This study examined language maintenance through a sociology study of K'iche', a Mayan language of Guatemala, where a stable diglossic situation is eroding in the face of an increasing tide of bilingualism. Seven K'iche'-speaking communities were examined, including an analysis of socioeconomic, demographic, and political data. Quantified observations were made of the speech transactions of 11,229 speakers. Qualitative data were examined in the framework of Ethnolinguistic Identity Theory. Quantitative data were subjected to statistical analysis using categorical models, maximum likelihood analysis, and chi-square to determine the effect of race, sex, age, and domain on language use. A Language Maintenance Index was calculated for each age group and domain of use. A Global Language Maintenance Index permitted comparison of intercommunity language maintenance levels. The communities were found to be at different levels of language maintenance in spite of an intact diglossic relationship between Spanish and K'iche'. The communities have different combinations of ethnolinguistic identity factors and the differences in language maintenance levels can be related to differences in demographic, institutional support, status, and subjective vitality factors. (Contains 19 references.)

(Author/NAV)
Measuring K'iche' (Mayan) Language Maintenance:
A Comprehensive Methodology

M. Paul Lewis
Summer Institute of Linguistics

January 7, 1995
1995 Annual Meeting
Linguistic Society of America
New Orleans, Louisiana
Abstract

Measuring K'iche' (Mayan) Language Maintenance: A Comprehensive Methodology

A study of the sociology of language of K'iche' (a Mayan language of Guatemala) was undertaken in order to examine language maintenance. This study examined seven K'iche'-speaking communities and included both an analysis of socioeconomic, demographic and political (i.e. qualitative) data, as well as quantified observations of 11,229 participants who were involved in speech transactions in the seven communities.

The qualitative data were examined within the framework of Ethnolinguistic Identity Theory (Giles and Johnson 1981), (Giles and others 1991) providing a profile of each community. The quantitative data were subjected to statistical analysis using categorical models maximum likelihood analysis and chi-square to determine the effect of race, sex, age and domain on language use. In addition, a Language Maintenance Index was calculated for each age group and domain of use. This index provided a means of ranking the age groups and domains of use within each community. A global Language Maintenance Index, calculated for each community, provided a means of comparing the language maintenance levels of the communities with each other.

The communities were found to be at different levels of language maintenance in spite of the existence of an intact diglossic relationship between Spanish and K'iche'. The communities have different combinations of ethnolinguistic identity factors and the differences in language maintenance levels can be related to these differences in demographic, institutional support, status and subjective vitality factors.
Introduction

The language use situation among the K'iche' of Guatemala is unclear. In spite of 500 years of Spanish dominance, K'iche' maintenance has been assumed. For most of that period Mayans have been denied access to most of the roles in society where Spanish is deemed appropriate and so have had little opportunity for the acquisition and use of Spanish. The language use situation could be characterized as one of diglossia without bilingualism (Fishman 1967; 1986). Over the last 60 years social changes have taken place in Guatemala which have opened the doors for Mayans to acquire Spanish and to participate in at least some of the domains from which they had previously been excluded. As a result there have been warnings of language shift among the K'iche's and other Mayan groups. The stable diglossic situation is eroding in the face of an increasing tide of bilingualism. Little quantitative research has been done on Mayan language use but increasingly, as the threat of language shift has risen to consciousness, such studies have been called for in order to allow both scholars and technicians of applied linguistics to gain an understanding of the dynamics of the current language use situation.

