



EUROPEAN LANGUAGE EQUALITY

META  NET
META  FORUM 2023

EuLTDom2023

European Language Technology Domains 2023

Diego Alves (University of Zagreb)
dfvalio@ffzg.hr

27-06-2023 META-FORUM 2023 – Digital Language Equality for a Multilingual Europe
<http://european-language-equality.eu>

Objectives

- **Map the usage of NLP/LT in various domains** and report findings regarding the fields that make regular use of NLP/LT
- **List domains** that infrequently use NLP/LT or do not use it at all
- **Output:** a survey of mainstream usage of NLP in domains and languages and details about underrepresented ones
- This snapshot gives an **insight into how domains and NLP tasks are distributed** over the languages

Methodology

- **Database:**

- ACL Anthology Corpus - Full Text
 - <https://github.com/shauryr/ACL-anthology-corpus>
 - Full-text and metadata of the ACL anthology
 - Total number of articles/posters (2010 – 2022): 49,466

- **Languages:**

- ELE languages with report update – <https://european-language-equality.eu/languages/>
- 39 languages

- **List of domains:**

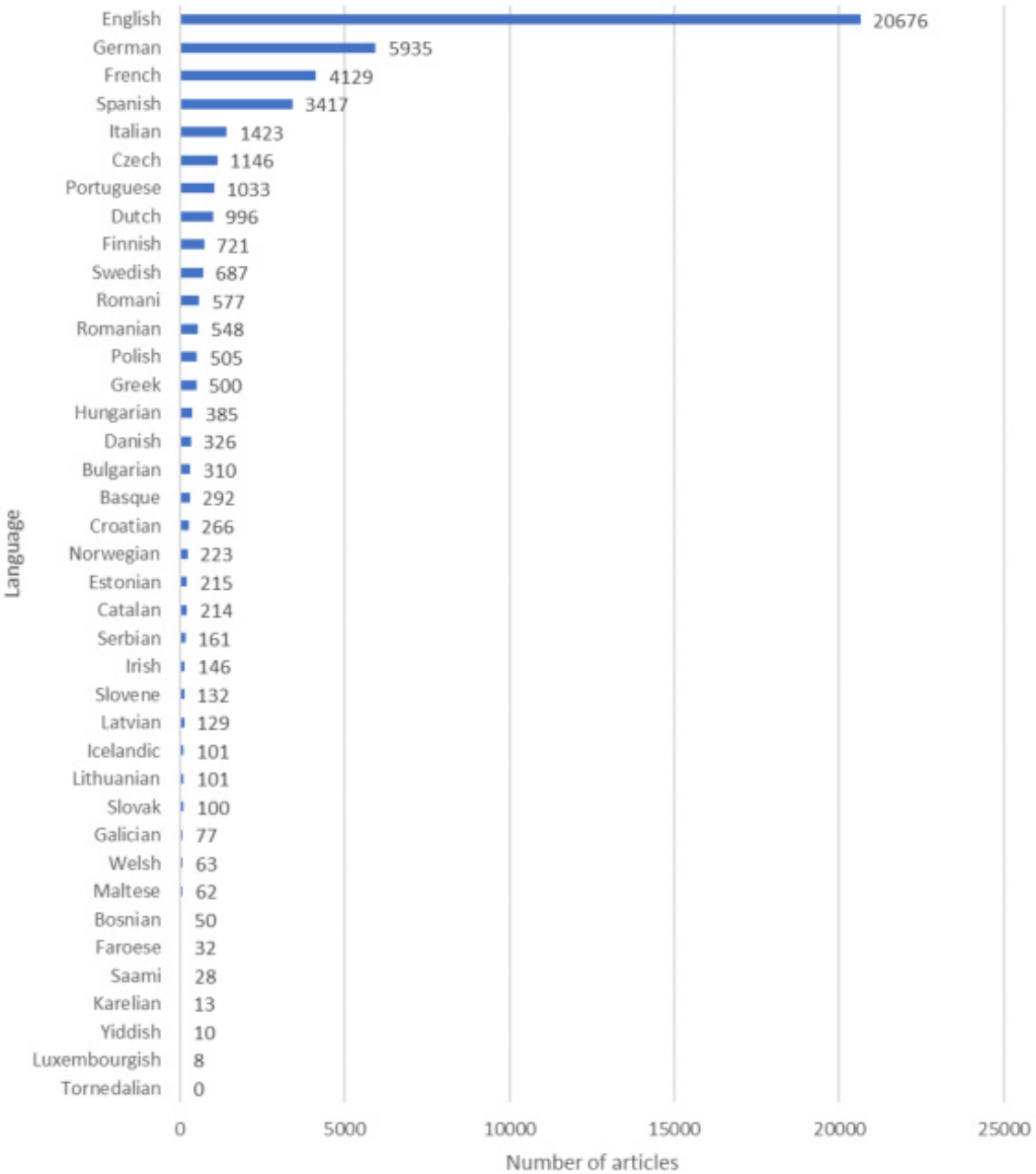
- FORD (Fields of Research and Development classification) + ELE
- Health and Media excluded
- 46 domains

