CAMBRIDGE
UNIVERSITY PRESS

Language in Ethiopia: implications of a survey for sociolinguistic theory and method[1]

M. L. BENDER

Southern Illinois University

R. L. COOPER

San Diego State College

C. A. FERGUSON

Stanford University

ABSTRACT

Work of the Language Survey of Ethiopia is drawn on to illustrate that social surveys can contribute to scientific theories and methods as well as to policy making. Findings with regard to four general problems are indicated: a metric of linguistic distance (considered here in terms of selection of diagnostic traits for linguistic areas, subgrouping of related languages, and dialect variation within a language; measurement of basic cognate frequencies; and testing of mutual intelligibility, a typology of the processes of language spread (lingua franca, mother tongue, superposed variety); verbal deference behavior (in use of Amharic second person pronouns); and pidginization (Simplified Italian of Ethiopia). Some comparisons to findings in sociolinguistics research elsewhere are made. (Language surveys; measurement of linguistic distance; language spread; social meaning and mode of address; pidginization.)

Sociolinguistically oriented language surveys on a national scale have generally been justified in terms of how useful the collected data will be for language policy makers (Ferguson 1966), but an equally persuasive case can be made in terms of their contribution to the theories and methods of sociolinguistics. The Language Survey of Ethiopia carried out in 1968–9 (hereafter LSE)[2] is a case in point. The information which was gathered provides useful material for the policy maker who wants to make informed decisions leading to valued outcomes,

but in the course of the year's work the surveyors and their associates also contributed substantially to basic issues in the general study of language in society.

One fundamental problem of all sociolinguistic research and theorizing is the measurement of linguistic difference. In order to discuss how similar or how different two varieties of speech or writing are, there must be some kind of metric and some kind of operational strategy for applying the metric. At the Sociolinguistics Seminar held at Indiana University in 1964[3] one of the first questions the sociologists asked the linguists was: 'How do you measure the degree of difference between two languages (or two varieties of the same language)?' They asked this question expecting to receive a straightforward technical answer which they could then put to work for their purposes, but to their surprise they discovered that the linguists had no ready measure of linguistic distance or diversity and indeed had relatively little interest in the notion.[4]

The LSE gave considerable attention to this problem, following three different approaches: (a) the selection of diagnostic traits in terms of a 'linguistic area' comprising a large number of related and unrelated languages, in terms of subgrouping within sets of related languages (e.g. the Ethio-Semitic languages), and in terms of dialect variation within a relatively homogeneous speech community (mother-tongue speakers of Amharic); (b) computer processing of basic cognate frequencies among a large number of related languages and comparison of these results with other measures of linguistic difference; and (c) testing of mutual intelligibility among a small number of related languages and comparison of results with other measures. While none of the three approaches yields anything close to the desired general metric, all contributed significantly toward this goal and they will be discussed in turn in (1) below.

A second basic issue in sociolinguistic research is the phenomenon of language spread, i.e. the extension, over a period of time, of the use of one language or variety at the expense of others. In spite of the rather sizeable literature on language maintenance and shift among immigrant groups,[5] the development of national languages,[6] and the use of lingua francas,[7] sociolinguistic theory lacks even an adequate formulation of the major types or processes of language spread. The LSE focused a considerable amount of its research effort on understanding the way in which Amharic, the officially designated national language of Ethiopia, has spread in past centuries and is now spreading throughout the country; some attention was also given to the spread of other languages. The results of the several different research strategies employed led to a typology of language

---

[3] Cf. Ferguson 1965.
[4] A few exceptional studies had been concerned with this question: see, for example, Chrétien & Kroeber 1937; Gudschinsky 1955; Hymes 1960. 19–22, 26–7, and references there.
[5] Cf. Fishman 1964.
[6] Cf. Guxman 1960; Fishman et al. 1968.
[7] Cf. Heine 1970.

spread. One of the research techniques, 'transaction count', offers sociolinguists a promising new tool, and the typology itself is a contribution to the theory of language use. This issue is presented in (2) below.

Verbal deference behavior, especially the use of forms of address, has constituted one of the principal research topics in sociolinguistics (e.g. Brown & Gilman 1960). It is assumed that all speech communities have grammatical and lexical differentiation in forms of address (e.g. pronouns, names, titles) the use of which directly symbolizes differences in status, role, setting, and affective factors. In particular, many languages have more than one second person pronoun (e.g. Spanish *tu, usted*), and the study of their use, variation in their use, and changes in their use through time offers a valuable means of access to the wider norms and values and the processes of social change in the speech community. In the case of the Ethiopian survey it was possible to include Susan Hoben's anthropological study of the pronouns of address in Amharic, which gave additional exemplification of some of the theoretical generalizations currently held by linguists and also some unique material on the use of the gender distinction, which extends the known range of variation in this aspect of languages. The findings of the study are summarized in (3).

At the very core of sociolinguistic research is study of the interplay between linguistic change in its narrow sense and the matrix of social change in which it moves, and one of the most telling specimens of this interplay is the process of pidginization – the linguistic reduction which arises in certain social settings of language contact, and may have results which persist in one form or another for long periods (Hymes 1971). The LSE looked for significant instances of this process, and in the work of an Ethiopian investigator (Habte-Mariam Marcos) documented a variety of modified Italian with startling similarities in form and function to the original Mediterranean Lingua Franca of centuries ago (Schuchardt 1909). The findings of this study and their implications are outlined below in (4).

