



EUROPEAN LANGUAGE EQUALITY



European Language Equality: Preliminary Results LRT Collection and Language Reports

Maria Giagkou (ILSP, R.C. "Athena")

15/16/17-11-2021 META-FORUM 2021 – Using the European Language Grid (virtual conference)

<http://www.european-language-grid.eu>

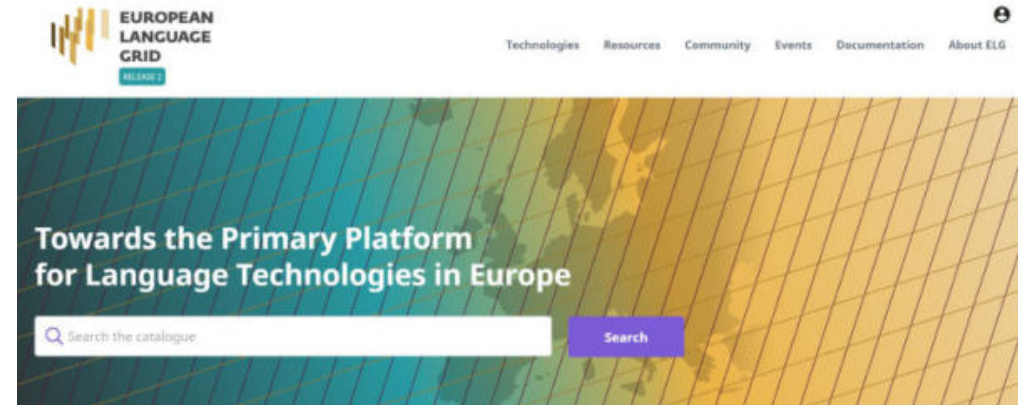
Assessing the current support of Europe's languages through Language Resources and Technologies (LRTs)

- **Methodology:**

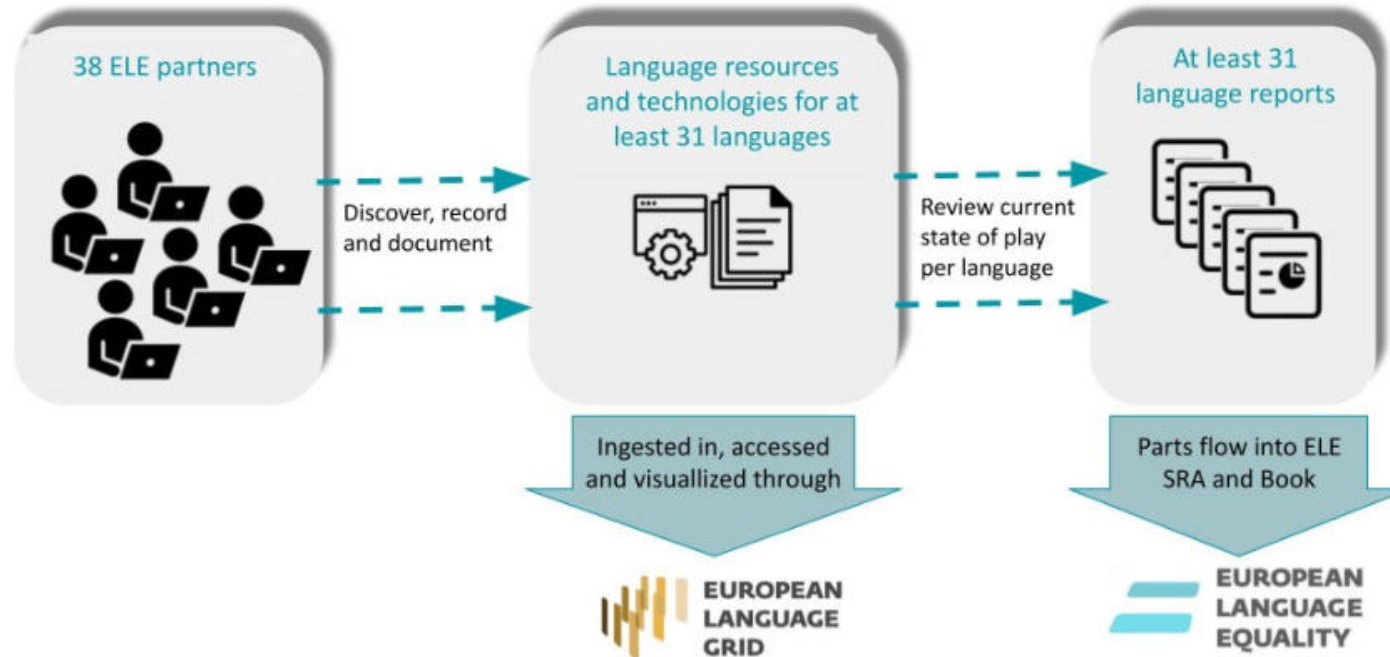
- Discover LRTs for each language
- Document LRTs
- Populate and update the European Language Grid platform
- Compile a report per language, based on evidence collected

