

Indigenous Tweets, Visible Voices, and Technology
Panel discussion at South-by-Southwest
March 9, 2013

I. The Big Picture

There are 7105 languages spoken in the world today, or at least that's the best estimate we have, from SIL (the Summer Institute of Linguistics), in the latest version of their [Ethnologue](#). Of these, roughly 2500 are endangered according to [UNESCO's Atlas of the World's Languages in Danger](#). And for about 1750 languages, intergenerational transmission has been interrupted; in other words, the language is not being passed on to children, or only very rarely. These languages can die out in as quickly as two generations. No matter how you look at it, the long-term outlook for the world's linguistic diversity is grim: most experts predict around half of the languages spoken today will be gone in the next 100 years, and in some regions: Australia and North America especially that number may be as high as 90%.

There's some good news to report, though, and really that's why we're all here today. Hundreds of groups around the world are working to revitalize their languages and ensure that they're passed on to the next generation -- that they'll survive until 2050, 2100 and beyond. Part of this work involves digital tools: indigenous and minority language speakers are writing blogs, they're tweeting, creating podcasts, youtube videos, etc. - all in their native language. Let's see some examples!

Here's a [blog](#) by my friend and collaborator Edmond Kachale from Malawi, written in the Chichewa language. Chichewa is a Bantu language with more than 10M speakers but it's rarely seen online; access to computers and the Internet aren't widespread in Malawi and the people that *are* online tend to use English.

This is [a site the Inuktitut language](#), spoken by about 35,000 people in Nunavut in Northern Canada. For a relatively small community in a remote area, there's a strong online presence in the language, most of it written in syllabic script, as you see here.

And a final example: this is the home page of the [Yuchi Language Project](#). Yuchi is spoken not too far from here, in northeastern Oklahoma. There are only five fluent speakers remaining, all elders. If you visit their site I think you'll be inspired by the work they're doing to reclaim the language and see that it's passed on to their children. They have a twitter account, youtube videos, and a radio show. It's incidentally the only language in the world that uses the at-symbol as part of its alphabet which presents [special challenges for them on Twitter!](#)

I've had [a project](#) running for more than ten years to crawl the web and gather texts in as many languages as possible - I use these texts in collaboration with native speakers for developing

basic computing resources like spell checkers, keyboard input methods, and online dictionaries. In all, I've found texts in more than 1500 languages on the web. [This image](#) shows a small subset of those languages, organized into a linguistic family tree.

One of the big obstacles we're trying to overcome is the attitude among many people that indigenous and minority languages are somehow "not suited" for computing or for modern life in general, and that English, French, Spanish, etc. are "the languages of the computer" and the languages of modern media.

There are technical and linguistic issues too:

- Terminology for modern concepts must be developed.
- Many indigenous language don't have a writing system at all (most in fact)
- Even when the language is written, speakers may not be comfortable doing so.
- There can be issues with fonts or keyboard input
- I mentioned connectivity and digital divide issues
- And given the lack of commercial interest in these languages, we often have to "go it alone" in trying to overcome all these difficulties.

Finally, I'd like to share [a quote I really like from Lance Twitchell](#) - he's a professor of Alaska Native Languages at the University of Alaska. "As people of nations and cultures, we need to speak our languages. In order to stop them from dying, we only need to speak them: in our homes, to our children, to each other, on our land. It will redefine who we are, and it will be the single largest act of defiance we can make today towards a past that tried to kill us off." What I like is that he talks about culture, language, and land; this is a common theme you hear from native people around the world: CULTURE + LANGUAGE and LAND go hand-in-hand: if you lose one you risk losing the other two as well. In Irish we say "Tír gan teanga, tír gan anam"; "a land without a language is a land without a soul".

II. Irish, Indigenous Tweets, Open Source

I speak Irish or "Irish Gaelic" as it's sometimes known. I do the work I do, and I'm here today not as an academic or researcher but as a member of that language community. Irish had millions of speakers in the 19th century but the number has been dropping steadily ever since.

In 1990, [this book](#) was published proclaiming once and for all the "Death of the Irish language". In fact, the language's obituary has been written many times, even going back to the 1800's.

The problem is, we're still here. This is a [tweet](#) from just a few weeks ago: "Tá an torramh fós ar siúl" means "The funeral is still underway!"

