THE MEASUREMENT OF PHONETIC SIMILARITY

Peter Ladefoged

University of California, Los Angeles

There are many reasons for wanting to measure the degree of phonetic similarity between members of a group of languages or dialects. The present study grew out of a research project which was designed to get data that might have a bearing on some of the practical problems which exist in Uganda. In the Southern part of Uganda, where two thirds of the nine million people live, there are numerous closely related Bantu languages or dialects. The official Ugandan census data lists 15 Bantu languages. The current study uses data on these and six others. We wanted to assess their phonetic similarity so that there would be data on which to base decisions on which languages to use for broadcasting (the government currently broadcasts in 8 or 9 of these languages, as well as in 10 non-Bantu languages), which to use in schools (3 are used officially and a further 5 unofficially, but with the connivance of the local education authorities), and which for other purposes.

One method of obtaining a measure might have been by devising a metric that could be applied to formal comparisons of phonological descriptions of each of these languages. This method was not attempted, largely because of time limitations. The data had to be collected and first analyses made within a period of one year. Furthermore, it soon appeared that the sound patterns of nearly all of these languages were very similar, and the phonological descriptions would have to be extremely detailed before systematic differences became apparent. Finally,

before we could quantify, in practical terms, the overall degree of phonetic similarity between a pair of languages, the phonological descriptions would have to be supported by counts of the frequency of occurrence of each rule. A difference between two languages due to, say, the addition of a rule in one but not the other would be more or less important depending on the number of times in which the rule was involved in ordinary utterances.

The technique which we chose to use instead was to measure the degree of phonetic similarity in a list of 30 common words in each language, all of which were historically cognate forms in at least 16 out of the 20 languages. The list was a subset of a list of 100 words which had been recorded so that lexico-statistical comparisons might be made. The complete lists had been recorded in a narrow phonetic transcription by the author, using IPA symbols except for the voiced and voiceless palatal affricates, which were transcribed j and c in accordance with the conventions of Ugandan orthographies. Long vowels and long consonants (both of which are phonemic in some of these languages) were transcribed with double letters. Tones were transcribed by acute accents (high), grave accents (low) and circumflex accents (falling); as far as is known these possibilities will account for nearly all the tonal contrasts that occur in these languages. Table 1 exemplifies the data for two words in each of the 20 languages.

The fundamental problem in making phonetic comparisons is how to line up two words, one in one dialect and one in another, in such a way that we can make a valid point by point comparison of all the things which affect phonetic similarity. In the Bantu languages with which we

were concerned, each noun consists of a stem, and a prefix indicating the noun class. Only the stems were used in these phonetic comparisons. In general, a stem begins with a consonant, C, followed by a vowel, V, and may contain additional alternations of consonants and vowels. The commonest form is CVCV. Some problems in lining up segments will be considered after we have considered how they may be compared.

There have been a number of attempts to devise measures of the degree of phonetic similarity of isolated segments. Some of these have been based on experimental studies showing, for instance, the degree of confusability of different segments (Miller and Nicely 1955, Peters 1963, Wickelgren 1965, 1966, Klatt 1968, Greenberg and Jenkins 1964, Mohr and Wang 1968); others have been based on more theoretical arguments (Austin 1957, Peterson and Harary 1961). All of these are of interest here, in that the knowledge of the degree of phonetic similarity between segments is a necessary prerequisite to a statement about the degree of phonetic similarity of languages as a whole.

Some of the studies cited above have discussed the possibility of quantifying the degree of difference between segments by counting the number of differences in their specifications in terms of features. Various ways of specifying segments in terms of features have been suggested, the most important being the early distinctive feature system of Jakobson, Fant, and Halle (1951), its revision by Jakobson and Halle (1956), and the system proposed by Chomsky and Halle (1968). All these features sets are intended for classifying the segments which occur in phonemic or phonological contrasts within a language. But it is by no

means obvious that the specification of the phonetic level in the way suggested by Chomsky and Halle, for instance, is directly related to the specification of the kind of phonetic similarity measure which is useful in cross language studies. Chomsky and Halle were certainly not trying to produce a phonetic specification of this kind. Accordingly for the purposes of the present study an ad hoc set of phonetic features was used.

For the sake of computational simplicity, the phonetic features were considered to be independent binary categories. This is obviously an invalid assumption which will be discussed further towards the end of this paper. Because vowels were being compared only with vowels, and consonants only with consonants, there was no need for features such as consonantal and vocalic; they would never have contributed anything to the cross language comparisons. Furthermore there was no need to use the same features for both consonants and vowels. The feature system which was set up was adequate for specifying all the phonetic differences which had been observed among Ugandan Bantu languages and seemed, on the basis of the experimental studies cited above, likely to be the best possible measure of segment similarity within the constraints previously noted.