1. MEASURES OF LINGUISTIC DIFFERENCE

Diagnostic traits

The technique most commonly used in linguistic classification has been the selection of diagnostic traits appropriate for particular sets of languages and language varieties. If a group of languages or varieties shares certain selected diagnostic traits (or in some cases even a single trait) they are classified together in the same unit (e.g. family, subgroup, language); if two varieties fail to share one or more selected traits, they are classified as belonging to different units. The basic analytic problem is, of course, how to select useful or valid criterial features. There is also the additional problem of how to assign degrees of difference to particular traits or to relative numbers of traits. In spite of the problems in this

approach to classification it is used at various levels, from large groups of languages to dialect variation within a language. Three examples of the use of diagnostic traits as the basis of a measure of linguistic difference will be given from the Ethiopian Survey.

(a) *The Ethiopian language area*. Ferguson has suggested the existence of an 'Ethiopian language area' (*Language in Ethiopia*, ch. 5 and Ferguson 1970). To say that the Ethiopian languages constitute a language area means that they tend to share a number of features which, taken together, distinguish them from any other geographically defined group of languages in the world. Some of these shared features are due to genetic relationship, i.e. they continue features present in a remote ancestral language. Others result from the processes of reciprocal diffusion among languages which have been in contact for many centuries. Not every language in Ethiopia has all these features, and a few languages do not fit into the pattern at all, but in general most of the languages in an area roughly coincident with Ethiopia's borders have features of pronunciation, grammar, vocabulary, and patterns of expression which are, taken together, distinctive and characteristic of the area.

Eight phonological and 18 grammatical features were found to be useful traits in demonstrating the existence of the area. It is the clustering of these traits which demarcates the area from non-area languages such as English and Anyuak, a Nilotic language chosen as representative of the non-area languages which are found in the westernmost part of Ethiopia. For example, Amharic (an area language) has 6 of the 8 phonological features and 16 of the 18 grammatical features, whereas Anyuak has 0 of 8 and 2 of 18 respectively and English 2 of 8 and 0 of 18. Examples of the features are: presence of glottalic consonants, a helping vowel occurring to break consonant clusters, basic sentence order of SOV (subject–object–verb), use of the singular noun with numerals, and an irregular imperative of the verb 'come'.

(b) *Classification of the Ethio-Semitic languages*. The Semitic languages of Ethiopia are generally accepted as a genetic subgroup of the Semitic languages, i.e. as a group having a common ancestor which at one time existed as a single language within Semitic. This is not only because of the known geographic isolation of Ethio-Semitic speakers over a long period of time but also because of the set of traits found common to all these languages but not found in other Semitic languages. In the same way, the Ethio-Semitic languages may be classified into subgroups on the basis of traits shared by some but not all the languages in question.

The Ethio-Semitic classification adopted by the Survey is that of Robert Hetzron, who suggests a more complex picture of Ethio-Semitic history than has been generally accepted (*Language in Ethiopia* and for a fuller account, Hetzron forthcoming). For one thing, Hetzron argues in favor of laying to rest the term 'Gurage' as an overall linguistic term. According to Hetzron, there is no Gurage

linguistic unit, and 'Gurage' should be used to refer only to 'a conglomeration of Semitic-speaking peoples surrounded by Cushitic (Sidamo)'. Some so-called Gurage languages are more closely related to Amharic and Harari than to other 'Gurage' languages and some belong with Gafat in a different branch of South Ethio-Semitic.

His classification is based on a careful study of traits known as 'shared innovations', i.e. changes which have taken place in subgroups of the overall language group (in this case Ethio-Semitic) which set them apart from other subgroups. For example, the first division in Ethio-Semitic is between the Northern and the Southern languages. The Northern languages share a common characteristic in the forms of the verb in the Imperfect (roughly: the present) and the Perfect (roughly: the past) as against the Southern languages. In the Northern forms, the next-to-last consonant is doubled only in the Imperfect, in the Southern forms only in the Perfect. In this case it is the Southern forms which have innovated: the Northern forms are believed to be the older pattern. Another innovation setting off Southern languages is that they make a sharp distinction in the selection of forms of main verbs and subordinate verbs, as against the Northern languages which have very much the same forms in both main and subordinate verbs.

(c) *Amharic dialects.* There has as yet been no large-scale or thorough study of the nature and range of dialect differentiation in Amharic, although it is a language of considerable socio-political importance, being a national language and one which is expanding from its present base of about 8,000,000 native-speakers plus many non-native speakers. The literature contains minor references to Amharic variation, and indeed most observers report a surprisingly small range of variation for a language spoken by mostly illiterate persons in a country of difficult communications and transportation. This lack of extensive variation may be partly explained by the existence of a well-developed common culture including the monarchy and the orthodox church.

The Ethiopian survey sponsored a small-scale dialect investigation of Amharic, carried out by eight members of the Department of Ethiopian Languages and Literature at the Haile Sellassie I University under the direction of Dr Getatchew Haile. This investigation was supplemented by a report on the Gonder area and some information on lexical variation by Roger Cowley (cf. *Language in Ethiopia,* ch. 6).

The main findings of this investigation were that '. . .phonologically, Menz and Wello form a subcluster within a cluster consisting of Gojjam, Menz, and Wello, and Addis Ababa–Gonder form a separate cluster, though Gonder shows some traits intermediate between the two larger clusters. On the syntactic level, however, the dialects of Addis Ababa, Menz, and Wello seem to belong to one dialect cluster, while Gojjam and Gonder are each separate dialects on their own.'

In general the diagnostic traits identified in this dialect study result from dif-

ferential incidence of diachronic changes. For example, there are different degrees of palatalizing in the various dialects reported, i.e. some dialects palatalize more consonants than others, or palatalize in more environments, or both. (Palatalization is a phonological process which may be thought of generally as inserting a -y- glide after a consonant, resulting in shifting t to č, s to š, l to y, etc.) The situation is fairly complex, but can be explained by assuming that the order of palatalizing rules differs in the grammars of the various dialects, or that rules differ in degree of generality among the dialects. It is this variation in order or generality of application of rules which accounts for the particular classification of the Amharic dialectal varieties.