- **Outputs:**

- META-NET White Papers series update
- Aggregated data searchable and visualized through the ELG interface



Compiling a collection of LRTs representative of the current state of digital equality of European languages



List of all languages investigated by ELE and partners involved:
<https://european-language-equality.eu/languages/>

What is in scope

- **Corpora:**

- textual, audio, video, images, multimodal
- monolingual, bi-/multilingual, parallel, comparable
- raw or annotated

- **Lexical/conceptual resources:**

- computational lexica, terminological databases, ontologies, thesauri, wordnets, gazetteers, term lists, etc.

- **Language models & computational grammars:**

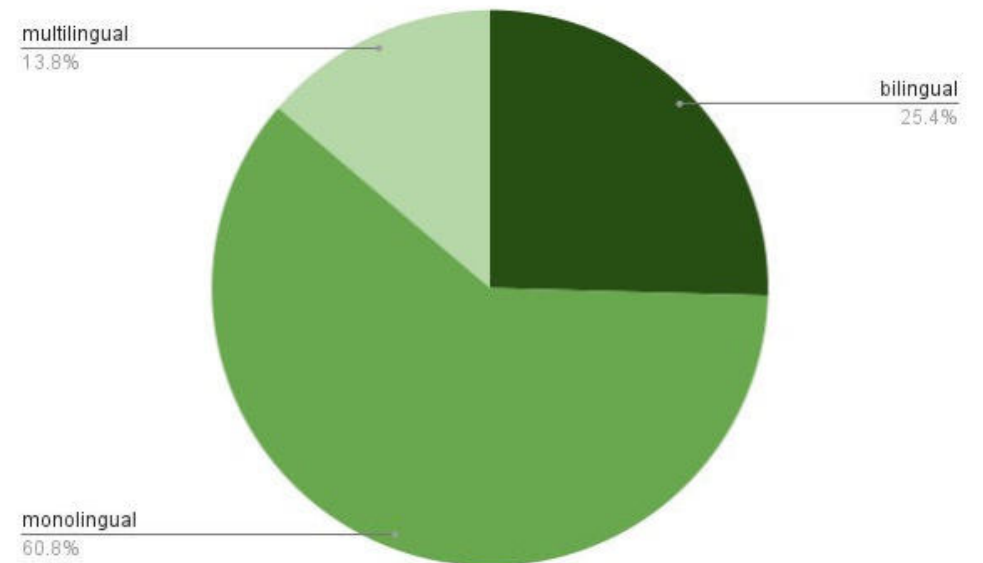
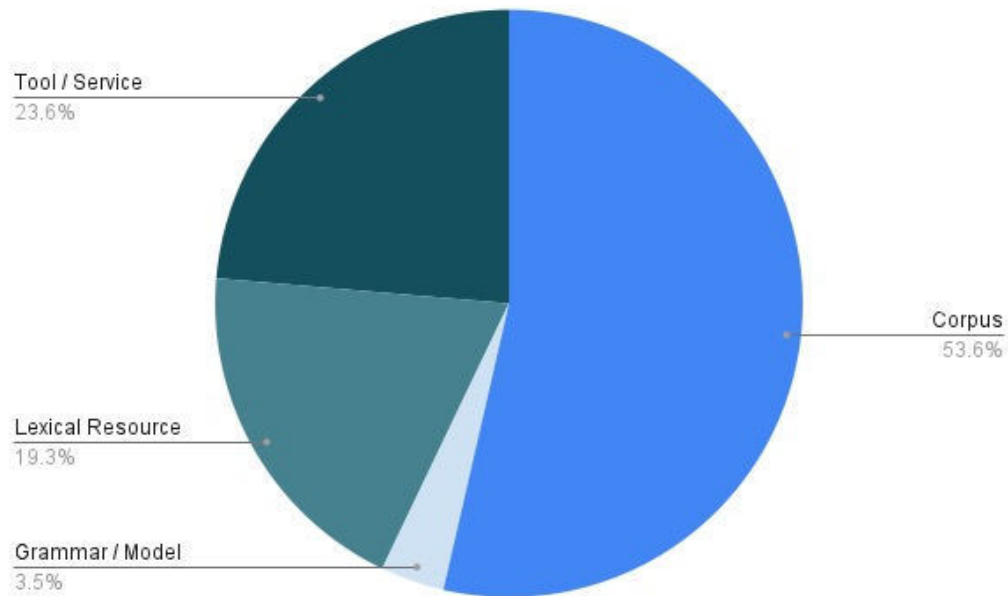
- ML models
- n-grams
- computational grammars

- **Tools/services**

- All the basic NLP tools for a language, from tokenization to syntactic parsing
- Text and speech analytics/mining
- Machine Translation
- NLU / NLG
- Conversational systems
- ...