The difficulty, of course, is the size of the task. While quantitative data on language use is required, simple counts of language use provide only superficial evidence of behavior that is motivated by sociological, economic, political, and cultural factors. Without reference to these more ethnographic types of data the frequency counts provide only a very flat view of the dynamics of language maintenance and shift.
This paper reports on one effort to provide an in-depth measure of the language use situation in the highland K'iche'-speaking area of Guatemala. The data collection was undertaken by members of the Summer Institute of Linguistics in Central America during the two year period from August 1986 to July 1988 as reported in Lewis (1987) and the bulk of the detailed analysis of the data was undertaken as my doctoral dissertation (Lewis 1994). The intended scope of the study was the entire K'iche'-speaking area but the primary focus was on seven communities, which are representative of the major dialects of K'iche'. These communities, Chichicastenango, Cunén, Joyabaj, Sacapulas, San Andrés Sajcabajá, Santa Cruz del Quiché and Totonicapán are both linguistic and demographic centers comprising more than 25% of the K'iche' population according to the 1981 Guatemalan government census figures.

The goal of our study was to cover as large an area as possible but to cover it in depth by examining two sets of data. We collected not only language use data but also data on ethnolinguistic vitality factors as well. Because of these goals we chose to become participant observers in the communities with our personnel residing in each community and learning the language and culture. We chose data gathering methodologies which included interviews with community leaders and change agents but which primarily stressed observation and the unobtrusive collection of attitudinal and language use data. Our goal was to minimize as much as possible the observer's paradox by doing our data gathering as part of our own participation in the community and by hiring residents of the community who would also participate in the observational process within their own social networks.
I will describe briefly the methodology used and the kinds of data collected but more importantly I will evaluate the feasibility and effectiveness of such a comprehensive approach to the study of language maintenance.

The Ethnolinguistic Vitality Data

The ethnolinguistic vitality data consist of an ethnographic and socioeconomic description of each community using a research guide developed by our literacy department in the Summer Institute of Linguistics in Central America. This instrument, called a Community Resource Profile, consists of 115 probe questions regarding geographic, linguistic, sociolinguistic, political, economic, worldview and cultural factors. The questions are not meant to be asked directly of interviewees but rather are designed to guide the researcher into areas which merit investigation in order to cover the broad range of qualitative factors which affect identity, attitudes and values, and prevailing social, economic and political pressures. Its purpose is to assist the analyst in designing literature development programs which are appropriate to the needs of the community. In the analysis stage of the study, I classified these research domains in terms of the type of information they provided regarding status, boundary maintenance, and subjective vitality, the rubrics identified as significant in Ethnolinguistic Vitality Theory as developed by Howard Giles and his numerous colleagues (Giles 1977; 1980; Giles and Saint-Jacques 1979; Giles and Smith 1979; Giles et al. 1985; Giles and Johnson 1986; 1987; Giles et al. 1991). Ethnolinguistic Vitality Theory provides a framework into which the data about a group can be placed and provides a certain degree of predictive power as well. A measure of a group’s ethnolinguistic vitality can be seen as a description of that group’s identity focus and can provide indications of the group’s probability of language maintenance. In addition,
Ethnolinguistic Vitality takes into account in a fairly structured way the factors mentioned by Fishman (1991) as being important dimensions in the assessment of the degree of dislocation of a minority language group. Fishman’s list of factors which constitute the causes of language and culture shift include language policies, and physical, demographic cultural and social disruption.

The resulting qualitative profiles of the seven communities provide a more structured description of each community and allows us to categorize the communities in terms of their ethnolinguistic vital signs. This structuring of the data in terms of a theoretical framework does not eliminate the analytical and interpretive difficulties associated with qualitative data, but it does provide a consistent set of parameters which can be used for comparative purposes between communities.

The Language Use Data

The purpose of this part of the data collection process was to collect quantifiable data on “who speaks what language to whom and when” (Fishman 1965). Although interviews, questionnaires, and self-report methodologies have been successfully used in other studies of language use (e.g. Showalter 1991) they were rejected because of the strong possibility that the results would not truly represent the actual language use patterns of the communities and for reasons of cultural appropriateness and sensitivity to a region traumatized by years of civil unrest in which the motivation for such questions could be easily misinterpreted.