- **List of NLP tasks:**

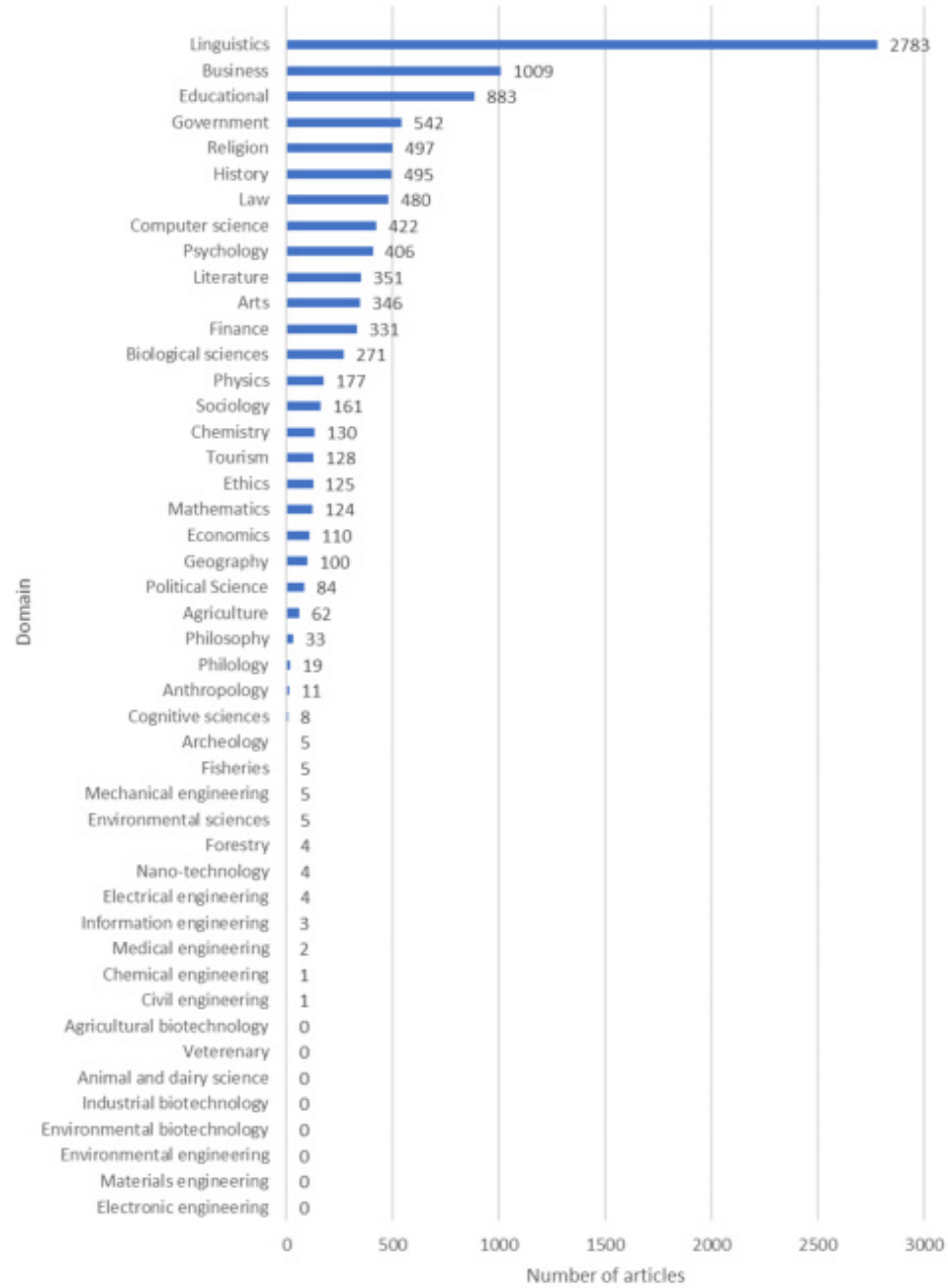
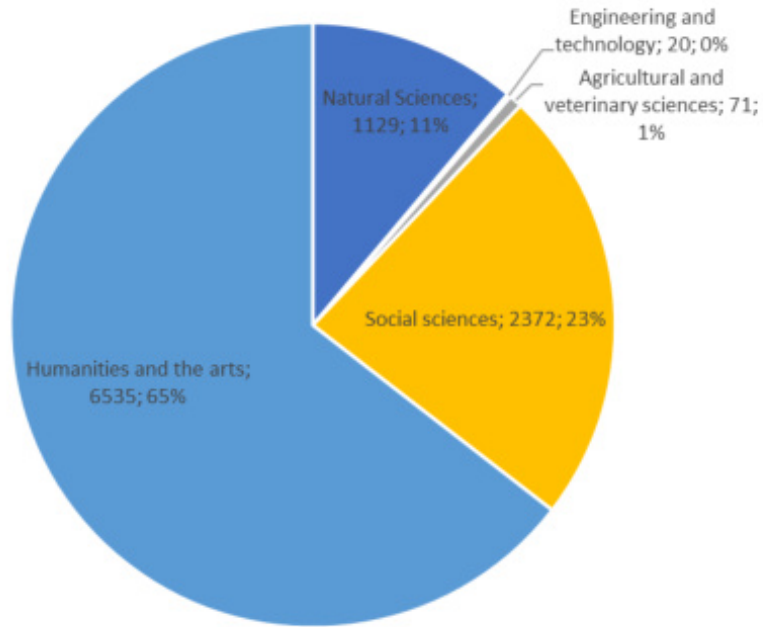
- The Oxford Handbook of Computational Linguistics + Wikipedia (Natural Language Processing) + IBM
- 51 NLP tasks

