



Digital Language (in)Equality in Europe

Teresa Lynn (DCU, ADAPT Centre)
 teresa.lynn@adaptcentre.ie
<http://european-language-equality.eu>

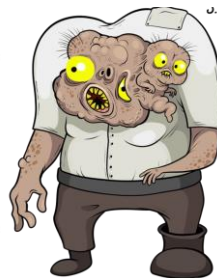
Factors influencing technological support



FUNDING



SKILL
SHORTAGE

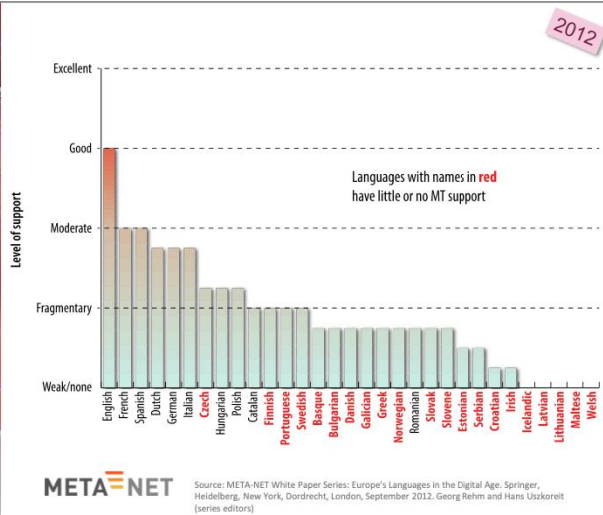


MORPHOLOGY



NUMBER OF
SPEAKERS

2012 META-NET White Paper Series



META-NET

Source: META-NET White Paper Series: Europe's Languages in the Digital Age. Springer, Heidelberg, New York, Dordrecht, London, September 2012. Georg Rehm and Hans Uszkoreit (series editors)

European Language Equality Project



“Language Equality” EP Resolution (2018)

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
Plenar session

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 (2018/2026(INI))

The European Parliament,

- having regard to Articles 2 and 3(1) 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 2013/378(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 329/01)⁴,
- having regard to the Council decision of 3 December 2011 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 Project



Consortium: 52 partners from all over Europe

Coordinator: Dublin City University

Co-coordinator: DFKI

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

Runtime: 18 months (January 2021– June 2022)



ELE



META-FORUM 2022 – June 8/9, Brussels, Belgium

ELE

STOA Workshop 2022

6

European Language Equality Project



- Main result: **Strategic agenda and roadmap**
- Research partners produce **updates** of the **31 META-NET White Papers**
- **Networks** and **initiatives** produce one report on: needs, wishes, demands, visions
- SME partners produce four technical **Deep Dives** for the main technology areas
- **EU citizens' survey** – 20,000+ responses
- **87 languages** taken into account

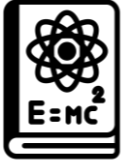
<https://european-language-equality.eu/deliverables/>

Digital Language Equality: Working Definition



Digital Language Equality (DLE) is the state of affairs in which all languages have the technological support and situational context necessary for them to continue to exist and to prosper as living languages in the digital age.

Digital Language Equality: Metric



- a measure that reflects the **digital readiness of a language** and its contribution to the state of **technology-enabled multilingualism**, tracking its progress towards the goal of DLE
- **dynamic** nature, score can be (regularly) updated
- computed for each language **on the basis of various factors**, grouped into

- **technological** factors and $TechDLE_{LangX} = \sum_{i=1}^N Score_{LRTi}$
- **contextual** factors

Digital Language Equality: Metric


Technological Factors

- Tools and Services
- Corpora
- Language Descriptions and Models
- Lexical and Conceptual Resources
- Projects
- Organisations

Contextual Factors

- Industry
- Online
- R&D&I
- Economy
- Education
- Funding
- Law
- Media
- Policy
- Public administration
- Society

Grounded in evidence



EUROPEAN LANGUAGE GRID


7213 Corpora 3669 Tools & Services 2713 Conceptual Resources 482 Models & Grammars 1760 Organizations 37 Projects

Browse ELG to find the resource you are looking for

The European Language Grid provides access to Language Technology resources from all over Europe. ELG contains tools and services, language resources and information on European LT companies and research organisations as well as their projects. You can try out the services or use our APIs. Make use of an LT service, download corpora or find your future business partner - all within the European Language Grid.


Do Europe's languages receive the same level of technology support? Check out our dashboard on Digital Language Equality to find out.

[Visit the dashboard](#)




TOOLS & SERVICES

ELG provides access to LT services that can be tested directly through the user interface or called through the REST API.



