Endangered Language Families

D. H. Whalen
Haskins Laboratories and Endangered Language Fund

Gary F. Simons
SIL International and Graduate Institute of Applied Linguistics

1st International Conference on Language Documentation and Conservation, University of Hawaii, 12-14 March 2009
Loss of 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.
  - The diversity that distinguishes families from each other has taken much longer to develop than that which distinguishes languages in the same family.
What is at risk?

- Losing the last language in a family could mean losing unique evidence about the range of phenomena that are possible in human language.
- For instance, some phenomena brought to our attention through highly endangered languages:
  - OVS default word order (Carib: Hixkaryána)
  - The most elaborated click inventories (Khoisan)
  - Grammatical metathesis (Salishan, Penutian)
  - Obligatory use of evidentials (Tucanoan)
Research question

- What proportion of the world’s linguistic diversity is at risk in the current endangerment crisis?

- Previous discussions have approached this by looking at how many individual languages are at risk.

- Since innovations tend to be shared within a family, we attempt to gauge a deeper level of diversity by asking what proportion of language families are at risk.
Methodology

- We use data from:

- We begin with the classification of the world’s languages into families as reported in *Ethnologue* to estimate the proportion of families that are endangered.

- We then identify branches that correspond to reconstructable linguistic stocks to get a more refined estimate of diversity at risk.
Defining “endangered”

- One definition: “The language is no longer learned by children.”
  - Plus “unsafe” = “most children speak the language, but it may be restricted to certain domains”
  - The atlas identifies more than 2,250 specific languages that are at risk by these definitions.
Endangerment as risk

- Our approach follows the lead of Krauss 1992 (Language 68:4-10) who defined endangerment in terms of foreseeable risk.
  - **Moribund:** The language is "no longer being learned by children as mother tongue." (p. 4)
  - **Endangered:** “Though now still being learned by children, [the language] will—if the present conditions continue—cease to be learned by children during the coming century." (p. 6)
  - **Safe:** All other languages
Possible versus likely

- Krauss 1992: Only 10% of languages are safe
  - A call to action based on a warning that “at the rate things are going, the coming century will see either the death or the doom of 90% of mankind’s languages.” (p. 7)
  - How does this compare to language size?
    - In *Ethnologue 16*, 10% of languages have more than 330,000 speakers and 90% have fewer.
- Emerging consensus: It looks likely that 50% will be lost.
  - In *Ethnologue 16*, the median language size is 7,560; 50% of languages are larger and 50% are smaller.
Estimating endangerment

- For our study we use the sizes of languages in a family as an estimator of its endangerment:
  - **Endangered family**: If the largest language in the family has fewer than 7,560 speakers, there is a reasonable likelihood that no language in the family will be spoken by children at the end of this century.
  - **Potentially endangered family**: If the largest language in the family has between 7,560 and 330,000 speakers, there is a possibility that no language in the family will be spoken by children at the end of this century.
Families in the *Ethnologue*

- Top-level categories in the genetic classification of languages:
  - 116 families and 7 special categories: creole, constructed language, deaf sign language, language isolate, mixed language, pidgin, unclassified languages

- See handout for estimates of endangerment
  - Table 1: Endangered and potentially endangered
  - Table 2: Unendangered
Endangerment of families

Estimated endangerment for the 116 families in *Ethnologue 16*

- **56** Endangered
- **38** Potentially endangered
- **22** Unendangered

Legend:
- Red: Endangered
- Yellow: Potentially endangered
- Green: Unendangered
Going deeper

- Linguistic stocks give a more accurate basis for considering the range of linguistic diversity than *Ethnologue* “families”.

- Sapir (1921), From chapter 7 of *Language: An introduction to the study of speech*

  All languages that are known to be genetically related, i.e., to be divergent forms of a single prototype, may be considered as constituting a “linguistic stock”. There is nothing final about a linguistic stock. When we set it up, we merely say, in effect, that thus far we can go no farther.
Identifying stocks

- We used the AUTOTYP genealogical classification (kindly shared with us by Johanna Nichols) to identify the stocks in the *Ethnologue* classification.

- In the *Ethnologue* data, stocks are of three types:
  - 100 of 116 families are stocks (e.g. Austronesian)
    - Regular type face in Tables 1 and 2
  - The 16 other families (e.g. Afro-Asiatic) are “quasi-stocks” which comprise 140 stocks
    - Italics in Tables 1 and 2; stocks enumerated in 3 and 4
  - 102 isolates and unclassified languages are stocks
    - Listed in Table 5
Endangerment of stocks

- **Endangered**
- **Potentially endangered**
- **Unendangered**

---

**Family stocks**

- 40% Endangered
- 40% Potentially endangered
- 20% Unendangered

**Subfamily stocks**

- 40% Endangered
- 40% Potentially endangered
- 20% Unendangered

**Singleton stocks**

- 60% Endangered
- 20% Potentially endangered
- 20% Unendangered
Linguistic diversity at risk

Estimated endangerment for the 342 linguistic stocks in *Ethnologue* 16

- **194** Endangered
- **106** Potentially endangered
- **42** Unendangered
Viewing the results by world area

Table 7 in the handout answers two questions:

- Where is the world’s linguistic diversity in terms of the five world areas by which *Ethnologue* reports data?
- What are the estimates of endangerment for each of these areas?
Where is the diversity?

Geographic distribution of the 342 linguistic stocks in *Ethnologue 16*
Stocks at risk by area

- Endangered
- Potentially endangered
- Unendangered

Global
- Endangered: 60%
- Potentially endangered: 30%
- Unendangered: 10%

Africa
- Endangered: 40%
- Potentially endangered: 50%
- Unendangered: 10%

Europe
- Endangered: 50%
- Potentially endangered: 40%
- Unendangered: 10%

Asia
- Endangered: 40%
- Potentially endangered: 50%
- Unendangered: 10%

Americas
- Endangered: 60%
- Potentially endangered: 30%
- Unendangered: 10%

Pacific
- Endangered: 60%
- Potentially endangered: 30%
- Unendangered: 10%
Conclusion

- Repring Krauss’s (1992) conclusion:
  - Obviously we must do some serious rethinking of our priorities, lest linguistics go down in history as the only science that presided obliviously over the disappearance of 90% of the very field to which it is dedicated.

- For the moral and human side of the issue, this rethinking calls for social and political action.

- For the scientific side of the issue, this rethinking calls for changing priorities in research.
  - Data on endangerment of linguistic stocks, combined with data on existing documentation for stocks, could be used as a basis for developing priorities for future field research.