EUROPEAN LANGUAGE EQUALITY

D2.10

Report from LIBER

Author	Oliver Blake
Dissemination level	Public
Date	28-02-2022

About this document

Project	European Language Equality (ELE)
Grant agreement no.	LC-01641480 – 101018166 ELE
Coordinator	Prof. Dr. Andy Way (DCU)
Co-coordinator	Prof. Dr. Georg Rehm (DFKI)
Start date, duration	01-01-2021, 18 months
Deliverable number	D2.10
Deliverable title	Report from LIBER
Type	Report
Number of pages	42
Status and version	Final
Dissemination level	Public
Date of delivery	Contractual: 28-02-2022 – Actual: 28-02-2022
Work package	WP2: European Language Equality – The Future Situation in 2030
Task	Task 2.2 The perspective of European LT users and consumers
Author	Oliver Blake
Reviewers	Natalia Resende, Jaroslava Hlavacova
EC project officers	Susan Fraser, Miklos Druskoczi
Contact	European Language Equality (ELE)
	ADAPT Centre, Dublin City University
	Glasnevin, Dublin 9, Ireland
	Prof. Dr. Andy Way – andy.way@adaptcentre.ie
	European Language Equality (ELE)
	DFKI GmbH
	Alt-Moabit 91c, 10559 Berlin, Germany
	Prof. Dr. Georg Rehm – georg.rehm@dfki.de
	http://www.european-language-equality.eu
	© 2022 ELE Consortium

Consortium

1	Dublin City University (Coordinator)	DCU	IE
2	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (Co-coordinator)	DFKI	DE
3	Univerzita Karlova (Charles University)	CUNI	CZ
4	Athina-Erevnitiko Kentro Kainotomias Stis Technologies Tis Pliroforias, Ton Epikoinonion Kai Tis Gnosis	ILSP	GR
5	Universidad Del Pais Vasco/ Euskal Herriko Unibertsitatea (University of the Basque Country)	UPV/EHU	ES
6	CROSSLANG NV	CRSLNG	BE
7	European Federation of National Institutes for Language	EFNIL	LU
8	Réseau européen pour l'égalité des langues (European Language Equality Network)	ELEN	FR
9	European Civil Society Platform for Multilingualism	ECSPM	DK
10	CLARIN ERIC – Common Language Resources and Technology Infrastructure as a European Research Infrastructure Consortium	CLARIN	NL
11	Universiteit Leiden (University of Leiden)	ULEI	NL
12	Eurescom (European Institute for Research and Strategic Studies in Telecommunications GmbH)	ERSCM	DE
13	Stichting LIBER (Association of European Research Libraries)	LIBER	NL
14	Wikimedia Deutschland (Gesellschaft zur Förderung freien Wissens e. V.)	WMD	DE
15	Tilde SIA	TILDE	LV
16	Evaluations and Language Resources Distribution Agency	ELDA	FR
17	Expert System Iberia SL	EXPSYS	ES
18	HENSOLDT Analytics GmbH	HENS	AT
19	Xcelerator Machine Translations Ltd. (KantanMT)	KNTN	IE
20	PANGEANIC-B. I. Europa SLU	PAN	ES
21	Semantic Web Company GmbH	SWC	AT
22	SIRMA AI EAD (Ontotext)	ONTO	BG
23	SAP SE	SAP	DE
24	Universität Wien (University of Vienna)	UVIE	AT
25	Universiteit Antwerpen (University of Antwerp)	UANTW	BE
26	Institute for Bulgarian Language "Prof. Lyubomir Andreychin"	IBL	BG
27	Sveučilište u Zagrebu Filozofski fakultet (Univ. of Zagreb, Faculty of Hum. and Social Sciences)	FFZG	HR
28	Københavns Universitet (University of Copenhagen)	UCPH	DK
29	Tartu Ulikool (University of Tartu)	UTART	EE
30	Helsingin Yliopisto (University of Helsinki)	UHEL	FI
31	Centre National de la Recherche Scientifique	CNRS	FR
32	Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)	NYTK	HU
33	Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)	SAM	IS
34	Fondazione Bruno Kessler	FBK	IT
35	Latvijas Universitātes Matemātikas un Informātikas institūts (Institute of Mathematics and Computer Science, University of Latvia)	IMCS	LV
36	Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)	LKI	LT
37	Luxembourg Institute of Science and Technology	LIST	LU
38	Università ta Malta (University of Malta)	UM	MT
39	Stichting Instituut voor de Nederlandse Taal (Dutch Language Institute)	INT	NL
40	Språkrådet (Language Council of Norway)	LCNOR	NO
41	Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)	IPIPAN	PL
42	Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)	FCULisbon	PT
43	Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)	ICIA	RO
44	University of Cyprus, French and European Studies	UCY	CY
45	Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)	JULS	SK
46	Institut Jožef Stefan (Jozef Stefan Institute)	JSI	SI
47	Centro Nacional de Supercomputación (Barcelona Supercomputing Center)	BSC	ES
48	Kungliga Tekniska högskolan (Royal Institute of Technology)	KTH	SE
49	Universität Zürich (University of Zurich)	UZH	CH
50	University of Sheffield	USFD	UK
51	Universidad de Vigo (University of Vigo)	UVIGO	ES
52	Bangor University	BNGR	UK

Contents

List of Figures

List of Tables

List of Acronyms

AI	Artificial Intelligence
CLARIN ERIC	Common Language Resources and Technology European Research Infrastructure Consortium
CNRS	The French National Centre for Scientific Research
ELE	European Language Equality *(this project)*
ELE Programme	European Language Equality Programme *(the long-term, large-scale funding programme specified by the ELE project)*
EU	European Union
euRobotics	euRobotics AISBL
LIBER	Association of European Research Libraries
LT	Language Technology/Technologies
NLP	Natural Language Processing
SRIA	Strategic Research and Innovation Agenda

Abstract

In this document, the views of language technology users and consumers from within LIBER's network of research librarians are presented, be that librarians themselves, library staff, or researchers associated with libraries. These opinions were collected through three channels: a survey, a workshop, and semi-structured interviews. Here, each methods is described and the findings are presented. Through this analysis, we show how different languages are used and to what degree, and that language technologies are not widespread and are an emerging issue for libraries. A desire to incorporate new languages into workflows is high within a sector that is multilingual in its work. Across all the data sources, there is an optimism around the future of language technologies, in their ability to preserve languages and become increasingly effective. We show that those who are using them, or are beginning to engage with them, see potential for their place within scholarly communication, an issue of specific relevance to those associated with research libraries. Language technologies have potential for publishing academic research with machine translation of articles helping to support a range of languages and increasing the visibility of research, no matter the language it is published in, which ultimately would help further the goals and aims of the Open Science movement. However, the best way to do this, the processes, the technologies, and the skills, are still severely lacking and the path to do this remains unclear.

1. Introduction

This document reports on the findings of a consultation with representatives from the Language Technology (LT) users community, conducted by the EU project European Language Equality (ELE). Its results will serve as an input for a strategic research, innovation and deployment agenda (SRIA) and roadmap, in order to tackle the striking imbalance between European languages in terms of the support they receive through LTs by 2030.

The ELE project sought to collect the views of European LT users and to consolidate their perspective on the differences in terms of technologies for the languages they work with and of the measures that need to be put in place so that all European languages are equally supported through technology by 2030.

Due to the interdisciplinary nature of the field of Language Technology, which stands at the intersection of Linguistics, Computational Linguistics, Computer Science and Artificial Intelligence, the ELE project brings together diverse groups of stakeholders including researchers, representatives of communities of LT users, language professionals (e. g., translators, lecturers and professors in the field of Linguistics and Computational Linguistics) and stakeholders from different economic sectors (e. g., banking, health).

Although the methodology and instruments utilised have been common to all ELE consortium members, this report covers and analyses the subset of responses of stakeholders contacted by LIBER, the Association of European Research Libraries.

About LIBER

LIBER is Europe's largest association of research libraries, consisting of over 400 national, university and other libraries from more than forty countries. It was founded as an association in 1971 under the auspices of the Council of Europe and became a Foundation under Dutch law in 2009. LIBER brings together European research libraries to support a functional network across national boundaries to ensure the preservation of European cultural heritage, to improve access to collections in European research libraries, and to provide more efficient information services. Enabling Open Science is a major priority for LIBER, as it

is promoting innovative scholarly communication, fostering digital skills and services, and engaging with world-class research infrastructures and e-infrastructures.

The LIBER 2018-2022 strategy "Powering Sustainable Knowledge in the Digital Age" addresses these areas with commitments to working groups, international project partnerships, policy development and coordination, lobbying and advocacy at the EC level, and supporting leadership and digital skills. We are a leading voice in Europe on copyright and research data management (RDM) and on Open Science, and hold a position on the European Open Science Policy Platform. We have strong visibility at the European political level, in the European Parliament and in the European Commission. Our Strategy is executed thanks to hundreds of volunteers who serve on our Executive Board, Steering Committees, Working Groups, and our journal LIBER Quarterly, as well as through our participation in international projects. LIBER has three Steering Committees on a) Innovative Scholarly Communication, b) Digital Skills & Services and c) Research Infrastructure.

2. Methodology and Instruments

2.1. Online Survey

The survey addressed to LT users sought to elicit the respondents' views in a way that facilitates the analysis, consolidation and integration of the collected feedback into the ELE SRIA and roadmap. It had 63 questions in total. Some of the questions depend on previous answers. As a result, a respondent was presented with 30 (minimum) to 63 (maximum) questions, including the "if other" questions. 46 questions were mandatory from which 33 were closed questions (single or multiple choice). Table 1 shows an overview of the types of questions.

Question types	Mandatory	Optional	Totals
Closed	20	13	33
Open-ended	26	4	30
Totals	46	17	63

Table 1: Type of survey questions

The survey was structured in four main parts. If any of the provided answers were not applicable, the respondents had the option to enter a different answer through the option "if other, please specify".