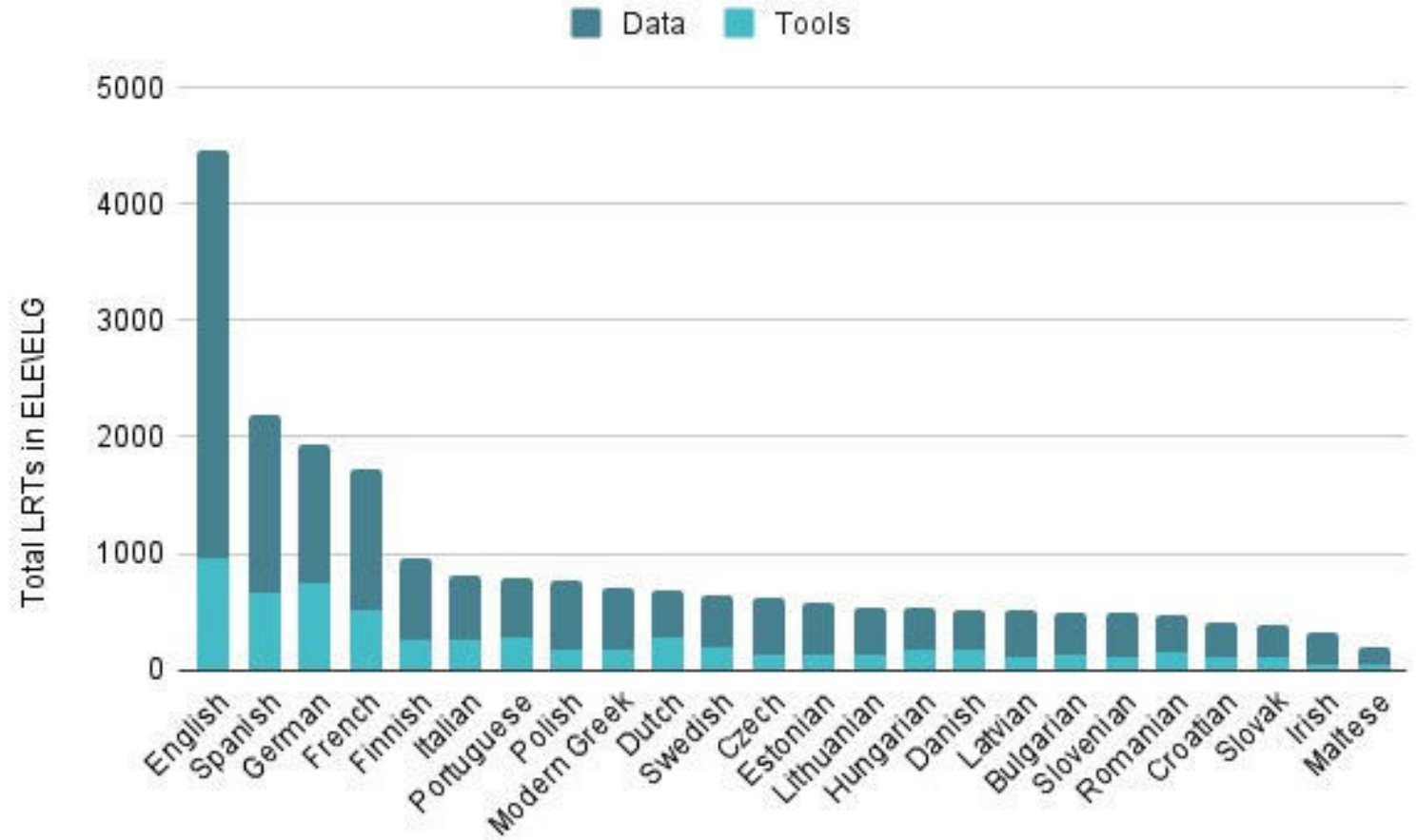
First results: overview

- **6.000+** new LRTs contributed by ELE partners (to appear in ELG)
- **11.000+** LRTs in the consolidated ELE & ELG collection

