Endangered languages in social media

Kevin Scannell
INNET Summer School
Gniezno, Poland
5 September 2013



CEANÁIN

Fotha Nuachta

Teachtaireachtaí

31 Ócáidí

Grianghraif

GRÚPAÍ

Citizen Media and...

11 African Languages

Gaeilge Amháin

Resource Network... 20+

11 NAACLT

Dunwoody Class o... 1

12 GaelSpell

Cruthaigh Grúpa...

FEIDHMCHI ÁIR

👔 Ionad na bhFeidhmchlár

Aistriúcháin

Fotha Cluichí 20+

LEATHANAIGH

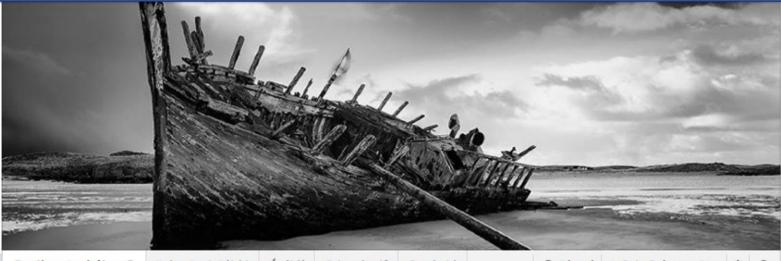
Pages Feed 20+

Leathanaigh is ma... 20+

Cruthaigh Fógra

TUILLEADH *

Tá an comhrá múchta. Chun comhrá a bheith agat le do chairde quir



Gaeilge Amháin 🚱

Eolas Faoi/Fúithi

Ócáidí

Grianghraif

Comhaid

⊘ Fógraí + Cuir Grúpa ar Bun

☆ Q

20+





Scríobh rud éigin...

TEACHTAIREACHTAÍ LE DÉANAÍ

Gearóid Ó HAnnaidh



Thóg mé griangraf den fhógraíocht 'Guinness' seo i nDún na nGall le déanaí.

An bhfaca sibh fógraíochtaí eile mar sin i nGaeilge?



Is maith liom · Déan trácht · Follow Post · Roinn · 13 uair ó shin

Is maith le 3 duine é seo.

Maitiú Ó Coimín Rud eicínt mar sin ar fud Chathair na

2.676 ball (134 nua) . Tabhair Cuireadh Trí Ríomhphost

+ Cuir Daoine Leis an nGrúpa

Grúpaí Molta

See All



Díograiseoirí na Gaeilge

Tá Brian Ó Conchubhair agus 79 cara eile páirteach ann





Cumann Gaeilge na hAstráile Iris...