- **Part A. Respondents' profile:** The first part of the survey included 13 questions for the demographic profile of respondents with emphasis on characteristics relevant to the task at hand, i.e.,
    - Country respondents are based in
    - Name of the organisation/representative body respondents work for
    - Communities they represent (if applicable)
    - Type of organisation respondents work for
    - Sectors or domains that respondents are active in (if applicable)
    - Role of respondents in the organisation (if applicable)
    - Organisation's estimated revenue (if applicable)

- **Part B. Language coverage:** looked into the European languages the respondents work with and the languages they intend to include in their workflow, i. e.,
  - Languages the organisations, associations, communities, professionals LT users work with
  - Languages planned to be supported in the short- or medium-term

- **Part C. Evaluation of current situation:** assessed the current situation by asking respondents to evaluate the level of technology support for the official European languages they work with and any minority, regional or lesser used language, i. e.,
  - Differences in availability of LTs for the official European languages they work with and, if applicable, differences in availability of LTs for the minority, regional or lesser-used languages they work with;
  - Gaps perceived in the technologies, tools or applications respondents work with, especially in relation to specific languages;
  - Respondents' opinion in relation to performance of LTs with regard to specific languages

- **Part D. Predictions and visions for the future:** Respondents are requested to share their needs and wishes for the future of language technologies, i. e.,
  - Policies or instruments that could contribute to speed up the effective deployment of LT in Europe equally for all languages
  - Prediction of future opportunities for LT in basic and applied research (scientific vision) and in innovation and the industry
  - Expectations of the community with regard to the challenges an ELE Programme can address by 2030

Follow-up: The last three questions requested the respondent's permission to be contacted for an interview and, given an affirmative answer, their contact details. Respondents were also requested to click on a confirmation question stating "By clicking on 'Submit', I agree that my personal data (email address and/or name) can be used according to the Privacy Policy of the European Language Equality (ELE) project".

The survey was designed, set up and published on the EU Survey platform.[1] The full survey, as published online, is presented in Appendix A (p. 15 ff.).

The survey was distributed by LIBER through emails to members of their network, through Working Groups, and social media channels. Additionally, it was advertised through the European Language Equality and European Language Technology websites, LinkedIn and Twitter.

The survey was opened on 21 June 2021 and closed on 18 October 2021. In total, 246 responses have been collected, out of which 29 from respondents contacted by LIBER. This subset of responses, representing the views of the stakeholders reached by LIBER is analysed in this report.

2.2. Interviews

Potential interview participants were sourced from the survey and a workshop organised by LIBER (Section 2.2.1). First, those who had replied to question 29 of the survey (see Appendix A, indicating interest in follow up contact about the project and the topic, were contacted. Of the eight people who responded *yes* to this question, all were contacted, and interviews were arranged with two persons. Second, multiple workshop participants were

---

[1]   https://ec.europa.eu/eusurvey/runner/LTusers-consumers

contacted for follow up-interviews. Of the three contacted, interviews were arranged with one person.

The three participants who agreed to take part in an interview are as follows:

- Will Lamb – Senior Lecturer, Celtic & Scottish Studies, University of Edinburgh (Survey)

- Ranka Stanković – Chair for Mathematics and Informatics, University of Belgrade (Survey)

- Simon Worthington – Researcher, Open Science Lab, German National Library of Science and Technology (Workshop)

The interviews were conducted using video conference software and lasted between 20 and 40 minutes. Each interview was recorded, with consent from the participant, to allow for the information to be accurately incorporated into this deliverable. Interviews took place between 22nd November 2021 and 3rd December 2021. All followed a similar structure, with slight variations between those who had taken part in the survey and those from the workshop. The questions were open-ended to garner detailed responses, in the participant's own words.

2.2.1. Workshop

On the basis of the information collected from the survey, a LIBER workshop was organised to facilitate a discussion on language technology within the library sector. The intention was threefold: (1) increase awareness of the ELE project; (2) collect input for this deliverable through audience discussion and comments; and (3) generate a bigger sample for follow-up interviews.

The workshop was advertised on LIBER's social media channels – Twitter, Facebook, and LinkedIn – and promoted amongst LIBER Working Groups focused on data science and digital humanities via targeted emails. The purpose was to attract an audience specifically of research librarians to guarantee that the discussion and feedback were relevant for this deliverable. This ensured we could collect more targeted input, compared to the survey.

The workshop lasted one hour and thirty minutes and took place on 18th November (2021)[2]. The workshop was moderated by Maria Eskevich, Central Office Coordinator at CLARIN ERIC, an expert on language and speech technologies. She was joined by two speakers: German Rigau, member of the Ixa research group and deputy director of HiTZ, and Jean-François Nominé, translator Institute for scientific and technological information (Inist). Inist is a research support unit of the French National Scientific Research Centre (CNRS). It must be noted that Maria Eskevich and German Rigau work for institutions which are members of the ELE project consortium while Jean-François Nominé has no specific ties to the ELE project.

In total 44 people registered for the online workshop. When the workshop began, 12 participants made up the workshop's audience. The final discussion, after the speakers' presentations, lasted 20 minutes and, throughout the session, the participants wrote notes on a collaborative Google Document to share their thoughts and reflections in real time. This document has been used to build the content of this deliverable, as well as comments and themes within the discussion, outlined in the next section. Furthermore, the workshop was recorded to capture information from the discussion for the purpose of this deliverable.

---

[2] To read the report of the workshop, watch the recording. Presentation slides: https://libereurope.eu/article/ele-workshop-report-achieving-digital-language-equality-2030-implications-for-libraries-collections-and-library-users/

3. Analysis of Responses

3.1. Survey Responses

3.1.1. Respondents' profiling:

We received 29 responses to the survey from institutions which are members of LIBER network.

Countries: The 29 respondents covered 13 different European countries. The most represented countries, in descending order, are as follows: Spain (7), France (6), United Kingdom (3), Finland (2), Malta (2), and Croatia (2). The remaining countries each had one respondent, these are: The Netherlands, Lithuania, Luxembourg, Hungary, Greece, Serbia, and Sweden. See Figure 1 and Table 2 in Appendix B.
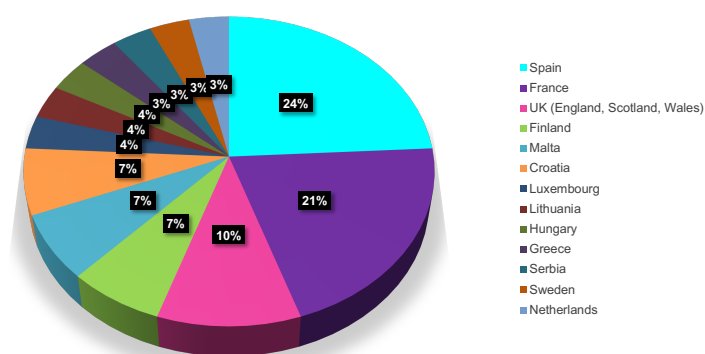
Figure 1: In which country are you based?

Organisations/Sectors: 27 respondents work for an education/research based organisations; 1 respondent was from a public sector heritage organisations and 1 from a professional association organisation. Figure 2 shows the breakdown by type of organisation.

When asked to specify the types of associations, communities, organisations and sectors of users they represent, helping to further understand the makeup of the group, 8 areas were selected, (participants could choose more than 1 response). Education was the most common response with 22 respondents. Following this sector, 7 respondents worked with Information and Communication Technologies, 7 with Digital Humanities, Arts, Culture and other services, 6 from the Research field sector, and 2 from within the Social Sciences. Publishing, Industry and Manufacturing, and Energy/Green Economy/Environment each had 1 response. See Table 3 in Appendix B.

Role or respondents: In terms of respondent's roles, at the time of the survey, 14 held research/teaching positions, 9 Professors, 4 Librarians/IT professionals, and the following groups have 1 respondent each: PhD student and Other.

Languages: The survey shows that our respondents currently use 26 different languages in their work. Respondents could select multiple languages which they used in the workplace. The breakdown of these languages across the 29 respondents is as follows: English (22), Spanish (12), French (8), German (6), Swedish (4), Basque (3), Italian (2), Croatian (2), Dutch (2), Finnish (2), Hungarian (2), Greek (2), Maltese (2), Norwegian (2), and Lithuanian (2). The following languages were selected by 1 respondent: Bulgarian, Alsatian, Turkish,
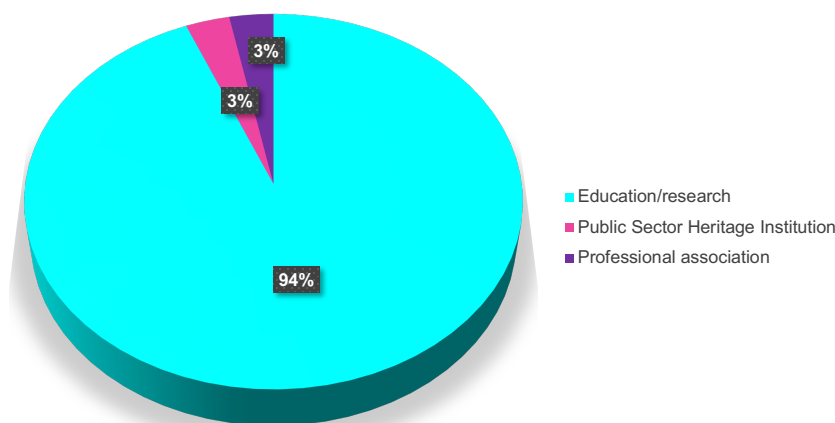
Figure 2: Which of the following best describes the type of organisation you work for?

Serbian, Scottish Gaelic, Catalan, Luxembourgish, Danish, Romanian, Portuguese, and Polish. (see Table 4 in Appendix B). Additionally, 4 respondents said that they try and work with as many languages as possible. Figure 3 shows the breakdown of languages selected.