Basic cognate frequencies

Given the linguistic complexity of Ethiopia and the problems of data collection imposed by the country's size and difficult terrain, it was quickly decided that a collection of uniform samples of basic vocabulary and minimal grammatical data on the lesser-known languages would be likely to lead to a useful classification most immediately. This collecting was restricted mainly to the complex and little-investigated areas of the south, south-west, and west (Sudan border). The resultant new data together with some previously-collected material was then processed by visual inspection and computer program. The results are found in Bender (1971) and will be discussed briefly below.

Essentially, the main result of the computer analysis is a tabulation of percentages of assumed lexical cognates among 101 languages of Ethiopia and nearby countries. The languages are then arranged into major families by previous knowledge of the main outlines of classification and by clusterings of high cognate percentages. Detailed study of the clustering of percentages then leads to a final suggested subgrouping, to be modified as further data on the languages in question and data on languages as yet uncollected become available.

The closeness of fit of this new classification and that proposed by Greenberg for the relevant languages as part of his general classification of African languages (Greenberg 1966) is significant because Greenberg's classification is not based on statistical methods. Thus we have a verification in this case that careful comparison of all available data by the mass comparison methods advocated by Greenberg (Greenberg 1957: chs 3, 4; Greenberg 1966: 1–5) and statistical analyses of uniform samples of basic lexicon lead to the same results within close margins. The slight differences in the two results raise interesting problems and point toward likely directions for further investigation.

In the process of subgrouping languages according to basic vocabulary shared, several indices were developed and used. These indices are similar to those used by Dyen (1965) (e.g. Dyen's *critical difference*, p. 19), but go beyond Dyen in one important respect, namely that they are generally ratios of one percentage to another (weighted ratios). Since the denominators of the ratios are the averages of

all cognate-percentages within given groupings, they take into account the relative 'tightness' or 'looseness' of the groupings. Several weighted ratios were calculated for each of three units of Afro-Asiatic: The Ethio-Semitic super-group (4th level in the hierarchy with super-family as first level), the Lowland East Cushitic group (level 5), and the Ometo subgroup (level 6). In general weighted ratios lead to significantly better subgroupings than unweighted ones, in terms of incisiveness of subgroupings and avoidance of overlappings, thus verifying the hypothesis that the more elaborate methods lead to sharper results.

One very interesting and unexpected result emerged. This is that the sharpest subgroupings of all are established by the use of the *numbers* of items which share phonological innovations. Thus, suppose languages A and B share 62 cognate items out of 98 and that 32 of the 62 share phonological innovations not found among all the languages of the grouping in question. It is the use of this 32 and all other such numbers found in comparing sets of languages which produces the best subgrouping. This result provides corroboration of the widely accepted idea that shared innovations in general provide the best basis for linguistic subgrouping.

The significance of this is that a fairly large-scale application of quantitative methods has led to the same answer as the usual method based on theoretical and common-sense considerations, very much in the spirit of Chrétien and Kroeber's quantitative study in Indo-European classification (Chrétien & Kroeber 1937). Furthermore, the finding in this case is probably the first attempt to overcome Greenberg's criticism of glottochronological methods in his essay on linguistic subgroupings (Greenberg 1957: 54). The idea behind the use of items with shared phonological innovations was Bender's, but the motivation for seeking an improved lexicostatistic method stemmed from Greenberg's criticism.

Mutual intelligibility

Probably the question most frequently asked of the Ethiopian survey team by laymen was: 'How many Ethiopian languages are there?' The answer depends on the criteria used for separating languages. Certainly mutual intelligibility is a fundamental criterion in this connection. Paul Bohannon is certainly correct when he suggests that the study of inter-intelligibility of African languages is important but still in its infancy (Bohannon 1964: 129).

Three of the five survey teams of the overall Survey undertook mutual intelligibility studies. The pioneering effort was by the Uganda team and it established the precedent of testing intelligibility by testing listening comprehension using uniform texts among speakers of related languages. The result of the experiment on a number of related Bantu languages of Uganda was that mutual intelligibility is substantially correlated with shared basic vocabulary as well as with the more readily determinable relative geographical proximities of the languages (Ladefoged, Glick & Criper 1971: 65–77).[8]

[8] In connection with the Uganda study, Ladefoged also observed that closeness of

The Ethiopia team, having the advantage of the Uganda team's efforts and benefiting from their insights, was able to carry out a more elaborate study on the Sidamo (or Highland) group of East Cushitic. The resulting correlations of intelligibility with shared basic vocabulary and geographical proximity replicated the Uganda results and suggested that they may be of universal validity.[9] In addition, significant correlations with in-text lexical and in-text combined lexical and grammatical morphemes were found, though the correlation with in-text grammatical morphemes alone was small. In addition, the study shows the Sidamo group to consist of five languages: Hadiyya, Kembata, Sidamo, Deresa, Burji. Alaba, at first considered a separate language, is now seen to be best considered a dialect of Kembata. Fuller results are to be found in Bender & Cooper (1971) and *Language in Ethiopia* (ch. 4).

A third intelligibility study is that of the Zambia team on several Bantu languages. (See Ohannessian & Kashoki, forthcoming.) One possible interpretation of their results is that strong correlation of intelligibility and geographical proximity depends on a relatively undisturbed situation after an initial set of dispersion migrations from the homeland of the proto-language speakers. A method of verification or disconfirmation would be to look at the historical backgrounds of migrations in the three areas in question (in so far as they can be determined) to see if the idea is supported by the evidence in all three cases.

The Ethiopia survey thus supplies a reasonable answer to the question: 'How many languages are there?' The answer is 70 to 80, discounting European imports, extinct languages (such as Giiz, the classical liturgical and literary language), and jargons and special 'secret languages', and allowing for the discovery of a few as yet unknown languages believed to exist in the south-west and west. The number of languages is in itself of both practical and theoretical value in the context of the usual naïve estimates ranging from 3 (Amharic, Tigrinya, and Giiz, the three literary languages of Ethiopia) to 200 (including many mutually intelligible varieties).

