Chapter 45
Strategic Research, Innovation and Implementation Agenda for Digital Language Equality in Europe by 2030

Georg Rehm and Andy Way

Abstract This chapter presents the ELE Programme (ELE Consortium 2022). Reacting to the landmark resolution (European Parliament 2018), its vision is to achieve digital language equality in Europe by 2030. The programme was prepared jointly with many stakeholders from the European Language Technology, Natural Language Processing, Computational Linguistics and language-centric AI communities, as well as with representatives of relevant initiatives and associations, and language communities. Europe still suffers from strong inequalities in terms of technology support of its languages. English is still by far the language with the best technological support, followed by a cluster of three languages (German, Spanish, French) that already have only half the technological support of English. More than half of the around 90 languages surveyed have either weak or no technological support at all. The ELE Programme is foreseen to be a shared, long-term funding programme tailored to Europe's needs, demands and values. For the EU we foresee the role of providing resources for coordinating the programme, for providing shared infrastructures, for maintaining the scientific goals and programme principles, etc. The participating countries have the role of providing resources for the development of technologies and datasets for their own languages. Key goals are to reduce the technology gap between English and all other European languages and to address the lack of available language data. The ELE Programme tackles the following overarching themes: *Language Modelling*, *Data and Knowledge*, *Machine Translation*, *Text Understanding* and *Speech*. These interconnected themes focus upon the socio-political goal of establishing DLE in Europe and on the scientific goal of Deep Natural Language Understanding, both by 2030.[1]

Georg Rehm
Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Germany, georg.rehm@dfki.de

Andy Way
Dublin City University, ADAPT Centre, Ireland, andy.way@adaptcentre.ie

on behalf of the whole European Language Equality consortium and all contributors.
European Language Equality EU Project, coordinator@european-language-equality.eu

[1] This chapter is a revised version of the ELE *Strategic, Research, Innovation and Implementation Agenda for Digital Language Equality* (ELE Consortium 2022), which is also available online: https://european-language-equality.eu/agenda/.

G. Rehm, A. Way (eds.), *European Language Equality*, Cognitive Technologies,
https://doi.org/10.1007/978-3-031-28819-7_45

1 Executive Summary

The overall vision of the ELE Programme is to achieve complete digital language equality (DLE) in Europe by 2030. The programme was prepared jointly with many relevant stakeholders from the European Language Technology (LT), Natural Language Processing (NLP), Computational Linguistics and language-centric Artificial Intelligence (AI) communities, as well as with representatives of relevant initiatives and associations, and language communities. The ELE Programme responds to the call "to establish a large-scale, long-term coordinated funding programme for research, development and innovation in the field of language technologies, at European, national and regional levels, tailored specifically to Europe's needs and demands", as specified by the European Parliament Resolution *Language equality in the digital age* (European Parliament 2018). The results of the ELE project show that English is still by far the language with the best and most thorough technological support, followed by a cluster of three languages (German, Spanish, French) that have only half the technological support of English. After yet another gap, the long tail of languages with fragmentary support starts with Finnish, Italian and Portuguese. More than half of the around 90 languages surveyed have either weak or no technological support at all. In comparison to previous results from 2012 (Rehm and Uszkoreit 2012), the gap between English and the other languages appears to be getting *bigger* instead of smaller. With the exceptions of English, German, French and Spanish, all languages we investigated exist in socio-political and economic ecosystems that do *not* incentivise, encourage or foster the development of technologies for these languages. While all 30 European countries we surveyed have put in place national AI strategies, almost all of these national strategies seem to have either ignored or left out the topic of languages and language-centric AI.[2]

The ELE Programme is foreseen to be a shared, long-term, coordinated and collaborative LT funding programme tailored to Europe's needs, demands and values, including multilingualism and language equality in general. For the EU we foresee the role of providing resources for coordinating the programme, for providing shared infrastructures, for maintaining the scientific goals and programme principles, etc. The participating countries have the role of providing resources for the development of technologies and datasets for their own languages. Key goals are to reduce the technology gap between English and all other European languages and to address the lack of available language data: this is true for all European languages except English. The ELE Programme focuses upon *openness*: open source, open access and open standards as well as interoperability and standardisation. A key emphasis is on the creation of large open access language models for all European languages, including

[2] Despite our original findings, in the interim, Spain has funded the 1.1B€ PERTE New Economy of Language programme to "maximize the value of Spanish and co-official languages in the new digital economy and artificial intelligence", see https://planderecuperacion.gob.es/como-acceder-a-los-fondos/pertes/perte-nueva-economia-de-la-lengua. Accordingly, rather than be seen as a laggard in this space, Spain now represents what could and should be done to support European languages and associated technology, and the PERTE programme stands as a template for other nations to adapt to their particular situation.

the creation of datasets and multilingual models, symbolic knowledge, models that include discourse capabilities as well as grounding and other sophisticated features currently out of reach for existing state-of-the-art technologies. The ELE Programme is expected to have a runtime of nine years. In addition to overall coordination, the ELE Programme tackles the following overarching themes: *Language Modelling*, *Data and Knowledge*, *Machine Translation*, *Text Understanding* and *Speech*. These interconnected themes focus upon the socio-political goal of achieving DLE in Europe and on the scientific goal of Deep Natural Language Understanding, both by 2030. The ELE Programme strengthens and makes optimal use of infrastructures, data spaces and services provided by other European initiatives.

The global NLP market is estimated to reach 341.7B$ by 2030. In contrast, the modest investment needed to implement the ELE Programme will not only bring about DLE in Europe but it will also move European research and industry in this field into a dominant position for years to come.

2 Multilingual Europe and Digital Language Equality

Languages are the most common and versatile way for humans to convey and access information. We use language, our most natural means of communication, to encode, store, transmit, share and manipulate information. We use language in everyday life to interact with others and our environment and as social glue, to express and to explain ourselves, to convince, agree with and rebut others. Our laws and constitutions are written in language. We use it in science, commerce, teaching and passing on knowledge to the next generations, for pleasure, creativity and aesthetic enjoyment in puns, jokes and literature. History and culture are recorded, interpreted and enjoyed through language. Our languages are a core part of our identities.

Human languages are incredibly complex: a single word (phrase, sentence, text) can have many meanings, a single meaning can be expressed by many different words (but meaning depends on linguistic and situational context), we can use language literally and metaphorically, language and knowledge are highly intertwined, we do not articulate important parts of a message if these parts are presumed shared knowledge by the community (this includes situational knowledge), important parts of meaning reside in what can be inferred from what has been said. At the same time, language changes. New words are invented, some old ones are dropped, even the structure (syntax and morphology) of languages and the meaning of words change over time. These aspects make human languages fundamentally different from the formal languages of mathematics, logic and computer science. This is also what makes human languages so efficient, elegant, flexible and enjoyable. Finally, there are many human languages (6,000+), not even counting regional and dialectical variants. All these aspects are at the core of human languages and they make it hard for computers to "fully understand" human language and to "properly" process human language in the context of "full and deep understanding".

Languages are at the heart of every aspect of life and their role is crucial to the future of European countries, citizens, businesses, and of the European Union as a whole. Full Digital Language Equality (see Chapter 3) in Europe can deliver an impact in the following four high-priority areas.

Digital Language Equality will have a positive and unprecedented impact on all European languages. We must ensure that no European languages remain under-resourced (see Chapter 4 for an overview and Chapters 5 to 37 for in-depth analyses), but that they are equipped with the same high level of technological support already enjoyed by very few of them (Chapter 2). This, in itself, will deliver a major impact on all European citizens and businesses: supporting all languages in the interest of equality and fairness empowers and brings advantages to their speakers, while reflecting the democratic and inclusive spirit of the EU.

