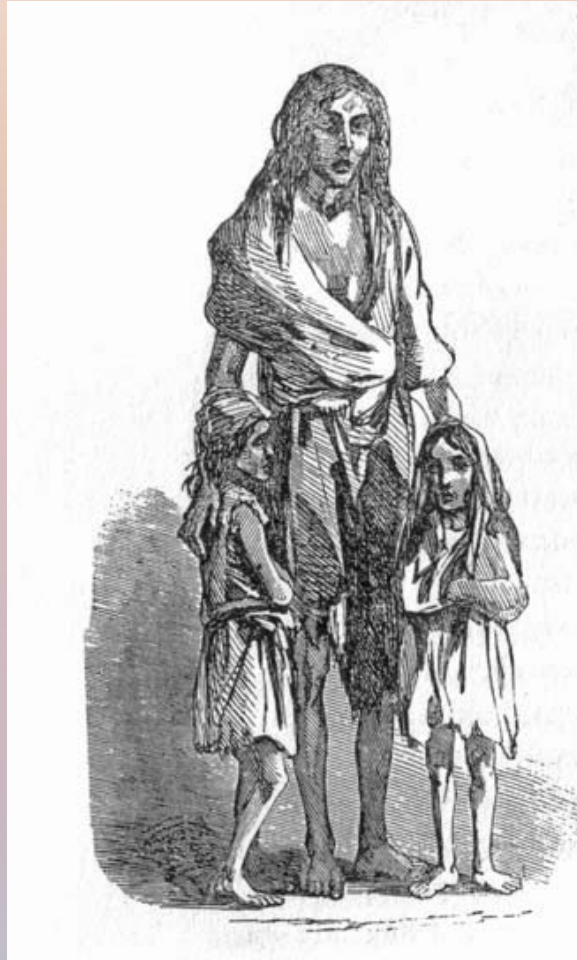


Turning the Tide: Terminology Creation and Free Software in Irish

Kevin Scannell
Saint Louis University
October 11, 2010

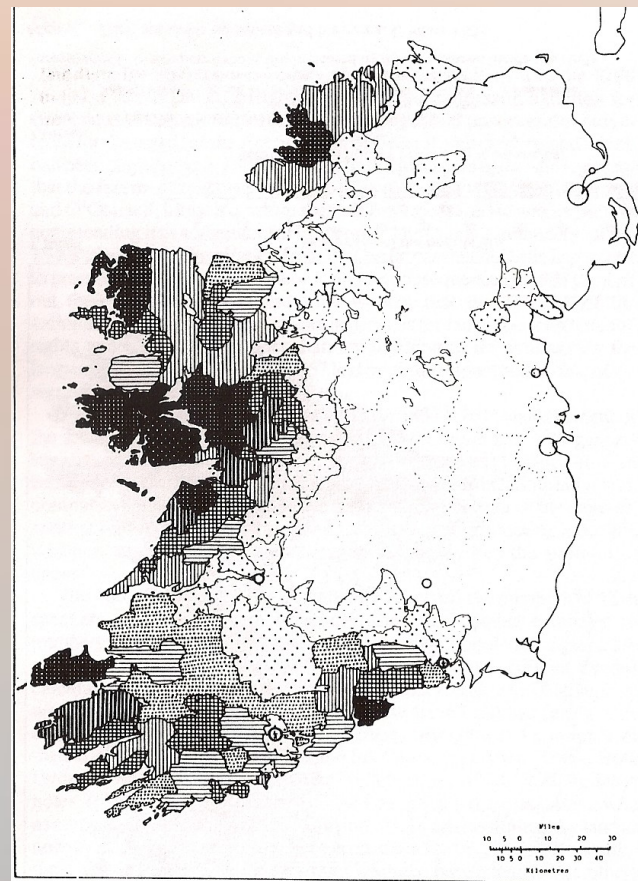
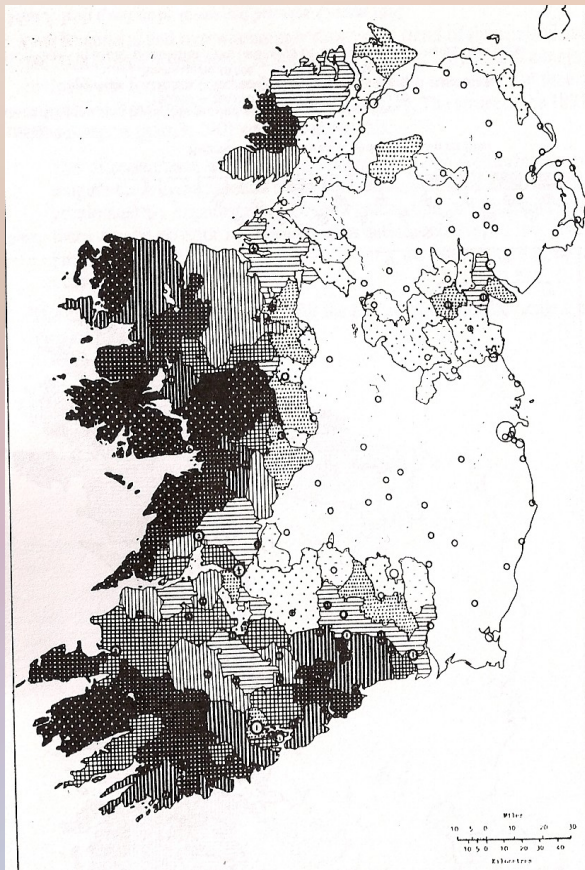
The Choctaw Irish Gift