2. LANGUAGE SPREAD

Typology

By the seventeenth century, Amharic had spread from its home base, east and north of Lake Tana, through much of central Ethiopia, and the extension of Ethiopia's boundaries in the nineteenth century has helped promote the language's continued expansion. Just as Giiz, the classical language of Ethiopia, spread

---

languages as measured by comparing phonological feature decompositions of the segments of corresponding basic vocabulary items gave no sharper results than simply deciding on possible cognates as units. (Personal communication, Peter Ladefoged.)

[9] But cf. the problems posed by Wolff (1959); Wurm & Laycock in New Guinea (1961); Voegelin *et al.* (1963: 24); and cf. Hymes 1968.

as a national language in the ancient Aksumite empire, so is Amharic steadily expanding today as a principal factor in national integration.

An attempt by the LSE to characterize the way in which Amharic is expanding led to the formulation of a general typology of language spread. The typology posits three ways in which the substantial use of a language can spread. One way is as a lingua franca, when the language becomes commonly used for purposes of between-group communication.[10] A second way is as a mother tongue (language first spoken), which occurs when people adopt the language as the language of their home. A third way is as a superposed variety, i.e. the language spreads for purposes of within-group communication, but not as a mother tongue.

When a language spreads as a lingua franca, it spreads as a second language. The people who learn it continue to speak their mother tongue for purposes of within-group communication. There have been many examples of the spread of lingua francas in Africa (Greenberg 1965). Some of these have been indigenous languages like Hausa and Malinke in western Africa and Swahili in eastern Africa. Others have arisen as pidginized forms of African languages in the European contact situation, such as Sango and Lingala in central Africa. Still others have been European languages, particularly English and French, in the former colonial areas. African lingua francas have arisen not only in the context of trade (e.g. Hausa, Dyula), but also in such contexts as administration and education (e.g. English), work (e.g. Fanagalo in the Zambian copperbelt), and religion (e.g. Arabic). For an extensive discussion of African lingua francas, see Heine (1970).

When a language spreads not as a lingua franca but as a mother tongue, people gradually abandon their mother tongue and adopt the new language for purposes of within-group communication. This type of language spread, language shift (Fishman 1964), may occur when migrant peoples adopt the dominant language of the area to which they have come, as in the case of most American immigrant groups. Language shift may also occur when the political and economic pressures exerted by a dominant minority induce a subordinate majority to abandon their mother tongue. There have been many times in which the territorial expansion of an empire has created wholesale language shift among its subject populations, as in the cases of Latin and Arabic.

The spread of a language as a superposed variety resembles both the spread of a lingua franca and language shift. It resembles the spread of a lingua franca because the new language is used for purposes of within-group communication. When a language spreads as a superposed variety, it displaces the original

---

[10] Other terms for lingua franca include trade language (langue de traite, langue commerciale, Handelssprache), vehicular language (langue véhiculaire, Verkehrssprache), and contact language (langue de relations, langue passe partout, Kontaktsprache). See Heine (1970: 15). In our usage, a lingua franca is any language which is used for purposes of between-group communication, between speakers for whom it is a second language and who do not share a common first language. See Greenberg (1965).

language in certain functions only, while the original language holds fast to other functions.[11] In such circumstances, the displacing language and the language it partially displaces come to coexist with different functional allocations in the same speech community. An example of such partial displacement can be found in Paraguay, where Spanish has replaced the indigenous Guarani for some purposes or in some contexts (e.g. in non-rural, formal, non-intimate contexts), but where Guarani is maintained for others (Rubin 1968). With respect to Amharic, it is likely that it is spreading in all three of the ways that have been outlined: as lingua franca, as mother tongue, and as superposed variety (Cooper, in press).

Household surveys

The LSE employed several methods in its study of the spread of Amharic. One method was the language census, a technique whereby the household head or other household representative answers questions about the language proficiency (and sometimes the language usage and language attitudes) of each member of his household (see Fishman 1969; and Lieberson 1966, 1969). The LSE procedure departed from that of the usual census in that the LSE adopted a sample survey approach, rather than attempting to interview a representative from all households in the population under study (Cooper, Singh, & Abraha, in press).

The LSE carried out sample surveys in two largely Galla-speaking areas.[12] In each area, surveys were taken in the towns as well as in the countryside adjacent to the towns. In all, over 2,400 respondents, drawn from a population of about 485,000 were interviewed. Each respondent was asked to identify those languages he could presently speak and the language he spoke first as a child, and he was also asked to identify the first and second languages of each of the other members of his household who was at least seven years old. The data obtained were analyzed for each of the 57 enumeration districts into which the survey areas had been divided.

The results of greatest interest from the point of view of sociolinguistic theory was the relationship between the bilingualism of mother-tongue groups and the relative size of each group. The relationship was a negative or inverse one. That is, the higher the proportion claiming a given language as mother-tongue, the smaller the percentage of bilinguals (here defined as people who knew *at least* two languages) among that mother-tongue group. Conversely, the smaller the mother-tongue group as a proportion of a district, the greater the proportion of bilinguals in that group. The same inverse relationship between relative size of a mother-tongue group and the incidence of bilingualism within it was found by Weinreich (1957) in his analysis of the 1951 census of India, and may represent a sociolinguistic universal.

[11] We are indebted to Joshua A. Fishman for pointing out that a language can spread by a process of partial displacement.

[12] These surveys were a joint undertaking of the Work-Oriented Adult Literacy Project (sponsored by Unesco and the Ethiopian Ministry of Education) and the LSE.

