EUROPEAN LANGUAGE EQUALITY

D2.17

Report on all external consultations and surveys

Authors	Andy Way, 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, Artem Revenko
Dissemination level	Public
Date	31-03-2022

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	D2.17
Deliverable title	Report on all external consultations and surveys
Type	Report
Number of pages	115
Status and version	Final
Dissemination level	Public
Date of delivery	Contractual: 31-03-2022 – Actual: 31-03-2022
Work package	WP2: European Language Equality – The Future Situation in 2030
Tasks	Task 2.1 The perspective of European LT developers (industry and research), Task 2.2 The perspective of European LT users and consumers
Authors	Andy Way, 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, Artem Revenko
Reviewers	Federico Gaspari, German Rigau
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

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
AI4EU	AI4EU (EU project, 2019-2021)
ACL	Association for Computational Linguistics
CLAIRE	Confederation of Laboratories for AI Research in Europe
CLARIN	Common Language Resources and Technology Infrastructure
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DLE	Digital Language Equality
DH	Digital Humanities
EAMT	European Association for Machine Translation
ECSPM	European Civil Society Platform for Multilingualism
EFNIL	European Federation of National Institutes for Language
ELE	European Language Equality *(this project)*
ELE Programme	European Language Equality Programme *(the long-term, large-scale funding programme specified by the ELE project)*
ELEN	European Language Equality Network
ELEXIS	European Lexicographic Infrastructure
ELISE	European Network of AI Excellence Centres
ELG	European Language Grid (EU project, 2019-2022)
ELRA	European Language Resource Association
ELRC	European Language Resource Coordination
HPC	High-Performance Computing
LIBER	Ligue des Bibliothèques Européennes de Recherche – Association of European Research Libraries
LR	Language Resources/Resources
LT	Language Technology/Technologies
META-NET	EU Network of Excellence to foster META
ML	Machine Learning
MT	Machine Translation
NEM	New European Media
NLP	Natural Language Processing
NLU	Natural Language Understanding
SME	Small and Medium Company
ST	Speech Technologies
SRIA	Strategic Research and Innovation Agenda
SSH	Social Sciences and the Humanities
TA	Text Analytics

Abstract

The primary objective of the ELE project is to prepare the European Language Equality Programme, in the form of a strategic research, innovation and implementation agenda (SRIA) as well as a roadmap for achieving full digital language equality in Europe by 2030. As a project from the community for the community, the consortium wants to ensure that all voices are heard and taken into account for the ELE SRIA and roadmap. This deliverable presents the insight gained from a number of online surveys and expert interviews targeting Language Technology (LT) developers and LT users and consumers. The surveys investigated language coverage, evaluated the current situation of LT in Europe and encouraged participants to share their predictions and visions for the future. More than 450 survey responses were collected and more than 65 expert interviews were conducted. In addition to these expert opinions, we also organised a short 3 minute survey, targeted at the European citizens, to investigate how they feel about the digital support for their languages. Further, four ELE industry partners compiled deep dive reports for the fields of Machine Translation, Speech, Text Analytics and Data and Knowledge.

1 Introduction

This deliverable D2.17 gives technical and processing details on the internal and external WP2 consultations and surveys. The aim of WP2 was to collect input for the Strategic Research and Innovation Agenda (SRIA) and to produce several reports by a broad spectrum of stakeholders – from research through industry to users – about their views, needs and perspectives related to Language Technologies (LT) and digital language equality, while having in mind specific developments over the coming 10 years. Special focus was placed on ways and means of achieving digital language equality during this period through the development, implementation and use of LT, in order to make Europeans of all regions and origins truly equal when accessing education, business, governments and public services in their own languages. Information was collected through a number of surveys, specifically tailored to target researchers, developers, innovators and users and their communities as well as through reports that were produced by a number of ELE partners. Given the COVID-19 restrictions during 2021, external consultations were either carried out in an asynchronous way (through effective online surveys or online questionnaires) or in a synchronous but remote way. The results described and summarised in this document are also used as one of the major sets of input to the development of the Strategic Agenda and roadmap that are prepared as part of WP3 (Deliverable D3.2).

The ELE consortium brings together a large number of partners that cover all areas that are relevant for the development of the SRIA. The project has benefited immensely from the partner's expertise and their community reach. The consortium-internal and external stakeholders' input and feedback was systematically collected throughout the project runtime.

