

**Summer Institute of Linguistics and  
The University of Texas at Arlington  
Publications in Linguistics**

Publication 110

**Editors**

**Donald A. Burquest  
University of Texas  
at Arlington**

**William R. Merrifield  
Summer Institute of  
Linguistics**

**Assistant Editors**

**Rhonda L. Hartell**

**Marilyn A. Mayers**

**Consulting Editors**

**Doris A. Bartholomew  
Pamela M. Bendor-Samuel  
Desmond C. Derbyshire  
Robert A. Dooley  
Jerold A. Edmondson**

**Austin Hale  
Robert E. Longacre  
Eugene E. Loos  
Kenneth L. Pike  
Viola G. Waterhouse**

# **Windows on Bilingualism**

**Eugene H. Casad**

*Editor*

**A Publication of  
The Summer Institute of Linguistics  
and  
The University of Texas at Arlington  
1992**

© 1992 by the Summer Institute of Linguistics, Inc.

Library of Congress Catalog No: 92-81102

ISBN: 0-88312-809-8

ISSN: 1040-0850

All Rights Reserved

No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form or by any means—electronic, mechanical, photocopy, recording, or otherwise—without the express permission of the Summer Institute of Linguistics, with the exception of brief excerpts in journal articles or reviews.

Cover design by Hazel Shorey

Copies of this and other publications of the Summer Institute of Linguistics may be obtained from

International Academic Bookstore  
7500 W. Camp Wisdom Road  
Dallas, TX 75236



# The Asia Area Survey Conference

Eugene H. Casad

The Asia Area Conference of the Summer Institute of Linguistics (SIL) on Survey Data Collecting and Interpreting was held at the Green Valley Country Club of Baguio City, Luzon, the Philippines on October 12-13, 1987. This conference was of primary importance both for the Asia Area's ongoing research and for SIL's goals in the area of language surveys. In addition, it was a broadening of my own exposure to the range of situations under which surveys are being carried out, as well as an opportunity to interact with a number of my Asia Area colleagues for the first time. I was not only able to contribute somewhat to the conference, but was asked to edit a subset of the papers given at this conference. All parties concerned with various aspects of language assessment could thus share the results of several years of intensive survey work in the Asia area, done under the capable guidance of Calvin Rensch and Carolyn Rensch.

The first day of the conference was devoted to the topic of bilingualism. Randy Kamp began with a paper titled *Bilingualism Testing in the Philippines* (in this volume titled *Inherent Intelligibility, Bilingualism, or Both?*) in which he detailed a comparison of four test instruments used during a survey of the Karao people of Benquet province, Luzon Island.

The test instruments were: a proficiency interview, a self-score evaluation, a self-test questionnaire, and a set of taped comprehension tests. The self-score test required subjects to rate their own ability in the languages that they spoke. This rating was in terms of a five-point scale. The self-test questionnaire consisted of a set of questions on the ways the subject claimed to be able to use a given language. These questions are grouped around a set of various levels of difficulty that are assumed to attach to

particular tasks. The proficiency interview followed the procedures in the *Second Language Oral Proficiency Evaluation* (SLOPE) syllabus, a work by Barbara Grimes (1987c) and others to adapt for SIL's use an Educational Testing Service approach to evaluating bilingual proficiency. The taped comprehension tests included a HOMETOWN test tape in Karao and two Ibaloi test tapes.

Kamp's conclusions are very significant. For one, he finds that he cannot always separate bilingualism from inherent intelligibility. He also observes statistically significant correlations between self-score evaluations and interview scores, and between the self-test questionnaire and the interviews. The strength and reliability of these correlations, however, is such as to require separate tests for intelligibility and bilingualism. Finally, he concludes that narrative tests for comprehension probably test no higher than an FSI level 3. This latter conclusion, however, cannot be generalized beyond the Karao survey—cf. the results of the recent study by James, Masland and Rand (1989).

