Seminar on Endangered Languages

Projects: at least one of the following

- Experience in documentation/collection
- Experience in resource creation
- Experience in LT development
Projects

- Final presentation (30mins)
- Writeup
- Plus artifacts (software, data, recordings)
Choosing your language

- **Must have accessible speakers**
  - *Not necessary physical (skype is fine)*
  - *Finding speakers can be hard*

- **Learn about issues in data collection**
  - *Political, social, etc*
  - *Why should they help you?*
Steven Bird UPenn/Melbourne

- Collect speech from indigenous peoples
- Respeak native speech (aligned)
- Respeak translated speech (aligned)
- Later transcribe/translate data
Writing Normalization

- **Script Normalization**
  - *Pacific Northwest Language*
  - *(Littell)*
Unwritten Languages

- **Use Inferred Phoneme** (Muthukumar)
  - Or SANTLR (Xinjian Li’s system)
- **Find symbol representation for unwritten language**
- **Use for pronunciation correction/documentation/translation**
Resource Creation

- Build lexicon/morphological analyzer
- From text (or from speech)
- Develop new tools or use existing tools
Collect Data in the Field

- Collect stories, respeaking, translation
- Interlinear gloss ....