Article prepared for the Elsevier *Encyclopedia of Language and Linguistics,* 2nd edition.

Ethnologue

Gary F. Simons

SIL International, Dallas, USA

Raymond G. Gordon, Jr.

SIL International, Dallas, USA

Ethnologue: Languages of the World is a reference work cataloging all known languages of the present-day world. Now in its fifteenth edition (2005), the *Ethnologue* identifies 6,912 living languages, both spoken and signed. These are distinct languages that have living mother-tongue speakers. A few hundred recently extinct languages are documented as well.

For over 50 years, the *Ethnologue* has been compiled and published by SIL International, a nonprofit, nongovernmental organization that studies, documents, and assists in developing the world's lesser-known languages. Information comes from a variety of sources including reliable published sources, a network of field correspondents, and numerous personal communications that are confirmed by consulting published sources or the network of correspondents. The editorial staff processes approximately 10,000 updates to the database every year.

History of the *Ethnologue*

The *Ethnologue* was founded by Richard S. Pittman who was motivated by the desire to share information on language development needs around the world with his colleagues in SIL International as well as with other language researchers. The first edition in 1951 was ten mimeographed pages and included information on 46 languages or groups of languages. Maps were first included in the fourth edition (1953). The publication transitioned from mimeographed pages to a printed book in the fifth edition (1958). Dr. Pittman continued to expand his research through the seventh edition (1969) which listed 4,493 languages.

In 1971 Barbara F. Grimes became editor. She had assisted with the *Ethnologue* since 1953 (fourth edition) and took on the role of research editor in 1967 for the seventh edition (1969). She continued as editor through the fourteenth edition (2000). In 1971 information was expanded from primarily minority languages to encompass all known living languages of the world. Between 1967 and 1973 Ms. Grimes completed an in-depth revision of the information on Africa, the Americas, the Pacific, and a few countries of Asia. During her years as editor, the number of identified languages grew from 4,493 to 6,809, and the information recorded on each expanded so that the published work more than tripled in size.

The fifteenth edition (2005) was edited by Raymond G. Gordon, Jr. It reflects an increase of 103 languages over the previous edition. Most of these are not newly discovered languages, but are ones that had been previously considered dialects of another language.

The problem of language identification

Due to the nature of language and the various perspectives brought to its study, it is not surprising that a number of issues prove controversial. Of preeminence in this regard is the definition of *language* itself. Since languages do not have self-identifying features, what actually constitutes a language must be operationally defined. That is, the definition of language one chooses depends

on the purpose one has in identifying a language. Some base their definition on purely linguistic grounds. Others recognize that social, cultural, or political factors must also be taken into account.

Every language is characterized by variation within the speech community that uses it. The resulting speech varieties are more or less divergent from one another. These divergent varieties are often referred to as dialects. They may be distinct enough to be considered separate languages or sufficiently similar as to be considered merely characteristic of a particular geographic region or social grouping within the speech community. Scholars do not all share the same criteria for what constitutes a "language" and what features define a "dialect." The *Ethnologue* applies the following basic criteria:

- Two related varieties are normally considered varieties of the same language if speakers of each variety have inherent understanding of the other variety at a functional level (that is, can understand based on knowledge of their own variety without needing to learn the other variety).

- Where spoken intelligibility between two varieties is marginal, the existence of a common literature or of a common ethnolinguistic identity with a central variety that both understand can be a strong indicator that they should nevertheless be considered varieties of the same language.

- Even where there is enough intelligibility between varieties to enable communication, the existence of well-established distinct ethnolinguistic identities can be a strong indicator that they should nevertheless be considered to be different languages.

Increasingly, scholars are recognizing that languages are not always easily treated as discrete isolatable units with clearly defined boundaries between them. Rather, languages are more often continua of features that extend across both geographic and social space. The *Ethnologue*

approach to listing and counting languages as though they were discrete, countable units does not mean to preclude a more dynamic understanding of the linguistic makeup of the countries of the world. In fact, particular language entries in the *Ethnologue* list known dialects and often comment on the similarity and intelligibility relationships among them. In the final analysis, however, the *Ethnologue* lists and counts languages as distinguished by the criteria named above because it serves as a baseline for those who are developing language policy and making plans for language development. It is also foundational for those, like librarians and archivists, who would classify written and spoken materials with respect to the languages they are in, or would organize pieces of language-related information with respect to the languages they are about.

