Translating Facebook into Endangered Languages

Kevin Scannell
Saint Louis University
September 14, 2012

Social media and endangered languages

- Social media: Facebook, Twitter, Google+, blogs, podcasts (audio and video), "content communities"

- 1000's of speakers of 100's of endangered languages are using their languages in social media

- Connect scattered speaker populations, diaspora

- Informal register; use the bits of language you know

- Make up for lack of "traditional" media

- Appealing to kids, the most important demographic!

Indigenous Tweets and Blogs

- Web site I started in March 2011 to help build indigenous language communities on Twitter

- We have found 7.5 million tweets by 46k users in 136 indigenous languages

- In total, roughly 230 languages used on Twitter

- Similar site called Indigenous Blogs tracking indigenous language blog posts

- Per language RSS feeds

"We must never forget that our culture is one of oral history. We will experience a final erosion of our languages if we're not careful and ensure that we have that face-to-face... that we go out on the land with our kids and our grandkids and we have these discussions. That we shut the BlackBerry off, we leave the iPad at home, and sit down by the lake." Chief Isadore Day, Serpent River First Nation

"Indigenous people can choose to resist these [technological] tools and allow the language to continue down its natural pathway to extinction, or embrace and utilize appropriate tools guided by the goals of the community for language documentation, revitalization, promotion and education."
Candace Galla

Facebook for language revitalization

- All major social media sites support Unicode
- Given good fonts and keyboards, one can create *content* in virtually any written language
- Facebook has some advantages for revitalization
- Good support for multimedia
- No 140 character limit
- Clearer "threading" of conversations
- Everyone is already on it!

Facebook's Translation Platform

- Facebook's menus and navigation are translated entirely by users of the site; volunteers

- Voting system to vote down bad translations, vote up good ones

- Ability to translate the site "live", while using Facebook as you normally would

- Sophisticated terminology support, quality checks

- Full translation is massive: tens of thousands of messages, probably more than 100,000 words

Official Translations

- Spanish was the first official translation: Jan. 2008

- Since then, at least partial translations into about 100 languages (I refuse to count "Hacker", "Klingon"...)

- Several appear in the UNESCO Atlas:

- Vulnerable: Aymara, Basque, Faroese, Frisian, Welsh

- Definitely Endangered: Cherokee, Irish, Northern Sámi, Rumantsch

The Problem

- Facebook have not added any new languages for quite a while (spam?)

- This despite petitions, agitation, and their stated goal to "make Facebook available in every language..."

- Same situation with Google's search interface

- Put plainly, neither Facebook nor Google really cares about endangered languages

- Upshot: they have great services, but it's a bad idea to rely on the goodwill of any for-profit company in devising a technology strategy for your language

The Solution, I

- My friend Neskie Manuel came up with a way to take Facebook out of the equation and translate the site into his language of Secwepemctsín

- Neskie passed away tragically in May 2011; I'm carrying this work forward in his honor

- Uses a technology called "Greasemonkey"

- Users can install a bit of JavaScript that "overlays" translations of any site, entirely in their browser

- Game changing idea! No need to rely on Facebook or Google to "allow" us to translate their sites

The Solution, II

- I rewrote Neskie's code to work with Facebook's new design, and generalized to all languages

- To get Facebook in your language, you just have to send me translations of about 200 short phrases

- Small fraction of the full site, but gives a convincing immersive monolingual experience

- Makes translation feasible for languages that have never attempted software translation before

Language Survey

- Overwhelming response: 61 language groups have begun translations, 34 are in a usable state

- Of these 61, previous software translations in just 22

- 40 of 61 are in the UNESCO atlas: 10 vulnerable, 14 def. endangered, 8 severely, 8 critically

- Bantu, Mayan, Australian Aboriginal, Polynesian, European minorities, Native American

- What about your language? Ask me!

Default Languages

- Since such a small fraction of the site is translated, each language group can choose a default "ambient language" from among certain official translations

- Currently available: English (US or UK), Spanish, French, Italian, Norwegian Bokmål, Russian

- Nice for learners and semi-speakers too

- Some languages give their users a choice; e.g. Haitian Creole with default English or French, or Chechen with default English or Russian

Terminology development

- Again, a major goal was to reach out to language groups that had never translated software before

- In such cases, terminology can be a major problem; one has to make up everything!

- I asked each translator to also provide "back translations" into English, for the benefit of others looking for the right term or metaphor

- Great opportunity for endangered language groups to share "different ways of thinking and knowing"

Back Translations Wiki

- These are all available on a wiki that new translators can consult; e.g. under the term "Activity Log":

- French: "Personal history"

- Gundjeihmi: "What have you been doing"

- Irish: "Activity history"

- Nawat: "See what has happened"

- Chichewa: "Frequent activities"

Announcing...

- Facebook in Māori!

- Translation team: Ian Cormack, Teanau Tuiono, Karaitiana Taiuru

- Free to download for Google Chrome, Firefox

- Get it here: http://chrome.google.com/webstore

- Install with a couple of clicks, no restart needed