In *Surveying Language Proficiency*, Steve Quackenbush describes two methods he used in a survey among the Agutaynen people of Palawan, Philippines. One was a self-reporting questionnaire which required subjects to give YES or NO responses to questions designed to sample language skills of varying difficulty. The other was an oral proficiency interview modeled on the approach of the Foreign Service Institute and the Educational Testing Service. Quackenbush found a fifty-six percent correlation between the two methods, which compares favorably with the findings of other researchers, including Kamp.

Quackenbush appropriately concludes that bilingual evaluation is not an exact science and that proficiency data must be construed as an approximation of the true state of affairs. The use of one instrument or another depends on the purpose and extent of a survey. Quackenbush prefers the direct test over the self-report interview, but retains both instruments for particular situations. Finally, he concludes that the FSI concept of levels of proficiency provides a usable, valid, and meaningful standard for evaluating proficiency.

Since Barbara Grimes was unable to attend, I presented material from the SLOPE syllabus which was prepared by Barbara and others in cooperation with Dr. Thea C. Brun, head of the language testing unit of the Foreign Service Institute. This had been presented at a workshop held in Dakar, Senegal, in April of 1987.

SLOPE is based on the assumption that bilingualism is most accurately modeled as interactive behavior. This view largely determines the form, content and mode of scoring of the test. A brief description of SLOPE, by Barbara Grimes, is included in this volume which should prove useful for

those who would like to know what the main outlines of the test are, without going into all the detail contained in the syllabus (B. Grimes 1987c).

The final paper on bilingualism was a paper by Carla Radloff, *The Sentence Repetition Test*, describing ongoing research in developing a sentence repetition test for evaluating bilingual proficiency. This paper represents work by a team consisting of Radloff, David Marshall, Charles Meeker, and several Pakistani evaluators. The overall project displays a methodological soundness, concern for reliability and validity, and meticulousness second to none in the field.

Radloff's paper was originally intended to be Part I of this volume. However, revisions and additional testing in the field led to a longer version which has been published as a separate monograph (Radloff 1991). In this volume, therefore, I confine myself to the following summary of the approach.

The sentence repetition test consists of a set of fifteen tape-recorded sentences which are played back to subjects individually through earphones. These sentences are determined beforehand to be maximally discriminating for particular levels of difficulty, and are not related semantically. The subject is expected to repeat each sentence verbally. He is scored as follows: perfect score on a sentence is 3; one error on a sentence gives him a score of 2; and two errors on a sentence gives him a score of 1. Three or more errors result in a 0.

The test is said to be a correlated one, i.e., one which takes its meaning from its relation to an independent instrument such as SLOPE or some other evaluation measure. Its function is that of screening for people who are of different proficiency levels.

Radloff has taken special care to tie the team's work in with other kinds of research. They place heavy emphasis on training their testers in order to guarantee the reliability of the results. They found no significant differences between the way both Urdu and Pashto testers scored subjects. Neither the age of the subject nor his level of education had any significant effect on his scores. However, whether a subject had been introduced to the test administrator by an acquaintance or relative, or had merely been encountered among bystanders did make a significant difference in the subject's scores.

Radloff points out that the Sentence Repetition Test (SRT) does not appear to distinguish between speakers who are at FSI levels of 3+ or higher in the evaluation language. In more recent field experimentation, they have tried to modify the SRT, making it more discriminating at higher levels of proficiency in a second language. The results and generalizations from those results are not yet known. There may well be an intrinsic

limitation on the discriminability of the method, as the definitions of the FSI levels themselves suggest. I do not see this limitation as sufficient grounds for rejecting the entire approach; rather, the SRT and SLOPE can be used to complement one another. In most cases, a community will probably turn out to be below 3+ in the evaluation language. The SRT could pick this up readily. SLOPE could then be employed for the cases in which the SRT reports a very high level of proficiency.

The Radloff team took consistent and meticulous care in its approach to do things properly. The steps they followed include, among other things, developing the test in the area where the evaluation language is spoken, selecting and training the personnel involved in the study, constructing and testing a long form of the test sentences (40–50), modifying the long form into a short 15-sentence version, developing an index of the discriminatory power of the sentences, and rating the participants by an external standard. The fact that the SRT is easy to administer once it is developed is an added bonus.

To summarize the first day of the seminar, we use Rensch's metaphor that there are "several windows on bilingualism."