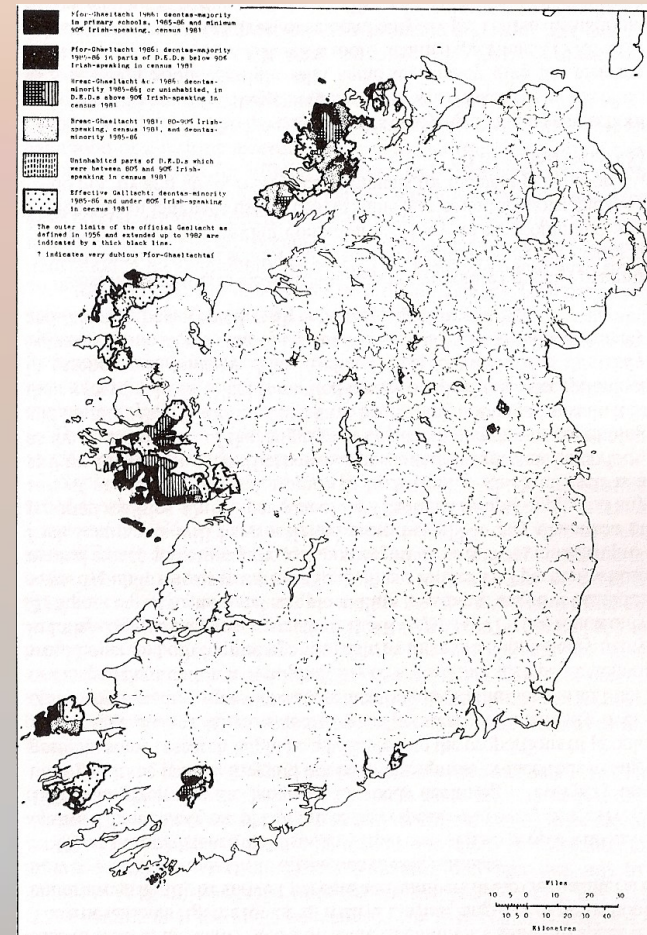
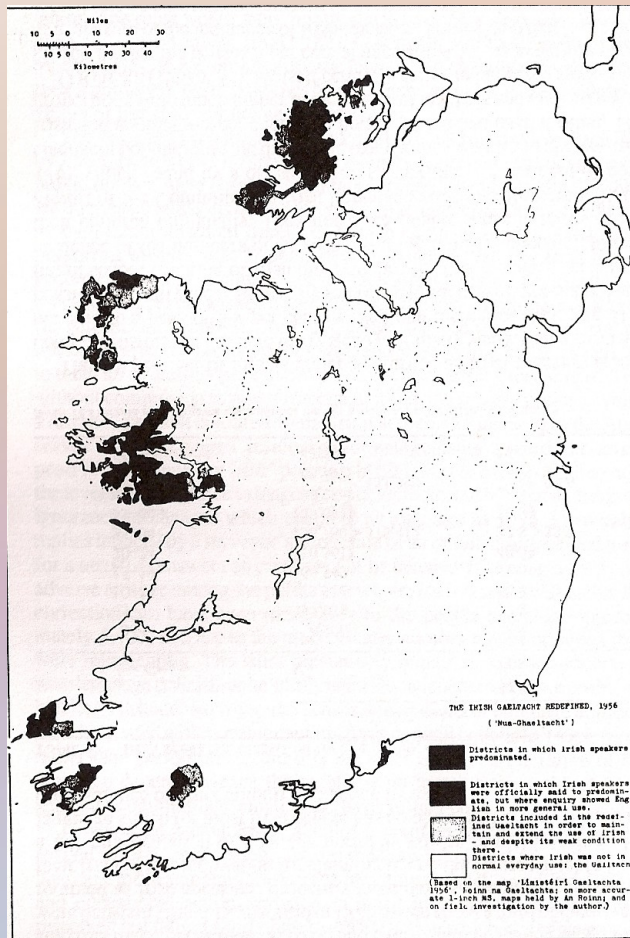
The Irish Language

