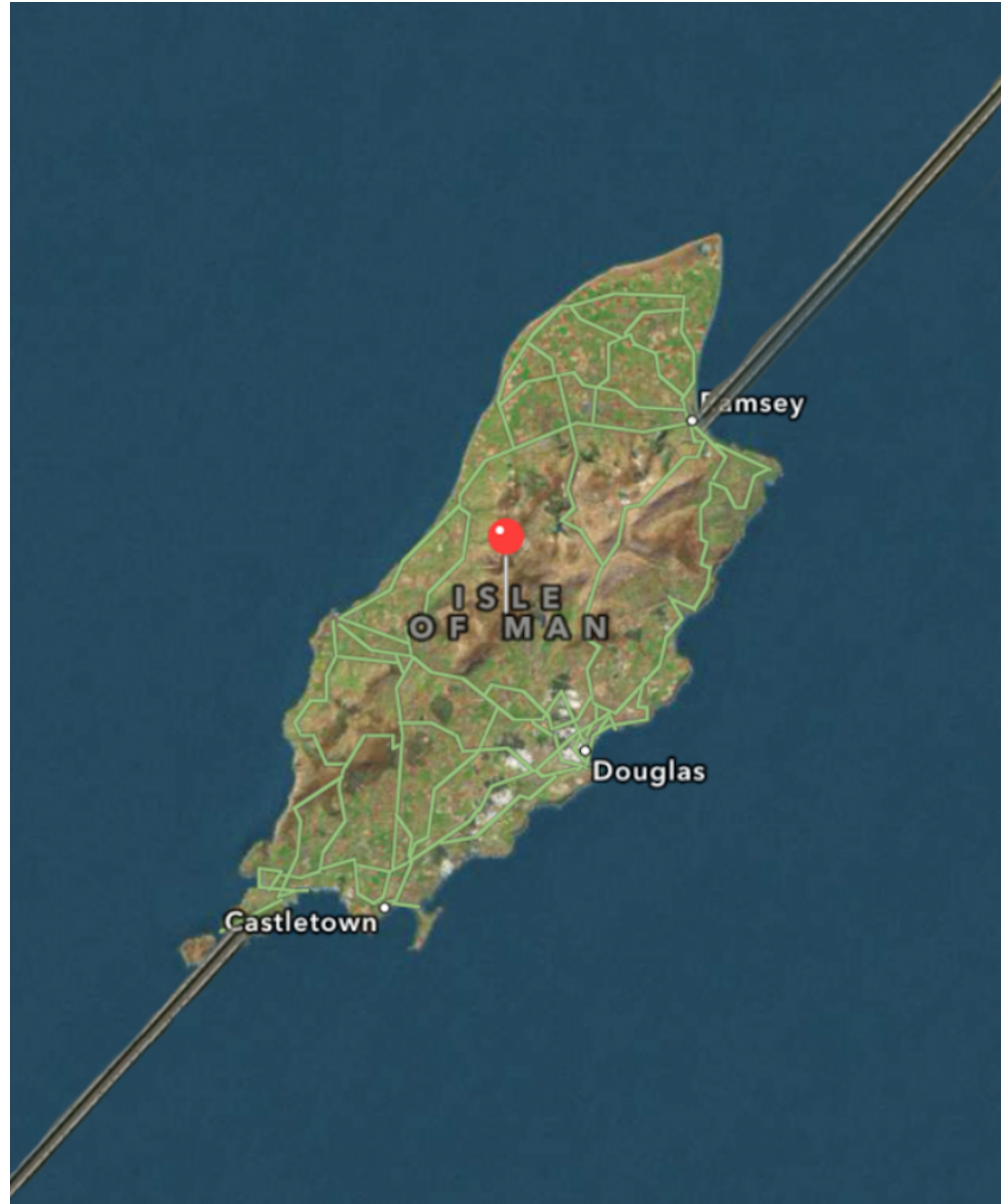


# New Manx language resources for Irish speakers

Kevin Scannell  
Saint Louis University  
10 June 2016

# Intergaelic: Building bridges





# Gaelg/Manx Gaelic

- Part of the Gaelic dialect continuum
- Orthography very different from ga/gd
- Declared “extinct” by UNESCO
- Being revived now, including immersion school
- Probably less than 500 fluent speakers

# Three-Way Comparison

- ga: “Níl leigheas agat air sin, arú,” a dúirt an Cat. “Táimid go léir as ár meabhair anseo.”
- gd: “O, chan eil atharrach ann,” ars an Cat. “Tha sinn uile às ar ciall an seo.”
- gv: “Ogh, cha nel niart ayd er shen,” dooyrt y Kayt. “Ta shin ooilley keoi ayns shoh.”

