



EUROPEAN LANGUAGE EQUALITY



European Language Equality: Preliminary Results European LT Developers Survey

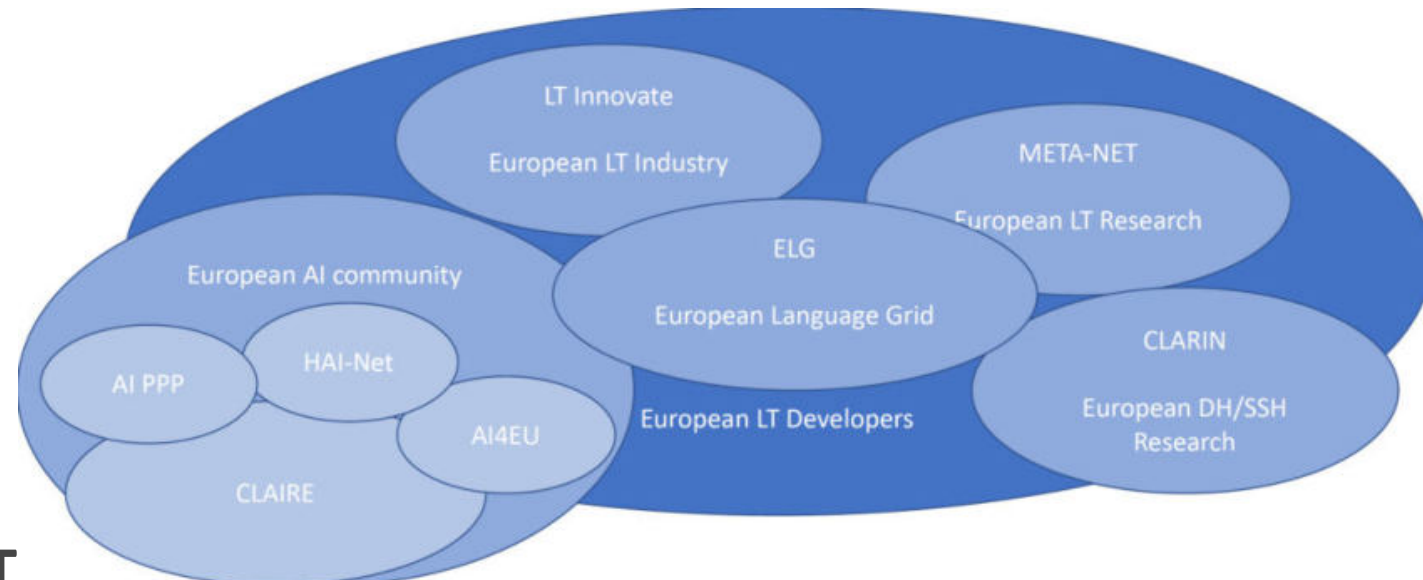
Maria Giagkou (ILSP, R.C. "Athena")

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

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

Consultation with LT developers: overview

- **Objective:** Collect and analyze the views of European LT developers and providers (research and industry) with regard to Digital Language Equality
- **Instruments:**
 - Survey
 - Focused interviews
- **Target groups:**
 - as per type of stakeholder: both **research** organisations and the **industry**, and
 - as per discipline/field: the **LT**, **AI** and **DH/SSH** communities




Survey overview (1/2)

- **Part A. Introduction:** profiles of respondents, country, organization, role, expertise etc. e.g.
 - Which LT areas do you mainly work in?
- **Part B. Language coverage,** e.g.
 - Which languages do you conduct research in or support in your products/services?
- **Part C. Evaluation of current situation:** current strengths, gaps and challenges for the European LT research and industry, e.g.
 - Please indicate if you agree with the following statements: “One of the main challenges and obstacles the European LT community currently faces is... ”
- **Part D. Predictions and visions for the future,** e.g.
 - Which are the top 3 research, technology development, or socio-economic opportunities that you personally envisage the ELE to bring about or to successfully address?

