Digital Language Equality in Europe: How are our languages doing?

Maria Giagkou (Institute for Language and Speech Processing, ATHENA RC, Greece)
mgiaigkou@athenarc.gr
08-11-2022 STOA Workshop Towards full digital language equality in a multilingual European Union
http://european-language-equality.eu
Digital Language Equality in Europe: How are our languages doing?

Digital Language Inequality in Europe: How unequal are our languages?
How (un)equal are our languages: evidence 1/n
How (un)equal are our languages: evidence 2/n
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
  • **contextual** factors

\[ \text{TechDLE}_{\text{Lang}X} = \sum_{i=1}^{N} \text{Score}_{\text{LRT}i} \]
Digital Language Equality: Metric

Technological Factors
- Tools and Services
- Corpora
- Language Descriptions and Models
- Lexical and Conceptual Resources

Contextual Factors
- Projects
- Organisations
- Industry
- Online
- R&D&I

- Economy
- Education
- Funding
- Law
- Media
- Policy
- Public administration
- Society
Grounding on evidence

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.

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

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

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

https://live.european-language-grid.eu/
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 (indication of the potential for LT development) with regard to a small number of basic types

- Four bands:
  - Good support (green)
  - Moderate (light green)
  - Fragmentary (yellow)
  - Weak or no support (light yellow)
Cross-language comparison

- Translation technologies: many languages are at least moderately supported
Cross-language comparison

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

- **EU official languages** better supported than other European languages
Cross-language comparison

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

![Diagram showing support levels for various languages](image-url)
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
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
Except for English and, to a certain extent, German, Spanish and French, all European languages are massively under-resourced in terms of data, tools and technologies.
Nearly all European languages exist in a socio-political ecosystem that does not incentivise, encourage or foster the development of technologies for these languages.
2022 vs. 2012

The gap between English and the next cluster of languages is getting **bigger** instead of smaller!
ELE dashboard

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).

Thank you very much

Maria Giagkou (Institute for Language and Speech Processing, ATHENA RC, Greece)
mgiaigkou@athenarc.gr

08-11-2022 STOA Workshop Towards full digital language equality in a multilingual European Union
http://european-language-equality.eu