Chapter 39
Results of the Forward-looking Community-wide Consultation

Emma Daly, Jane Dunne, Federico Gaspari, Teresa Lynn, Natalia Resende, Andy Way, Maria Giagkou, Stelios Piperidis, Tereza Vojtěchová, Jan Hajič, Annika Grützner-Zahn, Stefanie Hegele, Katrin Marheinecke, and Georg Rehm

Abstract Within the ELE project three complementary online surveys were designed and implemented to consult the Language Technology (LT) community with regard to the current state of play and the future situation in about 2030 in terms of Digital Language Equality (DLE). While Chapters 4 and 38 provide a general overview of the community consultation methodology and the results with regard to the current situation as of 2022, this chapter summarises the results concerning the future situation in 2030. All of these results have been taken into account for the specification of the project's Strategic Research, Innovation and Implementation Agenda (SRIA) and Roadmap for Achieving Full DLE in Europe by 2030.[1]

1 Introduction

Within ELE three complementary online surveys were designed and implemented in order to consult the Language Technology (LT) community with regard to the current state of play and the future situation in about 2030 in terms of Digital Language Equality (DLE). While Chapter 38 provides a general overview of the community consultation process and methodology and Chapter 4 in Part I gives a brief account

Emma Daly · Jane Dunne · Federico Gaspari · Teresa Lynn · Natalia Resende · Andy Way
Dublin City University, ADAPT Centre, Ireland, emma.daly@adaptcentre.ie, jane.dunne@adaptcentre.ie, federico.gaspari@adaptcentre.ie, teresa.lynn@adaptcentre.ie, natalia.resende@adaptcentre.ie, andy.way@adaptcentre.ie

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

Tereza Vojtěchová · Jan Hajič
Charles University, Czech Republic, vojtechova@ufal.mff.cuni.cz, hajic@ufal.mff.cuni.cz

Annika Grützner-Zahn · Stefanie Hegele · Katrin Marheinecke · Georg Rehm
Deutsches Forschungszentrum für Künstliche Intelligenz GmbH, Germany, annika.gruetzner-zahn@dfki.de, stefanie.hegele@dfki.de, katrin.marheinecke@dfki.de, georg.rehm@dfki.de

[1] This chapter summarises results reported in Way et al. (2022a) and Way et al. (2022b).

G. Rehm, A. Way (eds.), *European Language Equality*, Cognitive Technologies,
https://doi.org/10.1007/978-3-031-28819-7_39

of the results with regard to the current situation in 2022/2023, the present chapter summarises our results concerning the future situation. All of these results have been taken into account for the specification of the project's strategic recommendations (see Chapter 45).

Section 2 summarises the future-looking results with regard to the stakeholder group of European LT developers, introduced in Chapters 4 and 38, whereas Section 3 reports the findings with regard to the stakeholder group of European LT users and consumers. Section 4 describes the findings of the survey in which we reached out to Europe's citizens to gauge their expectations and desires in terms of DLE by 2030 (see Chapter 4, Section 3, p. 84 ff., and Chapter 38, Section 3, p. 231 ff.). Section 5 concludes the chapter.

2 The Perspective of European Language Technology Developers

The survey targeting LT developers and researchers generated a large number of responses between June and October 2021, representing more than 200 different organisations and more than 30 countries. The survey investigated topics like language coverage and evaluation of the current situation but also predictions and visions for the future. Detailed breakdowns of the results can be found in various ELE project reports (Thönnissen 2022; Eskevich and Jong 2022; Rufener and Wacker 2022; Hajič et al. 2022; Hegele et al. 2022). In addition to the survey, expert interviews with selected representatives from initiatives such as, among others, ELG and META-NET were conducted. The interviewees shared details on their work and related challenges, elaborating on how to do justice to all European languages, ways to position European LT on a global level and the key challenges towards establishing a long-term European LT programme.

2.1 Respondents' Profiles

One major goal of this survey was to bring the European LT community together and to reach a wide and demographically distributed audience. In total, the LT developers survey was filled in by 321 different respondents who represent 223 different organisations: 73% of the organisations were research or academic institutions (63% universities, 10% research centres) and 22% were companies (17% SMEs, 5% large enterprises). In 5% of responses the type "other" was indicated, i. e., freelancer, private practitioner, government agency, not-for-profit organisation, etc.

The headquarters of these organisations are located in 32 different countries, covering all EU member states and other European countries, such as the UK, Switzerland, Serbia, etc., but also other global regions, e. g., Brazil, the US and Israel. Most responses were contributed from Spain, Germany, Greece, the Czech Republic, and the Netherlands. The respondents cover a wide spectrum of the targeted groups

of stakeholders, as apparent from the range of networks, associations and relevant projects ongoing at the time the survey was circulated. The most established research networks in LT/AI, i. e., META-NET, CLARIN and CLAIRE are well represented in the survey responses with about 40 to 90 respondents each. ELG, ELE's sister project, is represented with more than 50 participants. Other related projects and networks focusing on LT or on neighbouring fields, such as AI4EU, ELISE, ELEXIS, and Nexus Linguarum are represented with around 10 to 25 survey respondents each (Table 1). Additional networks, associations and projects indicated by the respondents include ELRC, ELRA, ACL, EAMT, DARIAH and others.