The four largest languages used in respondent's workplaces, each accounting for over 20% of the sample, were English, Spanish, French, and German. This reflects not only the most spoken languages within the European continent, but also the fact that the membership of LIBER is particularly concentrated within Northwestern Europe. For instance, the four countries containing most LIBER members are: France, Germany, United Kingdom, and Spain. Furthermore, the research library community, which makes up the LIBER membership, are part of the wider academic community. English dominates this sector, therefore it is unsurprising that it is the most selected language, despite only having 3 respondents based in Europe's largest English speaking country, the United Kingdom, with an additional 2 from Malta, and none from Ireland.

Languages planned to be included in the workflow: 7 respondents said that they planned to add new languages to their work practices in the next three years. 3 plan to add English. The other languages that are going to be adopted by these respondents are as follows: Valencian dialect, Welsh, Breton, Catalan, German, and French. **Minority Languages**: Of the total sample, 9 respondents work with minority languages. 3 of them worked with Basque, the remaining minority languages only had one mention. These are: Scottish Gaelic, Montenegrin, Luxembourgish, Catalan, Alsatian, Turkish, and Bosnian. See Figure 3 above.

The remaining respondent, of the 9, indicated that they work with 'lots' of minority languages, but failed to offer further specification.

3.1.2. Evaluation of the Current Situation

The survey data shows how the official languages of Europe are supported by, and used in conjunction with, various types of language technologies within the workplaces of survey respondents. The following section will summarise this data and give an overview of the perceptions of the performance of these technologies.

Language tools: All the respondents use some language technologies. Search tools and translation tools are the most frequently used technologies within the sample of responses, with 22 respondents using each of these tools. When these respondents were asked to specify further what tools were used within these categories the following specific technologies were
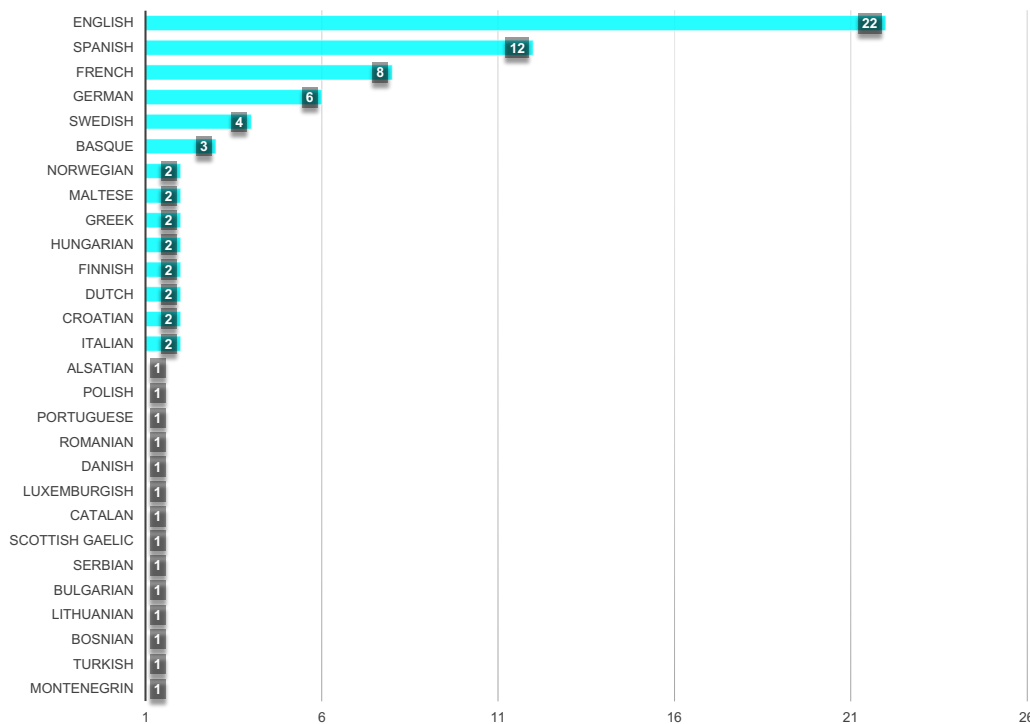
Figure 3: Which of the official European language(s) listed below do you or your organisation work with?

mentioned. Multilingual search engine, generic search systems, and web-based question answering systems were the most frequently recorded tools. Each of these items is used by over 50% of the respondents (see Table 5, Appendix B for full breakdown). As for translation tools, 62% of respondents use web-based, generic translation tools.

The next most commonly used language technologies, with European official languages, were language learning tools (16 respondents) and proofing tools (16 respondents). Upon specification by respondents, the latter is broken down into grammar checkers, spell checkers and auto correct tools, and each was reported as being used by around 50% of the initial 16 (see Table 5, Appendix B for full results). Language learning tools were reported as web-based thesaurus tools, web-based translation engines, and computer-assisted language learning tools. Beyond the four most commonly used tools, a significant drop in usage is reported. The remaining categories are used by 21% of respondents or less. This bracket includes speech recognition tools (6), parsing tools (5), sentiment and opinion analysis tools (5), text mining tools (5), text summarization tools (2), and other (1).

Performance of language tools: In relation to the usage of these language technologies, participants were also asked to rate the performance of each which they used with a major European language on a scale of 1 – 4, from very poor, poor, good, and excellent. In summary, search tools performed strongest with an average of 3.4 (between good and excellent), similarly proofing tools recorded an average score of 3.2. Also scoring favorably, was translation tools with a score of 2.5. Language learning, search tools, parsing, text mining, sentiment analysis and opinion analysis tools, and text summarization, each were deemed poor to very poor on average.

Technological gaps: When asked if technological gaps were perceived in the language technologies they used with official European languages, 16 respondents felt this was the
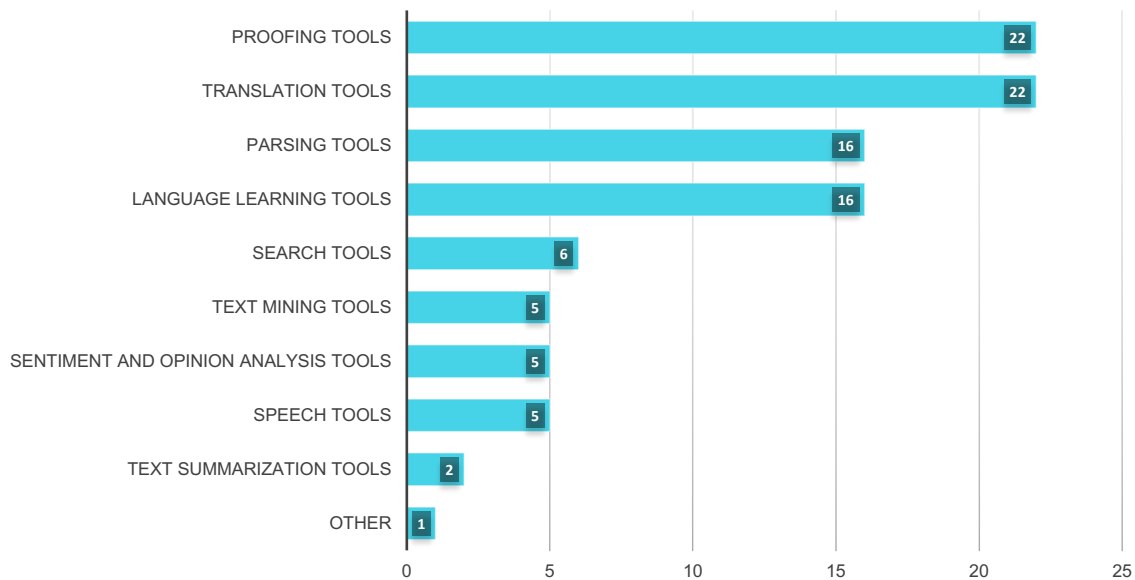
Figure 4: Which language technology tools/applications listed below do you or your organisation use with the official European language(s) you or your organisation work with?

case. These respondents could then select what issues they perceived with each of the languages they used across the 26 official European languages. In summary, across the 12 languages in this sample, the most cited issue was the amount and variety of available applications, selected 21 times. The next most recognised gap was quality of tools and applications (17). The variety of linguistic phenomena covered/text types covered was considered an issue by 13 respondents. Finally, adaptability to computer systems was identified as a gap by 10 respondents.

Technological support gaps: Of the total sample, 16 respondents felt there were gaps with the technological support for the tools they used with majority European languages. When they expanded on the gaps they encountered, with a multiple-choice selection, these options were selected with descending frequency: Gaps in the quality of the tool/application, e. g., delays in responding, difficulties with special characters, language-related errors in the output etc. (14); gaps in the variety of linguistic phenomena/text types covered (13); gaps in the amount and variety of available applications (13); and gaps in adaptability to systems, e. g., adaptability to iOS system (7).

Level of technological support: Respondents were asked to rate the level of technological support for the official European languages with which they work. Each language could be rated on a four-point scale (where 1 = *very poor*, 2= *poor*, 3= *good*, 4= *excellent*). The average score for each language is shown on Figure 5 .