- Wait, don't they speak English in Ireland?
- Sometimes called “Gaelic” (incorrectly)
- The indigenous language of Ireland
- Part of the Celtic branch of Indo-European

Na Gaeltachtaí: 1851, 1891



Na Gaeltachtaí: 1956, 1987



The Future

- A government-sponsored study in 2007 predicted that Irish will die out as a community language by 2027
- Even in the strongest Gaeltachtaí, there isn't a critical mass raising their children in Irish
- Only about 20,000 native speakers left
- English is everywhere, on signs, in shops, school textbooks, TV, radio, etc. Is immersion impossible?

Virtual Gaeltacht?

- Surface goal: a fully-Irish computing system
- Deeper goal: provide people an environment where they are comfortable socializing, writing, and being creative entirely in their native language
- Computers are tricky – it's fundamentally a written medium, and many native speakers aren't comfortable writing in the language, and they don't always know the “correct” computing terms

What's in a “fully-Irish” system?

- Desktop software: web browser, office software, email, games, multimedia players
- Web sites: search engines, social networking sites, blogs, Wikipedia
- Language tools: spell checker, grammar checker, online dictionaries, thesauri, translation aids

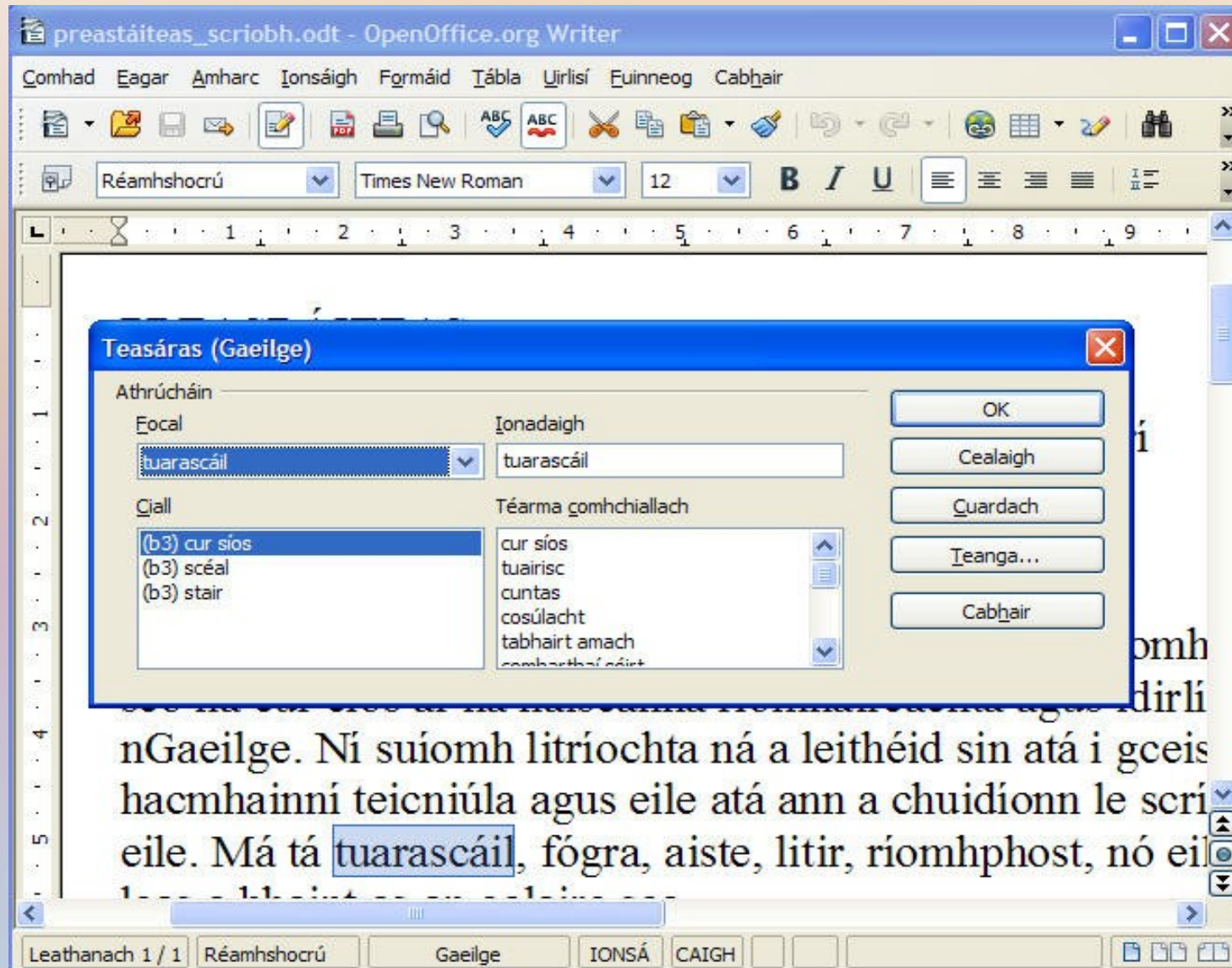
Ar son na cúise

- Purely volunteer effort, all at zero cost
- Everything we produce we give away for free
- About 10 contributors, only two active translators
- This is encouraging, not depressing! A small number of dedicated people can do a lot: 100+ packages, 2 million words translated since 2001

What is Free Software?

- Sometimes called “open source”
- Fewer bugs, more secure