Initiative	Responses	Interviews
CLAIRE	37	3
CLARIN	90	4
ELG	54	20
LT-Innovate	18	29
META-NET	61	5
AI4EU	16	–
BDVA	12	–
DIH4AI	1	–
ELEXIS	19	–
ELISE	4	–
HumanE AI	11	–
Nexus Linguarum	25	–
TAILOR	9	–
Other	31	–
None of the above	115	–

Table 1 LT developers survey – survey responses and interviews collected through the participating initiatives

The respondents were mainly active in the following areas: 1. Basic natural language processing services (POS tagging, parsing, named entity recognition etc.), 2. Text analytics and mining, information extraction, text classification, and 3. Language resources (LRs): data production, data aggregation (Figure 1).

The technologies, products or services offered by the respondents' organisations are used in various domains, a finding that demonstrates the applicability of LT in practically all economic sectors. The top three domains indicated by the respondents were 1. Information and communication technologies (ICTs), 2. Digital humanities (DH), arts, culture and other services and 3. Education.

2.2 Language Coverage

The respondents listed a wide range of languages they actively include in their research and development work and for which they offer services, software, resources, models etc. All official EU languages are covered as well as other state official, re-

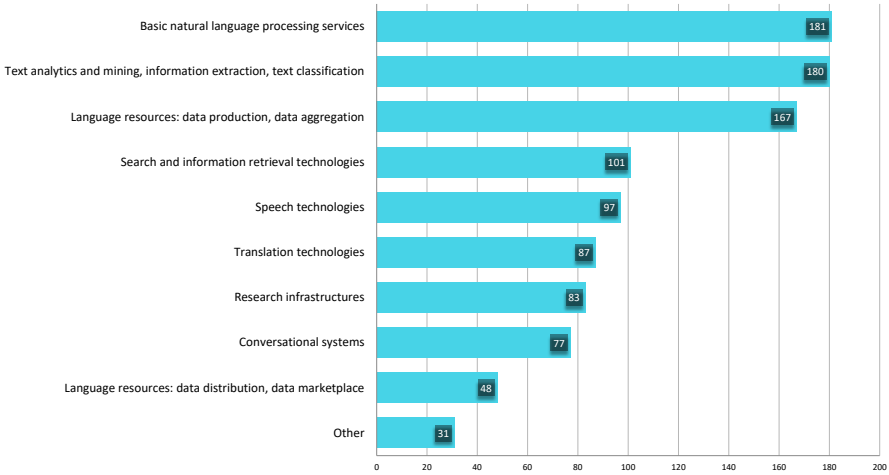

Fig. 1 LT areas in which the respondents conduct research or develop tools and services

gional or co-official European languages (see Figure 6 in Chapter 4, p. 86). The five most frequently mentioned languages are English, German, Spanish, French and Italian. A total of 80 respondents indicated "other" languages they support in their products or research, languages spoken in the Middle East and Asia with Arabic, Chinese, Japanese, Russian and Turkish being the five most frequently mentioned ones. Sign languages were also mentioned.

To get an idea about the focus of future work, the respondents were asked about the languages their organisation does not yet support, but plans to support in the next three years. Apart from some of the big languages, the respondents' future plans additionally include some regional and minority languages (RMLs), such as Basque, Catalan, Breton, Mirandese, Romani or Aromanian. Sign languages were mentioned five times, and it is worth noting the presence of regional and dialectal varieties in the respondents' future plans, e. g., Pontic Greek or Spanish varieties.

When considering the top three drivers for the decision to support additional languages (Table 2), the most frequently selected factor is research interest (212 mentions), followed by the availability of LRs (144) and market interest or demand (138). As expected, the prioritisation of these factors is different when the type of organisation the respondent represents is taken into account. For industry (including large enterprises and SMEs) market interest or demand by users or consumers play a pivotal role, while the availability of LRs follows at a distance. For research organisations and SMEs, more than big organisations, funding and investment opportunities are also to be considered. In terms of "other" reasons, these were often specified with an appeal for equality and the need for preserving all languages in the digital age, as for instance in the following answers: "Need for equality", "Ensure language rights in the digital economy, services, applications", "Supporting under-represented language communities to work towards the knowledge equity goals".

Drivers	Research organisation	Industry	Other	Total
Research or scientific interest	196	12	4	212
Availability of language resources	108	29	7	144
Market interest or demand	65	66	7	138
Available funding or investment	107	18	3	128
Availability of human experts	60	12	3	75
Availability of technologies or tools	44	18	5	67
Other	69	14	4	87

Table 2 LT developers survey – the top drivers for the decision to support additional languages

2.3 Predictions for the Future

We were also interested in the respondents' views on the measures and instruments that are deemed effective as well as the key challenges that a future large-scale ELE programme should address. The participants had the option to rate a number of policies and instruments as either very effective, effective, slightly effective or not effective at all. In addition, respondents were given the opportunity to elaborate on other policies or instruments, which they consider effective in speeding up the development and deployment of LT in Europe equally for all languages. The responses were provided as free text.

