

## PART I

# GENERAL INTRODUCTION AND THEORY OF FOREIGN LANGUAGE TESTING

## Chapter I

### LANGUAGE

#### **I.I INTERNATIONAL COMMUNICATION AND THE STUDY OF LANGUAGES**

The great advances in transportation and communication made by man have brought home to all but the most isolated of men the value of world languages for international communication. They have also shown the value of local languages for effective communication with the native speakers of these languages.

In previous centuries it was reasonable and proper to study languages exclusively for the purpose of reading their literatures. The jet airliner that spans the oceans in a matter of hours and flies regularly over the North Pole, radio, television, and the telephone, and safer and more comfortable ships, trains and automobiles have multiplied travel and international communication a thousandfold and have made the study of languages for basic communication with native speakers a mark of the twentieth century.

The speed and frequency of international communication have outstripped the speed of teaching and learning languages and demand more effective methods of teaching. The systematic investigation of science is needed to advance the teaching of languages.

With the need for more effective teaching of languages goes the need for more effective testing of their use. Language testing also requires the advances and methods of science to meet present and future demands. This book is a comprehensive discussion of foreign language testing, bringing together modern experience and scientific knowledge.

## 1.2 LANGUAGE AS A SYSTEM OF COMMUNICATION

Since language is so ever present in human activity and thinking and since it flows so easily from the tongue of native speakers, it is possible on one hand to oversimplify it in our thinking and in our plans to study it. Thus we often hear people who, upon meeting a native speaker of a different language, naively ask to be taught that language and wait right then for the process to begin. For the same reason we see people buying books that promise to teach a language with great ease even without a teacher.

Language is more than the apparently simple stream of sound that flows from the tongue of the native speaker; it is more than the native speaker thinks it is. It is a complex system of communication with various levels of complexity involving intricate selection and ordering of meanings, sounds, and larger units and arrangements.

On the other hand, in spite of the fact that language uses the same sound waves as the noise that the physicist studies, and writing the same ink and paper that can be studied as chemical substances, these chemical and physical properties of the substance in which language is cast are not human language strictly speaking, and it is confusing to think of them as part of language. Their study often helps us understand language and thus they are of interest to the student of linguistics, but they are not language.

Language has been defined in various ways by linguists and others. Rather than attempting a full definition here, we will identify language by pointing out some of its most distinctive features. Language is primarily an instrument of communication among human beings in a community. A community that speaks the same language is a *speech community*. Languages differ from each other in such ways that the members of one

speech community usually do not understand the speakers of other speech communities. The fact that languages differ from each other is accepted as a normal state of affairs, and when the speakers of one community wish to communicate with those of another they usually study their language or find someone who knows it.

Certain uses of language are felt to be aesthetically more satisfying than others and are called literature in a strict sense. These artistic uses of language are not independent of the ordinary uses of language for basic communication; they are related aesthetically to them.

Language in its most common, pervasive, representative and apparently central manifestation involves oral-aural communication. We therefore say that language consists of oral-aural symbols of communication, arbitrary in their association to particular meanings and units and arbitrary in their particular shape for a given language.

Language as an instrument of communication among the members of a speech community who are also members of the same culture is best suited to convey the meanings current in that particular culture. These *cultural meanings* are roughly uniform for the members of the community and are thus readily conveyed and understood. Beyond these cultural meanings are *individual meanings* not readily communicated through standard forms of a language. Individual meanings require explanations, metaphors, analogies, and other indirect approaches if they are to be communicated through language. Some meanings and features of meaning become so frequent in their use in language or so attached to language distinctions that they can be described in terms of a language without recourse to full cultural reference. These can be called *linguistic meanings*.

Linguistic, cultural, and individual meanings are independent of the facts that science discovers or proves. They are in a sense facts in their own right, to be studied by science as such.

Languages are often represented by various systems of writing. Certain aspects of these systems of writing can be studied somewhat independently of the languages they represent, but their essential features are relevant insofar as they relate to a language.