- “Free beer” vs. “free speech”
- Linux, OpenOffice.org, Mozilla Firefox, thousands of other software packages
- Ecosystem that future programmers can build on

OpenOffice.org, with thesaurus



Irish Facebook in Irish Firefox

The screenshot shows the Irish Facebook website in the Irish Firefox browser. The browser's address bar displays the URL <http://www.facebook.com/group.php?gid=114129875265581>. The page title is "Facebook | GaelSpell". The browser's menu bar includes "Firefox", "Comhad", "Eagar", "Amharc", "Stair", "Leabharmharcanna", "Uirlisí", "Fuinneog", and "Cabhair". The browser's toolbar shows various icons for navigation and search. The Facebook page itself is in Irish and features a search bar, navigation tabs for "Baile", "Próifil", and "Cuntas", and a profile picture of Kevin Scannell. The main content area displays a post by Micheál Johnny Ó Meachair about a dictionary project. The right sidebar contains a navigation menu and a public administration advertisement from South University.

Firefox Comhad Eagar Amharc Stair Leabharmharcanna Uirlisí Fuinneog Cabhair

Facebook | GaelSpell

Priority Inbox - kscanne@g... 31 Google Calendar Corpus building for minori... Facebook | GaelSpell

http://www.facebook.com/group.php?gid=114129875265581

Google Gmail Banner aimsigh KDE FF OOo Leabhair Teangacha CS Leabharmharc...

facebook Cuardaigh

Baile Próifil Cuntas

Kevin Scannell

Cuir Cairde In Eagar
Socruithe Cuntais
Socruithe Príobháideachais
Socruithe Feidhmchláir
Ionad Cabhrach
Logáil amach

Seol Teachtaireacht Chuig Gach Ball
Cuir an Grúpa chun Cinn le Fógra
Athraigh Socruithe Grúpa
Cuir Bail in Eagar
Tabhair Cuireadh do Dhaoine
Cruthaigh Ócáid Ghrúpa

Scríobh rud éigin faoi GaelSpell.