This breakdown shown in this graph highlights that the four languages considered to have good support or above, on average, are English, French, German, and Spanish. The English language sits considerably higher amongst this group with an average of 3.1. All of the remaining languages were considered to have no support to poor support. At the lower end of the scale, the languages with average score under 1.0 are as follows: Irish, Lithuanian, Italian, Bulgarian, Slovak, Romanian, Latvian, Estonian, Icelandic, and Danish.
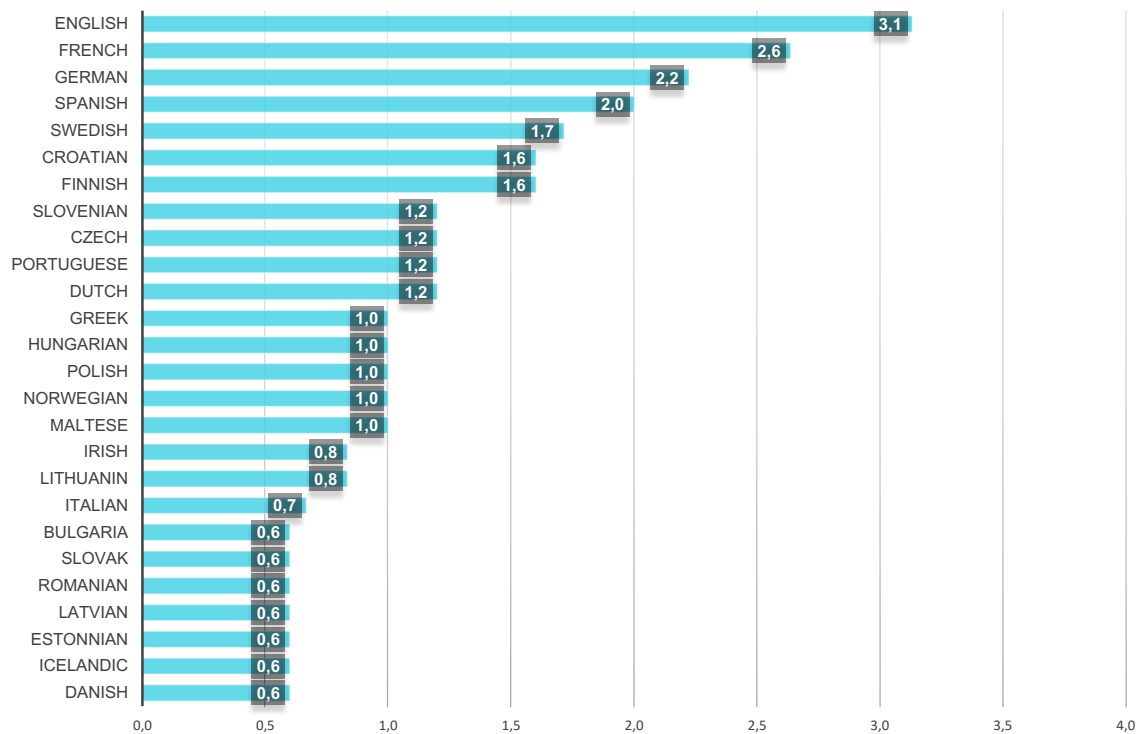
Figure 5: Please choose the option that best describes the level of language technology support for the official European language(s) you or your organisation work with.

Minority/regional/lesser-used languages

Language tools: 9 survey participants indicated that there are language technology tools or applications available for the minority languages they work with. In total, these 9 respondents pointed out 9 different tools. In descending order, they are as follows (including, how many times mentioned): Translation tools (9), proofing tools (7), language learning tools (5), parsing tools (5), search tools (5), speech recognition tools (3), sentiment and opinion analysis tools (2), text mining tools, e. g., IBM Watson (1), and text summarisation tools (1).

Technological support gaps: Survey participants were asked if they perceive gaps in technological support provided for minority languages that they work with, of which 8 respondents indicated they did. When they expanded on the gaps they encountered, with a drop down selection, these options were selected with descending frequency: Gaps in the quality of the tool/application, e. g., delays in responding, difficulties with special characters, language-related errors in the output etc. (8); gaps in the variety of linguistic phenomena/text types covered (8); gaps in the amount and variety of available applications (6); and gaps in adaptability to systems, e. g., adaptability to iOS system (5).

Comparison of LT tools used across languages

To compare the tools used for recognised European languages and minority languages, some broad observations are stated: For both recognised languages and minority languages, translation tools are the most frequently used language technology. Search tools were as frequently used by respondents, in reference to official European languages, however they were far less common for the minority languages. Proofing tools were the second most used

technology for minority languages, whereas respondents reported them slightly less often for European official languages. Language learning tools were in similar demand across both language brackets, being the third most popular in each. Similar to each was the peripheral usage of text mining tools and text summarization tools, both with least reported use.

3.1.3. Predictions and Visions for the Future

The survey asked respondents to imagine the next 10 years and imagine what they think would happen with the advance and possibilities of language technologies. Figure 6 shows the full results.

● In the next 10 years, language technology tools will help prevent the loss of linguistic diversity

● In the next 10 years, there will be a wider range of language tools for European Languages

● In the next 10 years, there will be higher-quality language tools that deal with all the languages that concern me, including minority languages
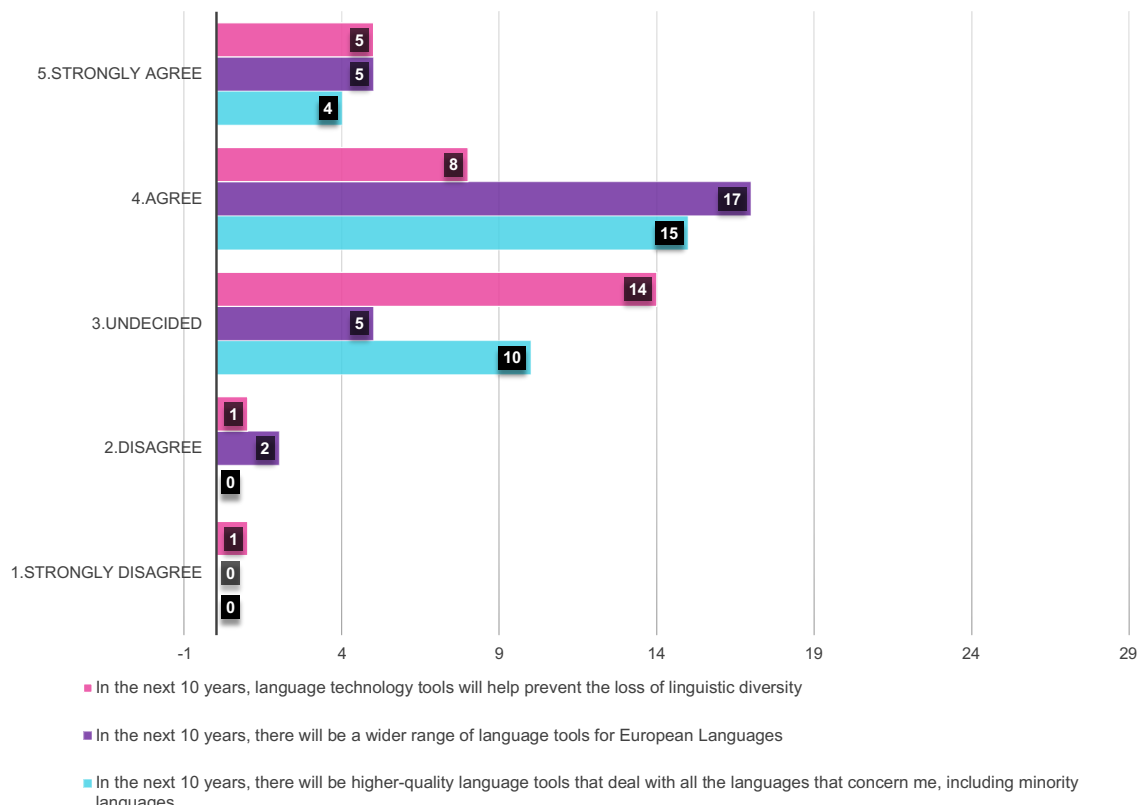
Figure 6: Please indicate the best option that describes your vision for the future of languages technology

Preventing the loss of linguistic diversity: 14 respondents were undecided on the capacity of language tools to prevent the loss of linguistic diversity, amounting to the mean response to this question, meaning that, of the provided options, this statement attracted the most uncertainty. However, 5 strongly agreed that they would prevent the loss of linguistic diversity, and 8 agreed.

Wider range of language tools: Survey respondents agreed that there would be a greater range of technologies available in the next 10 years, with 5 strongly agreeing and 17 agreeing. Only 5 stated they were undecided and 2 disagreed.

Increased quality of tools for languages (including minority languages): 19 respondents agreed that the quality of tools related to the languages they use would improve within

10 years (4 strongly agreed, 15 agreed). 10 remained undecided on this, but no respondent disagreed with the statement.

For further comment on the future situation Table 7 in Appendix B contains verbatim responses about potential language technologies for languages they work with.

3.2. Interview Responses

3.2.1. Workshop Feedback and Discussion

At the workshop on the 18th November 2021, participants were encouraged to give feedback after speakers had given their presentations, providing thoughts and reflections on what had been presented, use cases of digital language technologies, and general questions. The topic which dominated the discussion was how translation technologies can help the dissemination of research articles, opening access to new audiences, in new languages, and therefore creating more options for academic publishing in a range of languages.

The audience discussion focused on how research articles and academic publications can be affected by translation technologies. A series of practical questions were posed, asking about what services could provide translations, and how researchers could provide their publications for translation. In general, there was an interest in understanding best practices to utilise translation technology on this topic, led by research libraries. Open Science was also linked with this issue as it was anticipated that it would help to disseminate articles to wider audiences, and create more options for researchers to publish in languages other than English.

Some expertise on this issue was shared by the speakers. The importance of full document artificial intelligence (AI) was stressed. In addition, a case was brought up about individual Digital Object Identifier (DOI) for translated articles. For example, if a translated article has a new DOI, differing from the original, is this an issue? Should there be some way for DOIs, of the same piece of research in different languages, to be linked to the original text, in its original linguistic form. Finally, the example for post-editors was discussed. This was raised to show the audience that a layer of human validation can be used to confirm the work of the algorithm, also being used to train the AI. However, a point of caution was raised. One participant highlighted the complexity and cultural specificity of writing academic publications. The types of research papers that would be translated could include culturally grounded terms and concepts that are even difficult for expert translators to tackle, often requiring vast efforts with new translations and terminologies to be developed. Often experts who understand the subject have problems translating these topics as research output is highly specific by its very nature.