Languages which are no longer spoken have to be studied through written records left to us, but the written records do not in themselves constitute the central part of language.

The study of language is necessary for the effective study of literature. The study of language, on the other hand, can proceed without the study of literature until it reaches the level of the study of artistic creation in a language.

### 1.3 THE MOST COMPLEX OF MAN'S TOOLS

When we use language even in an ordinary conversation we are wielding a most complicated tool with amazing dexterity and ease. We keep track of the general thread of our conversation, select whether to say something, ask something, tell our listener to do something, admire something; we choose the particular words to fit the situation, the listener, the type of sentence; we give proper endings to the words, place them in the proper place in the sentence; we connect each sentence and part of sentence with what has gone on before; within each word we select the proper sounds and order them in the proper sequence; we place the stress where it belongs; we modulate the tone of what we say to make the sense we intend, and we add variations to our voice to show our changing mood and attitude. We do all this while considering the effect of what we say upon our listener. In trying to please, or to attack, we think ahead to support what we want to prove or to counter what we do not approve. And this is done at the rate of several hundred movements of the articulators per minute.

Man as we know him is incapable of performing the complex operation of language at its normal speed without reducing most of it to habit. Habit carries most of the operation of the complex mechanism of language, leaving our attention free to dwell on the message and the attitudes of speaker and listener. The lowly power of habit is the support of the distinctively human gift of language. We can in this sense speak of language as a conventionalized, highly complex system of habits which functions as a human instrument of communication.

## 1.4 LANGUAGE, CULTURE, AND THE INDIVIDUAL

When an individual communicates with another in a common language he emits certain sounds (or writes certain symbols) that in particular distributions have particular meanings some of which he intends to convey to his listener and others that go along through the habit patterns imposed by the language. The sounds and their arrangement and the units which they constitute we call *form*; the permitted environments in which each form appears constitute its *distribution*; and the meanings, intended and habitual, are the *meaning*. When the listener who speaks the same language hears these forms in this distribution he grasps the same linguistic meanings encoded by the speaker in his utterance.

If these two individuals are of the same culture, the cultural meanings that each will encode in language for communication will usually be those which are common to other members of the cultural community. The listener thus grasps from the linguistic utterance the cultural meanings encoded in it by the speaker.

Beyond the culturally bound meanings of language each person perceives meanings which are his own through the particular life experience that is his. The ordinary forms of a language and the ordinary meanings of a culture do not accommodate these individual meanings. If the word *dog* has a pleasant connotation for one who remembers a dog he had as a boy, when he uses this word in a message for one who remembers dogs as animals that frightened him or injured him, the individual meaning of the speaker will not be transmitted to the listener, who will on the contrary add his unpleasant individual meaning to the linguistic and cultural meaning of *dog*.

Figure 1.1 is intended to show these elements in a schematic way. Individual meaning is outside of the system of language and outside of the coverage of this book. Cultural meaning is part of language. Since cultural meanings, however, have not been fully described in any culture we will usually have to be content with minimal features of cultural meaning of wide currency in a particular culture. Linguistic meaning is part of

## LANGUAGE

language, for without linguistic meaning there is no language. The description of linguistic meaning has received much attention in dictionaries and grammars. Much of it remains vaguely described, yet we will have to deal with it if we are to test language in operation.