Three-letter language identifiers

A distinctive feature of the *Ethnologue* over the years has been its use of three-letter codes to uniquely identify the languages of the world. Any enterprise that would categorize language-related resources so that others might effectively retrieve those resources depends on the uniform identification of the languages to which they pertain. Simply using language names for this purpose is not adequate since the same language is typically known by many names and those names change over time. Furthermore, different languages may be known by the same name. Thus, the most effective approach is to use standardized language identifiers.

Standardized language identifiers were introduced into the *Ethnologue* in 1971 by then consulting editor, Joseph E. Grimes, when he transformed the typesetting tapes for the seventh edition (1969) into a computerized database on languages of the world. The work was done at the University of Oklahoma under a grant from the National Science Foundation. In 1974 the database was moved to a computer at Cornell University where Dr. Grimes was professor of linguistics; it was moved to a personal computer in 1979. Since 2000 it has been housed at the headquarters of SIL International in Dallas, Texas.

One feature of the database since its inception has been a system of three-letter language identifiers. Grimes explained this feature as follows in the 1974 final report for the grant: "Each language is given a three-letter code on the order of international airport codes. This aids in equating languages across national boundaries, where the same language may be called by different names, and in distinguishing different languages called by the same name." While the codes were used behind the scenes in the database that generated the eighth and ninth editions, it was not until the tenth edition (1984) that they appeared in the publication itself.

The fifteenth edition (2005) marked an important milestone in the development of the language identifiers, namely, their emergence as a draft international standard. In 1998, the International Organization for Standardization adopted ISO 639-2—its standard for three-letter language identifiers. That was based on a convergence of ISO 639-1 (its earlier standard for two-letter language identifiers adopted in 1988) and of ANSI Z39.53 (also known as the MARC language codes, a set of three-letter identifiers developed within the library community and adopted as an American National Standard in 1987). The current standard, ISO 639-2, has proven insufficient for many purposes since it has identifiers for fewer than 400 individual languages. Thus, in 2002, ISO TC37/SC2 invited SIL International to participate in the development of a new standard based on the language identifiers in the *Ethnologue.* The new standard was to be a superset of ISO 639-2 that would provide identifiers for all known languages. Consequently, hundreds of the *Ethnologue* language identifiers were changed in order to achieve alignment with ISO 639-2. In 2004 the proposed new standard, ISO 639-3, passed the first round of balloting by national standards bodies to attain the status of Draft International Standard. The three-letter language identifiers in the fifteenth edition of the *Ethnologue* are thus the codes of ISO/DIS 639-3.

Endangered languages

Language endangerment is a serious concern to which linguists and language planners have turned their attention in the last decade. For a variety of reasons, speakers of some languages are motivated to stop using their language and to use another. Parents may begin to use only that second language with their children. Eventually there may be no speakers who use the language as their first or primary language and frequently the language ceases to be used altogether and the language becomes extinct—existing, perhaps, only in recordings or written records and transcriptions. The concern about language endangerment is centered, first and foremost, around the factors which motivate speakers to abandon their language and the consequences of language death for the community of (former) speakers of that language. Since language is closely linked to culture, loss of language almost always is accompanied by social and cultural disruptions as well. Secondarily, those concerned about language endangerment recognize the implications of the loss of linguistic diversity both for the linguistic and social environment generally and for the academic community which is devoted to the study of language more specifically.

There are two dimensions to the evaluation and characterization of endangerment—the number of speakers of the language and the number and nature of the domains in which the language is used. A language may be endangered because there are fewer and fewer people who speak that language. It may also, or alternatively, be endangered because it is being used for fewer and fewer functions. The *Ethnologue* attempts to provide data on both of these dimensions whenever it is available.

Language endangerment is a matter of degree. At one end of the scale are languages that are vigorous and perhaps are even expanding in numbers of speakers or functional areas of use. At the other end are languages that are on the verge of extinction. In between are many degrees of greater or lesser endangerment. The *Ethnologue* does not attempt to identify the level of endangerment of each language listed but does specifically identify those languages at the far end

of the scale by indicating "Nearly extinct" at the end of the language entry. A language is listed as nearly extinct when the speaker population is fewer than 50 or when the number of speakers is a very small fraction of the ethnic group. In the fifteenth edition, 497 languages are so designated.

How to identify the level of endangerment of the remaining languages that are not designated as nearly extinct is not necessarily clear. Linguists seek to identify trends in language use, such as a decrease in the number of speakers or a decrease in the use of the language in certain domains or functions. An increase in bilingualism, both in the number of bilinguals and in their proficiency levels, is often associated with these trends. When data are available, the following factors which may contribute to endangerment are reported in the language entries: small population size, bilingualism, urbanization, modernization, migration, industrialization, the function of each language within a society, and whether or not children are learning it. Such factors interact within a society in dynamic ways that are not necessarily predictable. As a scholarly consensus forms that can be applied worldwide, a scale of endangerment is becoming increasingly possible. In the meantime, only brief statements about the above factors are given for each language as data becomes available.

