DOCUMENT RESUME

ED 024 920 AL 001 284

By- Ladefoged, Peter; And Others
The Measurement of Cross-Language Communication.
Survey of Language Use and Language Teaching in Eastern Africa, Nairobi (Kenya).
Pub Date 1 Feb 68
Note- 7p.; Uganda Language Survey Working Paper.
EDRS Price MF-$0.25 HC-$0.45
Descriptors- African Languages, *Comparative Analysis, Computational Linguistics, Data Analysis, Distinctive Features, *Field Interviews, Field Studies, *Language Research, Languages, *Multilingualism, *Mutual Intelligibility, Research Methodology, Sociolinguistics
Identifiers- Cognates, *Uganda

Of the approximately 40 languages in Uganda, some are very similar to one another and may be to some extent mutually intelligible. Because no one knows how to measure degrees of mutual intelligibility, the authors are attempting to establish reliable techniques which would be not only of practical value for the study of language problems in Uganda, but also theoretically interesting for sociolinguistics. Discussed in this paper are possible techniques for testing mutual intelligibility--asking a number of informants to assess the extent to which they can understand and follow instructions in the other languages, assessing linguistic similarities. The authors also propose using a system of comparison of lexical items which include common cultural phenomena (avoiding words which may produce a number of equivalent forms). A new approach to the finding of cognate words in related languages entails eliciting two kinds of word lists from each informant--(1) The informant is asked to give common equivalents for each word. (2) The informant, given the word in the neighboring languages, is asked if there is a similar sounding word with the same or a related meaning in his own language. Also described is a proposal for computing the degree of similarity between each pair of words. the distinctive features will be encoded in binary terms. The authors hope that this data will be processed at the Kampala Computer Centre. (AMM)

BWP12

Uganda Language Survey
Working Paper Feb. 1, 1968

The Measurement of cross-language Communication

There are probably about 40 languages in Uganda. Some of them are very similar to one another, and may be to some extent mutually intelligible. But nobody knows how to measure degrees of mutual intelligibility. If we could establish a reliable technique it would be not only of practical value for the study of language problems in Ugarda but also theoretically interesting for sociolinguistics.

There are basically three ways in which the mutual intelligibility of different languages can be assessed. Firstly, informants can be asked to estimate the extent to which they can understand neighbouring languages. Secondly, one can test the extent to which speakers of one language can follow instructions given to them in another. Thirdly, one can assess the degree of linguistic similarity (in syntax, lexicon and phonology) among the languages in question. We propose using various forms of all three of these basic techniques, and examining the correlations between them. There is no convincing evidence that any of them is a valid measure of the difficulty of cross-language communication. As we shall see, there are both practical and theoretical problems in connection with each of them, and it is not surprising that sociolinguists have been reluctant to accept any of them. But since each of them is independently motivated (albeit weakly), if we could demonstrate a good correlation between any two of them, it would be far more likely that they are in fact valid. When such diverse techniques produce similar results there are strong grounds for accepting them.

There are several difficulties in asking informants to assess the extent to which they can understand neighbouring languages. In the first place one is interested not in the ability of a single informant, but in the average ability of all the members of the speech community to which he belongs. The best way that I can think of for getting this information is to ask a number of informants, each of whom represents (in some sense) the local speech community. Any guidance on techniques of selecting such informants would be much appreciated. My own preference would be for choosing informants who represented the community in knowing their views, rather than in being themselves typical. It is obvious that the questions should not be of the form "How much can you....." but more of the form "How much an people around here..... understand of a speech in such and such language." If skilfully put to a knowledgeable informant such questions should produce data about the community rather than the individual. Hopefully, the answer should be placeable on the scale "None; a little; the main points; most; all". If a series of such questions were asked about different hypothetical situations, such as "If someone preached a sermon.....; If someone told a story.....; If someone from such and such a place started

working in this village...." and if a number of different people were asked, it should be possible to get a reliable estimate of the beliefs of a speech community concerning the degree of communication possible in other languages. These beliefs might, of course, be biassed. It has often been shown (e.g. by Wolff 1959, Hymes 1968) that the attitude of one group of people to another has an effect at least on their belief about the mutual intelligibility of their languages. Consequently, the informants should be asked additional questions which probe their attitude to other groups. It may turn out that the results obtained by this technique can be correlated with those obtained by other techniques only when some weighting factor corresponding to the attitudinal differences is taken into account.