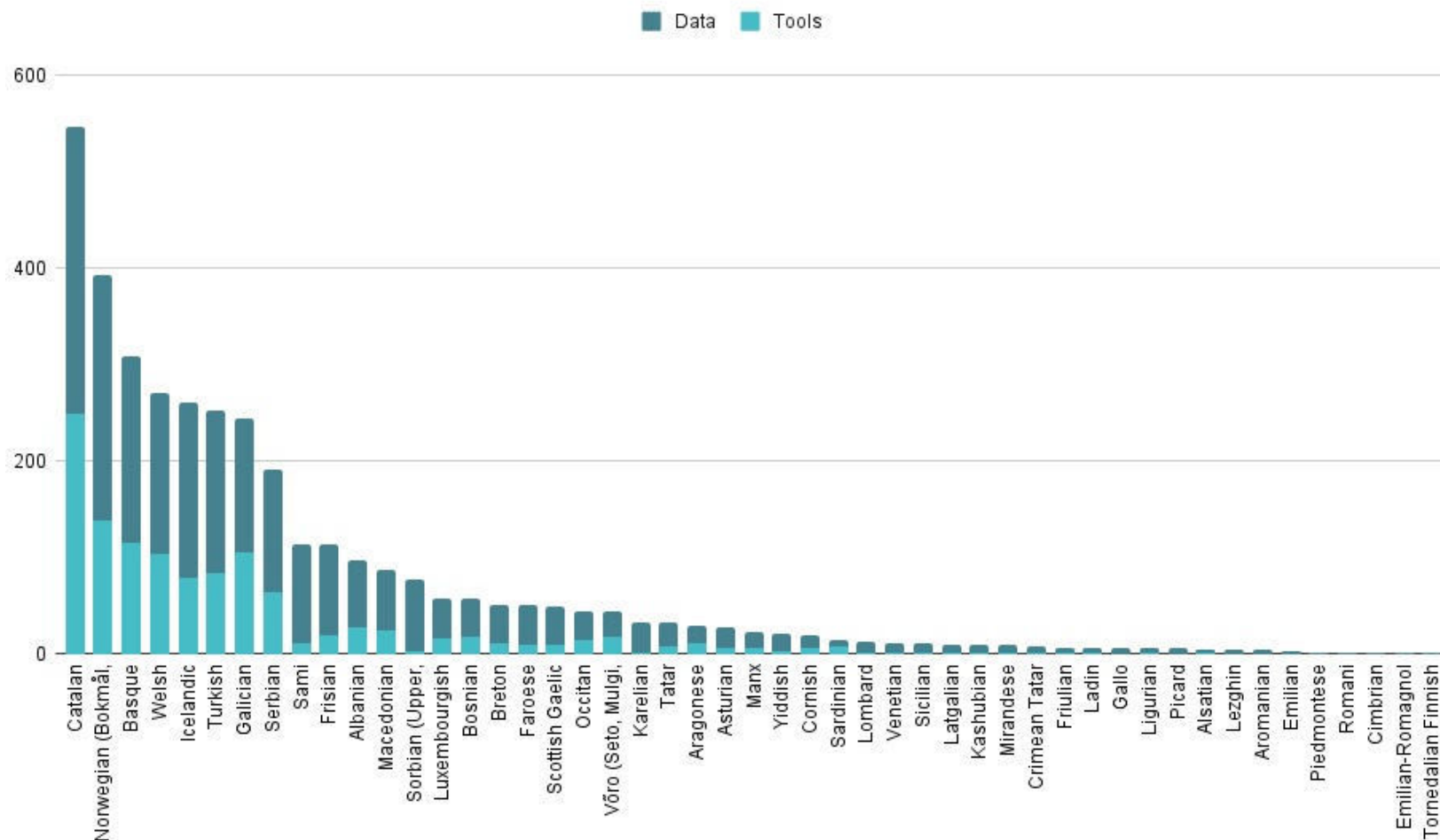


First results: EU-24 official languages

Languages	% of LRTs represented in / supported by
English	40%
Spanish	20%
German	17%
French	15%
Finnish Italian Portuguese Polish Greek Dutch Swedish Czech Estonian Lithuanian Hungarian Danish	5% to 10%
Latvian Bulgarian Slovenian Romanian Croatian Slovak Irish Maltese	<5%

