



EUROPEAN LANGUAGE EQUALITY



European Language Equality: Preliminary Results Overview of LT/AI Strategies in Europe

German Rigau (EHU)

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

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

WP3 Objectives

Evidence-based **strategic research agenda** and **roadmap** setting to ensure that digital language equality in the EU becomes a reality by 2030, through:

1. Task 3.1: Desk research – landscaping
200+ relevant documents and research papers so far!
2. Task 3.2: Consolidation and aggregation of all input received
53 deliverables!
3. Task 3.3: Final round of feedback collection
52 partners and many relevant stakeholders!

D3.1 Completed

D3.1 Report on existing strategic documents, white papers and projects in the LT area and related fields (M4: April 2021)

Dynamic document; updates every 6 months.

Includes:

150 pointers to relevant documents.

43 citations to relevant research papers.

Document updated in November 2021 with 20 additional pointers to relevant documents and 10 new citations.

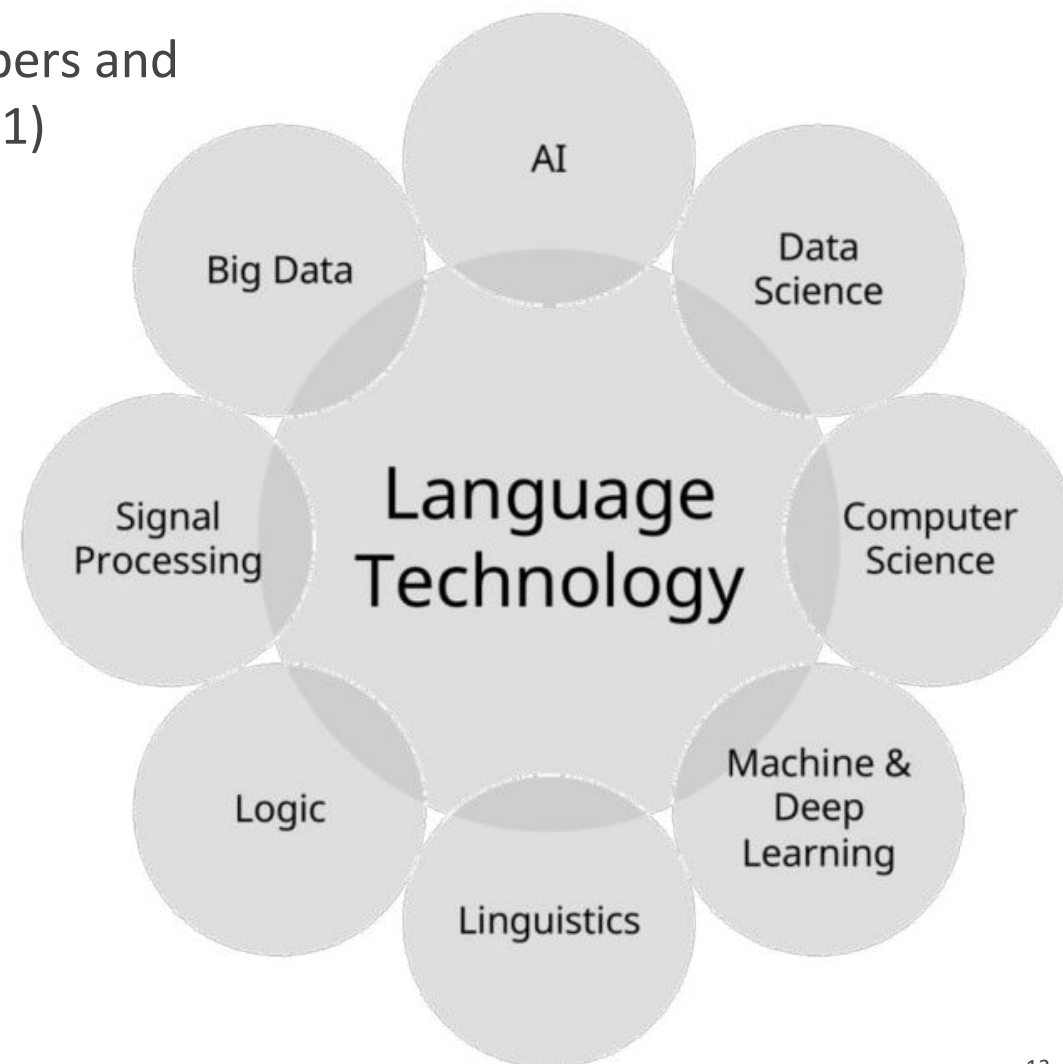
Your input to the next version is much appreciated!



