



Irish Language MT in the Public Sector

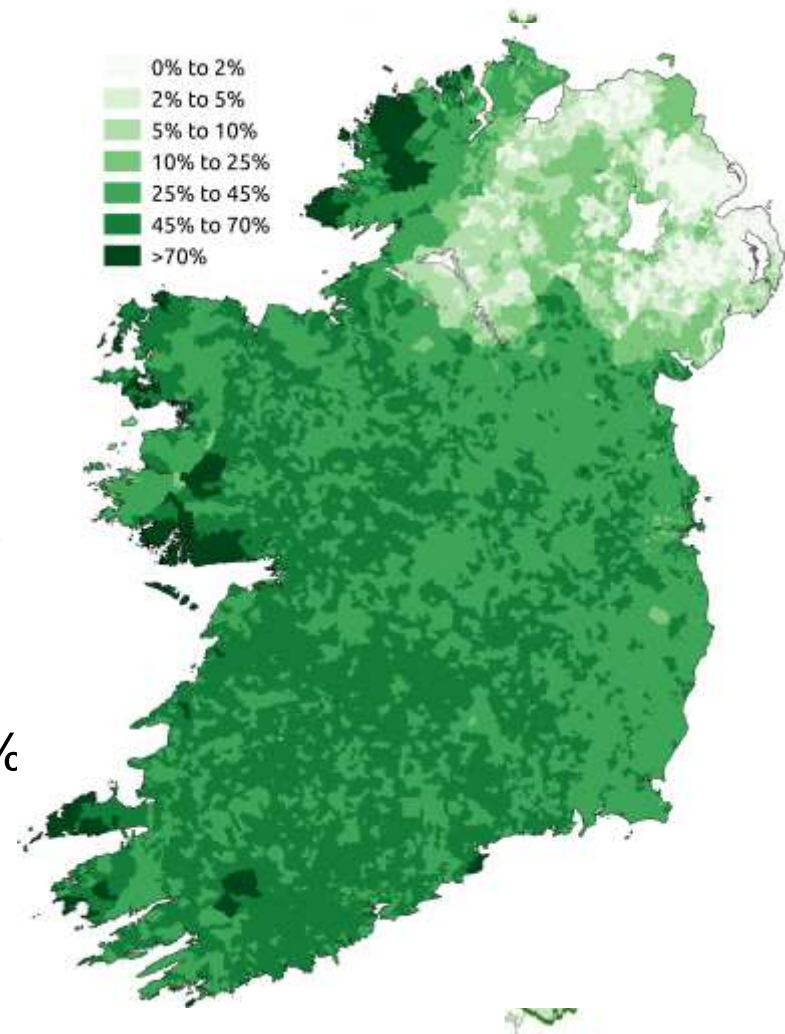
John Judge – ADAPT, DCU, Ireland

Aodhán Mac Cormaic – Dept. Arts Heritage and Gaeltacht,
Government of Ireland

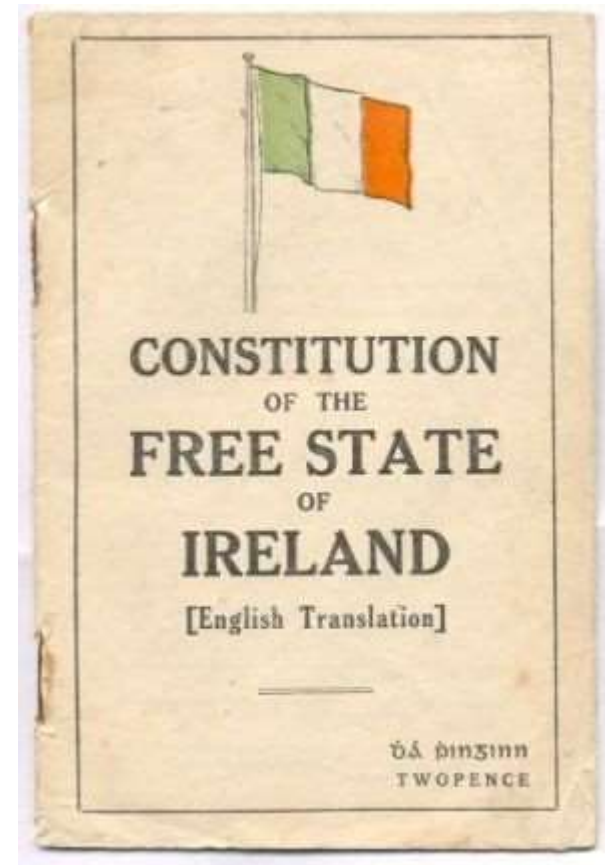
Ireland and the Irish Language



- Irish (Gaeilge) – Indo-European language, Celtic branch native to Ireland
- Ireland's first official language
- Once the most widely spoken language
- Historical factors led to 53% speakers by 1800
- Number of speakers peaked at approx 3.5 million in mid 1800s
- By 1911 just 580k speakers (13.3%)
- 2011 41.4% are speakers – BUT only 1.8% speak Irish daily outside of education
- Still spoken natively – Gaeltacht Regions
- “Urban” revival – Irish medium schools increasingly popular



- 1922 Establishment of the Free State
- Significant resources allocated to preservation and protection of Irish
 - Protected in the constitution (1922/37)
 - Required subject in education
 - Requirement for teaching, civil service, police, practising law...
- 1926 – 19.3% of the population speak Irish



1973 – Ireland joins EEC (EU): **Irish is a Treaty Language**, official procedural language of the European Court of Justice, correspond in written Irish with EU Institutions

1999 – British Irish Agreement Act: Recognised “the importance of respect understanding and tolerance in relation to linguistic diversity, the Irish language... **part of the cultural wealth of the island of Ireland**”

2003 – Official Languages Act: sets out rules regarding use of the Irish language by public bodies, **established the office of An Coimisinéir Teanga to monitor and enforce compliance** by public bodies