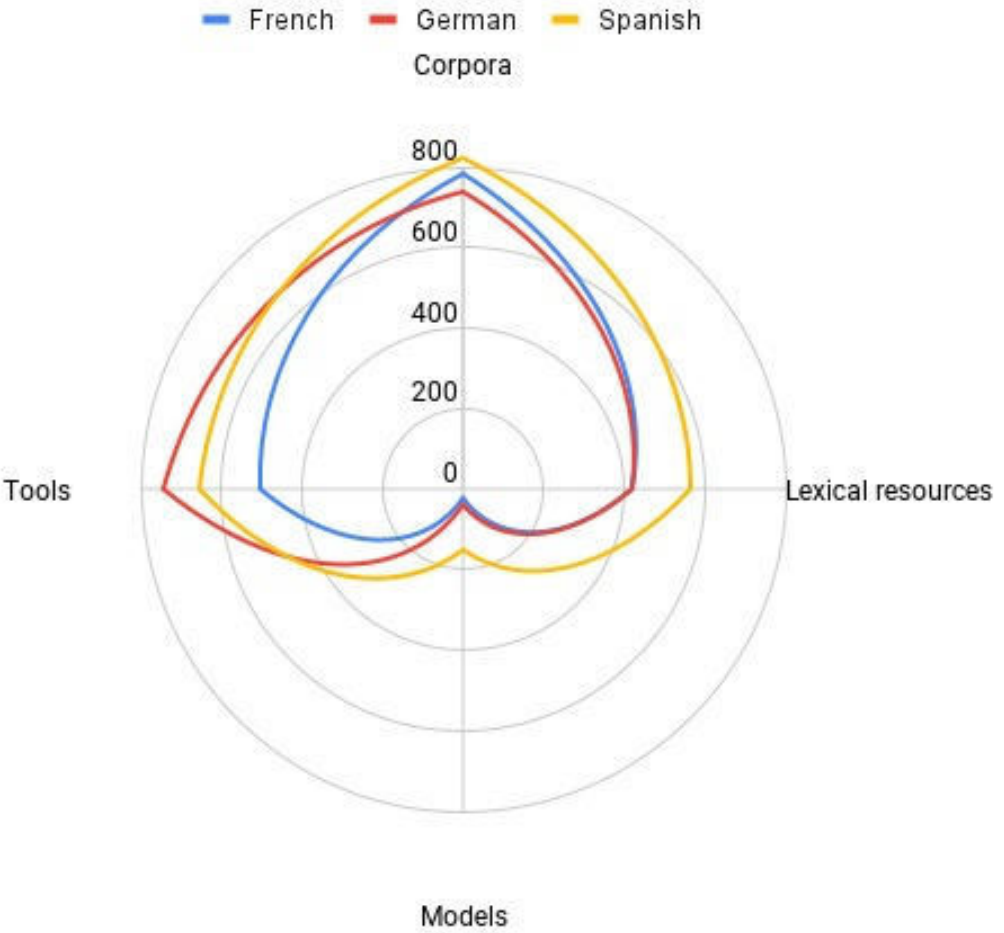


First results: non-EU-24 languages

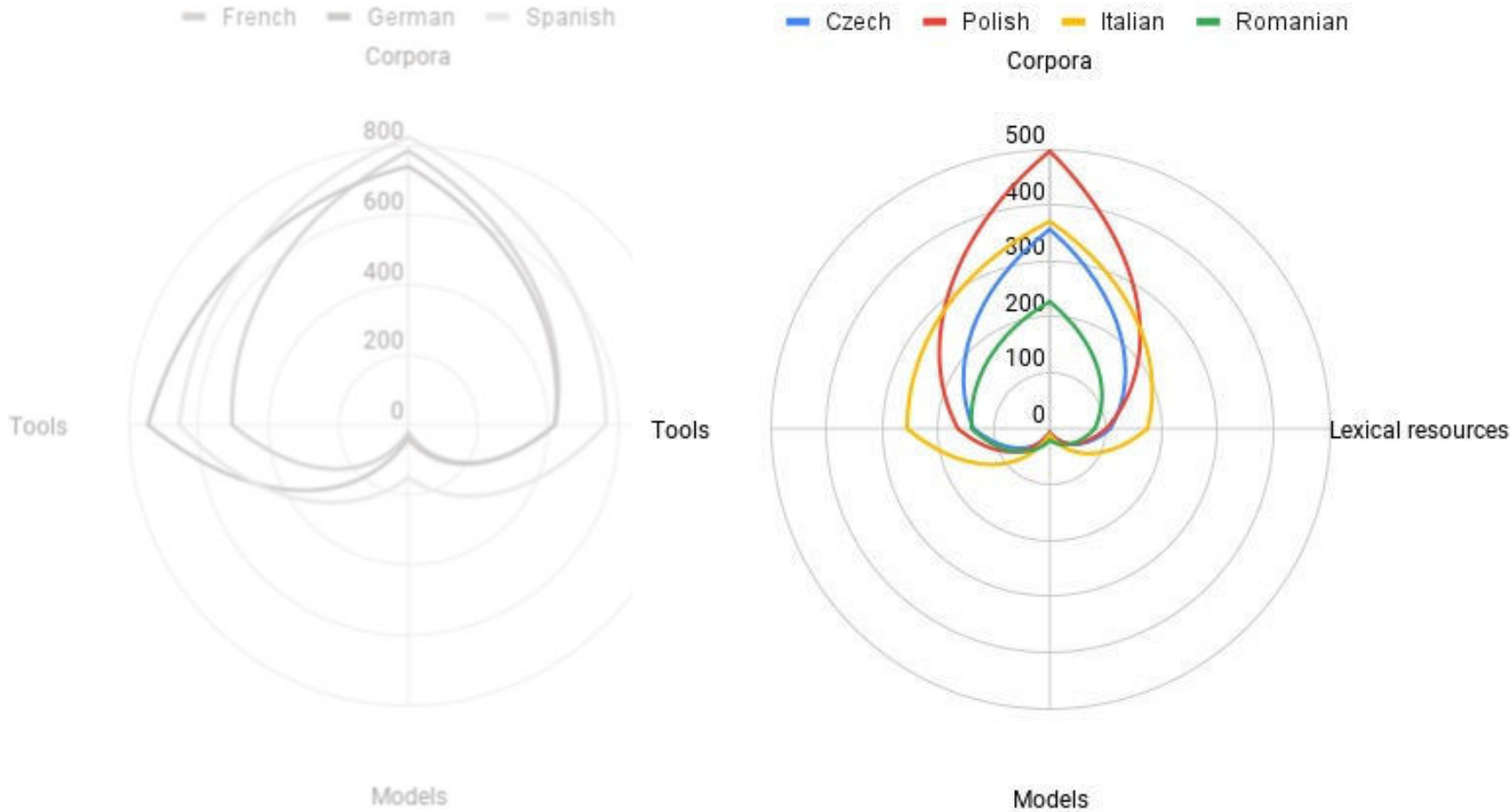


No LRTs identified for Arberesh, Carpathian-German, Carpato-Rusyn, Franco Provençal, Griko, Indigenous