The participants and panel agreed that it is vital to highlight the current limitations of translation technologies, especially in the field of academic writing. To summarise the response to this point, it was stressed that, at present, machine translation is not replacing a translator's job precisely for these reasons. But it is changing the job, as translation tools can be useful for a wider dissemination of research. Two things should be done to ensure this: collecting worthwhile and usable corpora as well as identifying new skill gaps and building training curricula to overcome these issues. Finally, participants agreed that more funding was needed to further this field. In summary, it is clear that the participants of this session mainly think about language equality and language technologies in relation to scholarly communication through the sharing and spread of academic research.

3.2.2. Usage of Language Technologies

Those interviewed had varied experience with LT and this largely reflected the nature of their work. Will Lamb's research specialisation is Scottish Gaelic, a minority language, and

he mainly works with natural language processors (NLPs), speech taggers, and handwriting and speech recognition software. Across the last eight years he has become increasingly involved with digital technologies. Ranka Stanković is a Professor of Mining and Engineering at the University of Belgrade, teaching and researching information sciences. She produces LTs, but also uses them with the Serbian languages, mainly with AI and NLP processes.

Simon Worthington, in his position at the German National Library of Science and Technology, has begun to explore translation technologies due to their potential for publishing scientific work. At the time of the interview, Simon had been experimenting with digital translation services, citing an online switch board called Translate Science.

3.2.3. Potential of Language Technology

The need and potential of translation technologies, for Ranka Stanković, had much to do with the Serbian language and academic output. "You can reach much higher impact if you publish something in English then if you bought your something with a Serbian dataset... that's the problem. Also, today all scientific work is published – I think more than 90 percent in English. Serbian scientific language, academic language, is maybe disappearing or becoming less important" (Ranka Stanković, Interview, 23/11/21). This also harms Serbian culture and understanding of Serbian resources. Similarly, Simon Worthington also focused on the issue of publishing but went further when describing the possibilities of translation technology. In his opinion, the only way that today's most important issues, chiefly the climate crisis, can be addressed is with knowledge and information. It is therefore essential that all scientific work is available, on an even footing, to speakers of all languages.

For Will Lamb, the potential was twofold: advanced language technologies for Scottish Gaelic would help him as a researcher and his academic field in general, and on a societal level, a closer integration with technology would enable the protection and furtherment of Scottish Gaelic. The second point centred on the concept of accessibility; if historic languages are not digitised, they will struggle to take hold in today's predominantly digital world.

3.2.4. Gaps and Challenges

Simon Worthington has encountered numerous questions and issues as he started to explore language technology. First, he stated that he is a yet to find accepted standards or best practices when it comes to running translations of scientific publications. His main questions include: "How does a research article know it's got translations; And then how do you know? How much of it has been translated? I'm trying to see how can you integrate translation into the publishing workflow?" (Simon Worthington, Interview, 01/12/21). As a more technical issue, he discussed issues with the processing of left to right, and right to left, written language with AI.

A key gap, identified by Ranka Stanković, is that Serbian is badly in need of modernisation. She explained how the language is not supported by language processing technology and how dictionaries and corpus are still totally on paper. This focuses more on the speed of change and slow progress that is made with the language, less to do with the technology itself, perhaps reflecting her dual position as a user/developer.

For Will Lamb, working with a minority language always means you are following the technological progress of bigger languages. However, the gaps were not only technological. On several occasions he mentioned a key skills gap that is hindering his work, knowledge of Scottish Gaelic, and the position of the language generally. Routinely, Gaelic speakers lack technological skills, and vice-versa, as Will explained: "Often people with the computational skills don't overlap with the people who have language skills" (Will Lamb, Interview,

22/11/21). This issue has led him to enrol in a masters' program in Speech and Language Processing to bridge this gap in his own skills. Overall, this reflects a level of frustration with the current technologies available and their technological capacity to work with Scottish Gaelic, and with a more in-depth knowledge he would not just be a frustrated user, but act as a user-producer, actively shaping the technology.

3.2.5. Future Situation

In the future, Will Lamb hoped for more funding to set up MA and PhD programs for speakers of Scottish Gaelic to develop technological training. Furthermore, he praised the premise of the ELE project, especially the strength of cooperation: "anytime you've got a lot of different people from different languages working towards common goals, then there's always the potential for sharing good practice" (Will Lamb, Interview, 22/11/21). Ranka Stanković also focused on funding difficulties. Her team had previously put forward funding proposals to the Serbian government, she recounted, but these were not granted. She expressed frustration that Serbia has one of the lowest levels of funding for language technology. Going forward, Simon Worthington felt like he needed more knowledge and widely accepted standards for how translation technologies can play a role in publishing. In other words, he misses accepted standards for how translation technologies can be used. He pointed out that this issue had become increasingly important as people tried to address imbalances between the Global North and South, and this issue is unlikely to go away. Therefore, the need for high-quality translation technologies and accepted processes would be vital.

4. Conclusions

The three sources of data presented in this report are broadly summarised in the following points.

The survey made it clear that English is the most widely used language in the European research library community. However, it is evident that other languages are also widely used within certain European regions despite not being primary languages. For example, 6 respondents reported using German but only 1 respondent comes from a country in which German is considered an official language. Following this, the importance of non-English languages should not be downplayed within academic communities. Participants in the workshop and those interviewed did not dispute the notion that English is the current lingua-franca in academia, however they all spoke about the importance of other languages, both majority and minority.

From the survey we can see that there is a desire to add more languages to workflows in the coming years, demonstrating a will to expand language diversity within workplaces. Digging deeper, these responses either related to large and established languages such as English, French, and German, or minority languages. It may be stated that either preserving/promoting small languages or adding widely spoken languages is the aim or participants. It is possible to say that although only 9 respondents currently worked with minority languages, 4 respondents hoped to add minority languages in the coming years: a significant increase considering the sample size. This also echoes with interviewee Will Lamb's work on Gaelic and the growth of interest in the language. The goals of research libraries, their staff, and those affiliated, could be to either expand by adding large languages to workflows to reach a large audience or become hyper-local adding minority languages.
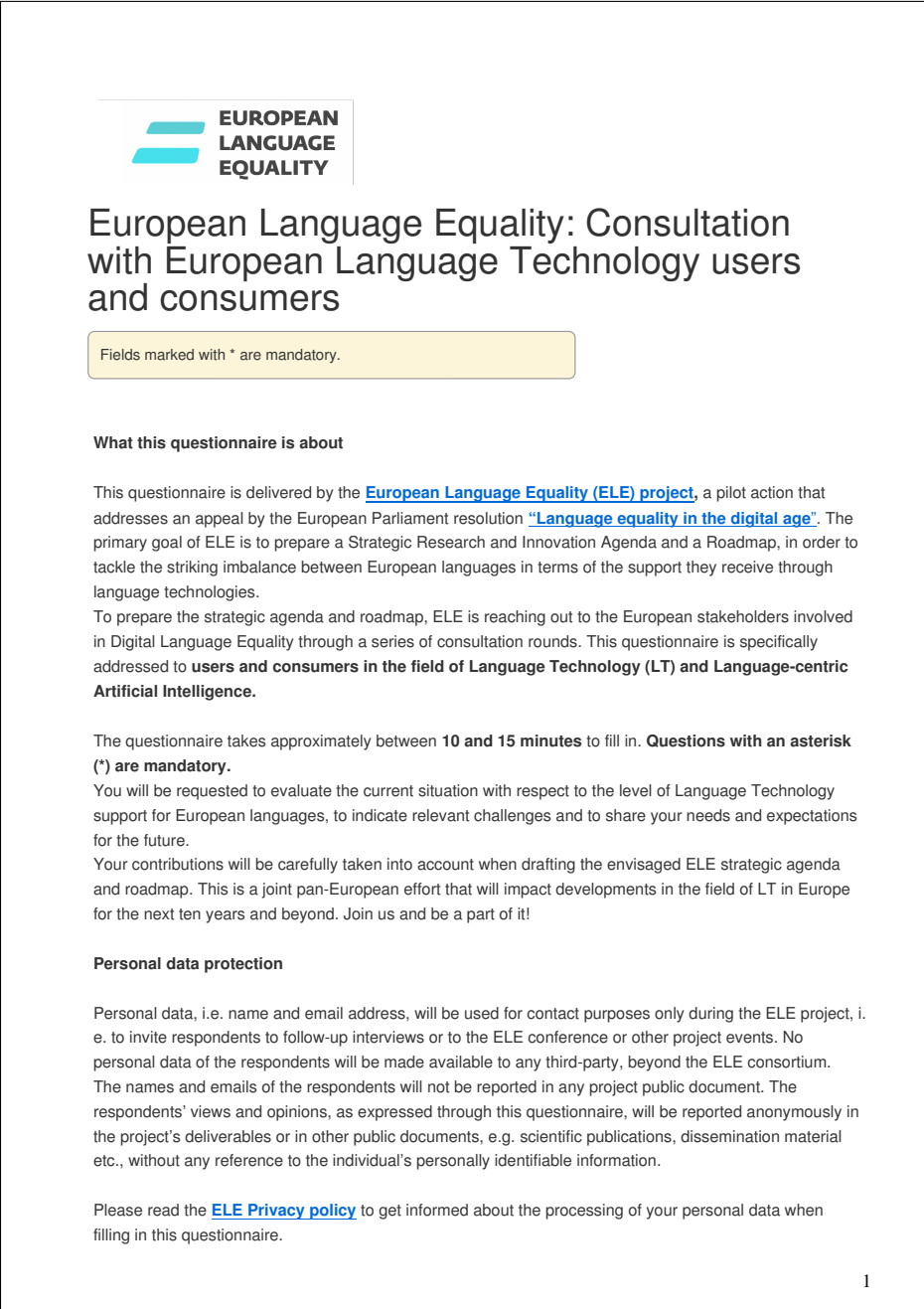
The levels of technological support for each language, as reported in the survey, demonstrates that the more widely spoken a language is the more technological support there is. English was by far the most used language within the group of respondents, and it also scored

most positively on the question regarding supporting technology. Indicative of this trend is the fact that Swedish, used by 4 respondents, performed well in the technological support question in comparison to the rest of the languages, ranking fifth behind English, Spanish, German, and French. From this we can draw two conclusions. First, the size of a language may lead to better services and support for it. Second, the relatively low sample size may affect the accuracy of our understanding for technological support for other languages than English, Spanish, French, and German.

