

## The Spoken Language Library: Filling A Gap

At NSA we have what may be the finest general-purpose linguistic library in the world. It contains at least 25,000 volumes, and its coverage of the world's languages is truly global, for more than 500 languages and dialects are represented in the collection. Many of its periodicals are in foreign languages and all but one or two deal exclusively with language and linguistics. Additional works on linguistic subjects are being acquired even after more than twenty-five years of careful purchases, and the value of the collection is enhanced by the presence of some volumes that were in scarce supply when we acquired them and have now become collectors' treasures.

But if one's interest in a language is in how it *sounds*, the library's learned treatises do not help much. Even a book that is deliberately directed toward the treatment of the spoken version of a language is incapable of describing with any accuracy how the language sounds. True, the library also contains quite a few cassettes. The FLACS (Foreign Language Cassette Series), an inspired contribution from Harry Rashbaum, is an important aid in language training and skill maintenance. The cassettes contain recorded broadcasts from [redacted]

[redacted] and there are also copies of the Voice of America broadcasts to serve the same training function.

Radio broadcasts, however, are a poor source if the criterion is how people really speak to each other, for something always happens to the voice when we concern ourselves with careful enunciation and projection, as is required (or should be) in broadcasting. Another less-than-satisfactory example of a language may be found in recordings made for training purposes, if they exist for the language in question. Here again the pace and the enunciation are altered, this time for the benefit of the student. No one really talks that way. If one wishes to hear [redacted] or any other language as it is actually spoken, it

is usually necessary to search out a native speaker and to convince him that it would be in his interest to talk.

There have been attempts in the past to capture examples of natural speech on some type of permanent record. A great deal of transcription was done using the International Phonetic Alphabet, an improvement over the standard Latin letters, but not nearly precise enough for most purposes. Cultural anthropologists, historians, and linguists are constantly on the lookout for speech types that are disappearing. The last speaker of Cornish, the language of Cornwall, died close to the beginning of the century. Other languages are disappearing for a variety of reasons, and there are a number of separate efforts, academic and governmental, to record them before they disappear entirely. Favorite subjects are groups like the Amazonian Indians, whose numbers are shrinking with each passing year, and American Indian tribes, whose speech is becoming increasingly English.

[redacted] Unlike most other problems involving unfamiliar elements, we have not had any satisfactory way of consulting appropriate reference materials.

An attempt was made several years ago in an NSA-related language research activity to record examples of languages as they are spoken around the world, but it was neither well conceived nor carefully controlled. Speakers claiming knowledge of various languages were hired to record examples of those languages. They recited stories and poems or quoted from the Gospel

or the Koran. Just think of what happens to your speech when you recite "The boy stood on the burning deck," or when you intone a passage from Genesis. Besides, if the speaker claimed several dialects of a language, he was encouraged to record them all. More often than not, his command of dialects other than that spoken in childhood would be less than satisfactory. The results of all these efforts are, unfortunately, of little value.

What is needed is a kind of snapshot of how people actually use their language in their daily lives, in conversation, in the country of interest. The speech should be natural, free-flowing and unselfconscious.

[redacted] it appears to be more important that the speech be authentic and representative of the people and their culture. Ideally, someone with a tape recorder should traverse each country, capturing examples of speech to provide a complete picture of the social, economic, regional, educational, and occupational variations for each language and dialect. While such an encyclopedic approach would certainly be sound academically, it just wouldn't be defensible as an investment for NSA.

What we need is an information resource designed both to fill a gap in our knowledge and yet practical enough to justify the necessary effort and investment.

[redacted] we should have a set of conversations, reasonably hi-fi in quality, accompanied by transcriptions and translations. And the building of such a library of recordings is under way and has been since early in 1980.