Each consonant segment in a Ugandan Bantu language was described as being, or not being: (1) a stop; (2) a nasal; (3) a fricative; (4) anterior -- made in the front of the mouth; (5) alveolar -- made near the teeth ridge; (6) coronal -- made in the centre of the mouth; (7) voiced; (8) long; (9) followed by a w-glide; (10) followed by a y-glide. The easiest way of appreciating the way in which these terms were used is

through the examples showing the partial characterization of some consonants given in Tables 2 and 3. A plus sign indicates the presence of a feature, and a minus sign shows its absence.

The degree of similarity between segments is exemplified in Table 4. Thus b and d̪ have nine out of the ten points in common; and b and ʃʸ differ in seven points, and have only three points in common.

In one or two details this measure is not entirely satisfactory. There is no reason why b should be considered to have seven points in common with l and only six points in common with r ; and, what is more important, there is no reason why h should have such varying degrees of similarity with b , d̪ , d , d- . These anomalies occur because segments were specified in terms of independent binary categories. With a classification system of this kind it is impossible to give a specification of h which is *equally* different from all the stop consonants. But these inequities probably did not have a significant effect. Among the 2,400 segments compared, h occurred only 31 times.

In specifying the vowels we stated whether each one was, or was not: (1) high; (2) mid; (3) low; (4) front; (5) central; (6) back; (7) long; (8) high tone; (9) falling tone. At one time we added the possibility: (10) low tone. But preliminary results showed that this gave too much importance to tonal similarity, and it was better to consider low tone as simply the absence of high or falling tone. The degree of similarity in vowels was measured by counting the number of features they had in common, in the same way as for consonants.

Using this measure of the degree of phonetic similarity, the features in each segment were compared with the corresponding features in

the corresponding segment in each of 30 words in each of the 20 Bantu languages. The 144,000 comparisons involved, the sums indicating the degree of phonetic similarity of each pair of languages, and the tabulations were all done on a computer.

A number of problems arose in the comparison of specific segments, two of which will be considered here. Both are due to the constraint of having to compare words segment by segment, a constraint which is necessary only because of the difficulties of formalizing the comparisons in any other way.

The first was that not all the stems to be compared were the same length. For example, the stem in the word for 'ear' has the form -tû or -twĺ in many of these languages; but in two languages it is disyllabic, being either -tóóyĺ or -túyĺ. One might guess that these are the older forms, and there has been some kind of shortening process in all the other languages. The solution that was adopted was to add dummy segments with entirely negative feature values to all the languages having a monosyllabic form. This did not affect the similarity measure within the monosyllabic group of languages; and it made the two languages having disyllabic forms more similar to the monosyllabic group than they would have been to another language which had a different second syllable.

The second problem arose when a phonetic feature such as palatalization was realized in one language in a consonant and in another in a vowel. The word for 'crocodile', for example, often has a stem of the form -góɲá; but sometimes, instead of the palatal nasal, the form is -góĺná. Note that if these two forms were lined up so that the consonants were compared only with the consonants and the vowels only with

the vowels, then there would be differences in both the last vowel and the last consonant. Consequently this pair would be counted as less similar than a pair such as -góónáá and -góódáá. This is not a desirable result. It was avoided by an ad hoc solution in which -in was arbitrarily specified as a consonant differing in one feature from the palatal nasal ɲ . Note also that the problem is not avoided by using the same features for consonants and vowels; it is simply a matter of the lining up of the segments to be compared.

The ad hoc approaches discussed above are, of course, unsatisfactory. They were adopted simply in the interests of expediency. Work is continuing on a better formalization of the problem of comparing whole words, but so far without success. Meanwhile, a computer program has been written which compares the features in each segment in each word in each language with the corresponding features in each word in every other language. The sums indicating the degree of phonetic similarity of each pair of languages are printed out in matrix form. The results for this particular group of 20 Ugandan languages are not particularly relevant here; they are given in detail elsewhere (Criper, Glick, and Ladefoged, forthcoming). It is sufficient to note that the relationships revealed suggested plausible and interesting groupings into dialect clusters.

What is of more interest here is the validation of the claim that this technique measures phonetic similarity between languages. We attempted to do this in two ways, first by assessing local opinion concerning the degree of similarity between one language and another, and secondly by testing the extent to which people actually understand

other languages. The first of these two methods did not produce reliable data; different local experts gave different figures, and even the same man gave different estimates when the questions were put to him in a slightly different way on different occasions. The second method produced limited but valid data. The procedures are described in full elsewhere (Criper, Glick, and Ladefoged, forthcoming). We conducted tests with speakers of two different languages. For each of these languages we used five groups of speakers, and played them recordings of stories in their own and four other languages, rotating stories, languages, and groups in a Latin square design. The group scores in answering questions about these stories were subjected to an analysis of variance, which showed that there were no significant differences between any of the listening groups, or between any of the stories; but there were very significant differences in the comprehension of the different languages. We therefore had valid scores on the comprehension of two languages relative to four other languages. These eight scores were compared with the degrees of phonetic similarity of the corresponding pairs of languages and, provided one score was left out for reasons discussed below, a high correlation was found ($r = 0.98$).