Digital Language Equality will make a contribution to establishing a fair, inclusive and sustainable multilingual Digital Single Market: this will be achieved by helping to future-proof all European languages through digital technologies, and especially preventing the threat of digital extinction for those that suffer from weak support. By fostering a more inclusive and cooperative business and social environment, companies and citizens will benefit from sharing knowledge, digital services and products on an equal footing, overcoming the fragmentation that is caused by many European languages lagging behind, which severely penalises their speakers as well as regional and local communities. Action in this vital area is particularly urgent due to the increasing range of economic, educational and social opportunities that are afforded online and delivered remotely, from e-commerce to online shopping, to web-based recruitment services, online teaching programmes and professional training courses, among others.

Digital Language Equality will help research in Europe, mobilising and leveraging their full potential to start reclaiming scientific and industrial leadership from US-based and Asian competitors, particularly large tech enterprises as well as academic institutions and research centres, that pose fierce competition in several fields. The ELE Programme will instigate regional, national and EU-wide collaboration among scientists from academia and industry covering a broad range of disciplines, ensuring the mix of competencies that is required to deliver substantial and lasting impact at the forefront of scientific and technological progress.

Digital Language Equality will act as a multiplier of opportunities. It will help to aggregate the players that are required to unlock the full potential of an EU-wide effort to exchange and share widely-agreed methodologies, resources and technologies with a focus on promoting the digital equality of European languages: this will benefit the use and promotion of all European languages, encouraging in particular those that have traditionally lagged behind.

3 What is Language Technology and How Can it Help?

Language Technology (LT) is concerned with studying and developing systems capable of processing human languages. Over the years, the field has developed different methods to make explicit the information contained in written and spoken language – and increasingly for other modalities such as sign language, for example – or to generate or synthesise written or spoken language (see Chapter 2 for more details). Despite the inherent difficulty of many of the tasks performed, current LT support allows many advanced applications which would have been unthinkable only a few years ago. LT is present in our daily lives, for example, through search engines, recommendation systems, virtual assistants, chatbots, authoring assistants, text predictors, automatic translation systems, automatic subtitling, automatic summarisation tools, etc. Its rapid development in recent years predicts even more encouraging and also exciting results in the near future. LT is providing solutions for the following main application areas: Machine Translation, Speech Processing, Text Analysis, Information Extraction and Information Retrieval, Natural Language Generation, Human-Computer Interaction (see Chapter 2 as well as Chapters 40 to 43 for in-depth analyses of the state-of-the-art).

4 A Shared European Programme for Language Technology and Digital Language Equality in Europe by 2030

Fully in line with the recommendations of the European Parliament resolution *Language equality in the digital age* (European Parliament 2018), our recommendations, as analysed in the chapters of the present book can be summarised as follows.

The vision described in this book is fully compatible with current EU policy, needs and demands; in fact, they are mission-critical. Missing investment in the underdeveloped areas of LT and language-centric AI will result in the digital extinction of languages, i. e., only global languages spoken by large numbers of speakers, including, crucially, outside the EU, will prevail and the global LT/NLP market will continue to be dominated by the US and China, while the European LT community will be pushed aside even further.

The main concept of the ELE Programme is a collaboration between the EU, and in particular the European Commission, and all participating countries and regions since funding and further investment are needed on all levels. Funding on the level of the EU should enable overarching coordination and EU-wide technological infrastructure. It should cover the topics which require pan-European coordination such as shared tasks, protocols, multilingual dataset creation based on the same principles in line with European values and priorities, etc. Coordination on the European level is needed because language communities are still too fragmented and mostly too small. Further effort should be invested into adequate policy-making, distributed research infrastructures and technological platforms like ELG (Rehm 2023) and the

Common European Language Data Space, with flexible access to sufficient High Performance Computing (HPC) facilities. Additionally, national and regional funding should complement the European funding with regard to language-specific research and development. The main gaps to be filled in these respects and the most important anticipated developments are described, among others, in the language reports (see Chapters 5 to 37).

This section summarises our main recommendations for this shared programme (more detailed recommendations are contained in the previous chapters of this book). First, we outline the possible cornerstones for suitable policy and infrastructure recommendations, as well as ideas for the realisation of a governance model. Second, we revise the technology and data recommendations suggested by the ELE consortium (derived from Chapters 39 to 44), which are closely related to those discussed in the *Language equality in the digital age* resolution (European Parliament 2018).

Further, in terms of our research recommendations, the ELE consortium together with the wider LT community has developed a clear vision for the different areas of LT. We see an urgent need to refocus and massively strengthen European LT/NLP research through a large-scale initiative as a shared, collaborative pan-European effort between the EU and those countries and regions that participate in the initiative, i. e., the *ELE Programme*. This endeavour should include the participation of research centres, academia, companies (particularly SMEs and startups), and other relevant stakeholders. As LT is aggregated and applied to more complex settings, interdisciplinary research and activities are becoming more relevant in order to further boost developments and allow synergies to become apparent. To achieve *Deep Natural Language Understanding*, we need to finance and investigate fields such as cognitive, neural and symbolic AI further.

The ELE Programme should boost pan-European long-term basic research as well as knowledge and technology transfer between research labs and industry. Frequently mentioned areas and tasks for basic and applied research where further investigation is needed include, among others, systematic language data collection (text, dialogue, vision, sign language and other forms of interactions), speech analysis, AI, human-computer interaction, machine learning, robotics, natural language understanding and processing tasks such as machine reading, text analysis, machine translation, chatbots, virtual assistants and summarisation.

4.1 Policy Recommendations

- Reinforce European leadership in LT by establishing the ELE Programme as a large-scale, long-term coordinated funding programme for research, development, innovation and education with the *societal goal* of achieving Digital Language Equality in Europe and the *scientific goal* of Deep Natural Language Understanding, both by 2030.
- Ensure comprehensive EU-level legal protection for the more than 60 regional and minority languages spoken in Europe.

- Empower recognition of the collective rights of national and linguistic minorities in the digital world (including sign languages).
- Encourage mother-tongue teaching for speakers of official and non-official languages of the EU.
- Safeguard sufficient funding to support new technological approaches, based on increased computational power and better access to sizeable amounts of data.
- Develop specific initiatives within current funding schemes, especially Horizon Europe and Digital Europe (including the Recovery Plan for Europe), to boost long-term basic research as well as knowledge and technology transfer between countries and regions, and between academia and industry.
- Support the coordination between research and industry to enhance the digital possibilities for LT and Open Access to language data.
- Define and develop a minimum set of language resources and capacities that all European languages should possess (see Krauwer 2003) .
- Develop common policy actions and protocols for language data sharing by public administrations at all levels. Language data should be included as a high-value data category in the Open Data Directive (2019/1024/EU).
- Enable and empower European SMEs and startups to easily access and use LT in order to grow their businesses independent of language barriers, also thanks to e-commerce and online marketplaces.
- Create the necessary appealing conditions to attract and retain qualified and diverse international LT personnel in Europe.
- Encourage all EU-funded projects to have a language diversity plan and to include direct or associated partners from a less-widely spoken language.
- Empower and encourage administrations at all levels to improve access to online services and information in different languages.
- Create a European network of centres of excellence in LT to increase industry visibility and to design national research agendas.
- Implement and maintain long-term an overall EU-wide policy framework to achieve European LT sovereignty.
- Facilitate EU Member States' acquisition of LT for their local industries without depending on non-European technology providers.

