MODERN APPLIED LINGUISTICS

John J. Ohala

Phonology Laboratory, Department of Linguistics
University of California
Berkeley, California 94720 USA

"Knowledge is power." --Francis Bacon

1. INTRODUCTION

In this paper I discuss and illustrate some of the actual and potential modern applications of linguistic knowledge, focussing in particular on phonetic and phonological aspects useful in the context of *man-machine* communication. Before doing this, however, it will be useful to review, very briefly, applied linguistics of the past.

1.1. Pre-19th Century Applied Linguistics

Although the term linguistics and its equivalent in other languages, e.g., German Sprachwissenschaft, Hindi bhasha vigyan, Arabic السنية) are relatively modern coinages, curiosity and speculation about the origin, structure, and behavior of language and its parts has been active since prehistoric times as evidenced by the myths found in most cultures regarding the origin of the names of things and of human language itself [1]. In fact, the history of successful applications--the "power" referred to in the epigraph--that follow from knowledge of the workings of language is long and distinguished. The invention of writing is perhaps the greatest accomplishment of applied linguistics (using this term quite broadly), especially in those cases where the symbols used were linked in some way to the phonetic composition of the words they represented, i.e., in alphabets and syllabaries. An even earlier application of the knowledge of the structure of language may be found in such devices as rhyme, assonance, alliteration, and meter. These are commonly thought of as serving only an esthetic end but they also have a very important utilitarian function. In non-literate societies, which, sadly, are still quite abundant, large quantities of the accumulated wisdom of the group are preserved only by being committed to memory. The oral "documents" containing this wisdom are more easily memorized when they are expressed in sentences using rhyme, fixed metrical structure, etc. For this reason the earliest literature of every culture is composed more as poetry than prose, e.g., the Indian Veda, Homer's Iliad, the Qur'an, Beowulf.

It remained for the grammarians among the Hindus, the Greeks, the Chinese, and the Arabs, among others, to initiate what may be regarded as the *formal* scientific analysis of language, i.e., where the study is objective rather than impressionistic [2]. The Hindu grammarian, Paṇini, some 23 centuries ago developed a system for describing the elements of Sanskrit and their behavior, which system became the model for modern descriptive linguistics when it was discovered by scholars in the 18th and 19th centuries. The Greeks made the useful modification of adding vowels to the Semitic alphabet. Grammars of the Greek language, notably that by Dionysius Thrax, served as models for Western grammars for two millenia afterwards. The Greeks also initiated the discipline of etymology, the study of the origin of words, although their first attempts at this were understandably naive. The Chinese grammarians did extensive lexicographic work and developed a phonological descriptive system that is still used by scholars of Chinese and other tone languages of East and Southeast Asia. In the eighth century the Arab grammarians, al-Khalīl Ibn Ahmad (الخليل بن احمد) and Sībawaihi (سيبويه), produced a very complete and in many respects a quite modern description of Arabic phonetics and phonology. From all of these efforts some very important applications resulted: dictionaries, improved orthographies (including shorthands), improved methods of second language instruction (via grammars, lexicons, etc.), and, in a few cases, improved therapeutic methods or prosthetic devices for those with speech and hearing defects [3,4].

It might be difficult for us to appreciate fully the importance of the achievements of applied linguistics just reviewed because they were done so long ago and because we now take them for granted. But considering the overall level of technology present in the world in the first few millenia of the historical era, these achievements were innovations which had a major impact on the world, furthering developments in all other areas of learning: science, technology, law, philosophy, etc. It is possible that some of their "invisibility" is due to the fact that their influence on other fields was (and is) indirect. These inventions did not by themselves permit, for example, improved navigation but the fact that written accounts of innovations in navigation were available allowed such developments to be quickly disseminated. Furthermore, the existence of dictionaries and grammars allowed accounts written in one language to be understood by native speakers of another, thus permitting the original discovery to be spread throughout the world. The tremendous influence these applications of linguistic science had is easy to understand when we take into consideration that language is, after all, the principal medium for the storage and dissemination of the accumulated wisdom of the human race. Naturally, any technical improvement which permits these functions to be accomplished more efficiently and more accurately will be of incalculable benefit to mankind. May the achievements of modern applied linguistics be as valuable as those of the past!

1.2 Nineteenth Century.

Naturally, linguistic science continued to refine its knowledge base and to accumulate a wider range of information on language structure, particularly in the area of phonetics. In this it was helped by the allied disciplines of physics and medicine. Thus it was possible to do an even better job at the applied linguistics tasks pioneered centuries earlier.

It was in the 19th century, however, that linguistics successfully addressed a completely new problem in the domain of language. Whereas most previous linguistic studies focussed on the relations that could be found in the elements of a single language, they tried now to account for the diversity of languages or equivalently, to show in a quite elegant way how many modern languages derive from a smaller number of parent languages no longer extant. This enterprise is called comparative or historical linguistics. Although this achievement did not immediately (or even yet) have any great practical application, at least not in technical areas, I believe it represents one of the most important scientific advances of the modern age. Furthermore I propose to describe below how the data from historical linguistics can, in fact, have practical application. Therefore, it will be useful to examine the methods and results of this sub-discipline in linguistics.

I should first justify my characterization of what comparative linguistics did as scientific, because there are those who would argue that there is no "science" outside of the so-called "hard" sciences such as physics, chemistry, and geology. I think what differentiates scientific undertakings from others like, say, literary criticism, which has no pretensions to being scientific, is the practice of supporting hypotheses or claims by objectively verifiable (i.e., empirical) criteria as opposed to such subjective criteria as the eloquence with which the claim is stated, the authority or prestige of the person or group making the claims, etc. What makes a criterion "objective" is its freedom from extraneous influences or distortions--including the biasses of the investigator--which reduce its evidential value. For example, astronomers know not to take the apparent position of a star near the horizon as an indication of its true position since refraction of its light due to the earth's atmosphere makes it seem higher than it really is. The objective value of its position is one which is corrected for atmospheric refraction.

Prior to the 19th century several people had called attention to the similarity that existed between certain words and grammatical processes in the modern and the classical European languages such that some genetic relation between them was suspected. For all anyone knew at the time, however, there could have been a number of other causes of such similarities: the words could have been borrowed, there might be some inherent property of the meaning of the words which motivated the choice of sounds used to form them (i.e., a manifestation of "sound symbolism" [5]), or the similarities might be due to chance. For a claim of genetic relationship between languages, all of these other possibilities represented uncontrolled extraneous factors and thus made the claim unscientific. Examples of chance similarity abound; consider, for example, the following false cognates from unrelated languages: Modern Greek /mati/[1] "eye" and Malay /mata/ "eye" [6], English /neɪm/ "name" and Japanese /namae/ "name". Although it is difficult to quantify this sort of thing, the probability of finding a few such phonetically and semantically similar pairs in any two languages seems to be rather high. The leaders of the comparative linguistics revolution in the 19th century, e.g., Rasmus Rask of Denmark, Franz Bopp and Jacob Grimm of Germany, had an intuitive (not a quantitative) understanding of such probabilities. However, they also realized intuitively that finding 5 cognate pairs by chance is less likely than finding one and finding 20 by chance is highly improbable, such that if a sufficient number of points of identity can be found between two or more languages one could rule out cognation due to chance [7]. The forms in Table 1 illustrate this. As mentioned, it is not completely unlikely that one cognate pair could be found by chance when comparing the vocabulary of any two languages but finding a cognate triple when comparing three languages is less likely and finding several such triples by chance is very much less likely. (In fact, many more such triples could be given than those in Table 1 and, moreover, cognate groups can be given involving many more languages than just these three.) To clinch the case, these linguists could point to a quite regular correspondence between the sounds in the cognate words. In the words given, Latin and Greek initial p always corresponds to English f, t to th (phonetic [θ]), and k (or the equivalent c) to h. After s the Latin and Greek stops are the same as those in the corresponding English words except in the case of the velars where the sk sequence corresponds to sh (phonetic [ʃ]) in English. The great number of intersecting points of similarity in vocabulary of such a basic nature (i.e., unlikely to be borrowed) and on words with such different meanings (thus ruling out the notion that the meaning somehow "determined" the sounds used), proves in a scientific fashion that there is a genetic relation between these languages.