Eolas

Catagóir:
An tIdirlíon Teicneolaíocht - Bogearraí

Cur Síos:
Is litreoir cumhachtach é GaelSpell atá ar fáil saor in aisce le deich mbliana anuas. Tá níos mó ná 332,000 focal ina bhunachar sonraí faoi láthair, agus focail nua ag dul isteach ann i gcónaí.

Cineál Príobháideachais:
Ar Oscailt: Tá gach rud os comhair an phobail.

GaelSpell

Balla Eolas Grianghraif Fiseán Ócáidí +

Scríobh rud éigin...

Ceangail: Comhroinn

Roghanna

Micheál Johnny Ó Meachair tá roinnt mhaith focail ar aidiachtaí iad, agus tá fadhb aige ríomhchlár nuair a ndéantar ainmfhocail astu. Mar shampla...
iolraithe - iolraitheoir [taispeánann sé "iolraithe oir" mar cheartúchán air]
ar 16 Bealtaine ag 11:58 · Déan trácht · Ní maith liom é · Tuairiscigh

Is maith leat é seo.

Micheál Johnny Ó Meachair agus, tugtar "idirbheartaíocht" mar cheartúchán ar idirbheartaíochtaí ...mar eolas
Tá samplaí cosúla eile dó seo ann chomh maith.
ar 16 Bealtaine ag 12:10 · Is Maith Liom É

Kevin Scannell GRMA as "iolraitheoir". Tá "idirbheartaíocht" ann cheana, san uimhir uatha amháin - aontaíonn sé seo le focal.ie mar shampla. "idirbheartaíocht" = "negotiations", b'fhéidir go mbaintear úsáid as an uimhir iolra faoi anáil an Bhéarla.
ar 20 Bealtaine ag 11:00 · Is Maith Liom É

Micheál Johnny Ó Meachair mo leithscéal muna raibheas soiléir faoi, cuireann sé spás roimh an "aíocht" sa leagan atá agam... agus deir sé gur botún é gan an spás sin. Bainfidh mé triail as an leagan nua, brb!
ar 21 Bealtaine ag 06:42 · Is Maith Liom É

Scríobh trácht anseo...

79% Comhrá (As Líne)

Aimsigh: Fami Ar Aghaidh Siar Aibhsigh gach ceann Meaitseáil an cás

Sonraí n-aistriú ó creative.ak.fbcdn.net...

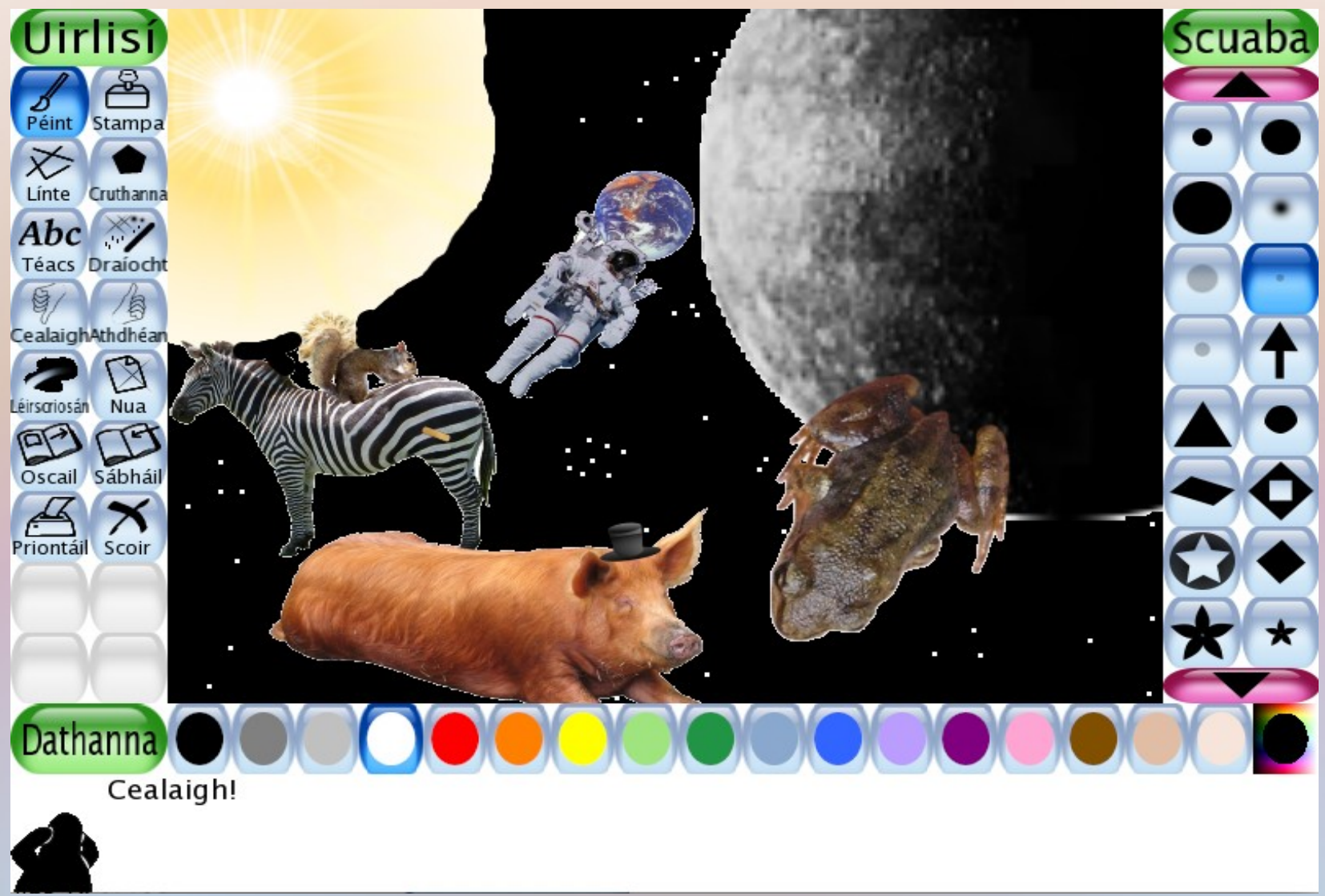
Travel with THE pioneer in Amazon River adventure!
Enjoy an award-winning 10-day small-ship expedition from \$3,298. Request a brochure.