Survey overview (2/2)


- 32 (minimum) to 45 (maximum) questions
- Published on EU Survey platform (open 17 June – 18 October 2021)
- Disseminated to:
 - **META-NET** (CUNI)
 - **ELG** (DFKI)
 - **CLAIRE** (Univ. Leiden)
 - **CLARIN** (CLARIN-ERIC)
 - **LT-innovate** (Crosslang)
 - **All others** (ILSP/CUNI)
- Full results to be published on the ELE website (Feb. 2022)

	Mandatory	Optional	Totals
Closed	24	3	27
Open-ended	2	16	18
Totals	26	19	



ELE
European Language Equality: Consultation with LT researchers and developers

Fields marked with * are mandatory.



**EUROPEAN
LANGUAGE
EQUALITY**

About this questionnaire

This questionnaire is delivered by the [European Language Equality \(ELE\)](#) project, a pilot action that addresses an appeal by the European Parliament resolution ["Language equality in the digital age"](#). The primary goal of ELE is to prepare a Strategic Research and Innovation Agenda and Roadmap, in order to tackle the striking imbalance between European languages in terms of the support they receive through language technologies.

To this end, ELE is reaching out to the European stakeholders involved in Digital Language Equality through a series of consultation rounds. This questionnaire is specifically addressed to **researchers and industry practitioners in the field of Language Technology (LT), Natural Language Processing (NLP), Speech Technologies and Language-centric AI**.

The questionnaire takes approximately 20 minutes to fill in. You are requested to evaluate the current situation with respect to the level of LT support for European languages, to indicate challenges and to share your needs and expectations for the future.

Your contributions will be carefully taken into account when preparing the ELE strategic agenda and roadmap.

This is a joint pan-European effort that will impact the field of LT in Europe for the next 10-15 year including the funding situation. Join us and be a part of it!

Personal data protection

*Which LT areas do you mainly work in?

- Basic natural language processing services (PDF tagging, parsing, named entity recognition etc.)
- Search and information retrieval technologies
- Text analytics and mining, information extraction, text classification
- Translation technologies (Machine Translation, translation resources management, CAT tools)
- Speech technologies
- Conversational systems
- Language resources: data production, data aggregation
- Language resources: data distribution, data marketplace
- Research infrastructures (e.g. catalogue, repository)
- Other

If "Other", please specify:

* Are you/your organization a member of one or more of the following associations/networks/projects?

- CLARIN
- META-NET
- ELG
- CLAIRE
- LT-innovate
- A4EU
- ELEXIS
- RDVA
- Ai PPP
- Horizon AI Network
- Horizon Linguamorph
- ELISE
- TALOR
- AIMedia
- VIGOR
- AHCopenhagen
- AIPan4EU
- EuroAPPs
- CHAI
- iNERGY
- StormAI
- Other
- None of the above

If "Other", please specify:

How many organizations participate in your national CLARIN consortium?