As a result the language use data consists of observations of speech interactions among members of each of the seven communities. These observations were made by trained observers, both expatriates and K’iche’s resident in the community. The participants in each speech interaction were categorized according to age, sex, and race and each interaction was classified
M. Paul Lewis

Measuring K'iche' Language Maintenance

according to the language used and according to topic/location of the interaction. Observations
were made as unobtrusively as possible. No recordings or transcriptions were made. Observers
had forms on which they were to record their observations but this was usually done after the
fact and not in the sight of those being observed. The intent was to study authentic language
usage with as little intrusion by the observer and the observation process as possible. Only the
pertinent characteristics of the participants were recorded on the forms with all participants
remaining unidentified other than in the general terms of age, sex, race and occupation or social
role (mother, father, teacher, merchant, client, etc.).

This technique enabled us to collect a great deal of data in a relatively short time. A total
of 4,920 observations were made in the seven communities which included more than 11,222
participant interchanges. The number of observations made in each community varies between a
low of 406 and a high of 898. Similarly, the number of participants involved in these
observations from each community ranges from 550 to 2,329. The sampling method was
unsystematic guided primarily by the topic/location categories (e.g. home, market, street, church,
school, etc.) which were suggested at the beginning of the research and by the daily and weekly
routines of the observers who were instructed to make their observations in their spheres of
activity and social interaction. These initial topic/location categories were augmented as data
collection progressed by new ones provided by the observers themselves. These raw
categorizations for analytical purposes were later classified, based in large part on Fishman's
broader identification scheme, into ten domains of use which include: home, personal
encounters, recreation, market, work, religious meetings, stores, mass media, formal education,
and government offices. These ten domains, in the order presented, progress from the most
Measuring K'iche' Language Maintenance

intimate and informal domains to the more public and formal domains. In addition, the
topic/location categories were characterized more generally as being either Formal or Informal
domains of use. This provides a less fine-grained means of measuring language use in each
community but also enables us to make some generalizations about the state of diglossia in these
K'iche'-speaking communities.

The participants in each speech interaction were also identified by age. This
identification was an estimate made by the observer. In order to maintain the unobtrusive nature
of the methodology, speakers were never interviewed and thus there was no opportunity to ask
them their age directly. The age estimates provided on the observation data forms have been
classified into 6 age groups, 1-12 years, 13-24, 25-34, 35-44, 45-54, and those 55 and older. The
differences in language use between these age groups have also been compared in order to
provide an indication of age grading in language use and/or the state of intergenerational
language transmission.

The participant identification data (race, sex, and age) were recorded for all participants
in a speech interaction whether speaker or interlocutor. In some of the observed interactions, one
or more of the interlocutors was a silent participant, never speaking. We felt it was important to
keep track of the role of the interlocutor's identity since Gal's experience (Gal 1978)
demonstrates clearly how important the interlocutor's identity can be in affecting language
choice.

Data Analysis

The most difficult part of the analysis of the data was the interpretation of the qualitative
data. Though the ethnolinguistic identity factors were identified and described for each
community, it was exceedingly difficult to express them in such a way as to make them comparable between communities. In addition the seven communities showed considerable diversity in their combinations of ethnolinguistic strengths and weaknesses. Where one community might evidence strong boundary maintenance and institutional support for language and identity maintenance, it might at the same time be characterized as having weak subjective vitality. Another community would have a very different configuration of these factors. This makes it quite difficult to use the ethnographic data to rank the communities in terms of their relative ethnolinguistic vitality.

