



Chapter 4

European Language Technology in 2022/2023

Maria Giagkou, Teresa Lynn, Jane Dunne, Stelios Piperidis, and Georg Rehm

Abstract This chapter presents the results of an extensive empirical investigation of the digital readiness of European languages, and provides a snapshot of the support they are offered through technology as of 2022. The degree of digital readiness was assessed on the basis of the availability of language resources and technologies for each language under investigation and a cross-language comparison was performed. As a complementary approach, the perspectives and opinions of LT users, developers and the regular citizen were acquired in order to fully understand the EU’s LT landscape. Both the objective empirical findings and the voice of the community clearly indicate that there is an extreme imbalance across languages when it comes to the individual levels of technological support. Although the LT field as a whole has demonstrated remarkable progress during the last decade, this progress is not equally evidenced across all languages, posing, more acutely than ever before, a threat of digital extinction for many of Europe’s lesser supported languages.¹

1 Introduction

More than ten years ago, the study “Europe’s Languages in the Digital Age” concluded that most European languages are under threat in the digital age. The study, prepared by more than 200 experts and documented in 32 volumes of the META-NET White Paper Series (Rehm and Uszkoreit 2012), assessed Language Technology (LT) support for each language in four different areas: automatic translation,

Maria Giagkou · Stelios Piperidis
R. C. “Athena”, Greece, mgiagkou@athenarc.gr, spip@athenarc.gr

Teresa Lynn · Jane Dunne
Dublin City University, ADAPT Centre, Ireland, teresa.lynn@adaptcentre.ie,
jane.dunne@adaptcentre.ie

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

¹ This chapter includes findings from Way et al. (2022) and makes use of the general sections written by the ELE consortium for the language reports (Giagkou et al. 2022).

speech interaction, text analysis and the availability of language resources (LRs). The results were alarming: most of the 32 European languages investigated were evaluated as severely under-resourced and some almost completely neglected.

During the last ten years since the publication of the META-NET White Papers, the LT field as a whole has seen remarkable progress. In particular, the advent of data-driven approaches such as deep learning and neural networks, together with the considerable increase in the number and quality of LRs for a number of languages, have yielded previously unforeseeable results. However, is this remarkable progress equally evidenced across all languages, or is the gap between “big” and “small” languages documented in 2012 still present in 2022/2023?

The question of whether languages can be considered digitally equal has become increasingly relevant in recent years, with a growing number of studies attempting to quantify digital readiness and compare languages in this respect. Methods have varied, with some assessing the level of technology support based on mentions of a language at NLP publication venues or language resource catalogues (e. g., Blasi et al. 2022; Joshi et al. 2020; Ranathunga and Silva 2022) or on websites describing LT tools and services (e. g., Simons et al. 2022). However, the overall conclusion is always the same; from a technological perspective, there is a striking imbalance across languages in terms of support, and it is clear that not all languages benefit equally and fairly from the overall progress in LT advances.

In the ELE project, we took an empirical approach to quantifying digital readiness of a language and providing an evidence-based grounding on which languages can be compared. We started by applying the Digital Language Equality (DLE) Metric (see Chapter 3) to examine both the current state of technology support and the potential for short- and mid-term development of LT (Section 2). We continued with a quantitative investigation of the various perspectives and dimensions of current technological support, as this is reflected in the Language Resources and Technologies (LRTs) collection of the European Language Grid (ELG, Rehm 2023). The results of this empirical assessment were then supplemented by surveys and consultations with a broad representation of LT developers and LT users and consumers, who provided feedback and insight as to their experiences with LTs for EU languages (Section 3). Furthermore and most importantly, we focused on a large number of European languages and provided updates of the META-NET White Papers in the form of the ELE Language Reports (Giagkou et al. 2022), condensed versions of which are presented in Chapters 5–37. It is only through such a holistic examination that a clear picture of the current status and future prospects of DLE can be gained.

2 How Do Europe’s Languages Compare?

In this section, we first describe our source of evidence and methodology (Section 2.1), followed by a presentation of our findings (Section 2.2).

2.1 Source of Evidence and Methodology

To compare the level of technology support across languages, we considered the language technology tools and resources in the catalogue of the European Language Grid (Rehm 2023; Piperidis et al. 2023; Labropoulou et al. 2020). The comparative evaluation was performed on various dimensions.

- The current state of technology support, as indicated by the availability of tools and services² broadly categorised into a number of core LT application areas:
 - Text processing (e. g., part-of-speech tagging, syntactic parsing)
 - Information extraction and retrieval (e. g., search and information mining)
 - Translation technologies (e. g., machine translation, computer-aided translation)
 - Natural language generation (NLG, e. g., text summarisation, simplification)
 - Speech processing (e. g., speech synthesis, speech recognition)
 - Image/video processing
 - Human-computer interaction (HCI, e. g., tools for conversational systems)
- The potential for short- and mid-term development of LTs, insofar as this potential can be approximated by the current availability of resources that can be used as training or evaluation data. The availability of data was investigated with regard to a small number of basic types of resources:
 - Text corpora
 - Parallel corpora
 - Multimodal corpora (incl. speech, image, video)
 - Language models
 - Lexical resources (incl. dictionaries, wordnets, ontologies, etc.)

We measured the LT support for 87 national, regional and minority European languages with regard to each of the dimensions mentioned above based on their respective coverage in the ELG catalogue. For the types of resources and application areas, the respective percentage of resources that support a specific language over the total number of resources of the same type was calculated, as well as their average. Subsequently, each language was assigned to one band per resource type and per application area and to an overall band, on a four-point scale, inspired by the scale used in the META-NET White Paper Series, as follows:

1. *Weak or no support*: the language is present (as content, input or output language) in <3% of the ELG resources of the same type

² Tools tagged as “language independent” without mentioning any specific language are *not* taken into account. Such tools can certainly be applied to a number of languages, either as readily applicable or following fine-tuning, adaptation, training on language-specific data etc., yet their exact language coverage or readiness is difficult to ascertain.

2. *Fragmentary support*: the language is present in $\geq 3\%$ and $< 10\%$ of the ELG resources of the same type
3. *Moderate support*: the language is present in $\geq 10\%$ and $< 30\%$ of the ELG resources of the same type
4. *Good support*: the language is present in $\geq 30\%$ of the ELG resources of the same type

The thresholds for defining the four bands (i. e., 3%, 10% and 30%) were informed by an exploratory *k*-means 4-cluster analysis based on all data per application and resource type, in order to investigate the boundaries of naturally occurring clusters in the data. The boundaries of the clusters were then used to define the bands per application area and resource type. The overall level of support for a language was calculated based on the average coverage of all dimensions investigated.