2007 – Official working language of the EU: Unanimous decision by EU foreign ministers (2005); **with a partial derogation** in place whereby only key legislation must be translated into Irish, i.e. Irish translations will only be carried out for documents covered by co-decision between the European Parliament and the Council of Ministers

2012 – Gaeltacht Act: Updated language planning and policies, **gives statutory effect to a 20 year strategy for Irish**



The Department of Arts Heritage and Gaeltacht (DAHG)





***An Roinn
Ealaíon, Oidhreachta agus Gaeltachta***

***Department of
Arts, Heritage and the Gaeltacht***

Department of the Government of Ireland

Built and Natural Heritage, Arts, Film, Music, Cultural Institutions as well as Irish Language, Gaeltacht Schemes and Offshore Islands.

Dedicated Minister of State with responsibility for Gaeltacht affairs – Irish language



- Considerations
 - Each department must provide content in Irish
 - Additionally must correspond in Irish
- Meeting the DAHG's need
 - In house translation team supported by fluent staff
 - DAHG is often the first port of call for other departments re Irish language matters
 - Additional, independent, translation team for Oireachteas
 - Outsourced translation work



Example Departmental EN>GA Translation Outsourcing Expenditure

Department	Cost in 2013 €
Communications, Energy and Natural Resources	23,189
Justice and Equality	18,308
Defence	20,682
Agriculture, Food and Marine	6,906
Health	44,687
Transport, Tourism and Sport	5,981
Foreign Affairs and Trade	93,004
Finance	10,202
Education and Skills	107,521
Public Expenditure and Reform (including OPW)	20,175
Jobs, Enterprise and Innovation	10,544
Social Protection	37,080
Arts, Heritage and the Gaeltacht	37,285
Taoiseach	20,888
Total	456,452

Many Depts. additionally employ their own translators

This isn't the whole story

Much government material CANNOT be outsourced

[1] Some programme expenditure not included.