Languages in Guyane, Jerriais, Meskhetian, Mocheno, Plattdeutsch, Réunion Creole, Southern Italian, Walser

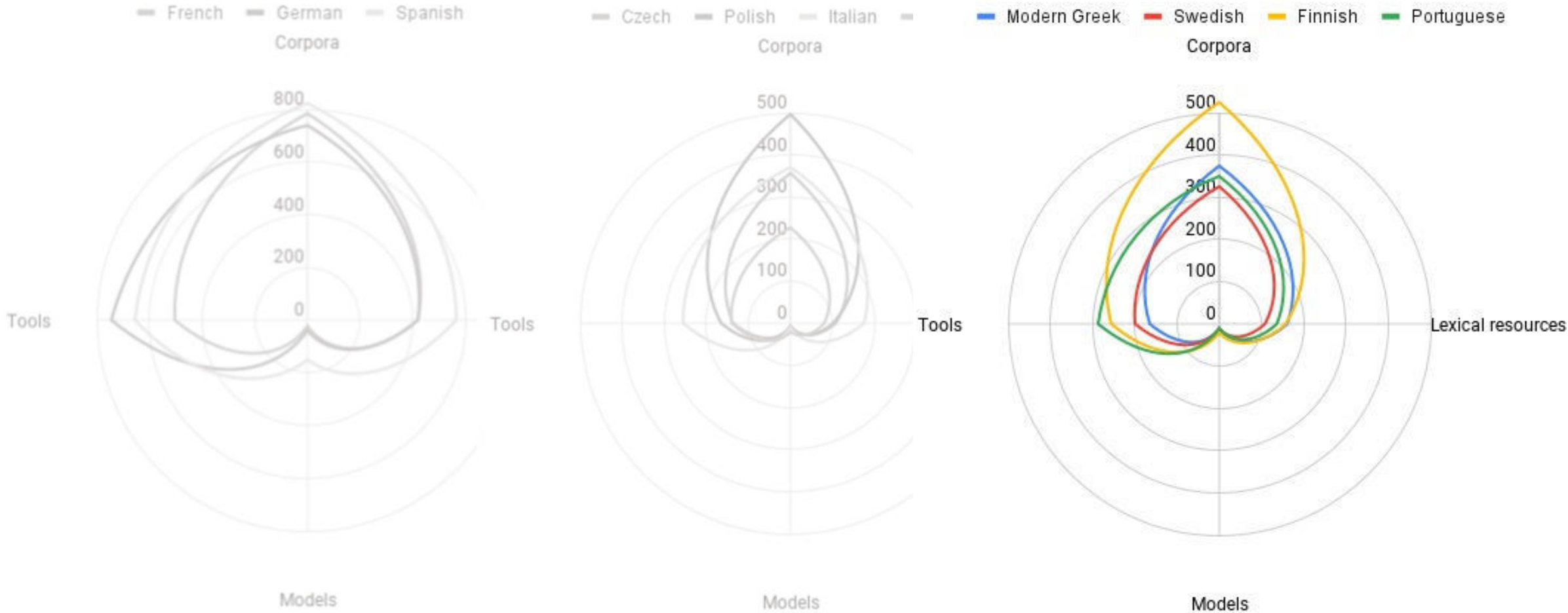
First results: comparing languages based on #LRTs