The ELG platform harvests several major LR/LT repositories³ and, on top of that, more than 6,000 additional LRTs were identified and documented by language informants in the ELE consortium. These records contain multiple levels of metadata granularity as part of their descriptions. At the time of investigation, the ELG catalogue comprised more than 11,500 metadata records, encompassing both data and tools/services, covering almost all European languages, both official and regional as well as minority ones.

It should be noted that due to the evolving nature of this extensive catalogue and differing approaches taken in documenting records, certain categories of metadata captured are not yet at the level of consistency required to carry out a reliable cross-lingual comparison at a granular level. For example, information provided on corpus size, annotation type, licensing type, size unit type, and so on, still varies across records for many languages, while numerous gaps exist for others. As the ELG catalogue is continuously growing, the comprehensiveness, accuracy and level of detail of the records are expected to improve over time.

For the purposes of a high-level comparison, the results presented here are based on relative counts of entries in the ELG for the varying types of data resources and tools/services for each language. As such, the positioning of each language into a specific level of technology support is subject to change as it reflects a snapshot of the available resources at the time of investigation.

That said, we consider the current status of the ELG catalogue and the higher-level findings below representative with regard to the current existence of LT resources for Europe's languages.

³ At the time, ELG harvested ELRC-SHARE, LINDAT/CLARIAH-CZ, CLARIN.SI, CLARIN-PL and the datasets section of Hugging Face (Labropoulou et al. 2023).

2.2 Results and Findings

As discussed above, our analysis takes into account a number of dimensions for data and tools/services. Table 1 reports the detailed results per language per dimension investigated and the classification of each language into an overall level of support.

The best supported language is, as expected, English, the only language that is classified in the *good support* group. French, German and Spanish form a group of languages with *moderate support*. Although they are similar to English in some dimensions (e. g., German in terms of available speech technologies and Spanish in terms of available models), overall they have not yet reached the coverage that English has according to the ELG catalogue. All other official EU languages are clustered in the *fragmentary support* group, with the exception of Irish and Maltese, which have only *weak or no support*. From the remaining languages, (co-)official at the national or regional level in at least one European country and other minority and lesser spoken languages,⁴ Norwegian and Catalan belong to the group of languages with *fragmentary support*. Basque, Galician, Icelandic and Welsh are borderline cases; while they are grouped in the *fragmentary support* level, they barely pass the threshold of the lowest level. All other languages are supported by technology either weakly or not at all. Figure 1 visualises these findings.

Looking into particular dimensions of data availability, it is evident that an abundance of training data for developing LTs is available only for a few languages with high commercial interest. For many (the majority of) European languages, this is not the case and only corpora which are minuscule in comparison to English are available. When investigating the current availability of some of the data types mentioned in the previous paragraph, as represented in the resources hosted in ELG in January 2023,⁵ it is apparent that even the best-supported languages in this dimension, Spanish and English, are still only moderately covered (Figure 2). With respect to multimodal data, all languages with the exception of English are weakly covered, with some, e. g., Maltese and Luxembourgish, severely underrepresented (Figure 3).

Although the data gaps per language are different, some data types are particularly sparse across many languages. These include: large language models, both monolingual and multilingual; multimodal data, especially speech in conversational settings (dialogues) from speakers of different ages, genders and linguistic/dialectal backgrounds, but also video corpora for sign languages; domain-specific data (e. g., medical, legal or media among many others of interest); data for language use on

⁴ In addition to the languages listed in Table 1, ELE also investigated Alsatian, Aragonese, Arberesh, Aromanian, Asturian, Breton, Cimbrian, Continental Southern Italian (Neapolitan), Cornish, Eastern Frisian, Emilian, FrancoProvençal (Arpitan), Friulian, Gallo, Griko, Inari Sami, Karelian, Kashubian, Ladin, Latgalian, Ligurian, Lombard, Lower Sorbian, Lule Sami, Mocheno, Northern Frisian, Northern Sami, Picard, Piedmontese, Pite Sami, Romagnol, Romany, Rusyn, Sardinian, Scottish Gaelic, Sicilian, Skolt Sami, Southern Sami, Tatar, Tornedalian Finnish, Venetian, Võro, Walser and Yiddish. The scores for all of these languages are very low, placing all of them in the *weak or no support* group.

⁵ The DLE dashboard enables more fine-grained comparisons. It dynamically visualises the contents of the ELG catalogue and offers an up-to-date snapshot of the current availability of LRTs (see Chapter 3): <https://live.european-language-grid.eu/catalogue/dashboard>.

		Tools and Services						Language Resources						
		Text Processing	Speech Processing	Image/Video Processing	Information Extraction and IR	Human-Computer Interaction	Translation Technologies	Natural Language Generation	Text Corpora	Multimodal Corpora	Parallel Corpora	Models	Lexical Resources	Overall
EU official languages	Bulgarian	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Croatian	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Czech	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Green	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Danish	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Green	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Dutch	Light Yellow	Light Green	Light Green	Light Yellow	Light Green	Light Green	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	English	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
	Estonian	Light Yellow	Light Green	Light Green	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Finnish	Light Yellow	Light Green	Light Green	Light Yellow	Light Yellow	Light Green	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	French	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green
	German	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green
	Greek	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Hungarian	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Green	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Irish	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Italian	Light Yellow	Light Green	Light Green	Light Yellow	Light Green	Light Green	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Latvian	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Lithuanian	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Maltese	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Polish	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
Portuguese	Light Green	Light Green	Light Green	Light Yellow	Light Green	Light Green	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	
Romanian	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Green	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	
Slovak	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	
Slovenian	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	
Spanish	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	Light Green	
Swedish	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Green	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	
National level	Albanian	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Bosnian	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Icelandic	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Luxembourgish	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Macedonian	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Norwegian	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Green	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Serbian	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
Regional level	Basque	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Catalan	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Green	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Faroese	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Frisian (Western)	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Galician	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Jerriais	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Low German	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Manx	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Mirandese	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
	Occitan	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow
Sorbian (Upper)	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	
Welsh	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	
<i>All other languages</i>		Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow	Light Yellow

Table 1 State of technology support, in 2022, for selected European languages with regard to core Language Technology areas and data types as well as overall level of support (light yellow: weak/no support; yellow: fragmentary support; light green: moderate support; green: good support)

