

9 **Ethnologue:** Languages of the world. 15th ed. Ed. by RAYMOND G. GORDON, JR. Dallas:
10 SIL International, 2005. Pp. 1,272. ISBN 155671159X. \$80(Hb).

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

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

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

27 *E* has the following chapters and sections: 'Introduction', 7–14; 'Statistical summaries', 15–36;
28 Part 1: 'Languages of the world', 37–648; 'References', 649–72; Part 2: 'Language maps',
29 673–888; Part 2: 'Indices', 889–1272 (including the 'Language code index', 1231–70).

30 The purpose of *E* 'is to provide a comprehensive listing of the known languages of the world'
31 (7). For this, *E* had to face the question of how to define a language, or more precisely how to
32 distinguish independent languages from dialects of the same language. As *E* says, 'not all scholars
33 share the same set of criteria for what constitutes a "language" and what features define a
34 "dialect"' (8). *E*'s criteria include: (i) 'two related varieties are normally considered varieties
35 of the same language if speakers of each variety have inherent understanding of the other variety
36 at a functional level'; so far so good, but *E* adds (ii) 'where spoken intelligibility between varieties
37 is marginal, the existence of a common literature or of a common ethnolinguistic identity with
38 a central variety that both understand can be a strong indicator that they should nevertheless be
39 considered varieties of the same language'; and (iii) 'where there is enough mutual intelligibility
40 between varieties to enable communication, the existence of well-established distinct ethnolingu-
41 istic identities can be a strong indicator that they should nevertheless be considered to be different
42 languages' (8). Linguists are usually driven to accept criteria (ii) and (iii) in certain situations,
43 but, given that they are inconsistent with (i), the application of (ii) and (iii) in any given instance
44 can be a delicate matter, raising serious questions of language identity. This is especially true
45 in numerous instances where *E* lists as independent languages entities that most linguists consider
46 dialects of a single language. This question of how to define a 'language', and in particular how
47 to apply the definition, is a serious problem for *E* (see below).

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

62 Most language entries contain a reference for the source of the information, though more than
63 10% have no indication of source (some 25% of these with no indication are listed as extinct).
64 This source information is valuable, but points to another problem, namely that *E*'s information
65 is very dated in a number of areas. For example, for Tukanoan languages of the Brazilian Upper

66 Amazon region, the sources for several are from 1986, several others from 1995; the newest (a
67 single one) is 2000, the oldest 1973. The populations reported here are mostly out of date;
68 linguists with interest in the area cannot rely on these figures; rather, they use the much more
69 accurate ones of ISA & FOIRN 2000, which *E* could have consulted. Hammarström (2005) finds
70 the earliest source year listed in *E* is 1922, with 183 entries whose source is before 1975, and
71 1,126 with sources from before 1985.¹

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

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

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

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

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

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

128 There are other problems of enumeration. *E* does not list most extinct languages, though some
129 360 extinct languages are included. These are mostly languages that have become extinct in the
130 last fifty years or so (‘Chorotega’ (Mangue), Matagalpa, Subtiaba, etc.), and some more ancient
131 extinct languages, mostly those for which there are Bible translations (Abipón, Chané, Classical
132 Greek, Classical Mandaic, Classical Nahuatl, Classical Quechua, Coptic, Ge’ez, Syriac). This
133 inconsistent treatment of extinct languages confuses the counts of how many languages there
134 are in the world today, as do also the instances of languages listed as ‘second language only’

135 (e.g. Latin, Sanskrit, and some twenty-five others).

136 Particularly disappointing and misleading to users is the language classification *E* uses, with
137 ninety-four language families (16–24). This is inadequate, since there are at least 350 independent
138 families (including isolates), which at present have no demonstrable affiliation with other families.
139 South America alone has 114 independent language families and isolates. In *E*'s classification,
140 one finds the following erroneous or highly disputed classificatory groups, as listed in 1.

- 144 (1) Altaic (see Campbell & Poser 2008)
- 146 Andamanese
- 148 Australian (but not Pama-Nyungan)
- Coahuiltecan (not in the list of language families in the statistical summaries, but
151 extinct Tonkawa and others are included)
- Gulf (not in the list of language families in the statistical summaries, but some lan-
154 guages in the entries are so classified)
- 156 Hokan (see Campbell 1997:290–305)
- Khoisan (now generally acknowledged as a linguistic area but not a genetic unity;
159 see Campbell & Poser 2008)
- Mataco-Guaicuru (two separate families, Matacoan and Guaicuruan, not demonstrably
162 related)
- 164 Na-Dene (with Haida included; see Campbell 1997:284–88)
- Nilo-Saharan (generally seen as Joseph Greenberg's wastebasket grouping of left-
167 overs; see Campbell & Poser 2008)
- Penutian (see Campbell 1997:309–20)

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

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

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

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

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

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

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

210 With regard to the intriguing question of how many languages there are, we note that of the
211 6,912 living languages listed in *E* (15), arguably the five pidgins and one artificial language
212 should be eliminated, and the 119 deaf sign languages should be given a special status. The
213 several languages known on independent evidence to be extinct but listed as living should be
214 eliminated from the list; some example are Chiapanec, Chicomuceltec, Lenca, Cacaopera, Ona,
215 and Pehuelche. *E* lists 263 Australian languages, with 224 given as not extinct, although Dixon

216 (2002:2) reports 'more than half of these [240–250 indigenous Australian languages] are no
217 longer spoken or remembered'.³

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

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

235 REFERENCES

236 CAMPBELL, LYLE. 1997. *American Indian languages: The historical linguistics of Native America*. New York:
237 Oxford University Press.

238 CAMPBELL, LYLE, and WILLIAM POSER. 2008. *Language classification: History and method*. Cambridge:
239 Cambridge University Press, to appear.

240 HAMMARSTRÖM, HARALD. 2005. Review of *Ethnologue*, ed. by Raymond J. Gordon. Online: [http://www](http://www.ling.ed.ac.uk/linguist/issues/16/16-2637.html)
241 [.ling.ed.ac.uk/linguist/issues/16/16-2637.html](http://www.ling.ed.ac.uk/linguist/issues/16/16-2637.html).

242 HAMMARSTRÖM, HARALD. 2007. *Handbook of descriptive language knowledge: A full-scale reference guide*
243 *for typologists*. (LINCOS handbooks in linguistics 22.) Munich: LINCOS Europa. Online:
244 www.cs.chalmers.se/~harald2.

245 ISA [Instituto Socioambiental] and FOIRN [Federação das Organizações Indígenas do Rio Negro]. 2000.
246 *Povos indígenas do alto e médio Rio Negro: Uma introdução à diversidade cultural e ambiental do*
247 *noroeste da Amazônia brasileira*. São Paulo: Instituto Socioambiental/Federação das Organizações
248 Indígenas do Rio Negro/Ministério da Educação.

249 ¹ The source of information for most references in this review to languages of the Americas is Campbell
250 1997.

251 ² For more detail and more examples on these and several other problems in *E* than we present here, see
252 Hammarström 2005, 2007.

253 ³ For very exact information on surviving Australian languages and their speakers, see the National Indige-
254 nous Languages Survey Report (2005; AIATSIS/FATSIL) at [http://www.dcita.gov.au/_data/assets/pdf_file/](http://www.dcita.gov.au/_data/assets/pdf_file/35637/NILS_Repport_2005.pdf)
255 [35637/NILS_Repport_2005.pdf](http://www.dcita.gov.au/_data/assets/pdf_file/35637/NILS_Repport_2005.pdf).

256

257

258

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

260 13043\$\$707 06-18-08 08:59:44

261