

LANGUAGE TECHNOLOGIES FOR IRISH

DR. TERESA LYNN, Dublin City University

24TH JUNE 2021



TOPICS OF THIS SESSION

- Is our language technologically ready for the AI era?
- Is there a language technology industry and research in Ireland or are we dependent on the big international players?
- What are the mature services and tools for Irish that can enhance digital services?

TOPICS OF THIS SESSION

NO

- Is our language technologically ready for the AI era?

Minimal

- Is there a language technology industry and research in Ireland
- What are the ~~mature~~ services and tools for Irish that can enhance digital services?

LET'S SEE...

THE IRISH AN
LANGUAGE IN GHAEILGE
THE DIGITAL SA RÉ
AGE DHIGITEACH

John Judge
Ailbhe Ní Chasaide
Rose Ní Dhubhda
Kevin P. Scannell
Elaine Uí Dhonnchadha

IRISH LANGUAGE TECHNOLOGY

- META-NET white paper series (Judge et al. 2012)
- EU-led survey
- Language resources and technologies
- 31 EU languages

	Excellent	Good	Moderate	Fragmentary	Weak/No Support
Machine Translation		English	French, Spanish	Catalan, Dutch, German, Hungarian, Italian, Polish, Romanian	Basque, Bulgarian, Croatian, Czech, Danish, Estonian, Finnish, Galician, Greek, Icelandic, Irish , Latvian, Lithuanian, Maltese, Norwegian, Portuguese, Serbian, Slovak, Slovene, Swedish, Welsh
Text Analysis		English	Dutch, French, German, Italian, Spanish	Basque, Bulgarian, Catalan, Czech, Danish, Finnish, Galician, Greek, Hungarian, Norwegian, Polish, Portuguese, Romanian, Slovak, Slovene, Swedish	Croatian, Estonian, Icelandic, Irish , Latvian, Lithuanian, Maltese, Serbian, Welsh
Speech Processing		English	Czech, Dutch, Finnish, French, German, Italian, Portuguese, Spanish	Basque, Bulgarian, Catalan, Danish, Estonian, Galician, Greek, Hungarian, Irish , Norwegian, Polish, Serbian, Slovak, Slovene, Swedish	Croatian, Icelandic, Latvian, Lithuanian, Maltese, Romanian, Welsh
Resources		English	Czech, Dutch, French, German, Hungarian, Italian, Polish, Spanish, Swedish	Basque, Bulgarian, Catalan, Croatian, Danish, Estonian, Finnish, Galician, Greek, Norwegian, Portuguese, Romanian, Serbian, Slovak, Slovene	Icelandic, Irish , Latvian, Lithuanian, Maltese, Welsh

RISK OF DIGITAL EXTINCTION

“The Printing Press resulted in
the extinction of many
Regional and Minority
Languages”



Will technology have the same impact on
Irish?

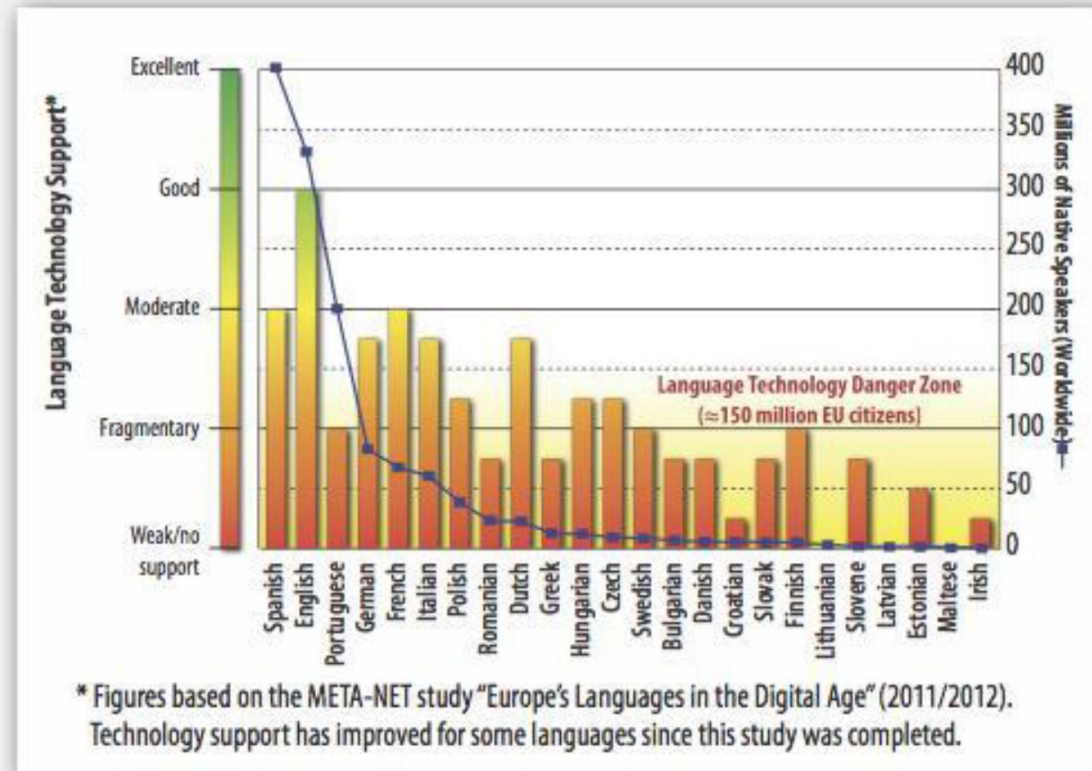
The native languages of approximately **140 million EU citizens** are in the **Language Technology Danger Zone**, where language technology is inadequate to support the DSM.



