

Ministère de l'Enseignement Supérieur, de l'Informatique  
et de la Recherche Scientifique

Institut des Sciences Humaines

Centre de Recherches et d'Etudes Anthropologiques

PRINCIPLES IN SOCIOLINGUISTIC SURVEYS

ACTUALITIES OF CAMEROON SIL

Alan Starr and Jürg Stalder

December 1990

Société Internationale de Linguistique  
B.P. 1299, Yaoundé  
République du Cameroun

# Principles in Sociolinguistic Surveys Actualities of Cameroon SIL

December 1990

By Alan Starr and Jürg Stalder

## 1. Introduction

Being in the work of indigenous language development since 1934, a work that has expanded to a worldwide scope in the last fifty years, the Summer Institute of Linguistics, SIL (Société Internationale de Linguistique) has been concerned with the proper assessment of the language situation in each location where it has worked. This concern has grown in light of the reality that as many as 2,500 languages of the world still require assessment as to their need for development. New emphasis and focus were given to the enormous task of language assessment, and it was declared a priority for the 80's. Thanks to this new emphasis, in 1987 the SIL Branch of Cameroon established a survey department which would oversee the task of assessment for the numerous complex language situations in need of such an effort in Cameroon.

SIL in Cameroon works under contract with the Ministry of Higher Education, Research and Computer Services (MESIRES) and the Institute of Human Sciences (ISH). The survey department in Cameroon has been privileged to work alongside CREA (Centre de Recherche et d'Etudes Anthropologiques), one of the Cameroonian research centers which is administered by ISH. From the very beginning collaboration between CREA and SIL was strong in the development of the methodology and strategy which would be used to do the momentous task of language assessment in Cameroon. The experience which CREA linguists brought to the survey task was very beneficial, having been involved in doing a general survey of Cameroonian languages from 1978 - 1983, which resulted in the publishing of the Atlas Linguistique du Cameroun, ALCAM, Inventaire Préliminaire in 1983. Combined participation in actual surveys has also been beneficial to the SIL survey department both linguistically and culturally. The strategy and approach in these first years in carrying out surveys has evolved and continues to do so, and the SIL survey team is grateful for the opportunity to collaborate by working with the linguists of CREA.

## 2. Aim of Sociolinguistic Surveys

Each linguistic situation constitutes a complex body of speech forms. While it is possible for each speech form to be developed it is necessary, for the sake of economy, to determine which speech forms can be grouped together so that fewer need to be developed. Often, people of slightly differing speech forms are able to use the same written material. This brings us to the question which sociolinguistic surveys seek to answer: Can the speech forms A and B use the same literature, or do two different

sets of literature need to be developed?». Thus, we see that the work of survey seeks to discover the complex realities of language in its local context.

In order to gain the needed understanding of the particular complexities in an existing situation, a survey focuses on three domains: dialectology, bilingualism and extra linguistic issues. The first domain is dialectology which essentially seeks to determine the boundaries of a given language and its dialects. Three areas for study are considered in order to gain clear indications concerning this linguistic relationship. The first is to look at the phonological relationship which exists at the word level, comparing the degree of lexical similarity shared between two or more speech forms. The second area deals with the amount of information which can pass between two speech forms independent of an interpreter, thus seeking to measure the level of intelligibility existing between the forms. The third area focuses on what the speakers themselves see as the relationship which exists between the speech forms, seeking in this way to determine what are recognized as the boundaries of a given language.

The second domain of study deals with bilingualism, which we define as acquired intelligibility gained through contact with another speech form as opposed to intelligibility that is inherent due to genetic linguistic similarities. The importance of this study is obvious when considering actual language use in developing countries. The reality is that one can rarely speak of monolingual societies, but rather of multilingual societies where a speaker may be able to communicate in several different languages. A measurement must be made as to a person's ability in speech forms which hold the potential for expanding the individual's integration into the larger societal structure. These speech forms are often already developed or are in the process of being developed.

The third domain concerns itself with extra linguistic issues of language. This domain is as vast as the culture in which the language is spoken, because it deals with all the complex interactions of language and culture in a given society. Most importantly, it seeks to determine the attitudes which the speakers of each speech form have with regard to the mother-tongue and surrounding speech forms, as well as considering the patterns of language use which in turn will reflect on the vitality of a given speech form.