The second day of the seminar was devoted to several different topics. It began with a paper by Francis A. Gray of Far Eastern Broadcasting Corporation (FEBC). Titled *The Use of Language Data in Broadcast Research and Project Development*, it was an informative presentation of both the research being carried out by FEBC, the role played by information gathered by other organizations, the nature of RICE (Radio In Church-planting Evangelism), the "World By 2000" declaration, and the scope of Christian broadcasting encompassed by Gray's research.

It is particularly interesting that the development of the personal computer was crucial to making Gray's project feasible. In addition, Gray has found *The Ethnologue* (B. Grimes 1984a) immensely useful and has benefited further from personal interaction with Joseph Grimes and Barbara Grimes.

The next paper was by Mark Taber. His *Survey: A Picture of Maluku* was an informative presentation of what had been done in Maluku up to 1987, the approaches employed in collecting survey data, the kinds of data collected thus far, and a summary of problems in training survey technicians.

Interspersed throughout the second day of the conference were three short, but important presentations by Cal Rensch, which I summarize here as a single block. The first paper, *Calculating Lexical Similarity*, mentions a few considerations in collecting word lists. It makes the point that the seemingly EASY tool of taking a word list in a language one does not know is not so simple after all. One reason is that phonologically similar words may be similar for several different reasons, e.g., they may either be related

historically or they may all be related due to borrowing from another language.

Rensch points out that, for our purposes, we do not need to apply the rigorous procedures of historical linguistics to our data. Nevertheless, I suggest that, with the accessibility of the personal computer and the development of programs such as John Wimbish's WordSurv and Don Frantz' Compass G (1970), which Joe Grimes is incorporating into a menu-driven family of programs, it is becoming feasible to describe historical patterns as part of our normal analysis of survey data. Even though some colleagues have mentioned that they do not ordinarily record word lists on tape, I would like to suggest that they do so, especially in view of the tentative state of our investigators' knowledge of the language.

Interestingly enough, the Asia area survey teams have found that they need to take two different word lists from different individuals. This frequently shows from five to ten percent of differences due to sampling error. The first step in calculating lexical similarity is to regularize the discrepancies. The actual calculation is based on applying a set of criteria to successive pairs of words. Rensch devised this set of criteria as a control for variability in researcher evaluations of lexical similarity. No hard and fast decisions are made with respect to words that fail to make the criterion. Such words could still be related historically. The purpose of the calculation at this preliminary stage is simply to identify the most obvious sound correspondences.

Rensch's second paper is called *Sociolinguistic Community Profiles*. Here he addresses the problems we fall into by trying to evaluate characteristics which are not uniformly distributed throughout a population. His observation is that the members of a society who are educated and travel frequently are more proficient in a given second language than their less fortunate fellow citizens. Rensch would like to identify the specific factors related to this greater proficiency. Assuming that a multiplicity of factors underlies second language proficiency, he notes that the size of the subgroups that are associated with these factors varies from factor to factor. His solution is to conduct a census in which a representative of each household is asked about a set of twenty categories outlined by Frank Blair in his *Survey on a Shoestring* (1990).

These data, then, can be used to construct a community profile based on age groups, sex, education level, etc. Profiles of different communities can then be compared, forming the basis for useful hypotheses and arriving at believable conclusions.

Rensch's third paper is a squib titled *Language Proficiency*. Because multilingualism is so widespread in Asia, and because extensive testing for bilingual proficiency is time consuming and costly, the Asia area team is

looking for diagnostic patterns of language use which show there is no need for extensive bilingualism testing.

The morning session ended with the presentation of my paper that appears in this volume and that was also presented at the International Language Assessment Conference held in Horsley's Green at the end of May 1989 (Casad 1990). It is titled *State of the Art: Dialect Survey Fifteen Years Later*.