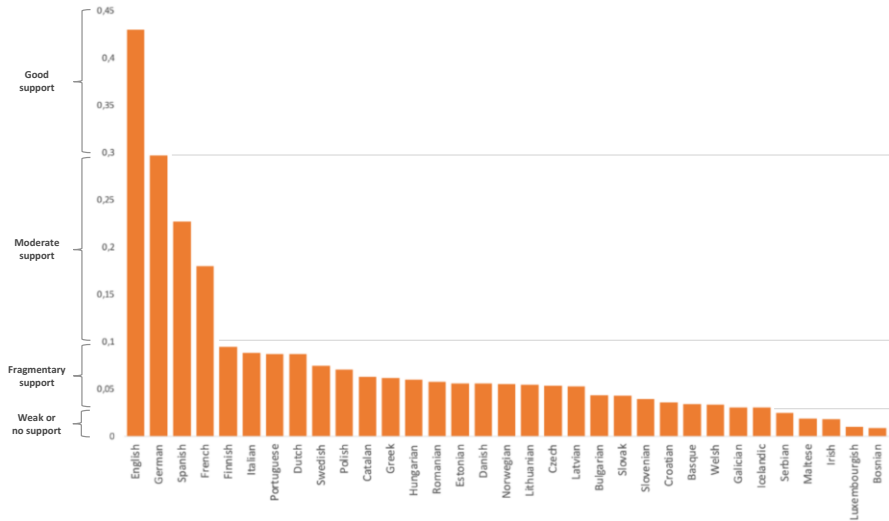


Fig. 1 Overall state of technology support for selected European languages (2022)

social media; semantic resources (e. g., semantic annotations and knowledge bases); data for language pathologies; benchmarks, i. e., well-designed gold-standard corpora for evaluating LT systems or fine-tuning language models.

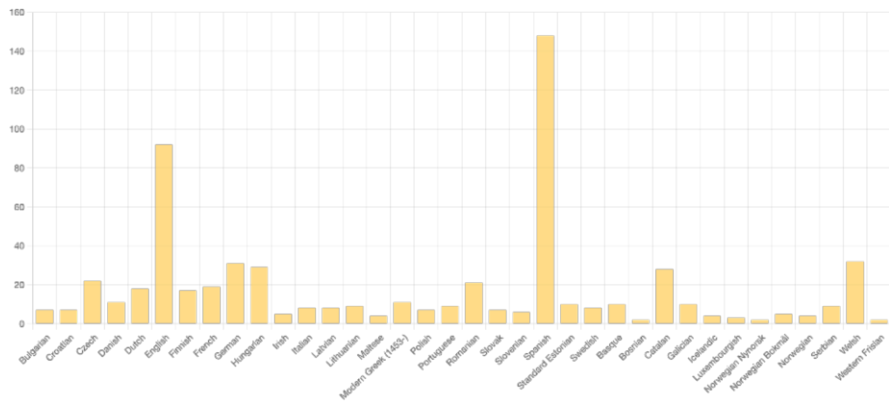


Fig. 2 Number of language models available in the catalogue of the European Language Grid for the EU official languages and for some indicative non-EU official ones (as of January 2023)

Similarly to data, the identified gaps for technologies are very diverse across languages. While overall LTs for English are numerous and at the state-of-the-art level, a number of very small minoritised languages lack even basic tools such as spell checkers. In the worst case, they are not even supported by operating systems. Nevertheless, there seems to be a generalised consensus that, when it comes to languages

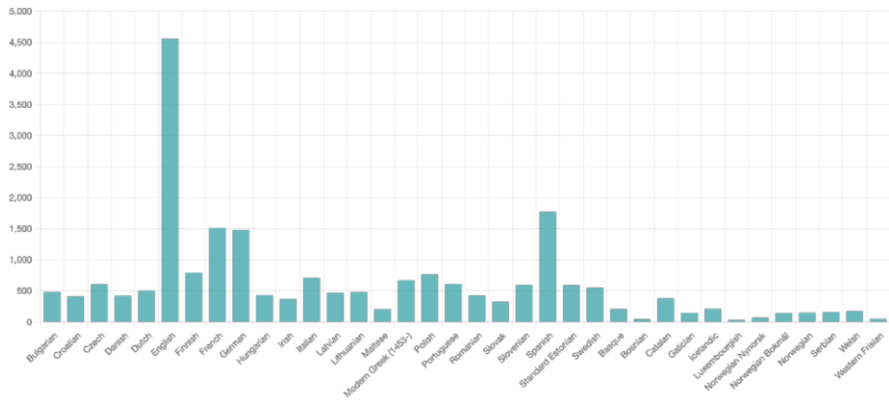


Fig. 3 Number of multimodal datasets (i. e., media type: audio, video or image) available in the catalogue of the European Language Grid for the EU official languages and for some indicative non-EU official ones (as of January 2023)

for which at least a minimum level of technological support has been achieved, the technologies most urgently needed include: discourse processing, bias detection and anonymisation, conversational systems and question-answering in the wider context of HCI, NLG (with summarisation mentioned frequently) and Natural Language Understanding (NLU), e. g., even English and German are currently supported by less than 100 HCI or NLG systems on ELG, while some languages like Bosnian and Norwegian Nynorsk are not supported at all (Figures 4 and 5).

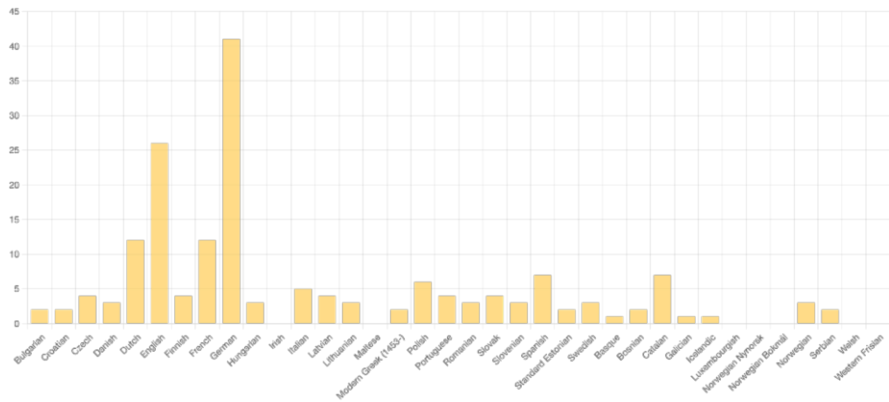


Fig. 4 Number of Human-Computer Interaction systems described in the catalogue of the European Language Grid for the EU official languages and for some indicative non-EU official ones (as of January 2023)