This method is simply an intuitive application of the statistical law that the probability, P, of the collection of events, X = x1, x2,...xi,...xn, is the product of the probabilities of the separate events:

$$\prod_{i=1}^{n} p(x_i) \qquad (1)$$

Although not used in a quantitative way by the historical linguists, it is a method capable of yielding a level of objectivity and accuracy of evidence comparable to that found in the "hard" sciences. Charles Darwin--several decades after the linguists--used this same method of marshalling evidence in support of his theory of evolution, i.e., proof by the sheer weight and interconnection of the accumulated bits of evidence. Any discipline faced with data that is difficult to quantify and which is subject to distortion by a great many uncontrollable extraneous factors, and this is a characteristics of the behavioral, social, and some of the biological sciences, must resort to this method of orchestrating the empirical support for its hypotheses. Linguistics was among the first, if

---

1 The phonetic transcriptions used throughout this paper follow the principles of the International Phonetic Association as of 1979. Square brackets, '[...]', enclose detailed "phonetic" transcriptions, and slashes, '/.../' or the absence of any enclosing punctuation, a less detailed or "phonemic" transcription. Underlined words represent the standard orthographic representation and are assumed to be phonetically accurate only with regard to the point for which they are cited. The arrows '>' and '<' stand for 'changes to' and 'comes from', respectively. Citation of cognate forms from sister dialects lack arrows because, strictly speaking, one is not derived from the other but both are derived from a common parent, not cited. In this case, however, it is assumed that the form on the left reflects the phonetic form of the parent language.

Table 1 Cognates in English, Greek and Latin [8]

ENGLISH	LATIN	GREEK
father	pater	patēr
foot	pēs	pous
flat	planus	platos
fathom	patere	patanē
sprout	sperma	spora
spoon	sponda	spaθē
spurn	spernō	spairō
thin	tenuis	tanu-
three	trēs	treis
thumb	tumere	tumbos
thresh	terere	terēdōn
stand	stāre	statos
stiff	stips	stibos
star	stella	(a)ster
heave	capīo	kōpē
horn	cornū	keras
hound	canis	kuōn
head	caput	kapia
shine	scintilla	skia
shave	scabere	skap^{h}e

not the first to successfully apply this method in order to treat in a scientific way an aspect of human behavior, a topic once thought (and by some, still thought) to be incapable of scientific study [9].

The next step taken by historical linguists is the reconstruction of as many details about the parent language as possible. This cannot be done with the same precision as in the case of establishing the genetic relation between languages, however, a whole arsenal of heuristic methods has been developed to accomplish the task. When two or more of these methods point to the same reconstruction, as they often do, confidence in the result is increased. In the data given in Table 1, for example, the principle of "majority vote" can be used. In the cognate forms where the initial consonant varies between the three languages, the Latin and Greek forms have the same consonant. As it happens, most other Indo-European languages (and for the sake of the most accurate reconstruction one would include as many as possible) also show p, t, k in initial position (in these words) where the Germanic languages, including English, generally show f, [θ], h (or related fricatives), respectively. On the basis of the sounds exhibited by the majority of the Indo-European languages, the linguists reconstruct *p, *t, and *k (the asterisk signalling the hypothetical nature of the sounds) for these initial consonants in the parent language, called "Proto-Indo-European". Geography is also taken into account: the Indo-European languages which usually exhibit the initial voiceless stops in words such as those in Table 1 are very distant from each other geographically, from the Celtic languages spoken as far west as Ireland to Sanskrit, spoken in ancient India. Other things being equal, geographically dispersed languages, even those from the same genetic stock, have a greater opportunity to undergo change than otherwise, so that if similarities are nevertheless found in them they probably represent a direct inheritance from the parent language. Comparable arguments are used by geneticists in accounting for similar behavior traits in identical twins raised apart.

Occasionally, ancient texts are available to assist in the reconstruction. In some of these cases the texts have been discovered after a reconstruction was ventured and they have, in most cases, provided dramatic support for the historical linguists' hypotheses.

The result of all this scholarship in historical linguistics, which has been going on for approximately the past 170 years, is a rich collection of data on the history of numerous language families. What is important for our purposes is that this data includes detailed information on the changes in pronunciation that languages are subject to over time, i.e., "sound changes".

2. SOUND CHANGE

In order to make use of the accumulated knowledge of sound change it is essential to know why it happens. Since language is a complex form of human behavior that has a physical phonetic aspect, a psychological aspect, and a social aspect, forces in all three domains could cause sound changes. Our primary interest here, however, is the phonetic domain. For the sake of completeness it may be useful to consider a couple of cases of change whose causes lie in other domains.

2.1. Non-Phonetic and Phonetic Sound Change

The word sound itself provides a good example of sound change due to non-phonetic factors. The word derives from Middle English soun which was a borrowing from Old French son (where, unlike Modern French, the final n was pronounced) which in turn was derived from Latin sonus. Where did the final -d come from? It was added by speakers who mistakenly thought the original d-less pronunciation was a substandard corruption of a form with a final -d. This was because other words which did possess a final -d, e.g., wound, hound, often had this final consonant deleted by "substandard" speakers.

Another example is the most common form of the plural inflection for nouns in English, s. Old English had several plural inflections, among them -an and -as (now -en and -(e)s). The irregular plural oxen and a few others are all that remain of the former. The ancestor of the word for name, i.e., nama used to take this form of the plural, too: naman. Gradually, however, speakers put more and more words into the class that took the -as inflection so that now the plural of this word is names [10].

Obviously, social and psychological influences of this sort vary from one culture to the next, from one time in history to the next, and from one speaker to the next. So sound changes of the sort just illustrated, /n#/ --> /nd#/ (where '#' stands for 'word boundary') or /ən/ --> /z/ would not be expected to crop up in different languages except by chance. In contrast, sound changes which are due to physical phonetic factors should be found in the history of many different unrelated languages since phonetic forces are universal and timeless. This, in fact, is what happens, as will be reviewed below. These common, cross-language sound changes, I propose, should be the focus of attention of the linguist interested in modern applications in man-machine communication. A proper reading of the vast record of sound changes accumulated by the historical linguists of the past two centuries provides us with hints and insights into the workings of speech production and perception. Sound change is, in a sense, a failure of the speech transmission process. But as with destructive testing in materials science or the creation of artificial mutations in the field of genetics, one of the best ways to attempt to understand the workings of a complex system is by seeing how it behaves when it breaks. The knowledge we gain in this way about the mechanics of speech, I maintain, will enable us to implement improved systems for speech synthesis and automatic speech recognition.

Two warnings are in order at this point. The first is that many linguists would not agree with my interpretation of sound change as being due to mechanical constraints of the speech production and perception systems. Rather, they view it as primarily a mentalistic phenomenon, i.e., motivated by a desire (subconscious, perhaps) by speakers to improve the phonetic or phonological shape of their language [11,12,13]. I have argued my own position in several papers and will not repeat those arguments here [14,15,16,17,18,19,20, 21,22,23,24]. I will, however, offer some evidence which should at least render my position plausible.

The second warning is that unfortunately the record of sound change is not easily accessible since so far historical linguists do not report their results in a standardized way nor is there a compendium of the common cross-language sound changes. For the present, then, this information can only be applied to problems in man-machine communication through close collaboration between linguists and communication engineers.

2.2. A View of Sound Change

To illustrate my view of sound change (where I use this term henceforth to refer to the common cross-language phonetically-caused sound changes) I find it useful to give the analogy of a transmission line, say that used in telephony. See Figure 1. A telephone transmission line begins with a message source (the caller's voice) which is then input to a transmitter. The transmitter is also a kind of transducer which converts the signal into a different form so that it may reach the receiver. The receiver is also a transducer which converts the signal into a form suitable for input to the message source, in this case the ear of the person at the other end of the line. Inevitably, somewhere between the message source and message destination, the signal gets distorted by noise. The noise may come from ambient sources or, equally, from the transmitter or receiver. In order to minimize the distortions from noise and to reconstruct the original message, communication engineers have developed various error-correcting or error-limiting codes. Even so, some noise and its consequent distortions to the signal do reach the message destination.