The Tuesday afternoon session was devoted to two papers on language attitudes. The first, *Language Attitude Test in a Multilingual Setting*, was presented by Ronald Krueger. This test was developed as a course requirement for the SIL course, *Sociolinguistic Surveys*, which was taught by Joe and Barbara Grimes at the University of Oklahoma SIL and has more recently been offered at the Texas SIL in Dallas. The test that Ronald and Joanne Krueger designed was intended to sample several areas of interest. These included the domains in which Hindi, Urdu, Gujarati, and English were used, the propriety of the usage of each of these languages in the presence of nonspeakers, the attitudes of people toward each of the languages in question, and their attitude toward the peoples who spoke these languages.

Krueger made a few significant generalizations from his study. For one, language attitudes do not have just simple plus and minus values, but rather reflect a whole range of strengths and meanings. Furthermore, attitudes are not uniform throughout a culture. This means that one must sample carefully and thoroughly if he is going to draw valid conclusions. Finally, controls can be built into an attitude test by including two forms of each question at different points in the test which allowed them to judge the consistency of the subjects' responses. The test consisted of thirty questions, some of which had several parts. Finally, it took about one-and-a-half hours to administer. Although Kreuger did not suggest this test as a model for others, Joe and Barbara Grimes and I have already used it this way in the SIL survey courses.

Rensch stated that in the Asia area surveys they had noticed two patterns of language attitudes. The first pattern is based on the idea of LIMITED GOOD. In this view, people tend to think that if one of the various languages they speak is good, then the other is bad. The second pattern is based on the idea of TOTAL GOOD. All the languages in the multilingual setting are held to be good; one serves for one purpose, the other serves for another. Needless to say, these two situations involve two different strategies for establishing and carrying out field programs.

The last paper in the conference was Roland Walker's *Towards a Model For Predicting The Acceptance of Vernacular Literacy by Minority-Language Groups*. Walker attempted to identify those sociolinguistic variables that

best predict if a given vernacular language group will accept literacy materials published in the mother-tongue. Walker's exploratory study cannot be taken to be the formulation of a strongly predictive model of the acceptance of literature by vernacular language groups. What first study of any sociolinguistic topic could turn out that way? Nevertheless, it suggested that certain constellations of variables may hinder the vernacular literature acceptance of vernacular literature within particular communities. One cannot help but ask whether there are not other constellations of variables that enhance such acceptance. Surely we would want to know about this, too. Walker did not address this question so directly.

In summary, the papers presented at the Baguio conference relate to important topics and reflect careful work on the part of well-trained people. There are several implications to be drawn from all this. Teamwork is one of the most salient—dedicated people working toward a common goal. Sound methodology—there is a consistent and concerted effort to develop test instruments reliable test instruments and validate them. These people are using fairly sophisticated statistical methods, but also know when to opt for their intuitions instead of being misled by the numbers. Realism—they recognize the complexity of the task, deem it doable, but know that it cannot be done overnight or even in three or four years. Continuity—rather than throw overboard everything else that previous researchers have done because of their exhilaration over the development of new research in the areas of bilingualism, language use and attitudes, they have taken the best from comparative work, lexicostatistics, and intelligibility testing and have melded it into a coherent program. In short, the members of the Asia area survey team have not only validated what their predecessors had developed in Mexico, but have gone beyond that, teaching us new and useful things.

In conclusion, we hope that this volume, based on the papers given at the Baguio conference and on the work done by the Asia area survey, will benefit others interested in language assessment.

Beyond this volume, I need to mention the fine paper written by Paul Kroeger (1986), called *Intellegibility Patterns in Sabah and the Problem of Prediction*, which was published in *FOCAL I: Papers from the Fourth International Conference on Austronesian Linguistics*; Paul Geraghty, Lois Carrington and S. A. Wurm, eds. This paper represents the finest statistical treatment of survey data that I have seen to date. Finally, Frank Blair (1990) has written a survey handbook called *Survey on a Shoestring*. Following its introductory chapter that gives definitions and outlines the scope of the work, chapters two through six treat the successive topics of survey planning, dialect areas, sampling, bilingualism, and oral proficiency testing. The second half of *Survey on a Shoestring* discusses additional

aspects of the assessment task, i.e., recorded text tests, observation, sentence repetition tests, self-evaluation questionnaires, and language use and language attitudes. This is a survey handbook in the full sense of the word and could easily be used in countries outside of the Asia area.