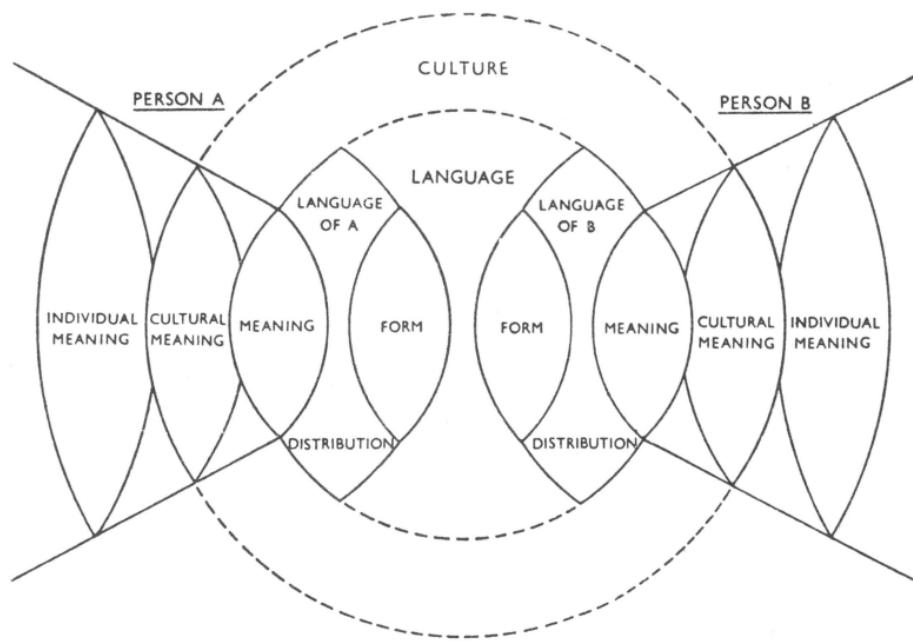


Fig. 1.1. Language, Culture, and the Individual

# VARIABLES AND STRATEGY OF LANGUAGE TESTING

4.1

## VARIABLES

**1. The elements.** The matter to be tested is language. Language is built of sounds, intonation, stress, morphemes, words, and arrangements of words having meanings that are linguistic and cultural. The degree of mastery of these elements does not advance evenly but goes faster in some and slower in others. Each of these elements of language constitutes a variable that we will want to test. They are pronunciation, grammatical structure, the lexicon, and cultural meanings. The first of these, pronunciation, is itself made up of three separate elements, namely sound segments, intonation and its borders, and stress and its sequences which constitute the rhythm of a language.

Within grammatical structure there are two main subdivisions, namely morphology and syntax. Syntax will be given priority in testing. Morphology will be treated as much as possible in connection with syntax.

The testing of the lexicon and of cultural meanings will be treated as separate variables although they are closely related.

**2. The skills.** The elements of language discussed can be profitably studied and described—and tested—as separate universes, yet they never occur separately in language. They are integrated in the total skills of speaking, listening, reading, and writing. That is, granted that a person's pronunciation may be ahead or behind his knowledge of vocabulary, when he speaks we can lump both together in stating what his ability to speak the foreign language is. There are then these four skills, the mastery of which does not advance evenly. We have thus four more variables to be tested, namely the degree of achievement in

speaking, understanding, reading and writing. A fifth skill is the ability to translate, which should be tested as an end in itself and not as a way to test mastery of the language. Each of these additional five skills will be treated in separate chapters.

#### 4.2 STRATEGY IN FOREIGN LANGUAGE TESTING

**i. Situation versus language.** If there were a high uniformity in the occurrence and use of the various variables and units of language we could simplify the testing of a foreign language a great deal. We could take any situation and engage the student in speaking, listening, reading, and writing and could give him an achievement score that would be the same if we chose a different situation, a longer topic, etc. This is not the case, however.

The situations in which language is the medium of communication are potentially almost infinite. No one, not even the most learned, can speak and understand his native language in any and all the situations in which it can be used. There is the technical language of the scientist in a particular field that is not understood by scientists from other fields of study. There is the language of the trained sailor which is not understood by the outsider. Even though all the classes in a large university may be taught in the same language it is doubtful that any professor could go into any and all the classrooms of his university and understand what is being explained. It is easy to find situations in which a person who speaks and understands a language natively fails to understand what is said because the subject matter is not within his experience.