A critical aspect of the respondents' visions for DLE, as brought up in multiple answers, is the availability of resources. By 2030 all European languages should have developed the critical mass of resources needed for developing LTs. These include not only raw data, but also large multilingual language models. The issue of data availability is often mentioned in relation to the legal framework for sharing them. Large amounts of data for all languages are expected not only to be available by 2030, but also available for free or at a reasonable cost for research and commercial purposes. Standardised training and evaluation data for all languages are deemed critical. In parallel, according to the survey respondents, LT developers will be working towards automated procedures for the construction, annotation and curation of language data, as well as to address the issue of data bias. Such achievements, combined with continuous work on improving transfer learning methods, are expected to contribute to a situation in which all languages, including small, minority and regional ones, enjoy technology support and a level of presence and use in the digital sphere that will ensure their preservation and prosperity.

A shared scientific goal of the LT community is the achievement of *Deep Natural Language Understanding by 2030*, brought up in numerous responses with various phrasings: "hybrid intelligence", "cognitive AI", "symbolic AI", etc. Nonetheless, all these mentions converge on the description of a future status of LTs where the leap from superficial language *processing* to language *understanding* has been achieved and seamless human-like interaction, viable discourse interpretation and ubiquitous natural language interfaces are a reality for all Europeans in their own language.

With respect to measures and instruments that can be employed to help achieve these goals and realise the visions, the respondents evaluated the effectiveness of a set of proposed measures. A long-term programme of ten or more years can potentially lead to groundbreaking research and subsequently to the desired leap from simple language processing to deep language understanding according to almost all respondents (average score 4.2 on a five-point Likert scale with 5: very effective and 1: not effective at all). Continuous investment in existing research infrastructures (RIs) that support LT was considered equally effective (average score 4.2). Among others, access to data and tools via distributed RIs is argued to allow for optimising both the storage space and processing power, as well as to compare the LTs in terms of their computational footprint.

At the technological level, investing in the development of new scientific methodologies for the transfer or adaptation of resources or technologies to other domains and languages is considered an effective measure to boost the digital readiness of less supported languages (average score 4.0). Given the importance of a strong foundation in basic research, it does not come as a surprise that a large majority of over 86% of respondents welcome an increase in the availability of qualified LT personnel and incentives for talent retention. This also included reinforcing training and education initiatives, including undergraduate and Master's programmes.

A number of elaborate answers focused on funding instruments as leverage to help Europe achieve global excellence and leadership in LT. Funding and investments should concentrate not only on the applied (computational) aspects of LT but also on basic research in linguistics and computational linguistics. Support of LR creation and sharing is an issue in many responses. With respect to the beneficiaries of funding, a number of respondents and interviewees expressed the opinion that incentives should be provided to language communities that strive to preserve their cultural and linguistic identities, especially with regard to enhancing a language's presence on the internet. Businesses and industry-research collaborations are noted as an additional target group.

In this context, some respondents perceive the role of national centres of excellence in LT as critically important. Such centers could collect and boost the voices of local players at a national level and increase industry visibility nationally and at the European level. Apart from designing the national research agendas in LT, they should be responsible for the collection, curation, sharing and standardisation of language data, and for following and implementing the European Data Strategy.

Regulatory aspects pertinent to the LT field, in the form of regulations, recommendations or guidelines, have additionally been highlighted. These include, e. g., the adoption of the FAIR principles in Europe, a revised legislative framework for facilitating the use of language data and the application of data mining techniques for both research and commercial purposes, guidelines for procurement beneficiaries and for public bodies to release their funded or public data, recommendations for big technology companies to open up their platforms for the lesser spoken languages and for the public and private sectors equally to provide multilingual websites. It could be also beneficial to impose content accessibility regulations, e. g., for multimedia subtitling, readability, dubbing, etc.

The role of the research community is often criticised for its bias towards publications on a small number of the world's languages. Raising awareness of equality issues in international LT fora and incentivising Open Access journals and conferences dedicated to less supported languages are among the suggested measures.

Awareness raising of the importance of LT for digital interactions and the role of training young LT professionals is mentioned in numerous responses. Finally, the social dimensions of DLE have been emphasised by respondents who argued that linguistic and social diversity go hand in hand: the more diverse our society is, the more there is an actual need for multilingual resources and technologies. Thus, large-scale policies against racism and discrimination are considered essential. In parallel, engaging minoritised language communities and supporting community building is argued to benefit the LT field, as it will increase demand for and the impact of LT.

European LT should foster and support multilingualism while strictly adhering to European values such as privacy by design, transferability, fairness, diversity and openness, transparency and accountability, public wealth, individual rights and collective purposes. Europe's strengths lie in catering for multilingual solutions covering all the European languages and serving all citizens of Europe. By supporting its linguistic diversity, Europe can achieve digital self-determination and sovereignty.

3 The Perspective of European Language Technology Users

For LT users, a similar survey was set up (see Chapter 4, Section 3, p. 84 ff., and Chapter 38, Section 4, p. 235 ff.) and generated almost 250 responses. Similarly to the LT developers survey, numerous additional interviews were conducted for more in-depth insights.

The survey brought together diverse groups of stakeholders including representatives of communities of LT users, academic and commercial stakeholders, language professionals (e. g., translators, lecturers and professors in the fields of linguistics and computational linguistics) and stakeholders from different economic sectors (e. g., banking, health, public administration, language services). The survey was disseminated mainly via email by the relevant ELE partners, namely, ELEN, LIBER, ECSPM, NEM, EFNIL and Wikipedia as well as through social networks. Table 3 shows the breakdown of responses collected through the survey.