The situation in speech is quite similar. For our purposes, the pronunciation conceived of in the brain of the speaker is the message source, the vocal tract, including the neuromuscular apparatus that controls it, is the transmitter, the listener's auditory apparatus and the neurological structures behind it are the receiver, and the listener's brain is the message destination. In the case of the speech signal, too, there may be noise added from ambient sources, e.g., wind, birds' cries, machines. But the signal may also be distorted by constraints of the vocal apparatus itself and by the auditory system. There is a form of error-correction or error-limitation in speech--accomplished as most such schemes are, by redundancy--but it is not perfect and is more effective at allowing the listener to extract the meaning of the transmitted signal, i.e., its semantics, than its pronunciation, i.e., what the listener needs to know in order to pronounce what he has just heard. If the listener-turned-speaker pronounces a word in a way that is different from that used by the original speaker then a sound change will have occurred. If other speakers copy this new pronunciation from him and alter other words in a like way, then this new pronunciation may spread and become typical of a whole community. It is usually at this stage that it comes to the attention of the linguist.

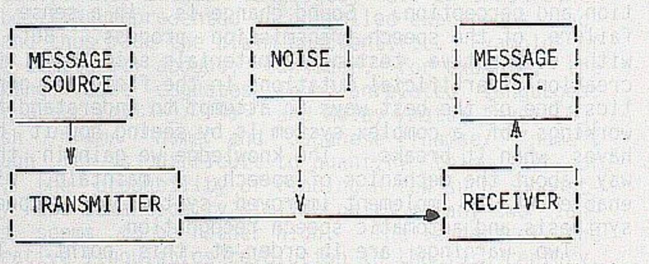

Figure 1 Schematic representation of transmission line used as analogy for speech communication.

3. APPLICATIONS

3.1. Confusion of Acoustically Similar Sounds

If speech were designed by communication engineers there would no doubt have been some effort to make each "cipher" in the code structurally equidistant from all others so that none of the them would be more subject to confusion than any other. Of course, speech wasn't designed by engineers and one of the first things that a study of sound change can tell us is that some sounds or sequences of sounds are much more subject to confusion than others. Between vocal tract shape and sound output there is a many-to-one relation so that two or more articulations give rise to the same or similar acoustic signal. If listeners cannot figure out which articulation produced the pronunciation they heard they may substitute another when they speak. Thus, a sound change would occur.

One example of this is the change of palatalized labial consonants (including those palatalized by virtue of being following by palatal glides or vowels) to dental, alveolar, or palato-alveolar consonants. Table 2 provides illustrations of this from the history of several languages. The large number of examples also demonstrate the point made above, that such physically-motivated sound changes do re-occur in languages the world over. (Due to limitations of space I will provide only one or two examples of other types of sound changes discussed.)

The same interchange of consonants, in the same environment, shows up with greater-than-average incidence in laboratory-based speech perception studies [25]. Spectrograms published by Fant [26], shown schematically in Figure 2, show why this substitution occurs. The F2 transition, which are important for the identification of place of articulation of the consonant, are more similar between /b^ja/ and /da/ than between /ba/ and /b^ja/. Presumably other cues, e.g., the bursts in the case of stops, are sufficient to differentiate palatalized labials from dentals, etc., but if these cues are missed listeners can easily become confused as to the true primary articulation of the consonant.

There is, in fact, independent evidence that the burst of labial stops is less salient than those of stops at other places of articulation. Many languages with a series of voiceless stops lack a voiceless labial stop entirely (except perhaps in loanwords or in "expressve" vocabulary). In addition to Arabic, the following are a few of the languages that lack a /p/:

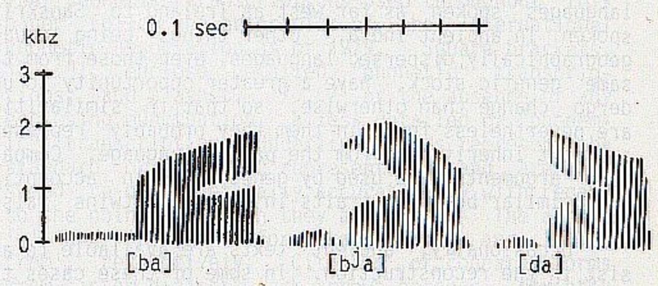

Figure 2 Spectrographic patterns of [ba], [b^ja], and [da] as spoken by a Russian speaker. The latter two syllables are more similar than the first two.

Table 2 Phonological data illustrating sound changes of palatalized labials to dentals, alveolars, and palato-alveolars (underlining marks segments showing the relevant variation) [16]

Pre-Classical Greek		Classical Greek	Translation
*gʷam-jo	>	baínɔ	"I come"
*kom-jo-	>	koinós	"common"

Proto-Bantu		Zulu	Venda	Translation
*pja	>	-tʃa		"new"
*piu	>		tsʷʰu	"knife"

Standard Czech	East Bohemian	Translation
mʲɛstɔ	nɛstɔ	"town"
pʲɛt	tɛt	"five"
pʲi:vɔ	ti:vɔ	"beer"
pʲɛknʲɛ	tɛknʲɛ	"nicely"

Lungchow	T'ien-chow	Translation
pjaa	tʃaa	"fish"
pjau	tʃuu	"empty"

Old Tibetan		Lhasa Tibetan	Translation
bja	>	tʃa	"bird"
bji-ba	>	tʃiwa	"rat"

Roman Italian	Genoese	Translation
pjeno	tʃena	"full"
pjanta	tʃanta	"to plant"
bjaŋko	dʒaɲku	"white"

Latin		French	Translation
sapius	>	saʒ	"wise"
rubeus	>	ʁuʒ	"red"
rabies	>	ʁaʒ	"rabid"

Yoruba, Bora, Vietnamese [27]. In its "regular" vocabulary, Japanese has /p/ only in restricted environments: never in initial position and medially only as a geminate, e.g., /nip:on/ "Japan". In most of these cases an original /p/ has become a weak fricative (either /f/ or /h/) through sound change. The cause of this pattern is the fact that labial stops, unlike the other stops, have no resonating cavity in front of them and thus are relatively low in intensity and very broad-band in their spectral characteristics. They are easily missed by the listener or, if heard, mistaken for a low-intensity fricative.

Another common sound change is the substitution of labials for back velars, especially labialized velars. Some examples are given in Table 3.

Again, these same substitutions are evident in the results of laboratory-based speech perception studies [25] and can be explained by reference to the misleading character of the F2 transitions.

Table 3 Phonological data illustrating sound changes of labialized velars to labials (underlining marks segments showing the relevant variation) [15]

Proto-Indo-European		Cl. Greek	Translation
*ekwōs	>	hippos	"horse"
*gwiwos	>	bios	"life"
Proto-Bantu		**West Teke**	
*-kumu	>	pfumu	"chief"

A quite striking example of sound change due to the auditory similarity of different articulations involves the perception of pharyngealization in Arabic. Among the primary acoustic correlates of pharyngealization (مفخّمة) is a generalized lowering of the formants. As it happens, the same effect can be achieved by rounding and protruding the lips. Because of this Arabic words are occasionally borrowed by other languages with the original pharyngealization replaced by labialization. See Table 4 [28,29].

Many other examples could be given of sound changes motivated by the acoustic-auditory similarity of certain speech sounds [14,21]. How can such data help us in designing better systems for man-machine communication via speech? First, we are pre-warned of the existence of speech sounds that will be difficult to differentiate in ASR systems. In fact, I have used such phonological data successfully in the design of a "difficult" and an "easy" vocabulary for an experiment investigating how users interact with an ASR device [30]. The ASR device had as much trouble in differentiating [bud] and [gud] in the "difficult" vocabulary as we know humans do based on the evidence from sound change. The phonological data, therefore, can aid us in the design of "robust" vocabularies in those ASR situations where a small vocabulary is sufficient. If the vocabulary cannot be limited to easily differentiated words, then the phonological data can tell us which contrasts may need special recognition algorthims to tell apart.

In the case of speech synthesis the same data can tell us which acoustic parameters to reinforce or even to exaggerate in order to guarantee greater intelligibility.