We distinguish between two main stakeholder groups – LT developers (industry and research) and LT users as well as consumers. Each group is very diverse, including many subgroups, representing a magnitude of domains. For the latter we looked at the interesting subdivisions of commercial and academic users as well as the EU citizen. Both groups are represented in ELE with several networks, initiatives and associations who also delivered one report each, representing the collective views of their respective constituencies, highlighting their own specific wishes, demands and needs towards digital language equality.

Targeted stakeholder-specific surveys were designed for both groups. They were distributed to the relevant stakeholders through the responsible ELE partners (in Task 2.1 and, partially, Task 2.2). The surveys conceptualised and designed by the consortium have proven to be an

effective means for collecting valuable input. Especially the open questions have generated a lot of valuable feedback and food for thought. The online surveys were shared between June and October 2021.

The survey targeting in particular LT researchers and developers generated a a large number of responses, representing more than 200 different organisations and more than 30 countries. The survey was divided into four main parts and entailed 45 questions in total. 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 interviews and additional input collected through additional expert interviews can be found in the Deliverables D2.1 (Specification of the consultation process, Hajič et al., 2021), D2.2 (Report from CLAIRE, Thönnissen, 2022), D2.3 (Report from CLARIN, Eskevich and de Jong, 2022), D2.4 (Report from LT Innovate, Rufener and Wacker, 2022), D2.5 (Report from META-NET, Hajič et al., 2022) and D2.6 (Report from ELG, Hegele et al., 2022). In addition to the survey, expert interviews with selected representatives from the different initiatives like ELG and META-NET were conducted. They shared details on their work and related challenges, elaborating on how to do justice to all European languages, ways how to position European LT on a global level and the key challenges towards establishing a long-term LT programme.

A similar questionnaire was set up for the LT users. The consultation with LT users and consumers additionally aimed at collecting input for the analysis and comparison of the level of technological support for the EU official languages, including all minority, regional and lesser-used European languages. The survey encompassed 63 questions in total. The dissemination was driven through the leaders of the European initiatives ELEN, NEM, EFNIL, ECSPM, LIBER and Wikimedia as well as the consortium. This survey generated almost 250 responses. Similarly to the LT developers survey, numerous additional interviews were conducted for a more in-depth insight.

These surveys have shown that there is still a huge gap between the LT support for English and the 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. Basic research is still urgently needed. The fragmentation of the LT industry remains a challenge. At the same time, the last decade has seen progress on a larger scale than could have been imagined 10 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 consortium has made every effort to ensure that all voices are heard and taken into account for the SRIA and roadmap. With the support of social media campaigns and an agency specialising in survey dissemination, we were able to reach many EU citizens to hear their thoughts on how well they feel their languages are digitally supported. This EU citizens survey included 12 questions and was structured to take less than 5 minutes to fill in. It was translated into 35 languages to ensure a broad set of respondents from across the EU and representing a wide variety of language and cultural backgrounds. The survey has already provided valuable insights especially into how language technology support and gaps are perceived by lay people. The survey will remain open until early April 2022. At the time of writing, a full analysis is not yet available; the results will contribute to inform the SRIA.

Further, the ELE industry partners generated, in various focus groups, four technology reports to illustrate the demands, wishes and visions of the European industry in a structured way. These so-called deep dives have been compiled for the fields of Machine Translation, Speech, Text Analytics and Data and Knowledge, and they offer an in-depth up-to-date analysis of the respective areas.

With all the valuable insights collected during the past year, a well-informed and comprehensive SRIA and roadmap will be crafted in the remainder of the ELE project.

2 European Language Equality: Stakeholder Analysis

2.1 Developers of Language Technologies

European Language Technology (LT) developers are a diverse group of stakeholders, comprising *academic and industrial entities* in the field of LT/NLP – beyond research, they develop pre-commercial prototypes, algorithms, applications and systems. An initial grouping is, thus, *LT research* and *LT industry*.

