



An Roinn
Ealaíon, Oidhreachta agus Gaeltachta
Department of
Arts, Heritage and the Gaeltacht

Language Technologies for The Irish Language (Gaeilge)

Dr Aodhán Mac Cormaic

Assistant Principal

Department of Arts, Heritage and the Gaeltacht

Ireland

Language Technologies for Irish

Achoimre / Summary

- Current status of LT for Irish.
- Sector driven by people with a passion for the Irish language.
- Details of work already underway.
- Research demonstrates that Irish, like some major world languages, is falling behind English in the Digital Age.
- Government efforts to tackle this problem:
 - Plean Digiteach don Ghaeilge / Digital Plan for the Irish Language

Investment by Department of Arts, Heritage and the Gaeltacht

€8.4m invested in Irish language digital and technology sector since 2006. Over €1m p.a. over last 3 years.

Projects Funded:

- Abair.ie – voice synthesis
- Tapadóir – Machine Translation project
- TechSpace as Gaeilge



*An Roinn
Ealaíon, Oidhreachta agus Gaeltachta*
*Department of
Arts, Heritage and the Gaeltacht*

Investment in Irish Language Corpora

Irish Language Terminology for the EU Terminology Database (IATE)

- Annual grant of €231,000 to Dublin City University.
- Irish is now the 13th largest of the languages in the database and the largest of the new languages!
- 72,000 terms translated into Irish do date.
- Important work due to strategy to end derogation by 2022.

Investment in Irish Language Corpora

www.gaois.ie

- Search engine on www.gaois.ie site allowing searches for legal texts.
- 9m words in this corpus, half in Irish and half in English.
- Developed by Dublin City University

Investment in Online Dictionaries

- Irish version of *Foclóir Béarla-Gaeilge*, 80% of which is complete, available on www.focloir.ie. For completion mid-2016.
- Three other dictionaries – Béarla/Gaeilge (1959), *An Foclóir Beag* (aon teangach) agus *Foclóir Gaeilge-Béarla* (1978) – all available on www.teanglann.ie.
- Royal Irish Academy – historical Irish language dictionary



Investment in Machine Translation

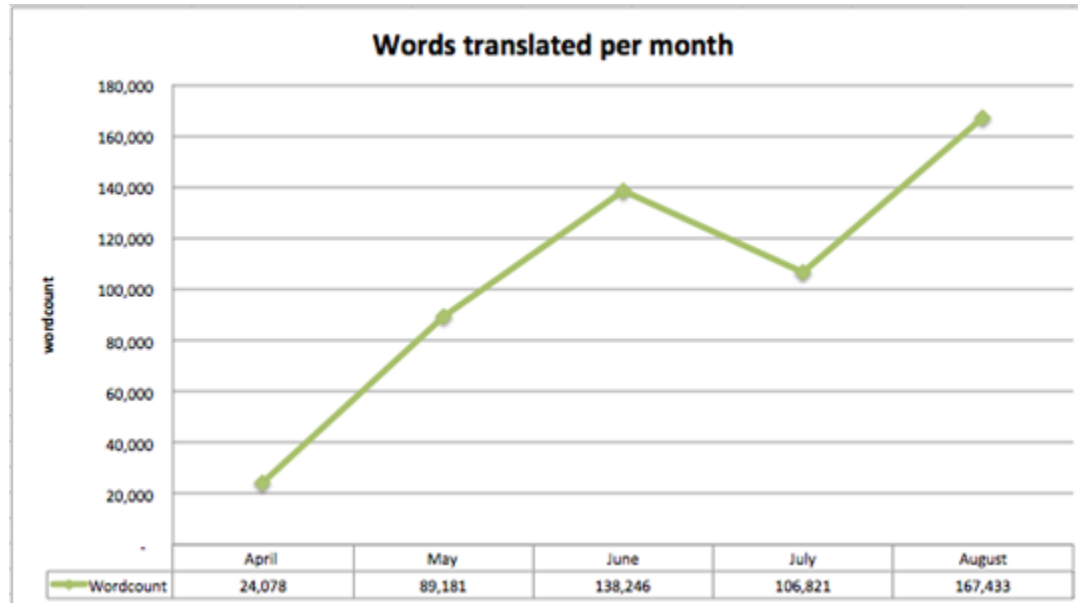
Tapadóir: Machine Translation System

- DCU research – statistic-based.
- Trinity College Dublin – rule-based.
- 2016: Hybrid system combining both.