Perhaps the most important lesson to be derived from this phonological evidence is the need to differentially weight and to integrate acoustic cues. We have seen that an F2 transition does not have the same value in all situations. What it tells us depends on what other cues it appears with. The integration of

Table 4 Examples of Arabic pharyngealized sounds being borrowed as labialized consonants (contrasted with plain sounds being borrowed unchanged)

Standard Arabic		Argobba	Translation
ṣabijj (< صبيّ)	>	sʷabijj	"baby"
ṣadaqa (< صدقة)	>	sʷadaqa	"alms, charity; death commemor."
BUT:			
risala (< رسالة)	>	risala	"letter"
dʒism (< جسم)	>	dʒism	"body"

the various pieces of information in the acoustic speech signal remains as a major problem in speech recognition work.

3.2 The Units of Speech

Ever since the invention of the alphabet there has been an assumption by language scholars that the ultimate unit of speech is something like the phoneme. Technically, there is a difference between phonemes and the letters used in orthography, but to a large extent they are the same conceptual entity. Certainly it is true that the phoneme is indispensible for the kind of things a grammarian deals with, that is, stating grammatically-based phonological regularities, many of the kind of pronunciation differences that exist when words are strung together in connected speech, dialect differences, etc. It would be extremely awkward to have to state in other than phoneme-sized units the phonological rule of English whereby final /n/ in some words assimilates to the place of articulation of the consonant beginning the next word, e.g., /wʌn tʰɑim/ "one time", /wʌm bɔi/ "one boy", /wʌŋ gɑi/ "one guy". It may also be the case that native speakers are capable of analyzing speech into segments of phoneme size for certain purposes, including the above-mentioned poetic devices of alliteration and assonance. (Also, see below, section 4.) Nevertheless, certain facts would suggest we be cautious before jumping to the conclusion that the phoneme is the appropriate unit for all speech-related tasks.

First, an examination of the independently-created orthographies of the world reveals that the alphabet, as such, was invented only once--by Semitic scribes some 3800 years ago in what is now Syria--and all other alphabets can be traced to that one influence. All other orthographies were syllabaries or, perhaps, systems of ideograms or logograms [31] It would be unwise to attach too much weight to this type of historical evidence but it at least undercuts the notion that there is something "intuitively correct" about analyzing speech in terms of phoneme-sized units. The evidence of sound change--which, I have tried to argue, reflects how speech is actually used--gives evidence against the phoneme. Most sound changes are what the linguist calls "conditioned" changes, i.e., they occur in specific phonological environments. Examples were given above. In Table 1 the Proto-Indo-European *p, *t, *k changed to voiceless fricatives in the Germanic languages primarily only in word initial position. The same sounds after *s did not change. Moreover, in medial position, these sounds often became voiced, not voiceless, fricatives, e.g., Latin pater, English father [fɑðɚ] not [fɑθɚ]. The data in Tables 2 and 3 likewise show contextually-conditioned changes: the labial consonants change to dentals, etc. only when followed by palatal segments; when followed by non-palatal segments this change did not occur. If the phoneme were the crucial unit in sound change (and thus language use) they would have undergone the same change in any and all environments they appear in, that is, in the linguists' terms, they would be "unconditioned changes". Although unconditioned changes do occur, they are not the "norm".

From this I think it is safe to conclude that units larger than the phoneme--larger in size and larger in number--can play an important part in the actual production and perception of speech and therefore also in man-made systems simulating these processes. In fact, so-called "diphones" or "demisyllables" are now receiving a great deal of attention in speech synthesis and speech recognition systems [32].

3.3. Segmenting Speech

The fact that many sound changes are conditioned by the environments provided by word-initial, word-medial, or word-final position has some relevance for the discovery of the inherent error-correcting or error-limiting mechanisms in speech. One of the ways errors are limited in digital communications is by breaking the entire message or data set into "clumps". On discs, tapes, and other storage media these are referred to as "sectors". "Checksums" or other error-checking measures are done in the domain of these sectors. If an error is detected it is localized to a single sector and does not affect the rest of the data stream. The fact, referred to and illustrated above, that sounds undergo different changes depending on where they are in the word, seems to show that word boundaries may be marked in speech, although this marking may be quite subtle (as evidenced by the difficulty speech scientists have in discovering them) and is certainly done in language-specific ways. Marking word boundaries has the same advantage to the listener as dividing the data stream up into sectors has in digital communications. If an error in transmission does occur, the error may be localized to one word, thus giving the listener the chance to correct it by use of redundancies at the pragmatic, semantic, or syntactic level.

3.4. Redundant Cues

Our examination of sound change can also give us some understanding of the status of redundant cues.

Over a century ago a linguist studying the history of the Chinese language noticed a probable correlation between the voicing of the syllable initial consonants in ancient words and the character of the tone in those syllables in the modern spoken language [33]. Similar patterns have been noticed in the history of other tone languages, most recently Northern Kammu, spoken in Southeast Asia, data from which is presented in Table 5. Here the ?Uu (Southern) dialect may be assumed to retain most closely the pronunciation of these two dialects' common parent language and that Yuan is the (Northern) dialect that has innovated. As can be seen, in the innovating dialect, a high tone (distinctively high fundamental frequency, F0) is present on the vowel following what was formerly (and still is in ?Uu) a voiceless stop, whereas a low tone sits on the vowel that followed a formerly voiced stop. The Yuan dialect has gained a tonal distinction on vowels but has lost a voicing distinction on stops. Where did the tone come from?

Modern phonetic studies on non-tone languages such as English, French, Russian, show that voiced and voiceless stops induce small but perceptually detectable perturbations in the F0 of the vowels following them: slightly higher F0 following voiceless stops and slightly lower F0 following voiced stops (vis-a-vis sonorants like /m, n, r, l/ which do not perturb F0 to any appreciable extent) [24]. Morever, there is evidence that listeners in these non-tone languages can make use of these F0 differences to classify stimuli as, for example, /ke/ or /ge,/ when all other differentiating cues have been removed [35]. Does this mean that we should revise our characterization of English,

Table 5 Tone Development in Northern Kammu

?Uu (Southern Kammu)	Yuan (Northern Kammu)	Translation
klɑɑŋ	klɑ́ɑŋ	"eagle"
glɑɑŋ	klɑ̀ɑŋ	"stone"

French, etc. as non-tone languages? No, because although listeners apparently can utilize F0 differences to differentiate words, the best evidence we have to date is that speakers, when articulating syllables such as /ke/ and /ge/ do not purposely or actively tense the muscles usually used to create F0 variations. Rather, the F0 perturbations are just accidental by-products created by the articulatory gestures needed to make the voiced/voiceless distinction. Listeners in non-tone languages, when they speak these syllables in their turn, treat the F0 variations as distortions of the intended signal and therefore factor them out. The basis for listeners' ability to factor out such distortions is their long experience in producing and perceiving speech and thus learning *inductively* of the predictable correlation between certain speech sounds and their associated distortions on other adjacent speech sounds.

If, for some reason, the listener fails to make the connection between a speech sound and the distortion it causes--for example, if he is a child learning the language for the first time, or if he fails to detect the feature or segment responsible for the distortion--then he will not factor out the distortion but will rather incorporate it into his own pronunciation. This must be what happened at some point in the history of Chinese and Kammu but which has not happened in English, French, etc. (In most cases where consonants have given rise to new tonal distinctions, some tonal distinctions were already present in the language or were used by closely neighboring languages. Conceivably, the listeners have to be "attuned" to tonal distinctions before new tones can be created out of what were previously F0 perturbations by consonants. This may be why this type of sound change is not found in English, French, etc.)

A great many sound changes have happened via the same mechanism, i.e., where a phonetically natural and predictable distortion of the speech signal becomes distinctive, often at the expense of the loss of the element of speech which created the distortion. The development of distinctive nasalization on vowels in French is a well know example. Early French /bon/ > Mod. French /bɔ̃/.

Another important change of this type involves shifts in the quality of vowels due to distortions of nearby segments. Table 6. gives some relevant data. Here Written Burmese is take to reflect the phonetic state of the parent language of Lisu (a member of the Tibeto-Burman language family), at least as far as the vowel + stop ending is concerned.

The original low vowel [ɑ] changed to mid back rounded [o] when it appeared before a labial consonant but to a mid front unrounded vowel [e] when it appeared before a dental consonant. A quite similar change occurred in Northern Lebanese and Northern Syrian dialects of Arabic [37].