Online Automatic Translation Quality

Current online automatic translation provided by US tech giants does not solve the “language problem”: **less than 30% of automatically translated content is truly useful** for online commerce.

Only three European languages (Spanish, English, and French) meet at least the “moderate” level of language technology support.



We need to ensure continued language usage through technology

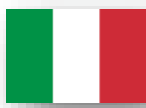
- Machine translation
- Social Media Use
- Multi-platform word processing tools
- Edutainment packages/ Computer assisted learning
- Search engine
- Text Generation (e.g. text summarisation)
- Automatic subtitling
- Chatbots



MACHINE LEARNING

**“DATA-DRIVEN”:
LEARNING PATTERNS
IN LANGUAGE**

LATEST AI TECHNIQUES @ ADAPT



Irish Language Technology

- Dictionaries, Terminologies, Corpora
- Thesaurus/ Wordnet
- Spell-checker/ Grammar checker
- Part-of-speech tagger, lemmatiser, tokenizer, morphological analyser
- Speech Synthesiser (www.abair.ie)
- IRIS machine translation



Fiontar & Scoil na Gaeilge



Irish Language Technology @ DCU

- Irish Syntactic Parsing (helping computers understand Irish syntax/grammar)
- Automatic processing of Irish Multiword Expressions
- Automatic processing of Irish Social Media Content
- Irish word-embeddings (language modelling)
- Machine Translation & Data collection (ELRC)



LANGUAGE TECHNOLOGY IN THE IRISH PUBLIC SECTOR

The screenshot displays a translation application window titled "Gnáthaistrúcháin an Aonaid, Iconic Plugin - Translation Results". The interface is split into several panes:

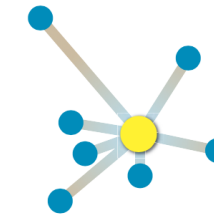
- Top Left:** Project settings and a toolbar with icons for project management.
- Top Center:** The source text: "Speaking at the [announcement](#) of the [Peace Proms](#), Minister [Heather Humphreys](#) said:"
- Top Right:** A "Term Recognition" sidebar listing terms like "speaking", "announcement", "peace", "prom", "PROM", and "Minister Humphreys".
- Main Area:** A table showing the translation process for two segments:

1	Speaking at the launeh announcement of the Peace book Proms, Minister Heather Humphreys said:	71%	Agus í ag labhairt ag seoladh an leabhair inniu, dúirt an tAire Mhich Unfraidh:
2	Speaking at the announcement of the Peace Proms, Minister Heather Humphreys said:	AT	Ag labhairt di nuair a fógraíodh an peace Proms, dúirt an tAire Heather Mhich Unfraidh:
- Bottom:** A split view showing the source text on the left and the target text in Irish on the right. The source text includes: "the young performers, who range in age from 10 to 19, have been drawn from across the island of Ireland, and are all exceptional and award-winning artists and performers." The Irish translation is: "tagair na taibheoirí óga, atá tuair 10 agus 19 mbliana d'aois, as gach cearn d'oileán na hÉireann, agus is ealaíontóirí agus taibheoirí den scoth, a bhfuil gradaim buaite acu, gach duine díobh."
- Bottom Right:** A "Translation Details" tooltip showing: Status: Draft, Origin: Interactive, Score: 0%.