LANGUAGE RESOURCES


ELG provides access to datasets, corpora, models and other resources from all over Europe for all European languages.



ORGANISATIONS

ELG provides access to information about all European LT organisations including companies, universities and research.


<https://live.european-language-grid.eu/>
STOA Workshop 2022



Cross-language comparison

- Comparison across languages as per
 - Tools/services broadly categorised into a number of core LT application areas
 - Resources that can be used as training or evaluation data
- Four bands:
 - Good support (green)
 - Moderate (light green)
 - Fragmentary (yellow)
 - Weak or no support (light yellow)

	Tools and Services							Language Resources				Overall	
	Text Processing	Speech Processing	Image/Video Processing	Information Extraction and IR	Human-Computer Interaction	Translation Technologies	Natural Language Generation	Text Corpora	Multimodal Corpora	Parallel Corpora	Models		Lexical Resources
EU official languages													
Bulgarian													
Croatian													
Czech													
Danish													
Dutch													
English													
Estonian													
Finnish													
French													
German													
Greek													
Hungarian													
Irish													
Italian													
Latvian													
Lithuanian													
Maltese													
Polish													
Portuguese													
Romanian													
Slovak													
Slovenian													
Spanish													
Swedish													
(Co-official languages)													
National level													
Albanian													
Bosnian													
Icelandic													
Luxembourgish													
Macedonian													
Norwegian													
Serbian													
Regional level													
Basque													
Catalan													
Faroese													
Frisian (Western)													
Galician													
Jerriais													
Low German													
Manx													
Mirandese													
Ocitan													
Sorbian (Upper)													
Welsh													
All other languages													



Cross-language comparison

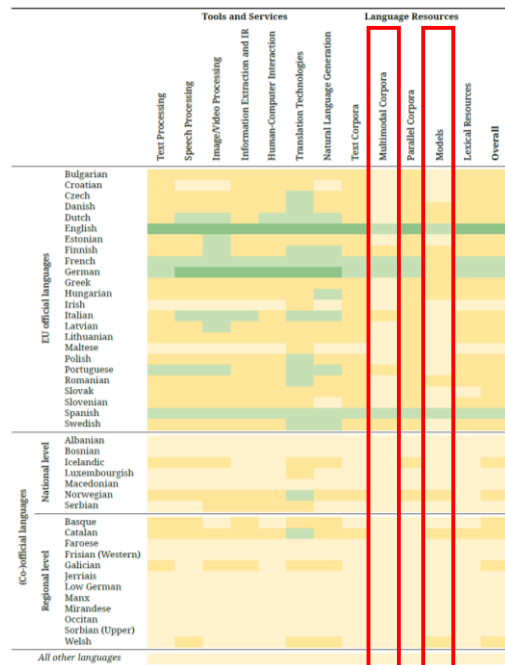
- **Translation technologies:** many languages are at least moderately supported



1
3

Cross-language comparison

- **Translation technologies:** many languages are at least moderately supported
- **Multimodal corpora and Models:** many languages are weakly supported



1
4

Cross-language comparison

- **EU official languages** better supported than other European languages



ELE

1
5

Cross-language comparison

- **EU official languages** better supported than other European languages
- **Best supported:** English, followed by Spanish, German and French



ELE

1
6

Cross-language comparison

- **EU official languages** better supported than other European languages
- **Best supported:** English, followed by Spanish, German and French
- **Least supported** among the EU official: Irish and Maltese



1
7

Cross-language comparison

- **EU official languages** better supported than other European languages
- **Best supported:** English, followed by Spanish, German and French
- **Least supported** among the EU official: Irish and Maltese
- Best supported among languages with official status at the national or regional level (but not EU): Norwegian, Icelandic, Serbian, Basque, Catalan, Galician, Welsh