In addition to the horizontal grouping into research and industry, a vertical categorisation can be performed with regard to the multi- and interdisciplinary nature of LT. LT is in the intersection of Linguistics/Computational Linguistics, Computer Science and Artificial Intelligence, while at the same time it encompasses methods and findings from Cognitive Science and Psychology, Mathematics, Statistics, Philosophy and other fields. As a result, the ELE stakeholder group of LT developers should be identified not only within the strict limits of *Language Technology* per se, but also in the neighbouring disciplines of *Artificial Intelligence* and *Digital Humanities/Social Science and Humanities* (DH/SSH), as partially visualised in Figure 1.

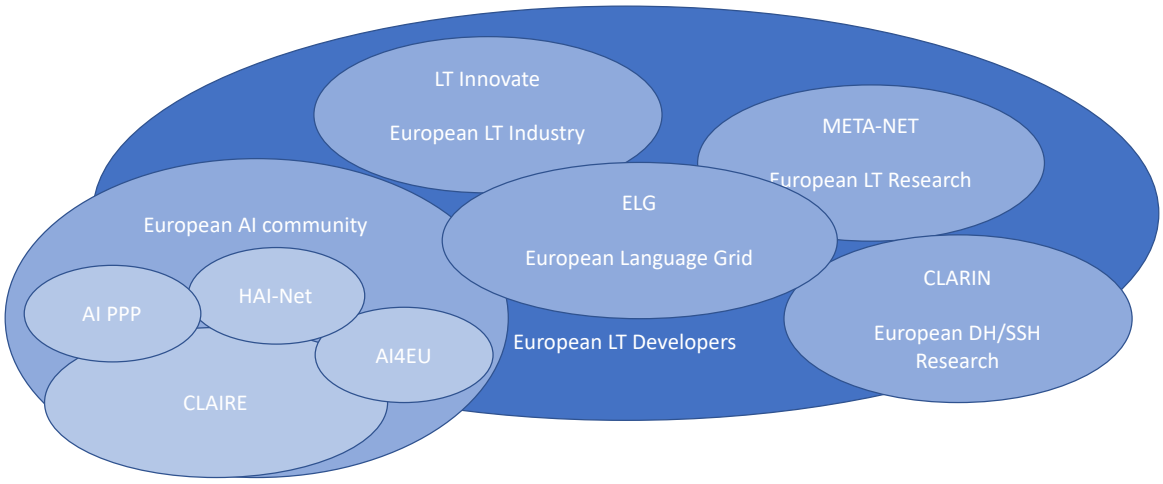


Figure 1: European LT developers communities (excerpt)

Companies

The European LT industry has been estimated to comprise 435 companies, according to LT-Innovate (2016) or 473 LT vendors in EU26 plus Iceland and Norway in 2017 (Vasiljevs et al., 2019). In March 2022, the ELG catalogue comprises more than 800 commercial entities, also including integrators and a certain number of user companies (Rehm et al., 2020, 2021).

Research Organisations

Europe has a long-standing research, development and innovation tradition in LT with over 800 centres performing excellent, highly visible and internationally recognised research on all European and many non-European languages.

2.2 Commercial or Academic Users of Language Technologies

The ELE project sought to bring together a diverse group of stakeholders deriving from interrelated research fields such as Linguistics/Computational Linguistics, Computer Science and Artificial Intelligence. Stakeholders from a diverse economic sector were also reached. The community of users and consumers that were addressed and reached in the surveys include mainly the following stakeholders:

Examples of commercial users that filled out the questionnaire are respondents representing companies in the sector of Information and Communication Technologies and eCommerce (e. g. Megabyte Ltd, A Capela group, Telecats) energy (e. g. Shell, Menai Science Park Ltd) and business services (e. g. Spencer Stuart, Inuits, Projectus grupa). We also considered users from the following groups:

- Self-employed language professionals (e. g. translators)
- Professionals working on different economic sectors (e. g. banking, health)
- Independent professionals/consultants
- Professionals working in public administration
- Media and publishing professionals

Academic users/research users

Academic users include:

- Researchers
- Data scientists
- University professors
- Language teachers
- Lecturers
- Master's and PhD students

Some non-governmental organisations were also represented in the survey such as Federal Lezghin National and Cultural Autonomy and representatives of public administration such as National Youth Service - Ministry of Education, Children and Youth (Luxembourg), Hungarian National Research, Development and Innovation Office and the Government of the Balearic Islands. It is also worth mentioning that Wikipedia partners collected responses from representatives of the various Wikipedia projects such as Wikimedia Community User Group Malta, Wikimedia Hungary, Wikimedia UK, Wikimedia Community Ireland, to name a few. The full list of stakeholders of the LT Users and Consumers survey is presented in Table 15, Appendix D.