- **Dictionary-based approach**

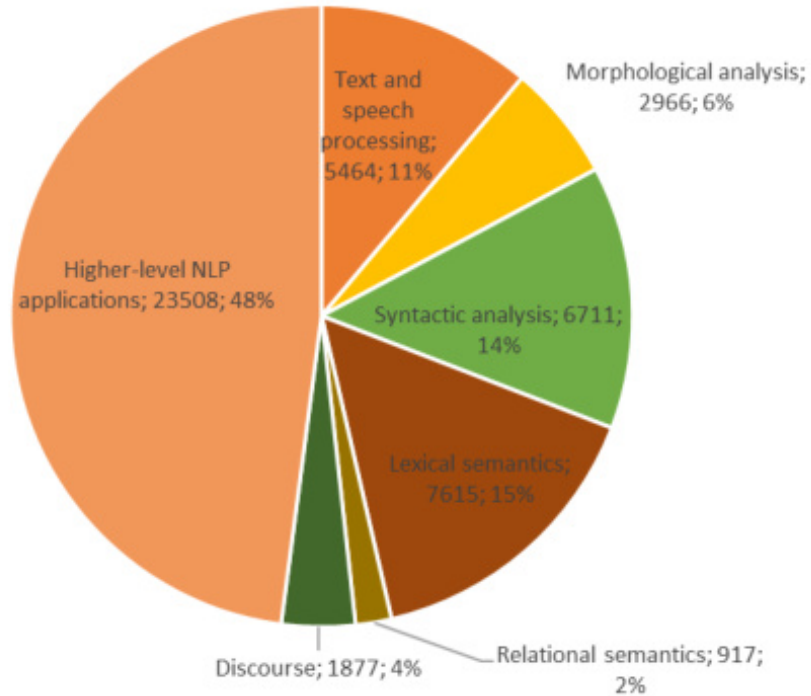
Languages Overview



Domains Overview



NLP Tasks Overview



Conclusions and Perspectives

- The EuLTDom report presents:
 - An overall analysis that allows the identification of the infrequent languages, domains, and NLP tasks in the research work done by the LT community.
 - Language-specific analysis of how different domains use LT which can be used for the identification of research opportunities (i.e.: domains and NLP tasks well-developed for some languages but deficient in others).
- A complementary study can be conducted using research databases such as Web of Science and Scopus.



European Language Equality



Thank you!



The European Language Equality project has received funding from the European Union under grant agreements № LC-01641480 – 101018166 (ELE) and № LC-01884166 – 101075356 (ELE2).

Parts of this research have been funded within the European Union's Horizon 2020 MSC ITN CLEOPATRA, grant agreement No 812997.

Diego Alves (University of Zagreb)
dfvalio@ffzg.hr

27-06-2023 META-FORUM 2023 – Digital Language Equality for a Multilingual Europe
<http://european-language-equality.eu>