The results of this analysis are only informative of the relative positioning of languages, but not of the technological progress achieved by a specific language. The

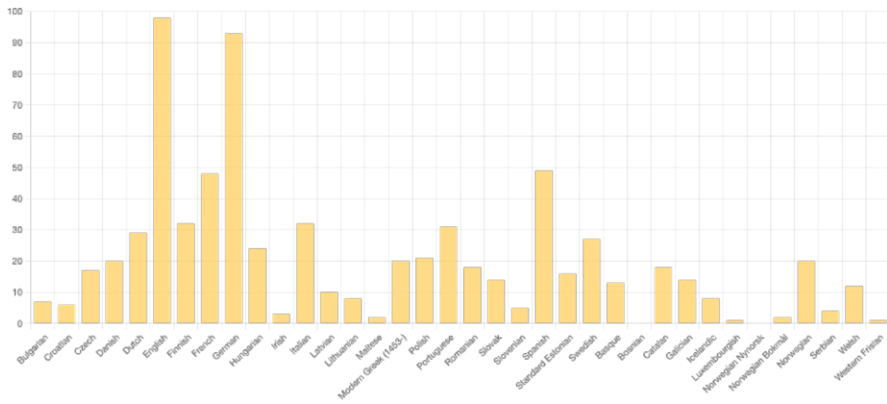


Fig. 5 Number of Natural Language Generation systems described in the catalogue of the European Language Grid for the EU official languages and for some indicative non-EU official ones (as of January 2023)

LT field as a whole has significantly progressed in the last ten years and remarkable progress has been achieved for specific languages in terms of quantity, quality and coverage of LRTs. It is at the same time undebatable that the technology requirements for a language to be considered digitally supported by today's standards have changed significantly in the last ten years (e. g., the prevalent use of virtual assistants, chatbots, improved text analytics capabilities, etc.). Nevertheless, the imbalance in distribution across languages which was documented in the META-NET White Papers in 2012 still exists, and the huge distance between the best supported languages and the minimally supported ones was still evidenced in 2022. It is exactly this distance that needs to be ideally eliminated, or at least reduced, in order to move towards DLE and avert the risks of digital language extinction.

It should be noted that this analysis does not include a fifth level, *excellent support*, for the grouping of languages, in addition to the four levels described in Section 2.1. Currently, no European language, not even English, is optimally supported by technology, i. e., the goal of *Deep Natural Language Understanding* has not been reached yet for any language. Although recently there have been many breakthroughs in AI, Computer Vision, Machine Learning and LT, we are still far from the grand challenge of highly accurate deep language understanding, which is able to seamlessly integrate modalities, situational and linguistic context, general knowledge, meaning, reasoning, emotion, irony, sarcasm, humour, culture, explain itself on request, and be effected as required on the fly and at scale. A language can only be considered excellently supported by technology if and when the goal of Deep Natural Language Understanding has been reached.

3 The Voice of the Community

The findings in Section 2 are extremely valuable in terms of highlighting the status quo across Europe with respect to LT support. However, facts and figures alone cannot paint the full picture. The perspectives and opinions of LT users, developers and the average citizen were also required in order to fully understand the EU's LT landscape. As a project from the community for the community, the ELE consortium wanted to ensure that as many voices as possible were heard and taken as input for the ELE strategic agenda and roadmap.

A broad spectrum of stakeholders was consulted to achieve this wider insight into the levels of LT support across European languages (also see Chapter 38, p. 229 ff.). We distinguish between three main stakeholder groups: *LT developers* (industry and research), *LT users* (commercial and academic users) and *EU citizens*, i. e., the general public who use and consume LTs in everyday personal and professional settings, often without even realising it. Each group is diverse, some including many sub-groups, representing a variety of sectors and domains. For the latter, we looked at the interesting subdivisions of commercial and academic users as well as EU citizens. The first two groups are represented in the ELE consortium with several networks, initiatives and associations, representing the views of their constituencies, highlighting their wishes, demands and needs towards full DLE in Europe.

Further insight was gained from a number of online surveys and expert interviews targeting LT developers, users and consumers. The surveys investigated language coverage, evaluated the current situation of LT in Europe and encouraged participants to share their predictions and visions for the future. In this section, we look, in particular, at the evaluation of the current situation to see how these opinions compare to the empirical results presented in Section 2 and also in Chapter 39 (p. 245 ff.).

3.1 Developers of Language Technologies

European LT developers are a diverse group of stakeholders, comprising *academic* and *industrial entities* in the field of LT. Beyond research, they develop pre-commercial prototypes, algorithms, applications and systems. An initial grouping is, thus, *LT industry* and *LT research* (also see Rehm et al. 2023, 2020). This section focuses on their view about the situation as of 2022, while Section 3 in Chapter 38 presents their forward-looking predictions going towards 2030.

In addition to the horizontal grouping into research and industry, a vertical categorisation can be performed with regard to the multi- and interdisciplinary nature of LT. LT is in the intersection of Linguistics and Computational Linguistics, Computer Science and AI, while at the same time encompassing methods and findings from Cognitive Science and Psychology, Mathematics, Statistics, Philosophy and other fields. As a result, the ELE stakeholder group of LT developers were identified not only within the strict limits of *LT per se*, but also in the neighbouring disciplines of *AI* and *Digital Humanities/Social Science and Humanities (DH/SSH)*.

Europe has a long-standing research, development and innovation tradition in LT with over 800 centres performing excellent, highly visible and internationally recognised research on all European and many non-European languages. In terms of companies, the European LT industry was estimated to comprise 435 companies (LT-Innovate 2016) or 473 LT vendors in the EU26 plus Iceland and Norway in 2017 (Vasiljevs et al. 2019). In January 2023, the ELG catalogue comprised more than 800 commercial entities including integrators and a certain number of user companies.

In order to disseminate the survey widely, we mobilised existing European networks, associations, initiatives and projects. Some of the well-established and long-standing pan-European LT networks were represented in the ELE consortium and they constituted the core ELE LT developers stakeholders groups (i. e., CLAIRE, CLARIN, LT-Innovate, META-NET and ELG). The ELE partners that represented these initiatives not only contributed their views to the project but also facilitated access to and elicitation of the views of their constituency and members. In particular, they coordinated the distribution of the survey to their members, conducted interviews and focused consultation meetings, where needed and appropriate, and consolidated their feedback (Thönnissen 2022; Eskevich and Jong 2022; Rufener and Wacker 2022; Hajič et al. 2022; Hegele et al. 2022).