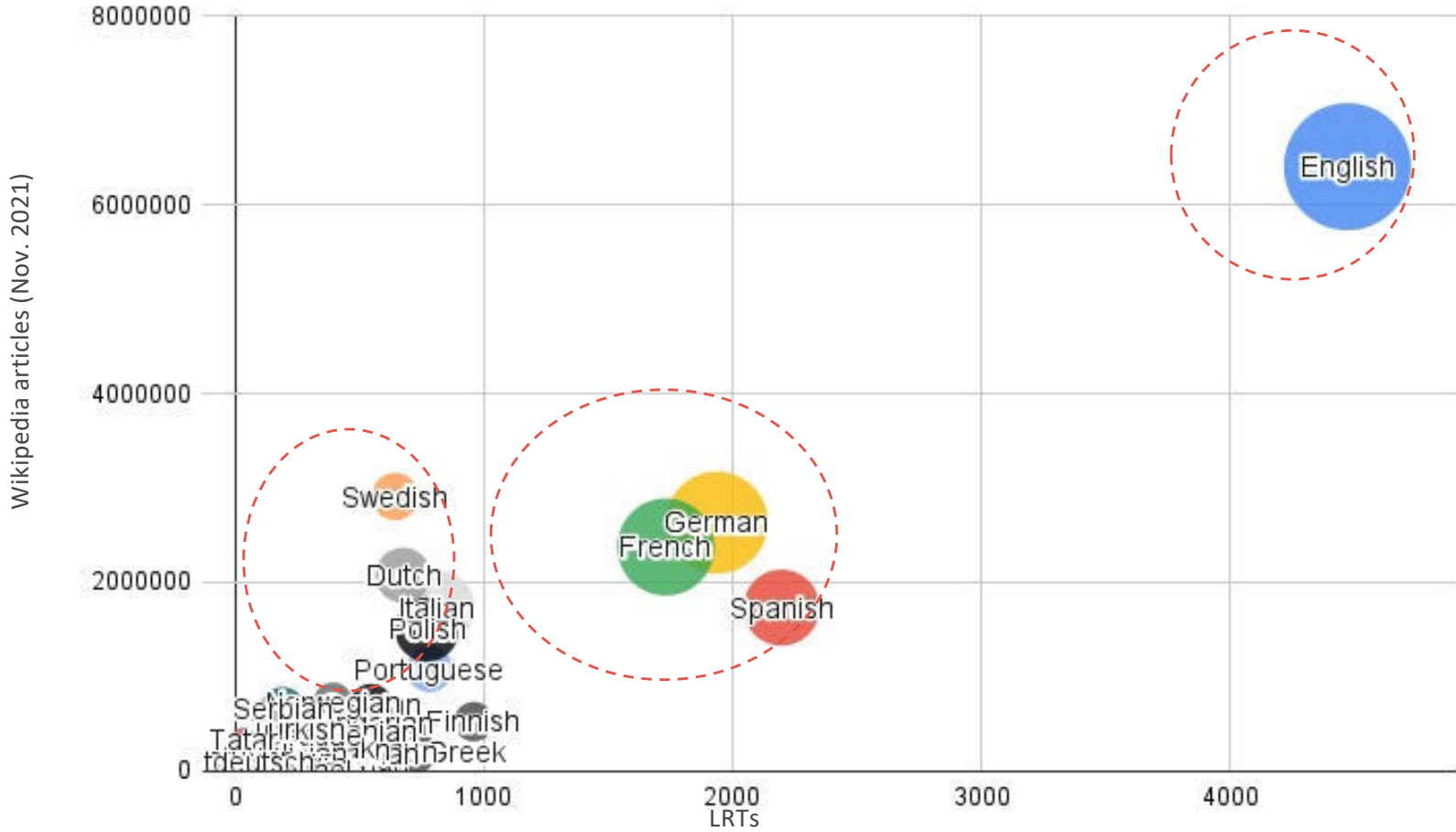
First results: comparing languages based on #LRTs