From the theoretical point of view, operational tests of cross-language communication seem open to fewer objections; but there are considerable practicable difficulties in using them on a country-wide scale. The work of the Summer Institute of Linguistics in Mexico is interesting in this respect. As far as I can recall (from conversations with John Crawford and Peter Landerman) their project involves going from village to village both making recordings, and also, in each village, playing recordings made in all the surrounding villages in order to test the extent to which they could be understood. The material which they record usually consists of a short account of some normal daily activity. They avoid using folk tales or any conventionalised material; and they also prefer not to use translations (which may be unreliable) of any of the standard English comprehension tests. This policy is no doubt the most appropriate for them; but it leads to difficulties in making up questions to test the degree of comprehension of each passage. In our circumstances, where we have skilled informants, I would prefer to use both passages and questions which have been carefully matched in their grammar and vocabulary, so that they might be presumed to be equal in their degree of difficulty of comprehension. But we are still left with the problem of how to select a group who would be capable of listening to a series of recorded tests, and somehow noting their answers. Classes in the primary schools, or groups of church members form a biassed sample of the local population. But they are probably the best we can get.

The third technique, the comparison of linguistic structures, has been widely used in historical linguistics and in language teaching methodology (where it is known as contrastive analysis). Since we are concerned with estimating cross-language communication, our approach will be slightly different. Ideally, we want to consider the phonology, the syntax, and the lexicon of each of the languages being compared. But nobody knows the relative contribution of each of these factors to mutual intelligibility; and in practice it is hard to quantify differences in syntactic and phonological patterns. Accordingly we propose

/p.3

using mainly a system of comparison of lexical items, albeit in a slightly new form. The first difference in our methodology lies in our choice of vocabulary items to be compared. A common technique (Swade 1950, 1955) is to use a 100 or 200 basic word list, carefully selected so as to be independent of any particular culture (of Gudschinsky 1954). But this is an unnecessary constraint when one is studying the cross-language communication among a limited group of languages whose speakers share many cultural phenomena. Thus elephants are common throughout the country, and are a proper subject of enquiry (but one would not, of course, ask questions about tigers or ice). The word list still has to be balanced so that it does not reflect one aspect of the culture unduly. It would be unwise for it to consist entirely of a list of animal names; a language might have borrowed these items and no others from its neighbours. But since, for assessing mutual intelligibility, we are not concerned with whether words have been borrowed from another language or not, we do not have to limit ourselves to the so-called basic vocabulary which is theoretically resistant to change through contact.

Because of the additional flexibility which is open to us for the choice of words to be compared, we can also avoid using words which may produce a number of equivalent forms. Thus for an English word like "come" there may be two or three words each of which might be regarded by different investigators or different informants as the correct common equivalent in a given language. But, to stick to our previous example, an elephant is an elephant, and there is unlikely to be a Ugandan language which has more than one generic term for "elephant". (They may, of course, have additional specific terms for "rogue elephant" etc.; but informants normally recognise these as specifics, and can easily be induced to give the general one.)

We are also going to try a new approach to the finding of cognate words in related languages. Nobody knows the part played in cross-language communication by the existence of cognate words with slightly different meanings (such as "mutton" in English and "mouton" in French); without further research I would not like to say whether they help or confuse. We propose getting two kinds of word lists from each informant. First we will elicit the common equivalents for each word. Then we will tell the informant the word in the neighbouring languages, and ask him if he has a **similar sounding** word with the same or a related meaning in his language. Again, we are trading on the fact that we are dealing with sophisticated informants, who are interested in language problems, and who can, with a little training, use their competence as native speakers to great advantage. Computing the degree of similarity between two languages using first the common equivalents and then the deliberately sought cognates should give us two different indices, either one of which might correlate with the other possible

/p.4

indices of cross-language communication.