3.1 Respondents' Profiles

Responses came from a diverse range of sectors and professional activities; most of the respondents work 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 non-governmental organisations (NGOs), large en-

Initiative	Responses	Interviews
ECSPM	10	2
EFNIL	28	6
ELEN	7	19
LIBER	29	3
NEM	29	6
Wikipedia	22	3
Other (e. g., social media)	121	–
Total	246	39

Table 3 LT users survey – survey responses and interviews collected through the participating initiatives

terprises, 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 (see Figure 2).

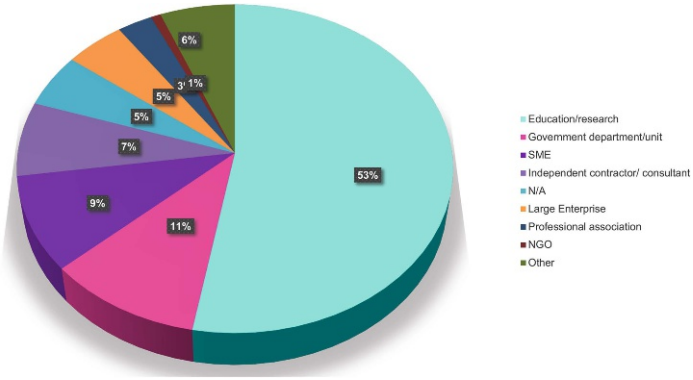

Fig. 2 LT users survey – types of sectors and professional activities

Contributions to the survey came from all over Europe and, due to social media sharing, some responses were provided by people based outside European countries such as the US, the Democratic Republic of Congo and the Russian Federation. In Europe, the most represented countries were Croatia (33 responses), Spain (23 responses), the UK (23 responses), Ireland (17 responses), Germany (16 responses) and France (14 responses).

3.2 Language Coverage

A total of 74% of the respondents indicated that they work with English, which is the dominant language followed by a well-balanced group of languages composed of German (31%), French (31%) and Spanish (30%). At the other end of the spectrum, many other European languages (e. g., Welsh, Catalan, Basque, Luxembourgish, Galician) are under-represented as few respondents (between one and three) indicated they work with them. Respondents who selected "other", mentioned that they work with Basque, Catalan, Macedonian, Luxembourgish, Moldovan, Welsh and Galician. Among the non-European languages respondents mentioned Japanese, Chinese (or Mandarin) and Russian. Figure 3 shows the breakdown of European languages the respondents work with in absolute numbers.

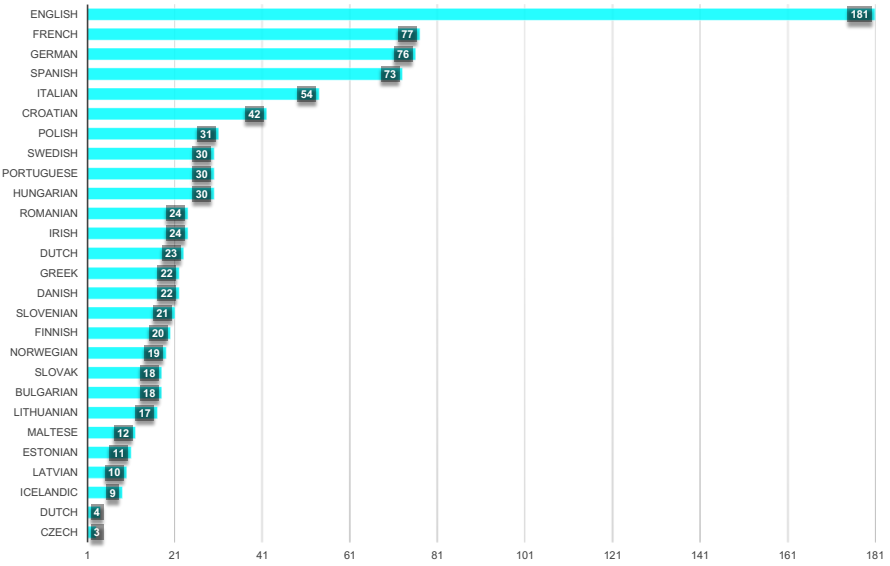

Fig. 3 LT users survey – European languages respondents work with (based on a set of 246 responses)

In relation to the languages respondents intend to include in their workflow, 50 respondents (20%) indicated that they plan to include English, German, Spanish and French. The survey shows, again, the English predominance over all languages followed by German, Spanish and French. Other official EU languages were mentioned by only a few respondents (between two and three respondents only) such as Italian, Portuguese and Greek as well as some minority, regional, and lesser-used languages such as Breton, Catalan, Faroese but only by one respondent each. These findings suggest a worrying scenario, where, in a multilingual and multicultural Europe, most minority, regional, lesser-used languages are disregarded either for not being commercially interesting or simply for lack of institutional investment.

3.3 Predictions for the Future

With regard to their predictions for the future, the range of opinions was very broad. In general, most respondents (68%) are confident that in the next ten years, there will be higher-quality tools for all European languages including minority, regional, and lesser-used languages and that there will also be a wider range of tools for all European languages (83%). However, fewer respondents (46%) believe that LTs will help to prevent linguistic loss, although 65% think that LTs can help to prevent RMLs from disappearing. Most respondents (64%) also agree that LTs can increase individuals' exposure to these languages and 60% believe that LTs can increase engagement with social, leisure and work activities in their own languages. Among other benefits mentioned in the open questions, respondents think that LTs can improve medical interactions between patients and clinicians and improve medical documentation. One respondent highlighted that LTs can help with the preservation of cultural heritage and improve its visibility. Another respondent pointed out that LTs can improve online and print publishing in minority, regional, and lesser-used languages, including academic publications and works of fiction.