The pressure to learn a second language is, of course, not exerted by the mere presence of the first-language group but rather by the interactions which its presence implies. For example, the correlation between size of mother-tongue group and proportion of others who speak it was higher for Amharic than for Galla. This can be attributed in part to the relative sociocultural or sociopolitical statuses of the mother-tongue groups in question. Thus, for example, an analysis by Lieberson (1970: 46–50) of bilingualism in the residential districts of six Canadian cities indicated that while the proportion of mother-tongue speakers was related to the proportions of others who knew that language, the pressure to learn English as a second language was greater than the pressure to learn French. For the same percentage of native speakers, a greater proportion knew English than French as a second language. Thus, a greater percentage of people knew English than French as a second language, not only because of the greater size of the English Canadian group but also because of its dominant socioeconomic position.

The findings of the Ethiopian sample surveys, then, are consistent with those obtained in two other sociolinguistic contexts: India and Canada. The results of these studies suggest that a sociolinguistic model can be built to predict the extent to which the members of one group will learn the language of another when the two groups are in contact. Such a model could take the form of a multiple-regression equation, with the relative proportions of each mother-tongue group in the population serving as two of the predictor variables. Additional variables might include for each group its degree of residential segregation, its degree of occupational segregation, the proportion of literates, its relative importance in the economic affairs of the community (estimated, for example, by indices such as *per capita* income, proportion of cultivated acres controlled, proportion of trading establishments operated, or proportion of corporate directorships held), and its relative political influence (judged, for example, by the proportion of civil servants, or the proportion of political appointees, or the proportion of heads of various administrative units). If estimates could be obtained for such variables for a number of districts, the relative importance of each variable in predicting the degree to which each language has been learned as a second language by the other group could be empirically determined, and the degree of explained variance would serve as a criterion for the success of the model. While estimates for some of these variables may necessarily be crude, an attempt to combine them via multiple-regression analysis should nonetheless yield information of considerable interest. Such analyses should help us to understand better those processes that impel one group to learn the language of another.

Market surveys

A common function of lingua francas is that of trade language. When buyers and

sellers have different mother tongues, as is often the case in linguistically hetero-
geneous settings, one language sometimes becomes established as the language
used for between-group communication in market contexts. The language spreads
along the trade routes and outward from market centers, and those who wish to
enter trade must learn to speak that language.

In the linguistically diverse Ethiopian setting, one might predict that either
Amharic or Arabic serves as a lingua franca for trade. Amharic would be pre-
dicted because it has been introduced throughout the country via schools and
other government agencies. Amharic, furthermore, appears to be spreading in
other contexts through the processes of industrialization and urbanization.
Arabic would be predicted because in Ethiopia Muslim merchants have tradi-
tionally dominated small retail trade and because one might expect Ethiopian
Muslims to have some acquaintance with Arabic through attendance at Quranic
schools. Furthermore, there is a widespread belief in Ethiopia that Arabic is used
as a market lingua franca by Islamic co-religionists. In the markets studied by the
LSE, however, neither Amharic, nor Arabic, nor indeed any other language
served to any significant extent as a lingua franca for trade (Cooper & Carpenter
1969).

The market surveys were carried out in twenty-three markets in eight towns.
None of these towns was located in the Amharic heartland of the country, which
is relatively homogeneous linguistically. Although most of the towns had a sub-
stantial proportion of Amharic mother-tongue speakers, other languages domi-
nated the countryside surrounding the towns. The survey procedure followed in
each market was the same. Enumerators, drawn from among the students at the
local high school, tallied the number of market transactions observed in various
languages on a single market day. A number of commodities and services was
chosen (e.g. cloth, onions, salt, tailoring), and each enumerator observed language
usage in each of these items. Thus, for each item, each enumerator noted the
number of transactions between buyer and seller carried out in Amharic, the
number carried out in Galla, etc. This technique permitted highly reliable esti-
mates to be made of the relative importance of various languages in the market
because representative items were chosen for observation and because large
numbers of observations were made. Hundreds of transactions were observed in
the smaller markets and thousands in the larger ones. Approximately 1,700
transactions were observed in the average market studied.

The principal finding of the market surveys was that the proportion of
transactions observed in a given language could be predicted on the basis of the
proportion of residents in the town who claimed that language as mother tongue
or home language (as determined by Central Statistical Office sample surveys).
The discrepancies between the proportion of transactions observed in a given
language and the proportion of urban residents claiming that language could be
explained in large part by the presence of people from the countryside who came

to the town on the market day. Thus the proportion of transactions observed in Amharic was generally smaller than the proportion of urban residents claiming it, and the proportion of transactions observed in the dominant language of the countryside was generally larger than the proportion of urban residents claiming it.

If a language were serving the function of trade language, presumably the proportion of transactions observed in that language would have been substantially greater than the proportion of respondents in the town who claimed it. Where such differences were in fact observed, they could be attributed to the influx from the countryside of people who spoke it natively. In no case was the proportion of transactions in Amharic or in Arabic substantially greater than the proportion of townspeople claiming it as first language. Thus these languages did not appear to be serving as trade languages in any substantial way.

One cannot infer from these results that the proportion of buyers who spoke each language natively was the same as the proportion of sellers who did so, or that the transactions were typically between people who spoke the same first language. On the contrary, the division of labor in Ethiopia, particularly as realized in handicrafts and in trade, has long been along ethnic lines. Amharas, for example, have traditionally eschewed trade as a livelihood. Thus, instead of the buyer and seller typically interacting in a common first language or in a common second language, it is likely that the seller typically accommodated himself to the buyer by speaking the buyer's first language. In the linguistically diverse contexts of these Ethiopian markets, therefore, it appears that transactions were facilitated by the multilingualism of the traders rather than by the emergence of a trade lingua franca.