Conradh Na Gaeilge joined





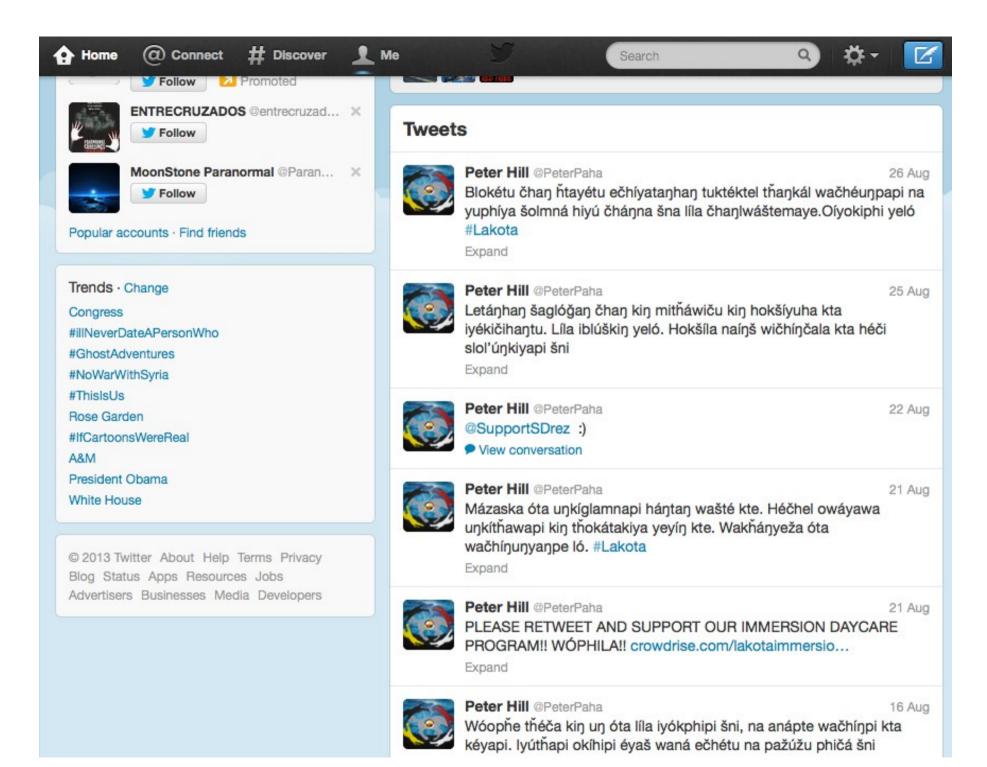
Folk Club Plancoet

Tá Aisteoirí le Gaeilge agus 8 cara eile páirteach ann

LE Glac Páirt sa Ghrúpa



72% Le Comhrá (múchta)





▼ hasfd∩sbD∩

Dsb/sb LD^b

A/A**JJOb Fch/45b

A* L<ACsb

Printable version

Permanent link

Page information

Data item

Cite this page

Ö

▼ ρ⇔Δρ Qafár af Ar₀cwea Acèh

Afrikaans

Akan

Alemannisch

L<∧C⁵b ▷ Search Read View source View history Search Q

$\mathsf{d}\mathsf{\Gamma}^\mathsf{\mathsf{Gb}}$

The interface and content is by default mostly in the Inuktitut syllabics. If you want to view it in the Latin script, you have to create an account or log in, go to Special:Preferences, select "ike-latn - inuktitut" in the drop-down box(es) and click on "Save". more info...

Δημα^ω οθημηση, Δημα^ω οθημηση.

ΔΦΦΩΥ ΡΔΡΛΩΦ ΛΓΕ 391 LΥΛΟΘ

DYSDSO DER DAPAND SOSSDSSO COSCOS

Coebo VD4ep TeVCep

baCΓ ΔΔΔ' ▷'b▷/°Γ' - ▷'b▷/Cnσ'Γ αΔαΔ'*C▷/'∜)> baCΓ ΔΔΔ' ▷'b▷/°Γ' Eskimo-Aleut ▷'b'&▷∩\' ΔαΓ' «language family» ΔΔ



APSID LIACID

Φροφονι Δηγοσιυσω ρερς Γρεγαώ

Watinkhani ta Luso ndi Umisiri

Tuesday, January 17, 2012

Chaka Chatsopano kwa Nonse

pa 11:23 Wolemba Edmond Kachale

Ndikulandireni mwapadera abale nonse amene mumakonda kuwerenga nkhani pa bwalo lino. Ndikukhulupirira kuti mwachimaliza bwino chaka cha 2011, ndipo mwayambapo kukwaniritsa ena mwa malingaliro amene muli nawo a mchilumika chino cha 2012.

Chaka cha 2011 chinali chopambana ku chiyankhulo cha Chichewa chifukwa pali zinthu zingapo zimene zinachitika zomwe zakupititsa patsogolo kugwiritsidwa kwa Chichewa pa utatavu ndi m'kopsuta. Mwa zina, mphunzitsi wa sukulu ina ya ukachenjede ku Amereka, a Kelvin Scannell, anatsegula madambwe awiri: http://indigenoustweets.com ndi http://indigenousblogs.com omwe akupititsa patsogolo kugwiritsidwa ntchito kwa ziyankhulo zachimidzimidzi pa utatavupa. http://indigenousblogs.com ndi ndi ndala chabe la indigenoustweets.com/blogs.