4.2 Governance Model

- Structure the ELE Programme as a shared, collaborative and coordinated programme between the EU and all countries and regions that participate.
- Allocate the area of multilingualism, linguistic diversity and language technology to the portfolio of a EU Commissioner.
- Set up a large lobby for EU regional and minority languages.
- Create a pan-European network of research centres to facilitate the coordination and also implementation of the ELE Programme at all levels.

- Promote a distributed centre for linguistic diversity that will strengthen awareness of the importance of lesser-used, regional and minority languages.
- Design and apply new forms of research funding and organisation to ease the transition from application-oriented basic research to commercially-focused technology development.

4.3 Technology and Data Recommendations

- Develop large open-source language models that work for all European languages, optimised in terms of compute time and cost.
- Address the lack of available data and define the minimum amount of language resources and capabilities that all European languages should possess.
- Add more focus on systematic and comprehensive language data collection (text, dialogue, multimodal) and exploit automatic data generation (synthetic data), crowd-sourcing and translation of high-quality data.
- Develop new methodologies for transfer and adaptation of resources and technologies to other domains and languages.
- Develop high-performance applications (in terms of speed and quality) for all languages that respect safety, security and privacy.
- Ensure efficient adaptations to applications, both in terms of language, domain, efficiency, power consumption, ease of maintenance, and quality assurance.
- Develop methods to overcome the unequal data availability, by focusing on, e. g., annotation transfer, multilingual models preserving quality, few-shot or zero-shot learning.
- Unleash the power of monolingual and multilingual public sector data, data from broadcasters, social media, publishers, etc.
- Enforce open ecosystems, open standards and interoperability (including Open Source and Open Access).
- Focus on research on bias for strengthening inclusiveness and accessibility, to respect and promote European values and principles.
- Focus upon Green LT with a small compute and carbon footprint (e. g., model compression).
- Foster publicly available resources that facilitate innovation and research for both commercial and non-commercial actors.
- Construct a multilingual LT benchmark, a European "SuperGLUE"-style (Wang et al. 2019) shared benchmark, that tracks progress.
- Define the minimum language resources that all European languages should possess in order to prevent digital extinction.

4.4 Infrastructure Recommendations

- Strengthen existing and create new research infrastructures and LT platforms that support research and development activities, including collaboration, knowledge sharing, and Open Access to data, tools and technologies.
- Fill the identified gaps in data, language resources and knowledge graphs and create a future path for Europe towards comprehensive and interlinked data infrastructures.
- Develop clear and robust protocols to ensure flexible access to sufficient GPU-based HPC infrastructure and robust protocols to process sensitive data.
- Ensure sufficient operational capacity, especially for Large Language Models (LLMs) and flexible access to GPU-based HPC facilities.
- Follow the idea of a Semantic Data Fabric including rich semantics for the development of an integrated and interoperable data infrastructure.

4.5 Research Recommendations

4.5.1 Recommendations for all Research Areas

- Gather and make available the critical mass of resources in terms of data, HPC facilities, and expertise from pan-European LT research labs and centres, with support from the EC as well as national and regional administrations.
- Create sufficient multilingual and multimodal data of quality (responsible, legal, diverse, unbiased, ethical, representative, etc.), in all European languages and domains (media, health, legal, education, etc.).
- Provide flexible access to HPC facilities for LT research and industry. HPC facilities should provide clear and robust protocols to process sensitive data.
- Develop better benchmarks and datasets (ethical, responsible, legal, etc.) for all languages, domains, tasks and modalities.
- Combine interactive LT (conversational AI) with text, knowledge, and multimedia technologies for a new generation of applications that can address the deeper questions of communication, common sense and reasoning.
- Encourage trustworthy, unbiased, inclusive, non-discriminatory LT/AI, making interpretability and explainability of AI models a priority.
- Develop further the areas of responsible AI by combining statistical and symbolic AI in multilingual environments to provide AI-based applications that deliver accurate results and benefits for research, industry, and society.
- Focus on methods and learning architectures to overcome the highly unequal data availability, such as annotation transfer, synthetic data and their proper use in machine learning, multilingual models preserving quality and coverage and few-shot or zero-shot learning.

- Focus on Green LT and investigate new efficient methods to extend, reuse and adapt existing pre-trained language models or develop new ones with much reduced carbon footprint.
- Develop language- and culture-specific technologies that cover more linguistic phenomena and text types, focusing on accessibility, through sign language, avatar technology, etc.
- Provide transparency of AI models with regard to accuracy and fairness.
- Reframe LT/NLP as a quantum computing problem.

4.5.2 Machine Translation

- Develop near-real-time MT across all modalities (speech, text, signs, etc.) and adaptive MT, where the system learns from interaction with users.
- Move towards context-aware methodologies that go beyond text data and include images, videos, tables, etc. by developing multimodal MT systems.
- Develop low-resource MT by deepening research on projection and structural organisation of embeddings to comprehend how structurally different languages and their respective embedding spaces can be mapped to one another.

4.5.3 Speech Processing

- Enhance speech resources and create acoustic models to cover all European languages, including non-standard varieties and dialects.
- Improve the handling of audio conditions currently perceived as difficult (e. g., multiple simultaneous speakers in noisy environments speaking spontaneously and highly emotionally in a mix of languages).
- Develop high-quality, natural synthetic voices, allowing users to obtain content in the language of their choice.
- Improve context modelling to handle the translation of speech models across larger volumes of text.
- Support research in the direction of combining speech, NLU and NLP with other modalities, such as image and vision.
- Address privacy and security threats in areas of speech synthesis, voice cloning and speaker recognition.

4.5.4 Text Analytics and Natural Language Understanding

- Create large Open-Access language models for all European languages (for fine-tuning and downstream tasks), datasets (for training and testing), multilingual models, models that include symbolic knowledge and discourse features.
- Increase the adoption of approaches based on self-supervised, zero-shot, and few-shot learning.

- Support research in NLU which integrates speech, NLP, and contextual information as well as additional modes of perception.
- Strengthen basic research in neurosymbolic approaches to NLP/NLU, including grounding and the use of human-understandable databases and sources.
- Strengthen progress in reinforcement-based learning, novel dialogue management strategies, and situation-aware natural language generation.
- Strengthen interdisciplinary research and enable better modelling of multimodal environments.

4.6 Implementation Recommendations

- Structure the ELE Programme into three phases of similar duration.
- Facilitate discussions between the EU, the European Commission in particular, and all participating countries to define the goals and the financial setup.
- Encourage participating countries to invest into the development of large language models, data sets, technologies, and tools for their own languages.
- Encourage the EU to establish legislation to promote participation.
- Encourage the EU to invest in the pan-European coordination of all language-specific projects and initiatives, support mechanisms, infrastructures, data procedures, cross-cutting projects, etc. and provide flex funds for bootstrapping poorly supported languages.
- Structure the ELE Programme into six themes covering: Language Modelling, Data and Knowledge, Machine Translation, Text Understanding, Speech, and Infrastructure and support each theme by coordination actions (CSAs), research actions (RIAs) as well as actions for innovation and deployment (IAs).

5 Roadmap towards Digital Language Equality in Europe

5.1 Main Components

Language Technologies have the potential to overcome the linguistic divide in the digital age. However, we need to define actions, tools, processes and actors that need to be involved. The ELE SRIA includes a roadmap with concrete steps for the implementation that carry tangible and measurable outputs.