At the same time it is easy to think of situations in which one would understand what a speaker means even without understanding the language. Leonard Bloomfield's example to illustrate the basic sequence of language between speaker and listener can also be used to illustrate a case in which communication takes place in spite of the lack of a common language. In Bloomfield's example, Jack and Jill are taking a walk. It is assumed that they speak the same language. Jill sees an apple in a tree nearby. She makes some noises with her vocal apparatus assumed to be language noises. Jack vaults the fence that separates them from the tree, fetches the apple and gives it to Jill who eats it. The

situation here is so obvious that we can imagine Jack and Jill speaking two different and unrelated languages, yet when Jill looks insistently at the apple and makes language noises that are unintelligible to Jack, he still vaults the fence, fetches the apple and gives it to Jill who eats it.

We have no assurance that we have tested language merely because a situation has been understood. And we have no assurance that it is lack of knowledge of the language when a situation is not understood.

Strategically a situation approach that does not specifically test language elements is not effective. It has only the outward appearance of validity.

Furthermore, even if we could pick only valid situations and even if we could be sure that understanding these situations occurred through the language used, we would still have the problem of the great variety of situations which must be sampled.

The elements of language on the other hand are limited, and it is more profitable to sample these elements than to sample the great variety of situations in which language can be used.

**2. Skills versus elements of language.** Even when we decide to test the language as directly as possible, we still are faced with choices between integrated skills and separate elements. It is economical and more comprehensive to test the sound segments systematically than to test only a few sounds and a few grammatical patterns in a general test of auditory comprehension, yet even in testing the sound segments systematically we have to decide whether to test this element in speaking or in listening.

The decision in the choice of skill versus element will depend to a considerable extent on the purpose for which the test is made and the available time for testing, plus the conditions under which the test must be given. If we want a comprehensive diagnostic test of pronunciation to be used for pinpointing specific problems yet to be mastered by the student we would obviously choose a test that ranges over the sound system, a specific test of that element of language.

If on the other hand we merely want a general level score to tell us whether or not this student may go on to study other subjects using the foreign language as a tool, we may choose a skill

test of auditory comprehension plus a test of reading, and perhaps one of writing. The test of auditory comprehension will have to choose a few sounds and a few structures at random hoping to give a fair indication of the general achievement of the student.

If we need to determine general proficiency without reference to listening alone or to any of the separate skills alone and if we have ample time for a comprehensive test we may choose to range over the several elements of language with some sections to test the integrated skills as well.

The conclusion here is that we need to test the elements and the skills separately and will choose one or the other or a combination depending on the purpose and conditions of the test.

**3. Group testing versus individual testing.** In general we will be interested in testing under practical teaching situations. In these we will have to handle considerable numbers of students under limited time and personnel. We will not be able to plan thirty-hour interviews for each student such as the anthropologist may conduct in his investigations. Even when we need to give individual tests we will attempt to devise techniques to do it in the shortest possible time to permit language teachers to use such a test.

A simple multiplication will show the necessity of using group tests or very short individual tests in a school situation. If we give a fifteen-minute individual test to one hundred pupils, the teacher would need to be giving these tests for three days working eight hours without rest. Three days of the examiner's time for a fifteen-minute test is not economical. On the other hand a modern group test of one hour duration will take only one hour of the examiner's time to administer. The matter of validity will be discussed fully in other chapters. It will not necessarily favor the fifteen-minute individual test over the one-hour group test.

**4. Objective versus subjective tests.** These terms are used to designate two types of scoring. Objective tests are those that are scored rather mechanically without need to evaluate complex performance on a scale. Subjective tests are those that require an opinion, a judgment on the part of the examiner.

In subjective tests there are differences in scoring by different examiners, hence the name subjective.

In general subjective tests permit the use of techniques that are natural and seem outwardly very valid. Objective tests have been criticized at times because in order to reduce the items to mechanical scoring they are rendered outwardly artificial.

This matter of objective versus subjective scoring is not one of two extremes; that is, tests are not either subjective or objective but range over a scale with completely objective scoring at one end and completely subjective scoring at the other. Often we have to choose between more apparent validity but less objectivity and more objectivity but less apparent validity.

This dilemma can be broken partly at least by locating and describing the linguistic problems to be tested with the accuracy that linguistic analysis makes possible. Knowing what an item is testing we are left free to choose objective techniques even if outwardly they seem less valid, for if they test the language problems in essentially valid linguistic situations they are valid items.