The survey also looked into the respondents' ideas for the future of LT. They had the chance to indicate applications that could potentially use LT they want to see that are not currently available for the languages they work with. There were several interesting responses. In general, we can see respondents wish for higher-quality tools for certain languages such as "better parsing of Danish than currently available" or the availability of tools that do not yet exist for some languages but exist for others such as "speech recognition for Welsh", "speech recognition for Catalan", "free spell check for Irish", "more reliable speech recognition, information extraction, summarisation, semantic parsing and semantic search for Greek", "a good Georgian-English Translator" and "better MT for Croatian". Other respondents indicated that they would like to see some of the existing tools and technologies available in more languages, for instance, "Text-To-Speech for low resource languages" or "more accurate speech2text, decent text summarization, GPT2 for Finnish".

Some ideas for new (currently non-existent) LTs were also provided. For instance, "case-sensitive tools or the creation of a tool that might provide more context, or warn the user if the same word means something completely different depending on the context. A tool that would be sensitive to connotative meanings" or "tools for collecting lexical data and speed up the process of dictionary building".

We can conclude that the most important finding of this survey is the respondents' concern regarding the differences in technological support between European languages, specifically the poor technological support of minority, regional and lesser-used languages. The differences in support are mainly reflected in differences in the quality and performance of tools between the languages as well as in the availability of tools for a small group of low-resource languages, while these same tools do not exist for many other European languages. In order to achieve full DLE as a crucial step to maintain linguistic diversity, the survey shows the necessity for action and an implementation agenda with the objective of fostering and supporting a multilingual and linguistically inclusive Europe that brings solutions to all European citizens.

4 The Perspective of Europe's Citizens as Consumers of LTs

The ELE project has made an effort to ensure that all voices were heard and taken into account in the preparation of the SRIA. With the support of social media campaigns and an agency specialising in survey dissemination, we were able to reach thousands of EU citizens to hear their thoughts on how well they feel their languages are digitally supported. The European Citizen survey included a total of 11 questions, six multiple-choice questions, four single-choice questions and one open-ended question which allowed respondents to include any comments or feedback they had. The survey was designed to take less than five minutes to fill in (see Chapter 4, Section 3, p. 84 ff., and Chapter 38, Section 4, p. 235 ff.). It was translated into 35 languages. To ensure the reliability of the survey data captured, a number of data cleaning steps were taken to remove responses that were deemed noisy or at risk of skewing the survey results. We analysed a total number of 20,586 valid responses, the largest public survey ever conducted to date among European citizens concerning LRTs.

4.1 Respondents' Profiles

We collected (anonymous) demographic information from respondents with the objective to ensure our sample was representative enough of the population for generalisation purposes. We asked respondents to state their level of education, age group and country of residence. We collected responses from 28 countries, and Figure 4 shows the breakdown of contributions per country.

The demographic of the respondents is as follows: 27% of the respondents were between 25-34 years old. A total of 23% accounted for both the 18-24 and 35-44 age brackets. The rest of the respondents were 45+ years old, 1% of the respondents preferred not to say. In terms of education, 35% of the respondents had reached high school level, 23% held a Bachelor's Degree, 17% held a Master's Degree, with the rest reporting vocational training (11%), only some high school completion (7%) and holding a PhD (5%), 2% declined to say.

4.2 Language Coverage

We asked respondents to select the languages they use both socially and professionally. Overall, results show that many respondents use their native language in addition to English even if they are not based in English-speaking countries. Therefore, we once again see a dominance of English over all other languages. Following English, German and French also appear as languages frequently used in non-German or non-French speaking countries. Figure 5 illustrates the comparison of the most represented languages in the survey.