Nevertheless, the community profiles have proved quite useful in identifying the common factors which seem to be at work in the region and which are placing pressure on the residents of the communities to make an identity shift (See Lewis 1993) which includes with it a positive evaluation of the acquisition and use of Spanish particularly in the formal and public domains of use. Briefly summarized and therefore stated in an overly simplistic way, those communities whose ethnolinguistic vitality profiles show a positive evaluation of economic activities based on cash rather than on subsistence farming, while ostensibly much stronger in their subjective vitality (i.e. they feel good about themselves) are at the same time generally weaker in their boundary maintenance (i.e. they are more willing to adopt Latin ways) and ascribe less status to their K'iche' identity (a strong sentimental attachment to K'iche' but an equal or greater instrumental attachment to Spanish (Kelman 1971). Frequently in these communities, the societal institutions (church, school, local development groups, government) are more strongly motivated to use Spanish than to promote the maintenance and use of K'iche'. This upwardly-mobile socioeconomic motivation, driven by the need to negotiate with and market
their products to the outside, non-K'iche'-speaking, world, as well as a positive evaluation of a
modern identity is an extremely pr-al one and is reshaping the social organization of these
communities and disrupting the intergenerational transmission of K'iche'.

This disruption can be seen in the analysis of the language use data. The quantitative data
is much more amenable to comparisons between communities and to the more rigorous
standards of statistical verification of significant differences between levels of language
maintenance. In spite of the not-insignificant difficulties with the raw data and the sampling
method (See Lewis 1994:133-136 for a fuller discussion of these problems.) the language use
observation data provide a clearer picture of how the sociopsychological factors of
ethnolinguistic vitality play themselves out sociolinguistically.

Several methods were used in analyzing the speech interactions that were observed. One
method was to apply the categorical models maximum likelihood statistical procedure to the
data to determine the effect of race and sex of both speaker and interlocutor on the choice of
language used in any given interaction. Again, briefly stated and simplified, not-unexpectedly
this analysis showed the strong influence of the race of the participants on language choice. In
most cases, the independent variables for race of speaker (RACE) and/or race of interlocutor
(IRACE) were not influential by themselves but only in combination with each other or with
either the sex of the speaker (SEX) or the sex of the interlocutor (ISEX). But in every
community race was shown to play a role. A summary of these findings is shown in Table 1
A second method of quantifying and measuring language maintenance was to calculate a Language Maintenance Index. This index is arrived at by assigning a weighting factor to each of the language varieties used (K'iche' = 2, Code-Mixed = 1, Spanish = 0). The frequency count for each of these varieties is then multiplied by the weighting factor and that number is divided by the total number of observations to arrive at an index number which lies between 0 and 2. A higher index number indicates greater K'iche' maintenance and a lower number indicates less K'iche' maintenance. This technique was used to arrive at Language Maintenance indices for each age group, each domain of use, as well as a global Language Maintenance index for each community. These indices together provide a profile of language maintenance for each community.

In addition to the calculation of the Language Maintenance Index, I also categorized the Language Maintenance Index scores as either strong, moderate or weak. This categorization was arrived at by categorizing all index numbers within one half standard deviation of the mean as being within the moderate range. Table 2 shows the ranges of index scores which fall into each category.
The comparisons of the Language Maintenance Indices for each community by age and domain produce yet another profile of the language use patterns. As with the qualitative data, the combinations of strong, moderate and weak age groups and domain groups show considerable diversity with some communities demonstrating strong maintenance in certain age or domain categories where others are evidencing moderate or only weak K'iche' maintenance.