The survey encompassed 45 questions in total. A respondent was presented with 32 (minimum) to 45 (maximum) questions, including “if other” questions. In all, 35 questions were mandatory and 27 were closed questions (single or multiple choice). The survey was structured into four main parts: Part A. Respondents’ profiling, Part B. Language coverage, Part C. Evaluation of current situation, and Part D. Predictions and visions for the future (see also Chapter 38, p. 229 ff., and Chapter 39, p. 245 ff.). For assessing the current situation from the perspective of LT developers, we focus on the findings based on responses to Parts B and C of the survey.

The LT developers survey was filled in by 321 different respondents who represent 223 different organisations (Way et al. 2022). 73% of the organisations are research or academic institutions and 22% are private companies. In 5% of responses the “Other” value was indicated as the type of organisation and this has been further specified as freelancer/private practitioner or currently unemployed, government agency, not-for-profit organisation, etc. Of note here is the response to the question “*What languages does your organisation conduct research in and/or for what languages do you offer services, software, resources, models etc.?*”. Figure 6 shows the languages supported by survey respondents’ organisations. All official EU languages are covered as well as other state official, regional and/or co-official European languages. The five most frequently mentioned languages are, yet again, English, German, Spanish, French and Italian.

In order to evaluate the current situation and to further grasp the main challenges and obstacles the European LT community faces, the survey participants were asked to indicate their level of agreement with a set of potential obstacles (Figure 7). As part of a free text question, respondents were also given the opportunity to elaborate on the obstacles and challenges indicated in the questions and/or add any other obstacle/challenge not previously listed.

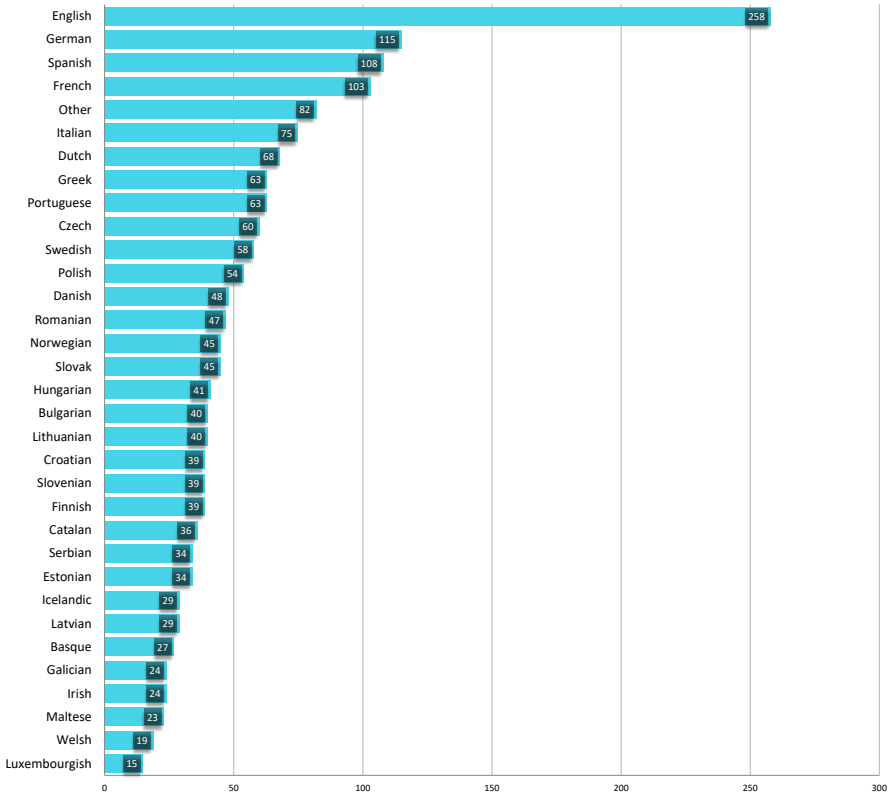


Fig. 6 LT developers survey – languages supported by the respondents’ organisations in their research and development activities

With respect to questions about the status quo of the languages, most of the participants agreed or strongly agreed that the importance of multilinguality in the European landscape does not always receive adequate recognition, and the smaller languages appear not to be attractive enough for industry and investors (74% agreed or strongly agreed on this point). This was backed up by comments relating to how industrial players can find a commercial interest in pre-competitive investments for “larger” languages, while this will rarely be the case for “smaller” ones. It was suggested that in that situation, the role of additional investors for the development of LTs for “smaller” languages should be played by bodies either at national or EU level. Moreover, it was noted that it is very often the case that small languages can rely on public funding only, which however is considered insufficient. For this reason, it was argued that public investments for small languages are necessary on a larger scale to really make them available to the wider community. It was also observed that the cost of developing LTs for a language is usually constant, regardless of the number of speakers of that language. Furthermore, for languages with larger numbers of speakers, it can often be easier to collect LRs: for instance, the larger

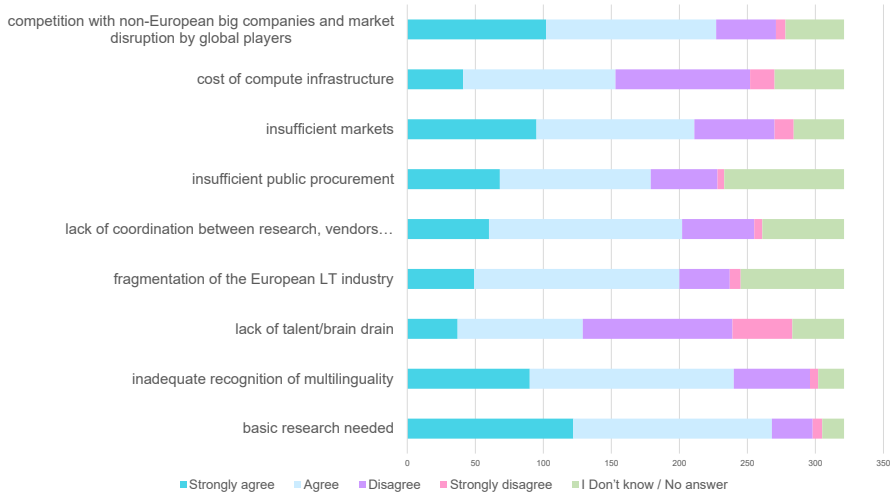


Fig. 7 LT developers survey – challenges the European LT community currently faces, according to LT developers

the number of speakers, the more online content is produced, which in turn can be collected and provide the raw language data necessary for the development of LRTs.