The main scientific goal of the ELE Programme is *Deep Natural Language Understanding in Europe by 2030*. Efficiency will be increased by sharing knowledge, infrastructures and resources, with a view to developing innovative technologies and services, in order to achieve the next scientific breakthrough in this area and help reduce the technology gap between Europe's languages with the collaboration of research centres, academic experts, industry and other relevant stakeholders. Crucially,

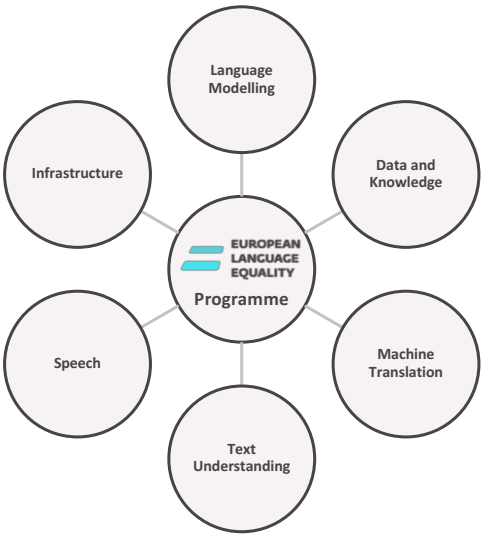

Fig. 1 The six main themes of the ELE Programme

the long-term ELE Programme will involve significantly intensified coordination between the participating countries and languages.

The main societal and economic goal of the ELE Programme is *Digital Language Equality in Europe in 2030*. The focus is on language equality and the provisioning of technologies, services and resources outside the often-preferred languages to achieve and maintain long-term technological sovereignty in this crucial application area. For regional, minority and lesser spoken languages, we need to find a (technological) way to consider Deep Natural Language Understanding within a common approach, to create synergies and increase efficiency of the solutions and their design and development. To narrow the digital divide, there is a pressing urgency for novel techniques that would bring less-resourced languages to a level comparable to state-of-the-art results for resource-rich languages. This includes the leveraging of multimodal and multilingual resources to support the development of applications for languages and varieties with scarce resources.

This roadmap towards Digital Language Equality in Europe by 2030 provides a path and the means to ensure that the two goals outlined above are met. To tackle this challenge, the ELE Programme combines the following six themes (see Figure 1).

Language Modelling This theme includes research, development and deployment activities regarding LLMs, especially multilingual and multimodal, generative LLMs that include text, speech, image, video, etc. Time and resources need to be invested for experiments, developing novel approaches, shared tasks, etc. For novel research approaches we need to combine national projects and data sets with international consortia. With regard to innovation and deployment, LLMs will be applied in industrial sectors and use-cases.

Data and Knowledge The Data and Knowledge theme is focused on the collection, production, annotation, curation, quality assessment, standardisation, etc. of text data, spoken data, video data, and other multimodal data, primarily with regard to their application as data for pre-training different sorts of LLMs.

Machine Translation The MT theme is focused on improving the automated translation from one natural language into another (including sign languages and other modalities). While Europe has a strong foundation in this field, research needs to combine novel, groundbreaking approaches with results of the Data and Knowledge as well as Language Modelling themes (see above). The results need to be applied in different industrial sectors and use-cases. Deployment needs to be fast, agile and driven by excellent teams.

Text Understanding The Text Understanding theme aims to improve the identification and labelling of information regarding all levels of linguistic analysis underlying any natural language text (or other modalities). This requires exploring new strands of research and building on synergies with the other themes. An equally important aspect is applicability in industry.

Speech The Speech theme addresses one big challenge of the European LT community, i. e., the shift from text-to-speech and multimodal processing (including research towards grounding). While progress in the area of speech applications has been made in the last decade, we also need novel research paradigms. This theme will benefit from the themes Data and Knowledge as well as Language Modelling. The development of relevant industry applications is another goal.

Infrastructure The Infrastructure theme involves the extension, maintenance and interoperability of platforms such as European Language Grid (ELG) and Language Data Space (LDS). ELG has the potential of functioning as one of the primary platforms to support the activities of the ELE Programme. Moreover, ELG can be further developed into the focal point for best practices and the development of bridges to other relevant platforms. New features and functionalities need to be implemented for a higher adaptability. Other important factors are the provisioning of GPUs and standardisation.

5.2 Actions, Budget, Timeline, Collaborations

The *Language equality in the digital age* resolution (European Parliament 2018) strongly encourages the "establish[ment of] a large-scale, long-term coordinated funding programme for research, development and innovation in the field of language technologies, […] tailored specifically to Europe's needs and demands".

As a direct response, the ELE project (Rehm et al. 2022a) has developed the DLE Metric (Gaspari et al. 2022a; Grützner-Zahn and Rehm 2022) as a measure to assess and track the advancement towards DLE in Europe empirically (Chapter 3) and, in parallel, an outline of necessary actions. These have been informed by 66 project

reports[3] that comprise more than 2400 pages with condensed findings, summarised in the form of the present book. A total of 92 languages have been taken into account. We have included voices from research, industry and civil society. In terms of research on Europe's languages, we prepared over 30 reports on the situation of individual languages (Chapters 5 to 37, Chapter 4 contains an overview analysis). In addition, we collected input through various surveys and more than 60 expert interviews (Chapters 4, 38 and 39). To cover the industry angle, our industry partners produced four technical deep dives and collected feedback in a number of surveys for further information (Chapters 40 to 43). Civil society was represented by the European citizen survey with about 20,000 responses (Chapters 4, 38 and 39).

The ELE Programme has a foreseen runtime of nine years, divided into three phases of three years each. Implementing the ELE Programme will significantly improve the state-of-the-art of LT and NLP and language-centric AI research (Chapter 2), create DLE in Europe and put Europe back into the global pole position of research and industrial applications of this type of technology (Chapter 44).

5.2.1 Actions

We foresee different types of projects, implemented using the different EC project types: coordination actions (CSAs), research actions (RIAs) as well as actions for innovation and deployment (IAs), see Table 1.

Coordination and Support Actions (CSAs) are needed to support research activities and policies (networking, exchange, access to research infrastructures, conferences, etc.). The ELE Programme envisages three CSAs for the overall programme coordination. These include, among others, the maintenance of the ELE principles, quality assurance approaches, shared tasks, etc. Additional CSAs are needed for the themes *Data and Knowledge* as well as *Language Modelling* as these are fundamental for all other themes as well. Another CSA is needed for supporting and further developing shared infrastructures.

[3] See Gaspari et al. (2021), Agerri et al. (2021), Gaspari et al. (2022b), Sarasola et al. (2022), Koeva and Stefanova (2022), Melero et al. (2022a), Tadić (2022), Hlavacova (2022), Pedersen et al. (2022), Steurs et al. (2022), Maynard et al. (2022), Muischnek (2022), Lindén and Dyster (2022), Adda et al. (2022), Sánchez and García-Mateo (2022), Hegele et al. (2022a), Gavriilidou et al. (2022), Jelencsik-Mátyus et al. (2022), Rögnvaldsson (2022), Lynn (2022), Magnini et al. (2022), Skadiņa et al. (2022), Gaidienė and Tamulionienė (2022), Anastasiou (2022), Rosner and Borg (2022), Eide et al. (2022), Ogrodniczuk et al. (2022), Branco et al. (2022), Păiș and Tufiș (2022), Garabík (2022), Krek (2022), Melero et al. (2022b), Borin et al. (2022), Prys et al. (2022), Krstev and Stanković (2022), Ćušić (2022), Giagkou et al. (2022), Moshagen et al. (2022), Robinson-Jones and Scarse (2022), Hajič et al. (2021), Thönnissen (2022), Eskevich and Jong (2022), Rufener and Wacker (2022), Hajič et al. (2022), Hegele et al. (2022b), Gísladóttir (2022), Kirchmeier (2022), Hicks (2022), Blake (2022), Hrasnica (2022), Heuschkel (2022), Bērziņš et al. (2022), Backfried et al. (2022), Gomez-Perez et al. (2022), Kaltenböck et al. (2022), Way et al. (2022b), Way et al. (2022a), Aldabe et al. (2022b), Aldabe et al. (2022a), ELE Consortium (2022), Hegele et al. (2021a), Hegele et al. (2021b), Rehm et al. (2022b), Marheinecke et al. (2022) and Rehm et al. (2022c).