# Goals

- Resources to help Irish speakers learn Manx
- Manx to Irish dictionary
- Machine translation engine
- Translations of social media posts
- Manx texts annotated with Irish “hints”
- Talked about similar stuff for Gàidhlig last year

# Why Machine Translation?

- One big community from three small ones
- Break down communication barriers
- Open new markets for Manx writers
- Tools for learners, w/o going via English
- “Transfer” Irish language tech to Manx
- Enables comparative linguistic study
- Rehabilitate a much-maligned technology!
- Deliver MT results in creative, tailored ways

# Parallel corpus

- Aligned Irish-Manx Gaelic texts
- A little bit of everything!
- The full Bible, old and new testaments
- Six novels translated from Irish (!)
- ECRML
- Software translations (Minecraft, e.g.)
- 74k segments, just over 1M words on each side
- Unusual in that very little is direct translation



# How to translate?

- Work through the source sentence left-to-right
- Incorporate limited reordering via a “phrase table”
- e.g. “er dty eiyrts” → “sa tóir ort” (or “ar do lorg”)
- Each word/phrase has multiple possible translations
- Hence each sentence has *many* possible translations
- 20 word sentence,  $\sim 2$  translations/word  $\Rightarrow \sim 1000000!$
- Goal is to find the *most probable* translation
- Prune the possibilities via a “Markov model”

# Web Corpora

- No linguistic analysis of source lang needed
- No statistics for source language either!
- Entirely driven by statistics of target (Irish)
- Of which there is A LOT online
- About 150 million words of Irish in total
- Some care needed, but basically more=better

# Disambiguation, I: orthography

- Challenging b/c of reduced spelling
- “Diu y kayt ooilley'n bainney oor”
- Manx “diu” is either Ir. “d'ól (d'ibh)” or “daoibh”
- Manx “oor” is either Ir. “úr” or “uair”
- “D'ól an cat an bainne úr go léir”
- “Beeym ersooyl tree kerroo oor”
- “Beidh mé ar siúl trí cheathrú uair (an chloig)”

# Disambiguation II: grammar

- Irish dependent verbs don't survive in Manx
- Manx “Ass ferg **ren** eh eh”
- Ir. “As fearg a **rinne** sé é”
- “...boayl raad **ren** y Çhiarn conaant rish cloan Israel, tra haink ad magh ass Egypt”
- “...áit a **ndearna** an Tiarna conradh le clann Iosrael, nuair a tháinig siad amach as an Éigipt”

# Disambiguation III: hyperambiguity

- Manx “ny” = Ir. na, ina, ná, nó, níos
- Manx “ta fer **ny** croshey **ny** lhie **ny** ta eh **ny** s’aa na mish; **ny** bee aggle ort roish”
- Ir. “tá fear **na** croise **ina** luí **nó** tá sé **níos** óige ná mise; **ná** bíodh eagla ort roimhe”

# Disambiguation IV: conditional

- Manx “dy” = Ir. dá/go (and more!)
- Preceding context required
- “**Dy** beign my phaitçhey aeg jiu, foddee **dy** beagh”
- “**Dá** mbeinn i mo pháiste óg inniu, b’fhéidir **go** mbeadh”

# Disambiguation V: cur

- Cur, m. – act of putting, sending, sowing, raining, discussing, burying, vomiting, hammering into the ground, throwing through the air, rejecting, shooting, the setting or clamp in a rick of turf, selling, addressing, the crown of cast-iron buttons which have been made bright by contact with cliff-faces, the stench of congealing badger’s suet, the luminance of glue-lice, a noise made in an empty house by an unauthorised person, a heron’s boil, a leprachaun’s denture, a sheep-biscuit, the act of inflating hare’s offal with a bicycle pump, a leak in a spirit level, the whinge of a sewage farm windmill, a corncrake’s clapper, the scum on the eye of a senile ram...
- Worse in Manx: “cur” can mean Ir. “cur” or “tabhairt”
- Mx. “Cur er y choontey eh” = Ir. “Cuir ar an chuntas é”
- Mx. “Cur dou dty laue” = Ir. “Tabhair dom do lámh”