It was reported that this situation was even worse for non-standard languages: local dialects, non-standard written language on social media platforms, non-standard language for speech recognition, and non-standard language as used by migrants or citizens with a migration background. There is hardly ever funding available for creating LRs for non-standard varieties. There is equally little incentive for researchers to publish their work on small languages, resulting in the dominance of the English language in scientific literature.

3.2 Users of Language Technologies

Commercial users were those respondents representing companies in the sector of Information and Communication Technologies (ICTs) and eCommerce (e.g., Megabyte Ltd, A Capela group, Telecats), energy (e.g., Shell, Menai Science Park Ltd) and business services (e.g., Spencer Stuart, Inuits, Projectus grupa). They also included respondents from the following groups: self-employed language professionals (e.g., translators); professionals working on different economic sectors (e.g., banking, health); independent professionals/consultants; professionals working in public administration; media and publishing professionals.

Academic users included researchers, data scientists, university professors, language teachers, lecturers, and Master’s and PhD students. Some non-governmental organisations (NGOs) were also represented in the survey, such as Federal Lezghin

National and Cultural Autonomy, and representatives of public administration, such as National Youth Service (Ministry of Education, Children and Youth, Luxembourg), Hungarian National Research, Development and Innovation Office and the Government of the Balearic Islands. In addition, Wikipedia partners collected responses from representatives of the various Wikipedia projects, such as Wikimedia Community User Group Malta, Wikimedia Hungary, Wikimedia UK, and Wikimedia Community Ireland, to name a few. The full list of stakeholders of the LT users and consumers survey is presented in Way et al. (2022).

Six well-known European initiatives disseminated the survey within their networks and produced one report each, based on their respective constituencies. These include the European Federation of National Institutions for Language (EFNIL, Kirchmeier 2022), the European Language Equality Network (ELEN, Hicks 2022), the European Civil Society Platform for Multilingualism (ECSPM, Gísladóttir 2022), the New European Media initiative (NEM, Hrasnica 2022), the Association of European Research Libraries (LIBER, Blake 2022) and Wikipedia (Heuschkel 2022).

The survey obtained a total of 246 responses. The results show that contributions came from a diverse range of economic sectors and professional activities, but most of the respondents worked in the education and research sector with 130 responses (53%) out of 246, that is, most respondents were researchers, university professors, assistant professors, lecturers or held other academic positions. The survey was also filled out by representatives of NGOs, large enterprises, SMEs, government departments and independent contractors and consultants in diverse economic sectors. The 15 (6%) respondents who selected the option “other” represented non-governmental bodies, non-profit organisations, public sector organisations, social organisations and independent government departments.

Of relevance to assessing the current situation, we note here the responses to the question “*In general terms, how do you evaluate the performance of the tools you use for the official European language(s) you work with*”. Responses were captured through a 4-point Likert scale (where 1 indicated *very poor support*, 2 *poor support*, 3 *good support* and 4 *excellent support*). The list of LTs evaluated can be seen in Way et al. (2022). Figure 8 shows the average score for each of the European languages evaluated. The results show striking differences in technological support between European languages. Unsurprisingly, English is very well supported with a mean score of 3.4, while the group formed by German, French and Spanish follows with a mean score between 2.4 and 2.5. All other European languages were considered to have either *poor support* (mean scores ranging from 1 to 1.3), *very poor support* or *no support* at all with scores below 1.

3.3 European Citizens as Consumers of Language Technologies

In addition to the consultation with stakeholders that represent communities of users and consumers, a survey targeting European citizens was carried out to make sure that their voices also play a decisive role in the pursuit of full DLE in Europe. This

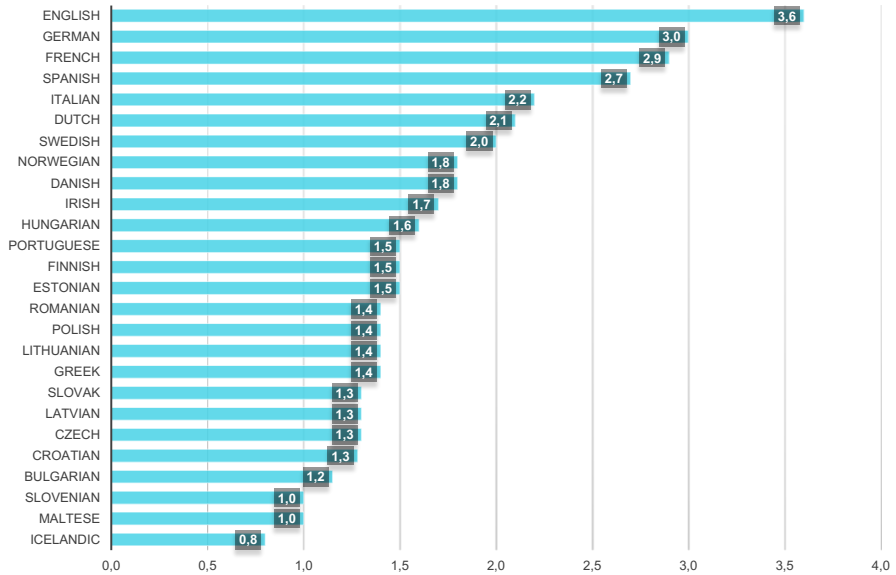


Fig. 8 LT users survey – level of technological support: average scores for the European language(s) that respondents work with

consultation with a larger and more diverse cohort of consumers allowed us to obtain a more accurate picture of the current scenario in terms of LT support across European languages and have a more representative basis for a technological and scientific forecasting on how LTs can be deployed and applied in Europe by 2030.

The citizens' survey was launched in January 2022 and closed on 01 May 2022. It was made available in 35 languages and disseminated across 28 countries.⁶ For each country we created a standalone survey so that respondents only saw the version in the language of the country in which they were based. For countries with more than one official language, we created a standalone version of the survey in each language spoken in the country, e. g., four surveys were set up in Spain (in Spanish, Catalan, Galician and Basque). This approach allowed us to specifically target regions where we were more likely to find communities of respondents that were speakers of that language. More details on this survey and the community consultation methodology are presented in Chapter 38 (p. 229 ff.).

In total, 21,108 complete responses were collected. However, as the collection of survey responses through commercial online services is known to present some known issues that can render results unreliable (Lawlor et al. 2021), closer inspection revealed a number of flags indicating unreliable responses. These responses were filtered from the dataset, and as such, a final 20,586 responses were analysed.