[2] Only outsourcing costs are included in figure provided. DJE also employs a full-time translator to translate Rules of Court on an ongoing basis.

[3] Average annual cost over period 2007 to 2013.

[4] Includes cost of staff translator grade III.



- All content fully human translated
- CAT tools – Trados
- No language tools
- No shared resources
- No terminology management
- Certainly no MT



A Shared Translation Service



- Government Decision (April 2014) – Official Languages (Amendment) Bill
- Authorised initiatives including the establishment of a Shared Translation Service to meet the Irish language requirements of Government Departments
- STS will operate within DAHG's Irish Language Policy division and be available for all government. departments
- Initially creating 4 new jobs
- Scale STS to other languages for other Depts. eg. Health, Justice, Social Protection, Children



- Prioritise the day-to-day translation requirements of Departments
 - Parliamentary Questions
 - Ministerial speeches
 - Press releases
 - General correspondence
- Assist with the translation of other material,
 - Website material,
 - Annual reports,
 - Audited accounts
 - Strategy statements
- Provide an advisory service to avoid duplication and assist in managing any existing translation contracts



“... Central to the successful implantation of this plan and to the achievement of value for money and economies of scale is the development of an effective machine translation system. ...”

- DAHG report to Government on the STS



Developing MT for the Shared Translation Service



- No in-house knowledge or experience of MT in practice
 - save free online services
- Significant negative perception
- However, strong will to modernise and establish best practice in translation → result of META-NET Whitepapers, SRA, outreach campaigns
- Scale STS to other languages for other Depts. eg. Health, Justice, Social Protection, Children...



- Met with translators
- Discussed: workloads, problematic cases, repetitive tasks, consistency issues
- Examined workflows using only TM (or nothing in some cases)
- Established a use case for initial MT trials
- Followed up with presentation of mocked up TM/MT integration → Translators convinced



- Severe shortage of quality data for Irish
 - Legislative texts available – very long winded, constrained vocabulary, specific style → Not useful for the use case for the STS
 - DAHG TMs available, but
 - Disparate locations
 - No management or reuse
 - Often “reinventing the wheel”
- Put the case for sharing public data (eg. TMs)
 - Improve the development of LT tools and resources for Irish
 - Establish a community of interest (public/industry/R&D)
 - Consistency and control
 - Stop “reinventing the wheel”
- Principle of data sharing wherever possible is at the core of the STS and associated MT development
 - Initial investigations uncovered ca 30k parallel segments which can be shared
 - New workflow will generate MUCH more





- Short term pilot project: Tapadóir
 - Establish the viability of developing an MT engine for DAHG/STS use case
 - Test integration into existing translation workflows to establish usability for STS
 - Measure impact

- Established use case and accompanying test set documents with DAHG translators
- Combined existing resources into baseline system → Poor
- Crawled public information sources online
- Harvested and cleaned TM data from DAHG
- Integrated new data
- Very significant improvement in automated testing
- Outperforms free online services for the DAHG/STS use case



- Small scale integration test with DAHG translator
- Test feasibility/usability of MT-based solution in STS
- Measure impact on productivity

- Key results
- User impressions overwhelmingly positive
- No negative impact on productivity
- Indicate >50% productivity increase is possible on certain texts



- DAHG updating translation tooling, policies and infrastructure
- Data sparsity – DAHG data collection, clean up and curation
- In house deployment of MT engine from pilot phase
- Partnering with ADAPT to do further R&D → Improved engine
- ADAPT identifying STS resources which could/should be shared and reused



Thanks!

www.adaptcentre.ie

Tapadóir Team:

Dr. John Judge

Eimear Maguire

Teresa Lynn

Supported by:

Dr. Jennifer Foster, Prof. Andy Way, Prof. Qun Liu and Team

Contact: jjudge@computing.dcu.ie

Thank You