# 20,000 word bilingual dictionary

Իրաճօր  
ՄԱՌԱՌՈՒՐ-ՇԱԵԼՇԵ

ՇԱՌՄԻՆ Օ ՏՇԱՌԱԼԼ

ՇԱԾԱՌ ԱՌԱՌՐ




# intergaelic.com

INTERGAELIC

Gaelg ► Gaeilge ▼

FOCLÓIR **AISTRIÚCHÁN**

Tra va shin ayns thie carrey dooin yn chiaghtyn shoh chaie, daag my ven e kishthey tombaghey. Cha row fys eck derrey va shin drappal magh ass my ghleashtan cheumooie nyn dhie. Dooyrt ee, "Vel monney tombaghey ayd? Ta mee er n'aagail my chishtey ec Knocksharree."

 Aistrigh »

Tra va shin ayns thie carrey dooin yn chiaghtyn shoh chaie, daag my ven e kishthey

Nuair a bhíomar i dteach cara dúinn an tseachtain seo caite, d'fhág mo bhean a bosca

tombaghey. Cha row fys eck derrey va shin drappal magh ass my ghleashtan cheumooie nyn

tobac. Ní raibh a fhios aici go dtí go raibh muid dreapadh amach as mo charr lasmuigh ina

dhie. Dooyrt ee, "Vel monney tombaghey ayd? Ta mee er n'aagail my chishtey ec

dteach. Dúirt sí, "An bhfuil mórán tobac agat? Tá mé tar éis fágáil mo bhosca ag

Knocksharree."

Knocksharree."

Ghow shin nyn greieyn-kiaullee ass y ghleashtan as hug mee yn gleashtan ersooyl harrish y

Ghabhamar a ionstraimí as an charr agus chuir mé an carr imithe thar an

raad. Beign dou cur lhiams shiartanse dy chostraylyn follym as v'ad ayns my ghaa laue

mbóthar. B'éigean dom cur liomsa roinnt bhuidéal folamh agus bhí siad i mo dhá lámh

tra va mee cheet er ash harrish y raad reesht. Tra haink mee shaghey yn barney eddyr

nuair a bhí mé ag teacht ar ais thar an mbóthar arís. Nuair a tháinig mé thar an bearna idir

# Twitter stream



**greinneyder** Greinneyder

She scorcher t'ayn reesht ayns Mannin jiu [#manx](#) [#gaelg](#) [#isleofman](#)

3 lá ó shin Freagra Atweetáil Réiltín



**FeerVaynrey** Chris Williamson

Varkee shin dys y faaie voulleragh ayns Balley Chashtal as honnick shin nyn garrey Stewie as eshyn ve ny speeikear er'n skimmee shen.

[#Gaelg](#)

4 lá ó shin Freagra Atweetáil Réiltín



**greinneyder** Greinneyder

S'mie lhiam ratçhal - I like racing

<https://t.co/L1SzGxrBRN> [#manx](#) [#gaelg](#) [#iomtt2016](#)

5 lá ó shin Freagra Atweetáil Réiltín



**Carsalagh** Robard y Carsalagh

Lesh Radio TT er AM 1368, cha bee Traa dy Liooar ry chlashtyn derrey 13ss Mean Souree. S'aighoil nagh vel peiagh erbee eaishtagh ree. [#gaelg](#)

5 lá ó shin Freagra Atweetáil Réiltín



**Carsalagh** Robard y Carsalagh

[@jamysjeheiney](#) Lesh Radio TT er AM, cha bee Jamys Jeheiney goll er creeley rish kegeesh. Y nah chlaare, bee shen 17/06/16.

5 lá ó shin Freagra Atweetáil Réiltín



**FloydGaelgayr** Floyd gaelgayr

[@FeerVaynrey](#) [@JamysJeheiney](#) Share lhiam Vegemite mish-hene

6 lá ó shin Freagra Atweetáil Réiltín



**FloydGaelgayr** Floyd gaelgayr

S'mie shen! <https://t.co/ynKhozCV3t>

6 lá ó shin Freagra Atweetáil Réiltín

# Ar Ghuailí na bhFathach

- Ciarán Dunbar
- Rob Teare
- Phil Kelly
- Chris Sheard
- Marion Gunn
- Caoimhín Ó Donnaíle
- Michal Boleslav Měchura
- Adrian Cain
- Brian Stowell
- Gearóid Ó Néill