The results of this study are relevant to sociolinguistic theory in several ways. First, they demonstrate that heterogeneous market settings do not necessarily give rise to trade lingua francas. Second, they demonstrate that a language need not spread uniformly through all domains or contexts of language use. Whereas Amharic is spreading for some purposes and in some contexts, it is not expanding as a trade language, at least not in the markets observed. These findings, therefore, raise interesting questions. Why do trade lingua francas arise in some linguistically heterogeneous settings but not others? Why do languages spread in some contexts but not others? Why, for example, did Hausa spread with Muslim traders in West Africa, whereas Arabic did not spread with the Lebanese in West Africa? The construction of a unified theory that can answer such questions and accommodate these findings would be a fruitful enterprise.

Finally, these results are relevant to sociolinguistic theory because they demonstrate the usefulness of a new technique of sociolinguistic inquiry, that of the transaction count, which permits the systematic and unobtrusive observation of language usage as it occurs. This technique appears to be a promising one. It can be used as a survey method in settings where household canvassing is not

possible. Where such canvassing can be performed, the technique makes it possible to cope with a problem found in all social science, the validity of self-reported behavior. Thus, a group's self-reported language usage (what languages people *say* they use in particular contexts) could be compared with their actual usage. Although there is evidence that in bilingual communities self-reported frequency of usage of a given language in specified contexts can be a valid indicator of global *proficiency* in that language (Fishman & Cooper 1969), there have been few if any attempts to assess the validity of such reports as estimates of relative *usage*. To account for the agreement or discrepancy between self-reported and actual usage, we would need to know (1) the degree to which language usage is prescribed for a given situation or, conversely, the degree of variability in language usage in that situation, (2) the degree to which the norms of language usage are explicitly recognized or known by members of the community, and (3) the willingness of respondents to tell what they know. The relationships and interactions among such variables are of considerable sociolinguistic interest. In addition to permitting a comparison between reported and actual usage, the transaction count technique, by being used in an array of societal contexts, can help build sociolinguistically valid descriptions of the functional specialization of languages in multilingual settings, as well as explanatory models of language spread.

3. VERBAL DEFERENCE

A study of the meanings of Amharic second-person pronouns, as they are understood by Amharas, yielded findings which are of general interest. First, the use of second-person pronouns in Amharic was found to conform in general outline to the western European usage described by Brown & Gilman (1960), wherein the dimensions of relative status (power) and intimacy (solidarity) account for the pronouns chosen. Second, Amharic pronominal usage demonstrates a rare phenomenon, that of gender fluctuation in direct address. The observations which are summarized here were made by Hoben (in press) via the anthropological techniques of participant observation.[13]

There are ten independent or self-standing pronouns in Amharic. These can be characterized by the intersection of five features – number (singular, plural), person (first, second, third), gender (masculine, feminine), and deference (familiar, polite). Four of these are second-person pronouns: *antɛ* (singular, masculine, familiar), *anči* (singular, feminine, familiar), *irswo* (singular, unspecified gender, polite), and *innantɛ* (plural, unspecified gender, unspecified deference).[14]

---

[14] Note that not all of the theoretically possible distinctions occur. For example, there

In Amharic, the person of the subject is shown by pronominal affixes on the verb, and in the second person either the familiar or the polite form must be used. If an independent pronoun is also used to mark the subject of the sentence in direct address, the pronoun and pronominal affix must agree with respect to the form which is chosen. The choice between the familiar and polite forms in Amharic is determined in general by the speakers' social position relative to one another and by the solidarity of their relationship.

Relative social status among the Amhara is determined by the interaction of many personal attributes, including religious purity, title, education, and age. Each of these attributes, in turn, is complex. Purity, for example, may be gained in several ways, among which are monastic vows or ordination. There are also many kinds of title, which may be conferred in terms of the traditional systems of ecclesiastical, military, or court ranks, or in terms of the modern administrative, military, and police bureaucracies. One person may have many claims to high position and deference. He may, for example, be old, religiously pure, educated, and titled. Another may be high on some scales but low on others. Furthermore, not all of a person's attributes may count in every situation. His claims to deference will depend upon those factors which are considered important in the particular situation. For example, the presence of a third person may cause some speakers who would normally use the familiar with one another to use the polite form.

Amharas normally use the familiar form with family members, neighbors, and friends. However, the attributes which are associated with polite address may sometimes cause speakers to use the polite form in circumstances in which the familiar form would otherwise be expected. Thus, even among blood relatives, the age, religious purity, rank or education of one may cause another to address him with the polite form.

There are other uses of the second-person pronoun besides deference marking. These pronouns may also be used to indicate the emotional reaction of the speaker to the person addressed. For example, the familiar form may be substituted for the polite form as an insult or to express anger. In addition, the feminine form may be applied to a man if the speaker wishes to insult him. In Addis Ababa, but not in the countryside, a man may also use the feminine form when addressing another man if he wishes to indicate affection. Thus, men who are close friends may use the feminine form with each other. Hoben found, for

---

is no second-person, plural, polite form in Amharic. In addition to the ten forms described, there is an eleventh independent pronoun, *antu*, which is said to show a degree of formality midway between *antɛ* and *ɨrswo*, to be used, for instance, for untitled, non-office-holding elders (male or female). Hoben believed that her data were insufficient to show any reliable semantic differences between *antu* and *ɨrswo*, and for purposes of this report, the forms are not distinguished. In the province of Wello, *antu* replaces *ɨrswo* for the second-person singular polite form, and *ɨrswo* is used for the third-person singular polite form (Hailu, Getatchew & Cowley, in press).

example, that a group of school teachers was divided into two subgroups which were clearly marked by the mutual use of the feminine form. Within each subgroup, the members used the feminine form with one another, but across subgroups they used the masculine form. She also reported that the smallest man in a group of friends may sometimes receive the feminine form but use the masculine form when addressing the others.