The domain of particular interest for language maintenance is the home domain, since that is the primary locus of intergenerational language transmission. In addition the age groups most closely connected with intergenerational transmission, the younger age groups and those age groups corresponding to young married adults are also of interest. Language use patterns evidenced in these categories it would be hoped could provide a window on the future of language use in each community. Table 3 summarizes the Language Maintenance Index scores for each community by age group and Table 4 summarizes the Language Maintenance Index scores by domain group.
<table>
<thead>
<tr>
<th>TOWN</th>
<th>HOME</th>
<th>STREET</th>
<th>PLAY</th>
<th>MARKET</th>
<th>WORK</th>
<th>RELIGION</th>
<th>STORES</th>
<th>MEDIA</th>
<th>SCHOOL</th>
<th>GOVT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chichicastenango</td>
<td>1.53</td>
<td>1.54</td>
<td>1.21</td>
<td>1.55</td>
<td>1.76</td>
<td>1.13</td>
<td>1.21</td>
<td>1.51</td>
<td>1.13</td>
<td>1.06</td>
</tr>
<tr>
<td>Cusán</td>
<td>1.99</td>
<td>1.70</td>
<td>1.89</td>
<td>1.80</td>
<td>1.88</td>
<td>1.43</td>
<td>1.77</td>
<td>0.97</td>
<td>1.80</td>
<td>1.07</td>
</tr>
<tr>
<td>Juyubaj</td>
<td>----</td>
<td>----</td>
<td>----</td>
<td>1.32</td>
<td>1.89</td>
<td>1.05</td>
<td>0.70</td>
<td>----</td>
<td>----</td>
<td>----</td>
</tr>
<tr>
<td>Sacapulti</td>
<td>1.81</td>
<td>1.82</td>
<td>1.50</td>
<td>1.72</td>
<td>1.55</td>
<td>1.21</td>
<td>1.60</td>
<td>1.41</td>
<td>1.71</td>
<td>1.64</td>
</tr>
<tr>
<td>San Andrés Salcabajá</td>
<td>1.46</td>
<td>1.73</td>
<td>1.80</td>
<td>1.78</td>
<td>1.77</td>
<td>1.12</td>
<td>1.46</td>
<td>1.79</td>
<td>0.79</td>
<td>1.29</td>
</tr>
<tr>
<td>Sta Cruz del Quiché</td>
<td>1.46</td>
<td>1.65</td>
<td>0.62</td>
<td>1.70</td>
<td>1.54</td>
<td>0.87</td>
<td>1.32</td>
<td>1.76</td>
<td>0.36</td>
<td>1.53</td>
</tr>
<tr>
<td>Totonicapán</td>
<td>1.89</td>
<td>1.55</td>
<td>0.66</td>
<td>1.75</td>
<td>1.33</td>
<td>0.91</td>
<td>0.96</td>
<td>0.91</td>
<td>1.08</td>
<td>1.23</td>
</tr>
</tbody>
</table>

<----MORE INTIMATE------------------------LESS INTIMATE----->
An alternative, but less revealing, statistical method of analysis of the age and domain data is to use chi-square to determine if the differences between categories are statistically significant.

Evaluation of the Methodology

The methodology presented here, as with all research methods, has advantages and disadvantages, and is very much shaped by the field situation in which it was developed and by the goals of the research project of which it was a part. It is offered as an example of an attempt to evaluate language use in a comprehensive fashion taking into account not only overt observable behavior, but the social context in which that behavior is situated. This study is an attempt to mesh both qualitative and quantitative data in order to arrive at a sociolinguistic profile of each community.

The strong point of the methodology is that it relates language behavior from the sociological perspective of "who speaks what to whom and where" to the social psychological perspective of Ethnolinguistic Vitality Theory with its analysis of societal pressures which affect individuals and groups in their choice of language variety. The scope of this study also provides us with an in-depth analysis of the seven target communities which represent a good cross-section of the highland K'iche' area of Guatemala. It is a strength of the methodology that it can be applied to such large scale investigations provided there are sufficient personnel available for the amount of time required to complete the ethnographic research. In addition the collection of the language use observations can be carried out, with supervision, by residents of the communities after only very little training.
There are some improvements that could be made as well which would affect both the reliability and validity of the data. This study in Guatemala was very much a learn-as-you-go effort. Any replication of this work should greatly benefit from the lessons learned from this first experience.

A first area of concern is the need for a research instrument more specifically designed to investigate ethnolinguistic vitality. While the Community Resource Profile covers the topics of interest, it is an instrument that was designed for a different purpose (the evaluation of literature development prospects). The data could have been collected more efficiently and more uniformly had a "less blunt" instrument been designed. A better research guide might also result in a greater level of comparability and connectability of the data collected with the language use observations.