European initiatives

The group of European users and consumers was represented by six main European initiatives, who disseminated the survey within their networks and produced one report each, based on their respective constituencies. These initiatives are listed below:

- The European Federation of National Institutions for Language (EFNIL)

- The European Language Equality Network (ELEN)
- European Civil Society Platform for Multilingualism (ECSPM)
- The New European Media (NEM)
- Association of European Research Libraries (LIBER)
- Wikipedia

In order to broaden the overall reach of the survey and complement and enhance the results obtained through the group of European initiatives listed above, 66 additional external stakeholders were also contacted by the ELE coordinator Dublin City University (DCU). The dissemination was carried out through emails to points of contact identified as leads of organisations that represent important communities of users and consumers of LTs. DCU also disseminated the survey through its communication and marketing channels that promoted the survey on social media posts between August and September 2021. The communication and marketing team also promoted the survey within their contacts in the Government department of the Gaeltacht.

2.3 European Citizens as Users of Language Technologies

In addition to the consultation with stakeholders that represent communities of users and consumers, a survey targeting European citizens was carried out with the aim of taking into account their opinions, individual needs, wishes and general demands as well as to make sure that their voices play a decisive role in the pursuit of full digital language equality supported by LTs. This consultation with a larger and more diverse cohort of users and consumers allowed us to obtain an accurate picture of the current scenario in terms of LT support across European languages and have a more representative basis for a technological and scientific forecasting on how LTs can be deployed and applied in Europe by 2030. The methodological approach and dissemination strategy of this consultation is detailed in section 3.2.3 of this report.

3 Stakeholder Consultation: Objectives and Methodology

In this section, the objectives of the process of all the consultations (surveys, interviews, deep dives) are described, together with the methodology that has been used for the various modes of soliciting the facts, figures, opinions and suggestions.

3.1 Objectives

The objective of ELE is the preparation of the strategic research, innovation and implementation agenda and roadmap for achieving full DLE in Europe by 2030. Since the goal of achieving DLE involves a large number of stakeholders, the process of preparing, discussing and finalising the different parts of the strategic agenda and roadmap is carried out together with all 52 partners of the consortium and the wider European LT community, which we communicate with via the consortium partners' many strong networks and connections.

This broad network has been assembled and approached in order to collect representative, structured, and broad and overarching input to produce a convincing, sustainable and evidence-based agenda and roadmap. The engaged stakeholders cover different areas of our core field of Computational Linguistics and Language Technology and also on the borders to other fields such as, among others, Cognitive Science, AI, Machine Learning, Data Science

and Knowledge Technologies. Specific focus has been on companies active in the field of LT: the involvement of industry makes sense in the grander scheme of things, especially regarding the inclusion of their needs and wishes and goals towards growth. The same holds for the non-industrial, but strong stakeholders as users and consumers, in the areas such as Digital Humanities/Social Science and Humanities (DH/SSH) research, policymaking, normative care of languages (including minority ones), education and others.

3.2 Setup and Methodology

In order to achieve the aforementioned objectives, two main stakeholder groups have been approached: LT developers (industry and research) and LT users as well as consumers. Both groups are represented in ELE with several networks, initiatives and associations who produce one report each, together with their respective constituencies, highlighting their own individual needs, wishes and demands towards DLE. The industry partners of the consortium have produced, in various (sub)groups, four technology deep dives to provide the needs, wishes, demands and visions of European industry, structured into Machine Translation, Speech, Text Analytics and Data. A large number of surveys and consultation meetings with stakeholders who are not represented in the consortium have been administered and described detailed reports. The methodology applied here is thus based on a number of stakeholder-specific surveys inspired by Rehm and Hegele (2018) as well as collaborative document preparation ("deep dives") that also involves technology foresighting. Both approaches are complemented through the collection of additional input and feedback through various online channels.

The two stakeholder groups addressed in the surveys differ in one substantial way: while the group of commercial or academic LT developers is, in a certain way, closed and well represented through relevant organisations, networks and initiatives in the ELE consortium, the group of LT users is an open set of stakeholders that is only partially represented, albeit very broadly - it includes, for example, areas such as Digital Humanities and Social Sciences, policymaking, normative care of languages (including minority ones), education etc.