Overview of contents

The *Ethnologue* begins with an "Introduction" and "Statistical Summaries." The latter give a summary view of the world language situation in terms of numerical tabulations of living languages and number of speakers by world area, by language size, by language family, and by country. Then follows the main body of the work in "Part 1: Languages of the World." This section provides detailed information on all known living languages of the world organized by area and by country. An extensive bibliography is located at the end of this section. "Part 2: Language Maps" provides 208 color maps locating the languages in most countries of the world. Finally, "Part 3: Indexes" consists of three indexes: a language name index listing all of the

39,491 distinct names that are associated with the languages described in Part 1, a language code index listing all of the three-letter language identification codes that are used in Part 1, and a country index listing the pages on which the section for that country begins in Part 1 and Part 2.

In Part 1, languages are listed by country under the five major geographic areas of Africa, Americas, Asia, Europe, and Pacific. The entry for a country begins with a header giving summary information about the country including official name, total population, a listing of national or official languages, a listing of recent immigrant languages, country literacy rates, and a count of languages indigenous to the country. This header is followed by an entry for each language of the country that is not a recent immigrant.

Entries are alphabetized by the primary name of the language. This is followed by all known alternate names and the three-letter identification code. An estimate of the speaker population is then given; there may also be an estimate of monolingual speakers, or of the size of the ethnic group (including those who no longer speak the language). Next, the location of the language group within the country is described, followed by the genetic classification of the language. Where dialects are known to exist, these are listed along with alternate names. Comments on intelligibility and similarity relationships among dialects or with neighboring languages may follow. Next are notes on language use, including functions of the language (such as official language or language of wider communication), viability, domains of use, age range of speakers, attitudes toward the language, and bilingual proficiency in other languages. These are followed by notes on the status of language development, including literacy rates, use in elementary or secondary schools, scripts used for writing, existence of published literature, and use in media. The entry closes with information in miscellaneous categories including general remarks, linguistic typology, geological and ecological environment, subsistence type of the speakers, and religions.

Further reading

Gordon, R. G., Jr. (ed.) (2005). *Ethnologue: Languages of the World,* (15th edn.). Dallas: SIL International. Web edition at: http://www.ethnologue.com.

Hale, K., Krauss, M., and others. (1992). Endangered languages. *Language* 68(1):1–42.

Abstract

Ethnologue: Languages of the World is a reference work published by SIL International that catalogs all known languages of the present-day world. Now in its fifteenth edition, the *Ethnologue* identifies 6,912 living languages, both spoken and signed. This article recounts the 50-year history of the *Ethnologue*, describes the criteria used for language identification, traces the development of the three-letter language identification codes that have been a distinctive of the *Ethnologue*, discusses ways in which the *Ethnologue* documents the problem of language endangerment, and gives an overview of the contents of the complete work.

Biographies

Gary F. Simons is Associate Vice President for Academic Affairs, SIL International. In this role he oversaw the production of the fifteenth edition of the *Ethnologue* as Executive Editor. In recent years he has been actively engaged with the linguistics community in matters relating to digital language archiving and standards for language identification. Before taking up his current role in 1999, he spent 15 years directing SIL's Academic Computing Department where he oversaw the development of language software tools and of the SIL web site. Prior to that he did field work with SIL in Solomon Islands (1977-1983) as language surveyor and as translation advisor for North Malaita. In 1979, he received a Ph.D. in general linguistics from Cornell University (with minor emphases in Computer Science and Classics). His dissertation, *Language Variation and Limits to Communication,* was supervised by Joseph E. Grimes and deals with the problem of language identification.

Raymond G. Gordon, Jr. has worked with SIL International since 1962. Most recently he has served as editor for the *Ethnologue* (2000-2004). From 1984 to 1995 he served in field administration overseeing linguistic and translation work in North America. He had previously served as an SIL instructor and consultant in linguistics and literacy. From 1967 to 1974 he did linguistic research and translation work on the Crow Indian Reservation in Montana. In 1981, he received a Ph.D. in general linguistics from Cornell University.

Key words and cross-references

Ethnologue

SIL International

language identification

language vs. dialect

endangered languages, language endangerment

three-letter codes, language codes

ISO 639

(any of the above that have entries in the Encyclopedia are also candidates for cross-references)