Public Administration
olp.southuniversity.edu

South University
Established 1999

Get more with a public administration degree from South University. Flexible, online programs fit your life.

TuxPaint: drawing program for kids



Community Ownership

- Free software is not controlled by any one company
- Irish translation of Apple OS in the 1990's
- Google search interface, gmail
- “Never beg to save the language” - what we translate is ours and can't be taken away at the whim of a for-profit company. Self-determination.
- “In a way, we are anarchists” Abandon bad systems!
- Free software offers the only viable long-term strategy for languages with limited resources

Corpora

- Modern language technology (spelling and grammar checkers, machine translation) is based on statistics
- Need large databases of text, monolingual or bilingual, to use these approaches
- I am running a web crawler that is gathering texts for almost 500 minority/indigenous languages around the world
- <http://borel.slu.edu/crubadan/>

“Pick a point on the globe, and the picture's the same”

- We (speakers of minority languages) *must* share tools and resources
- Crawler was originally developed for Irish, but now I'm providing data to dozens of language groups around the world
- 25 new open source spell checkers created since 2004, in collaboration with native speakers
- DEMO

Terminology: Guiding Principles

- First, use or reuse native elements when possible
- Second, learn from other indigenous languages
- Third, we can be better/clearer than English
- Fourth, our primary audience is native speakers, but we strive to be understood by semi-speakers too

Native elements

- *lann* is an old word meaning “house” or “floor”
- Used in combinations *leabhar+lann*, *amharc+lann*
- *comhad* is a paper file, hence *comhad+lann*
- *ríomh* means to count, hence *ríomhaire*, *ríomheolaí*, *ríomhphost*, *ríomhlann*, etc.
- Or “cookie” = *cuach* or *ubh chuaiche*

Learn from other languages

- Many other languages are creating terms too: share!
- Chichewa, spoken in Malawi (E. Kachale)
- A directory is *chikwatu* (mixed bag)
- A cached page is *mkute wa tsamba* (page leftovers – *mkute* is leftovers of *nsima*, a Malawian staple food)
- A URL is *ulalo* (a bridge across a river)
- Domain = *dambwe* (a secret initiation camp, entry by secret phrases)



- Firefox addon that lets you type special characters without a special keyboard or extra keystrokes
- Another tool that was first developed for Irish
- Now supports 114 languages
- Created together with Michael Schade
- DEMO