Kwa amene sanamvepo za madabwewa, tingonena mopaza. indigenoustweets.com imatolera mitongoliro pa Twitter kaundula wa akadyolidyoli a ziyankhulo zosiyanasiyana. Apapa dzina loti akadyolidyoli silikutanthauza akavuwevuwe osokoneza, koma awa ndi anthu amene amakhala akutumiza timitu tankhani (mitongoliro) pa Twitter kuthandiza enafe kumva zochitika mmaderamu. Momwemo, dambwe la http://indigenousblogs.com limatolera zibaluwa zolembedwa mziyankhulo zimenezi. Dambwe limeneli ndi njira yothandiza kuti mudzitha kupeza nkhani m'ziyakhulo ngati zathuzi mosavuta. Mukudziwa kuti nkhani zambiri pa utatavupa zimakhala m'Chingerezi, m'Chifalaansa ndi ziyankhulo zina zomwe zili ndi mphamvu pa chuma ndi chitukuko. Nkhani za m'Chichewa zimapezeka pa http://indigenousblogs.com/ny (kapenanso ulalo wotanimphirako: http://indigenoustweets.com/blogs/ny). Ndili wokondwa kuti mwa ziyankhulo zina zomwe zilipo, Chichewa chinali chimodzi mwa zoyambirira kuyikidwapo pa m'ndandanda owoneka nawo pa madambwe amenewa.

Followers





Edmond Kachale

I don't know how to break a record. All I know is how to build it. Now, I let others break mine!

View my complete profile





Language Revitalization

- I am not a linguist!
- I come to this work as an Irish speaker
- Began working to help my own community 1999
- Proofing tools, dictionaries, software translation
- Tools to help languages flourish online
- Have developed resources for 150+ languages

Social media: advantages

- Often solves the "critical mass" problem
- Connects learners and native speakers
- Informal medium; welcoming for learners
- Short bursts: use the words, phrases you know

Social media: disadvantages

- Written medium only
- Anxiety about spelling and grammar
- Difficulties typing the language (diacritics, etc.)
- Shyness in a public forum, w/ fluent speakers
- Impossible to immerse oneself in the language

Some Perspective

"The internet and digital world cannot save us. They cannot save Indigenous languages. Of course these things have benefits but they are not the Messiah. We don't need another website or DVD or multi-media application, these are short term, quick fix solutions. What we really need is sustainable initiatives, to create opportunities for Indigenous language users to communicate with each other in their native tongue. To get people speaking again."

John Hobson, University of Sydney, 16 May 2013

http://au.artshub.com/au/news-article/news/arts/digital-not-always-the-answer-195370

Social media and linguistics

- Many of you are linguists interested in research
- "Live" examples of language in use
- "Real" language; not Bible translations
- Readily accessible text, often in great quantity
- Allows social network analysis

Disadvantages

- Privacy and consent
- Facebook, Twitter terms of service
- Written language only
- Informal language, code-switching
- Non-standard spelling, encodings, other "noise"
- Tweets, e.g., represent a "sublanguage"

Demo of Twitter API

- Twitter allows developers access to tweets
- Streaming API lets you follow tweets live
- Can follow set of users, words, hashtags
- \$ perl live.pl się
- \$ perl live.pl @BarackObama
- \$ perl live.pl '#Syria'

100 million Twitter project ideas

- Offers a host of possibilities for ling. research
- Regional dialect analyses (geocoded tweets)
- Compare dialect and social network (more l8r)
- Measure relative vitality of minority languages
- Twitter as a sublanguage; hashtags, etc.
- Code-switching in tweets
- Lexicography

Tweetálaithe

Indigenous Tweets

- http://indigenoustweets.com/
- Launched March 2011
- Community-building on Twitter
- Tracks all tweets in 149 languages
- 61828 accounts, 12+ million tweets
- 33 languages with only a single tweeter