The language use component of the study could have benefited from more careful thought in terms of its design and implementation. There are several design features that might have been dealt with in order to make the data more reliable. The sampling method, as mentioned above, was unsystematic thus reducing the generalizability of the conclusions of the study. In addition, most of the observations were made in the town centers making the results less representative of the rural areas of each township, though there are sociological reasons to expect that this might not be as serious a deficiency as might be expected. A stratified sampling method based on accepted statistical procedures and the demographic profile of each community could very well have saved time by reducing the number of observations needed while at the same time achieving a greater level of reliability. This would enable the conclusions to be more broadly generalizable.
Better operational definitions would also have reduced some of the complexity in the analysis of the collected data. We used "common sense" definitions for most of the independent variables. Race and age are two relatively trivial examples of variables for which we provided no guidelines for our observers. Fortunately, the non-linguistic markers of race in Guatemala are fairly obvious reducing the deleterious effect of that particular lapse. By categorizing age in terms of broader age groupings, we also were able to overcome the lack of precision in our observers' estimates.

A more difficult problem was the definition of which language was being spoken. The language used was identified as being either K'iche', Spanish, or Code-Mixed. While the first two might seem fairly clear-cut categories, the third, is not nearly so easy to identify nor to distinguish from one or the other of the two languages. The identification of an utterance as being K'iche' becomes quite subjective when the number of assimilated Spanish loan words is taken into account. What one observer might consider to be "pure" K'iche' could be considered code-mixed by another. Because of the priority we placed on being unobtrusive we made no recordings nor transcriptions. We therefore could not analyze each speech transaction at our leisure in order to apply some set of criteria to identify the utterances as being "pure" or code-mixed. Were we to do it again, we would almost certainly not expect our observers to be able to do such an analysis on the spot, but we would deal with the topic in our training sessions in an attempt to socialize the observers around a norm and thus increase our inter-rater reliability.

This particular problem has implications for the comparability of the data from the seven communities as well. Since the communities chosen were representative of different dialects of K'iche' and since they also were chosen because of their different levels of exposure to and
contact with outsiders, the level of assimilated loan words can be expected to be different from community to community. The "pure K'iche'" variety spoken in one community may very well be quite similar linguistically to the code-mixed variety spoken in another. We did not attempt to establish any absolutes in the identification of language variety used but rather relied on the fact that most of our observers were working within their own speech communities and would thus apply the locally salient criteria in their categorizations of the language used. A study of how contact with Spanish has affected the regional dialects of K'iche' is a much needed one, but was not within the realm of our possibilities.

In summary, the K'iche' study demonstrates that the use of qualitative and quantitative approaches is both useful and feasible. The observed language behavior can be placed within the context of the social, economic, political and demographic forces which surround and shape it. The combination of the two sets of data allows for the construction of fairly detailed sociolinguistic profiles of the communities which are more instructive than simple counts of language use or self-reports of language preferences. In addition the study demonstrates the feasibility of such a methodology for both a large-scale and, at the same time, an in-depth study of the ethnolinguistic vitality of a region. While this methodology, because of its reliance on ethnographic techniques, does require trained personnel and more time than a questionnaire or interview methodology might, it also makes use of local, minimally-trained investigators for the collection of the language use observations. With proper consideration of the caveats mentioned above and with appropriate adaptations for the contexts in which it will be used, I believe it can be a valuable tool in providing better quality and more useful documentation of the status of endangered languages.
M. Paul Lewis

Measuring K'iche' Language Maintenance

References Cited

Kelman, Herbert C. 1971. Language as a aid and barrier to involvement in the national system. The motivation and rationalization for language policy. Can language be planned?
M. Paul Lewis

Measuring K'iche' Language Maintenance


Lewis, M. Paul. 1987. Un estudio de la sociología del lenguaje del idioma Quiché. WINAK: Boletín Intercultural. 2.4.249-255.

