
Reviewed by LYLE CAMPBELL, University of Utah, and VERÓNICA GRONDONA, Eastern Michigan University

What could be more natural for the linguistics profession than a catalogue of the world’s languages? In Ethnologue we have one, highly valuable, yet not produced by an organization of academic linguistics but by the Summer Institute of Linguistics (SIL), whose primary focus is Bible translation. Ethnologue (henceforth E) has become the standard reference, and its usefulness is hard to overestimate. It deserves the high commendation we emphasize in this review. It is to E that scholars and laypersons typically turn to answer such questions as, how many languages are spoken in the world?, what languages are spoken in a given country?, how many speakers are there of a particular language?, and so on. E is unquestionably the best source for answering these questions even if it does not always provide adequate answers. We assume E’s high merit is beyond debate, and therefore concentrate here on matters that we hope can lead to future improvements in this catalogue of the world’s languages.

An unusual feature of this volume is that the information has also been made available on an associated website (www.ethnologue.com) that can be searched easily. The editors are to be commended for making this service available at no cost to the public.


The purpose of E is to provide a comprehensive listing of the known languages of the world’ (7). For this, E had to face the question of how to define a language, or more precisely how to distinguish independent languages from dialects of the same language. As E says, ‘not all scholars share the same set of criteria for what constitutes a ‘language’ and what features define a ‘dialect’ ’ (8). E’s criteria include: (i) ‘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’; so far so good, but E adds (ii) ‘where spoken intelligibility between 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’; and (iii) ‘where there is enough mutual 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 varieties of the same language’; and, these criteria can be a delicate matter, raising serious questions of language identity. This is especially true in numerous instances where E lists as independent languages entities that most linguists consider dialects of a single language. This question of how to define a ‘language’, and in particular how to apply the definition, is a serious problem for E (see below).

‘Layout of language entries’ (10–13) explains the information included in the volume: primary language name, alternate names, language identification code, speaker population, location, linguistic affiliation, dialect names, intelligibility and dialect relations, lexical similarity, language function, viability, domains (of use), age (of speakers), language attitudes, bilingual proficiency, literacy rates, writing scripts, publications and use in media (especially Bible translations), general remarks, typology (especially basic word order), geological and ecological information, religion, and status (e.g. as a ‘second language only’). The information, however, is not consistent across language entries, often only information from this list down to ‘affiliation’ or ‘dialect names’ is present. For example, information on basic word order (typology) is present for only about 15% of the languages, on ‘religion’ for about 38%. We might speculate that SIL’s interests in Bible translation may account for the higher figure for religion than for typology, though perhaps religious affiliation is just known in more cases than basic word order is. For those with interests in endangered languages, we note that E lists a language ‘as nearly extinct when the speaker population is fewer than 50’ (8).

Most language entries contain a reference for the source of the information, though more than 10% have no indication of source (some 25% of these with no indication are listed as extinct). This source information is valuable, but points to another problem, namely that E’s information is very dated in a number of areas. For example, for Tukanoan languages of the Brazilian Upper
Amazon region, the sources for several are from 1986, several others from 1995; the newest (a single one) is 2000, the oldest 1973. The populations reported here are mostly out of date; linguists with interest in the area cannot rely on these figures; rather, they use the much more accurate ones of ISA & FOIRN 2000, which E could have consulted. Hammarström (2005) finds the earliest source year listed in E is 1922, with 183 entries whose source is before 1975, and 1,126 with sources from before 1985.¹

A major part of E is its enumeration of languages of the world. Frequently E counts more languages than specialists do, and sometimes fewer. Problems are observed of the following sorts: (i) cases where multiple languages are incorrectly assumed to be a single language; (ii) cases where a name has mistakenly been assigned to a language that does not exist; and (iii) instances where known, named languages are not represented. In addition, (iv) there are cases where the name chosen to represent a language in E is not the name by which the language is most commonly known (either by speakers or by scholars).²

There are numerous cases where a single language is assumed to be multiple languages and so has multiple-language names assigned in E. For example, the number of ‘living languages’ is given as sixty-eight for Mayan languages, where the standard linguistic view has thirty-one (living and dead). Otomanguean is said to have 172, compared to specialists’ count of under fifty (Zapotec is disputed; E lists fifty-eight Zapotec languages, against other specialists’ eight to twenty; E gives fifty-two Mixtec languages, opposed to only three to eight of other linguists).

The number of indigenous (‘living’) languages of different countries is inflated, and inconsistently reported in several cases. For example, Argentina is said to have thirty-nine living languages, twenty-five of them indigenous (23); but in Part 1, ‘Languages of the world’, listed by region and country, Argentina has only twenty-seven languages, and some of these are not living (Abipón long extinct; Chané, which has ‘not been spoken for 300 years’, 219; Ona; and Puelche). The immigrant (nonindigenous) languages include ‘Argentine Sign Language’, Central Aymara, South Bolivian Quechua, Spanish, and Welsh. There are not three distinct Wichí languages (Wichí Lhamtés Güisnay, Wichí Lhamtés Nocnent, Wichí Lhamtés Vejoz), but only one, and not two separate Chorote languages (Iyohwaja and Iyo’wujwa), but rather only one, leaving a maximum of nineteen indigenous languages, only fifteen of them living. El Salvador is presented as having five living indigenous languages (25, 252) when in fact it has only one, Pipil. Cacaopera and Lenca are extinct; ‘Keekhi’ (Q’eqchi’) is a recent immigrant language. Germany (538–40) has some thirteen varieties of German listed as distinct languages, though the independent status of many of these would certainly be denied by specialists.