Language	\$	Users	\$	Tweets	•	Top User	\$	Tweets	\$	First Tweet	\$
Euskara		17042		4629179		berria		43464		eastigarraga	
Kreyòl Ayisyen		14267		3801664		RoodHT509		49006		tichrist	
Cymraeg		14218		2461608		newyddcymraeg		57703		meigwilym	
Kapampangan		1379		1560984		itsmeshellie		19449		desperada	
Frysk		2048		571851		omropfytsban		61042		vogelzang	
Setswana		314		564575		sesutho		30409		WameDre	
Gaeilge		7388		472825		aonghusoha		24336		imeall	
Asturianu		749		254257		iyangc		16439		Pingarates	
Hausa		300		236400		bbchausa		20128		Richee_Rich	
Soomaaliga		558		141671		Weedhsan		15895		HaPpYMaXaMeD	
Gàidhlig		652		93423		sconewt		27303		ruighean	
Māori		335		89391		maonewt		29810		DigitalMaori	
Gaelg		26		85533		glvold		40479		greinneyder	
Chamoru		9		61894		chanewtest		25699		jenccamacho	
Tamasheq		3		57363		tmjnew		27484		tmjbbl	
Aragonés		80		54292		Purnas		13655		jesmar	
de		25		50931		mvblogs		9788		muraasil	
Winaray		33		40969		ilovetacloban		6665		piamaemendz	
Malagasy		252		36356		Ainjah		4450		elsifaka	
Chicheŵa		142		32994		skinnybrizzo		6642		caseynesbit	
ייִדיש		260		31102		jewish_wire		11553		ZackBergerMDPhD	
Occitan		98		29001		occitania		8736		sbagudanch	
Cántabru		13		28523		Xoselnaciu		5786		r_macho	
Emiliàn e Rumagnòl		53		28078		ingiro_concerti		2339		Ze_Beps	
Tshivenda		25		26892		mirimat		8875		its_tshima	

Indigenous Tweets.com

Indigenous Blogs

Kevin Scannell

© ① ①

EY SA





\$	Úsáideoir	\$		Gaeilge 🕶	lomlán ¢	% Gaeilge	•	Leantóirí \$	Á Leanúint [‡]	+	Giolc Is Déanaí 💠
1	aonghusoha			24336	28340	85.9		788	536		2013-08-23 21:04:23
2	Tuigim		1	13316	19090	69.8		833	981		2013-08-23 23:04:57
3	maitiuocoimin		O	8162	10277	79.4		676	990		2013-08-24 00:38:22
4	KingDonncha1			7010	8699	80.6		255	181		2013-08-24 04:13:50
5	murchadhmor			5892	7196	81.9		1006	1686		2013-08-23 08:24:43
6	Foinse		Foinse !	5454	6199	88.0		4407	4325		2013-08-23 15:55:59
7	TG4TV		TG 4	4979	8549	58.2		8184	2527		2013-08-24 08:55:53
8	Tuairisceoir		T	4891	5900	82.9		3356	3684		2013-08-23 16:15:31
9	SachiEika_Bot		4	4811	85341	5.6		541	293		2013-08-24 05:13:12
10	GAA_BEO		GAA BEO	4340	79142	5.5		16703	2401		2013-08-23 20:41:08
11	Gaelport		g	4121	6036	68.3		3092	2508		2013-08-23 11:29:28

INDIGENOUS TWEETS.COM

Gaeilge

I mbéal an phobail:

- #ilaughed

- blicht#igcomhrálemairtín

Daoine ar iarraidh? Ainm Twitter:





Inis don domhan go bhfuil tú

Indigenous Blogs

Kevin Scannell

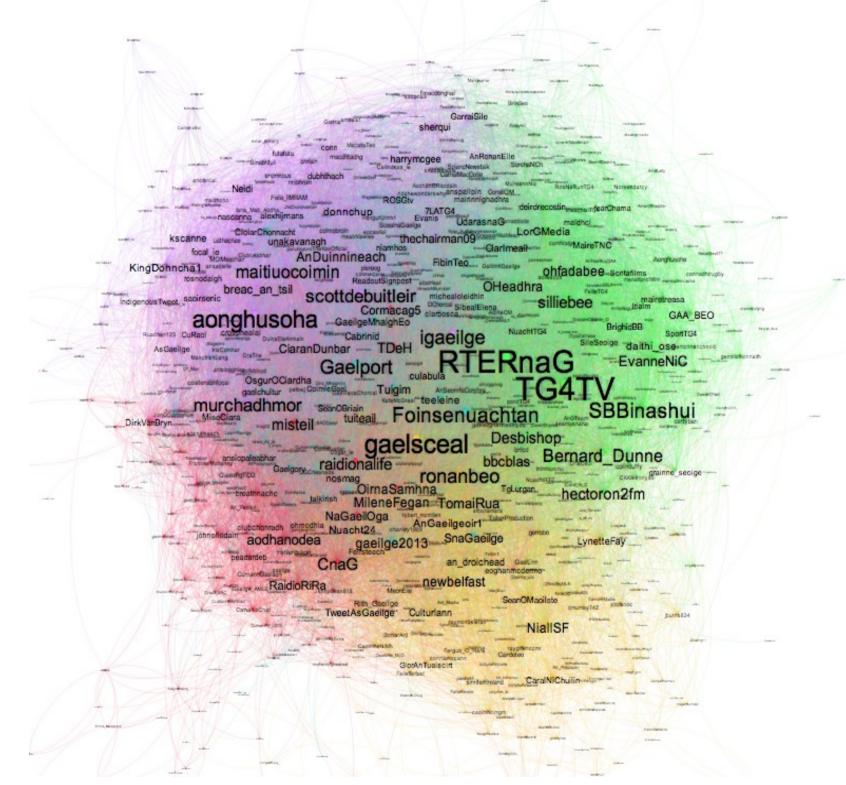


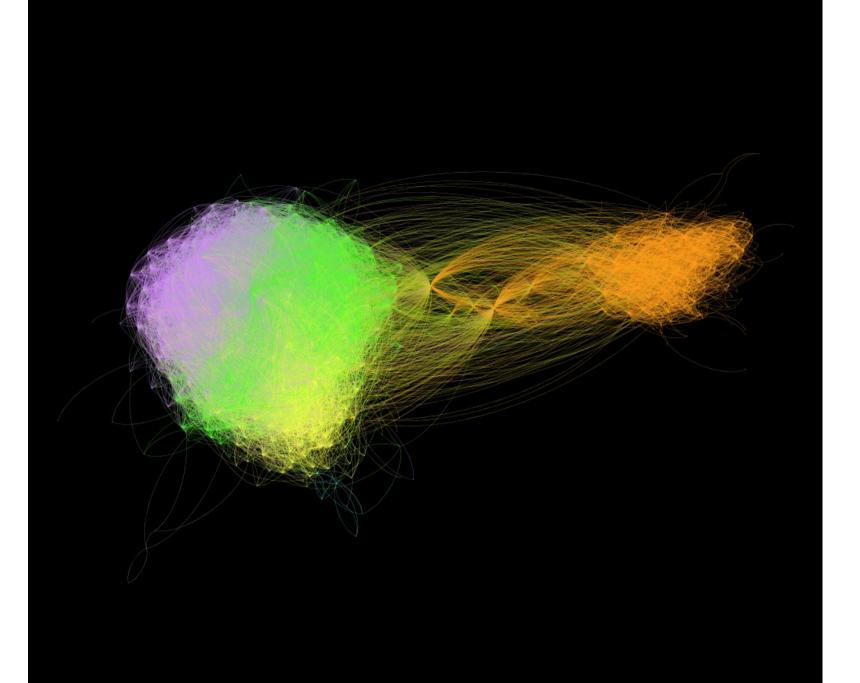




Network analysis

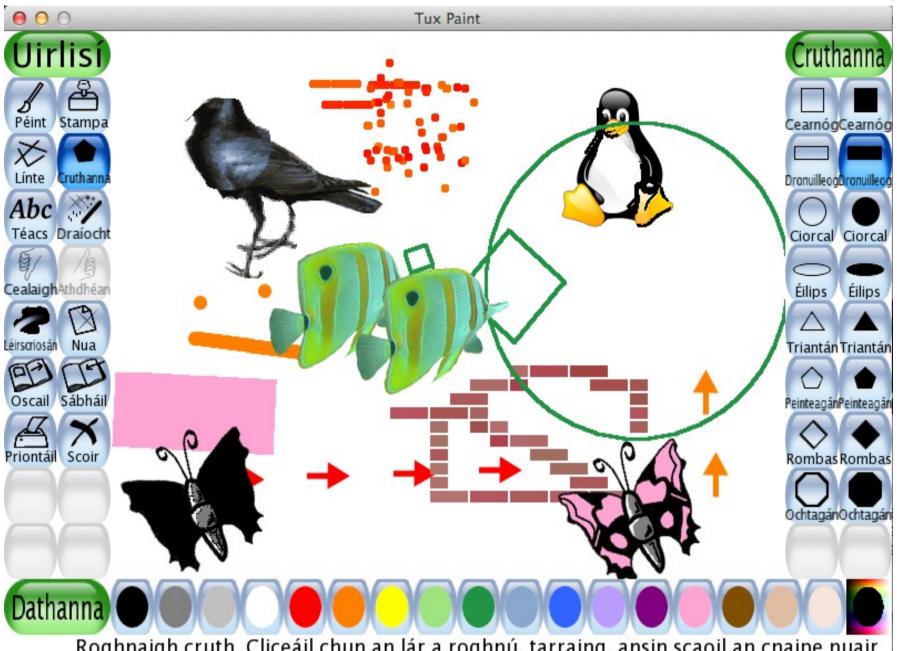
- Twitter contains a couple of social graphs
- Nodes correspond to Twitter accounts
- The "follow graph"
- The "mention graph"
- Both are directed graphs





