, 1962; Ohala, 1971) abounds with the observation that soft palate height varies directly with the "height" of the vowel. This is probably not due to any direct mechanical linkage between soft palate and tongue as was suggested by Moll, since recent electromyographic investigations of the muscles controlling the soft palate reveal that these muscles actively produce a more elevated position of the soft palate for high vowels than for low vowels (Fritzell, 1969; Lubker, 1968). In any case, because of this correlation we might simply hypothesize that distinctive nasalization occurs on low vowels first, because they are likely to have a somewhat lowered soft palate.

But this explanation is weakened when we examine in more detail the results of the House and Stevens (1956) study. Their research reveals that for a given amount of nasal coupling there is less of an effect acoustically and perceptually on low vowels than there is on high vowels. This, in fact, as has been suggested by Lubker (1968), may explain why there is the correlation between soft palate height and vowel height: a somewhat lowered velum can be tolerated during a low vowel because nasal coupling will have less of an effect on the acoustic quality of the vowel. But this does not help us to account for the greater tendency for nasalization of low vowels over high vowels, in fact, it would predict jus· the opposite. This, then, is a point that requires further study, but it is clear that the ultimate explanation for it will be a physical one.

"Ease of Articulation"

Note that I have not appealed to the notion of "ease of articulation" in trying to provide a physical account for sound change. This is because I do not think it is a very useful concept for our purposes. To the extent that it attributes sound change to laziness or sloppiness of speakers it is probably wrong or at least unproven. The kinds of fortuitous deviations in pronunciation which I have described are not caused by laziness or carelessness; they are caused by the inherent anatomical, physiological, and neurophysiological constraints characteristic of all vocal tracts - even those of hard-working speakers.

"Leveling"

Another idea prevalent in historical phonology for which the evidence is quite poor is the notion that all physically-caused sound changes are largely a matter of articulatory factors tending to reduce or level the contrasts in language while perceptual needs counteract this by maintaining them or by causing compensatory contrasts to be introduced to replace lost ones. This is an appealing notion but it hardly presents a very accurate picture of sound change. Not all sound changes which are caused by articulatory constraints result in loss of contrasts, e.g., the case of the epenthetic stops, considered above, and some of the sound changes which are most likely due to auditory factors result in reduction of contrasts, e.g. replacement (and loss) of [x] by [f] in the history of English. This is not to deny that some sound changes reduce the contrasts in language nor that the need for intelligibility in speech necessitates some more or less constant level of distinctiveness in language, but these are hardly the most important patterns in sound change.

Choice of Explanatory Models for Phonology

There is in the current phonological literature considerable debate over the choice of feature systems for the representation and explanation of the way speech sounds behave. I have indicated in this paper the kinds of features or parameters it is necessary to specify for these purposes, namely, all of the traditional parameters used by physical phonetics, e.g., the geometry of tongue-palate contacts for consonants, where a consonantal constriction is with respect to the opening to the nasal cavity, the effect varying degrees of nasal coupling has on the formant frequencies, the contribution of formant frequencies for the perception of vowel quality, the aero-dynamic effects of various consonantal articulations, aspects of the transitions between sounds, etc., etc. These are obviously not exclusively articulatory or exclusively acoustic features; both are needed. They are obviously not all binary and they obviously must be incorporated in formal systems which reveal their complex interactions, that is, not in a two-dimensional matrix. Two-dimensional feature matrices using a small number of exclusively acoustic or exclusively articulatory binary features may serve gross taxonomic functions adequately but they are of little use beyond that--at least they explain very little about why speech sounds pattern the way they do. Fortunately there is much recent phonetic research, using sophisticated models, which has direct applicability to phonological problems (see, e.g., Lindblom, 1972, Liljencrantz and Lindblom, 1972).

Part II: Propagation of a Sound Change within a Speaker's Lexicon.

The spreading of sound changes through the vocabulary of a single speaker is a complex area of study because there are many types of sound change, some conditioned, some unconditioned, some involving morphological factors, some not. Such differences may entail differences in the mechanisms of sound change propagation. I will therefore concentrate here on one type of sound change: that which is manifested today in morphological alternations.

An Experimental Technique to Study the Psychological Side of Sound Change

Generative phonologists have typically assumed that if the linguist finds a systematic sound pattern in the language that there is then no question that the native speaker has also noticed it and can be productive in extending it. This w be nice (and amazing) if true, but it is not, as Zimmer (1 and others have shown.[4] So the first task is to find out the particular sound pattern is productive and then, if it to try to find out how it is productive, that is, to disco the psychological mechanisms used by the speaker.

Take, for example, the sound change that turned [k] i [s] before high front vowels, named "velar softening" by s linguistic poet. It is examplified in such word pairs as public, publicity; mystic, mysticism. We can see if speakers are productive in extending this sound change by seeing if they can manifest it in new environments where i has never appeared before but should appear if it is an ac pattern. I therefore approached 26 American English spea all students of art or architecture at Berkeley, and asked them to help me prepare an extrapolated or extended dictio of English (J. Ohala, forthcoming). They were to do this using a common rule of English word formation, that of suffixation with existing word stems. They were to prono the newly invented word, tell its meaning, and tell if the would be likely to use it. I was not interested in what thought the word would mean or whether or not they would u it, but feigning interest in these things possibly made th task a bit more believable and perhaps diverted the subjec attention somewhat from the pronunciation of the word, ins perhaps, a more unstudied response. The entire test was conducted orally. First a suffix was given and then the words to which the suffix was to be added. In the cases interest these were stem + suffix combinations that would likely to yield some phonetic change in the stem. A fe "filler" words and suffixes were also included, that is, combinations of stems plus suffixes which would not be expected to lead to any phonetic changes in the stem, e.g. normal + hood, fad + ity, brain + ism. This was done to insure that the subjects were answering as naturally as possible and did not think that they were required to prov a phonetic change in the stem but rather that they should so only if they were so inclined.

Velar Softening

The results of the investigation on velar softening i given in table 1.

RESPONSE / STIMULUS	k	s	other
"DOMESTIC+ISM"	14	10	2
"TOXIC+ISM"	15	4	7
"PUBLIC+ISM"	19	6	1

In this table and in the two to follow, the stimuli or input, consisting of the particular stem plus suffix, are given in the left-most column. The types of responses or outputs the subjects could give are listed in the top row, where, in this table, "k" indicates the retention of the [k] in the stem and "s" indicates change of the [k] to [s] in the new derivation. "Other" indicates anomalous responses. The numbers at the intersection of particular rows and columns indicates how many subjects gave a particular pronunciation for a particular stem + suffix combination. This table reveals that velar softening - at least with these types of derivations - is only marginally productive: only about 30% of those cases where the responses could be clearly interpreted actually exhibited velar softening; most of the pronunciations were of the type: [dowmέstIkIzm].

Should we conclude from this that velar softening is largely dead as a sound change in Modern English? Yes, if we imagine that the propagation of a sound pattern is done by the kind of independent phonological rules such as one finds in modern generative grammars. But independent phonological rules, that is, those which effect a change in a form given only its grammatical and phonetic characteristics, are not the only kind of rules a speaker can use. There is also analogy, which I take to be a shorter label for analogical phonological rules. These are rules which for their application require not only information about the given phonological item, but also information culled from the lexicon as a whole. It can be illustrated by some of the other data obtained from this investigation.

Evidence for Analogy

Early in this test derivations with -ion were elicited with the words obtain and pertain. In this case, however, a "leading" example of the use of this suffix was first given to the subjects, namely detain, detention. Later, towards the end of the test, derivations with -atory were elicited, again, with the words obtain and pertain. This time the example given them was explain, explanatory. The results for obtain are shown in table 2.

STIMULUS \ RESPONSE	ɛ	ej	æ
"OBTAIN+ION" E.G.:"DETAIN-DETENTION"	18	8	
"OBTAIN+ATORY" E.G.:"EXPLAIN-EXPLANATORY"		16	10

As is shown, when primed with detain, detention, 18 of the 26 subjects responded with [ʌbtʰɛ́nʃən] and 8 left the stem unchan ed as [ʌbtʰéjnjən]. When primed with explain, explanatory, most subjects left the stem phonetically unchanged as [ʌbtʰéjnətɔrij], but 10 gave [ʌbtʰǽnətɔrij], and of these 10, 9 were among the 18 earlier who had given [ʌbtʰɛ́nʃən].

This shows among other things[5]: that the assumption by generative phonologists of unique underlying forms for words is not supported because they would apparently posit a differ ent underlying form for those words showing the [ej] -[æ] alternation versus those showing the [ej] -[ɛ́] alternation (Chomsky and Halle, 1968). But here we have the same word showing both alternations in the speech of some subjects. S these derivations cannot be based on a single underlying form (or, perhaps, - as I will suggest below - on any abstract underlying form). Also it shows that the particular form o the derivations, contrary to the assumptions of generative phonology, does depend on other words or pairs of words in t lexicon of the speaker. Having found, or, in the present c having been provided with suitable existing models, the speak can pattern new derivations after them, that is, he can analogize.