Research and Innovation Actions (RIA) are collaborative projects funding research activities that allow the exploration of new technologies, new methods, new products, or improvements of existing ones. Research is the fundamental prerequisite for DLE. Over the last decade, the community has developed a clear vision of the work needed in the different areas of LT. To achieve Deep NLU, we need to invest in and further research the areas of language modelling, machine translation, text understanding and speech.

Innovation Actions (IAs) consist of activities directly aiming at producing improved products, processes or services. They may include prototyping, testing, demonstrating, piloting, large-scale product validation and market replication.

	Type	Number
ELE Programme – overall coordination	CSA	3
Theme Data and Knowledge – coordination	CSA	3
Theme Language Modelling – coordination	CSA	3
Theme Language Modelling – research	RIA	15
Theme Language Modelling – innovation and deployment	IA	15
Theme Machine Translation – research	RIA	12
Theme Machine Translation – innovation and deployment	IA	12
Theme Text Understanding – research	RIA	12
Theme Text Understanding – innovation and deployment	IA	12
Theme Speech – research	RIA	12
Theme Speech – innovation and deployment	IA	12
Theme Infrastructure – support	CSA	3

Table 1 Different types and number of projects foreseen in the ELE Programme

5.2.2 Budget

As a shared programme between the EU and the participating countries, the final financial setup needs to be discussed between all involved parties. For the EU part of the budget, we suggest the breakdown shown in Table 2. In addition to these investments, which relate to the overarching coordination, research and innovation projects, the participating countries and regions are expected to invest in their languages themselves, while the languages with fragmentary, weak or no technical support can request funding from the European Union (*flexible funds*, see below).

In addition to the sum of 690M€ for the actions implementing the theme-related projects of the ELE Programme, we envisage investing an additional 150M€ as *flexible funds* for languages with fragmentary, weak or no technical support since we anticipate that a number of participating countries will require complementary fund-

ELE Programme (overall coordination)	60M€
Theme Data and Knowledge	45M€
Theme Language Modelling	195M€
Theme Machine Translation	120M€
Theme Text Understanding	120M€
Theme Speech	120M€
Theme Infrastructure	30M€
Sum	**690M€**
Flexible funds	150M€
Total	**840M€**

Table 2 Budget breakdown of the ELE Programme (EU contribution only; numbers are indicative)

ing from the EU. A more detailed breakdown of the different themes with their associated project types and runtime is shown in Table 4.

The complementary national/regional investments required on the individual language level are difficult to predict. We group the languages into three clusters (see Table 3) and provide indicative investments, which relate to the whole duration of the ELE Programme. Other factors (e. g., number of speakers, etc.) can be taken into account to arrive at more precise numbers.

Languages with *weak or no support*	40-50M€ each
Languages with *fragmentary support*	30-40M€ each
Languages with *moderate support*	20-30M€ each

Table 3 Indicative investments required by language, provided by the participating countries

This language-specific funding is foreseen to be provided by the participating countries. However, the EU should help bootstrap the development of technologies for languages that are not doing well digitally, using the suggested flexible funds.

5.2.3 Timeline

The ELE Programme is foreseen to have a runtime of nine years, divided into three phases of three years each (Table 4). The CSA and RIA projects are expected to run for three years each while the IA projects have a runtime of two years so that they can focus on the innovation and deployment aspects.

Phase 1: 2024-2026 Phase 1 lays a strong foundation for the overall ELE Programme. All projects start in Phase 1, except for the Innovation Actions.

Phase 2: 2027-2029 Phase 2 drives forward all projects of all types while continuing the Coordination Actions.

Phase 3: 2030-2032 Phase 3 continues the Coordination Actions and finishes off all projects in 2032.

	Type	Num.	Phase 1 2024	2025	2026	Phase 2 2027	2028	2029	Phase 3 2030	2031	2032	Budget Each	Sum
ELE Programme – overall coordination	CSA	3										20M€	60M€
Theme Data and Knowledge – coordination	CSA	3										15M€	45M€
Theme Language Modelling – coordination	CSA	3										15M€	45M€
Theme Language Modelling – research	RIA	15										5M€	75M€
Theme Language Modelling – innovation and deployment	IA	15										5M€	75M€
Theme Machine Translation – research	RIA	12										5M€	60M€
Theme Machine Translation – innovation and deployment	IA	12										5M€	60M€
Theme Text Understanding – research	RIA	12										5M€	60M€
Theme Text Understanding – innovation and deployment	IA	12										5M€	60M€
Theme Speech – research	RIA	12										5M€	60M€
Theme Speech – innovation and deployment	IA	12										5M€	60M€
Theme Infrastructure – support	CSA	3										10M€	30M€
													690M€
Flexible funds for languages with fragmentary, weak or no technological support.													**150M€**
													840M€

Table 4 Project types, timeline and indicative budget breakdown of the ELE Programme (EU)

5.2.4 Collaborations with Related Initiatives

The ELE Programme complements related initiatives, projects and organisations and it will make use of the services, resources and infrastructures provided by these initiatives. We can group these different stakeholders into several broader categories:

Data spaces and data infrastructures:
Various EU/EC Data Spaces including the Common European Language Data Space (LDS), Media Data Space and others; Big Data Value Association (BDVA) and Data, AI and Robotics (DAIRO);[4] Gaia-X;[5] International Data Spaces Association (IDSA);[6] etc.

Research and research data infrastructures:
European Open Science Cloud (EOSC);[7] German National Research Data Infrastructure (NFDI);[8] CLARIN ERIC;[9] Research Data Alliance (RDA);[10] etc.

Various AI initiatives:
ADRA;[11] CLAIRE;[12] LEAM;[13] HumanE-AI;[14] OpenGPT-X;[15] etc.

AI on Demand Platform:
AI-on-Demand Platform;[16] European Language Grid (ELG);[17] etc.

High performance computing:
EuroHPC Joint Undertaking;[18] etc.

Standardisation:
World Wide Web Consortium (W3C);[19] DIN;[20] etc.

[4] https://www.bdva.eu, https://www.bdva.eu/DAIRO
[5] https://gaia-x.eu
[6] https://internationaldataspaces.org
[7] https://eosc.eu
[8] https://www.nfdi.de
[9] https://www.clarin.eu
[10] https://www.rd-alliance.org
[11] https://adr-association.eu
[12] https://claire-ai.org
[13] https://leam.ai
[14] https://www.humane-ai.eu
[15] https://opengpt-x.de
[16] https://www.ai4europe.eu
[17] https://www.european-language-grid.eu
[18] https://eurohpc-ju.europa.eu
[19] https://www.w3.org
[20] https://www.din.de

6 Concluding Remarks

Large-scale studies such as the META-NET White Paper Series (Rehm and Uszkoreit 2012), the STOA study (STOA 2018) and the ELE language reports (see Chapter 4 for an overview and Chapters 5 to 37 for in-depth analyses) have shown that many languages are in danger of digital extinction because they are not sufficiently supported through Language Technologies. Digital Language Equality 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 (Chapter 3). In alignment with what the Language Technology community has promoted for more than a decade, the European Parliament adopted a resolution on *Language equality in the digital age* that suggested initiating a large-scale European LT research, development and innovation programme and to intensify research and funding to achieve Deep Natural Language Understanding and also Digital Language Equality (European Parliament 2018).