LANGUAGE TECHNOLOGY IN THE IRISH PUBLIC SECTOR



shutterstock.com • 103492922



**European Language
Resource Coordination**
Connecting Europe Facility

ELRC WHITE PAPER

**Sustainable Language Data Sharing
to Support Language Equality
in Multilingual Europe**

WHY LANGUAGE DATA MATTERS

LANGUAGE TECHNOLOGY IN THE IRISH PUBLIC DOMAIN?



Parliamentary questions

Naturalisation Application Forms

[Home](#) / [Citizenship](#) / [Naturalisation Application Forms](#)

Automatic subtitling – reduce time/costs

Subtitle translation – reduce time/costs

Improve translation turnaround times

LANGUAGE TECHNOLOGY IN THE IRISH PUBLIC SECTOR?

ABAIR Web Reader for Irish

The ABAIR *Web Reader* is an online service that provides Irish text-to-speech synthesis for websites.

Why use text-to-speech synthesis?

If your website is connected to the ABAIR *Web Reader*, users will have the option of having the content read out to them.

Why this is important

By providing text-to-speech synthesis, you make the content of your website accessible to visually impaired users. Public bodies are even required by law to do so (cf. *Disability Act 2005*, §26 (1)(a)).

Even users that are not visually impaired like to have content read out to them sometimes.

Easy to integrate

The ABAIR *Web Reader* is easy to integrate. One line of HTML code is enough:

```
<script type="text/javascript" src="http://abair.tcd.ie/webreader/scripts/abair_tts_webui.js"></script>
```

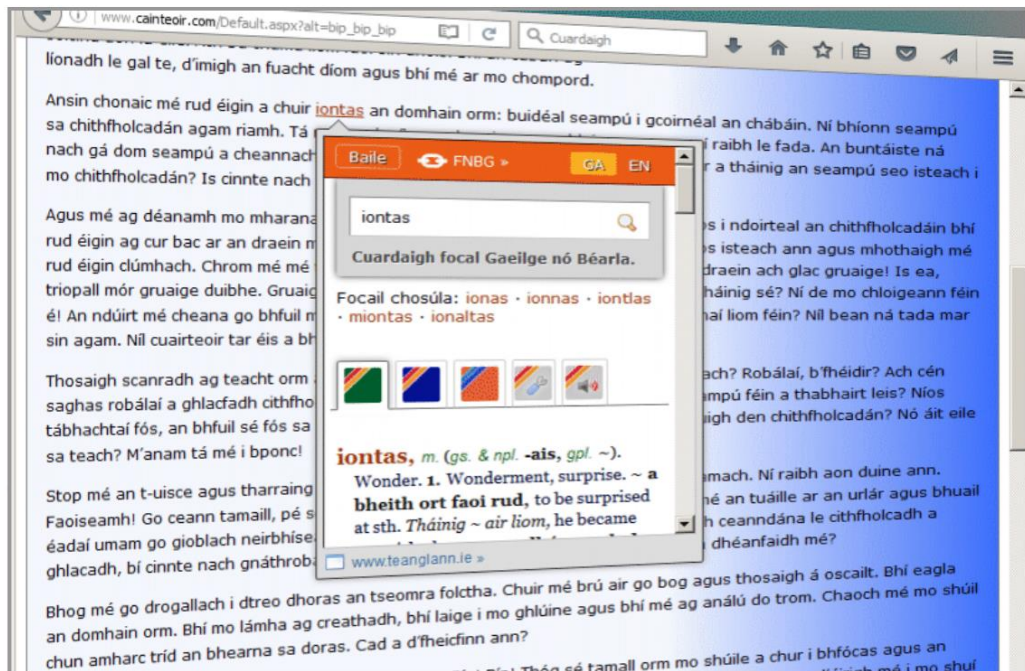
The code above links a website to the ABAIR *Web Reader*. See also the [user guide](#).



LANGUAGE TECHNOLOGY IN THE IRISH PUBLIC SECTOR?

An Foclóirí

If you have a website, you are welcome to include in it a plugin we call *An Foclóirí*. What *An Foclóirí* does is turn every word on your webpage into a clickable link which leads to a search for that word on teanglann.ie. The links are opened in a small window beside the text so your users do not have to interrupt their reading.



Widgets and Plugins
underused in public sector

CONNECTING EUROPE FACILITY FUNDED PROJECTS @ ADAPT



ELRC-SHARE Repository

52 language resources at your disposal (and counting)

Search

EUROPEAN LANGUAGE RESOURCE
COORDINATION (ELRC)

- Sharing & Contributing Language Resources
- Promotion of **eTranslation**
- Cooperation with public bodies

Sharing Data at National Level

Gaeilge English

Home Usage Help Register Login

ELRI European Language Resource Infrastructure

ELRI National Relay Station Ireland

Welcome to ELRI in Ireland!