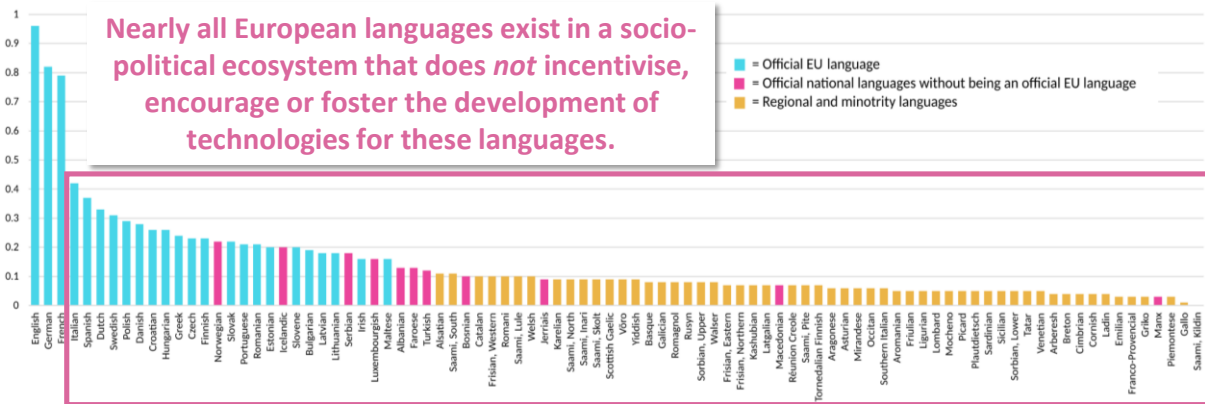


1
8

DLE metric – technological scores



DLE metric – contextual scores



ELE dashboard

<https://live.european-language-grid.eu/catalogue/dashboard>

EUROPEAN LANGUAGE GRID

Catalogue Documentation & Media About

Go to catalogue

About DLE Scores Cross-language comparison Within-language comparison

EUROPEAN LANGUAGE EQUALITY

About European Language Equality / European Language Grid Dashboard

This dashboard facilitates interactive explorations and visualisations of the Digital Language Equality metric and its components.

Digital Language Equality (DLE) is the state of affairs in which all languages have the technological support and situational context necessary for them to continue to exist and to prosper as living languages in the digital age.

The **DLE Metric** is a measure that reflects the digital readiness of a language and its contribution to the state of technology-enabled multilingualism, tracking its progress towards the goal of DLE. The DLE Metric is computed for each language on the basis of various factors, grouped into **technological factors** and **contextual factors**. (See SLE Deliverable D1.1 and D1.2.)

The dashboard has been developed as part of the European Language Grid Catalogue in the framework of the European Language Equality project.

Contact us

GitHub LinkedIn Twitter Youtube

The European Language Grid has received funding from the European Union's Horizon 2020 research and innovation programme under grant

ELE

Risk of Digital Extinction

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



Will technology have the same impact?

Impact on Societies



- The lack of digital support for a language can lead to:
 - **language shift** and eventual language decline, particularly amongst younger generations.
 - A **divide arises in levels of information accessibility** and economic progression across language communities.
- Online data mining is often used by governments and media to gather information on events, political issues and public sentiment. The **voices of many will be left unheard**, unrepresented and unaccounted for.
- **Divide in the levels of education** on offer across language communities - contributing further to **societal inequalities**.

Impact on Societies



- Court and criminal justice systems use multimodal content retrieval to find evidence in audio and video content. **Disadvantage for certain language communities.**
- Societal divide in level of **access to healthcare**
- **Marginalised** groups in speech processing (dialects, elderly, disabilities)
- **Selective** customer **support**
- **Fragmented EU market** (businesses at disadvantage)

Next Steps and Recommendations

- Around least 75 (out of 87) European languages in danger of **digital extinction**
- Europe wants AI to be “based on and respecting European values”. Right now, **European Language Technology development and support is not equal, i.e., not based on European values.**
- We advocate for a **shared, long-term European funding programme** – between the EC and member states. No more sporadic, partial support.
- Prerequisites: **powerful data and compute infrastructure and vast amounts of FAIR, open language data** for all European languages, uses, apps. etc.
- **Global LT/NLP market size by 2025 is \$US 45B:** we want Europe to be a **key player.**
- European digital language and data processing infrastructure crucial for **digital sovereignty.**

EU Parliament STOA Workshop: Nov '22



Relevant ELE reports for UniDive

EU Language Reports:

<https://european-language-equality.eu/deliverables/>

Digital Language Equality (specification) – explaining DLE metric

https://european-language-equality.eu/wp-content/uploads/2022/03/ELE_Deliverable_D1_3.pdf

Summary of survey and consultation process

https://european-language-equality.eu/wp-content/uploads/2021/05/ELE_Deliverable_D2_1.pdf

NLP SOTA, issues and impact on society (2 UniDive authors)

https://european-language-equality.eu/wp-content/uploads/2021/05/ELE_Deliverable_D2_1.pdf

Strategic Agenda and Roadmap (towards digital equality)

https://european-language-equality.eu/wp-content/uploads/2022/11/ELE_Deliverable_D3_4_SRIIA_and_Roadmap_final_version_-1.pdf



European Language Equality

Go raibh maith agaibh
Thank you



The European Language Equality project has received funding from the European Union under the grant agreements № LC-01641480 – 101018166 (ELE) and № LC-01884166 – 101075356 (ELE 2).

Teresa Lynn (DCU)
teresa.lynn@adapcentre.ie
<http://european-language-equality.eu>