Again, modern phonetic studies have shown that vowels are distorted in systematic ways depending on the phonetic quality of adjacent consonants [38,39]. Some of these results are represented graphically in Figure 3, which shows the familiar vowel space where the auditory dimension of frontness correlates with the

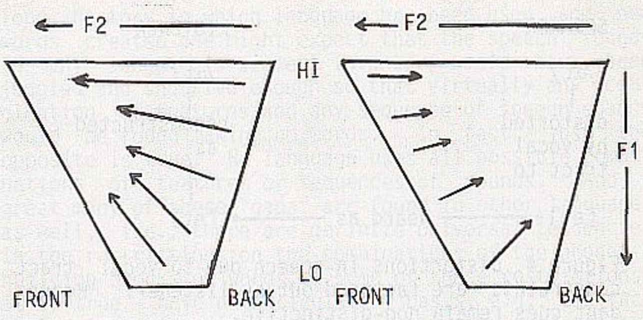

Figure 3 The direction of the perturbations of vowel quality caused by flanking consonants. On the left, the effect of dentals, alveolars, etc.; on the right, the effect of labials.

second formant (F2) and vowel height correlates inversely with F1. Labial consonants, especially labial-velars like [w], shift vowels towards the back, and dental, alveolar and palatal consonants shift vowels toward the front. The different length of the arrows in the two vowel diagrams are intended to represent the fact that labials distort vowels less than those which cause the fronting.

Vowels also assimilate to vowels and glides not immediately juxtaposed to them but in an adjacent syllable. The process called umlaut, so prevalent in the Germanic languages, exemplifies this. Some relevant data from the history of English is given in Table 7. Pre-historic Old English back vowels became fronted if the front vowel /i/ or the front glide /j/ was in the next syllable.

Languages which have "vowel harmony", e.g., Turkish, Mongolian, Hungarian, Yoruba, Somali, i.e., where all vowels in a word tend to agree in certain phonetic properties such as position of the tongue body, position of the lips, or position of the tongue root, acquired this feature through a process similar to that responsible for umlaut [40].

A simple way to formulate a generalization which covers all of these tendencies is: the second formant of vowels *tends to assimilate to that of adjacent segments.*

The examples just given, of tonal development, nasalization of vowels, change in vowel quality, are just a few of the many types of conditioned sound changes in which a previously redundant phonetic feature *becomes distinctive. As a way of summarizing* this, it may be useful to represent the perceptual process whereby the redundant features are either recognized as redundant or changed to be distinctive in the following schematic way. Figure 4 represents the case where a speaker intends to utter /ɑt/ but due to *vocal tract constraints* the vowel is phonetically more like [e]. This utterance is detected by the listener as [et] but is "reconstructed" as the correct /ɑt/ by reference to the knowledge base that incorporates the information that a vowel like /ɑ/ may be fronted by a flanking dental consonant.

Table 6 Sound Change Data in Tibeto-Burman [36]

Written Burmese	Lisu	English translation
kʰɑp	kʰò	"draw water"
hnjɑp	njo	"pinch"
sɑt	ʃè	"kill"
tshɑt	tshɛ	"deer"

Table 7 Umlaut in English

Prehistoric Old English		Middle English	Translation
*fo:d-jan	>	fe:d	"feed"
Compare: *fo:dɑ	>	fo:d	"food"
*mu:si	>	mi:s	"mice"
Compare: * mu:s	>	mu:s	"mouse"

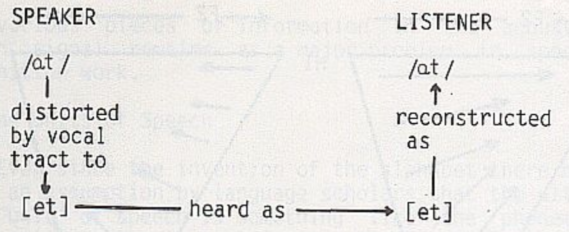

Figure 4 Distortions in speech due to vocal tract constraints are factored out by listener. Redundant cues remain non-distinctive.

Figure 5 represents the case which is similar to that in Figure 4, i.e., the intended /ɑt/ is realized phonetically as [et], but this time the "reconstruction" by the listener fails to occur (for any of a variety of reasons), so the listener interprets the pronunciation as /et/ and thus a sound change occurs. Although the original speaker, in his most careful articulations could show little or no fronting of the vowel /ɑ/ near dental consonants, the listener in Figure 5, even when giving careful pronunciations, will produce the vowel /e/ in these words--no longer due to the influence of the dental consonants but because that is the lexically-specified pronunciation for these syllables [18].

3.5. Further Evidence for Error-Correction

The existence of error-correction in speech perception has obvious relevance and importance for attempts at the mechanization of speech perception. Therefore the more evidence we have of its existence and its form, the better. One of the most dramatic pieces of phonological evidence of this sort consists of cases where these reconstructive, error-correcting rules are used inappropriately. (As mentioned above, this is in accord with the strategy of learning about the workings of a machine by seeing what it does when it breaks down.) Imagine, first, a situation, represented in Figure 6, similar to those in Figures 4 and 5 but where a speaker intends to utter /et/ and it is realized phonetically as [et]. The listener now behaves like the listener in Figure 4 and, thinking the [e] is a distortion caused by the [t], reconstructs the utterance /ɑt/. When this listener speaks this syllable, he will utter, in his most careful pronunciations, /ɑt/ = [ɑt].

Figure 6 does not represent a hypothetical occurrence. Virtually the same sound change did happen in the history of the Slavic languages: the low front vowel /a/ was changed to a back vowel /ɑ/ only when it occurred in the environment of palatal consonants, precisely the phonetic environment that would normally cause a fronting, not a backing, of vowel quality. (See Reference 19 for details.)

This type of sound change, where the conditioned change goes in the opposite direction to that expected

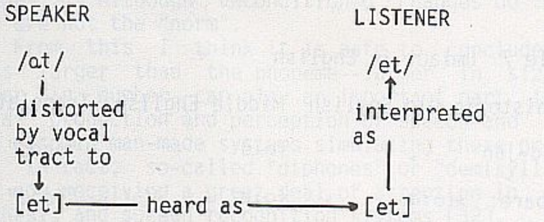

Figure 5 Distortions of speech signal by vocal tract not factored out by listener. Previously redundant features become distinctive.

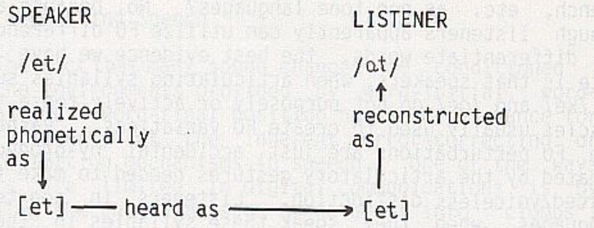

Figure 6 Listener creates sound change by inappropriately applying error-correcting rules.

by normal assimilative processes, is known as dissimilation. Although not as common as assimilation, there are still abundant examples of it.

One form of dissimilation that puzzled linguists for a long time is so-called "dissimilation at a distance", some of examples of which are given in Table 8.

In all the cases in Table 8, the co-occurrence of the same phonetic feature within the same word (usually in adjacent syllables or at opposite ends of the same syllable) causes the occurrence of this feature on one of the two segments to be removed. For example, in the last item in the table, the feature of [palatal] occurs on the initial ʃ and the medial geminate ʒʒ . In the dialectal form, that feature was removed from the initial sound, thus yielding s. What seemed puzzling was how these two sounds could affect each other when they left no trace on the intervening vowel. "Action at a distance" is as much abhorred by linguists as it is by physicists. The mystery is resolved, however, by modern phonetic studies which, as has been alluded to above, show that certain phonetic features though distinctive on a given segment, may "spill over" (via assimilation) onto adjacent segments. We may assume, then, that all of the features dissimilated in the above examples, e.g., aspiration, labialization, glottalization, and palatalization, also "colored" the intervening segments, whether vowel, glide, or other sonorant. The listener detected the feature of palatalization throughout the first syllable of

Table 8 Examples of dissimilation at a distance

Pre-Greek		Classical Greek	Translation
*tʰrikʰos	>	trikʰos	"hair"
Latin		Italian	
kʷīnkʷe	>	tʃiŋkʷe	"five"
Ancient Chinese		Cantonese	
*pjam	>	pin	"diminish"
Proto-Quechua		Quechua	
* t'ant'a	>	t'anta	"bread"
Standard Arabic		Dialectal Arabic [37]	
?ib?i («إبقي»)	>	hib?i	"remain"
?a?ʕud («اقعد»)	>	haʕud	"shall I sit down?"
ʃaʒar («شجر»)	>	saʒar	"trees"
ʃaʒʒaʕ («شجّع»)	>	ʃaʒʒaʕ	"be encouraged"

but erroneously attributed all of it--even that emanating from the initial consonant--as a distortion caused by the medial geminate. He therefore factored out all palatalization except that due to the medial sound.