I realize that to some readers this may appear to be a demonstration of the obvious, but since analogical processes in phonology have received little or no attention in modern generative phonological grammars, perhaps it is occasionally necessary to demonstrate the obvious.

The details of analogical derivation may be something like the hypothetical algorithm given in figure 11 (from J. Ohala, forthcoming).

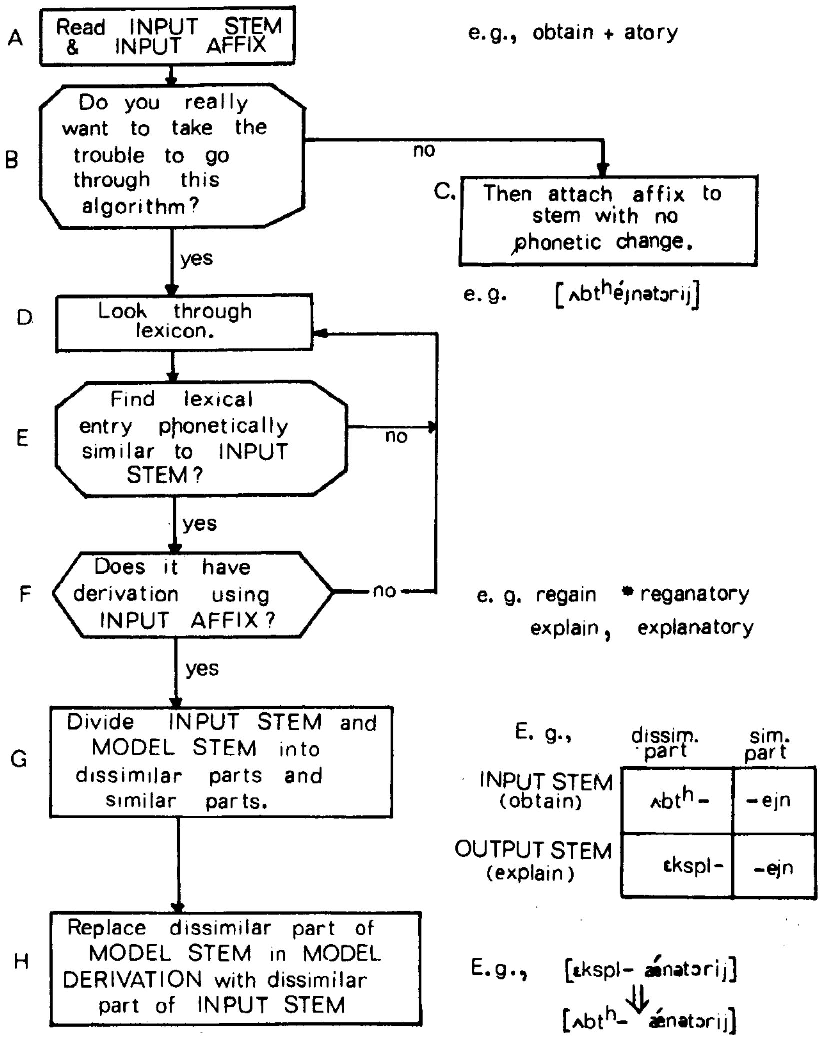


Figure 11. Representation in flow chart form of hypothetical algorithm for analogical derivation.

Finding a model in the lexicon on which to base the new derivation, i.e. Steps D, E, and F, is the most important feature of this model. Also, since most subjects did not change the stem phonetically when making the derivations, it seems necessary to include the branching BC which allows the affix to be attached to the stem without creating any phoneti change in the stem. I realize that this algorithm ignores certain problems, e.g., how does the speaker determine that two words are phonetically similar? Some phonetic analysis is necessary, but is this done by phoneme or feature comparison or what? Also, how does the speaker know, for example, that *explain* and *explanatory* are derivationally related? Perhaps it is by their similar phonetic, semantic, or orthographic features.

All of this is pure speculation but interesting speculation, I think, because this is a superficially viable model and yet it differs so greatly from the system proposed by generative phonology as accounting for this same type of phonological behavior. I will mention just a few of these differences here (see J. Ohala, forthcoming, for further more detailed discussion). First, these analogical rules do not operate on highly abstract forms; to derive [ʌbtʰǽnətɔrij] only the forms *obtain*, *explain*, and *explanatory* have to be posited. The vowel [æ] in [ʌbtʰǽnətɔrij] was derived not from the same source as the vowel [ej] in *obtain*, [ʌbtʰéjn], but rather from the vowel [æ] in *explanatory*, [ɪksplǽnətɔrij] that is, a "surface" vowel. Second, separate linearly (or non-linearly) ordered rules are unnecessary with analogical rules. For example, according to Chomsky and Halle, a word such as *criticize* [krÍtIsajz] , is said to require for its derivation a sequence of rules such as follows: given an underlying form of /krItIk + īz/, there is first a change of the stem-final /k/ to /s/ in the environment of the high fron vowel /ī/, then a change of the /ī/ to /aj/ via vowel shift, diphthongization, etc. But a speaker could produce a novel *-ize* derivation with velar softening, say, *sputnicize* [spÚtnIsajz], from *sputnik*, [spÚtnIk], in one step by simply copying, analogically, the *-cize*, [-sajz], part from the derived member of some suitable pair of existing words, say *critic*, *criticize*. Third, the speaker would not have to store in his memory long lists of phonological rules such as is posited by generative phonologists, because analogical rules can be made up on the spur of the moment by reference t the lexicon and after they have been used they can be forgotten. According to this, speakers might be expected to show some variability in their derivations as they might pick different models for their derivations on different occasions. And, of course, speakers do exhibit such variability on new and unfamiliar derivations. From this we can conclude that had I provided the subjects in my study with more appropriate examples, they might have shown more productivity with velar softening or, in fact, with any sound pattern.

There is a large literature on analogy and most of the standard linguistic textbooks contain a chapter or two with many examples of its apparent use.[6] But most of the examples in the area of phonology are those of the analogical extension of phonetically implausible sound patterns. I would suggest, however, that the process can be used for any sound pattern, even phonetically plausible ones such as vowel laxing and velar softening.

The term "analogy" occasionally seems to have bad connotations in modern linguistic discussions, but for reasons which are irrelevant to the present use of it. For example, the facile appeal to "analogy" to explain speakers' productive use of the syntactic patterns of their language has been criticized. But this has nothing to do with the present discussion. Also philologists have used "analogy" and "dialect borrowing" - usually with no supporting evidence - to explain away all the exceptions to their posited laws of sound change. This, again, does not apply to the present case since I am attempting to provide evidence for use of analogy (and moreover, exceptions to sound change are no longer something that needs to be "explained away" (Wang, 1969)).

Stress Patterns

I am not in a position to prove that all productive extensions of sound patterns are done by analogical phonological rules; all I can do, given my results, is to suggest that some cases quite clearly involve them, and others quite plausibly do. Consider the results of my investigation of part of the stress rules. These are given in table 3. Three representative input items are given: thermos + ian, human + ian, and methane + ity. The response categories are indicated along the top: the S's represent the two syllables of the stem and the superposed stress mark indicates which of the syllables received the stress. The input stems all have stress on the first syllable in the underived form, but should have stress put on the second syllable in the derived forms, assuming the productivity of the stress pattern on similar pairs of existing words, e.g., Dárwin, Darwínian. The responses show that the stress rules are generally

RESPONSE / STIMULUS	´s s	s ´s	other
"THERMOS + IAN"	10	12	4
"HUMAN + IAN"	2	19	5
"METHANE + ITY"		24	2

productive - very much so in the case of *human* + *ian* and *methane* + *ity*, a little less so but still quite productive in the case of *thermos* + *ian*. How can the differences in the distribution of the responses be accounted for? I suggest that it may be a question of whether or not the speaker can make any guess about the phonetic quality of the second vowel, which would have to be clearly pronounced if stressed. Information from the lexicon may aid him in making that guess in some cases. *Methane* is no problem because although the second vowel is unstressed, it is not reduced. In *human*, however, the second vowel is unstressed and reduced, but there is a related word, *humanity*, [hjuwmǽnItij], which might suggest that [æ] would be an appropriate vowel quality for the second vowel in the stem if stressed. (In fact, 14 of the 26 subjects gave the pronunciation as [hjuwmǽniən], 8 others gave other lax vowels and only 2 gave [hjuwméjniən], as Chomsky and Halle would have predicted by their vowel tensing rule.[7]) *Thermos* has a reduced vowel in the second syllable and it does not have any obviously related words which might suggest what the quality of that vowel should be if stressed. As a consequence speakers are rather less productive in extending the stress pattern to this word. It seems plausible that it is the availability of certain types of information from the lexicon which accounts for the differences in responses here. As for the fact that stress was shifted to the second syllable in most cases, there is no way of telling how this is done - whether the subjects reached for their mental dictionaries, looking for words that would give them a clue how to stress the derived word, or whether they reached for their mental equivalent of chapter 5 of Sound Pattern of English, that is, a list of rules. Only further experimentation, though, not armchair speculation, will reveal what goes on in speakers' heads. Fortunately, some very promising experimental work is being done presently.[8]