How many languages will the library contain? Dr. Ladefoged, head of the Phonetics Laboratory at UCLA, estimates that there are about 8,000 spoken languages in the world. But that number may easily be challenged, since no one really knows how to count languages. If we were recording English, for example, would we consider the speech of a Vermonter and a Georgian as the same, or should we record them separately and label one American English (New England) and the other American English (Deep South)? It's all very well to say that if dialects are mutually intelligible, they should be classed together, and if not, they should be regarded as distinct languages. Intelligible to whom? [redacted]



In spite of these uncertainties, the collecting of tapes, transcripts, and translations that will make up the Spoken Language Library is under way. Using the services of a contractor to recruit and pay the speakers, we have made about 30 recordings as of this writing. The ground rules call for two speakers, both of whom must have the language as a mother tongue, both preferably only recently out of their native land to avoid the inevitably contaminating influence of English. Ideally we should record as many as five different pairs of speakers to diminish the influence of individual speech mannerisms and to allow male-male, male-female, and female-female pairs of speakers. In most societies these pairings will elicit quite different speech patterns.

The subjects are placed in separate rooms and converse for ten to twelve minutes over a specially-constructed telephone-microphone setup that permits them to speak to each other and simultaneously to record on a two-channel tape recorder. By eliminating gestures and other nonverbal signals characteristic of face-to-face conversations, and by limiting audio contact to that which is transmitted by the telephone, we are able to capture the entire communication on tape. Also, whatever shyness or self-consciousness the speakers might experience by talking into a microphone appears to be quickly dispelled by the comfortable familiarity of the telephone, a device that seems to provide a natural setting, even a stimulus, for free-flowing speech. After the purpose of the project is explained and the subjects dismiss their lingering suspicions about the possibly harmful effects of allowing themselves to be recorded, most throw themselves into the project with enthusiasm and pride, and the conversation flows after perhaps one or two minutes of hesitation while topics are raised and discarded.

Occasionally the process goes less smoothly, as when one subject, [redacted] stationed at the [redacted] was reluctant to record via the telephone. He said he would prefer a face-to-face discussion with his conversation partner, a young and pretty [redacted] student from George Washington University. (Perhaps I was wrong in assuming that he was suspicious of the phones—she was good to look at.) But after I pointed out to him that the phones didn't go anywhere except between the speakers, and that being in separate rooms meant that all the communication had to be of the spoken, and therefore

UNCLASSIFIED

recordable, variety, he relented and we went on with the recording.

In another instance we were not so fortunate. One of a pair of [redacted] speakers decided, after he had listened to the sales pitch, that he would "rather not leave his voiceprint behind," whatever that meant. So we rescheduled the session when we found another, more cooperative [redacted] subject.

Sometimes the subject can be too cooperative. One [redacted] pair, a husband-wife combination, both with PhDs in literature, tried to be more helpful than I wanted them to be. First, the husband asked me if I wanted him to speak on [redacted] history, or current economic conditions, or some other topic on which he felt well informed. No, I told him several times, I wanted the conversation to be natural and not inhibited as to subject matter. What I got on tape is quite satisfactory, although the conversation did begin with a mini-drama consisting of lines like "Oh, I didn't recognize your voice," and "I haven't heard from you for a long time," and more of such rubbish.

So far our recording activities have been limited to the local area, using foreign students from the nearby universities and other likely subjects. But we plan to extend our search for appropriate subjects to the West Coast, where there are not only many foreign students

but immigrant communities as well, affording us the opportunity to broaden the social-economic-educational sampling of the recordings, which has been strongly biased toward the upper classes and the more highly educated segment of each country's population.

The Spoken Language Library is and will be unclassified. It is conceivable that some academic institution may prefer to see the Library complete, and will arrange some day to add such items as Eskimo languages, regional American dialects, languages spoken by fewer than fifty people, etc. The recording of these languages by us would surely qualify NSA for Senator Proxmire's Golden Fleece Award, but interest in the subject transcends [redacted]

[redacted] and it is possible that someday there will be copies of the Spoken Language Library in institutions of higher learning in this country and abroad.

[redacted] (b) (3)-P.L. 86-36 working in R5 as a reemployed annuitant in the field of speech research. He contributes frequently to this and other Agency publications and is a strong advocate of more and better information exchange within and among Agency elements.

- (b) (1)
- (b) (3)-50 USC 403
- (b) (3)-18 USC 798
- (b) (3)-P.L. 86-36