The perceptual error that is committed in cases of dissimilation has an analogue in the domain of visual perception, namely, camouflage. A viewer may be looking directly at a white rabbit on a brightly lit snowdrift and yet not "see" him because he attributes the patch of white he is looking at to the background white of the snow. The viewer, in effect, is using his knowledge of what snowdrifts look like to interpret--erroneously, in this case--the patch of white which characterizes the rabbit. Analogously, in sound change the feature which is dissimilated, i.e., removed, is camouflaged by the presence, elsewhere in the word, of a source of the same feature.

I should admit, at this point, that the analysis I have just given of dissimilation is my own and deviates from other linguists' previous attempts at explaining it. There is, however, some strong evidence which favors my account over others. For example, from what I have said above it follows that the only phonetic features which could be dissimilated at a distance, i.e., across intervening segments, are those which are known to spread out beyond the immediate domain of the segments on which they are distinctive. Such features are: aspiration, glottalization, palatalization, labialization, retroflexion, pharyngealization, etc. All of these have been involved in well-documented cases of dissimilation. In contrast, features such as stop, fricative, and affricate should not be dissimilated at a distance since they are not "spreadable". In fact, these features are rarely involved in dissimilation [19]. Second, recall that it is often the case that contextually-conditioned sound changes of an assimilative character will lose the conditioning context at the same time as the conditioned change occurs, e.g., Pre-French /bon/ > Modern French /bɔ̃/. That is, the nasal consonant which created the nasalized vowel dropped at the same time as the vowel became distinctively nasalized. In dissimilation, however, it must not be the case that the conditioning context disappear at the same time the conditioned change occurs. This is because the conditioning context--the other segment with the same phonetic feature--must still be around to be "blamed" by the listener for what he imagines is the spill-over of that feature onto surrounding segments. So far, all the cases of dissimilation that I know of conform to this prediction.

Finally, there are laboratory-based studies which supports my account of dissimilation. One of these involved obtaining listeners' judgements of the degree of nasalization present in stimuli of the sort /mɪm/ and /mɑm/ when the nasal consonants flanking the vowels had been attenuated in amplitude by various degrees or removed entirely. In the processing of these stimuli the vowels themselves, which were, of course, heavily nasalized, were not changed in any way. Nevertheless, the listeners gave lower ratings for perceived nasalization on the vowels when the nasal consonants had their full amplitude than when they were attenuated or removed. In other words, when the nasal consonants were around to be blamed for the nasalization on the vowels, the vowels themselves seemed less nasal; when the nasals were removed, they couldn't be blamed and so the nasalization on the vowel was heard more clearly [41]. Other studies involving different phonetic entities also show that listeners "take into account" the kind of effect a segment X has on an adjacent segment Y before making a judgement about the phonological nature of Y [42,43].

Another important manifestation of what is in essence a dissimilatory effect can be found in constraints on the sequencing of speech sounds and on the co-occurrence of certain phonetic features. Over the long history in which language has been used and new words created one might expect that the speech sounds in any language's segment inventory would have been juggled and shuffled enough so that virtually any combination of features and any sequence of speech sounds would be found making up words. In fact, just the opposite is true. No language uses all possible combinations of features or sequences of sounds. And a great many of these "gaps" are found in other languages as well, i.e., there are definite universal tendencies in the restrictions on the combinations of the phonetic building blocks of words. Some of these restrictions, it is true, have an articulatory basis, e.g., the fact that obstruent clusters tend not to mix voiced and voiceless segments is due to aerodynamic and other physiological constraints [22,44]. What I wish to focus on here are restrictions that have a perceptual basis.

One example is the fact that distinctive pharyngealization in Arabic does not sit well on labials. Although "emphatic" labials are attested in some dialects of Arabic, some of these attestations are disputed and many dialects definitely do not have any. The reason, presumably, is that the acoustic-auditory characteristics of pharyngealization and those for labial place of articulation are too similar. That they are similar has already been evidenced by the fact--mentioned above--that certain non-Arabic communities when borrowing Arabic words, change pharyngealization into labialization (see Table 4). On the other hand the presence of distinctive pharyngealization on dental consonants is very common among Arabic dialects. The reason is that dentals have acoustic-auditory features which are just the opposite of pharyngealization, namely, relatively high F2, such that the addition of pharyngealization on these sounds creates a clearly detectable auditory difference. Speech--in fact, all forms of communication from semaphore to Morse code--consist of modulations of some carrier medium. The bigger the modulations, the better and the more easily detected the signal.

For similar reasons, labialized labials, i.e., [pʷ], and palatalized palatals, i.e., [ʃʲ], though found in some languages, are rare in comparison to labialized velars, or palatalized velars or labials.

The same principles are responsible for restrictions on the sequences of sounds, the so-called "phonotactics" of languages. For example, in English, although many obstruents form initial clusters with following /w/, e.g., those at the beginning of twin, dwarf, swear, quick, labial consonants do not (except in the case of a few loanwords). The /w/'s acoustic cues, lowered F1, F2, and F3, are too similar to the acoustic cues for labial obstruents. On the other hand, initial two consonant clusters with /j/ (the palatal glide) as the second component tend to favor labials and velars, but not alveolars or palatals, as the first component, e.g., beauty, puce, few, music, cute ([bjuri pʰjus fju mjuzɪk kʰjut]). Sequences of [sj-] long ago changed to [ʃ], e.g., sure is now [ʃuɹ] and many American English dialects have changed [tju], etc. to [tu] and so on. The cue for palatal glide, high F2, is evidently too similar to the cue for alveolar and palatal place of articulation; the one cue camouflages the other.

English also prohibits the lateral /l/ as second member of initial clusters having alveolar stops as the first member, that is, dlove, tlack are not possible English words. The lateral creates a clearer acoustic modulation when it follows labials and velars, e.g, bleed, plead, flee, clear, glib.

These last few examples have illustrated that labials and velars form a natural class as opposed to the consonants articulated between those regions, i.e., dentals, alveolars, and palatals. This, of course, is a division that has been traditional in Arab linguistics for centuries: the so-called "sun letters" (حروف الشمس), which include most dentals and palatals and the "moon letters" (حروف القمر), which include everything else. As with English, this division is motivated by the way they interact with the consonant /l/, especially in the definite article /al/. When joined to sun letters this /l/ disappears (dissimilates) and causes gemination of the following consonant, otherwise it remains as /l/.

3.6. Implications for Speech Technology

The preceding evidence from sound change and, in the case of phonotactics, the results of sound change, have important implications for the mechanization of speech perception. If we accept that human listeners make use of an extensive and detailed knowledge base on how speech sounds interact, i.e., on how one sound influences and distorts another, then it follows that if we hope to build machines capable of duplicating the human's ability in this domain, we should also construct and make use of the same knowledge base. Proposals to use such a "top-down" approach in automatic speech recognition are not new, but most previous attempts have emphasized a knowledge base involving pragmatic, semantic, and syntactic information. These are definitely needed, but what I am maintaining here is that a knowledge base incorporating some very fine-grained details about speech sounds themselves is also necessary.

Moreover, a list of the common cross-language sound changes created by perceptual mistakes can give us a good idea of the kind of rules that must go into this knowledge base. Phonetic studies have uncovered hundreds of systematic interactions--I call them 'distortions'--that exist between certain classes of speech sounds. The way that consonants perturb the F0 of following vowels has been mentioned above. There is also evidence that vowels themselves perturb the F0 they are uttered on, i.e., high vowels like /i, u/ have higher F0 than low vowels like /ɑ,ɔ/, other things being equal. Nevertheless, although the consonant-induced F0 variation may lead to a sound change, the vowel-induced F0 variation apparently never has [24]. Although the reasons for this asymmetry are not entirely clear yet, we may tentatively conclude that rules which take into account the former F0 perturbation would be useful but rules on the latter would not be.