Natural Rules

The sound patterns I investigated in this test are velar softening, vowel laxing and tensing, vowel shift, stress assignment, s-voicing, and some others. Some of these patterns, especially velar softening and perhaps vowel laxing, would be considered "natural" sound patterns in that they have some phonetic motivation and have been observed in many other languages. Some of these patterns, especially the stress patterns, would be considered "unnatural" in that they are quite arbitrary phonetically and are characteristic only of English. But speakers were more productive with the unnatural sound pattern, stress, than they were with the other, allegedly more natural, sound patterns. We could at least conclude from this that natural sound patterns are not propagated more easily than unnatural ones. This conclusion, if borne out in further tests, undermines the claim of generative phonologists that speakers prefer or can more easily learn natural than unnatural sound patterns. And, in fact, this conclusion is supported by the results of two psycholinguistic experiments which specifically investigated the question "are natural sound patterns more easily learned than unnatural ones?" (Schane and Tranel (1970) and personal communication, and J. Ohala, forthcoming.) (Cf. also Sherman (1973) who demonstrated the rapid spreading in English of the noun-verb stress alternation.)

Origin vs. Spread of Sound Change

In this paper I have given separate treatment to the origin of sound changes and to their propagation. This is not a novel proposal today, but I think the reasons for it are worth elaborating. Formerly, many philologists expressed the view that phonetically-plausible sound changes happened because the pronunciation of every member of the speech community in question experienced the same physical disturb at the same rate over the same period of time. This is bo extremely unlikely statistically, and also unnecessary. I is unnecessary because the transmission or "popularization" of a sound change is more easily accounted for by essential psychological and sociological means. The origin of velar softening may have involved the misperception of a [ki] sequence as [tʃi] or [si] but we do not have to suppose that those subjects in the test who did extend velar softening a gave forms such as [pʰʌ́blIsIzm] and [tʰɑ́ksIsIzm] also suffered some kind of misperception. This has an analogue evolutionary theory: when we note the origin of a new spec as evidenced by some new physical attribute, we can assume origin of this attribute was due to some genetic mutation i some individual--perhaps cosmic rays or some other influenc caused a change in his DNA. But we do not believe that ev member of the new species acquired this trait by having his DNA accidently re-arranged. Rather, we assume they got th trait in the normal way by inheritance.

Generative phonology does not make this mistake of assuming an exclusively physical character to sound change; rather it seems to have gone too far in the other direction and has assumed an almost exclusively mentalistic character sound change, ignoring the clear evidence of the physical origin of sound change. Postal (1968), for example, sugges that sounds change due to the same whims of fashion that ca clothing and car styling to change. This would make it an amazing coincidence that so many different language communit over the ages and in distant lands experienced the same whi of fashion in pronunciation. Separation of the issue of th origin of sound change from the issue of its propagation av these problems and eliminates the absurdities.

Conclusion

I have tried to show that the mechanisms of sound chang can be studied experimentally. Questions of the origin of those widely attested, so-called "natural" sound changes can be answered by reference to the physical phonetic facts of speech. There is a large body of data in the literature on physical phonetics, some of which I have cited in this paper which can already be used for this purpose. Questions of t spread of sound change can be answered in part by reference psychological factors. Some experimental techniques which may help us to discover these psychological processes were presented.

Acknowledgements

Many of the ideas in this paper developed out of enlightening discussions on sound change with my former and present colleagues in the Phonology Laboratory at Berkeley, especially Matthew Chen and Bill Wang.

Discussion

Richard C. DeArmond. Your results to the experiment of vocalic alternations parallels to some degree the results shown by Danny Steinberg at the summer meeting of the L.S.A. earlier this year. His paper generated a considerable amount of discussion, including one comment that perhaps experiments might be extended to test acceptability of the hearer. It was noted that there seems to be a varying amount acceptability of resulting vowel alternations of Romance or polysyllabic stems, but a change of the vowel of a monosyllabic stem is generally unacceptable, to which there was general agreement; e.g. from flake/flejk/,/fléjkítıj / is acceptable, even if strange, whereas /flækit ij/, the result of the vowel laxing rule is not acceptable.

John Ohala. In the Krohn, Steinberg, and Kobayashi (1972) experiment, subjects were told they were being tested for the speed with which they could select an appropriate suffix to add to a given word given the context of a little story that they were told orally, e.g., (not an actual example) "My uncle bought a sapphire ring. Someone said it was just a piece of blue glass, but I was convinced of its ______." The story was to be completed orally with the word sapphire plus, say, either the suffix -ity or -ic, which the subject had to choose between. There were a score or so of such trials. This is a more subtle way than the one I used to get subjects to make novel stem + suffix combinations. Compared to my results, they reported far less productivity of the vowel laxing and vowel shift which one might expect with such derivations (that is, only a very small number of subjects gave forms such as [səfIrəDi] or [səfIrIk]; most kept the stressed vowel as [aj]). I would guess from this that they were more successful in distracting their subjects' attention from the pronunciation task itself than I was. I suppose the reason productivity varies with the amount of conscious attention a speaker gives to the task is because with more time and attention he has a better opportunity to search his lexicon in order to find a model to analogize on.

More than likely the strata of vocabulary the stem represents, i.e., whether Romance or Germanic, and the numb of syllables in the stem does affect the eligibility of a stem for vowel laxing and the like (although one wouldn't have guessed this by reading most recent generative phonolo of English). This is evident in other results I obtained from this experiment. I found that "between" (Gmc., bisyllabic), "fleece" (Gmc., monosyllabic), and "space" (Romance, monosyllabic) were pronounced without lax vowels l all 26 of my subjects when derived with the "laxing" suffix -ity, i.e., the stem was phonetically unchanged, whereas "methane", "subline", and "supreme" (all bisyllabic and fro the learned Greek or Latin vocabulary) were rendered with l vowels by 12, 9, and 7 subjects, respectively, when derived with laxing suffixes (either -ity or -ify). (For further details, see Ohala, forthcoming.) However, I don't think such factors affected the results I reported here.

Richard C. DeArmond. I have questions to ask as a Slavist. In the history of the Slavic languages if a labial consonant preceded jod, an epenthetic [l] was inserted, or possibly, less likely, jod became lateralized. Whatever the histori process, can you explain the lateralizations that occurred?

John Ohala. No, I cannot explain it.

References

Krohn, R., Steinberg, D., and Kobayashi, L. R. (1972). The psychological validity of Chomsky and Halle's vowel shift rule. Abstract guide of XXth International Congress of Psychology. 1972, Tokyo, p. 405.

Ohala, J. (forthcoming). On phonological experiments.

Notes

1 I accept the view that sound change does indeed spread from word-to-word, gradually moving through the lexicon (of the individual speaker and thus of that of the speech community as a whole), as proposed and/or documented by Sommerfelt (1923), de Groot (1941), Wang (1969), Chen (1972), Sherman (1973) and others.

2 I refer to Sweet's (1888) notion of sound change due to faulty acoustic imitation, Jakobson's (1931) notion of phonologization, and Andersen's (forthcoming) notion of abductive inference.

3 S. R. Anderson (1972) concludes that the glottal consonants in Sundanese are not nasalized but he still allows nasalization to leapfrog over them in some mysterious way. But this conclusion is based on his misreading of Robins' (1957) kymograms: these records show that nasal airflow *does* continue through the glottal consonants.

4 See also: Kiparsky (1971) and Skousen (1972).

5 That subjects treated *detain*, *detention*, as a better model than *explain*, *explanatory*, for the derivation of *obtain* may be due to the fact that there is more in common phonetically and orthographically between *detain* and *obtain*, namely, -*tain* [t^{h}ejn], than there is between *explain* and *obtain*, -*ain*, [ejn].

Niels Ege has also pointed out that the lack of such responses as [ʌbt^{h}ɛnətɔrij] and [ʌbt^{h}ænʃə n] may be because there are no words in English having the sequence [ɛnətɔrij] and relatively few with [ænʃə n], e.g., *mansion*, *expansion*.

6 See, for example, Bloomfield (1933:404-424), Hockett (1958:356, 389), Lehmann (1962:177-192).

7 One might have expected subjects to give [hjuwmejniən] due to the existence of *humane*, [hjuwmejn], except for the fact that there is nowadays only a very tenuous semantic (and thus derivational) connection between *human* and *humane*.

8 See Esper (1925), Greenberg and Jenkins (1964), Hsieh (1970), Krohn, Steinberg, and Kobayashi (1972), Ladefoged and Fromkin (1968), Moskowitz (forthcoming), J. Ohala (forthcoming), M. Ohala (1972 and forthcoming), Schane and Tranel (1970), Sherzer (1970), Zimmer (1969).