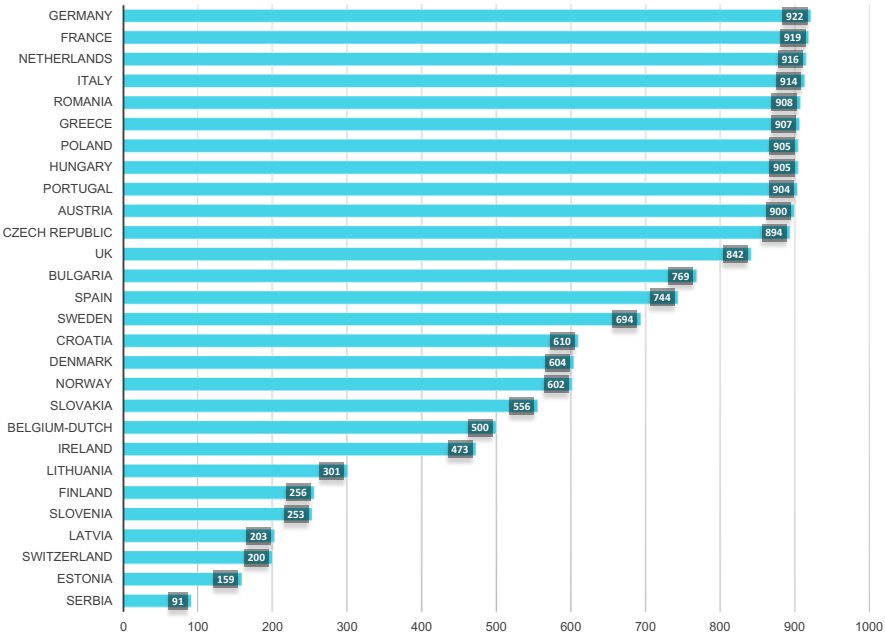

Fig. 4 European citizens survey – number of responses collected

4.3 Predictions for the Future

The following discussion concentrates on the forward-looking questions of the EU citizens survey and the responses concerning anticipated or hoped for future developments with regard to the development and consolidation of LTs for Europe's languages. In one question we asked the respondents "What would be the top 3 advantages of improving apps and tools for all languages? Please select the three most important advantages in your opinion." The purpose of this question was to assess respondents' views on the benefits of LTs. Notably, as seen from Figure 6, LTs are regarded as key to enhancing multilingual societies from a linguistic diversity perspective. Of seemingly less importance to the average citizen is the economic advantage that arises from LT support.

With regard to the question "What holds you back from using some of these apps or tools in your languages?", based on the answers received, it is reasonable to assume that if the reported barriers that are currently holding users back from using apps or tools in their languages were removed, and tools more adequately supported, then there would be more uptake in the number of people using language tools in their own preferred language (see Figure 7). It was somewhat surprising that the top response was "I don't need to use any apps or tools for this language", which might suggest that the poor support for some languages may condition users into believing that technologies do not apply to some chronically underserved languages. This

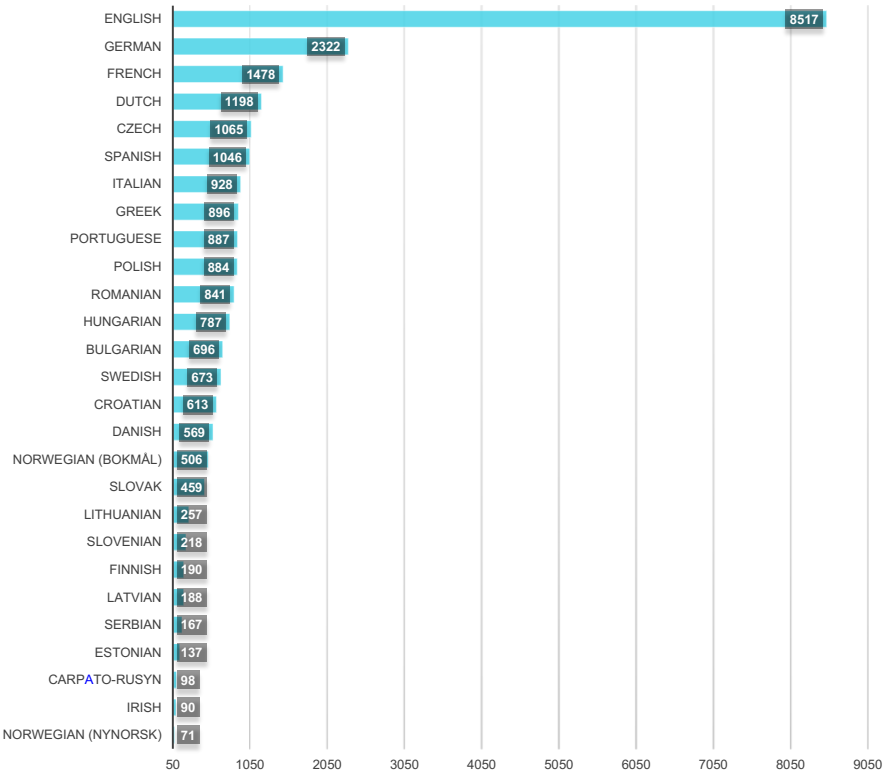

Fig. 5 European citizens survey – most represented languages

may apply in particular to users who also speak a dominant language that is well supported by tools and apps, in addition to one that is scarcely supported.

In other words, these responses suggest that there is a real risk that some users have become so accustomed to using apps in or for better supported languages that they no longer see the need for similar apps to be developed and made available in or for their own language; at the same time, this disappointing perception may stabilise a situation where users default to using apps and tools in an additional language that is better supported, also due to their overall superior quality. Another popular response was "Issues with the quality of the available apps or tools", indicating that people will not use an app or tool if they perceive its quality to be insufficient or inadequate. This suggests that once the quality of the tools is improved to a sufficient standard, more people would be inclined to use the app or tool in their language in the future.

Concerning the query "Please select the tools that you currently do not use but would like to use in the future.", one tool that people are calling for in particular among those to be made available for their languages is automatic subtitling (Figure 8). Having this available for more languages would improve communication