Guatemala (26, 253–57) is listed with fifty-four living languages, all of them given as indigenous, including American Sign Language (253) and Spanish (256), though Chicomucelteco was never spoken in Guatemala (rather in Mexico) and has long been extinct. However, this figure of fifty-four languages is highly inflated. It gives two Achi languages (Cubulco Achi and Rabinal Achi), which are not distinct, and in fact on linguistic criteria are considered by many to be dialects of K’iche’. Similarly, two varieties of Chuj are distinguished, three of Ixlil, two of Jakalteko, eleven of Kaqchikel, six of K’iche’, six of Mam (including Tzacanec), three of Poqomam, two of Poqomchi’, and two of Tz’utujil; other Mayanists recognize only one distinct language for each of these. By contrast, Xinca (Xinka) is given as a single language and as extinct, though there were four Xinka languages, two (Guazacapán and Jumaytepeque) still spoken by a few. A retally for Guatemala, then, leaves twenty-four indigenous languages recognized by specialists, counting generously—less than the number given by E.

For Mexico (27), 297 living languages are listed, 291 of them indigenous. Mesoamerican linguists, however, recognize only some seventy-five to eighty (allowing for several Chinantec and Zapotec languages). Some also are extinct (Chiapanec, Chicomucelteco). For example, the twenty-eight Nahuaual languages (270–72) listed as distinct constitute an inflated number, even taking out Classical Nahuaual; on linguistic grounds specialists would insist there is only one or a very few Nahuaual languages; similarly, listing fifty-two distinct Mixtec languages is highly inflated. Ocuituco (Tlahuica) is mistakenly and misleadingly given as Atzingo Matlatzinca, though Ocuituco is a distinct language from Matlatzinca. Though E has eight Totonac languages, most specialists have only one, and even the most generous ones list no more than four.

E lists three indigenous languages for New Zealand (27), but Maori, with perhaps New Zealand Sign Language, is the only truly indigenous language there (Pitcairn-Norfolk is immigrant). Nicaragua (27) lists seven living languages, none immigrant (not even Spanish, 282); of the seven, however, three are extinct (Matagalpa, Subtiaba, and Monimbo). Panama (28) is given with eighteen languages, fourteen indigenous; but in Part 1, only a total of fourteen are listed, of which only eight are indigenous. For Peru, with ninety-four living languages (ninety-three indigenous, 28), thirty-three distinct Quechua languages are given (including Classical Quechua, listed as extinct), though for a the more realistic figure, this would need to be reduced.

There are other problems of enumeration. E does not list most extinct languages, though some 360 extinct languages are included. These are mostly languages that have become extinct in the last fifty years or so (‘Chorotega’ (Mangue), Matagalpa, Subtiaba, etc.), and some more ancient extinct languages, mostly those for which there are Bible translations (Abipón, Chané, Classical Greek, Classical Mandaic, Classical Nahuatl, Classical Quechua, Coptic, Ge’ez, Syriac). This inconsistent treatment of extinct languages confuses the counts of how many languages there are in the world today, as do also the instances of languages listed as ‘second language only’
Particularly disappointing and misleading to users is the language classification E uses, with ninety-four language families (16–24). This is inadequate, since there are at least 350 independent families (including isolates), which at present have no demonstrable affiliation with other families. South America alone has 114 independent language families and isolates. In E's classification, one finds the following erroneous or highly disputed classificatory groups, as listed in 1.

(1) Altaic (see Campbell & Poser 2008)

Hokan (see Campbell 1997:290–305)

Khoisan (now generally acknowledged as a linguistic area but not a genetic unity; see Campbell & Poser 2008)

Mataco-Guaicuru (two separate families, Matacoan and Guaicuruan, not demonstrably related)

Na-Dene (with Haida included; see Campbell 1997:284–88)

Nilo-Saharan (generally seen as Joseph Greenberg's wastebasket grouping of leftovers; see Campbell & Poser 2008)

Penutian (see Campbell 1997:309–20)

By contrast, E failed to represent some undisputedly accepted groups. For example, it lists as independent Subtiaba-Tlapanec (which belong to Otomanguean), and Panoan and Tacanan (members of the Pano-Tacanan family).

Some of the names given are problematic in that they do not follow the standard convention of distinguishing individual language names (with no suffix) from family names (which typically bear the -an suffix), for example, Carib (where Cariban is expected), Choco (for Chocoan), Tupí (for Tupían). This is not a consistent pattern, however, since several do follow the standard convention, for example, Barbacoan. The convention is an important one, to keep individual languages, such as Tupí, from being confused with the name of the family as a whole (thus Tupían rather than Tupí for the family name, Makuan not Maku, Muran not Mura, etc.).