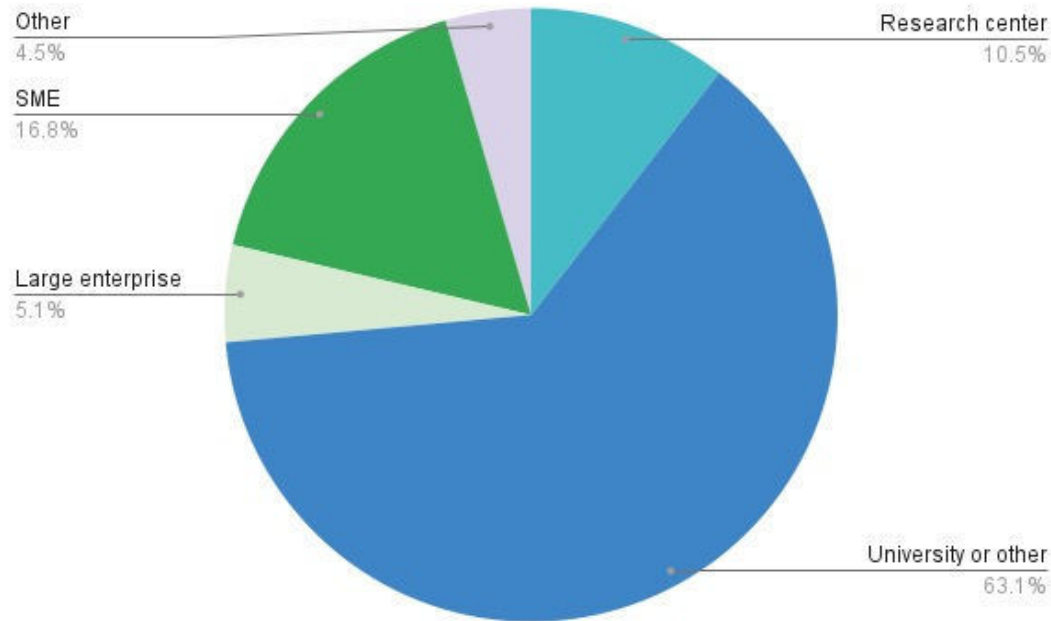
How many LT researchers/experts/students are employed and/or actively contribute to the national CLARIN consortium?
Please do not report the number of students using the resources in education only. Only the number of adult contributors is relevant.

In which sectors are your technologies, products or services used?

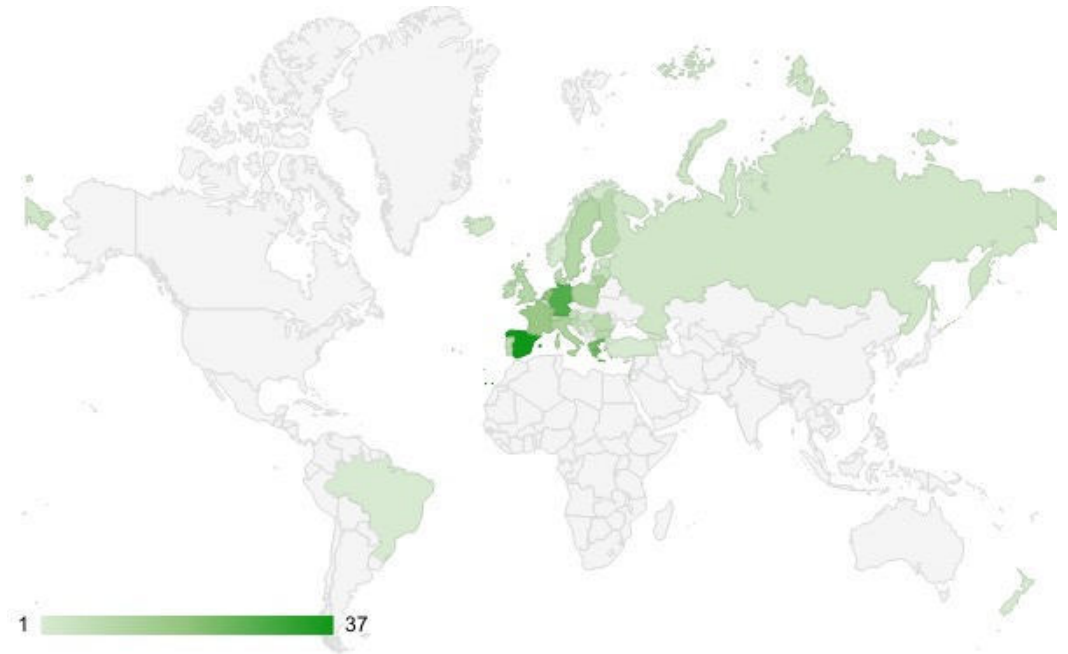
- Agriculture and fisheries
- Insurance industry