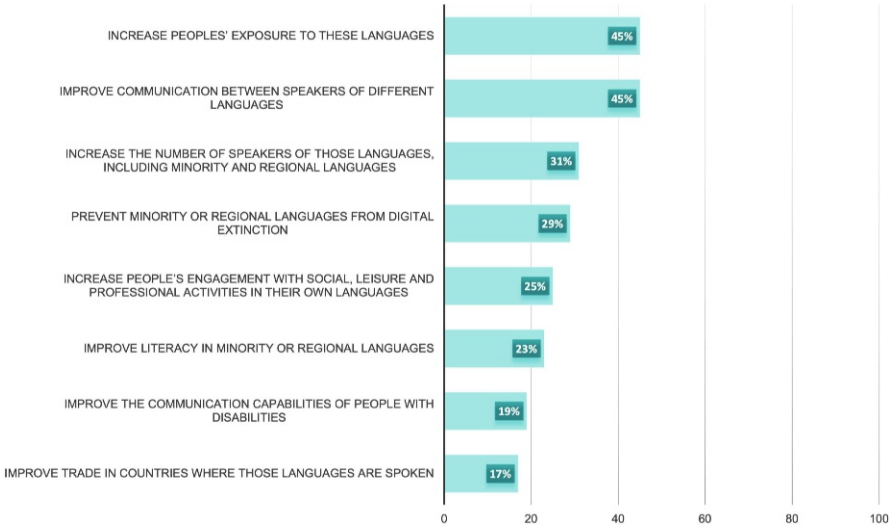

Fig. 6 Responses to the question "What would be the top 3 advantages of improving apps and tools for all languages?" in the EU citizen survey

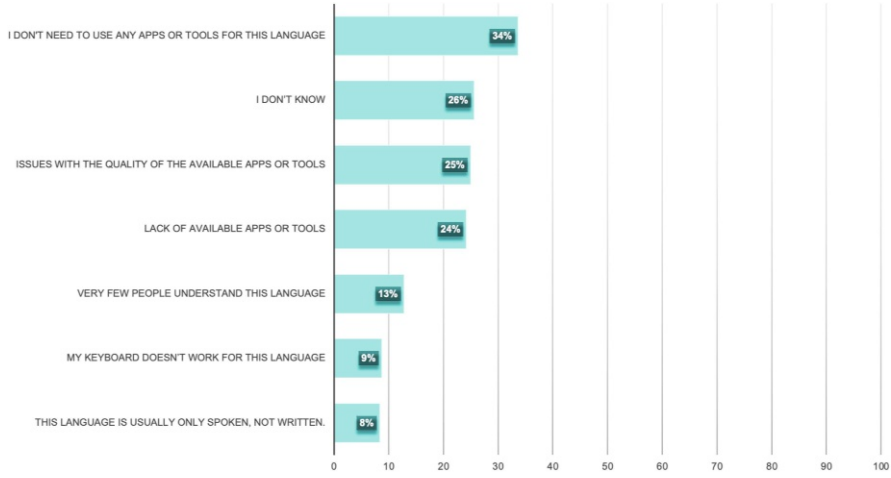

Fig. 7 Responses to the question "What holds you back from using some of these apps or tools in your languages?" in the EU citizen survey

and accessibility of multimedia content for an ever-increasing range of European citizens (e. g., disabled people, elderly users, etc.). Relevant examples include automatic subtitles being made available to those who are hearing-impaired, so they can watch videos and read subtitles in their own language. Translation apps are also in very high demand, which is not particularly surprising. However, even for those language-pairs that are serviced by MT, we need to be vigilant as many of the freely

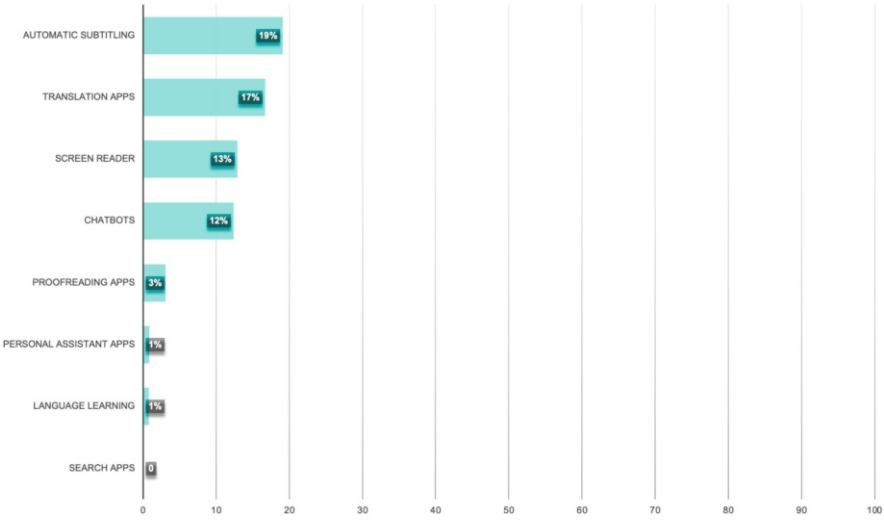

Fig. 8 Responses to the question "Please select the tools that you currently do not use but would like to use in the future." in the EU citizen survey

available translation tools are not owned or resourced by EU companies. Screen readers are another tool that is quite popular, with obvious relevance to visually impaired people. If screen readers were available in more languages, accessibility would be substantially increased for several language communities across Europe.

Finally, in the analysis of the responses to the survey, a number of interesting comments made by ordinary EU citizens were found in the section that elicited more general reactions at the end of the questionnaire. In particular, the very last question of the survey asked the participants to enter any comments they had about the survey or LTs in general. Here follows a selection of the most insightful comments that we feel encapsulate some of the most relevant opinions on the matter.