It is virtually impossible to test the relative comprehension of all possible pairs of a large number of languages, because of the complexities in the experimental design which are necessary. But it would appear that, at least in the case of these Ugandan Bantu languages, valid predictions may be made on the basis of the phonetic similarity measure described above. There are, however, circumstances in which our predictions would be wrong. The degree of comprehension of one

language to another is not always a reversible relationship; speakers of a prestige language do not understand a minor language as well as speakers of the minor language understand the prestige language. It is this discrepancy which accounts for our having to leave out one score in order to get a high correlation as described above. Phonetic similarity is a good predictor of intelligibility only if questions of prestige are not involved.

Finally we must consider ways in which we could improve the metric used for comparing the phonetic similarity of segments. Perhaps the most obvious improvement is to allow for variations in the importance of different features. The experimental studies cited above generally agree in finding that differences in manner of articulation contribute more to perceptual distance than differences in voicing, and both contribute more than differences in place of articulation. Accordingly features must be assigned different weights.

The situation is, however, more complicated. We must also allow for the interaction of features. For example, the experimental studies cited above have shown that there is a greater difference between the members of the set pa - ta - ka than there is between the members of the set ba - da - ga ; and the members of the set ma - na - ŋa are even less different from one another. Consequently differences in place of articulation, however coded, must be made to have less effect when the feature voiced is also present; and even less effect when the feature nasal is also present.

It seems that it would also be advisable to allow for non-binary specifications of features. Multivalued feature specifications can be

treated in either of two ways. In one way, each value is regarded as being equally different from all others. Thus if the consonants p , t , c , k are assigned the values 1, 2, 3, 4 on a feature of articulatory place, they will each be regarded as being one point different from each other with respect to this feature, assuming it has been given a weight of 1. Alternatively multivalued specifications can be treated as scalar quantities. If this is done and, for example, the vowels i , e , a are specified as having the values 1, 2, 3 on a feature of vowel height, then e would be counted as one point different from i and a , but i and a would be two points different from each other (assuming this feature has a weight of 1). If they had been specified as 1, 4, 7 then e would have been three points from i and a and they would have been six points different from each other.

The use of independent multivalued feature specifications allows us to correct an anomaly which was mentioned above. It will be remembered that using the previous system it was impossible to specify h in a way such that it was equally different from all stop consonants. But if place of articulation is an independent multivalued feature, and if h is assigned a value different from any of the stop consonants, then it can be made equally different from all of them. In other words, this type of specification allows us to formalize within the metric the notion of an irrelevant feature.

A computer program has now been written which compares segments which may be specified in terms of weighted, interacting, multivalued, independent or scalar, features. It is hoped that results of experiments using this program will be available for reporting to the conference.

References

Austin, W.M. (1957) 'Criteria for phonetic similarity' *Language* 33, 538-43.

Chomsky, A.N. and Halle, M. (1968) *The Sound Pattern of English* Harper and Row, New York, New York.

Criper, C., Glick, R., and Ladefoged, P. (forthcoming) *Language in Uganda.*

Greenberg, J.H. and Jenkins, J.J. (1964) 'Studies in the psychological correlates to the sound system of American English' *Word* 20, No. 2, 157-77.

Jakobson, R., Fant, G., and Halle, M. (1951) *Preliminaries to Speech Analysis* (sixth printing, 1965) Cambridge, Mass., M.I.T. Press.

Jakobson, R. and Halle, M. (1956) *Fundamentals of Language* Mouton, The Hague.

Klatt, D.H. (1968) 'Structure of confusions in short-term memory between English consonants' *J. Acoust. Soc. Amer.* 44, No. 2, 401-7.

Miller, G.A. and Nicely, P.E. (1955) 'An analysis of perceptual confusions among some English consonants' *J. Acoust. Soc. Amer.* 27, 338-52.

Mohr, B. and Wang, W. (1968) 'Perceptual distance and the specification of phonological features' *Phonetica* 18, 31-45.

Peters, R.W. (1963) 'Dimensions of perception for consonants' *J. Acoust. Soc. Amer.* 35, 1985-9.