### 3. Methods Used in Survey, <The Tools of the Trade>

#### Lexicostatistics

This is a primary tool in dialectology and often the first step in a language survey, being the most basic method at our disposal to determine linguistic similarity between speech forms. Once the wordlists of each of the speech forms included in the study have been collected, the distance between them can be measured by deciding which words are related and then carrying out statistical calculations based on those decisions. SIL in Cameroon

uses the standard ALCAM wordlist of 120 items which are considered to represent a core vocabulary that would contain few if any borrowed/loan words. The cognate decisions are made using the inspection method that considers the phonetic similarity between words, taking a synchronic approach in analysis which considers the speech form in its present context. In this age of computers, two programs are available to aid in analysis. Both work only from the decisions that the linguist has already made, but the program *Wordsurv* also has a phonostatistic component which aids in confirming the reasonableness of the decisions made, therefore giving a check of the work accomplished. Another benefit of the *Wordsurv* program is that once the data is entered, the wordlist becomes a computerized database. *Lexistat* is the second program. The advantage of this is that it can generate tree diagrams showing the percentage of cognicity between speech forms. Another advantage is that it is relatively quick and easy to apply as compared to the more elaborate *Wordsurv* program.

Deciding whether two words are indeed related is sometimes difficult (even with the assistance of the phonostatistic component of *Wordsurv*). Because further testing is always done if the level of lexical similarity exceeds a standard threshold, it is advantageous in cases of doubt to count a pair as lexically related. To err (if indeed it is an error) in this positive direction means that doubtful cases will be tested further. The threshold is determined by empirical values discovered as a result of many years of application of the method. SIL considers two speech forms to be possible dialects of the same language if the percentage of cognates between them is 70% or greater. Below 70%, two speech forms are considered to be two distinct languages. Since a cognate percentage based on lists of only 120 words per speech form may differ slightly from the actual lexical similarity and since potential errors have to be considered, a range of error is calculated. The actual upper limit of the range of error is the value which must be compared with the empirical value of 70%.

#### Recorded Text Testing (RTT)

If the empirical value of 70% or more is met for lexical similarity, the relationship between the speech forms is further examined by carrying out dialect intelligibility testing or RTT. This method allows us to measure the amount of information that can pass between two speakers of the speech forms being tested, without the aid of an interpreter. This test method was developed during the 1950s, was later perfected in Latin America and is presently used by SIL around the world. Eugene Casad outlined the main principles of the method in Dialect Intelligibility Testing which is the standard for the RTT that SIL does.

Where two speech forms are considered possible dialects of the same language, stories are recorded in each form. These stories must be personal in nature, dealing with actual experiences lived by the storyteller, and not folklore or history. Neither should they deal too directly with a culturally-charged topic such as a

story dealing with a viper, because the outcome of such a story can already be guessed by what is known concerning vipers. For each speech form to be tested, two texts should be recorded; one of a minimum length of one minute and the other about three minutes. Once recorded, a literal written translation into a language known by the surveyor is made and from this translation, questions are developed. The questions are to be interspersed throughout the story as evenly as possible and should be varied in form so as to test different domains of comprehension. Attention should be given to avoid questions whose answer would be a loan word or would have a simple <yes> or <no> answer. Also, the questions must be given on information that is explicitly stated in the story and not on implied information.

These questions are placed in the story immediately following the sentence/segment in which the answer is found. Verification of the developed test tapes is next in the process of preparing for testing comprehension. This is done by testing the tapes with about five speakers of the language, in order to see what their response to the questions would be. This allows us to determine what answer to expect when it is tested in other locations. This procedure usually reveals one or two questions which must be omitted from the final test tape because of poor quality or faulty placement. The questions are then translated into each dialect where it will be tested so that the testee will always be able to understand them.

Once the above procedure is carried out, testing can begin. First, speakers in the region of the dominant dialect are tested for their comprehension of the subordinate dialect. The identification of dominant vs. subordinate is arbitrary when there is no real difference in dominance between the speech forms, but in some situations there is a very clear distinction. The reason for this procedure is that we are only intending to test inherent intelligibility using RTT, and not intelligibility attributed to contact or acquired intelligibility. Since the dominant is less likely to acquire the subordinate dialect than vice versa, a truer measure of inherent intelligibility is obtained. In keeping with the desire to test inherent intelligibility, the present strategy for the administration of RTT is to test children's comprehension of the surrounding dialects, assuming that their contact with these speech forms will be minimal, and thus the comprehension would more likely be inherent. An introduction tape consisting of a translated introduction text and the short story is the first tape that is listened to by the testee. It serves as an orientation to the whole test procedure. The second tape listened to is called the <Hometown>, being the story in the language of the testee which is the control for the test. It also is the control of the testee by showing whether he is capable of taking the test. Once the Hometown tape shows that the testee is capable, he will listen to up to three other tapes of stories in neighbouring dialects which actually test intelligibility.

