Minority Languages & Technology: Openness and Cross-Language Collaboration

Kevin Scannell
Saint Louis University
9 November 2015

Outline

- Full-stack language technology
- Case Study #1: Biigtigong Language Project
- Case Study #2: Intergaelic.com
- Common themes
- Recommendations

Full stack language technology

- Metaphor from software engineering

- Basic resources

  - Lexical database, corpora, terminology

- Foundational technologies

  - Word analysers, taggers, parsers

- User-facing applications

  - Speech, translation, dictionaries, proofing tools

Baile Chreimeadh na hAbhann

Biigtigong

mls
1

Biigtigong Language Project

- First Nation on the N. shore of Lake Superior
- Traditionally Nishnaabemwin-speaking
- Dialect reconstruction project
- Online dictionary at biigtigong.com
- Recordings of native speaking elders
- Immersion video for language acquirers
- Advanced language technology

Puzzle Pieces

- Community buy-in, volunteer contributors
- Institutional support
- Full-time lang tech developer JP Montano
- Ojibwe People's Dictionary
- Patricia Ningewance Dictionary
- "Grammar-bridging" from related dialects
- Open source technologies (e.g. FST)

FOCLÓIR **AISTRIÚCHÁN**

B' ann à Beàrnaraigh na Hearadh a bha e. B' e mac Dhòmhnaill MhicAsgaill agus Anna Mhoireasdan a bh' ann. Bha piuthar aige, Effie. Bha aon bhràthair, Alexander, aige a chaidh ainmeachadh air a chlach-uaighe. Chaidh e do Sgoil Bheàrnaraigh eadar 1903 agus 1912. Ann an 1911, bha e agus a phiuthar Effie a' fuireach còmhla ri pàrantan am màthar, Calum agus Effie Moireasdan aig

🔍 Aistrigh »

B'ann à Beàrnaraigh na Hearadh a bha e. B'e mac Dhòmhnaill MhicAsgaill agus Anna
Ba   as   Beàrnaraigh   na   Hearadh   a bhí sé. Ba é mac   Dhòmhnaill   Mhicasgaill   agus   Anna

Mhoireasdan a bh' ann. Bha piuthar aige, Effie. Bha aon bhràthair, Alexander, aige a chaidh
Mhoireasdan   a bhí ann. Bhí   deirfiúr   aige, Effie.   Bhí   aon   bhráthair,   Alexander,   aige a chuaigh

ainmeachadh air a chlach-uaighe. Chaidh e do Sgoil Bheàrnaraigh eadar 1903 agus 1912.
ainmniú   ar a   chlach-uaighe.   Chuaigh sé do Scoil   Bheàrnaraigh   idir   1903 agus 1912.

Ann an 1911, bha e agus a phiuthar Effie a' fuireach còmhla ri pàrantan am màthar, Calum
I   1911, bhí sé agus a dheirfiúr Effie   ag fanacht in éineacht le tuismitheoirí a   máthar,   Calum

agus Effie Moireasdan aig Croit 3, Borbh.
agus   Effie   Moireasdan   ag   Croit 3, Borbh.

Intergaelic.com

- Resources for Irish speakers learning Gàidhlig
- Machine translation Gàidhlig to Gaeilge
- Bilingual dictionary
- Google Translate-like interface
- Twitter stream of Gàidhlig tweets with "help"
- Integrated into Clilstore and Multidict (SMO)
- Manx Gaelic (coming soon!)

Puzzle Pieces

- Core engine written for NEID and repurposed
- Lexical databases shared freely
  - Michael Bauer, Caoimhín Ó Donnaíle, Phil Kelly
- Bilingual corpus, ~1M words in each language
  - Open source translations, community contributors
- "Technology transfer" from Irish to Gàidhlig
- Previous work by Ciarán Ó Duibhín
- Beautiful front-end by Michal Měchura

Common Themes

- Just two out of dozens of examples
- Open source software, open databases
- Language-independent tools, repurposed
- Bridging resources between related languages
- Collaborative approach
  - Community
  - Academia
  - Government
  - Industry

Recommendations

- Government investment in basic research
  - PhDs in linguistics, language technology
- Formalize/recognize community contributions
  - Acadamh um Logánú
- Publicly-funded research must be open source
- Tech companies must support local languages
  - Irish localisation of Gmail as a model
- Cross-language collaboration
- R&D driven by community needs