References

Ainsworth, W. A. (1972). Duration as a cue in the recognit: of synthetic vowels. Journal of the Acoustical Society America 51:648-52.

Andersen, H. (forthcoming). Abductive and deductive change

Anderson, S. R. (1972). On nasalization in Sundanese. Linguistic Inquiry 3:253-268.

Bendor-Samuel, J. T. (1966). Some prosodic features in Tere In memory of J. R. Firth. London: Longmans, Green and pp. 30-39.

Bloomfield, L. (1933). Language. Holt, Rinehart and Winston

Chen, M. (1972). The time dimension: contribution toward a theory of sound change. Foundations of Language.

Chen, M. (1973). Nasals and nasalization in the history of Chinese. University of California, Berkeley, Doctoral Dissertation. Unpublished.

Chen, M. and Wang, Wm. S.-Y. (1973) Lexical diffusion and phonetic actuation. (This volume).

Chistovich, L. A. (1969). Variations of the fundamental voice pitch as a discriminatory cue for consonants. Soviet Physics: Acoustics 14:372-378.

Chomsky, N. and Halle, M. (1968). Sound pattern of English. Harper and Row.

Czermak (1857). Über das Verhalten des weichen Gaumens beim Hervorbringen der reinen Vokale. Sitzb. d. K. Akad. d. Wiss. Wien, math-naturwiss. Kl. (Cited by Scripture, 1904).

de Groot, A. W. (1941). Structural linguistics and phonetic law. Archives neerlandaises de phonétique expérimentale 17:71-106.

Delattre, P. (1967). Coarticulation and the locus theory. (Abstract). Journal of the Acoustical Society of America 42:1163.

Delattre, P. (1970). Rapports entre la physiologie et la chronologie de la nasalité distinctive. *Actes du X^e Congrès International des Linguistes*, Bucarest, 28 Aug-2 Sept. 1967. Editions de l'Academie de la république socialiste de Roumaine, Bucharest. Vol. 4; pp. 221-227.

Esper, E. A. (1925). A technique for the experimental investigation of associative interference in artificial linguistic material. *Language Monographs*, No. 1.

Fant, G. (1960). *Acoustic theory of speech production.* The Hague: Mouton.

Fritzell, B. (1969). The velopharyngeal muscles in speech. *Acta Otolaryngologica*. Supplement 250.

Fry, D. B. (1960). Linguistic theory and experimental research. *Transactions of the Philological Society (London) 1960*: pp. 13-39.

Fujimura, O. (1971). Remarks on stop consonants--synthesis experiments and acoustic cues. In *Form and Substance*. Akademisk Forlag. pp. 221-232.

Grammont, M. (1939). *Traite de phonétique*. Paris: Libraire Delagrave.

Grandgent, C. H. (1896). Warmpth. *Publications of the Modern Language Association* 11 (new series 4): 63-75.

Greenberg, J. H. and Jenkins, J. J. (1964). Studies in the psychological correlates of the sound system of American English. *Word* 20:157-177.

Haden, E. F. (1938). The physiology of French consonant changes. *Language Dissertations* No. 26.

Haggard, M., Ambler, S., and Callow, M. (1970). Pitch as a voicing cue. *Journal of the Acoustical Society of America* 47:613-617.

Hetzron, R. (1969). Two notes on Semitic laryngeals in East Gurage. *Phonetica* 19:69-81.

Hockett, C. F. (1958). *A course in modern linguistics.* Macmillan.

Hockett, C. F. (1965). Sound change. *Language* 41:185-204.

House, A. S. and Stevens, K. N. (1956). Analog studies of the nasalization of vowels. Journal of Speech and Hear Disorders 21:218-232.

Hsieh, H.-I. (1970). The psychological reality of tone san rules in Taiwanese. Papers from the 6th regional meeti of the Chicago Linguistic Society. pp. 489-503.

Hyman, L. (ed.) (1973). Consonant types and tone. Southe California Occasional Papers in Linguistics No. 1.

Jakobson, R. (1931). Principes de phonologie historique. Trauvaux du Cercle Linguistique de Prague 4:247-267.

Javkin, H. R. and Ohala, J. J. (1972). Epenthetic stops between [l] and following sibilants. (Abstract) Journal of the Acoustical Society of America 53:378.

Joos, M. (1948). Acoustic phonetics. Language Monographs No. 23.

Kiparsky, P. (1971). Historical linguistics. In Survey of Linguistic Science. Ed. by Dingwall, W.O., Department of Linguistics, University of Maryland.

Krohn, R. and Steinberg, D., and Kobayashi, L. R. (1972). The psychological validity of Chomsky and Halle's vowel shift rule. Abstract guide of the XXth International Congress of Psychology, 1972, Tokyo. p. 405.

Ladefoged, P. (1967). Three areas of experimental phonetics Oxford University Press.

Ladefoged, P. and Fromkin, V. (1968). Experiments on competence and performance. IEEE Transactions on audio and electro-acoustics. AU-16:130-136.

Lea, W. (1973). Segmental and suprasegmental influences on fundamental frequency contours. In Hyman (ed.) (1973).

Lehmann, W. P. (1962). Historical linguistics. Holt, Rinehart and Winston.

Liljencrants, J. and Lindblom, B. (1972). Numerical simulation of vowel quality systems: the role of perceptu contrast. Language 48:839-862.

Lindblom, B. (1972). Phonetics and the description of language. Proceedings of the 7th International Congress of Phonetic Sciences, Montreal, 22-28 August 1971. Mouton. pp. 63-98.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. Language 33:42-49.

Lubker, J. (1968). An EMG-cinefluorographic investigation of velar function during normal speech production. Cleft Palate Journal 5:1.

Malkiel, Y. (1966). Genetic analysis of word formation. Current Trends in Linguistics, vol. 3. Mouton. pp. 305-364.

Malkiel, Y. (1973). Phonological irregularities vs. lexical complexity in diachronic projection. In Issues in Linguistics. Papers in Honor of Henry and Renée Kahane. Ed. by Kachru, B.B., Lees, R.B., Malkiek, Y., Peitrangeli, A., and Saporta, S. University of Illinois Press.

Martinet, A. (1955). Économie des changements phonétiques. Editions A. Francke S. A. Berne.

Matisoff, J. (1973). The grammar of Lahu. University of California Publications in Linguistics, No. 75.

Moll, K. L. (1962). Velopharyngeal closure on vowels. Journal of Speech and Hearing Research 5:30-37.

Moskowitz, A. (forthcoming). The acquisition of phonology. Mouton.

Ohala, J. (1971). Monitoring soft palate activity in speech. (Abstract). Journal of the Acoustical Society of America 50:140. (Printed in full in Project on Linguistic Analysis Reports, Second Series (Berkeley), 13:J01-J015.)

______. (1972a) How to represent natural sound patterns. Project on Linguistic Analysis Reports, Second Series, (Berkeley) 16.

______. (1972b). Physical models in phonology. Proceedings of the 7th International Congress of Phonetic Sciences, Montreal, 22-28 August 1971. Mouton. pp. 1166-1171.

Ohala, J. (1973). The physiology of tone. In Hyman (ed.) (1973). pp. 3-14.

______. (forthcoming). On phonological experiments.

Ohala, M. (1972). Topics in Hindi-Urdu phonology. University of California, Los Angeles, Doctoral Dissertat: Unpublished.

______. (forthcoming). The abstractness controversy: experimental input from Hindi. Language.

Passy, P. (1891). Étude sur les changements phonétiques. Paris: Librairie Firmin-Didot.

Postal, P. M. (1968). Aspects of phonological theory. Harper and Row.

Robins, R. H. (1957). Vowel nasality in Sundanese. Studies in Linguistics Analysis. Oxford: Blackwell.

Rousselot, L'abbe (1891). Les modifications phonétiques du langage. Paris: H. Welter.

Schane, S. and Tranel, B. (1970). Experimental design on the psychological reality of a natural rule in phonology. University of Michigan Studies in Language and Language Behavior.

Schourup, L. C. (1973). A cross-language study of vowel nasalization. Working Papers in Linguistics (The Ohio State University) 15:190-221.

Scripture, E. W. (1904). The elements of experimental phonetics. New York: Charles Scribner's Sons.

Sherman, D. (1973). Noun-verb stress alternations: an example of the lexical diffusion of sound change in English. Project on Linguistic Analysis Reports, Second Series, (Berkeley) 17:46-81.

Sherzer, J. (1970). Talking backwards in Cuna: the sociological reality of phonological descriptions. South-western Journal of Anthropology 26:343-353.

Skousen, R. (1972). On capturing regularities. Papers from the 8th regional meeting of the Chicago Linguistic Society. pp. 567-577.

Sommerfelt, A. (1923). Note sur les changements phonétiques. Bulletin de la Société de Linguistique de Paris. 24:138-141.

Sweet, H. (1888). History of English sounds. Oxford.

