



EUROPEAN LANGUAGE EQUALITY

META[≡]NET
META[≡]FORUM 2022



Defining Digital Language Equality

Federico Gaspari (DCU), Annika Grützner-Zahn (DFKI)
Federico.Gaspari@adaptcentre.ie, Annika.Gruetzner-Zahn@dfki.de

08/09-06-2022 META-FORUM 2022 – Joining the European Language Grid (hybrid conference)
<http://www.european-language-grid.eu> – <http://european-language-equality.eu>

Overview

- Defining Digital Language Equality
- Digital Language Equality Metric
 - Technological Factors
 - Contextual Factors
- Conclusion and way forward

Defining Digital Language Equality

- Joint work of the ELE Consortium
- Rooted in the **state-of-the-art**, but it will also serve the **future needs** of the languages targeted by ELE and the expectations of the relevant language communities for the **digital age**
- The challenge is to enable **all European languages**, regardless of their specific circumstances, to **realize their full potential**, supporting them in **achieving full digital equality by 2030**

*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*

- Theoretical foundation to achieve **DLE in Europe by 2030**
- Key component of the **evidence-based SRIA** and **roadmap** produced by ELE

Digital Language Equality Metric

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

*It is computed for each language on the basis of various factors, grouped into **technological factors** (e.g., the available language resources, tools and technologies) and a range of situational **contextual factors** (e.g., societal, economic, educational, industrial)*

- The DLE Metric is **modular** and **flexible** with well-defined quantifiers, measures and indicators
- It is **dynamic**, i.e. its scores can be **updated and monitored** over time:
 - at regular intervals or whenever one wishes
 - to check the **status and progress** of one or more European languages **towards DLE**
- This enables **direct comparisons** between (groups of) European languages...
 - ... and helps to **identify gaps** and **prioritize future interventions** based on evidence