In Amharic, the gender of the noun fluctuates. A noun may be considered masculine or feminine according to whether its referent is considered to be large and distant (masculine) or small and intimate (feminine). Thus, the optional gender fluctuation found in the second-person pronoun is analogous to the gender fluctuation of the noun.

The study of second-person pronoun usage provided data of general interest to sociolinguistic theory as well as information of particular relevance to Amharic and to Ethiopia. Though there may be nothing new in the use of participant observation for the description of second-person pronoun usage, the usefulness of new data against which previously made generalizations can be checked is clear. While the set of features which command respect is specific to Amhara society, the power and solidarity semantics were found to be operative. That these dimensions should be observed in a context quite different from that of western Europe raises some interesting questions. First, are power and solidarity semantic universals? Second, is deference a syntactic universal? Third, does the spread of a language imply the spread of the deference usage associated with that language? It would be of interest, for example, to compare the Amharic deference usage of sociolinguistically differing groups. The investigation of second-person pronoun variation among speakers of Amharic, then, helps us learn not only about linguistic and social variation among the Amhara but also about the more general processes underlying the interpenetration and interaction between linguistic and non-linguistic social behavior.

4. PIDGINIZATION

The process of pidginization is widely attested throughout the world in various situations of language contact. Although there are divergent views on the exact nature of the process and the conditions for its appearance and maintenance, it is generally agreed that pidginization involves such features as reduction of vocabulary and elimination of inflectional affixation of a source language, restriction in use to particular communication situations, and contribution from two or more languages in phonology and grammar. Among the best known examples are a number of pidgins and creoles with vocabulary drawn from a European language (English, French, Portuguese, Spanish) as base language. Surprisingly, however, there seems to be no well-attested example of an Italian-based pidgin in current use even though the earliest known and most famous instance of pidginization, the original lingua franca of the mediterranean area, was based on an Italianate

variety of Romance (Schuchardt 1909). Accordingly, the LSE attempted to determine whether a pidginized variety of Italian is in use in Ethiopia, since the historical conditions seem likely to have led to such a phenomenon and there are casual reports of people speaking 'broken Italian'.

Investigation showed that pidginized varieties of Italian do exist, differing greatly from the mother-tongue Italian and approximations to it also spoken in Ethiopia. These pidginized varieties, called Simplified Italian of Ethiopia (Habte-Mariam Marcos, in press) are apparently used only in certain limited functions (e.g. between Europeans and Ethiopians in certain occupational settings) and have probably been dying out ever since the Italian presence as a colonial power or occupying force was ended (although some of Habte-Mariam's informants had apparently acquired their SIE after the Italian occupation). Of special interest was the discovery that SIE is sometimes used as a lingua franca between Ethiopians of different mother tongues. A specific example was the report of a Tigrinya-speaking functionary that he regularly used SIE in talking with Kunama speakers with whom he shared no other language. The demonstration of the existence of SIE and Habte-Mariam's description of some of its salient characteristics constituted a contribution to sociolinguistic data, but perhaps of greater importance was the evidence it offered on the nature of the pidginization process.

Among factors given different degrees of emphasis in various sociolinguistic explanations of the pidginization process are the following: a natural (possibly innate, universal) process of simplification, use of 'baby talk' by speakers of the base language, 'interference' from the phonological and grammatical structures of the other language(s), mutual adaptation to a lowest common denominator between languages, 'relexification', i.e. replacing of one base vocabulary by another one. Ferguson (1971 and forthcoming) has recently reemphasized the possible role of a conventionalized 'foreigner talk' register used by speakers of the base language, differing in many respects from a 'baby talk' register.

Evidence is fairly substantial for the existence of an Italian foreigner talk regarded as appropriate (a) for talking to Africans and other 'native' people, and (b) as representing the broken Italian of such people. It ranges from the language used in a sixteenth-century play about a Gypsy woman to the language attributed to Africans in modern Italian cartoons and comic strips. A striking feature of this foreigner talk is the use of only two verb forms, the infinitive and past participle, of which the latter refers to past time and the former to all other times. This feature also appears in SIE, and seems to be a very strong piece of evidence, since Ethiopian languages do not have any parallel to such a use, and normal Italian does not work at all in this way. The reasonable explanation would seem to be that Italians speaking with Ethiopians sometimes used this well-established feature of foreigner talk, which was also found in the original lingua franca. Indeed Schuchardt's argument here deserves verbatim quotation:

No one disagrees that an Arab who knows the verb *mangiar* in the meaning 'to eat' must have learned it directly or indirectly from an Italian, but that he also uses *mangiar* for '(I) eat', '(you) eat', '(he) eats', etc., is usually set to his own account. But even if there is special effort on both sides to make oneself understood by the simplest means, and especially to give up the inflectional complexity of the base language, how could it happen that the Arab who is as yet ignorant of Italian chooses *mangiar* as representative for *mangio*, *mangi*, *mangia*, etc.? Only by very great familiarity with Romance could he be able to recognize the statistical predominance and functional generality of the Romance infinitive, and even then, since in his own language there is nothing corresponding to this infinitive, he would probably take the third person singular and say, for example, not *mi voler mangiar* but *mi vuole mi mangia*. It is the European who puts the stamp of general usage on the infinitive . . . (Schuchardt 1909: 443–4, tr. CAF.)

The line of argument seems just as valid today for a speaker of Tigrinya or Amharic, and the Ethiopia data provided valuable confirmatory evidence for the importance of the factor of foreigner talk in the formation of pidginized languages.

REFERENCES

Bender, M. L. (1971). The languages of Ethiopia: a new lexicostatistic classification and some problems of diffusion. *AL* **13**. 165–288.