⁶ While ELE investigated about 90 European languages, we only produced translated versions for those languages for which native speaker post-editing was available. The 35 languages covered by the survey represent the support offered through the ELE consortium members.

Respondents provided profiling questions and were asked to list all of the languages they speak. Of particular interest in our examination of the current situation is the response to question 6 “Please rate all the types of software applications, apps, tools or devices you use for your language(s)”.

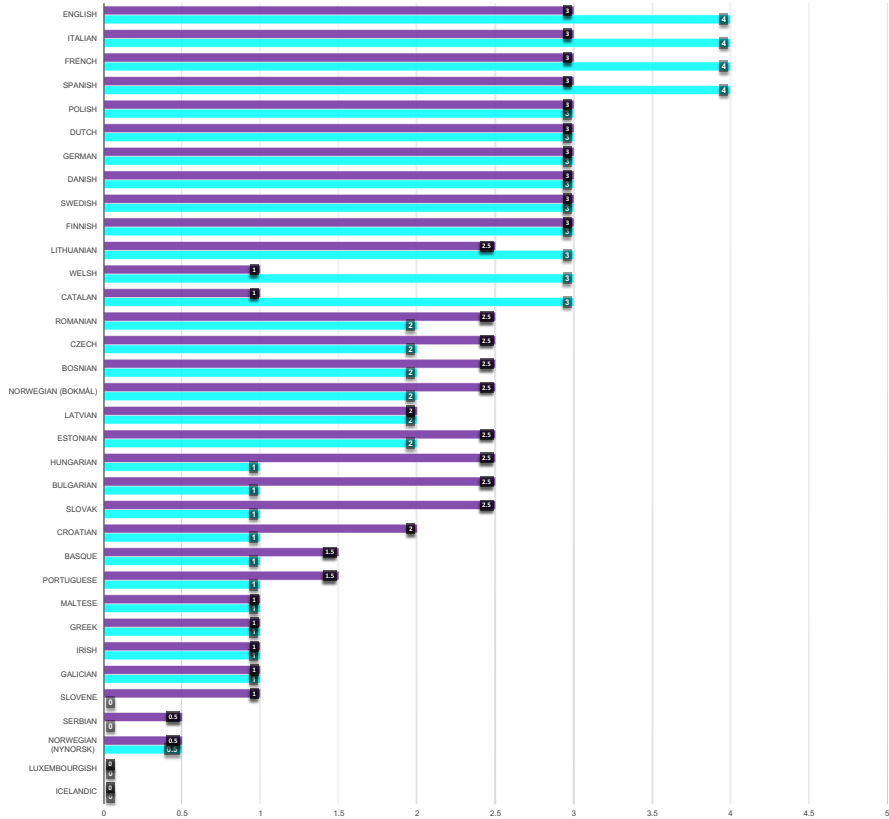


Fig. 9 EU citizens survey – responses to question 6: *Please rate all the types of software applications, apps, tools or devices you use for your language(s). Tools you do not use for your language(s) do not need to be rated.* Note that purple indicates the median and blue the mode.

The list of eight tools presented was: Search apps (e.g., Google, Bing); personal assistant apps (e.g., Siri, Alexa); proofreading apps (e.g., spelling and grammar checkers, autocorrect); translation apps (e.g., Google Translate, DeepL); automatic subtitling (e.g., news report, YouTube); language learning apps (e.g., Babbel, Rosetta Stone); chatbots (e.g., for customer support) and screen readers. The aim of this question was to understand the perception of the average EU citizen and LT user of the quality of the tools that they use for each language they speak.

The ratings were based on a 5-point Likert scale, i.e., respondents had the option of rating 1-star (*poor*) through to 5-stars (*excellent*) for each of the eight tools

presented, and for each language they had selected in the previous question. In the interest of space, Figure 9 presents only the languages for which language reports were produced (see Chapters 5–37) and only shows responses from the perspective of each language, as opposed to each tool. Due to the large size of the dataset and the varying proportion of responses for each language, the figures presented here are based on the calculation of the median score (purple) and the mode (blue). Tools that were not available or used by a respondent did not receive a score. In these instances, the tool was assigned a rating of zero, as a penalty for lesser-used tools across all languages. This explains the low scores for languages such as Serbian, Luxembourgish and Icelandic, which either have very few available or low-rated existing LTs.

To some degree, the results reflect the trend presented for the technological DLE scores of the relevant languages (see Chapter 3) in terms of the quantification of the technological factors of the DLE Metric. The difference between the median score for English and the next well-resourced languages is not as stark, however. This could be explained by the fact that the ratings of the tools are bound to an upper limit of five and as a result, the scores are “flatter” and closer to each other. On the other hand, we can see that the mode score reveals that tools for English, French, Spanish and Italian received more frequent higher ratings. Nevertheless, the results provide a clear insight into the average European user’s perception of the quality of LT support for their languages.

4 Conclusions

We examined around 90 European languages with the goal of creating a snapshot of their digital readiness in 2022. We made use of the inventory of LRTs in the European Language Grid and assessed the technological readiness of each language based on the availability of LRTs. From this, we carried out a cross-language comparison on this empirical basis, as well as an analysis of feedback from developers and users of LTs across Europe, including input from over 20,000 EU citizens.

The status as analysed in 2022 is very clear: there is an extreme imbalance across languages when it comes to the individual levels of technological support. While the META-NET White Paper Series reported a similar imbalance ten years ago, what is surprising is the little comparative change seen across the board since then. The same trend of acute digital inequality continues, and worse still, the gap between English and the rest of the EU languages is getting wider. Even though some of the widely spoken languages in Europe and beyond (Spanish, German and French) have demonstrated considerable progress and are among the top performers, their distance from English is intolerable. Moreover, a striking asymmetry is evidenced between official and non-official EU or EEA languages.

Our results reiterate that digital language *inequality* poses a direct threat to Europe’s linguistic and cultural diversity. Europe has become or is about to become a continent where *digital diglossia* is the de facto context for many EU citizens, with the exception of English native speakers. When going about their online lives, EU

citizens too often find it more efficient or even absolutely necessary to rely on other, more widely supported languages (predominantly English) for certain services and information because this gives them greater access to high-quality and reliable content to a broader audience, and allows them to use more advanced technologies. This is true particularly for the younger generations, thus increasing the generational language gap and bringing lesser-resourced languages ever closer to digital extinction.