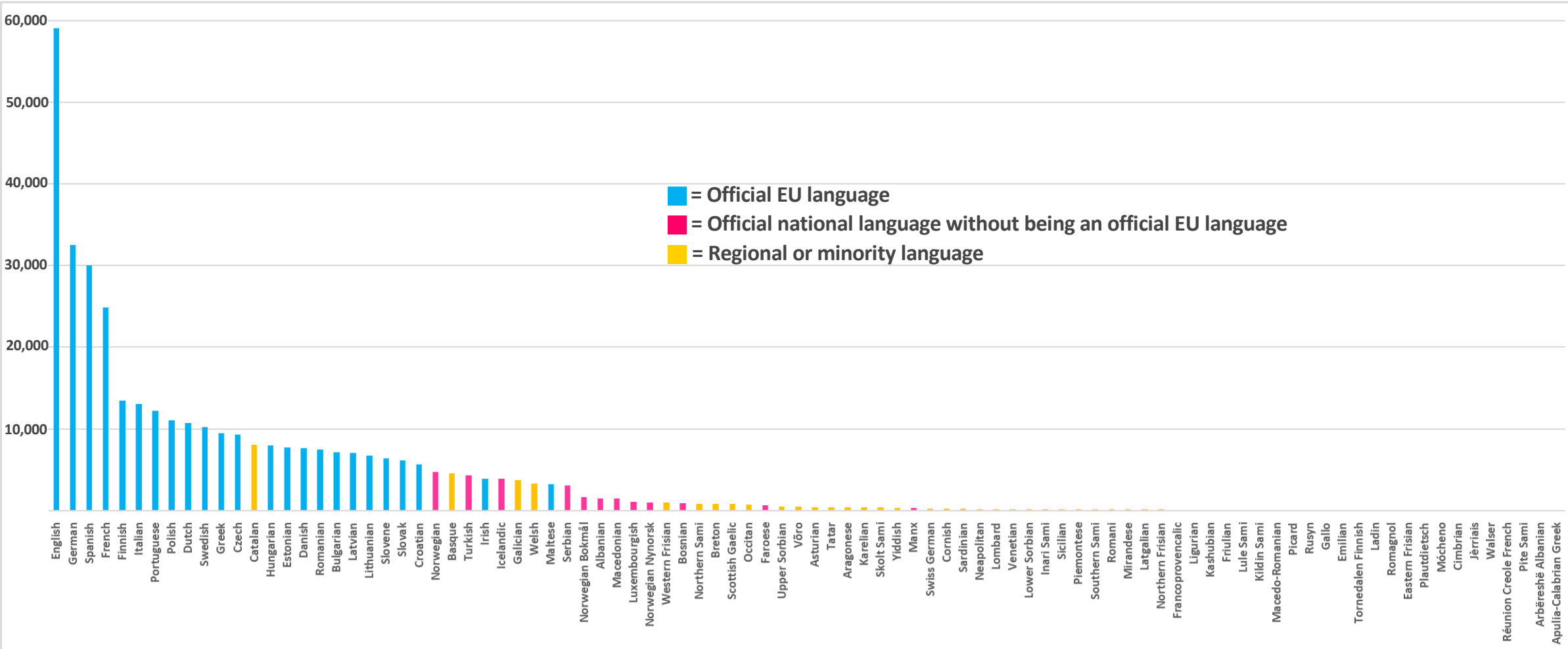
DLE Metric: Technological Factors

- For the TFs, the DLE Metric is fully aligned with the metadata schema of the ELG
- TFs are indicators of linguistic digital readiness, and can be directly compared across languages:
 - **corpora**
 - **language models and computational grammars**
 - **lexical and conceptual resources**
 - **tools and services**
 - (projects)
 - (organizations)
- All LRs and tools have features based on the accompanying metadata in ELG
 - weights are assigned to the feature values
 - higher weights reflect higher complexity or costs, wider availability and applicability

DLE Metric: Computing the Technological DLE Score

- The weights were refined iteratively to reflect a realistic, fair and bias-free scoring mechanism
 - each resource in the ELG Catalogue (dataset or tool) obtains a score (Score_{LR}), which is equal to the sum of the weights of its relevant features
 - e.g. for the features *Annotation Type* and *Domain*, the respective weight is multiplied by the number of unique feature values possessed by the LR in question, as multiple are possible
- To compute the technological DLE score for language X ($\text{TechDLE}_{\text{LangX}}$), including all LRs that support language X (LR1, LR2, ... LRN), we sum up the Score_{LR} of all LRs for language X (LR1, LR2, ... LRN):
 - **$\text{TechDLE}_{\text{LangX}} = \text{Score}_{\text{LR1}} + \text{Score}_{\text{LR2}} + \dots + \text{Score}_{\text{LRN}}$**
- The same process applies to tools and services, with the relevant weights assigned to feature values
- The partial scores for LRs and tools for language X are then summed up to give the overall technological DLE score for the language in question