- "No language is inferior to others. All languages are worthy of survival as long as there is at least one person who speaks that language."
- "Usually I google things in English because more information is available in English."
- "It is extremely important to have more language technology tools for the national minority languages in Sweden. It is a rights issue to access everything from speech synthesis, machine translation, language apps, proofing programs, etc. At the moment, there are no opportunities for this for Roma, Meänkieli and to some extent for Sami and Yiddish."
- "It would be great to have a little more guidance on what ordinary people (without great technological resources such as universities and companies) can do to 'feed' or develop those technological resources for our minority languages."

These comments clearly indicate that some European citizens are eager to have more LT tools and apps made available to them in their language in the future, as

this is related to the role that individual speakers and their communities can play going forward in the digital age in the interest of equality. At the moment many people seem to be resorting to using search apps and personal assistants particularly in English or other well-resourced languages, as they are currently unavailable in their own language or are not perceived to perform equally well. This suggests that if required LTs were developed and made available as tools or apps, people would use them in their own language rather than English; at the very least they would have a choice, depending on the type of tasks that they need to perform in different circumstances (e. g., for professional purposes as opposed to personal or social reasons, with colleagues, within the family or with circles of friends and acquaintances, etc.).

The survey also revealed that some European citizens want to see technology for their languages improved and maintained, and some are willing to get involved themselves, as shown by the comment asking what the ordinary citizen can do to help the development of these much-needed technologies. Overall, citizens are concerned about the technological status of their language, and are willing to help to ensure that their language is technologically well supported in the future for the digital age, especially if otherwise there is a threat of extinction. We were particularly pleased at respondents' willingness to take ownership of these issues, and act not only as users of tools but also as developers. We take this as a strong endorsement of the ELE project, and further evidence of the need for the ELE programme to be fully funded throughout Europe to ensure DLE for all Europeans, as reflected in the ELE SRIA.

5 Summary and Conclusions

The surveys and expert interviews discussed here targeted LT developers, users and the EU citizens. We investigated language coverage and encouraged participants to share their predictions and visions for the future of LTs in Europe with respect to achieving full DLE. The results show that there is still a huge gap between the LT support for English and all other European languages, with dramatic differences in several cases. Even though there is an increased interest in bridging this gap and in expanding technological support to more languages, limited funding, demand and obstacles with regard to available resources make it a challenging endeavour. While basic research is still urgently needed, the last decade has seen progress on a larger scale than could have been imagined ten years ago. Many experts highlight European excellence, also on a global level and consider leadership in LT and language-centric AI to be possible if the necessary conditions are created by political decision-makers.

The LT developers survey addressed the European LT community, reaching a wide and demographically distributed audience. It was answered by 321 respondents who represent 223 organisations in 32 countries. The respondents were recruited by the research networks, i. e., META-NET, CLARIN and CLAIRE, projects like ELG and other related initiatives focusing on LT or neighbouring fields, such as ELISE, ELEXIS, and Nexus Linguarum. Additional networks, associations and projects represented by the respondents include ELRC, ELRA, ACL, EAMT, DARIAH and oth-

ers. The areas in which the respondents are active covered the full range of LT. The languages they focus on have a skewed distribution that reflects current imbalances in the field in Europe as well as elsewhere, with English first by a large margin, followed by the big official EU languages. The two main concerns expressed were the insufficient support for basic research in NLP and LT and the fierce competition of non-EU companies with the market disruption they cause. The survey answers to the open-ended questions and views of the interviewed experts brought a host of opinions and suggestions in several important directions, in particular: the higher and even elementary education area, research funding, legal and regulatory obstacles, biases and privacy issues of various types, commercialisation difficulties and ways of supporting such efforts, the need to coordinate efforts between national centres of excellence vs. pan-European ones, etc.

The LT users and consumers survey brought together academic and commercial stakeholders, language professionals and stakeholders from different sectors. It was disseminated by the relevant ELE partners, i. e., ELEN, LIBER, ECSPM, NEM, EFNIL and Wikipedia who promoted the survey targeting representatives of organisations and communities of users and consumers. Based on the results, it can be concluded that the most important finding is the respondents' concern regarding the differences in technological support between Europe's languages, specifically the poor technological support of minority, regional and lesser-used languages. The differences in support are mainly reflected in differences in the quality and performance of tools between the languages as well as in the availability of tools for a small group of languages, while these same tools do not exist for many other European languages. To achieve full DLE as a step to maintain and promote linguistic diversity, the survey shows the necessity for action and calls for an implementation agenda with the objective of fostering and supporting a multilingual and linguistically inclusive Europe that brings solutions to all European citizens that are relevant in the digital age.

An additional survey was carried out targeting EU citizens with the aim of taking into account their opinions, individual needs, wishes, general demands and, importantly, to make sure that their voices play a decisive role in the pursuit of full DLE supported by LT. The survey was disseminated in 28 countries with the help of a service provider. Additional dissemination was carried out with the help of ELE partners who promoted the survey on social media, within their networks and through the ELE project website. While structured very differently than the stakeholder group surveys, there are several similarities not only in terms of the scope of the analysis, but also of the key results that were obtained: languages other than English are poorly supported (with only a few exceptions) – something evident even from the distribution of languages that the respondents considered in their responses. These answers show that raising awareness for the LT potential in Europe on a political and institutional level is more important now than ever before. The European LT community is in a position where change is needed in order to compete with innovative systems and tools built elsewhere. On a political level, this involves more commitment from the European institutions as well as those of the Member States.

References

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.

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, 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.

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.

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.

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.