Bender, M. L. & Cooper, R. L. (1971). Mutual intelligibility within Sidamo. *Lingua* **27**. 32–52. [Also ch. 4, Mutual intelligibility and classification: Sidamo languages. In Bender et al., *Language in Ethiopia*.]

Bender, M., Bowen, J. D., Cooper, R. L. & Ferguson, C. A. *et al*. In press. *Language in Ethiopia*. Nairobi: Oxford University Press.

Bohannon, P. (1964). *Africa and Africans*. New York: American Museum Science Books.

Brown, R. & Gilman, A. (1960). The pronouns of power and solidarity. In T. A. Sebeok (ed.), *Style in language*. Cambridge and New York: MIT Press and John Wiley. 253–76. [Also in J. A. Fishman (ed.) (1968). *Readings in the sociology of language*. The Hague: Mouton. 252–75.]

Chrétien, C. D. & Kroeber, A. L. (1937). Quantitative classification of the Indo-European languages. *Lg* **13**. 83–104.

Cooper, R. L. & Carpenter, S. (1969). Linguistic diversity in the Ethiopian market. *JAfrL* **8** (3). 160–8. [Revised as ch. 18, Language in the market. In Bender *et al*, *Language in Ethiopia*.]

Cooper, R. L. (in press). The spread of Amharic. In Bender *et al*., *Language in Ethiopia*. Ch. 23.

Cooper, R. L., Singh, B. N. & Ghermazion, A. (in press). Mother tongue and other tongue in Kefa and Arusi. In Bender *et al*., *Language in Ethiopia*. Ch. 17.

Dyen, I. (1965). *A lexicostatistical classification of the Austronesian languages*. (Memoir 19 of *IJAL*). Bloomington: Indiana University Press.

Ferguson, C. A. (1965). Directions in sociolinguistics. *Items* **19** (1). 1–4.

—— (1966). On sociolinguistically oriented language surveys. *Linguistic Reporter* **8** (4). 1–3.

—— (1971). Absence of copula and the notion of simplicity: a study of normal speech, baby talk, foreigner talk and pidgins. In Hymes 1971.

—— (forthcoming). Toward a characterization of English foreigner talk.

Fishman, J. A. (1964). Language maintenance and language shift as a field of inquiry; a definition of the field and suggestions for its further development. *Linguistics* 9. 32–70.
—— (1969). A sociolinguistic census of a bilingual neighborhood. *AJS* 75. 323–39.
Fishman, J. A. & Cooper, R. L. (1969). Alternative measures of bilingualism. *Journal of Verbal Learning and Verbal Behavior* 8. 276–82.
Greenberg, J. H. (1957). *Essays in linguistics*. Chicago: University of Chicago Press.
—— (1965). Urbanism migration, and language. In H. Kuper (ed.), *Urbanization and migration in West Africa*. Berkeley: University of California Press. 50–9, 189. Reprinted in Y. A. Cohen (ed.) (1968). *Man in adaptation, the biosocial background*. Chicago: Aldine. 259–67.
—— (1966). *The languages of Africa*. Bloomington: University of Indiana Press.
Gudschinsky, S. C. (1955). Lexico-statistical skewing from dialect borrowing. *IJAL* 21. 138–49.
Guxman, M. M. (1960). *Voprosy formirovanija i razvitija nacionalnyx jazykov*. Moscow: Akademija Nauk.
Habte-Mariam M. (in press). Italian. In Bender *et al.*, *Language in Ethiopia*. Nairobi: Oxford University Press. Ch. 13.2.
Hailu, F., Haile, G. & Cowley, R. (in press). Dialect variation. In Bender *et al.*, *Language in Ethiopia*. Ch. 6.2.
Heine, B. (1970). *Status and use of African lingua francas*. Munich: Weltforum Verlag.
Hetzron, R. (forthcoming). Ethiopian Semitic: studies in classification. *Journal of Semitic Studies Monograph Series*.
Hoben, S. J. (in press). The meaning of the second-person pronouns in Amharic. In Bender *et al.*, *Language in Ethiopia*. Ch. 22.
Hymes, D. (1960). Lexicostatistics so far. *Current Anthropology* 1. 3–44.
—— (1968). Linguistic problems in defining the concept of 'tribe'. In J. Helm (ed.), *Essays on the problem of tribe*. Seattle: University of Washington Press. 23–48.
—— (ed.) (1971). *Pidginization and creolization of languages*. London: Cambridge University Press.
Ladefoged, P., Glick, R. & Criper, C. (1971). *Language in Uganda*. Nairobi: Oxford University Press.
Lieberson, S. (1966). Language questions in censuses. *Sociological Inquiry* 36. 262–79.
—— (1969). How can we describe and measure the incidence and distribution of bilingualism? In L. G. Kelly (ed.), *Description and measurement of bilingualism*. Toronto: University of Toronto Press. 285–348.
—— (1970). *Language and ethnic relations in Canada*. New York: Wiley.
Ohannessian, S. & Kashoki, Mubanga E. (forthcoming). *Language in Zambia*. Nairobi: Oxford University Press.
Rubin, J. (1968). *National bilingualism in Paraguay*. The Hague: Mouton.
Schuchardt, H. (1909). Die lingua franca. *Zeitschrift für Romanische Philologie* 33 441–61.
Voegelin, C. F., Voegelin, F. M., Wurm, S., O'Grady, G. & Matsuda, T. (1963). Obtaining an index of phonological differentiation from the construction of non-existent minimax systems. *IJAL* 29. 4–28.
Weinreich, U. (1957). Functional aspects of Indian bilingualism. *Word* 13. 203–33.
Wolff, H. (1959). Intelligibility and inter-ethnic attitudes. *AL* 1 (3). 34–41.
Wurm, S. A. & Laycock, D. C. (1961). The question of language and dialect in New Guinea. *Oceania* 32. 128–43.