Languages are at the heart of every aspect of life. Understanding language is key for building intelligent systems. Over the coming years, AI is expected to transform every industry and society as a whole. There are trends and megatrends that bear closely on digital technologies. Among others, these include accelerating hyperconnectivity, shifts in the nature of work, increasing digitalisation, new modes of learning, expanding consumerism, novel approaches to politics and governance, changes in healthcare, etc. LT and NLP are, by now, considered important driving forces. Language Technology will play a deciding role in how these unfold.

Language tools and resources have increased and improved since the end of the last century, a process further catalysed by the advent of deep learning and neural networks over the past decade. We find ourselves today in the midst of a significant paradigm shift in LT and language-centric AI. This revolution has brought noteworthy advances to the field along with the promise of substantial breakthroughs in the coming years. However, this transformative technology poses problems from a research advancement, environmental, and ethical perspective. Furthermore, it has also laid bare the acute digital inequality that exists between languages. In fact, many sophisticated NLP systems are unintentionally exacerbating this imbalance due to their reliance on vast quantities of data derived mostly from English-language sources. Other languages lag far behind English in terms of digital presence and even the latter would benefit from greater support. Moreover, the striking asymmetry between official and non-official European languages with respect to available digital resources is very worrisome. The unfortunate truth is that European Language Technology is failing to keep pace with the newfound and rapidly evolving changes in the field.

One need look no further than what is happening today across the diverse topography of state-of-the-art LT and language-centric AI for confirmation of the current linguistic unevenness. The paradox at the heart of recent LT advances is evident in almost every LT discipline. Our ability to reproduce ever better synthetic voices has improved sharply for well-resourced languages, but dependence on large volumes of high-quality recordings effectively undermines attempts to do the same for low-resource languages. Multilingual NMT systems return demonstrably improved

results for low- and zero-resource language pairs, but insufficient model capacity continues to haunt transfer learning because large multilingual datasets are required, forcing researchers to rely on English as the best-resourced language. A similar language discrepancy is also found in several of the domain sectors: medical corpora, models and knowledge bases suffer from this disparity, as do users of under-resourced languages in education, where access to language-related tools is limited for most smaller language communities.

However, this time of transition also represents an opportunity to right the ship. Now is the moment to seek balance between European languages in the digital realm. There are ample reasons for optimism. Although there is more work that can and must be done, Europe's leading language resource repositories, platforms, libraries, models and benchmarks have begun to make inroads in this regard.

Over the last decade, the community has developed a clear vision of the work needed in the different areas of LT. The ELE project has devised an outline of necessary actions in the form of concrete recommendations. The ELE Programme, specified in the form of the SRIA and roadmap presented in this chapter, will serve as the blueprint for achieving DLE in Europe. While the political and societal goal is reaching full *Digital Language Equality across all European languages* (and, at the same, preventing digital extinction of many of our languages in Europe), the scientific goal envisioned to be reached by 2030 is *Deep Natural Language Understanding*.

Deep Natural Language Understanding is still an open research problem far from being solved since all current approaches have severe limitations. The development of new LT systems would not be possible without sufficient resources (data, experts, compute facilities, etc.). Creation of carefully designed evaluation benchmarks and annotated data sets for every language and domain of application is needed to foster technological progress, while encouraging deeper understanding of the mechanisms by which they are achieved. All these efforts will then lead to long-term progress towards multilingual, efficient, accurate, explainable, ethical and unbiased language understanding and communication, to create transparent digital language equality in Europe in all aspects of society, from government to businesses to the citizens.

We foresee an ELE Programme of nine years (2024-2032). This period will be divided into three phases of three years each, combining coordination actions (CSAs), research actions (RIAs) as well as actions for innovation and deployment (IAs). The whole community, meaning all relevant scientific and industrial stakeholders from all Member States and Associated Countries, need to be involved. The ELE Programme will tackle the following central themes: Language Modelling, Data and Knowledge, Machine Translation, Text Understanding, and Speech.

As a shared programme between the EU and the participating countries, we suggest an EU budget of 690M€, plus 150M€ of flexible funds to help bootstrap the development of technologies for languages with fragmentary, weak or no technical support. This will be supplemented by national and regional funding.

The ELE Programme is meant to develop into the focal point in which all coordinated developments come together. In this regard, the European Institutions and national as well as regional governments and language institutes must be involved in creating resources, tools and technologies for their own languages. It is exactly

the large scale of the effort that will accelerate the developments and advance the state-of-the-art that will make it possible to join forces that have so far never been joined. This will make it possible to address all European and other relevant languages, all cultures with their particular background and framing of the world, all relevant domains, and all stakeholders by means of a substantial number of use cases. We are convinced that this initiative, built around a coordinated giant pool of shared data sets, open evaluations, open competitions, shared tasks, standardisation efforts, etc. in the literal sense of Open Science, will have a much-needed, substantial and lasting impact in terms of interoperability, development costs, quality and, thus, uptake of the truly game-changing technologies developed in the ELE Programme. Research in Europe must focus on creating the new paradigm of Language Technology, fully harnessing the power of current and emerging AI methods that are based on vast data sets and knowledge bases. With a concerted effort and significant funding, digital language equality will be achieved, for the benefit of *all* Europeans.

References

Adda, Gilles, Annelies Braffort, Ioana Vasilescu, and François Yvon (2022). *Deliverable D1.14 Report on the French Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-french.pdf.

Agerri, Rodrigo, Eneko Agirre, Itziar Aldabe, Nora Aranberri, Jose Maria Arriola, Aitziber Atutxa, Gorka Azkune, Arantza Casillas, Ainara Estarrona, Aritz Farwell, Iakes Goenaga, Josu Goikoetxea, Koldo Gojenola, Inma Hernaez, Mikel Iruskieta, Gorka Labaka, Oier Lopez de Lacalle, Eva Navas, Maite Oronoz, Arantxa Otegi, Alicia Pérez, Olatz Perez de Viñaspre, German Rigau, Jon Sanchez, Ibon Saratxaga, and Aitor Soroa (2021). *Deliverable D1.2 Report on the State of the Art in Language Technology and Language-centric AI*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/LT-state-of-the-art.pdf.

Aldabe, Itziar, Aritz Farwell, and German Rigau (2022a). *Deliverable D3.3 Report on the final round of feedback collection*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/feedback-collection.pdf.

Aldabe, Itziar, Georg Rehm, German Rigau, and Andy Way (2022b). *Deliverable D3.1 Report on existing strategic documents and projects in LT/AI (second revision)*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/LT-strategic-documents-v3.pdf.

Anastasiou, Dimitra (2022). *Deliverable D1.24 Report on the Luxembourgish Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-luxembourgish.pdf.

Backfried, Gerhard, Marcin Skowron, Eva Navas, Aivars Bērziņš, Joachim Van den Bogaert, Franciska de Jong, Andrea DeMarco, Inma Hernaez, Marek Kováč, Peter Polák, Johan Rohdin, Michael Rosner, Jon Sanchez, Ibon Saratxaga, and Petr Schwarz (2022). *Deliverable D2.14 Technology Deep Dive – Speech Technologies*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/speech-deep-dive.pdf.