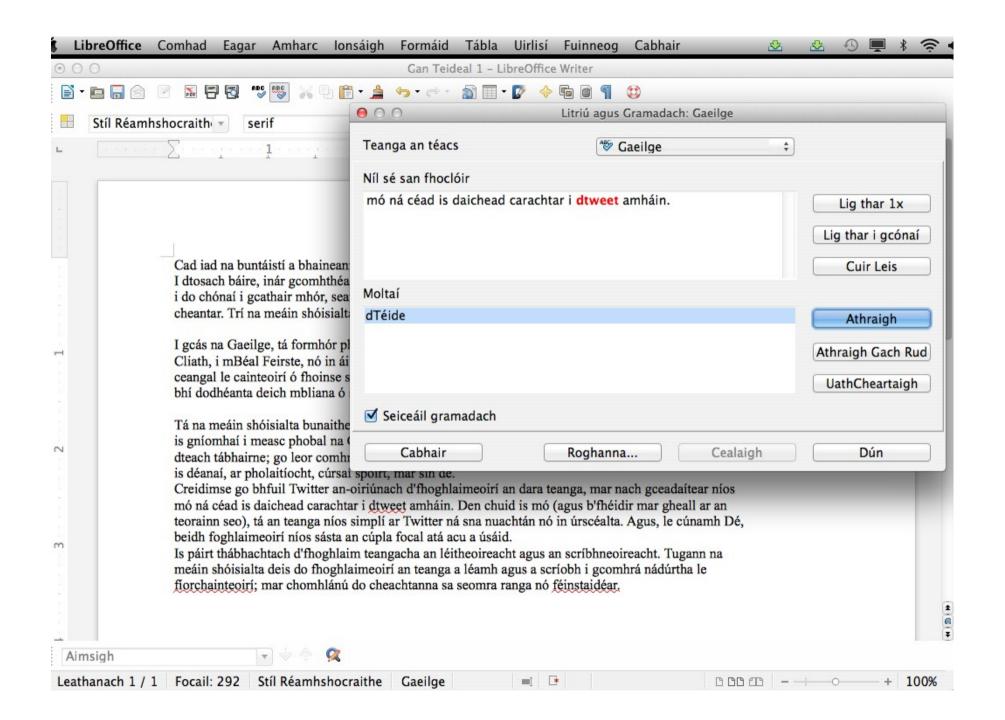
Software Localization

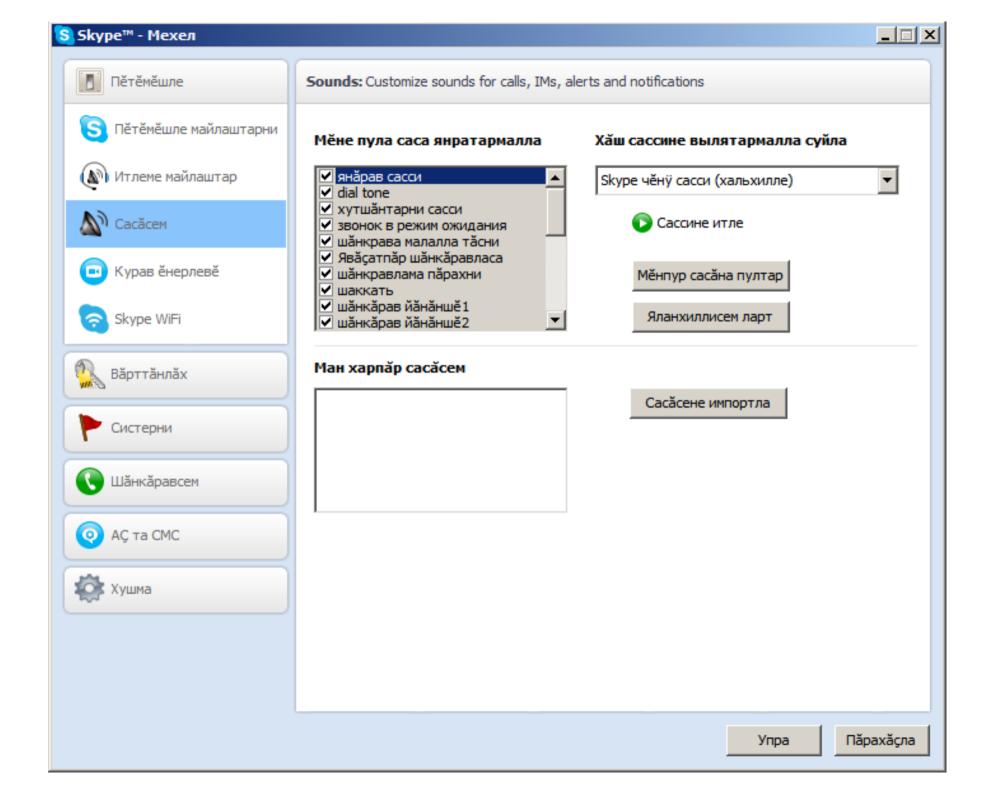
- Translate menus, navigation, etc.
- Critical for language revitalization
- Otherwise, English is language of the computer
- Hopefully influences people's language use too



Roghnaigh cruth. Cliceáil chun an lár a roghnú, tarraing, ansin scaoil an cnaipe nuair atá an méid socraithe agat. Bog chun é a rothlú, agus cliceáil arís chun é a dhearadh.







Software as a Service

- Above examples are standalone software
- Also, excepting Skype, open source!
- More and more software is online-only: "SaaS"
- Gmail, Google Docs, Facebook, Twitter, ...
- Need special permission to translate these sites

Facebook translations

- Facebook has about 100 "official" translations
- Almost all "crowd-sourced" translations
- Translate "live", with a voting system for QA
- Spanish was the first non-English locale, 2008
- A few endangered languages were "lucky"
- Irish, Welsh, Basque, N. Sámi, Cherokee, ...
- No new languages added in 2+ years
- This despite great demand, petitions, etc.

Solutions?

- Pressure from each group individually?
- Each lang works their contacts "on the inside"?
- Legislation, rights-based approaches?