Wang, Wm. S.-Y. (1969). Competing change as a cause of residue. Language 45:9-25.

Weinreich, U., Labov, W., and Herzog, M. (1968). Empirical foundations for a theory of language change. In Directions for Historical Linguistics. Ed. by Lehmann, W. and Malkiel, Y. Austin: University of Texas Press. pp. 97-195.

Zimmer, K. (1969). Psychological correlates of some Turkish morpheme structure conditions. Language 45:307-321.

PROCEEDINGS of the FOURTH ANNUAL MEETING of the BERKELEY LINGUISTICS SOCIETY

FEBRUARY 18-20, 1978

JERI J. JAEGER — ANTHONY C. WOODBURY
FARRELL ACKERMAN, CHRISTINE CHIARELLO
ORIN D. GENSLER, JOHN KINGSTON,
EVE E. SWEETSER, HENRY THOMPSON,
KENNETH W. WHISTLER
EDITORS

BERKELEY LINGUISTICS SOCIETY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA
~ 94720 ~

Southern Bantu vs. the world: The case of palatalization of labials

John J. Ohala

University of California, Berkeley

Introduction.

Applying typological data to the analysis of particular languages is not new nor is the general principle behind it: the inductive method. Having observed a particular pattern in many languages, positing it in yet another language (given appropriate circumstances) is not wholly unjustified. The purpose of this paper is to demonstrate that one can do a much better job of linguistic analysis if one uses a combination of the inductive and deductive methods, i.e., if one's expectation of a pattern is determined not only by the fact that it has been encountered previously but also because one knows the underlying principles which give rise to it. To illustrate this, I will examine the so-called palatalization of labials, i.e., the shift in place of articulation of palatalized labials or labials followed by a palatal (off-) glide to dentals, alveolars, or palatals (henceforth, for ease of reference, simply 'dentals').

The inductive approach.

A survey of the phonologies of many languages around the world turns up quite a few independent cases of the type of sound pattern exemplified in (1).

(1) [pj, pʲ] → [t, ts, tʃ]
[bj, bʲ] → [d, dz, dʒ]
[mj, mʲ] → [n, ɲ]
etc.

(Occasionally, but not necessarily, intermediate stages may be found, e.g., [ptʃ], [bdʒ], [mn].) Some examples of this pattern are listed in (2) through (10).[1]

(2) Czech (data from Bělič 1966 and Andersen 1973).

Standard Czech	East Bohemian	English gloss
[mʲɛstɔ]	[nɛstɔ]	'town'
[pʲɛt]	[tɛt]	'five'
[pʲi:vɔ]	[ti:vɔ]	'beer'
[pʲɛknʲɛ]	[tɛknʲɛ]	'nicely'

(3) Tai (data from Li 1977).

Siamese	Lungchow	T'ien-chow	English gloss
plaa	pjaa	čaa	'fish'
plau (plaaw)+	pjau	čuu	'empty'
phaai (phaj)+	phjaai	čaai	'to walk'

(+ Accepted current phonemic transcription in Thai.)

Evidently the post-consonantal /l/ changed to /j/ first, then /p (h) j/ changed to the palatal affricate. This same pattern of development is also attested in the Romance languages; see (5), (6) below.

(4) Tibetan (data from Thomas 1948, Benedict 1972, and Chang and Chang 1975).

Old Tibetan	Tzu-ta	Wassu	Mi-li		English gloss
mig~myig	temňa	temniak	nie +		'eye'
byi-ru	ptsyeru				'coral'
	Gyarong	Lha-sa	Lolopho	Ahi	
bya	pyē-pyē	ca [tʃa]	byo	do	'bird'
byi-ba		ci-wa [tʃiwa]			'rat'

(+ If Thomas' 'Mi-li' refers to the Tibetan language known as 'Muli', then this word should rather be given as /nöö/ (Nagano 1957).)

(5) Spanish and Portuguese (data from Malkiel 1963).

Latin	Spanish	English gloss
amplu	ancho	'large, spacious'
	Old Spanish	
implēre	(f)enchir	'to fill'
	Portuguese	
plōrāre	chorar	'to weep'
flamma	chama	'flame'
plānu	chão	'floor; level'
plumbu	chumbo	'lead (metal)'

Other data provide evidence for the following separate stages of development of the Latin pl- cluster in these languages: pl-> pj-> tʃ- > ʃ-.

(6) Italian (data from Jaberg and Jud 1928-1940; transcription simplified).

Roman dialect	Genoese and neighboring dialects	English gloss
[pjeno]	[tʃena]	'full'
[pjanta]	[tʃanta]	'to plant'
[er fjato]	[uʃa]	'breath'
[bjanko]	[dʒanku]	'white'

(7) French

Latin	French	English gloss
sapius	sage [saʒ]	'wise'
rubeus	rouge [ʁuʒ]	'red'
rabies	rage [ʁaʒ]	'rabid'
cavea	cage [kaʒ]	'cave'
Proto-Germanic		
laubja	loge [loʒ]+	'arbor'; 'small house'

(+ Cf. English 'lobby' and 'lodge', the first having a Germanic origin, the second French.)

(8) Bantu (data from Guthrie 1967-1970).

Proto-Bantu	Tonga	Xhosa, Zulu		English gloss
*pia	phya	-tʃha		'new'
	Kaonde	Sena	S. Sotho	
*biad	-βyal-	-bzaɽ	dʒal	'plant'
	Venda			
*piu	tswhu			'knife'

(9) Classical Greek (data from Meillet and Vendryes 1924).

Pre-Classical Greek	Classical Greek	English gloss
*g^wam-yo	βαίνω (cf. Latin venio)	'I come'
*kom-yo- (cf. Latin cum)	κοινός	'common'
*χαλεπ-yω	χαλέπτω	'provoke'
*θαφ-yω	θάπτω	'bury'

(10) Gwari (No. Nigeria) (data from Hyman and Magaji 1970).

Kuta	Ganagana	Nupe	English gloss
byè	dywè	dzò	'sow'
byì	dywī	dzū	'bury'
ōpyá	ēpʃā	ētswā	'moon'

The deductive approach.

To understand why such changes occur, it is instructive to examine the spectrographic pattern of palatalized labials and compare this with the patterns of plain labials and dentals. Figure 1 shows tracings of spectrograms (the originals published by Fant 1960) of the Russian CV syllables [ba], [bʲa], and [da]. In examining this figure it is necessary to keep in mind the fact that place of articulation cues reside in the second formant (F2) transitions and in the noise bursts. That being the case, it is of interest to see in the figure that the F2 transition for the palatalized labial is more similar to that for the dental than it is to that for the plain labial.[2] Undoubtedly in this instance the noise burst from the release of the stop is a sufficient cue to the labiality of the palatalized labial in spite of the dental-like F2 transition. If a listener were to miss the noise burst cue, however, the consonant would very likely be taken for a dental. Moreover, the impression that such stops were dentals or palatals would be reinforced by any fricative noise generated from the rush of air through the narrow palatal constriction. A sound mistaken for a dental or palatal is likely to be repeated as such. Thus a sound change could occur.

But why should the palatal constriction, a secondary articulation, have a greater influence on the consonantal F2, than the labial constriction, the primary articulation? The beginnings of an answer to this question can be seen in the nomogram in Figure 2 (again, from Fant 1960). The nomogram shows the formant frequencies that would be produced as one varies both the position of the constrictions in the vocal tract and the accompanying lip opening. As can be seen, although the F2 frequency is generally susceptible to change due to both variations in place of constriction and variations in lip opening, its frequency due to a constriction in the palatal region (see arrow) is largely independent of the lip opening. A palatal constriction, even though a secondary articulation, will be the primary determinant of the F2 frequency and will produce a frequency much like that of a dental consonant.

In the case of nasal consonants, there are related but slightly different reasons why the nasal murmur of a palatalized [m] or an [m] coarticulated with the palatal vowel [i], would be acoustically similar to an [n] or [ɲ] (Ohala 1975).

The acoustic similarity of palatalized labials (or labials followed by or coarticulated with palatal vowels) and dentals is

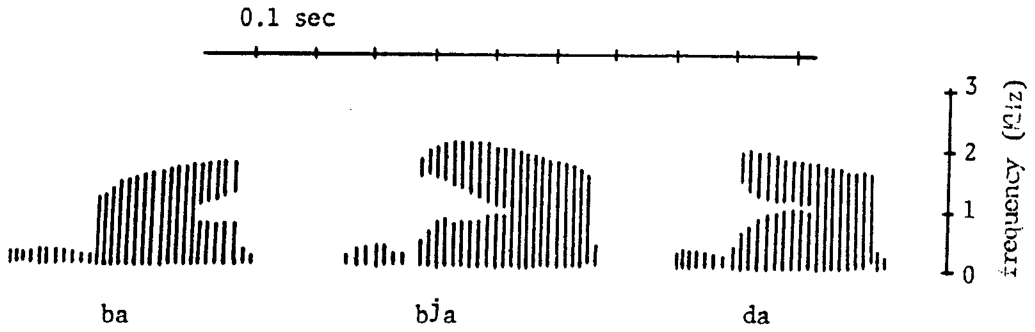


Figure 1. Tracings of spectrographic patterns for the Russian syllables [ba], [bʲa], and [da] (from Fant 1960).