Both stakeholder groups have been addressed with targeted and stakeholder-specific surveys that have been distributed to the relevant stakeholders through the responsible ELE partners. They have been asked to distribute the survey among their constituencies, but especially in the case of the consumers and users of LT, also to others. In addition, we have communicated with a large number of additional stakeholders, with shorter surveys and targeted interviews, mostly organized as online meetings.

While the stakeholder groups identified by ELE provide very structured, informed and consistent input from various relevant perspectives, it has been decided to additionally address also the European citizen, through a special, large scale (yet simple) survey mapping the awareness, views, experience and opinions of LT among the true end users of LT. This survey is still ongoing, with preliminary results reported here in Sect. 3.2.3.

3.2.1 LT Developers Survey and Interviews

With the aim of informing the ELE strategic agenda and roadmap with the opinions, views and demands of the widest possible, directly or indirectly involved group of stakeholders, we engaged with the research community and industry representatives and, with regard to their field, with LT proper, AI and DH/SSH communities. We mobilised existing European networks, associations, initiatives and projects. Some of the well-established and long-standing pan-European LT networks are represented in the ELE consortium and they constitute the core ELE LT developers stakeholders groups (see Section 4). The ELE partners that represent

these initiatives not only contributed their views to the project but they also facilitated access to and elicitation of the views of their constituency and members. In particular, they coordinated the distribution of the survey to their members, they conducted interviews and focused consultation meetings, where needed and appropriate, and they consolidated their feedback in Thönnissen (2022); Eskevich and de Jong (2022); Rufener and Wacker (2022); Hajič et al. (2022); Hegele et al. (2022).

The survey encompassed 45 questions in total. A respondent was presented with 32 (minimum) to 45 (maximum) questions, including the "if other" questions. 35 questions were mandatory and 27 were closed questions (single or multiple choice) (see Table 1).

	Mandatory	Optional	Total
Closed	24	3	27
Open-ended	2	16	18
Total	26	19	45

Table 1: Types of questions in the LT developers survey

The survey was structured in four main parts:

- **Part A. Respondents' profiling**: The first part of the survey included 13 questions for the demographic profiling of respondents, with an emphasis on characteristics relevant to the task at hand, i. e.
  - Country
  - Affiliation
  - Type of organisation
  - LT areas that the respondent is mainly active in
  - Participation/membership in networks/associations
  - Sectors/domains that the respondent is active in (if relevant)
- **Part B. Language coverage**: The second part investigated the degree of coverage of the European languages by the respondents' current research and development activities, i. e.
  - languages currently supported in research/products/services
  - languages planned to be supported in the short-/medium-term
  - factors that influence the respondents' decision with regard to language coverage/support
- **Part C. Evaluation of current situation**: This part included questions that sought to elicit the respondents' evaluation of the current situation of LT research and development, the strengths, gaps and challenges that the European LT community is facing, i. e.
  - gaps in terms of: a) technologies, b) tools/applications, and c) resources, especially with regard to specific languages
  - LT areas where the European LT community excels
  - main perceived challenges and obstacles that should be overcome

- **Part D. Predictions and visions for the future**: The fourth part of the survey is the forward-looking section that investigated ideas, predictions and wishes of the LT community about how the LT field as a whole will be able to equally support all European languages by 2030, i. e.
  - policies/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 asked the respondent's permission to be contacted again for a follow-up interview and, if so, his/her contact details.

The survey was designed, set up and published for circulation on the EU Survey platform.[1] The full survey, as published online, is presented in Appendix A. To supplement the survey responses and to collect more detailed feedback, but also with the aim to further engage with the LT community and to spread the message about ELE's mission, a number of consultation meetings with targeted informants were conducted.

The interviews were designed and executed by the ELE partners who acted as deliverable leads for D2.2 to D2.6 (reports from the networks of stakeholders). The interviewees were selected based on either the quality of their input to the survey or their prominence in and impact on the European LT landscape. The candidates for an interview were additionally selected to represent mainly the industry, but also research and some European Institutions, to cover as many European countries as possible and different LT areas in terms of their expertise.