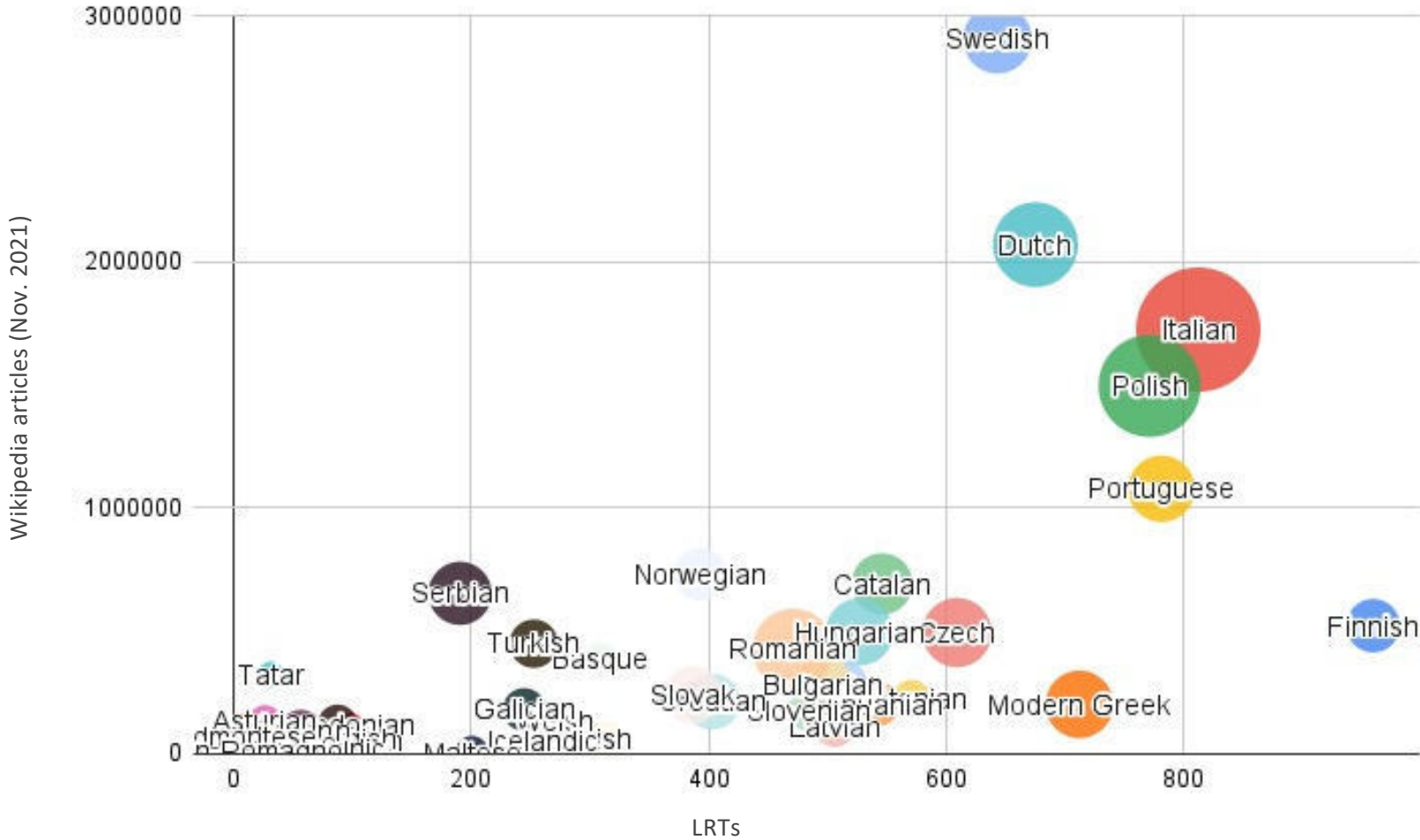
First results: comparing languages based on #LRTs



First results: introducing a contextual factor: presence on the internet (1/2)



First results: introducing a contextual factor: presence on the internet (2/2)



Language Reports

- Tradition of META-NET White Paper Series
- Quantitative and qualitative evaluations of the level of support each language receives through technology:
 - brief introduction to the LT field
 - general facts about the language
 - availability of LRTs based on the combined collection of ELG and ELE resources
 - support it receives through dedicated funding programs and projects
 - participation in research infrastructures
 - ...
- Stand-alone digital publications; parts will be included in the final ELE agenda and book.



Thank you!



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

Maria Giagkou (ILSP, R.C. “Athena”)

15/16/17-11-2021 META-FORUM 2021 – Using the European Language Grid (virtual conference)

<http://www.european-language-grid.eu>