The workshop and interviews highlighted that there exists general optimism and excitement about how language technologies can both protect and promote language diversity, and support scholarly communication. Data from the survey paints a similar picture. Overall respondents were overwhelmingly neutral to positive that the future of language technology involves: 1) a wider range of digital language tools; 2) digital language tools of better quality; and 3) preservation of language and promotion of linguistic diversity. From this, you can clearly summarise that there is not pessimism about the future advancement of language technology. Qualitative data, presented in this report, shows that there are still many questions about the best path to these goals, with little clarity about best practice in this field. Despite this, the potential of language technologies is evident.

A. LT users and consumers survey

Figures 7 to 24 show the complete LT research and developers survey.

Figure 7: Full survey as published (page 1/18)

Introduce yourself and your organisation

* **In which country are you based?**

- Austria
- Germany
- Poland
- Belgium
- Greece
- Portugal
- Bulgaria
- Hungary
- Romania
- Croatia
- Ireland
- Slovak Republic
- Cyprus
- Italy
- Slovenia
- Czechia
- Latvia
- Spain
- Denmark
- Lithuania
- Sweden
- Estonia
- Luxembourg
- Other
- Finland
- Malta
- France
- Netherlands

* If "other', please specify.

* **Which association(s)/community(ies)/organisation(s)/sector(s) of users and consumers do you represent?**

Please, select as many as apply

- Agriculture and fisheries
- Finance/banking
- Publishing
- Digital Humanities, arts, culture and other services
- Health
- Research
- Broadcasting
- Industry and manufacturing
- Security (threat detection in general)
- Business services
- Information and Communication Technologies
- Social Sciences
- Construction
- Insurance industry
- Tourism, accommodation and food services
- eCommerce
- Justice and legal
- Trade and repair
- Education
- Media
- Transportation, logistics and storage
- Energy/green economy /environment
- Public administration
- Other

* If "other", please specify.

2

Figure 8: Full survey as published (page 2/18)

* **What is the name of the organisation/representative body you work for? (if you are self-employed or if you are not employed, please specify)**

* **How many members are there in the association(s)/community(ies)/organisation(s)/sector(s) of users and consumers you represent in this survey? (total number of full-time employees)**
  - ○ 1-10
  - ○ 11-100
  - ○ 101-500
  - ○ 501-5000
  - ○ More than 5000
  - ○ N/A
  - ○ Not sure

* **Which of the following best describes the type of organisation you work for?**
  - ○ Professional association
  - ○ Government department/unit
  - ○ SME
  - ○ Large Enterprise
  - ○ Independent contractor/ consultant
  - ○ Education/research
  - ○ N/A
  - ○ Other

* **If "other", please specify.**

* **What is your main role at the organisation where you work? (if you are self-employed or if you are not employed, please specify)**

  If you are the representative of a community of users and consumers, please enter your role at the representative body you work for.

3

Figure 9: Full survey as published (page 3/18)

Language Coverage

* **Which of the official European language(s) listed below do you or your organisation work with?**

if you represent an organisation/community of users and consumers please select the languages this organisation /community work with. Otherwise, please select the languages you work when using language technologies.

☐ Bulgarian ☐ German ☐ Norwegian
☐ Croatian ☐ Greek ☐ Polish
☐ Czech ☐ Hungarian ☐ Portuguese
☐ Danish ☐ Icelandic ☐ Romanian
☐ Dutch ☐ Irish ☐ Slovak
☐ English ☐ Italian ☐ Slovenian
☐ Estonian ☐ Latvian ☐ Spanish
☐ Finnish ☐ Lithuanian ☐ Swedish
☐ French ☐ Maltese ☐ Other

* If "other", please specify.

* **Do you or your organisation plan to include additional languages in your workflow in the next 3 years?**

○ Yes
○ No
○ Not sure

* **Which language(s)?**

☐ Bulgarian ☐ German ☐ Norwegian
☐ Croatian ☐ Greek ☐ Polish
☐ Czech ☐ Hungarian ☐ Portuguese
☐ Danish ☐ Icelandic ☐ Romanian
☐ Dutch ☐ Irish ☐ Slovak
☐ English ☐ Italian ☐ Slovenian
☐ Estonian ☐ Latvian ☐ Spanish
☐ Finnish ☐ Lithuanian ☐ Swedish
☐ French ☐ Maltese ☐ Other

* If "other", please specify.

4

Figure 10: Full survey as published (page 4/18)

Is any of the languages you selected considered a minority/regional/lesser-used language?

○ Yes
○ No

* **Do you or your organisation work with any minority/regional/lesser-used language(s) not included in the list of EU languages provided above?**

Minority languages/regional/lesser-used languages are languages that are traditionally used within a given territory of a state by nationals of that state who form a group numerically smaller than the rest of the state's population and [are] different from the official language(s) of that state" (Council of Europe, 1992, p. 2)

○ Yes
○ No

* **Which minority/regional/lesser-used language(s)?**

Evaluation of the current situation

* **Which language technology tools/applications listed below do you or your organisation use with the official European language(s) you or your organisation work with?**

If you are the representative of a organisation/community of users and consumers, please select the tools used by the organisation/community. Otherwise, select the tools you use with the languages you work with.
For examples of these types of tools/applications, click on boxes and select as many as apply.

☐ Proofing tools   ☐ Sentiment and opinion analysis tools
☐ Translation tools   ☐ Text summarization tools (e.g. Quilbot AI)
☐ Speech recognition tools   ☐ Text mining tools (e.g. IBM Watson)
☐ Parsing tools   ☐ Language learning tools
☐ Search tools   ☐ Other

* Proofing tools

Please, select as many as apply.

☐ Spell checkers
☐ Grammar checkers
☐ Autocorrect tools

* Translation tools

☐ Computer-assisted translation tools (e.g. translation memories)
☐ Terminology management applications

5

Figure 11: Full survey as published (page 5/18)

Generic translation tools freely available on the web (e.g. Google Translate)

Custom-built translation engines

* Speech recognition tools

Voice user interfaces (e.g. Siri, native android, native iOS, smart speakers [Google home, Alexa, ...], Bose Headphones, Adobe Acrobat reader, Amazon Polly, Chromevox, Wordreference)

Text-to-speech systems (i.e. systems that turn text into speech for reading texts out loud (e.g. Amazon Polly, Adobe Acrobat reader)

* Parsing tools

Dependency or constituency parsing systems to automatically analyse the syntax of textual or spoken data (e.g. Stanford NLP's CoreNLP java framework, Stanford NLP Stanza, AllenNLP parsing, UDPipe, MaChAmp)

Part-of-speech taggers of any type (e.g. NLTK python library, NLPdotnet)

* Search tools

Web-based question-answering systems (e.g. Stack exchange, StackOverflow, Quora, Google search)

Ontology tools for extracting the corresponding domain's terms and the relationships between the concepts that these terms represent in a text (e.g. Robot tool)

Generic search systems freely on the web (e.g. Google search)

Customer-build search engines (e.g. organisations or vendors create search engines themselves)

Domain-specific search engines (focusing on domain-specific topics, e.g. PubMed, Copernic, CC search)

Multilingual search engines (e.g. Google, Wikipedia)

Cross-language search engines (e.g. eBay, Aliexpress)

Language-focused search engines (e.g. Baidu)

Multimedia search engines (e.g. plantnet, or applications like 'Snooth')

Private search engines (e.g. Search Encrypt and OneSearch, use different encryption methods to keep your query private)

* Language learning tools

Computer-assisted language learning tools (e.g. Duolingo, FluentU, SKELL)

Web-based thesaurus tools (help users to find synonyms of words)

Intelligent systems to aid and assess reading comprehension (e.g. Whooo's Reading, Storia)

Web-based translation search engines (e.g. Linguee)

* If "other" tool(s), please specify.

* Do you perceive gaps in technological support for the official European language(s) you work with?

By gaps in technological support we mean, for instance, gaps in the variety of available applications for certain languages, gaps in the quality of tools for certain languages, among other gaps listed in the next questions.

Yes

No

6

Figure 12: Full survey as published (page 6/18)

Please, indicate the language(s) you perceive the gaps below.
Please, select as many gaps and languages as apply.

	Amount and variety of available applications	Quality of the tool/application (delays in responding, difficulties with special characters, language-related errors in the output etc.)	Variety of linguistic phenomena /text types covered	Adaptability to systems (e.g. adaptability to iOS system)	Other
Bulgarian	☐	☐	☐	☐	☐
Croatian	☐	☐	☐	☐	☐
Czech	☐	☐	☐	☐	☐
Danish	☐	☐	☐	☐	☐
Dutch	☐	☐	☐	☐	☐
English	☐	☐	☐	☐	☐
Estonian	☐	☐	☐	☐	☐
Finnish	☐	☐	☐	☐	☐
French	☐	☐	☐	☐	☐
German	☐	☐	☐	☐	☐
Greek	☐	☐	☐	☐	☐
Hungarian	☐	☐	☐	☐	☐
Icelandic	☐	☐	☐	☐	☐
Irish	☐	☐	☐	☐	☐
Italian	☐	☐	☐	☐	☐

7

Figure 13: Full survey as published (page 7/18)

Latvian	☐	☐	☐	☐	☐
Lithuanian	☐	☐	☐	☐	☐
Maltese	☐	☐	☐	☐	☐
Norwegian	☐	☐	☐	☐	☐
Polish	☐	☐	☐	☐	☐
Portuguese	☐	☐	☐	☐	☐
Romanian	☐	☐	☐	☐	☐
Slovak	☐	☐	☐	☐	☐
Slovenian	☐	☐	☐	☐	☐
Spanish	☐	☐	☐	☐	☐
Swedish	☐	☐	☐	☐	☐

8

Figure 14: Full survey as published (page 8/18)

*If "other", please specify.