The minimum sample for RTT is ten individuals from each speech form. The individual scores are added up and then the mean is

calculated. The standard deviation of all the scores for a tested speech form is also calculated. This measurement is really a tabulation of the spread of the scores which may indicate whether the scores show inherent or acquired intelligibility. If the standard deviation is above the empirical value of 12-15%, it is believed that the high scores represent a learning process and actual inherent intelligibility is somewhat lower. Each response is scored as right, wrong, or half-right. Once the calculations are made, the end result is a percentage of intelligibility between speech forms which gives a closer indication of linguistic similarity. If the percentage is from 75-85%, we consider this in the critical range of comprehension, and other factors must be considered if there is to be any unified language development. Eighty-five percent and above is considered adequate comprehension for the use of a common literature. Intelligibility below 75% is generally considered inadequate, but in any case other sociolinguistic factors must be examined.

### Questionnaires

In the beginning stages of the present effort of survey work in Cameroon, linguists from CREA and SIL worked together in standardizing the questionnaire format, and over time this has been modified and refined as experience was gained. There are now three primary questionnaires used to collect sociolinguistic data and gain insight into those areas of less accessible information, being embedded in layers of culture. The information that is collected tends to be either very general or very specific in nature. Questionnaire #1 gives a general profile of a particular speech form in its local context, asking for demographic information etc. It can be used for preliminary research, as well as being used in contact with local administration, churches, village leaders and so on. Questionnaire #2, often filled out in a group interview setting, seeks more specific information which defines the actual geographic boundaries of the language and helps to identify its dialects. There is also a Group Questionnaire which has been developed by Dr. Etienne Sadembouo, which uses the idea of variation or shades of colour to help the people understand the variations present in spoken language. The main focus of the questionnaire is to allow the people to indicate what they would consider to be the dialect of reference within their language. Questionnaire #3 is primarily an individual questionnaire which seeks to gain insight into the specific attitudes and language-use habits of speakers, which in turn can shed light on the vitality of the language surveyed. This questionnaire is completed with a representative sample of people from each speech form being studied.

Another issue researched through the use of questionnaires is bilingualism. A self-evaluation questionnaire has been developed which asks a number of questions that have <yes> or <no> responses. The questions begin very simply, but they become more difficult as one progresses through the questions. The scale of the questionnaire is matched to the SLOPE test which is explained below; therefore as one progresses through the questions he moves

from one level to another (the levels go from 0+, someone who knows only 30 words, to 5, someone who speaks as if it were his mother-tongue). This method of measuring bilingualism is only decisive when the community demonstrates a relatively low level of bilingualism. If results of this questionnaire indicate that bilingualism is widespread and at a reasonable level of proficiency, then more extensive testing procedures must be used to determine more accurately the level of bilingualism.

#### Interview Test

SLOPE, Second Language Oral Proficiency Evaluation, is a <more objective> method for testing levels of bilingualism. It was developed by modifying the procedure used by the Foreign Service Institute (FSI) of the U.S. government which tests the oral proficiency of government employees. The SLOPE test was developed in Dakar, Senegal, in 1987, using an interview context with as many as four people involved. There are the tester, the testee, an assistant to the testee who is able to support the testee in his mother-tongue, and the linguist who is monitoring the interview. SLOPE considers five different aspects of an individual's proficiency in a second language such as: comprehension, discourse, structure, lexicalization, and fluency, not necessarily being given the same weight. The interview is well structured and consists of three parts: a warm-up phase, a phase in which the testee elicits information, and a phase in which the testee provides information. In each phase, the tester raises the level of difficulty of conversation which helps to reveal the testee's language proficiency.

At this stage of development, SLOPE is considered the most valid, accurate and precise means of measuring bilingual proficiency. The main disadvantages to the method are the time that it takes to administer the test and the special training which is required for the people administering the test.

#### 4. Decision Making in Survey, <How to do it>

The basic procedure that is followed by the SIL survey department in Cameroon concerning surveys can be seen in the flowchart that follows. The details of each method referred to in the chart have already been given, but it is this chart that gives clarified direction on the strategy.

