

***Ethnologue* as a global sourcebook on linguistic diversity and language vitality**

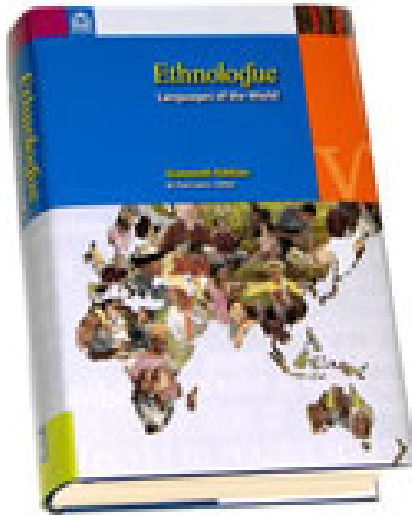
Gary Simons
SIL International

12th International Congress of Ethnobiology
Tofino, British Columbia, 9–14 May 2010

What is *Ethnologue*?

- Began in 1951 as 10 mimeographed sheets listing languages in areas where SIL might soon begin language development projects
- During the 1960s it grew to become a comprehensive catalog of all known languages
- In the 1970s it developed a system of three-letter codes to uniquely identify each language
- In 2007 these were adopted as ISO 639-3 after reconciling with the 400 codes in ISO 639-2

Latest edition



- *Ethnologue: Languages of the World, 16th edition.* M. Paul Lewis, ed.
- Published spring 2009
- ISBN 978-1-55671-216-6
- 1,248 pp., \$100.00
- Lists 6,909 living languages + 421 that have become extinct since 1950
- All content also accessible freely in a web version:
 - <http://www.ethnologue.com>

Nootka

A language of Canada

ISO 639-3: noo

Population	200 (2002 W. Poser). Includes 30 Nitinat (1991 M. Dale Kinkade). Canada Census (2001) lists 505. Ethnic population: 3,500 (1977 SIL).
Region	Southwest British Columbia, Pacific side of Vancouver Island, Nitinat, Nitinat Lake.
Language map	Southwestern Canada
Alternate names	Nutka, Nuuchahnulth
Dialects	Nitinat (Ditinat, Didinaht, Nitinaht).
Classification	Wakashan, Southern
Language use	Older adults. Also use English.
Language development	Dictionary. Grammar.

Languages by size

Population range	Living languages		Number of speakers	
	<i>Count</i>	<i>Percent</i>	<i>Count</i>	<i>Percent</i>
Over 1,000,000	389	5.6%	5,607,366,063	94.08%
100,000 to 999,999	895	13.0%	283,116,716	4.76%
10,000 to 99,999	1,824	26.4%	60,780,797	1.02%
1,000 to 9,999	2,014	29.2%	7,773,810	0.13%
1 to 999	1,510	21.8%	474 331	0.01%
Unknown	277	4.0%		
<i>Totals</i>	6,909	100.0%	5,959,511,717	100.0%

From *Ethnologue* 16th ed, "Summary by Language Size"

Greenberg's diversity index

- *Ethnologue* has reported a linguistic diversity index at the country level for the last 3 editions:
 - the probability that any two people selected at random would have different mother tongues
- For instance:

– Papua New Guinea	.990
– Vanuatu	.974
– Canada	.599
– United States	.319
– Cuba	.001
– North Korea	.000

Dimensions of diversity

- *Ecological diversity and its measurement*, by Anne Maguran (Princeton U Press, 1988)
- Diversity of species has two dimensions:
 - **Richness**: the total number of species
 - E.g., having 200 species is greater diversity than having 100 species
 - **Evenness**: the relative abundance of the species
 - E.g., having 10 species with equal populations is greater diversity than 10 species in which one accounts for 90% of all individuals

More measures of diversity

- Greenberg's Diversity Index converts to a measure of evenness as $1 / (1 - \text{GDI})$
 - E.g., Canada's .599 \rightarrow 2.49 = number of equal-sized languages that would give the same probability of having different mother tongues
- A simple metric of unevenness:
 - **Dominance:** the proportional abundance of the most abundant species
 - E.g., when the most abundant species accounts for 20% of all individuals there is greater diversity than when it accounts for 80%