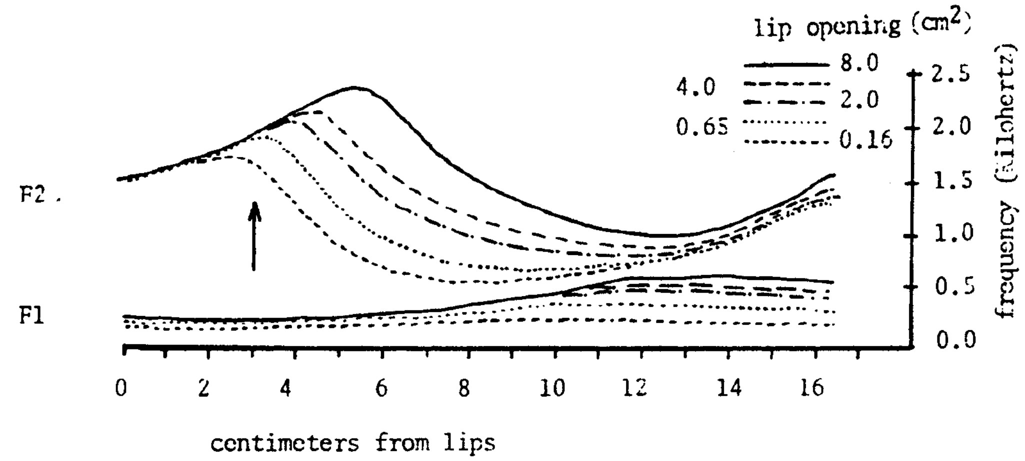


Figure 2. A nomogram (from Fant 1960) showing the frequencies of the first two formants (F1 and F2) that would result from constrictions in various places in the vocal tract (horizontal axis) and from variations in the lip aperture (the parameter). The arrow marks the approximate position of a palatal constriction.

also suggested by the results of various speech perception studies. Lyublinskaya (1966) in studying the confusability of VC transitions (i.e., where there were no bursts to aid identification of the consonant) found palatalized labials were up to 30% more likely to be confused with dentals than were plain labials. Winitz, Scheib, and Reeds (1972) published confusion matrices of CV sequences (where C = p, t, k, and V = i, a, u,) obtained under two conditions: stop burst only and stop burst plus 100 msec of the following vowel. The sequence /pi/ was one of the few stimuli that showed a relatively large decrease in identifiability when 100 msec of vowel was included. In accord with the phonological evidence presented above, most of the confusions of this syllable were with the syllable /ti/. (These two studies, as well as that of Wang and Fillmore 1961, also reveal a strong tendency for labials preceded or followed by /u/ to be misheard as dentals. This effect will be discussed further below.)

In addition, the auditory similarity of the syllables [mi] and [ni] has been revealed in other perceptual studies by House (1957), and Gay (1970).

Analysis of the Southern Bantu Palatalization of labials.

We can now attempt to apply this information to an analysis of some rather unusual cases of palatalization of labials in several Southern Bantu languages, e.g., Zulu, Xhosa, Tswana, Sọthọ, Venda, Pedi. In these languages the palatalization of labials (and some non-labials) plays an important role in a number of morphological processes, all of which, however, are manifested in about three or four types of (surface) phonological environments. (The data to follow on Southern Bantu are taken from Doke 1926, Meinhof and Warmelo 1932, Cole 1955, Tucker 1929, Jacottet 1927, Guthrie 1967-1970).

One of these types is exemplified in (11). Addition of the causative suffix -ya triggers palatalization. Although we might

(11)		verb stem + a	causative	English gloss of stem
	Sọthọ	ɛta	etsa	'go'
		kɛna	keɲa	'go in'
	but:	ɬapa	ɬatswa	'wash'
	Tswana	-tlhaleɸa	-tlhaletʃhwa or -tlhaletsha	'become wise'
		-nat'eɸa	-nat'etʃhwa or -nat'etsha	'become pleasant'

wonder a bit as to where the w came from in the derived forms of the last three verb stems (a point that will be addressed later)[3], this seems to be a normal development; it parallels the kind of change we have seen before in many other languages.

Another environment for palatalization is that exemplified in (12) viz., addition of the diminutive suffix -ana. This is a

(12) Diminutive formation of nouns by suffixation of -ana.

	Noun stem	Diminutive	English gloss of noun stem
Sotho	lebese	lebesana	'milk'
but:	moriɸi	moritsʰwana	'saucer'
Zulu	u:pha:phɛ	u:phaʃa:na	'feather'
	u:ʃu:bu	u:ʃudʒa:na	'meal-water'
	iŋk'a:bi	iŋk'atʃ'a:na	'ox'
Tswana	tlʰap'i	tlʰatʃ'wana (dial. tlʰap'ʃana)	'fish'

very unnatural and unexpected environment for this process--given what we have reviewed above. There is no apparent reason why the vowel /a/ or any other phonetic property of the suffix /ana/ should lead to palatalization.

Finally,[4] the most unexpected case of palatalization of labials is that triggered by the addition of the passive suffix -wa to verb stems. Examples are given in (13). Not only should a

(13) Passive formation by suffixation of -wa.

	Verb stem + a	Passive	English gloss of verb stem
Sọthọ	lesa	leswa	'leave'
	rɛka	rɛkwa	'bury'
but:	bɔɸa	boʃwa	'tie on back'
	tseba	tseʒwa	'know'
	thopa	thotʃwa	'capture'
	ɬɔpha	ɬotʃwha	'heap up'

following /w/ not cause palatalization of labials, it is the one segment most likely to reinforce the labiality of labials.

A digression on [w].

It might be asked how I can justify that last statement given the evidence mentioned above that in at least 3 perceptual studies labials in the environment of the vowel /u/ were often misheard as dentals. The justification is that /u/, although phonetically close to /w/, differs in the important respect that, unlike /w/, in the environment of a labial it need not have a rapid change in formant frequency. Lyublinskaya presents evidence that one of the cues for place of articulation of a consonant is direction of F2 transition; presumably a negative (downward) transition for labials and a lack of transition for dentals. Thus the lack of strong F2 transitions between /u/ and a labial could lead to the consonant being taken for dental. Since /w/ necessarily has rapid formant transitions this factor should not apply. There is, moreover, phonological evidence that /w/ should enhance the labiality of labials.[5]

A labial (velar) glide (sometimes < u + V) adjacent to non-labial obstruents generally precipitates a shift to the labial place of articulation (if it leads to any change at all). Examples of this are not uncommon; (14) and (15) provide a few of many cases that could be cited.

(14) Classical Greek (data from Meillet and Vendryes 1924, Meillet 1964).[6]

PIE root	Latin	Greek	English gloss
*yĕkʷr̥	iecur	ἧπατος	'liver'
*ekwōs	equus	ἵππος	'horse'
	Sanskrit		
*gʷiwos	gayaḥ	βίος	'life'

(Cf. also, βαίνω in (9) above.)

(15) Gujarati (data from Turner 1921).

Middle Indic	Gujarati	English gloss
dvāra	bār	'door'
dvē	bɛ̄	'two'
-tvana	-pan	(suffix)

References to further such data can be found in Ohala and Lorentz (1977) and Ohala (forthcoming). These same sources as well as Durand (1956) discuss some of the auditory-acoustic reasons for these sound patterns.

Such data would lead us to expect [w] to reinforce the labiality of labials since it has the power to cause non-labials to become labials.

Previous analyses.

There is a fairly extensive literature on these Southern Bantu languages and I have not had access to most of it. Nevertheless, if there has been any analysis which succeeds in pulling all instances of palatalization together under one rule, it has not found its way into the standard reference works on Bantu. Meinhof and Warmelo (1932) and even Guthrie (1967-1970), for example, two major contributors to the reconstruction of Proto-Bantu, simply list for these Southern Bantu languages (what amounts to) sound changes of the type *p → t / _j, w; they apparently saw nothing unusual in the presence of the w in the environment. Meinhof does offer an explanation for the w or labialization that often remains after the palatalization of labials occurs:

> A very peculiar process is that by which sounds to some extent exchange their quality, each giving up some of its own and assuming those of the other. Thus in Sotho ƒya becomes swa. The first sound, ƒ, is a voiceless labial fricative, the second y is a lingual (or more accurately palatal) semi-vowel. The first sound becomes s, i.e. a lingual (strictly an alveolar) voiceless fricative, the second becomes w, i.e. a labial semivowel [16].

However, he seems not to have applied this analysis to cases such as, e.g., Sotho bɔΦa + wa > boʃwa, where there is no surface y to account for the ʃ.