- Develop your own social network?
- Don't let your kids onto Facebook?

Secwepemc Facebook

- Idea of Neskie Manuel, orig. for his language
- Greasemonkey scripts; a kind of "overlay"
- Runs in your browser, no permission needed
- Translations into 31 languages ready for use
- 25 other languages underway
- Majority are endangered
- Majority have never translated any software
- Goal: more unofficial than official translations!

Sharing Metaphors

- "Computer" (sorter, lightning brain)
- "Link" (bridge)
- "Domain" (secret initiation camp)
- "Cookie" (cuckoo's egg)
- "Cache" (leftovers)
- "Keyboard" (fingerboard vs. musical instrument)
- "De-interlacing" (de-cross-weaving)

Terminology Wiki

- Most languages have had to coin new terms
- Technical terms ("mobile upload", "cookie")
- Facebook-specific ("like", "poke", "news feed")
- Western concepts ("privacy", "advertising", etc.)
- http://github.com/kscanne/secwepemc-facebook/wiki/Back-Translations
- Please add your own!!

Chichewa



Kriol



Manx Gaelic



K'iche'



Workshop: Gephi

- http://www.gephi.org/
- Sample data; mention graphs from Twitter
- Māori, Gàidhlig, Welsh, Chichewa, Irish, Basque, Asturian
- http://borel.slu.edu/obair/gdf.zip