The interview script flexibly followed the structure and contents of the LT researchers and developers survey, adapted as appropriate to the background of the interviewee and to the dynamics of the discussion.[2] In total 61 interviews were conducted from November 2021 to January 2022 mainly through 30 to 60 minutes phone calls or video conferences, while some interviewees submitted their answers in writing.

3.2.2 LT Users and Consumers Survey

The LT user and consumer group consisted of professionals and communities who use LT on a regular basis. The consultation with aimed at collecting input for the analysis and comparison of the level of technological support for the EU official languages and minority, regional or lesser-used European languages.

The survey was disseminated through the leaders of the six European initiatives, namely, ELEN, NEM, EFNIL, ECSPM, LIBER and Wikipedia (see section 2.2), who brought together diverse groups of stakeholders including researchers, representatives of communities of LT users and consumers, language professionals (e. g. translators, lecturers and professors in the field of Linguistics and Computational Linguistics) and stakeholders from various economic sectors (e. g. banking, health). The data collected through this survey will serve as input for the preparation of the Strategic Research and Innovation Agenda (SRIA) and Roadmap to be produced by the ELE project, in order to tackle the striking imbalance between European languages in terms of the support they receive through LTs by 2030.

[1] https://ec.europa.eu/eusurvey/runner/ELE-LTdevs

[2] The interviews conducted with LT-innovate members divert from the common methodology applied throughout all other ELE deliverables with respect to the set of questions used (Rufener and Wacker, 2022).

The survey was designed, set up and published on the EU Survey platform.[3] The full survey, as published online, is presented in Appendix C. The data was collected between 21st of June 2021 and 17th September 2021.

The survey addressed to LT users and consumers 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. The LT Users and Consumers survey encompassed 63 questions in total. Some of the questions depended 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, and 33 of them were closed questions (single or multiple choice). Table 2 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 2: Type of survey questions

The survey was structured in four main parts, Part A, Part B, Part C and Part D. If none of the answers provided as options were applicable, then the respondents had the option to enter a different answer through the option "if other, please specify".

- **Part A. Respondents' profiling:** the first part of the survey included 13 questions for the demographic profiling of respondents, with an emphasis on characteristics relevant to the task at hand, i. e.,
  - Country in which respondents were based
  - Name of the organisation/representative body respondents work for
  - Communities they represent (if applicable)
  - Type of organisation respondents work for
  - Sectors or domains in which respondents are active (if applicable)
  - Role of respondents in the organisation (if applicable)
  - Organisations' 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 of 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 languages, i. e.,
  - Differences in availability of LTs between the official European languages they work with and, if applicable, differences in availability of LTs between 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

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

  - Respondents' opinion in relation to performance of LTs with regard to specific languages

- **Part D. Predictions and visions for the future:** respondents were requested to share their needs and wishes for the future of LTs, 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 a follow-up interview and, if so, 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"; see Appendix, p. 63 ff.

3.2.3 EU Citizens Survey

Before setting up the survey, different survey platforms were tested to choose the most suitable platform for ELE survey needs. The *QuestionPro*[4] platform was chosen as it offers the features needed for the structure of the survey logic. After setting up the survey on *QuestionPro* it was disseminated in 28 European countries and in 38 European languages from January 2022 to March 31st 2022.[5] The survey included a total of 12 questions, 6 of which were single-choice questions, 5 were multiple-choice and 1 open-ended question which allowed respondents to include any comments or feedback they had. These 12 questions could be answered in approximately 5 minutes via computers or mobile devices.

The first question of the survey aimed at understanding the level of familiarity of respondents with terms from the field of LTs. This information was collected through a multiple-choice question that asked respondents to select the terms (e. g. *Information Retrieval*, *Natural Language Processing*, '*Natural Language Understanding*) that they were familiar with or could immediately recognise.

The remaining 11 questions were distributed into four main groups with a structure similar to the setup for LT Users and Consumers survey described in section 3.2.2. The complete survey can be found in Appendix E. The groups of questions are described below.

- **Respondents' profiling**. The first part of the survey included 3 questions for the demographic profiling of respondents, including country of residence, age group and education level.
- **Language coverage**. This part included 1 multiple-choice question to find out which European languages respondents used both socially and professionally.
- **Evaluation of the current situation**. This section of the survey was designed with 5 questions to capture information regarding the differences in availability of LTs between the official European languages and European minority, regional, lesser-used languages; gaps perceived in the technologies, tools/applications respondents use with respect to specific languages as well as respondents' perceptions in relation to the performance of LTs with regard to specific languages.