Finally, the phonological evidence which suggests that "good" sound sequences are those that create a large acoustic modulation (in spectral shape, amplitude, periodicity), can aid us in the design of special purpose vocabularies for automatic speech recognition tasks. In such cases, we would do well to avoid such acoustically "flat" words as wool and try to use acoustically "contoured" words like swindle.

4. NEW GOALS, NEW METHODS IN LINGUISTICS

As mentioned earlier, although there had been interest in the differences and similarities between languages long before the 19th century, it wasn't until the early 1800's that a method was developed to actually deal with this issue in a rigorous way. Similarly, there has always been considerable interest in how a native speaker's knowledge of his language is represented, i.e., mentally, but rigorous methods to address this question have not been available until relatively recently. Even now, the methods developed are still rather primitive and much more work is needed to refine them. Nevertheless, this is an area where, in my opinion, some of the most promising new research is currently being done in phonology. The potential for applications of such research in areas of man-machine communication is quite high. If we believe that ultimately the best way to process speech--whether in synthesis or recognition--is to duplicate to the extent possible the way humans do these tasks, then we must find out how language and all its parts are represented in the minds of native speakers.

4.1. Accessing the Lexicon

As discussed above, every language has certain constraints on how speech sounds may be strung together to form words. It is not that the forbidden sequences are humanly impossible to pronounce, just that one language exhibits a systematic absence of them. Thus /mwa/ is not a well-formed English word but it is perfectly good French. It is a curious but quite commonplace fact that native speakers are able to tell a "native" form from a non-native form, even when it is a string of phonemes they have never heard before. Thus, given made-up forms like /glus/ and /gdɷb/, native speakers of English might easily believe that the first was a rare English word but they would not believe that of the second. How do they do this? Two hypotheses have been put forth to account for this behavior. One, from the early generative phonologists [45], maintains that native speakers formulate very general rules on permissible phoneme sequences. The English speaker might have a rule such as "In initial obstruent clusters, the first segment must be /s/." But they would not have a rule such as "In the context of an initial /kl/ cluster and a following /m/, the vowel /I/ may not appear." Only the former rule is a general one; the effect of the latter is just to rule out the non-existing /klIm/. The second hypothesis was offered by Greenberg and Jenkins [46] (G & J) and holds that speakers do a systematic comparison of the hypothetical form with the entries--potentially, all of them--in their mental lexicons. This comparison involves trying to change one or more of the phonemes in the made-up word to see if it can be transformed into an existing word. The more such substitutions succeed in yielding an existing word, the closer the form is to the native pattern. According to this speakers would be able to detect general as well as very particular, fine-grained anomalies. A crucial pair of "words" are the nonsense forms /klɛb/ and /klæb/. Presumably no one would claim that either one violates a general rule. All the partial sequences, /kl-/, /klɛ-/, /-ɛb/, /klæ-/, /-æb/ are attested in one or more English words, e.g., click, Clem, reb, clam, grab. But according to the comparison procedure of G & J, /klɛb/ is further from English than /klæb/. Thus the generative phonologists' prediction would be that English speakers would judge these forms as equal in "strangeness" whereas the G & J would predict that /klɛb/ would be judged as stranger than the other form.

These predictions (as well as similar ones about comparable pairs of made-up forms) were tested [47]. The results overwhelmingly supported the predictions of G & J. What this result implies is the speakers do not need to formulate general rules on how phonemes may be strung together to form words. All they need, to make the kind of judgements just discussed, is their lexicon-consisting of the words they know from having heard or read them--and a very general kind of comparison procedure. Furthermore, the need for these two is felt in ordinary speech perception as well. How is it that words are recognized in the stream of speech except by the listener finding a match in his lexicon for the particular string of sounds just heard?

In addition to having important implications for second language teaching--"don't teach rules, teach lots of vocabulary and lots of stock sentences and phrases"--these results also have implications for optimal ways to do speech recognition. It suggests we would have more payoff by refining the ways that comparisons are made between the acoustic signal and the stored vocabulary, than by formulating general "phonotactic" rules. Given the preliminary nature of these experiments, this must be only a tentative conclusion, however.

4.2. Categories of Speech Sounds

One of the most well-known "discoveries" of phonology--since approximately the turn of the century--is the phoneme. There are many semi-orthogonal claims that have been made about phonemes. One, as mentioned, is a claim about the approximate size these ultimate building blocks of speech. Another is that phonetically distinct sounds are psychologically the "same". For example, in English one finds voiceless aspirated stops in syllable intial position before stressed vowels ([kʰɪt]) but voicelss unaspirated stops after [s] in initial clusters ([skɪt]). These two sound never contrast since they occur in mutually exclusive environments. Because of this they are claimed to be allophones, i.e., contextually-caused variants of a single sound or phoneme. This is a quite reasonable view to take in cases where the variation can be accounted for as a mechanical effect--a distortion in the sense used earlier--of the context. Unfortunately, this does not seem to be the case here; there is nothing about syllable initial position that causes aspiration and nothing, so far as we know, about the position after [s] that causes de-aspiration. But strictly speaking the claim is not a phonetic one, it is psychological. A proper test of it, then, must also be psychological.

Psycholinguistic tests have been devised to test this and similar claims about how native speakers categorize speech sounds [48,49]. The method used is called concept formation (CF) and has been borrowed from experimental psychology where it has been used for many years on non-phonological questions.

Briefly, a CF experiment would proceed as follows. The subject (S), seated in a quiet room, is instructed as follows: "You'll hear a series of taped words, some of which belong to a certain category due to the way they sound; the rest do not belong to the category. If the word is in the category, respond 'yes', if not, respond 'no'. After you respond, we'll tell you what the right answer was. You'll have to guess on the first few words but eventually you should figure out how to anticipate the right answer. When this happens we'll give you a test. In this part we won't tell you what the right answer was after you respond." Then the experimental session might proceed as in Table 9, which should be read from left to right, top to bottom.

As can be seen, except for focussing Ss' attention on the pronunciation of the stimulus words, no other hints are given about the defining attributes of the target category. Ss must figure this out on their own inductively. It is possible in this way to "teach" linguistic concepts to linguistically naive Ss without using any verbal mediation, i.e., metavocabulary like "distinctive", "complementary distribution", etc.

As is evident from inspecting Table 9, the target category is words containing a [kʰ]. In the first part of the experiment, only clear, uncontroversial exemplars of category and non-category items are given along with feedback. In order to make sure Ss don't inadvertantly form some unwanted category using orthographic criteria (based on their own mental image of the spelled word), the category words have spellings which represent [kʰ] in diverse ways: ch, c, k, qu. Moreover, some of the non-category words may be spelled with those same letters representing different sounds as in, e.g., knife, chip, ceiling. Subjects would therefore find that orthographic cues are of no help in the task. When Ss get a certain pre-set number of trials correct in a row then the second part of the experiment is administered where feedback is withheld. Of interest is how Ss categorize words such as square

Table 9 Sample CF session; target category: [kʰ]

Trial No.	Stimulus (taped)	Sample response from S	Correct answer (taped)
1	keep	no	yes
2	fan	yes	no
3	cash	yes	yes
4	ghost	yes	no
5	lip	no	no
6	knife	no	no
7	choir	yes	yes
8	chip	no	no
9	ceiling	no	no
10	kerosene	yes	yes
11	occur	no	yes
12	cool	yes	yes
.			
.			
.			
75	gnat	no	no
76	science	no	--
77	kitchen	yes	--
78	square	yes	--
79	step	no	--
80	school	yes	--

([skʷæɹ]). If they respond 'yes' (putting this word in the same category with words starting with [kʰ]) then it would imply that [kʰ] and [k] are psychologically the same.

In fact, this is what American English Ss do, in conformity with the traditional phonemic analyses of English. The result may not surprise anyone, but it represents a validation of the experimental procedure. The CF technique has also been used to verify that English Ss regard affricates such as [tʃ] and [dʒ] as units not clusters. The method has the potential of being able to resolve most of the traditional but as yet untested claims about the psychological representation of speech.