Doke (1926:139ff) attributed palatalization in the passive to the process of dissimilation since, as it happens, most of these Southern Bantu languages do not have (or permit?) sequences of labial + w. In order to avoid this supposedly forbidden sequence (which would result upon addition of the passive suffix -wa to a stem ending in a labial) speakers, he reasoned, must have shifted the labials to dentals.

In 1970 Talmy Givon and Erhard Voeltz recognized the need to unify all the various instances of palatalization in Southern Bantu under one process triggered by a palatal glide. They found evidence for the 'missing' palatal glides in all the relevant cases of palatalization. They incorporated their views and evidence in various lectures (T. Givon, personal communication).

Stahlke (1976) in arguing for the notion of 'segmental fusion' (essentially that expressed by Meinhof in the above quote[7]), cited Tswana data such as that in (11) and (12) and presented evidence that the causative, passive, and diminutive formations could all be accounted for by one basic rule which involved exchange of features between the labial consonant and a following (sometimes reconstructed) palatal segment.

Herbert (1977) took issue with Stahlke's analysis and rather argued that these cases were morphologically not phonologically conditioned (that is, not phonetically based). His arguments were based primarily on three points: a) that the alternations observed are phonetically unnatural, b) that they applied only in certain morphologically, not phonologically-defined environments, and c) that they show many exceptions and much free variation.

In what follows I propose to provide additional evidence for the Givon-Voeltz-Stahlke analysis. I do not dispute Herbert's claim that these alternations are <u>now</u> activated by specific morphological environments (e.g., diminutive, passive formations) not by phonetic environments. But this is in no way contradictory to the claim that they all had a common phonetically natural origin.

I have already presented evidence for the phonetic naturalness of labials shifting to dentals when followed by palatal glides. What is necessary, then, is to assemble the evidence that palatal glides can be found in earlier stages of the passive and diminutive suffixes.

The history of the passive.

As was pointed out by Stahlke the passive suffix in Tswana has two forms: -wa and -iwa. One can make the case that there was once only one, -iwa. Prosodic factors presumably contributed to the creation of two forms. Zulu, Tswana and Venda for example, use the -iwa form primarily with monosyllabic verb stems but -wa with most polysyllabic stems. We can guess that the -iw coalesced in the latter case, perhaps to [j], perhaps to [ɥ]. Indeed, with simultaneous labialization there is not much difference between these two. As Tucker points out, the labialization of consonants, although indicated in conventional phonetic transcriptions as a w after the sound, is in fact more a prosody of labialization that persists through the segment. If we indicate this prosody as a superscripted line above the consonant, as is done in the Firthian Prosodic tradition, it is easier to see why [j̄ʷ] is equivalent to [ɥ]. Pedi probably still shows some remnants of this early development; see (16).

(16)	-rɛka + wa	>	-rɛkwa	'buy'
	-ripa + wa	>	-ripɥa	'cut'

Assuming that all labial consonants are intrinsically labialized in these languages (which explains why labials cannot support distinctive, i.e., extrinsic, labialization[8]), we may speculate that the two passive forms originated more or less in the sequence indicated in (17).

(17)	[rɛkiwa]	[rip̄iw̄a] ≡	[ripywa]	
	↓		↓ [ripyɥa]	via glide assimilation
	[rɛkwa]		↓ [ripɥa]	via vowel deletion

(On this point one wonders what use to make of the comment by Jacottet (1927:110) that an older, rarer form of the Sotho passive suffix exists: -uwa, and that it is used mainly after labials? It could validate the claim that labials tend to round the following vowel--especially when it is assisted in this by the -w- --but unfortunately it does not prove that the resulting rounded vowel would be [y].)

Another possible route for the development of a palatal glide after labials is suggested by the passive forms in Venda given in (18) (from Ziervogel and Dau 1961:37).

(18)	-bɛba + wa > -bɛbɣwa	≡	[bɛb̄ɣ̄ʷa] ~ -bɛbja	'bear'

The [ɣw] is described as having very little frication.

Since the listeners would expect inherent labialization after the labial b, they may have ignored (i.e., treated as redundant or non-distinctive) the labial part of the following labial velar glide w. What would remain to be considered distinctive in this glide, then, would be the velar component, [ɣ]. This is quite speculative, of course, and, in any case, does not reveal how [ɣ] could change to [j] in the environment given or, indeed, whether the [j] is from the [ɣ].

The history of the diminutive.

In the case of the diminutive formation, as Stahlke pointed out, there is another common form of the suffix which has a palatal initial, -ɲana, and it is possible to trace this and the -ana form to -jana. This morpheme exists in many Bantu languages and means 'child'.

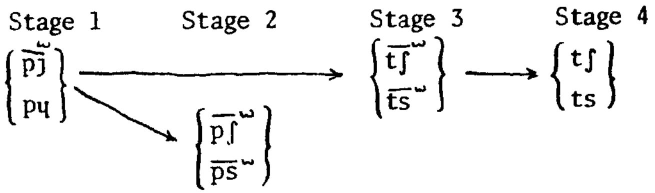
Further evidence for the elided j comes from traces of its influence on non-labials, e.g., in Tswana logoŋ 'piece of wood' + dim. suffix > logoɲana. The change of [ŋ] to [ɲ] is plausible only if we assume a palatal glide in the suffix.

Possible additional evidence for this point comes from the participation of the diminutive suffix in the pattern of vowel harmony in Sǫthǫ. In that language it seems that the lower mid vowels shift to higher mid vowels when followed in the next syllable by a high vowel (either i or u) or--and this is the curious part--by labialized consonants formed by passivization or diminutivization, e.g., selɛpɛ 'axe' + dim. > seletswana (transcription simplified). If we can be sure that simple labialization itself doesn't effect vowel harmony (the evidence isn't very clear on this point, but it seems not to) and if we can trust Tucker's transcription of the Sǫthǫ vowels (Cole 1955:xxvii, warns us not to) it would seem that the shift of [ɛ] to [e] in the above example could be attributed to the once-present palatal segment.

The source of the w.

Returning to the question raised earlier, 'where did the w or labialization come from after the palatalization had shifted the labial to a dental?', we can agree with Meinhof and Stahlke that it came from the original labial consonant.[9] From this it follows that the w in the phonetic transcription of passive forms such as tseʒwa has a different source from the w in forms such as rɛkwa. In the former case it comes from the labial in the stem tseba; in the latter it comes from the passive suffix itself -(i)wa. It is not necessary, though to view the process as entirely one of featural exchange as Meinhof and Stahlke do, at least as far as preservation of the labialization is concerned. It is simpler to assume that labialization spanned the stem-final consonant originally (since, as was mentioned, these labials were probably intrinsically labialized) and it spanned the stem-final consonants after the shift in place of articulation as well. Using the prosodic transcription of labialization, (19) presents a probable scenario for the change, taking up where (17) and (18) left off.

(19)

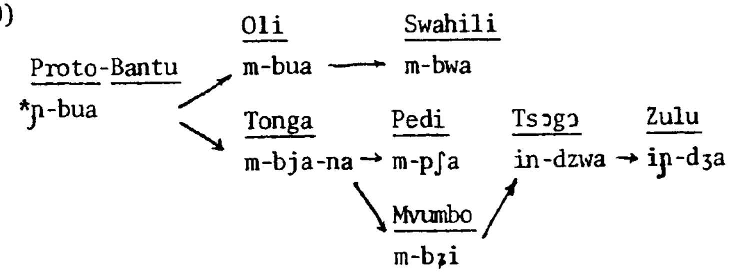


Evidence for Stage 2 comes from, among other things, the passive forms in 'old fashioned' Sotho, e.g., bɔɸa + wa > bɔɸʃwa, as well as some of the variants given above in (12). It should be emphasized that Stage 2 is a possible but not a necessary intermediate stage between Stages 1 and 2. In the perceptual tests reviewed above, subjects' misperceptions, which may be regarded as the stuff sound changes are made from, were abrupt--hearing a p as a t--and did not involve any intermediate stages.

Evidence for the development from Stage 3 to 4, i.e., loss of labialization, was presented earlier (see 11); Cole (1955:43) testifies that among forms showing this variation, those with labialization are older.

Some support for the scheme in (19) comes from the variant reflexes of the Bantu word for 'dog' as given in (20). Here it

(20)



was a palatal nasal noun class prefix ɲ which sometimes exerted a palatalizing influence on the following consonant. The arrows here do not imply direct genetic development but rather that the form near the head of the arrow represents in some sense a further development than the one near the tail.

Conclusion.

Why is it better to combine the inductive and deductive approaches in linguistic analysis? Because they complement each other. In the present case, the inductive approach could tell us that j is a catalyst in the shift of labials to dentals and w helps to shift dentals and velars to labials but it could not tell that in principle it is unlikely that w could also help shift labials to dentals. The deductive approach, in this case reference to the underlying phonetic factors which cause perceptual ambiguity, can tell us which kinds of misperceptions (which might lead to sound changes) are more likely than others. Nevertheless, the deductive approach works perfectly only if our knowledge of the underlying principles of speech production and perception is also perfect. Since this undoubtedly will never be the case, our deductions based on current knowledge may sometimes be erroneous. Only when we find a match between our deductions and inductions can we have some increased confidence that we are on the right track.