Respondents' profiles (1/2)

- 333 responses
- 74% from research/academia



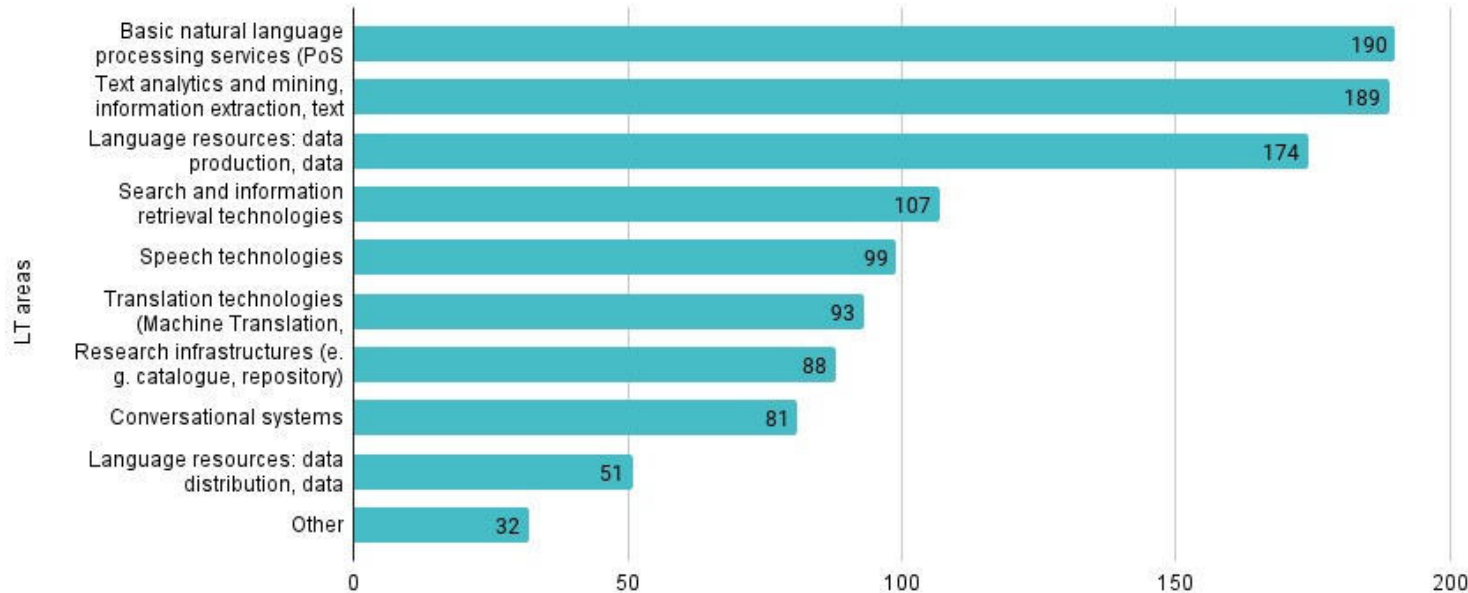
- 85% from EU countries



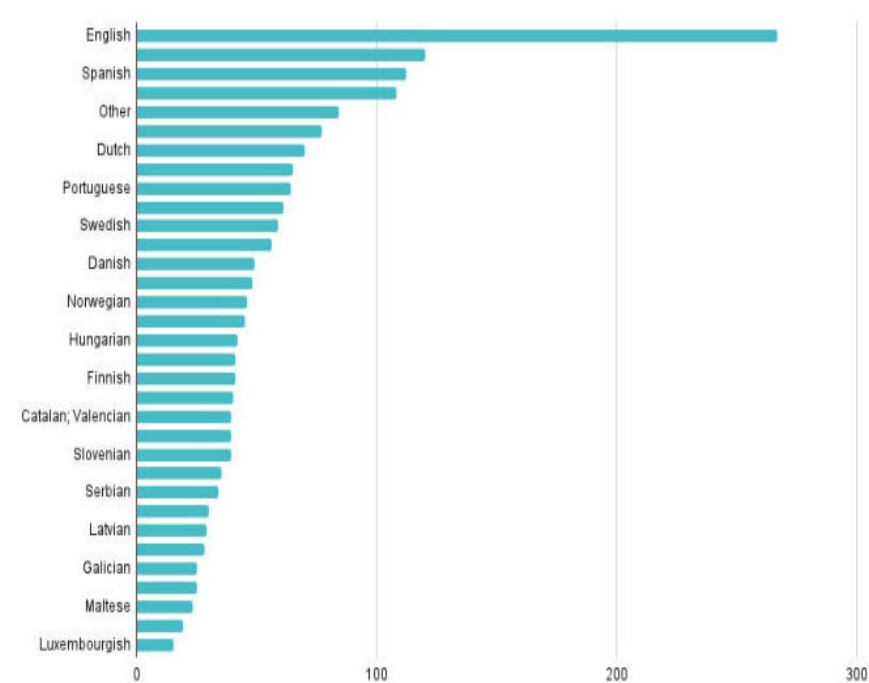
*Respondents to the interviews are not included in this analysis

Respondents' profiles (2/2)

Respondents active in all LT areas



All EU-24 languages covered, plus co-official, regional and minority



Evaluation of current situation

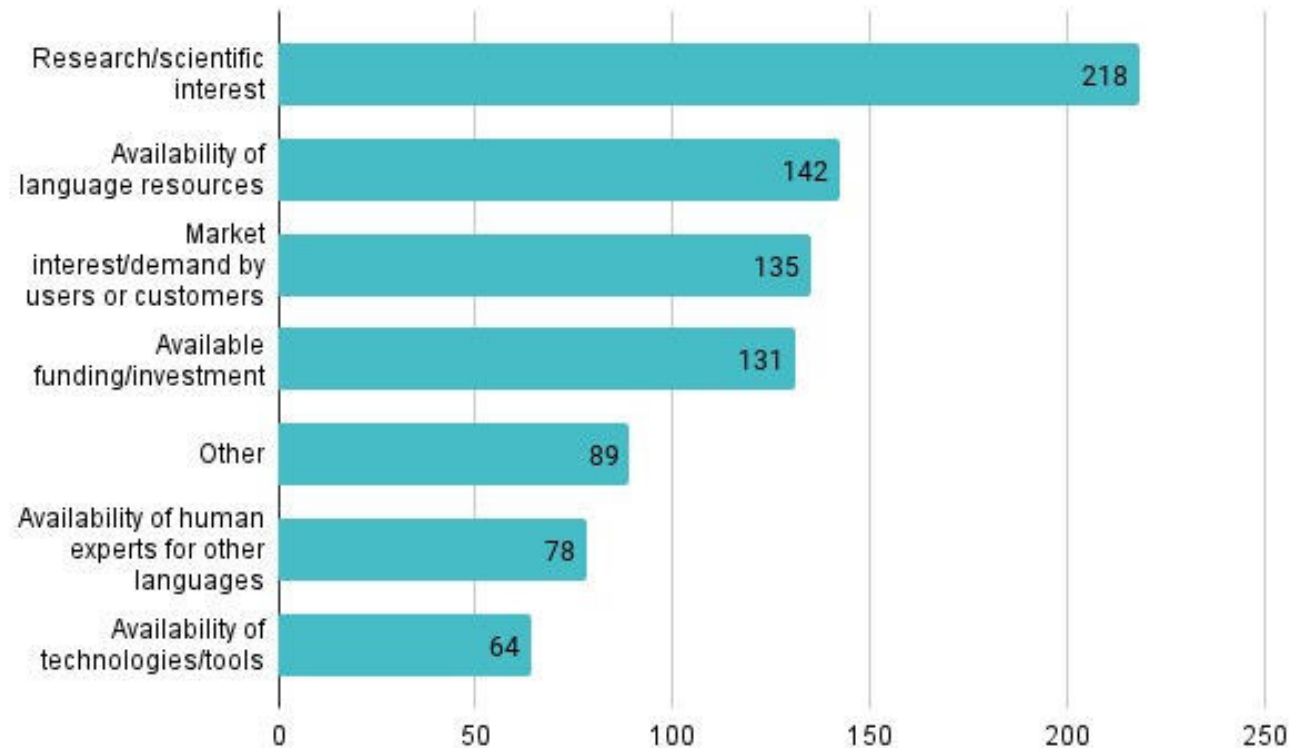
<i>Please indicate if you agree with the following statements: “One of the main challenges and obstacles the European LT community currently faces is...”</i>	Avg. Score (4: Strongly agree, 1: Strongly disagree)	Strongly agree & Agree %
basic research is still needed	3,3	88%
inadequate recognition of the importance of multilinguality	3,1	79%
lack of talent/brain drain	2,4	46%
fragmentation of the European LT industry	3,0	82%
lack of coordination and missing links between research, LT vendors, integrators and customers	3,0	78%
insufficient public procurement	3,0	77%
insufficient markets to justify investments in LTs for smaller languages	3,0	75%
cost of access to compute infrastructure	2,7	57%
competition with non-European big companies and market disruption by global players	3,2	82%

