Aikuma: A mobile app for collaborative language documentation

Steven Bird\textsuperscript{1,2}, Florian Hanke\textsuperscript{1}, Oliver Adams\textsuperscript{1}, Haejoong Lee\textsuperscript{2}
\textsuperscript{1}Department of Computing and Information Systems, University of Melbourne
\textsuperscript{2}Linguistic Data Consortium, University of Pennsylvania

Supported by NSF Grant to LDC, UPenn “Language Preservation 2.0: Crowdsourcing Oral Language Documentation using Mobile Devices”

Google Play: install Aikuma
Postponing the Transcription Task

Source audio → Oral translation

Proposed workflow

Translation

Typical workflow

Transcription
Thought Experiment: The future philologist

- time is shorter than we think
- what should we archive today?
- Boasian trilogy: texts, grammar, lexicon
- the primary resource?
Careful Respeaking

- Listen to an original recording and repeat what was said, slowly and carefully
- Recordings made “in the wild”
- Phonetic analysis (Bettinson 2013)
- Enables us to postpone the transcription task
Language Documentation

- what it consists of
- who it involves
- the challenge
- ensuring interpretability
  - what was said; what it meant
Design Principles

› text free
› synchronisation
› metadata
Transcription (mk 1)
Transcription (mk 2)

- Certo dia um homem foi para o mato.
  One particular day a man went into the forest

- Ele gostava muito de trabalhar de tucumã, cipó e piassaba.
  He very much liked to collect tucumã, cipó and piassaba

- Ele passa durante o seu trabalho mais ou menos uma...duas ou três semanas.
  He went out to the forest more or less every one, two or three weeks

- Trabalha muito longe, quase no meio do...do centro da mata.
  He works very far away almost in the middle of the...in the center of the forest

- Era muito longe.
  It was very far away
Sharing and Archiving
Architecture 1:
Backup and synchronization for several devices

Android devices

Personal Google Drive
Architecture 2: Village-based networking
Architecture 3: Cloud-based sharing

- Cloud-hosted metadata
- Android devices
- Personal Google Drives

▸ demonstrate cloud server
Architecture 4: Preservation

- not yet implemented
Ongoing work: More features

- video recording and playback
  - commentary not yet supported
- Informed consent
  - storage and dissemination
  - documenting informed consent
- Lexical elicitation
Future work: applying MT algorithms
Conclusion

- what does it take to preserve a language?
- the allure of MT
- MT in a week
- contribute: lp20.org