In fact, there's a vibrant, healthy, growing community of speakers, including many young people. This is a graph of the number of tweets per week in Irish. It certainly doesn't look like a dead language to me. That big spike is St. Patrick's Day, by the way! It's a special annual event we call Seachtain na Gaeilge = Irish language week.

The problem with minority languages is that they can seem "invisible". Minority language voices are easily drowned out by the millions of people who speak English and other major languages.

Two years ago I created a web site called [Indigenous Tweets](#) to try and raise the profile of indigenous and minority language groups on Twitter, to make it easier for people to find other members of their community, and to encourage people to tweet more in their native language. The site was launched with 35 languages and we now track almost 150. These range from Haitian Creole and Basque with around 14,000 tweeters each, to 32 languages that have just a single person tweeting! There is a similar site aimed specifically at Basque called [UMAP](#).

This is a screenshot of [the Indigenous Tweets page for Irish](#). For each user it has stats on number of tweets in the language, percentage of tweets in the language, number of followers, and then Trending Topics specific to the language.

Using the data from Indigenous Tweets, I created [this visualization](#) of the Irish language community on Twitter, with two users linked when one mentions the other in a tweet. Usually these correspond to conversations. What you see a tightly knit community, and an active one, with tens of thousands of conversations.

We've even seen entirely new channels of communication emerging on Twitter. As an example, here's [another visualization](#) that shows both the Irish and Scottish Gaelic Twitter communities, Irish on the left and Scottish Gaelic on the right, in orange. These are related, but mutually unintelligible languages, comparable to Spanish and Catalán. You can see a handful of bilingual people in the middle, but also an amazing number of "direct" conversations between the two languages. I don't know of any other context in which this kind of bilingual conversation is occurring between the two communities.

Translating software and creating online content is a critical part of language revitalization. It's certainly not *the* most important thing but it's one important piece in a very big puzzle. If young people never see their native language online, it only reinforces the attitudes I mentioned earlier that "English is the language of the computer" and that their native language is somehow "useless" in the modern world.

I started translating Firefox into Irish in 2003 with a group of friends; we've done translations of OpenOffice (now LibreOffice), KDE, and lots of other open source packages since then. It's

now possible to live your computing life entirely through the medium of Irish. What I want to emphasize here is that an open source approach is absolutely essential for indigenous languages, given the lack of commercial interest. We don't have to ask permission from Microsoft or Apple to translate their software, and the results "belong" to the community and not to a big software company.

The organization [Mozilla México](#) has done incredible work over the last three years providing translations of Firefox into *ten* indigenous languages of Mexico. We'd hoped to have Julio Gómez Sánchez, the leader of that effort, here to talk a bit about their work but he was unable to make it at the last minute. There are Mozilla translation teams in almost 120 languages now, and these ten from Mexico are the first translations into any indigenous language of the Americas. I'm hoping to see Cherokee, Navajo, Ojibwe, or other languages from the US and Canada sometime soon.

Open source is great, but we're at a something of a crossroads at the moment. More and more of the software we use is "Software as a service": Gmail instead of Thunderbird, Google docs instead of LibreOffice, and of course Facebook, Twitter, etc. In these cases, the big companies behind the services are in complete control of the linguistic landscape online. Google and Facebook have crowd-sourced translation platforms - all of the translations are done online by volunteers - the problem is that neither company is letting any new languages in. A translation of something even as simple as Google's search interface has huge symbolic value for indigenous language groups; but these groups have been effectively "locked out" and there's no sign that this will change any time soon.

About three years ago, my friend Neskie Manuel created a browser add-on based on Greasemonkey to translate Facebook into his native language of Secwepemctsin - a First Nations language spoken in the interior of British Columbia. Since Facebook wasn't allowing new translations onto their site, Neskie's idea was just to take them out of the equation entirely. Tragically, Neskie passed away in 2011 at just 30 years of age, not long after [releasing his code on github](#). I decided to continue work on the Facebook project in his honor. I generalized things so in theory any group can now provide translations and pretty quickly have Facebook up and running in their language. The response has been overwhelming: 62 indigenous languages have started translations and many of them are already available for download from userscripts.org and the Google Chrome store. This is a screenshot of the [Facebook translation in K'iche'](#). My hope is that before long we'll have more unofficial translations than Facebook has official ones!

That's all the time I have. Go raibh míle maith agaibh! Thank you all!