SOCIOLINGUISTIC SURVEY

This basic procedure for a sociolinguistic survey, has the aim of determining whether the same literature could be used or whether separate literature is needed for speech forms A and B.

1. Linguistic differences: Are they big?

Generally speaking, big linguistic differences mean clear boundaries whereas small linguistic differences mean unclear boundaries.

Example: Clear boundaries between Ewondo and French, unclear boundaries between Ewondo and Bulu.

2. Cognicity?

Calculation of the degree of cognicity, applying lexicostatistics, based on standardized wordlists and with the aid of the computer.

3. Intelligibility greater than 75%?

Investigation of intelligibility using Recorded Text Testing (which was described by E. Casad).

Basic principle: A speaker of dialect A listens to a recorded story in dialect B. The story is interspersed with questions on the preceding text. (The questions are asked in dialect A, the mother tongue of the person being tested).

4. Bilingualism adequate?

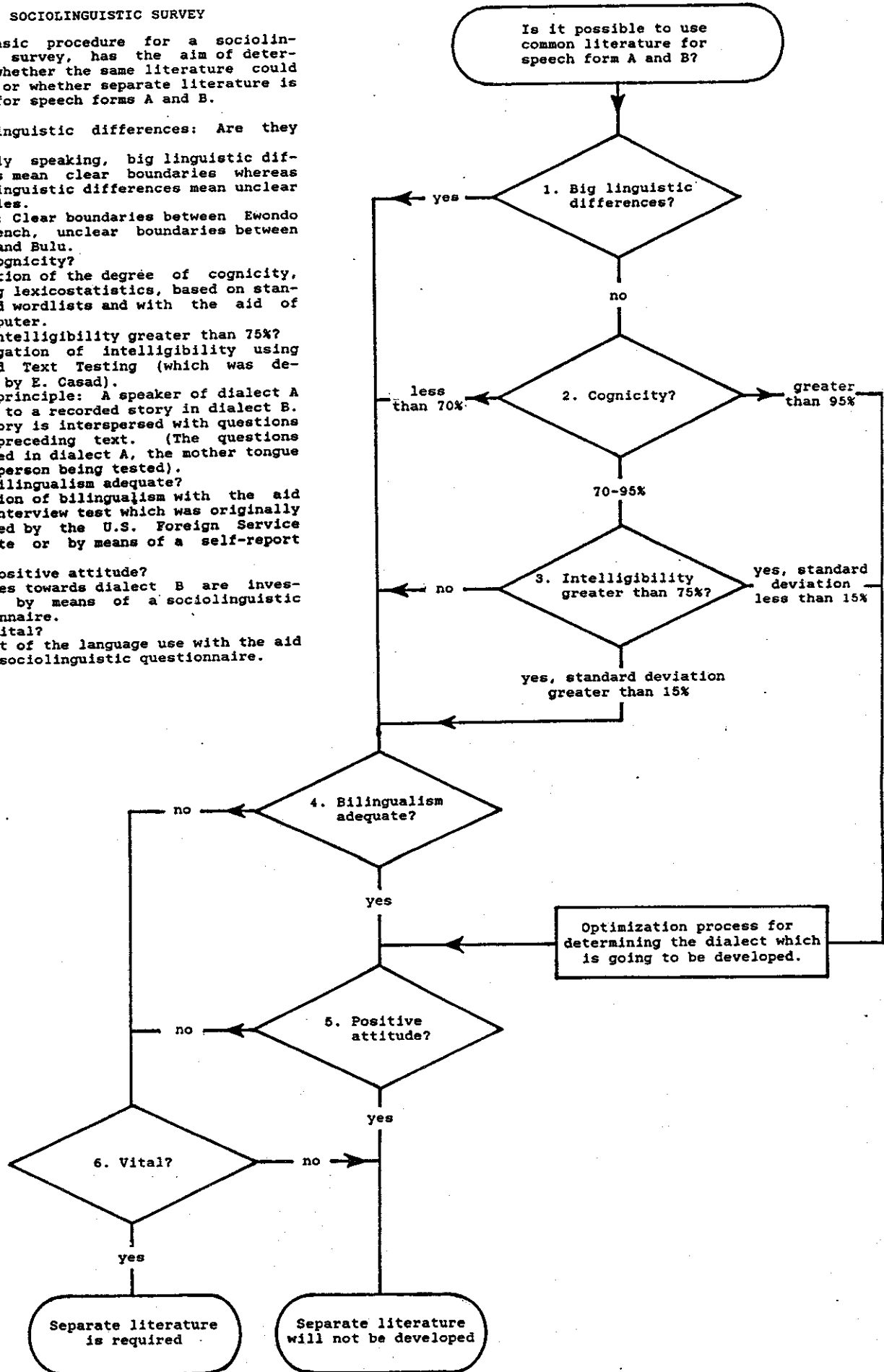
Evaluation of bilingualism with the aid of an interview test which was originally developed by the U.S. Foreign Service Institute or by means of a self-report test.