References

- 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>.
- Blasi, Damian, Antonios Anastasopoulos, and Graham Neubig (2022). “Systematic Inequalities in Language Technology Performance across the World’s Languages”. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. Dublin, Ireland: Association for Computational Linguistics, pp. 5486–5505. DOI: 10.18653/v1/2022.acl-long.376. <https://aclanthology.org/2022.acl-long.376>.
- 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>.
- Giagkou, Maria, Stelios Piperidis, Georg Rehm, and Jane Dunne, eds. (2022). *ELE Language Report Series*. Project deliverables; European Language Equality (ELE); EU project no. LC-01641480 – 101018166. <https://european-language-equality.eu/deliverables/>.
- 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>.
- Hajić, 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, Katrin Marheinecke, and Georg Rehm (2022). *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>.
- 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>.
- LT-Innovate (2016). *The LT-Innovate Innovation Agenda*. http://www.lt-innovate.org/sites/default/files/2904-LTi_Innovation_Agenda.pdf.
- Joshi, Pratik, Sebastin Santy, Amar Budhiraja, Kalika Bali, and Monojit Choudhury (2020). “The State and Fate of Linguistic Diversity and Inclusion in the NLP World”. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL 2020)*. Online: Association for Computational Linguistics, pp. 6282–6293. DOI: 10.18653/v1/2020.acl-main.560. <https://aclanthology.org/2020.acl-main.560>.

- 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>.
- Labropoulou, Penny, Katerina Gkirtzou, Maria Gavriilidou, Miltos Deligiannis, Dimitris Galanis, Stelios Piperidis, Georg Rehm, Maria Berger, Valérie Mapelli, Michael Rigault, Victoria Aranz, Khalid Choukri, Gerhard Backfried, José Manuel Gómez Pérez, and Andres Garcia-Silva (2020). “Making Metadata Fit for Next Generation Language Technology Platforms: The Metadata Schema of the European Language Grid”. In: *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*. Ed. by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Christopher Cieri, Khalid Choukri, Thierry Declerck, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis. Marseille, France: ELRA, pp. 3421–3430. <https://www.aclweb.org/anthology/2020.lrec-1.420/>.
- Labropoulou, Penny, Stelios Piperidis, Miltos Deligiannis, Leon Voukoutis, Maria Giagkou, Ondřej Košarko, Jan Hajič, and Georg Rehm (2023). “Interoperable Metadata Bridges to the wider Language Technology Ecosystem”. In: *European Language Grid: A Language Technology Platform for Multilingual Europe*. Ed. by Georg Rehm. Cognitive Technologies. Cham, Switzerland: Springer, pp. 107–127.
- Lawlor, Jennifer, Carl Thomas, Andrew T Guhin, Kendra Kenyon, Matthew D Lerner, UCAS Consortium, and Amy Drahotá (2021). “Suspicious and fraudulent online survey participation: Introducing the REAL framework”. In: *Methodological Innovations* 14.3. DOI: 10.1177/20597991211050467. <https://doi.org/10.1177/20597991211050467>.
- Piperidis, Stelios, Penny Labropoulou, Dimitris Galanis, Miltos Deligiannis, and Georg Rehm (2023). “The European Language Grid Platform: Basic Concepts”. In: *European Language Grid: A Language Technology Platform for Multilingual Europe*. Ed. by Georg Rehm. Cham: Springer, pp. 13–36. DOI: 10.1007/978-3-031-17258-8_2. https://doi.org/10.1007/978-3-031-17258-8_2.
- Ranathunga, Surangika and Nisansa de Silva (2022). “Some Languages are More Equal than Others: Probing Deeper into the Linguistic Disparity in the NLP World”. In: *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*. Online only: Association for Computational Linguistics, pp. 823–848. <https://aclanthology.org/2022.aacl-main.62>.
- Rehm, Georg, ed. (2023). *European Language Grid: A Language Technology Platform for Multilingual Europe*. Cognitive Technologies. Cham, Switzerland: Springer.
- Rehm, Georg, Katrin Marheinecke, Rémi Calizzano, and Penny Labropoulou (2023). “Language Technology Companies, Research Organisations and Projects”. In: *European Language Grid: A Language Technology Platform for Multilingual Europe*. Ed. by Georg Rehm. Cognitive Technologies. Cham, Switzerland: Springer, pp. 171–185.
- Rehm, Georg, Katrin Marheinecke, Stefanie Hegele, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Khalid Choukri, Andrejs Vasiljevs, Gerhard Backfried, Christoph Prinz, José Manuel Gómez Pérez, Luc Meertens, Paul Lukowicz, Josef van Genabith, Andrea Lösch, Philipp Slusallek, Morten Irgens, Patrick Gatellier, Joachim Köhler, Laure Le Bars, Dimitra Anastasiou, Alina Auksoiriütė, Núria Bel, António Branco, Gerhard Budin, Walter Daelemans, Koenraad De Smedt, Radovan Garabik, Maria Gavriilidou, Dagmar Gromann, Svetla Koeva, Simon Krek, Cvetana Krstev, Krister Lindén, Bernardo Magnini, Jan Odijk, Maciej Ogrodniczuk, Eiríkur Rögnvaldsson, Mike Rosner, Bolette Pedersen, Inguna Skadina, Marko Tadić, Dan Tufiş, Tamás Váradi, Kadri Vider, Andy Way, and François Yvon (2020). “The European Language Technology Landscape in 2020: Language-Centric and Human-Centric AI for Cross-Cultural Communication in Multilingual Europe”. In: *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*. Ed. by Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Christopher Cieri, Khalid Choukri, Thierry Declerck, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis. Marseille, France: ELRA, pp. 3315–3325. <https://www.aclweb.org/anthology/2020.lrec-1.407/>.

- 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.
- 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>.
- Simons, Gary F., Abbey L. L. Thomas, and Chad K. K. White (2022). “Assessing Digital Language Support on a Global Scale”. In: *Proceedings of the 29th International Conference on Computational Linguistics (COLING 2022)*. Gyeongju, Republic of Korea: International Committee on Computational Linguistics, pp. 4299–4305. <https://aclanthology.org/2022.coling-1.379>.
- Thönissen, 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>.
- Vasiljevs, Andrejs, Khalid Choukri, Luc Meertens, and Stefania Aguzzi (2019). *Final study report on CEF Automated Translation value proposition in the context of the European LT market/ecosystem*. DOI: 10.2759/142151. <https://op.europa.eu/de/publication-detail/-/publication/8494e56d-ef0b-11e9-a32c-01aa75ed71a1/language-en>.
- 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 (2022). *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>.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