We also propose trying a new way of computing the degree of smililarity between each pair of words. Since we are not concerned with measuring genetic similarity between languages, we need not stick to the criteria which are usual in lexicostatistics or in traditional historical studies, such as the necessity of regular, explainable, sound changes. All we are worried about is the degree of similarity in the sound of the words in the particular pair in question. On the basis of previous studies (reference to Wickelgreen (sp?)etc.) it would appear that the degree of similarity between two sounds depends on the number of phonetic features which they have in common. Accordingly we would like to develop a way of comparing each sound in each word in each language with the corresponding sound in each corresponding word in all the other languages. We are a long way from getting all the bugs out of the procedure at the moment. But it might work as follows.

Consider the data in Table 1.

TABLE 1. Unchecked data in orthography on 9 interlacustrine Bantu languages

English	fire	water	sun	moon	two	three	four
Luganda	omuliro	amaazi	enjuba	omwezi	biri	ssatu	nnya
Lusoga	omuliro	amadhi	endhuba	omwezi	ibiri	isatu	ina
Lugisu	kumuliro	kamezi	inyanga	kumwezi	zibiri	zidatu	zine
Lunyole	omuliro	amaji	elyuba	omwesi	ebiri	edatu	ene
Lugwere	omusyo	amaizi	isana	omwezi	ibiri	isatu	inna
Lusamia-Lugwe	omuliro	amachi	eryuba	omwoozi	chibiri	chidatu	chine
Rukiga	omuriro	amaizi	eizoba	okwezi	ibiri	ishatu	ina
Runyarwanda	umuriro	amazi	izuba	ukwezi	ebyiri	eshatu	enye
Runyoro	omurro	amaizi	izooba	okwezi	ibiri	isatu	ina

These are obviously closely related languages. The major problem is to ensure that when we quantify the degree of similarity in sound, we do so by comparing each segment with the appropriate segment in the corresponding word in the other languages. There are two steps to the solution we have chosen. Firstly, each word will be normalised so that the first consonant of the stem begins in a given location. Secondly, each word will be transcribed so that it consists of an alternating sequence of vowels and consonants. Double vowels and double consonants will be marked by a superimposed length mark; and the consonant compounds will be treated as single units with features of prenasility, or palatalisation, or labiovelarisation. Both these steps involve untested, but seemingly reasonable, assumptions about the subjective degree of phonetic similarity between cognate words in closely related languages. It also seems reasonable to assume that when the same word in two different languages has a similar form but a different number of segments in each language, then the extra segments should be considered

/p.5

as contributing only slightly to the perceptual differences. This will be achieved by adding dummy segments (unmarked segments, in the Chomsky-Halle sense) so that all words in a set have the same length. We might also consider the possibility of calculating differences in stems or roots separately from those in affixes. It is at least possible that speakers of related Bantu languages will be less put out by differences in prefixes than by differences in stems.

Each segment in each word will be fully specified in terms of a set of distinctive features. The distinctive features chosen will not be those of Jakobson & Halle (1955) or Chomsky & Halle (1968) or Ladefoged (1967), but will be a set appropriate for the population to be examined, namely, Ugandan languages. All the features will be binary. This is not to be taken as a reversal of any previous position stated in Ladefoged (1967 and elsewhere), but is partly for ease in computer processing, and partly because, at this stage of our ignorance, I would rather encode ternary and quaternary features in binary terms so as to avoid prejudging the issue of whether or not they are linearly ordered sets. The features are also set up so that certain redundancies are built in. This will result in some segments which might have been considered to differ in only one feature (in, say, the Jakobson-Halle system), nevertheless being counted as different in two. The redundancies have been used so that segments are characterised in a way which seems intuitively (an on the basis of the experimental evidence for English) to correspond to the degree of phonetic difference between them'. The possible segments for Luganda are shown in Table 2 on page 6. The difference between each of them is shown in the matrix in Table 3 on page 6.

It is hoped that all the data will be processed at the Kampala Computer Centre. Table 4 (page 6) exemplifies the inputs corresponding to two of the words in Table 1. It may be seen that these two words differ by a total of 7 points. If we added the total differences in 100 words, we should have a fairly sensitive measure of the degree of difference between two languages. For each pair of word lists we should get at least two counts, one in which only the stems were compared (say columns 5-10, if the first consonant of the stem had always been placed in column 5), and one in which whole words were compared (say columns 1-10). This assumes that no word has a prefix longer than 4 segments, or a stem longer than 6.