This site is the official National Relay Station in Ireland, where language resources can be collected, prepared and shared amongst public institutions and translation centres.

National Relay Stations are developed within the [ELRI](#) European project, supported by the [Connecting Europe Facility](#) programme.

Help promote language equality in Europe by joining the ELRI network!

Join ELRI

Members of public institutions and translation centres can join user groups to download shared resources, contribute their own data and benefit from translation memories or other language resources automatically prepared by the ELRI engines.

[Register](#) to become a member of ELRI.

Already a member? [Sign in](#).

Search and share

[Browse](#) available language resources.

[Contribute](#) your own data and retrieve prepared language resources.

ELRI National Relay Station Ireland
Managed by ADAPT and hosted at DCU
powered by [META-SHARE](#) and [ELRC-SHARE](#)



CC-BY-NC-SA 4.0
[Privacy Policy](#)

[Cookie Policy](#)



PRINCIPLE



An Roinn Dlí agus Cirt
agus Comhionannais
Department of Justice
and Equality



Foras na Gaeilge



NUI Galway
OÉ Gaillimh



THE FUTURE OF IRISH LANGUAGE TECHNOLOGY - WHAT LIES AHEAD?



EUROPEAN LANGUAGE EQUALITY

EP Resolution “Language equality in the digital age”
P8_TA(2018)0332 – partially based on the STOA study

Voting (11 Sept. 2018): **592 yes** – 45 no

Selected Recommendations addressed by ELE:

25. Establish a large-scale, long-term coordinated funding programme for research, development and innovation in the field of language technologies, at European, national and regional levels, tailored specifically to Europe’s needs and demands
29. Create a European LT platform for sharing of services
27. Europe has to secure its leadership in language-centric AI

European Parliament
2014-2019



TEXTS ADOPTED
Provisional edition

P8_TA-PROV(2018)0332

Language equality in the digital age

European Parliament resolution of 11 September 2018 on language equality in the digital age (2018/2028(INI))

The European Parliament,

- having regard to Articles 2 and 3(3) of the Treaty on the Functioning of the European Union (TFEU),
- having regard to Articles 21(1) and 22 of the Charter of Fundamental Rights of the European Union,
- having regard to the 2003 UNESCO Convention for the Safeguarding of the Intangible Cultural Heritage,
- having regard to Directive 2003/98/EC of the European Parliament and of the Council of 17 November 2003 on the re-use of public sector information¹,
- having regard to Directive 2013/37/EU of the European Parliament and of the Council of 26 June 2013 amending Directive 2003/98/EC on the re-use of public sector information²,
- having regard to Decision (EU) 2015/2240 of the European Parliament and of the Council of 25 November 2015 establishing a programme on interoperability solutions and common frameworks for European public administrations, businesses and citizens (ISA2 programme) as a means for modernising the public sector³,
- having regard to the Council resolution of 21 November 2008 on a European strategy for multilingualism (2008/C 320/01)⁴,
- having regard to the Council decision of 3 December 2013 establishing the specific programme implementing Horizon 2020 – the Framework Programme for Research and

¹ OJ L 345, 31.12.2003, p. 90.

² OJ L 175, 27.6.2013, p. 1.

³ OJ L 318, 4.12.2015, p. 1.

⁴ OJ C 320, 16.12.2008, p. 1.



EUROPEAN LANGUAGE EQUALITY

Consortium: 52 partners from all over Europe

Coordinator: ADAPT Centre (Dublin City University)

Objective: *development of a strategic research, innovation and deployment agenda to achieve digital language equality in Europe by 2030*

Runtime: 18 months

Started on 1 January 2021



Language technologies

LT services, tools, components, downloadable or deployed directly through the grid

[Browse Technologies](#)



Language data and resources

The collection of data sets, corpora, language models and other language resources

[View Resources](#)



Community

Organizations, projects, events etc in European Language technology field

[Explore Community](#)

Digital plan for the Irish Language 2021

Linguistic Resources	Corpora	Knowledge Bases	NLP Tools	NLG Tools
Speech Models	Speech Synthesis	Speech Recognition	Spoken Dialogue Systems	Machine Translation
Information Retrieval	State and Public Use	CALL	Disability and Access	Synergies (Industry and Public)

QUESTIONS/ CEISTEANNA?