[4] http://www.questionpro.com

[5] Note that the preliminary results presented here are based only on data collected up until 29th March, as the survey remains open past the delivery of this report.

- **Future of LTs in Europe**. In this section, 2 questions requested respondents to share the tools they would like to use in the future if not currently available in their languages and also to rate the top 3 advantages of improving LTs for all languages.

The first round of the dissemination process was carried out through Lucid's survey solutions[6] which offers online market tools and access to a large community of respondents world wide. For each country we created a standalone survey so that respondents only saw the version in the language of the country in which they were based. For countries with more than one official language, we created a standalone version of the survey in each language spoken in the country. In Spain, for instance, four different surveys were setup as they were disseminated in four languages, namely, Spanish, Catalan, Galician and Basque. Setting up separate language versions of the survey for multilingual countries allowed us to specifically target regions in the country where we were more likely to find communities of respondents that were speakers of that language.

We collected a total of 18,850 responses through Lucid's services. These responses were divided into quotas established by country, to ensure that the responses collectively provided a fair representation of European citizens. The quota established for multilingual countries was divided between the languages spoken in the country. The sample size established per language was based on the size of the population speaking that language in the country.[7] In Spain, for instance, as Spanish is the most widely spoken language, 83% (750 responses out of 900) of total the quota was setup for the survey disseminated in Spanish, while the remaining 17% was distributed into the smaller regional languages in Spain such as Basque, Catalan and Galician. Table 3 shows all countries where the European Citizens Survey was disseminated, the sample size per country and the sample size per language.

For countries and languages not covered by Lucid's services, the survey was disseminated with the help of ELE partners as, through their networks, they could target the communities of speakers of these languages. The countries not covered by Lucid included Luxembourg, Macedonia, Malta, Turkey, Iceland, regions of Ireland with larger amount of Irish speakers and Bosnia.

Survey Translation

The survey was first set up in English and it was automatically translated using the eTranslation tool[8] and post-edited by native speakers from the ELE consortium into 42 languages. [9]. After setting up the translated versions on the QuestionPro tool, the same native speakers were requested to visualise the survey translated on the tool and provide a final review for translation quality assessment. Most of the relevant languages (38 languages out of 42) are presented in Table 3. The six additional surveys were set up and translated into Bosnian, Icelandic, Luxembourgish, Macedonian, Maltese and Turkish languages. These additional surveys are not presented in Table 3 as they were not covered by Lucid's services but instead disseminated via consortium partners, who, at the time of writing are promoting the survey via their networks and their social media channels.

[6] https://luc.id
[7] Guidance on sample size to this effect provided by Lucid based on their previous campaigns.
[8] https://ec.europa.eu/digital-building-blocks/wikis/display/CEFDIGITAL/eTranslation
[9] While ELE covers 85 European languages, we only produced translated versions for those languages for which native speaker post-editing was available. The 42 languages covered represent the support offered through the ELE consortium members.

Countries	Total Sample Size	Language(s)	Sample per Language
Austria	900	German	900
Belgium	900	French	350
		German	50
		Flemish Dutch	500
Bulgaria	750	Bulgarian	750
Croatia	600	Croatian	600
Czech Republic	900	Czech	900
Denmark	600	Danish	600
Estonia	150	Estonian	150
Finland	300	Finnish	250
		Swedish	50
France	900	French	900
Germany	900	German	900
Greece	900	Greek	900
Hungary	900	Hungarian	900
Ireland	550	English	450
		Irish	100
Italy	900	Italian	900
Latvia	200	Latvian	200
Lithuania	300	Lithuanian	300
Netherlands	900	Dutch	900
Norway	600	Norwegian	600
Poland	900	Polish	900
Portugal	900	Portuguese	900
Romania	900	Romanian	900
Serbia	100	Serbian	100
Slovakia	550	Slovak	550
Spain	900	Spanish	750
		Catalan	50
		Galician	50
		Basque	50
Sweden	900	Swedish	900
Switzerland	400	French	150
		German	200
		Italian	50
United Kingdom	1000	Welsh	100
		English	900
Slovenia	250	Slovenian	250