We hope to gather matched data in the form of 100-word lists from each of about 40 languages. These languages will be split into 4 groups, so that we compare only languages which belong to the same family. It would be virtually impossible to use this technique for languages which differ greatly. But this is not a source of concern, since we would not expect to find any cross-language communication between these languages. If this technique for assessing mutual

/p.6

intelligibility can be shown to be valid (by correlating it with the other techniques), it will be extremely convenient. It is obviously far more reliable than any technique which depends for its success on the careful selection of informants. To complie a good 100-word list one needs a single intelligent informant who knows the language well; one does not need to find a group of informants whose I.Q.'s and linguistic backgrounds are a representative sample of their speech community.

TABLE 2: Distinctive features of some of possible consonantal segments in Bantu languages. (To be extended later.) nny is the orthographic representation of a long palatal nasal. These features have been chosen and the values for segments assigned as indicated not for the purpose of setting up appropriate natural classes for use in phonological descriptions, but only so as to provide a measure of the degree of a phonetic similarity between segments (as shown in Table 4.) 0 = +; 1 = —

	p	b	m	pp	mp	py	pw	f	t	c	k	nny
stop	o	o	1	o	o	o	o	1	o	o	o	1
nasal	1	1	o	1	o	1	1	1	1	1	1	o
fricative	1	1	1	1	1	1	1	o	1	1	1	1
approximant	1	1	1	1	1	1	1	1	1	1	1	1
voiced	1	o	o	1	1	1	1	1	1	1	1	o
frontal	o	o	o	o	o	o	o	o	o	1	1	1
coronal	1	1	1	1	1	1	1	1	o	o	1	o
long	1	1	1	o	1	1	1	1	1	1	1	o
labialised	1	1	1	1	1	1	o	1	1	1	1	1
palatalised	1	1	1	1	1	o	1	1	1	1	1	1

TABLE 3: Number of differences in the feature specifications of some pairs of consonants in Bantu languages

	p	b	m	pp	mp	py	pw	f	t	c	k	nny
p	0	1	3	1	1	1	1	2	1	2	1	6
b		0	2	2	2	2	2	3	2	3	2	5
m			0	4	2	4	4	3	4	5	4	3
pp				0	2	2	2	3	2	3	2	5
mp					0	2	2	3	2	3	2	5
py						0	2	3	2	3	2	6
pw							0	3	2	3	2	6
f								0	3	4	3	6
t									0	1	2	5
c										0	1	4
k											0	5
nny												0

TABLE 4: Matrices for the comparison of the Lusoga and the Lugisu words for "three" (unchecked data)

Lugisu	Lusoga	z	i	d	a	t	u	-	i	s	a	t	u
Segment number		1	2	3	4	5	6	1	2	3	4	5	6
odd numbers	even numbers												
stop	high	1	o	o	1	o	o	1	o	1	1	o	o
nasal	mid	1	1	1	1	1	1	1	1	1	1	1	1
fricative	low	o	1	1	o	1	1	1	1	o	o	1	1
approximant	front	1	o	1	1	1	1	1	o	1	1	1	1
voiced	central	o	1	o	o	1	1	1	1	1	o	1	1
frontal	back	o	1	o	1	o	o	1	1	o	1	o	o
coronal	long	o	1	o	1	o	1	1	1	o	1	o	1
long	y-glide	1	1	1	1	1	1	1	1	1	1	1	1
labialised	high tone	1	1	1	1	1	1	1	1	1	1	1	1
palatalised	low tone	1	1	1	1	1	1	1	1	1	1	1	1

TABLE 5: Distinctive features of some of the possible vocalic segments in Bantu languages (Tonal features would be added to this set when appropriate)

	i	e	a	o	u	aa	ay
high	o	1	1	1	o	1	1
mid	1	o	1	o	1	1	1
low	1	1	o	1	1	o	o
front	o	o	1	1	1	1	1
central	1	1	o	1	1	o	o
back	1	1	1	o	o	1	1
long	1	1	1	1	1	o	1
y-glide	1	1	1	1	1	1	o

TABLE 6: Number of differences in the feature specifications of some pairs of vowels in Bantu languages

	i	e	a	o	u	aa	ay
i	0	2	4	4	2	5	5
e		0	4	2	4	5	5
a			0	4	4	1	1
o				0	2	5	5
u					0	5	5
aa						0	2
ay							0