In general terms, how do you evaluate the performance of the tools you use for the official European language(s) you work with?
Please evaluate based on a four-point scale.

Please, evaluate as many tools as apply. If you do not know one or more tools, please select non-applicable (N/A).

	1.Very poor	2. Poor	3. Good	4. Excellent	5. N /A
Proofing tools (e.g. Spell checkers, Autocorrect)	○	○	○	○	○
Translation tools (e.g. Google Translate)	○	○	○	○	○
Speech recognition tools (e.g. Siri, Alexa)	○	○	○	○	○
Parsing (e.g. PoS taggers)	○	○	○	○	○
Search tools (e.g. Google search)	○	○	○	○	○
Sentiment analysis and opinion analysis tools	○	○	○	○	○
Text summarization (e.g. Quillbot)	○	○	○	○	○
Text mining (e.g. IBM Watson)	○	○	○	○	○
Language learning (e.g. Duolingo, thesaurus, bilingual dictionaries)	○	○	○	○	○
Other	○	○	○	○	○

*If "other", please specify.

Please choose the option that best describes the level of language technology support for the official European language(s) you or your organisation work with.

Please, choose as many languages as apply.

	1. No support	2. Poor support	3. Good support	4. Excellent support	5. I do not know
Bulgarian	○	○	○	○	○
Croatian	○	○	○	○	○
Czech	○	○	○	○	○

9

Figure 15: Full survey as published (page 9/18)

	1.	2.	3.	4.	5.
Danish	○	○	○	○	○
Dutch	○	○	○	○	○
English	○	○	○	○	○
Estonian	○	○	○	○	○
Finnish	○	○	○	○	○
French	○	○	○	○	○
German	○	○	○	○	○
Greek	○	○	○	○	○
Hungarian	○	○	○	○	○
Icelandic	○	○	○	○	○
Irish	○	○	○	○	○
Italian	○	○	○	○	○
Latvian	○	○	○	○	○
Lithuanian	○	○	○	○	○
Maltese	○	○	○	○	○
Norwegian	○	○	○	○	○
Polish	○	○	○	○	○
Portuguese	○	○	○	○	○
Romanian	○	○	○	○	○
Slovak	○	○	○	○	○
Slovenian	○	○	○	○	○
Spanish	○	○	○	○	○
Swedish	○	○	○	○	○

Please indicate based on a five-point scale how frequently you use the language technology tools /applications listed below for the languages you work with.

Please, select as many tools as apply.

	1. Never	2. Rarely	3. Sometimes	4. Frequently	5. Every day
Proofing tools (e.g. Spell checkers, Autocorrect)	○	○	○	○	○
Translation tools (e.g. Google Translate)	○	○	○	○	○

10

Figure 16: Full survey as published (page 10/18)

Speech recognition tools (e.g. Siri, Alexa)	○	○	○	○	○
Parsing (e.g. PoS taggers)	○	○	○	○	○
Search tools (e.g. Google search)	○	○	○	○	○
Sentiment analysis and opinion analysis tools	○	○	○	○	○
Text summarization (e.g. Quillbot)	○	○	○	○	○
Text mining (e.g. IBM Watson)	○	○	○	○	○
Language learning (e.g. Duolingo, thesaurus, bilingual dictionaries)	○	○	○	○	○
Other	○	○	○	○	○

* If "other" tool(s), please specify.

Please indicate for which language(s) you or your organisation use the language technology tools /applications listed below.

Please, select as many tools and languages as apply.

	Proofing tools (e.g. Spell checkers, grammar checkers)	Translation tools (e.g. Google Translate)	Speech Recognition tools (e.g. Siri, Alexa)	Search tools (e.g. Google search, Wikipea)
Bulgarian	☐	☐	☐	☐
Croatian	☐	☐	☐	☐
Czech	☐	☐	☐	☐
Danish	☐	☐	☐	☐
Dutch	☐	☐	☐	☐
English	☐	☐	☐	☐
Estonian	☐	☐	☐	☐
Finnish	☐	☐	☐	☐
French	☐	☐	☐	☐
German	☐	☐	☐	☐
Greek	☐	☐	☐	☐
Hungarian	☐	☐	☐	☐
Icelandic	☐	☐	☐	☐

11

Figure 17: Full survey as published (page 11/18)

Irish	☐	☐	☐	☐
Italian	☐	☐	☐	☐
Latvian	☐	☐	☐	☐
Lithuanian	☐	☐	☐	☐
Maltese	☐	☐	☐	☐
Norwegian	☐	☐	☐	☐
Polish	☐	☐	☐	☐
Portuguese	☐	☐	☐	☐
Romanian	☐	☐	☐	☐
Slovak	☐	☐	☐	☐
Slovenian	☐	☐	☐	☐
Spanish	☐	☐	☐	☐
Swedish	☐	☐	☐	☐
Other	☐	☐	☐	☐

* If "other" language(s), please specify.

Are there language technology tools/applications available for the minority/regional/lesser-used language(s) you or your organisation work with?
- ○ Yes
- ○ No
- ○ I do not know

* **Which tools/applications do you use with these minority/regional/lesser-used languages?**
For more examples of these types of tools, click on the boxes and select as many tools as apply.

- ☐ Proofing tools
- ☐ Translation tools
- ☐ Speech recognition tools
- ☐ Parsing tools
- ☐ Search tools
- ☐ Sentiment and opinion analysis tools
- ☐ Text summarization tools (e.g. Quilbot AI)
- ☐ Text mining tools (e.g. IBM Watson)
- ☐ Language learning tools
- ☐ Other

* Proofing tools
Select as many as apply.
- ☐ Spell checkers
- ☐ Grammar checkers
- ☐ Autocorrect

12

Figure 18: Full survey as published (page 12/18)

* Translation tools
Select as many as apply.
- ☐ Computer-assisted translation tools (e.g. translation memories)
- ☐ Terminology management applications
- ☐ Generic translation tools freely available on the web (e.g. Google Translate)
- ☐ Custom-built translation engines

* Speech recognition/synthesis tools
Select as many as apply.
- ☐ Voice user interfaces (e.g. Siri, native android, native iOS, smart speakers [Google home, Alexa, ...], Bose Headphones, Adobe Acrobat reader, Amazon Polly, Chromevox, Wordreference)
- ☐ Text-to-speech systems (i.e. systems that turn text into speech or for reading text out loud (e.g. Amazon Polly, Adobe Acrobat reader)

* Parsing tools
Please, select as many as apply.
- ☐ Dependency or constituency parsing systems to automatically analyse the syntax of textual or spoken data (e.g. Stanford NLP's CoreNLP java framework, Stanford NLP Stanza, AllenNLP parsing, UDPipe, MaChAmp)
- ☐ Part-of-speech taggers of any type (e.g. NLTK python library, NLPdotnet)

* Search tools
Please, select as many as apply.
- ☐ Web-based question-answering systems (e.g Stack exchange, StackOverflow, Quora, Google search)
- ☐ Ontology tools for extracting the corresponding domain's terms and the relationships between the concepts that these terms represent in a corpus (e.g. Robot tool)
- ☐ Generic search systems freely on the web (e.g. Google search)
- ☐ Customer-build search engines (e.g organisations or vendors create search engines themselves)
- ☐ Domain-specific search engines (focusing on domain-specific topics, e.g. PubMed, Copernic, CC search)
- ☐ Multilingual search engines (e.g. Google, Wikipedia)
- ☐ Cross-language search engines (e.g. eBay, Aliexpress)
- ☐ Language-focused search engines (e.g. Baidu)
- ☐ Multimedia search engines (e.g. plantnet, or applications like 'Snooth')
- ☐ Private search engines (e.g. Search Encrypt and OneSearch, use different encryption methods to keep your query private)

* Language learning tools
Please, select as many as apply.
- ☐ Computer-assisted language learning tools (e.g. Duolingo, FluentU, SKELL)
- ☐ Web-based thesaurus tools (help users to find synonyms of words e.g. thesaurus.com)
- ☐ Intelligent systems to aid and assess reading comprehension (e.g. Whooo's Reading, Storia)
- ☐ Web-based translation search engines (e.g. Linguee)

* If "other", please specify.

13

Figure 19: Full survey as published (page 13/18)

Do you perceive gaps in technological support for the minority/regional/lesser-used language(s) you work with?

By gaps in technological support we mean, for instance, gaps in the variety of available applications for certain languages, gaps in the quality of tools for certain languages, among other gaps listed in the next questions.

○ Yes
○ No

Please, indicate the gap(s) you perceive.

Please, select as many as apply.

☐ Gaps in the amount and variety of available applications
☐ Gaps in the quality of the tool/application (delays in responding, difficulties with special characters, language-related errors in the output etc.)
☐ Gaps in the variety of linguistic phenomena/text types covered
☐ Gaps in adaptability to systems (e.g. adaptability to iOS system)
☐ Not sure
☐ Other

* If "other", please specify.

In general terms, how do you evaluate the performance of the language technology tools for the minority/regional/lesser-used language(s) you work with? Please evaluate based on a four-point scale.

Please, select as many tools as apply. If you cannot evaluate for any reason, please select not applicable (N/A).

	1. Very poor	2. Poor	3. Good	4. Excellent	5. N /A
Proofing tools (e.g. Spell checkers, Autocorrect)	○	○	○	○	○
Translation tools (e.g. Google Translate)	○	○	○	○	○
Speech recognition tools (e.g. Siri, Alexa)	○	○	○	○	○
Parsing (e.g. PoS taggers)	○	○	○	○	○
Search tools (e.g. Google search)	○	○	○	○	○
Sentiment analysis and opinion analysis tools	○	○	○	○	○

14

Figure 20: Full survey as published (page 14/18)

	1. Very poor	2. Poor	3. Good	4. Excellent	5. N/A
Text summarization (e.g. Quillbot)	○	○	○	○	○
Text mining (e.g. IBM Watson)	○	○	○	○	○
Language learning (e.g. Duolingo, thesaurus, bilingual dictionaries)	○	○	○	○	○
Other	○	○	○	○	○

If "other", please specify.