DLE Metric: Technological Factors – Results



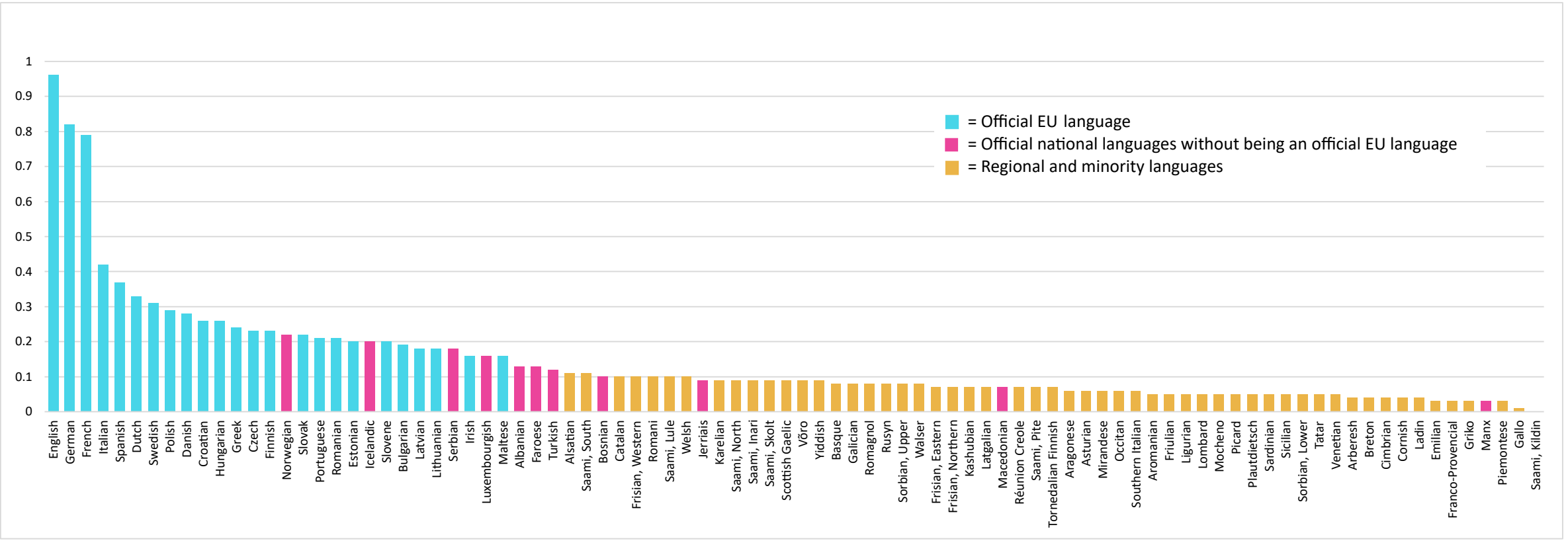
DLE Metric: Contextual Factors

- CFs represent the **general condition and situation of the broader context** regarding the development of Language Resources and Technologies (LRTs)
- CFs indicate the **potential of a language to achieve Digital Language Equality**
 - Low score → context with limited possibility to evolve: *without* political will, funding, innovation and economic interest in the region
 - **Lower probability** that LRTs will be developed in the coming years
 - High score → good context with the possibility to evolve: *with* political will, funding, etc.
 - **Higher probability** that LRTs will be developed in the coming years

DLE Metric: Contextual Factors – Selection and Calculation

- 72 CFs with an impact on the development of LRTs
- Selection of CFs based on the following criteria:
 - Data availability
 - Automatic updatability
 - Data quality
 - Balance of the factor classes → not more than two factors per class (economy, industry, etc.)
 - Balance of the data types → approximately equal share of total and proportional numbers
- Evaluation of the results in the ELE consortium → inclusion of three additional factors
- Calculation: relative score per factor per language → average of all scores per language = overall score of the language

DLE Metric: Contextual Factors – Results



Conclusion and way forward

- The DLE Metric is a tool to achieve digital equality for all of Europe's languages by 2030
 - it combines the technological aspects with the contextual situation of the languages
 - it can guide and help to prioritize future LT development and LR creation, collection and curation
- Balance between theoretical rigor and intuitiveness
 - for a range of stakeholders such as policy-makers, industry leaders, researchers, developers, etc.
- Essential component of the evidence-based SRIA and roadmap produced by ELE
 - to guide future work to achieve full DLE in Europe by 2030
- DLE Metric scores (TFs and CFs) for all European languages can be interactively visualized through the online Dashboard
 - *more on this in the next presentation!*

Details at the TDLE Workshop at LREC 2022 in Marseille ...

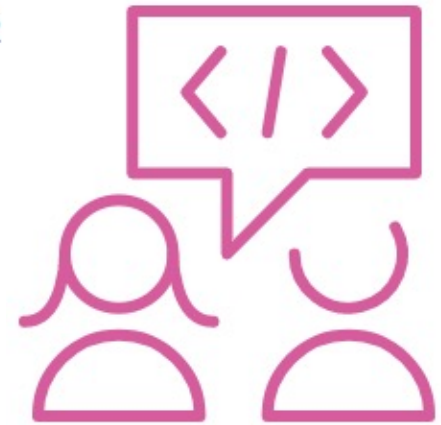


Towards Digital Language Equality

TDLE 2022 – CO-LOCATED WITH LREC 2022

20 June 2022

Palais du Pharo, Marseille, France



european-language-equality.eu/tdle-2022/



European Language Grid
European Language Equality



Thank you!



The European Language Grid has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement № 825627 (ELG).

The European Language Equality project has received funding from the European Union under grant agreement № LC-01641480 – 101018166 (ELE),

Federico Gaspari (DCU), Annika Grützner-Zahn (DFKI)
Federico.Gaspari@adaptcentre.ie, Annika.Gruetzner-Zahn@dfki.de

08/09-06-2022 META-FORUM 2022 – Joining the European Language Grid (hybrid conference)
<http://www.european-language-grid.eu> – <http://european-language-equality.eu>