Peterson, G.E. and Harary, F. (1961) 'Foundations of phonemic theory' in *Structure of Language and its Mathematical Aspects* (ed. R. Jakobson) American Mathematical Society, Providence, Rhode Island.

Wickelgren, W.A. (1965) 'Distinctive features and errors in short-term memory for English vowels' *J. Acoust. Soc. Amer.* 38, 583-8.

Wickelgren, W.A. (1966) 'Distinctive features and errors in short-term memory for English consonants' *J. Acoust. Soc. Amer.* 39, 388-98.

Table 1: Phonetic transcriptions of the words for 'bee' and 'bone' in 20 Ugandan Bantu languages. IPA symbols are used, except that j and c are used for the voiced and voiceless palatal affricates. Doubled letters denote long sounds. The stems (which are all that were used in the comparisons) are separated from the noun class prefixes by a vertical line.

Language	'bee'								'bone'							
Lumasaba	ɪ̀	n		z	ù	k	ɪ̀				ì	ì	g	ùù	mb	á
Lunyole	è	n		j	ù	h	ɪ̀		é				g	ùù	mb	à
Lusamia	è	n		j	ú	h	ɪ̀		é				k	ùù	mb	à
Lugwe	è	n		j	ú	h	ɪ̀		é				k	ùù	mb	à
Lugwere	ó	n		z	ò	k	ɪ̀		é		ì		g	ùù	mb	à
Lukenyi	e	n		j	ò	k	ɪ̀		é		ì		g	óò	mb	á
Lusoga	é	n		d̲	ù	k	ɪ̀		è		ì		g	ùù	mb	á
Luganda	è	n		j	ú	k	ì		è		`		gg	úú	mb	à
Ruruli	`	m		b	ò	c	ɪ̀		è		ì		g	ùù	mb	à
Runyoro	è	n		j	ó	k	ì		è		ì		g	ú	f	à
Rutooro	ò	r	ù	j	ó	k	ì				ì		g	ù	f	à
Ruhororo	è	n		ʒ	ó	c	ì				ì		g	ú	f	à
Rutagwenda	`	n		ts	ù	x	ɪ̀		è		ì		k	ùù	mb	à
Runyankore	ò	r	ù	ʒ	ó	k	ì		è		ì		g	ú	f	à
Rukiga	è	n		ʒ	ó	k	ɪ̀		è		ì		g	ú	f	á
Lubwisi	`	n		j	ó	k	ì		è				k	ú	w	à
Rukonjo	è	n		z	u	c	ì		è	r	ì		k	ú	h	à
Rugungu		k	à	h	ò	k	i		ì				k	ú	h	á
Runyarwanda	`	n		z	ú	k	ì		ì				g	ú	f	à
Rwamba	`	n		j	óó	k	ì			n			k	ú	w	é

Table 2: The classification of the places of articulation required for the description of Ugandan Bantu languages.

Example	Phonetic term	Characteristic Features		
		anterior	alveolar	coronal
b	labial	+	-	-
d̪	dental	+	-	+
d	alveolar	+	+	+
d-	postalveolar	-	+	+
ǰ	prepalatal	-	-	+
g	velar	-	-	-

Table 3: The classification of some manners of articulation required for the description of Ugandan Bantu languages.

Example	Phonetic term	Characteristic Features		
		nasal	stop	fricative
n	nasal	+	-	-
nz	prenasal fricative	+	-	+
nd	prenasal stop	+	+	-
d	stop	-	+	-
j	affricate	-	+	+
z	fricative	-	-	+
l	approximant	-	-	-

Table 4: The degree of similarity between some consonant segments in Ugandan Bantu languages.

	ɗ	d	d-	j	g	dʸ	dʷ	d:	dz	z	nz	l	r	h	s	ʃ	sʸ	ʃʸ
b	9	8	7	7	9	7	7	7	7	6	5	7	6	7	5	4	4	3
ɗ		9	8	8	8	8	8	8	8	7	6	8	7	6	6	5	5	4
d			9	7	7	9	9	9	9	8	7	9	8	5	7	6	6	5
d-				8	8	8	8	8	8	7	6	8	9	6	6	7	5	6
j					8	6	6	6	8	7	6	6	7	6	6	7	5	6
g						6	6	6	6	5	4	6	7	8	4	5	3	4
dʸ							8	8	8	7	6	8	7	4	6	5	7	6
dʷ								8	8	7	6	8	7	4	6	5	5	4
d:									8	7	6	8	7	4	6	5	5	4
dz										9	8	8	7	4	8	7	7	6
z											9	9	8	5	9	8	8	7
nz												8	7	4	8	7	7	6
l													9	6	8	7	7	6
r														7	7	8	6	7
h															6	7	5	6
s																9	9	8
ʃ																	8	9
sʸ																		9