Please, choose the option that best describes the level of language technology support for the minority/regional/lesser-used language(s) you or your organisation work with.

Please, select as many tools as apply. If you do not know one or more tools, select not applicable (N/A).

	1. Very poor	2. Poor	3. Good	4. Excellent	5. N/A
Proofing tools (e.g. Spell checkers, Autocorrect)	○	○	○	○	○
Translation tools (e.g. Google Translate)	○	○	○	○	○
Speech recognition tools (e.g. Siri, Alexa)	○	○	○	○	○
Parsing (e.g. PoS taggers)	○	○	○	○	○
Search tools (e.g. Google search)	○	○	○	○	○
Sentiment analysis and opinion analysis tools	○	○	○	○	○
Text summarization (e.g. Quillbot)	○	○	○	○	○
Text mining (e.g. IBM Watson)	○	○	○	○	○
Language learning (e.g. Duolingo, thesaurus, bilingual dictionaries)	○	○	○	○	○
Other	○	○	○	○	○

* If "other", please specify.

15

Figure 21: Full survey as published (page 15/18)

Please indicate based on a five-point scale how frequently you use the language technology tools /applications listed below for the minority/regional/lesser-used languages you work with.

Please, select as many tools as apply.

	1. Never	2. Rarely	3. Sometimes	4. Frequently	5. Every day
Proofing tools (e.g. Spell checkers, Autocorrect)	○	○	○	○	○
Translation tools (e.g. Google Translate)	○	○	○	○	○
Speech recognition tools (e.g. Siri, Alexa)	○	○	○	○	○
Parsing (e.g. PoS taggers)	○	○	○	○	○
Search tools (e.g. Google search)	○	○	○	○	○
Sentiment analysis and opinion analysis tools	○	○	○	○	○
Text summarization (e.g. Quillbot)	○	○	○	○	○
Text mining (e.g. IBM Watson)	○	○	○	○	○
Language learning (e.g. Duolingo, thesaurus, bilingual dictionaries)	○	○	○	○	○
Other	○	○	○	○	○

* If "other" tool, please specify.

Predictions and visions for future

* **In your opinion, what provision of resources would increase the use of language tools for the specific languages you or your organisation use?**

Please, select as many as apply.

☐ A wider range of language tools for the languages I work with
☐ Higher-quality tools for the languages I work with
☐ More training of personnel dealing with such tools
☐ Other

* If "other", please specify.

16

Figure 22: Full survey as published (page 16/18)

Which tools or applications that could potentially use language technology do you want to see that is not currently available for the languages you work with (we welcome any suggestion, even ideas that are not possible with current technology)?

Please indicate the best option that describes your vision for the future of languages technology.

	1. Strongly disagree	2. Disagree	3. Undecided	4. Agree	5. Strongly Agree
* In the next 10 years, there will be higher-quality language tools that deal with all the languages that concern me, including minority languages	○	○	○	○	○
* In the next 10 years, there will be a wider range of language tools for European Languages	○	○	○	○	○
* In the next 10 years, language technology tools will help prevent the loss of linguistic diversity	○	○	○	○	○

* **In your opinion, what would be the most relevant benefits of improving technologies for the languages you or your organisation work with (including minority/regional/lesser-used languages)?**

Please, select as many as apply.

☐ Increase individuals' exposure to these languages
☐ Prevent minority/regional languages from disappearing
☐ Increase the number of speakers of those languages, including minority/regional languages
☐ Improve communication between native speakers
☐ Improve literacy for minority/regional languages
☐ Enhance the communication capabilities of people with disabilities
☐ Increase engagement with social, leisure and work activities in their own languages
☐ Improve online trade in countries where those languages are spoken
☐ Improve offline trade (i.e. not e-commerce) in countries where those languages are spoken
☐ Other

* If "other", please specify.

17

Figure 23: Full survey as published (page 17/18)

If you have any comments/suggestions, please let us know.

* Can we contact you to arrange a possible follow-up discussion?

○ Yes
○ No

* What is your e-mail address?

What is your name?

☐ **By clicking on 'Submit', I agree that my personal data (email address and/or name) can be used according to the Privacy Policy of the European Language Equality (ELE) project.**

ELE_Privacy_Policy.pdf

18

Figure 24: Full survey as published (page 18/18)

B. Additional tables and graphs

Country	Answers count	%
Spain	7	24.1
France	6	20.7
Finland	2	6.9
Malta	2	6.9
Croatia	2	6.9
UK (England, Scotland, Wales)	3	10.3
Luxembourg	1	3.4
Lithuania	1	3.4
Hungary	1	3.4
Greece	1	3.4
Serbia	1	3.4
Sweden	1	3.4
Netherlands	1	3.4

Table 2: Breakdown of answers count to questions "In which country are you based? if "other", please speficy"

Types of organisations	Answers count	%
Education/research	27	93.1
Public Sector Heritage Institution	1	3.45
Professional association	1	3.45

Table 3: Breakdown of answers count to the question "Which of the following best describes the type of organisation you work for?"

Languages	Answers count	%
English	22	75.9
Spanish	12	13.1
French	8	27.6
German	6	20.7
Swedish	4	13.8
Basque	3	10.3
Italian	2	6.9
Croatian	2	6.9
Dutch	2	6.9
Finnish	2	6.9
Hungarian	2	6.9
Greek	2	6.9
Maltese	2	6.9
Norwegian	2	6.9
Lithuanian	1	3.4
Bulgarian	1	3.4
Serbian	1	3.4
Scottish Gaelic	1	3.4
Catalan	1	3.4
Luxembourgish	1	3.4
Danish	1	3.4
Romanian	1	3.4
Portuguese	1	3.4
Polish	1	3.4
Alsatian	1	3.4

Table 4: Breakdown of answers to the question "Which of the official European language(s) listed below do you or your organisation work with? if "other", please speficy"

Language Technologies	Answers counts	%
Parsing tools		
Part-of-speech taggers of any type	5	17.2
Dependency or constituency parsing systems	4	13.8
Proofing tools		
Grammar checkers	16	55.2
Spell checkers	15	51.7
Autocorrect tools	14	48.3
Search tools		
Multilingual search engines	20	69
Generic search systems freely on the web	20	69
Web-based question-answering systems	15	51.7
Domain-specific search engines	13	44.8
Ontology tools	4	13.8
Customer-build search engines	4	13.8
Multimedia search engines	3	10.3
Cross-language search engines	2	6.9
Language-focused search engines	1	3.4
Speech technologies		
Voice user interfaces	4	13.8
Text-to-speech systems	3	10.3
Translation tools		
Generic translation tools freely available on the web	18	62.1
Computer-assisted translation tools	11	37.9
Terminology management applications	7	24.1
Custom-built translation engines	1	3.4
Language Learning tools		
Web-based thesaurus tools	12	41.4
Web-based translation search engines	10	34.5
Computer-assisted language learning tools	8	27.6
Other		
Information extraction tools in the medical domain, medical ontologies	1	3.4

Table 5: Breakdown of answers to the question: "Which language technology tools or applications listed below do you or your organisation use with the official European language(s) you or your organisation work with? if "other", please speficy"

Language Technologies	Answers counts	%
Parsing tools		
Part-of-speech taggers of any type	3	10.3
Dependency or constituency parsing systems	2	6.9
Proofing tools		
Grammar checkers	5	17.2
Spell checkers	7	24.1
Autocorrect tools	5	17.2
Search tools		
Generic search systems freely available on the web	5	17.2
Multilingual search engines	4	13.8
Customer-build search engines	3	10.3
Web-based question-answering systems	2	6.9
Private search engines	2	6.9
Domain-specific search engines	2	6.9
Ontology tools	1	3.4
Multimedia search engines	1	3.4
Language-focused search engines	1	3.4
Cross-language search engines	1	3.4
Speech technologies		
Voice user interfaces	3	10.3
Text-to-speech systems	2	6.9
Translation tools		
Generic translation tools freely available on the web	8	27.6
Custom-built translation engines	5	17.2
Computer-assisted translation tools	5	17.2
Terminology management applications	2	6.9
Language Learning tools		
Web-based thesaurus tools	2	6.9
Web-based translation search engines		
Intelligent systems to aid and assess reading comprehension	5	17.2
Computer-assisted language learning tools	5	17.2

Table 6: Breakdown of answers count to the question: "Which language technology tools or applications listed below do you or your organisation use with the minority, regional, lesser-used languages you or your organisation work with? if "other", please speficy"

Machine translation that translates directly between Finnish and other languages instead of going through a pivot (assumed English)
Instantaneous subtitled/audio translations so that our events can be in multiple languages but accessible to all.
Tools to help teachers to communicate with parents and pupils with whom they do not share a language.
Tools which take into account regional variations in pronunciation, vocabulary... (e.g. spoken Arabic can be very different in different areas of the world).
Pronunciation practice and immediate feedback
Improvement of speech recognition for Greek.
We are currently working to bring a suite of Gaelic NLP tools to users of the language, including a PoS tagger, a lemmatiser and a parser – are tools that have already been developed - as well as an orthographic normalisation tool and automatic speech recognition. The last two tools are still in development, but we expect to provide freely available prototypes by the end of Nov 2021.
I'm working on such applications in my research although, once again, I'm wondering how sustainable they might be: collaborative terminography, collaborative corpus processing ...
Promote communities of learning languages in europe through technology
Text summarisation, QA tools, Multilingual resources and supporting tools (including APIs), Computational semantics

Table 7: Full list of answers to the question "Which tools or applications that could potentially use language technology do you want to see that is not currently available for the languages you work with (we welcome any suggestion, even ideas that are not possible with current technology)?"