Global language diversity

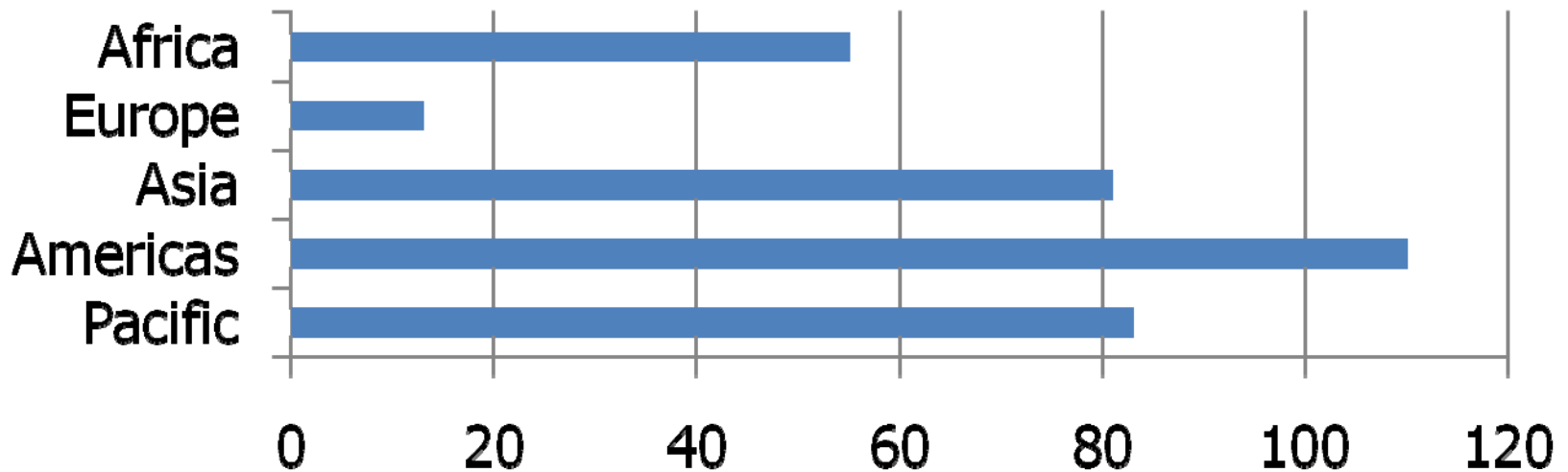
	GDI	Richness	Evenness	Dominance
World	0.967	6,909	30.30	14.2%
Africa	0.986	2,162	71.43	7.2%
Europe	0.934	376	15.15	16.6%
Asia	0.933	2,355	14.93	23.4%
Americas	0.723	1,187	3.61	38.0%
Pacific	0.464	1,310	1.87	73.1%

Loss of genetic diversity

- While the loss of any language without good documentation leaves a significant gap in the knowledge base of humankind, the loss of a whole family of languages without documentation leaves an even bigger gap.
- The typological diversity that demonstrates what is possible in human language has taken many millennia to unfold.
 - Losing a whole family loses an entire line of typological development that may be unlike any other.

Genetic richness

- *Endangered language families*, Whalen and Simons:
<http://scholarspace.manoa.hawaii.edu/handle/10125/5017>
- We identified 342 distinct linguistic stocks, i.e., groupings that are reconstructable to a common protolanguage, distributed as follows:

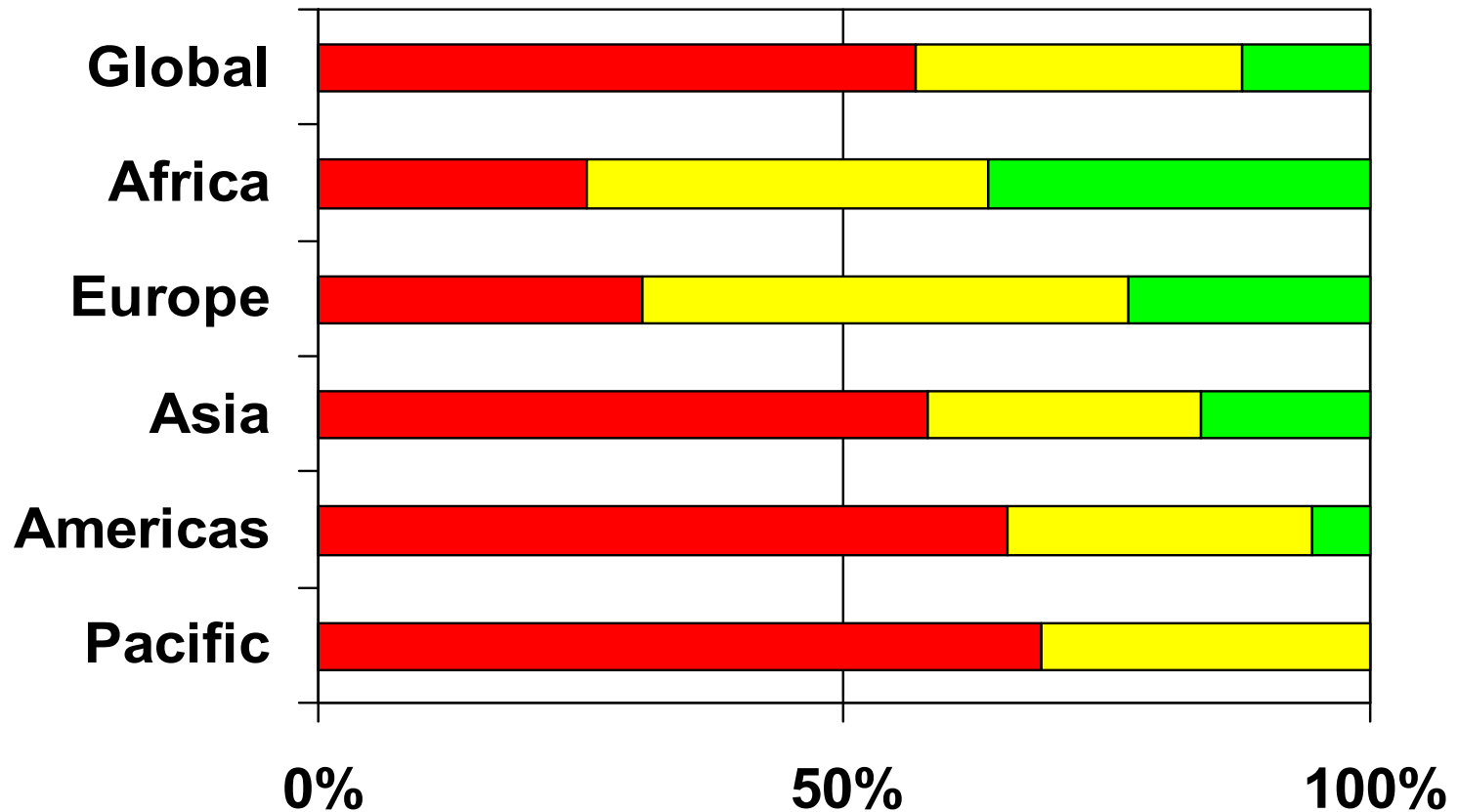


Estimating endangerment

- Endangerment data is not available for all languages; thus we estimated using population as a proxy
- The linguistics community consensus is that:
 - 50% are likely to become extinct or moribund by 2100
 - 40% are potentially in danger
 - 10% are safe
- The estimate for a linguistic stock is based on the population of its largest (i.e. safest) language:
 - Safe if in top 10% of languages ($> 330,000$)
 - Endangered if in bottom 50% ($< 7,500$)
 - Potentially endangered otherwise

Stocks at risk by area

■ Endangered ■ Potentially endangered ■ Safe



Toward global vitality data — EGIDS: Expanded GIDS

- “Graded Intergenerational Disruption Scale”,
from Fishman 1991, *Reversing Language Shift*
- EGIDS developed by Lewis and Simons for use in
Ethnologue to assess development or endangerment
 - Preprint: <http://www.sil.org/~simonsg/preprint/EGIDS.pdf>
- A 13-level scale which harmonizes:
 - Fishman’s 8-level GIDS
 - UNESCO’s 6-level scale of language endangerment
 - *Ethnologue’s* 5-level scale of language vitality
- Any language of the world can be categorized

EGIDS: Safe levels

LEVEL	LABEL	DESCRIPTION
0	Inter-national	The language is used internationally for a broad range of functions.
1	National	The language is used in education, work, mass media, government at the nationwide level.
2	Regional	The language is used for local and regional mass media and governmental services.
3	Trade	The language is used for local and regional work by both insiders and outsiders.
4	Educational	Literacy in the language is being transmitted through a system of public education.
5	Written	The language is used orally by all generations and is used in written form in parts of the community.
6a	Vigorous	The language is used orally by all generations and is being learned by children as their first language.