Acknowledgements.

The research for this paper was supported by grants from the National Science Foundation and the Committee on Research, University of California. For useful tips and advice I thank Andy Allen, Steve Baron, James Bisso, Talmy Givon, Leon Jacobson, Jeri Jaeger, Ian Maddieson, Yasuhiko Nagano. I especially thank my colleagues Madison Beeler, Carol Justus, and Yakov Malkiel for help with the literature on Greek and Romance. Any errors of fact or interpretation are my responsibility.

Footnotes.

1 In general, I retain the transcription used by the source of the data. This creates potential problems only when palatal glides are sometimes represented as y and sometimes as j. Given the emphasis of this paper and the context of a particular symbol, no ambiguity should arise. When square brackets surround a word IPA transcriptional conventions are followed, i.e., [j] for a palatal glide, [y] for a high front rounded vowel, [ɥ] for a labial palatal glide.

2 The same high F2 is characteristic of labials adjacent to the palatal vowel [i] (Lehiste and Peterson 1961, Fant 1973, Öhman 1966).

3 Leaving aside, for the moment, where the w came from in these forms, the presence or absence of the labialization in the latter two (the Tswana) cases provides a ready explanation for the alternation between the phonetic quality of the resulting affricates [tʃ hw] and [tsh]. As is well known, the presence of lip rounding effectively lowers the resonant frequencies of

the vocal tract. Thus the sibilant noise generated will have a higher or lower center frequency depending on whether or not there is labialization. From such an initial allophonic difference it is quite plausible to find the development of distinctive [ʃ]-type vs. [s]-type fricative releases to the affricates since the primary difference between these fricatives is in their low vs. high center frequencies. In this regard we can note that the lip rounding accompanying English [ʃ] and [tʃ] may not be entirely coincidental: it helps to keep these sounds as distinct as possible from [s].

4 I pass over three other morphological processes that involve palatalization of labials: the formation of the perfect tense of verbs, the formation of the locative of nouns, and the action of the singular prefix of the 5th noun class li > le (and occasionally some other similar prefixes). In general, the analysis of these cases is less controversial than that of the passive and diminutive and, in some cases, was well understood by Bantuists early on (cf. Tucker 1929:85ff).

5 [m] in the environment of [w] or [u] is liable to shift to [ŋ] but not [n] or [ɲ]. The reasons for this, which are relevant only to nasals, are given by Ohala and Lorentz 1977a, b. In view of this it is interesting to note the varying fate of stem-final [m] when it is subject to the same derivations that palatalize the obstruents. The result seems to hinge in part on whether labialization is retained or not. If it is not, we find only [ɲ]; if it is retained we can find either [ɲ] or [ŋ]. Thus, e.g., Zulu int'a:mɔ 'neck' + dim. > int'aɲa:na, but Sotho lelene 'tongue' + dim. > leleŋwana. One could venture the prediction that [ɲw] sequences are unstable and will shift either to [ɲ] or [ŋw].

6 Before front vowels, however, PIE labial velars generaly become Greek dentals, e.g., cf. Latin que 'and', but Greek τε. See Allen (1957) for an interesting discussion of this exception and of exceptions to the exception.

7 See also Henderson 1975 for similar views.

8 The Southern Bantu languages are in no way unusual in not having distinctively labialized labials or labials followed by w. Even in English, clusters of labial + w exist only by virtue of some rather uncommon loanwords, e.g., bwana. Many other examples could be cited (see Ohala, forthcoming).

9 Muddying the issue somewhat is the fact that some non-labial consonants also show labialization in these derivations. Such cases, however, can be attributed to a stem-final rounded vowel, e.g., Tswana lekoto 'leg' + dim. > lekotwana.

References.

Allen, W. S. 1957. Some problems of palatalization in Greek. Lingua 7.113-133.

Andersen, H. 1973. Abductive and deductive change. Lg. 49.765-793.

Bělič, J. 1966. K otázce zaniklého litomyšlského t, d, n za některejší p', b', m'. Slavica Pragensia 8, Acta Univ. Carolinae, Philologica, 1-3.

Benedict, P. K. 1972. Sino-Tibetan: a conspectus. Cambridge Univ. Press.

Chang, B. S. and Chang, K. 1975. Gyarong historical phonology. Bull. of Inst. of History and Philology. Academia Sinica. 46.3.391-524.

Cole, D. T. 1955. An introduction to Tswana grammar. London: Longmans, Green.

Doke, C. M. 1926. The phonetics of the Zulu language. Bantu Studies 2, Special Number.

Durand, M. 1956. Du rôle de l'auditeur dans la formation des sons du language. J. de Psych. Norm. et Path. 52.347-355.

Fant, G. 1960. Acoustic theory of speech production. The Hague: Mouton.

Fant, G. 1973. Stops in CV-syllables. In G. Fant. Speech sounds and features. Cambridge: MIT Press. Pp. 110-139.

Gay, T. 1970. Effects of filtering and vowel environment on consonant perception. JASA 48.993-998.

Guthrie, M. 1967-1970. Comparative Bantu. 4 vols. Gregg.

Henderson, E. J. A. 1975. Feature-shuffling in Southeast Asian languages. Paper presented at 8th Sino-Tibetan Conference, Berkeley.

Herbert, R. K. 1977. Morphophonological palatalization in Southern Bantu: a reply to segmental fusion. Studies in African Linguistics 8.143-171.

House, A. S. 1957. Analog studies of nasal consonants. JSHD 22. 190-204.

Hyman, L. M. and Magaji, D. L. 1970. Essentials of Gwari grammar. Occasional Papers No. 27, Institute of African Studies, Univ. of Ibadan.

Jaberg, K. and Jud, J. 1928-1940. Sprach und Sachatlas Italiens und der Südschweiz. Zurich.

Jacottet, E. 1927. A grammar of the Sesuto language. Bantu Studies 3, Special Number.

Lehiste, I. and Peterson, G. E. 1961. Transitions, glides, and diphthongs. JASA 33.268-277.

Li, F. K. 1977. A handbook of comparative Tai. Univ. Press of Hawaii.

Lyublinskaya, V. V. 1966. Recognition of articulation cues in stop consonants in transition from vowel to consonant. Soviet Physics--Acoustics 12.185-192.

Malkiel, Y. 1963. The interlocking of narrow sound change, broad phonological pattern, level of transmission, areal configuration, sound symbolism. Archivum Linguisticum 15.2.144-173; 3.1-33.

Meillet, A. 1964. Introduction à l'étude comparative des langues indo-européennes. University of Alabama Press.

Meillet, A. and Vendryes, J. 1924. Traite de grammaire comparée des langues classiques. Paris: Librairie Ancienne Edouard Champion.

Meinhof, C. and Warmelo, N. J. V. 1932. Introduction to the phonology of the Bantu languages. Berlin: Dietrich Reimer/ Ernst Vohsen.

Nagano, Y. 1975. Phonemics of the Muli language. Annual Report of Japan Assoc. for Tibetan Studies No 21. Pp. 10-11.

Ohala, J. J. 1975. Phonetic explanations for nasal sound patterns. In: C.A. Ferguson, L. M. Hyman, and J. J. Ohala (eds.) Nasálfest. Papers from a Symposium on Nasals and Nasalization. Stanford: Language Universals Project. Pp. 289-316.

Ohala, J. J. Forthcoming. The origins of sound patterns in language.

Ohala, J. J. and Lorentz, J. 1977a. The story of [w]: an exercise in the phonetic explanation for sound patterns. BLS 3. 577-599.

Ohala, J. J. and Lorentz, J. 1977b. Acoustic reasons why labio-velars are both labials and velars. JASA 61. 932.

Öhman, S. E. G. 1966. Coarticulation in VCV utterances: spectro-graphic measurements. JASA 39. 151-168.

Stahlke, H. F. W. 1976. Segment sequences and sequential fusion. Studies in African Linguistics 7. 41-63.

Thomas, F. W. 1948. Nam. An ancient language of the Sino-Tibetan borderland. Publications of the Philological Society [of London]. 14.

Tucker, A. N. 1929. The comarative phonetics of the Suto-Chuana group of Bantu languages. London: Longmans, Green.

Turner, R. L. 1921. Gujarati phonology. JRAS, 1921, part 3. Pp. 329-365; 505-544.

Wang, W. S-Y. and Fillmore, C. J. 1961. Intrinsic cues and consonant perception. JSHR 4. 130-136.

Winitz, H. Scheib, M. E., and Reeds, J. A. 1972. Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech. JASA 51.1309-1317.

Ziervogel, D. and Dau, R. S. 1961. A handbook of the Venda language. Pretoria: Univ. of South Africa.