## Chapter 5

# CRITICAL EVALUATION OF TESTS

### 5.1 CRITERIA FOR THE EVALUATION OF LANGUAGE TESTS

The questions we ask about a test will vary in each case depending on purpose, time, subject, etc. In general, however, we must ask if a test is valid, reliable, scorable, economical, and administrable.

**1. Validity.** Does the test measure what it is intended to measure? If it does, it is a valid test. Validity is not general but specific. If a test of pronunciation measures pronunciation and nothing else, it is a valid test of pronunciation; it would not be a valid test of grammar or vocabulary because it does not test grammar or vocabulary.

Validity in language tests depends on the linguistic content of the test and on the situation or technique used to test this content. A test that uses a perfectly valid conversational situation but does not test the elements of the language is not valid. On the other hand, a test that tests the elements of the language but does it by lists or rules or technical names rather than in use in essentially communicative situations is not a valid test either.

Also, a test that has good language content and appears to use essentially linguistic situations will not be valid if it introduces a very heavy intelligence factor, or a heavy memory factor, or some other element that cancels out the language content and the valid situation.

Validity can be achieved and verified indirectly by correlating the scores on a test with those of another test or criterion which is valid. If the two sets of scores correlate highly, that is, if students who make high scores on the valid criterion test also score high on the experimental test and if those who score low on one also score low on the other, we say that the test is valid.

**2. Reliability.** Does a test yield the same scores one day and the next if there has been no instruction intervening? That is, does the test yield dependable scores in the sense that they will not fluctuate very much so that we may know that the score obtained by a student is pretty close to the score he would obtain if we gave the test again? If it does, the test is reliable.

Reliability is necessary for validity, because a test with scores which fluctuate very much does not test anything. Reliability, however, is general rather than specific. If the scores on a test are steady, that is, reliable, they are reliable regardless of what we test. If the scores of a pronunciation test are reliable they will remain reliable even if we use the scores as a measure of vocabulary although they are not valid as a measure of vocabulary and should not be used as such.

Reliability is measured by a correlation between the scores of the same set of students on two consecutive administrations of the test. This is known as the re-testing coefficient of reliability. Reliability is also measured by computing separate scores for two halves of the same test. The correlation between the sets of scores for the two halves of the test is the coefficient of reliability for one-half of the test. By a special formula<sup>1</sup> we can compute from the reliability of half of the test a close estimate of the reliability of the entire test.

**3. Scorability.** Can the test be scored with ease so that the users may be able to handle it? Subjective tests are not easy to score. The examiner who is conscientious hesitates, wonders if this response is as good as another he considered good, if he is being too easy or too harsh in his scoring. Objective tests are easy to score. Even objective tests vary in their scorability. A test that collects all the responses on a separate sheet that can be scored by a superimposed punched stencil or by a machine is more scorable than one which has the responses scattered in the pages of the test and not on the same column. The difference will affect the accuracy of the scoring and the time it takes to score a test.

**4. Economy.** This is a practical criterion. Does the test measure what we want it to test in a reasonable time considering the testing situation? If it does, the test is practical and economical.

<sup>1</sup> The Spearman-Brown correction formula is generally used.

## CRITICAL EVALUATION OF TESTS

**5. Administrability.** Can the test be given under the conditions that prevail and by the personnel that is available? That is, if we do not have a record player, a test on a record cannot be administered. In general we will think of the more representative situations. If tape-recorders are generally available in the schools of Japan, a tape-recorded test is satisfactory even if in a particular school there does not happen to be a tape-recorder at hand. We can assume that in Japan the school might easily obtain a tape-recorder.

If a test requires highly trained engineers to handle electronic equipment it is not administrable since most schools would not have the services of an engineer available every time the test is to be given.

Much of the above is simply a matter of good practical common sense. Some of the criteria, however, will require technical formulas that will be discussed in later chapters.