E counts only thirty-six language isolates (21), which is misleading since there are 116 isolates in the world (about 25% of these are extinct).

E also frequently confuses language isolates (languages with sufficient attestation but no evidence to demonstrate a genetic relationship with any other language) and unclassified languages (where the available data are insufficient to determine whether they have relatives or not). Thus, for example, contrary to E, Beothuk, Kunza, Puquina, and several others are generally considered language isolates, not unclassified languages (See Hammarström 2005, 2007 for additional examples). By contrast, some languages treated by E as isolates or unclassified are in fact small families of languages; for example, Lenca (El Salvador, Honduras) is not unclassified, but is a family of two languages; Xinca (Guatemala) is not unclassified, but a family of four languages; Tol (Jicaque) (Honduras) is not an isolate, but a family of two languages. And of course, several isolates were erroneously lumped into Hokan, Penutian, Gulf, or Na-Dene—for example, Alsea, Atakapa (really a family of two languages), Chimariko, Chitimacha, Esselen, Haida, Kalapuya, Karok (Karuk), Molale, Natchez, Salinan, Seri, Takelma, Tunica, Yana, and so on.

E's treatment of language classification is perhaps its weakest and most disappointing aspect; it is to be hoped that future editions will treat this topic much more seriously.

By contrast, the language maps are highly valuable. Most represent a single country, though some group a few countries into one map (e.g. Algeria-Morocco-Tunisia, Argentina-Chile, El Salvador-Honduras). Some countries are split into more than one map, for example, Northern Cameroon, Southwestern Cameroon (with a separate map of an enlarged portion of Southwestern Cameroon), Southeastern Cameroon, Northern Colombia, and Southern Colombia. Djibouti, Eritrea, and Ethiopia are on a single map, but an enlargement of Southwestern Ethiopia is on a separate map. Nigeria has ten maps plus an index map; Mexico has four maps.

Most of the maps are of high quality and are user-friendly, though not all are of equal quality. Brazil, for example, one of the largest countries with many languages, is all crowded onto a single map with two small insets. Somewhat smaller Canada, with far fewer languages, has three maps (and two insets), and the US has six maps and an index map. For most of the maps, the different language families are color-coded, but Brazil is unfortunately monochrome; China has three maps and several language families, and Australia has two maps and several language families, but all these maps are also monochrome.

With regard to the intriguing question of how many languages there are, we note that of the 6,912 living languages listed in E (15), arguably the five pidgins and one artificial language should be eliminated, and the 119 deaf sign languages should be given a special status. The several languages known on independent evidence to be extinct but listed as living should be eliminated from the list; some example are Chiapanec, Chichimec, Lenca, Cacaopera, Ona, and Pehuenche. E lists 263 Australian languages, with 224 given as not extinct, although Dixon
(2002:2) reports ‘more than half of these [240–250 indigenous Australian languages] are no
longer spoken or remembered’.  

On another front, we need to restore to the living those nonextinct languages that are listed
as extinct (for example two Xinkan languages of Guatemala). And we need to eliminate some
spurious languages, for example, Monimbo (Nicaragua), a present-day ethnic group with no
known language, though probably descendants of Mangue speakers (see Hammarström 2005 for
other examples).

To conclude we reiterate again, forcefully, our assessment that E is truly excellent, highly
valuable, and the very best book of its sort available. Given the amount of work it represents
and the amount and kind of information it provides, it might seem greedy to ask for more. We
have, nevertheless, attempted to point out areas where E can and should be improved. We feel
this is even more important in this case than it might normally be, since the languages and their
codes represented in E have been adopted by ISO (the International Organization for Standardiza-
tion) as the international standard. This means that mistakes in E are carried over to these ISO
codes and are compounded as libraries, archives, and funding agencies use these names and
codes. ISO has given SIL ‘administrative authority’ for these language names and codes, and
SIL is eager to make reasonable corrections. For that reason, we urge the editors of E to address
seriously the problems of the sort identified here and to ensure that the changes also make it
into the ISO code standards.

REFERENCES

Oxford University Press.


ling.ed.ac.uk/linguist/issues/16/16-2637.html.

for typologists. (LINCOM handbooks in linguistics 22.) Munich: LINCOM Europa. Online:

Povos indígenas do alto e médio Rio Negro: Uma introdução à diversidade cultural e ambiental do
noroeste da Amazônia brasileira. São Paulo: Instituto Socioambiental/Federação das Organizações
Indígenas do Rio Negro/Ministério da Educação.

1 The source of information for most references in this review to languages of the Americas is Campbell
1997.

2 For more detail and more examples on these and several other problems in E than we present here, see

3 For very exact information on surviving Australian languages and their speakers, see the National Indige-

***** END OF FILE *****

13043S707 06-18-08 08:59:44