Tapadóir



2012 META NET Report:

An Ghaeilge sa Ré Dhigiteach (The Irish Language in the Digital Age)

Language Processing: level of support for language technology for 30 European languages				
Excellent Support	Good Support	Reasonable Support	Intermittent Support	Poor or No Support
	Béarla (English)	Gearmáinis Iodáilis Fionlainnis Fraincis Ollainnis Portaingéilis Spáinnis Seicis	Bascais Bulgáiris Danmhairgis Eastóinis Gailísis Gréigis Gaeilge Catalóinis Ioruais Polainnis Sualainnis Seirbis Slóvaicis Slóivéinis Ungáiris	Íoslainnis Cróitis Laitvis Liotuáinis Máltais Rómáinis



Tuarascáil META NET foilsithe i 2012: An Ghaeilge sa Ré Dhigiteach

Machine Translation: level of support for language technology for 30 European languages

Tacaíocht den scoth	Tacaíocht mhaith	Tacaíocht réasúnta	Tacaíocht bhriste	Tacaíocht lag nó gan tacaíocht
	Béarla (English)	Fraincis Spáinnis	Gearmáinis Iodáilis Catalóinis Ollainnis Polainnis Rómáinis Ungáiris	Bascais Bulgáiris Danmhairgis Eastóinis Fionlainnis Gailísis Gréigis Gaeilge Íoslainnis Cróitis Laitvis Liotuáinis Máltais Ioruais Portaingéilis Sualainnis Seirbis Slóvaicis Slóivéinis Seicis



An Roinn
Ealaíon, Oidhreachta agus Gaeltachta
Department of
Arts, Heritage and the Gaeltacht

Tuarascáil META NET foilsithe i 2012: An Ghaeilge sa Ré Dhigiteach

Text Analysis: level of support for language technology for 30 European languages				
Tacaíocht den scoth	Tacaíocht mhaith	Tacaíocht réasúnta	Tacaíocht bhriste	Tacaíocht lag nó gan tacaíocht
	Béarla (English)	Gearmáinis Fraincis Iodáilis Ollainnis Spáinnis	Bascais Bulgáiris Danmhairgis Fionnlainnis Gailísis Gréigis Catalóinis Ioruais Polainnis Portaingéilis Rómáinis Sualainnis Slóvaicis Slóivéinis Seicis Ungáiris	Eastóinis Gaeilge Íoslainnis Cróitis Laitvis Liotuáinis Máltais Seirbis



Tuarascáil META NET foilsithe i 2012: An Ghaeilge sa Ré Dhigiteach

Speech and Text Resources: level of support for language technology for 30 European languages				
Tacaíocht den scoth	Tacaíocht mhaith	Tacaíocht réasúnta	Tacaíocht bhriste	Tacaíocht lag nó gan tacaíocht
	Béarla (English)	Gearmáinis Fraincis Iodáilis Ollainnis Polainnis Sualainnis Spáinnis Seicis Ungáiris	Bascais Bulgáiris Danmhairgis Eastóinis Fionlainnis Gailísis Gréigis Catalóinis Cróitis Ioruais Portaingéilis Rómáinis Seirbis Slóvaicis Slóivéinis	Gaeilge Íoslainnis Cróitis Laitvis Liotuáinis Máltais



Our New Approach: A Digital Plan for the Irish Language

- Long-term plan required in order to improve technologies in various sectors.
- Expert team from DCU and Trinity College.
- Drew on Welsh language *Digital Survival Kit*.
- Research commenced in September 2015 and to be published in summer 2016.



An Roinn
Ealaíon, Oidhreacht agus Gaeltachta
Department of
Arts, Heritage and the Gaeltacht

Aims of the Plan

- To set up a long-term research and development infrastructure that will into the future deliver those state-of-the-art technologies that are increasingly vital for language maintenance.
- In the plan, the basic linguistic and phonetic research is seen as providing the essential resources for the technology development. These technologies include machine translation, text-to-speech synthesis, speech recognition, and dialogue systems that enable speech-based human computer interaction.
- These core technologies will enable the development of the growing number of applications that will serve the Irish speaking public.
- These technologies are particularly vital for the teaching/learning of Irish, as well as for those with disabilities.



Contents of the Digital Plan?

- **Digital documentation and linguistic analysis of the written and spoken dialects**
- **Language Resources: Resources, Data and Knowledge Bases**
- **Natural Language Processing (NLP)**
- **Natural Language Understanding (NLU)**
- **Speech Synthesis**
- **Speech Recognition: Conversion of spoken word to text**
- **Machine Translation Systems**
- **Dialogue Systems**
- **Information Retrieval Systems**
- **Educational Applications**
- **Access for people with disabilities**
- **Role of national and multi-national companies and of Government and the public**



*An Roinn
Ealaíon, Oidhreachta agus Gaeltachta
Department of
Arts, Heritage and the Gaeltacht*

Next Steps

- Publication of Plan in summer 2016
- Ministerial support essential
- Funding plan
- Review and update every 5 years

Críoch / End