- **A very short summary of open-ended answers:**

- *Support basic research in Linguistics and language modelling, cross-lingual transfer learning, multimodal communication, including speech and sign languages*
- *incentives for research on smaller languages are lacking not only because of the lesser market interest, but also because research publications for smaller languages in the LT fora are often not considered impactful enough*
- *smaller languages rely on public funding to balance the lack of market interest*
- *the threat of digital extinction has not been realized by the policy makers or by the wider public*

Future perspectives (1/3)

Considering your development plans with respect to language coverage, what are the top three drivers for your decision to support additional languages? (multiple choice)



Future perspectives (2/3)

<i>In your opinion, how effective can the following policies/instruments be in speeding up the development and deployment of LT in Europe equally for all languages?</i>	Avg. Score (5: Very effective, 1: Not at all)
Initiate large-scale, long-term funding programme for European LT development	4,24
Continuous investment in the Research Infrastructures that support LT	4,23
Invest in the development of new (scientific/technological) methodologies for transfer/adaptation of resources/technologies to other domains and languages	4,05
Increase availability of qualified personnel on LT and incentives for talent retention	4,03
Reinforce training and education initiatives, including undergraduate and masters programs and vocational training in LT	4,02
Initiate investment instruments and accelerator programs targeting LT start-ups	3,84
Public procurement of innovative technology and pre-commercial public procurement	3,79
Raise awareness of the benefits for companies, public bodies, and citizens of the availability of on-line services, contents and products in multiple languages	3,74
Impose content accessibility regulations, e.g., multimedia subtitling, readability, dubbing, availability of content in multiple languages etc.	3,70

Future perspectives (3/3)

A very short summary of conclusions:

- **Resources:** By 2030 all European languages should have developed the critical mass of resources that are adequate for developing Language Technologies:
 - raw data
 - massive multilingual language models
 - standardized training and evaluation data
- **Reuse terms and legal framework:**
 - Resources should be available for free or at a reasonable cost for research and commercial purposes
 - revised legislative framework for facilitating the use of language data and the application of data mining techniques
- **LT development:**
 - automations for the construction, annotation and curation of language data
 - address the issue of data bias
 - continuous work on improving transfer learning methods
- **Scientific goal:**
 - Deep Language Understanding by 2030 (“hybrid intelligence”, “cognitive AI”, “symbolic AI”, etc.): a leap from language processing to language understanding and seamless human-like interactivity, viable discourse interpretation and ubiquitous natural language interfaces for all Europeans in their own language.



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>