EGIDS: Endangered levels

LEVEL	LABEL	DESCRIPTION
6b	Threatened	The language is used orally by all generations but only some parents are transmitting it to children.
7	Shifting	The child-bearing generation knows the language well enough to use it among themselves but none are transmitting it to their children
8a	Moribund	The only remaining active speakers of the language are in the grandparent generation.
8b	Nearly Extinct	The only remaining speakers of the language are members of the grandparent generation or older who have little opportunity to use the language.
9	Dormant	The language serves as a reminder of heritage identity for an ethnic community. No one has more than symbolic proficiency.
10	Extinct	No one retains a sense of ethnic identity associated with the language.

Decision tree

1. What is the current identity function of the language?
2. What is the level of official use?
3. Are all parents transmitting the language to their children?
4. What is the literacy status?
5. What is the youngest generation of proficient speakers?

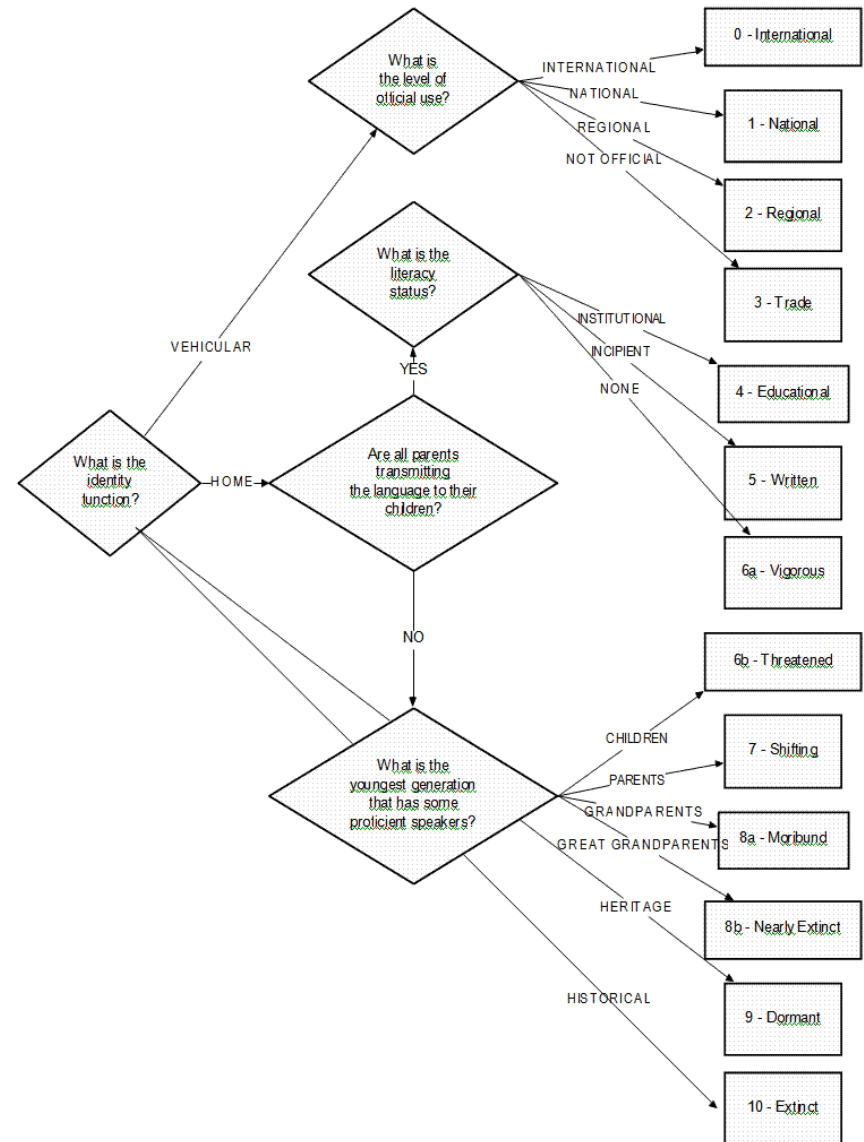


Figure 1 - Extended GIDS Diagnostic Decision Tree

Future developments

- We hope to include EGIDS assessments for most languages in the next edition
- We are working on plans to add a “crowd sourcing” element to data collection on the web:
 - for population estimates (L1 speakers, L2 speakers, ethnic population)
 - for language vitality facts like age range of speakers and domains of use
- In this way we hope to improve the coverage and the currency of the data on global ethnolinguistic diversity and vitality.