Bērziņš, Aivars, Mārcis Pinnis, Inguna Skadiņa, Andrejs Vasiļjevs, Nora Aranberri, Joachim Van den Bogaert, Sally O'Connor, Mercedes García–Martínez, Iakes Goenaga, Jan Hajič, Manuel Herranz, Christian Lieske, Martin Popel, Maja Popović, Sheila Castilho, Federico Gaspari,

Rudolf Rosa, Riccardo Superbo, and Andy Way (2022). *Deliverable D2.13 Technology Deep Dive – Machine Translation*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/MT-deep-dive.pdf.

Blake, Oliver (2022). *Deliverable D2.10 Report from LIBER*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/consultation-LIBER.pdf.

Borin, Lars, Rickard Domeij, Jens Edlund, and Markus Forsberg (2022). *Deliverable D1.33 Report on the Swedish Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166 ELE. https://european-language-equality.eu/reports/language-report-swedish.pdf.

Branco, António, Sara Grilo, and João Silva (2022). *Deliverable D1.28 Report on the Portuguese Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-portuguese.pdf.

Ćušić, Tarik (2022). *Deliverable D1.36 Report on the Bosnian Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-bosnian.pdf.

Eide, Kristine, Andre Kåsen, and Ingerid Løyning Dale (2022). *Deliverable D1.26 Report on the Norwegian Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-norwegian.pdf.

ELE Consortium (2022). *Deliverable D3.4 Digital Language Equality in Europe by 2030: Strategic Agenda and Roadmap*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/SRIA-and-roadmap.pdf.

Eskevich, Maria and Franciska de Jong (2022). *Deliverable D2.3 Report from CLARIN*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/consultation-CLARIN.pdf.

European Parliament (2018). *Language Equality in the Digital Age. European Parliament resolution of 11 September 2018 on Language Equality in the Digital Age (2018/2028(INI)*. http://www.europarl.europa.eu/doceo/document/TA-8-2018-0332_EN.pdf.

Gaidienė, Anželika and Aurelija Tamulionienė (2022). *Deliverable D1.23 Report on the Lithuanian Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-lithuanian.pdf.

Garabík, Radovan (2022). *Deliverable D1.30 Report on the Slovak Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-slovak.pdf.

Gaspari, Federico, Owen Gallagher, Georg Rehm, Maria Giagkou, Stelios Piperidis, Jane Dunne, and Andy Way (2022a). "Introducing the Digital Language Equality Metric: Technological Factors". In: *Proceedings of the Workshop Towards Digital Language Equality (TDLE 2022; co-located with LREC 2022)*. Ed. by Itziar Aldabe, Begoña Altuna, Aritz Farwell, and German Rigau. Marseille, France, pp. 1–12. http://www.lrec-conf.org/proceedings/lrec2022/workshops/TDLE/pdf/2022.tdle-1.1.pdf.

Gaspari, Federico, Annika Grützner-Zahn, Georg Rehm, Owen Gallagher, Maria Giagkou, Stelios Piperidis, and Andy Way (2022b). *Deliverable D1.3 Digital Language Equality (full specification)*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166 ELE. https://european-language-equality.eu/reports/DLE-definition.pdf.

Gaspari, Federico, Andy Way, Jane Dunne, Georg Rehm, Stelios Piperidis, and Maria Giagkou (2021). *Deliverable D1.1 Digital Language Equality (preliminary definition)*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/DLE-preliminary-definition.pdf.

Gavriilidou, Maria, Maria Giagkou, Dora Loizidou, and Stelios Piperidis (2022). *Deliverable D1.17 Report on the Greek Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-greek.pdf.

Giagkou, Maria, Penny Labropoulou, Stelios Piperidis, Miltos Deligiannis, Athanasia Kolovou, and Leon Voukoutis (2022). *Deliverable D1.37 Database and Dashboard*. European Language

Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/DLE-dashboard.pdf.

Gísladóttir, Guðrún (2022). *Deliverable D2.7 Report from ECSPM*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/consultation-ECSPM.pdf.

Gomez-Perez, Jose Manuel, Andres Garcia-Silva, Cristian Berrio, German Rigau, Aitor Soroa, Christian Lieske, Johannes Hoffart, Felix Sasaki, Daniel Dahlmeier, Inguna Skadiņa, Aivars Bērziņš, Andrejs Vasiļjevs, and Teresa Lynn (2022). *Deliverable D2.15 Technology Deep Dive – Text Analytics, Text and Data Mining, NLU*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/text-analytics-deep-dive.pdf.

Grützner-Zahn, Annika and Georg Rehm (2022). "Introducing the Digital Language Equality Metric: Contextual Factors". In: *Proceedings of the Workshop Towards Digital Language Equality (TDLE 2022; co-located with LREC 2022)*. Ed. by Itziar Aldabe, Begoña Altuna, Aritz Farwell, and German Rigau. Marseille, France, pp. 13–26. http://www.lrec-conf.org/proceedings/lrec2022/workshops/TDLE/pdf/2022.tdle-1.2.pdf.

Hajič, Jan, Maria Giagkou, Stelios Piperidis, Georg Rehm, and Natalia Resende (2021). *Deliverable D2.1 Specification of the consultation process*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/consultation-process.pdf.

Hajič, Jan, Tea Vojtěchová, and Maria Giagkou (2022). *Deliverable D2.5 Report from META-NET*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/consultation-META-NET.pdf.

Hegele, Stefanie, Rémi Calizzano, Annika Grützner-Zahn, Katrin Marheinecke, and Georg Rehm (2021a). *Deliverable D4.1 Promotional materials and PR package*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/promotional-materials.pdf.

Hegele, Stefanie, Barbara Heinisch, Antonia Popp, Katrin Marheinecke, Annette Rios, Dagmar Gromann, Martin Volk, and Georg Rehm (2022a). *Deliverable D1.16 Report on the German Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-german.pdf.

Hegele, Stefanie, Katrin Marheinecke, Jens-Peter Kückens, and Georg Rehm (2021b). *Deliverable D4.2 Communication and dissemination plan*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/communication-dissemination-plan.pdf.

Hegele, Stefanie, Katrin Marheinecke, and Georg Rehm (2022b). *Deliverable D2.6 Report from ELG*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/consultation-ELG.pdf.

Heuschkel, Maria (2022). *Deliverable D2.12 Report from Wikipedia*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/consultation-Wikipedia.pdf.

Hicks, Davyth (2022). *Deliverable D2.9 Report from ELEN*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/consultation-ELEN.pdf.

Hlavacova, Jaroslava (2022). *Deliverable D1.8 Report on the Czech Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-czech.pdf.

Hrasnica, Halid (2022). *Deliverable D2.11 Report from NEM*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/consultation-NEM.pdf.

Jelencsik-Mátyus, Kinga, Enikő Héja, Zsófia Varga, Tamás Váradi, László János Laki, and Győző Yang Zijian (2022). *Deliverable D1.18 Report on the Hungarian Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-hungarian.pdf.

Kaltenböck, Martin, Artem Revenko, Khalid Choukri, Svetla Boytcheva, Christian Lieske, Teresa Lynn, German Rigau, Maria Heuschkel, Aritz Farwell, Gareth Jones, Itziar Aldabe, Ainara Estarrona, Katrin Marheinecke, Stelios Piperidis, Victoria Arranz, Vincent Vandeghinste, and Claudia Borg (2022). *Deliverable D2.16 Technology Deep Dive – Data, Language Resources, Knowledge Graphs*. European Language Equality (ELE); EU project no. LC-01641480 – 1010-18166. https://european-language-equality.eu/reports/data-knowledge-deep-dive.pdf.

Kirchmeier, Sabine (2022). *Deliverable D2.8 Report from EFNIL*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/consultation-EFNIL.pdf.

Koeva, Svetla and Valentina Stefanova (2022). *Deliverable D1.5 Report on the Bulgarian Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-bulgarian.pdf.