5. Positive attitude?

Attitudes towards dialect B are investigated by means of a sociolinguistic questionnaire.

6. Vital?

Judgment of the language use with the aid of the sociolinguistic questionnaire.





## 5. Closing Remarks

Having considered survey procedure and the tools at our disposal in the language assessment task, an evaluation of existing methods would be beneficial. There is satisfaction and confidence given to what we do with regard to dialectology. Lexicostatics, dealing with a language synchronically at the word level, provides an accurate estimation of similarity at that level. Intelligibility testing, still dealing with the language synchronically, has more variables that are not always easily controlled. While this increases the possibility for error, it does not invalidate the method and it is felt that if technique is properly followed, the measure is an accurate one. Group interviews which focus on clarifying the dialect situation are considered to be a very useful tool, and complete the domain of dialectology.

The two domains of bilingualism and extra linguistic issues have need of further development. We are confident that we are able to collect needed data using the tools at our disposal, but there is a need to refine and develop other methods and tools which will improve our confidence in the final results. In the area of extra linguistic issues, at present we try to answer basic questions concerning language attitude and viability. Sociolinguistic questionnaires are therefore being developed with regard to these basic questions. This is of vital importance to proper assessment of a language. On the one hand, there is a danger to the subjectiveness of questionnaires. This can be avoided to a degree by asking good questions and choosing a good sample. On the other hand, the opinions expressed by the local people are extremely valuable. Because of this, when considering <pure linguistic> factors versus sociolinguistic (extra linguistic) factors, such as attitude and vitality, preference is given to the latter.

As for bilingualism testing, there is a need for the development and application of good methods and tools to measure bilingualism in African contexts. In addition to SLOPE, a new method called the Sentence Repetition Test (SRT) is being considered for further development. The benefit of this method is that one can test a very large sample since the test's administration is simple and rapid, thus allowing for good sampling of a given population. Finally, testing comprehension of existing literature such as Scripture is considered an area which needs to be developed. This involves the use of written texts with different discourse styles.

The need to know how one speech form relates to another in present-day multilingual society is very real. Language assessment, as described in this paper, is a means to acquire the desired and required knowledge that in turn affects decisions concerning the development of a given language entity. The Société Internationale de Linguistique in Cameroon seeks to do an effective job in the analysis of language situations where development is a possibility, and therefore encourages the work done by its Survey Department. For further information and answers to

questions that this paper may have brought up, one should not hesitate to contact the Survey Department office located on Mbalmayo Road in Yaoundé at the SIL Training Center.

## BIBLIOGRAPHY

- Bergman, Ted G., 1989. Survey Reference Manual: A Collection of Papers on the Assessment of Bible Translation Need. Dallas: Summer Institute of Linguistics.
- Casad, Eugene H., 1974. Dialect Intelligibility Testing. Summer Institute of Linguistics Publications in Linguistics and Related Fields, 38. Norman Oklahoma: Summer Institute of Linguistics of the University of Oklahoma.
- Grimes, Barbara F. 1987. <How Bilingual is Bilingual?>. Notes on Linguistics, #40A, pp.34-54.
- Schadeberg, Thilo C., 1987. Lexistat, version 2.0. Afrikaanse Taalkunde, Rijksuniversiteit de Leiden.
- Simons, Gary, 1977. <Tables of Significance for Lexicostatistics.> In Loving & Simons, eds. Workpapers in Papua New Guinea Languages. Volume 21, pp.75-106.
- Summer Institute of Linguistics. 1987. <The SIL Second Language Oral Proficiency Evaluation>. Notes on Linguistics, #40A, pp.3-33.
- Wimbish, John S. 1987. WORDSURV: A Survey Word List Analysis Program. Dallas: SIL.