5. CONCLUSION

There is currently a great deal of interest in developing ways whereby humans may communicate more easily with machines by the medium of speech, the natural form of communication of the former--if not the latter. To accomplish this task we must understand how speech works. We don't know this right now, so we should seek clues to its nature wherever we may find them. In this paper I have tried to show that the traditional and the newer findings of linguistics provide us with some valuable insights in this exciting domain.

REFERENCES

1. Borst, A., Der Turmbau von Babel, Stuttgart, 1957-63.
2. Robins, R. H., A Short History of Linguistics, Longmans, London, 1967.
3. van Helmont, F. M., Alphabeti vere naturalis Hebraici brevissima delineatio, Sulzbach, 1667.
4. Panconcelli-Calzia, G., 3000 Jahre Stimmforschung, N. G. Elwert Verlag, Marburg, 1961.
5. Jakobson, R. and Waugh, L., The Sound Shape of Language, Indiana Univ. Press, Bloomington, 1979.
6. Bloomfield, L., Language, Holt, Rinehart and Winston, New York, 1933.

7. Meillet, A., La méthode comparative en linguistique historique, H. Aschehoug, Oslo, 1925.
8. Cowan, W. and Rakusan, J., Source Book for Linguistics, Language Resource Center, Carleton Univ., Ottawa, 1980.
9. Lass, R., On Explaining Language Change, Cambridge Univ. Press, Cambridge, 1980.
10. Sweet, H., A New English Grammar, Part I. Clarendon Press, Oxford, 1891.
11. Halle, M., Phonology in Generative Grammar, Word, vol. 18, pp. 54-72, 1962.
12. Kiparsky, P., Linguistic Universals and Linguistic Change, in Universals in Linguistic Theory, ed. E. Bach and R. T. Harms, pp. 170-202, Holt, Rinehart and Winston, New York, 1968.
13. Martinet, A., Elements of General Linguistics, Univ. of Chicago Press, Chicago, 1964.
14. Ohala, J. J., Experimental Historical Phonology, in Historical Linguistics II: Theory and Description in Phonology, ed. J. M. Anderson and C. Jones, pp. 353-389, North Holland, Amsterdam, 1974.
15. Ohala, J. J. and Lorentz, J., The Story of [w]: an exercise in the phonetic explanation for sound patterns. Proc., Ann. Meeting, Berkeley Linguistic Soc., vol. 3, pp. 577-599, 1977.
16. Ohala, J. J., Southern Bantu vs. the world: the case of palatalization of labials, Proc., Ann. Meeting, Berkeley Linguistic Soc., vol. 4, pp. 370-386, 1978.
17. Ohala, J. J., The contribution of acoustic phonetics to phonology, in Frontiers of Speech Communication Research ed. B. Lindblom and S. Ohman, pp. 355-363, Academic Press, London, 1979.
18. Ohala, J. J., Articulatory Constraints on the Cognitive Representation of Speech, in The Cognitive Representation of Speech, ed. T. Myers, J. Laver, and J. Anderson, pp. 111-122, North Holland, 1981.
19. Ohala, J. J., The listener as a source of sound change, in Papers from the Parasession on Language and Behavior, ed. C. S. Masek, R. A. Hendrick, and M. F. Miller, pp. 178-203, Chicago Ling. Soc., Chicago, 1981.
20. Ohala, J. J., The Direction of Sound Change, in Abstracts of the Tenth International Congress of Phonetic Sciences, ed. A. Cohen and M. P. R. v.d. Broecke, pp. 253-258, Foris, Dordrecht, 1983.
21. Ohala, J. J., The Phonological End Justifies Any Means, in Preprints of the Plenary Sessions, 13th International Congress of Linguists, Tokyo, 1982, pp. 199-208, ICL Editorial Committee, Tokyo, 1982.
22. Ohala, J. J., The Origin of Sound Patterns in Vocal Tract Constraints, in The Production of Speech, ed. P. F. MacNeilage, pp. 189-216, Springer-Verlag, New York, 1983.
23. Ohala, J. J., Calibrated Vocabularies, in Proc., Workshop on Standardization for Speech i/o Technology, pp. 79-85, National Bureau of Standards, Gaithersburg, MD, 1983.
24. Hombert, J.-M., Ohala, J. J., and Ewan, W. G., Phonetic Explanations for the Development of Tones, Language, vol. 55, pp. 37-58, 1979.
25. Winitz, H., Scheib, M. E., and Reeds, J. A., Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech, J. Acous. Soc. Am., vol. 51, pp. 1309-1317, 1972.
26. Fant, G., Acoustic Theory of Speech Production, Mouton, The Hague, 1960.
27. Sherman, D., Stop and Fricative Systems: a Discussion of Paradigmatic Gaps and the Question of Language Sampling, Working Papers on Language Universals, Stanford, vol. 17, pp. 1-31, 1975.
28. Leslau, W., The Phonetic Treatment of the Arabic Loanwords in Ethiopic, Word, vol. 13, pp. 100-123.
29. Jakobson, R., Mufaxxama the 'Emphatic' Phonemes in Arabic, in Selected Writings I: Phonological Studies, Roman Jakobson, pp. 510-522, Mouton, The Hague, 1962.
30. Ohala, M., The Machine as an Addressee: When Paralinguistics Fails, in Abstracts of the 10th Int. Congr. of Phonetic Sciences, Utrecht, ed. A. Cohen and M. P. R. v. d. Broecke, p. 428, Foris, Dordrecht, 1983.
31. Diringer, D., Writing, Frederick A. Praeger, New York, 1962.
32. Olive, J. P., Rule Synthesis of Speech from Dyadic Units, Conf. Rec. of IEEE Int. Acous. Speech Signal Process., pp. 568-570, Hartford, 1977.
33. Edkins, J., A Grammar of the Chinese Colloquial Language, Presbyterian Mission Press, Shanghai, 1864.
34. Ohala, J. J., Physiological Mechanisms Underlying Tone and Intonation, in Preprints, Working Group on Intonation, 13th International Congress of Linguists, Tokyo, ed. Hiroya Fujisaki and Eva Garding, pp. 1-12, Tokyo, 1982.
35. Fujimura, O., Remarks on Stop Consonants--Synthesis Experiments and Acoustic Cues, in Form and Substance, ed. L. L. Hammerich, R. Jakobson, and E. Zwirner, pp. 221-232, Akademisk Forlag, Copenhagen, 1971.
36. Thurgood, G. and Javkin, H., An acoustic explanation of a sound change: *-ap to -o, *-at to -e, and *-ak to ae, J. of Phonetics, vol. 3, pp. 161-165, 1975.
37. Garbell, I., Remarks on the Historical Phonology of an Eastern Mediterranean Arabic Dialect, Word, vol. 14, pp. 303-337, 1958.
38. Lindblom, B., Spectrographic Study of Vowel Reduction, J. Acous. Soc. Am., vol. 35, pp. 1773-1781, 1963.
39. Stevens, K. N. and House, A. S., Perturbation of Vowel Articulations by Consonantal Context: An Acoustical Study, J. Speech and Hearing Res., vol. 6, pp. 111-128, 1963.
40. Ultan, R., Some Reflections on Vowel Harmony, Working Papers on Language Universals, Stanford, vol. 12, pp. 37-67, 1973.
41. Kawasaki, H., The perceived nasality of vowels with gradual attenuation of adjacent nasal consonants, to appear in Experimental Phonology, ed. J. Ohala, Academic Press, New York.
42. Ohala, J. J., Kawasaki, H., Riordan, C., and Caisse, M., The influence of consonantal environment upon the perception of vowel quality. Forthcoming.
43. Mann, V. A. and Repp, B. H., Influence of vocalic context on perception of the [s]-[ʃ] distinction. Perception and Psychophysics, vol. 28, pp. 213-228, 1980.
44. Kawasaki, H., An Acoustical Basis for Universal Constraints on Sound Sequences, Doc. Diss., Univ. of California, Berkeley, 1982.
45. Halle, M., The Sound Pattern of Russian, Mouton, The Hague, 1959.
46. Greenberg, J. H. and Jenkins, J. J., Studies in the Psychological Correlates of the Sound System of American English. I and II., Word, vol. 20, pp. 157-177, 1964.
47. Ohala, J. J. and Ohala, M., Testing Hypotheses Regarding the Psychological Manifestation of Morpheme Structure Constraints, Rep. of the Phonology Lab. [Berkeley], vol. 2, pp. 156-165, 1978.