Krauwer, Steven (2003). "The Basic Language Resource Kit (BLARK) as the First Milestone for the Language Resources Roadmap". In: *Proceedings of the International Workshop Speech and Computer (SPECOM 2003)*. Moscow, Russia.

Krek, Simon (2022). *Deliverable D1.31 Report on the Slovenian Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-slovenian.pdf.

Krstev, Cvetana and Ranka Stanković (2022). *Deliverable D1.35 Report on the Serbian Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-serbian.pdf.

Lindén, Krister and Wilhelmina Dyster (2022). *Deliverable D1.13 Report on the Finnish Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-finnish.pdf.

Lynn, Teresa (2022). *Deliverable D1.20 Report on the Irish Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-irish.pdf.

Magnini, Bernardo, Alberto Lavelli, and Manuela Speranza (2022). *Deliverable D1.21 Report on the Italian Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-italian.pdf.

Marheinecke, Katrin, Annika Grützner-Zahn, and Georg Rehm (2022). *Deliverable D4.4 Report on ELE Conference*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/conference.pdf.

Maynard, Diana, Joanna Wright, Mark A. Greenwood, and Kalina Bontcheva (2022). *Deliverable D1.11 Report on the English Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-english.pdf.

Melero, Maite, Blanca C. Figueras, Mar Rodríguez, and Marta Villegas (2022a). *Deliverable D1.6 Report on the Catalan Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-catalan.pdf.

Melero, Maite, Pablo Peñarrubia, David Cabestany, Blanca C. Figueras, Mar Rodríguez, and Marta Villegas (2022b). *Deliverable D1.32 Report on the Spanish Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-spanish.pdf.

Moshagen, Sjur Nørstebø, Rickard Domeij, Kristine Eide, Peter Juel Henrichsen, and Per Langgård (2022). *Deliverable D1.38 Report on the Nordic Minority Languages*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-nordic-languages.pdf.

Muischnek, Kadri (2022). *Deliverable D1.12 Report on the Estonian Language*. Reports on European Language Equality (ELE) | Coordinator: Prof. Dr. Andy Way, Co-Coordinator: Prof. Dr. Georg Rehm, received funding from the European Union (EU project no. LC-01641480 – 101018166). https://european-language-equality.eu/reports/language-report-estonian.pdf.

Ogrodniczuk, Maciej, Piotr Pęzik, Marek Łaziński, and Marcin Miłkowski (2022). *Deliverable D1.27 Report on the Polish Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-polish.pdf.

Păiș, Vasile and Dan Tufiș (2022). *Deliverable D1.29 Report on the Romanian Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-romanian.pdf.

Pedersen, Bolette Sandford, Sussi Olsen, and Lina Henriksen (2022). *Deliverable D1.9 Report on the Danish Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-danish.pdf.

Prys, Delyth, Gareth Watkins, and Stefano Ghazzali (2022). *Deliverable D1.34 Report on the Welsh Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-welsh.pdf.

Rehm, Georg, ed. (2023). *European Language Grid: A Language Technology Platform for Multilingual Europe*. Cognitive Technologies. Cham, Switzerland: Springer.

Rehm, Georg, Federico Gaspari, German Rigau, Maria Giagkou, Stelios Piperidis, Annika Grützner-Zahn, Natalia Resende, Jan Hajic, and Andy Way (2022a). "The European Language Equality Project: Enabling digital language equality for all European languages by 2030". In: *The Role of National Language Institutions in the Digital Age – Contributions to the EFNIL Conference 2021 in Cavtat*. Ed. by Željko Jozić and Sabine Kirchmeier. Budapest, Hungary: Nyelvtudományi Kutatóközpont, Hungarian Research Centre for Linguistics, pp. 17–47.

Rehm, Georg, Stefanie Hegele, and Katrin Marheinecke (2022b). *Deliverable D4.3 Report on EP/EC Workshop*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/EC-workshop.pdf.

Rehm, Georg, Stefanie Hegele, and Katrin Marheinecke (2022c). *Deliverable D4.6 ELE Book Publication*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/book-publication.pdf.

Rehm, Georg and Hans Uszkoreit, eds. (2012). *META-NET White Paper Series: Europe's Languages in the Digital Age*. 32 volumes on 31 European languages. Heidelberg etc.: Springer.

Robinson-Jones, Charlie and Ydwine R. Scarse (2022). *Deliverable D1.39 Report on the West Frisian Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-frisian.pdf.

Rögnvaldsson, Eiríkur (2022). *Deliverable D1.19 Report on the Icelandic Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-icelandic.pdf.

Rosner, Mike and Claudia Borg (2022). *Deliverable D1.25 Report on the Maltese Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-maltese.pdf.

Rufener, Andrew and Philippe Wacker (2022). *Deliverable D2.4 Report from LT-innovate*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/consultation-LTInnovate.pdf.

Sánchez, José Manuel Ramírez and Carmen García-Mateo (2022). *Deliverable D1.15 Report on the Galician Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-galician.pdf.

Sarasola, Kepa, Itziar Aldabe, Arantza Diaz de Ilarraza, Ainara Estarrona, Aritz Farwell, Inma Hernaez, and Eva Navas (2022). *Deliverable D1.4 Report on the Basque Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-basque.pdf.

Skadiņa, Inguna, Ilze Auziņa, Baiba Valkovska, and Normunds Grūzītis (2022). *Deliverable D1.22 Report on the Latvian Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-latvian.pdf.

Steurs, Frieda, Vincent Vandeghinste, and Walter Daelemans (2022). *Deliverable D1.10 Report on the Dutch Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-dutch.pdf.

STOA (2018). *Language equality in the digital age – Towards a Human Language Project*. STOA study (PE 598.621), IP/G/STOA/FWC/2013-001/Lot4/C2. https://data.europa.eu/doi/10.2861/136527.

Tadić, Marko (2022). *Deliverable D1.7 Report on the Croatian Language*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/language-report-croatian.pdf.

Thönnissen, Marlies (2022). *Deliverable D2.2 Report from CLAIRE*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/consultation-CLAIRE.pdf.

Wang, Alex, Yada Pruksachatkun, Nikita Nangia, Amanpreet Singh, Julian Michael, Felix Hill, Omer Levy, and Samuel R. Bowman (2019). "SuperGLUE: A Stickier Benchmark for General-Purpose Language Understanding Systems". In: *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*. Ed. by Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett, pp. 3261–3275. https://proceedings.neurips.cc/paper/2019/hash/4496bf24afe7fab6f046bf4923da8de6-Abstract.html.

Way, Andy, Georg Rehm, Jane Dunne, Maria Giagkou, Jose Manuel Gomez-Perez, Jan Hajič, Stefanie Hegele, Martin Kaltenböck, Teresa Lynn, Katrin Marheinecke, Natalia Resende, Inguna Skadina, Marcin Skowron, Tereza Vojtěchová, and Annika Grützner-Zahn (2022a). *Deliverable D2.18 Report on the state of Language Technology in 2030*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/LT-in-2030.pdf.

Way, Andy, Georg Rehm, Jane Dunne, Jan Hajič, Teresa Lynn, Maria Giagkou, Natalia Resende, Tereza Vojtěchová, Stelios Piperidis, Andrejs Vasiljevs, Aivars Berzins, Gerhard Backfried, Marcin Skowron, Jose Manuel Gomez-Perez, Andres Garcia-Silva, Martin Kaltenböck, and Artem Revenko (2022b). *Deliverable D2.17 Report on all external consultations and surveys*. European Language Equality (ELE); EU project no. LC-01641480 – 101018166. https://european-language-equality.eu/reports/external-consultations.pdf.