D3.1 Contents

D3.1 Report on existing strategic documents, white papers and projects in the LT area and related fields (M4: April 2021)

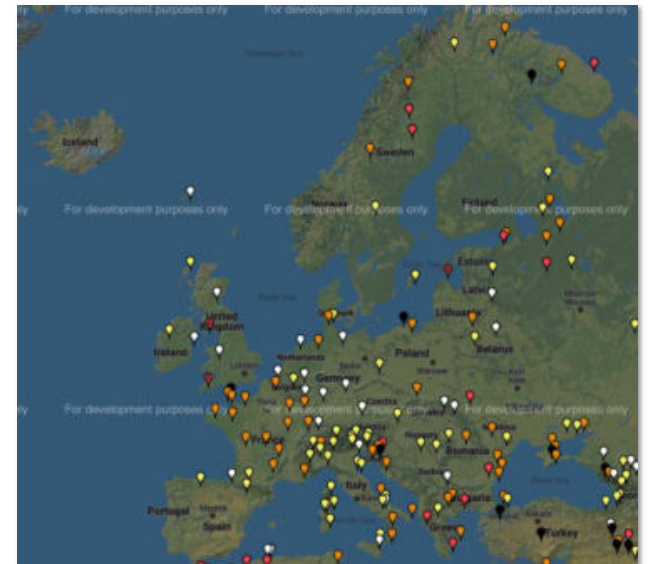
- Abstract and Introduction
- LT: General Overview
- LT in International Organizations
- LT in European Initiatives
- National LT Initiatives in Europe
- Non-EU National Initiatives
- SWOT Analysis
- Recommendations



D3.1 Main Findings

- **LT** one of three core application areas in **AI**, together with **Vision** and **Robotics**.
- **LT** allows many advanced applications, **unthinkable** only a few years ago.
- Even more **exciting results** in the near future.
- **LT** is one of the **most relevant** technologies for society.
- **LT** has a fast-growing **economic impact**.
- However, LT is **not equally** developed for all European languages.

Endangered European languages according to the UNESCO Atlas of the World's Languages in Danger (2010)



D3.1 SWOT Analysis: Strengths

- Powerful new **deep learning** tools are revolutionizing **LT**.
- Important LT applications are **available**.
- Existence of multiple and relevant **stakeholders**. Existence of very useful **data resources** that can be exploited by current **LT**.
- Increasing number of **companies** in **LT**
- Good level of **readiness** for deploying **LT** in production environments.
- **LT** contributes to **inclusive** digital societies.

D3.1 SWOT Analysis: Weaknesses

- Deep learning LT and large pre-trained language models have **shortcomings** and **limitations**.
- Large **non-EU** actors dominate the multilingual European LT market.
- LT plays a minor role in the **political agenda** and **public debate** of the EU and its member states.
- General **misconception** and over-hyping of the actual **AI** and LT capabilities.
- No common **EU policy** has been proposed to address the problem of language barriers.
- **GDPR/Copyright** a major barrier to access and re-use of LRs → Data Governance Act (Nov. 2021)
- Lack of adequate LT policies.
- Scarce and limited LT support for **non-official** EU languages.
- **Computing divide** between large firms and non-elite universities. → EuroHPC JU (Nov. 2021)
- **Uneven** distribution of resources by country, region and language.

D3.1 SWOT Analysis: Opportunities

- Many new powerful **deep learning** LT capabilities are available.
- LT is key to support the European **multilingualism**.
- LT is used in practically **all** everyday digital products and services.
- LT can impact on sectors of fundamental importance to the **well-being** of all European citizens, such as health, administration, justice, education, culture, tourism, etc.
- LT offers effective **solutions** to facilitate language-related disabilities or impairments.
- LT is one of the most important AI application areas with a fast-growing **economic impact**.

D3.1 SWOT Analysis: Threats

- 21+ European languages in danger of **digital extinction** (META-NET White Papers)
- Development of **non-explainable** deep learning models without any **common-sense** knowledge, with social **biases**, containing **personal** data, with a very heavy impact on **carbon footprint**, etc.
- **AI** overshadows the importance, benefits and contributions of LT.
- Loss of **LT skills** and human capital trained in Europe.
- The **complexity** of copyright/GDPR/Open Data EU directives makes the access to resources too costly, unclear and risky → Data Governance Act (Nov. 2021) **deep**

Timeline

Sep. 2021	M9	T1.2 D1.2 Report on the SoTA in LT and Language-centric AI
Nov. 2021	M11	T1.3 D1.4-D1.35 drafts ready for WP3
		T2.1 D2.2-2.6 drafts ready for WP3
		T2.2 D2.7-2.12 drafts ready for WP3
Dec. 2021	M12	T2.3 D2.13-D2.16 drafts ready for WP3
Jan. 2022	M13	T1.1 D1.3 DLE – full specification of the concept
Mar. 2022	M15	D3.2 Strategic agenda including roadmap – initial version
May 2022	M17	D3.3 Report on the final round of feedback collection
June 2022	M18	D3.4 Strategic agenda including roadmap – final version



Thank you!



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

German Rigau (EHU)

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

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