Table 3: Sample size per country and language

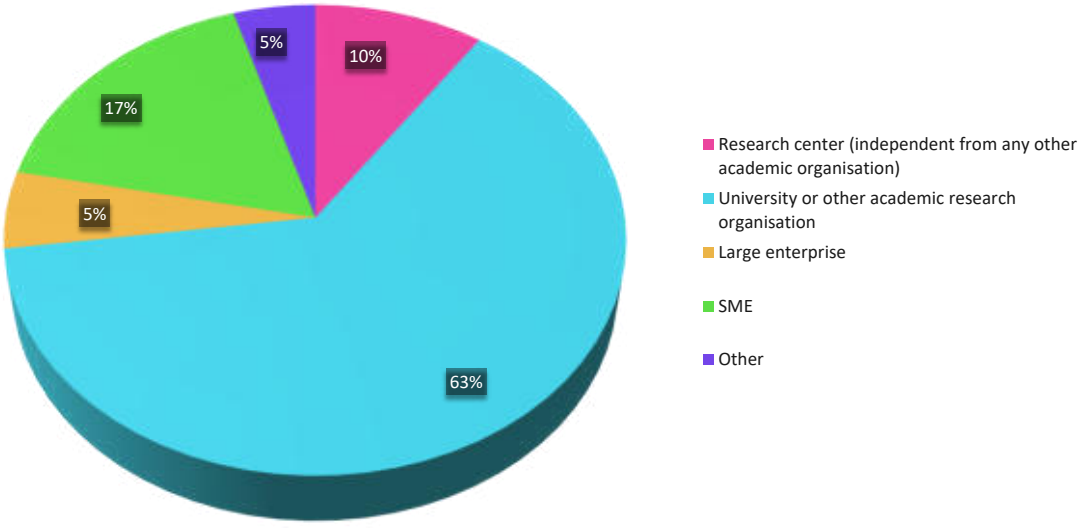


Figure 2: Type of organisation

4 Results of the Stakeholder Consultation

In this section, the results of the consultation process (Sect. 3) performed in cooperation with the large number of stakeholders (Sect. 2) are described.

4.1 Developers of Language Technologies

4.1.1 Respondents' Profiles

One major goal of the LT developers survey was to bring the European LT community together and hence reach a wide and demographically distributed audience. The LT Developers survey was filled in by 321 different respondents[10] who represent 223 different organisations (see Appendix B, Table 8). 73% of the organisations are research or academic institutions and 22% are private companies (Figure 2). In 5% of responses the "Other" value was indicated as the type of organisation and this has been further specified as freelancer/private practitioner or currently unemployed, government agency, not-for-profit organisation, etc.

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, Russia, Serbia, etc., but also other global regions, e. g. Brazil, the USA and Israel. Most responses were contributed from 1) Spain, 2) Germany, 3) Greece, 4) the Czech Republic, and 5) The Netherlands (see Appendix B, Table 9). 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 very well represented in the survey responses with approximately 40 to 90 respondents each. ELG, ELE's sister project, is repre-

[10] In total, 333 responses have been collected. However, some respondents filled in the survey twice. These respondents are not taken into account in this analysis.

sented 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 approximately 10 to 25 survey respondents each (Table 4). Additional networks, associations and projects indicated by the respondents include the ELRC initiative, ELRA, ACL, EAMT, DARIAH and many more (see Appendix B, Table 10 for the full list of entities).

Initiative	Answers counts	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 Network	11	–
Nexus Linguarum	25	–
TAILOR	9	–
Other	31	–
None of the above	115	–

Table 4: Main networks, associations or projects the respondents participate to.

The respondents are mainly active in the following LT areas (by order of frequency): 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: data production, data aggregation (Figure 3). When the value “other” was selected, the respondents had the possibility to further specify their areas of expertise as “NLP in Education”, “corpus construction”, “language modelling”, “text generation”, “sign language”, etc.

The technologies, products or services offered by the respondents’ organisations are used in a number of diverse domains, a finding that demonstrates the applicability of LT in practically all economic sectors. The top 3 domains indicated by the respondents were 1) Information and communication technologies, 2) Digital humanities, arts, culture and other services and 3) Education. See also Appendix B, Table 11 for a detailed list of all sectors.

4.1.2 Language Coverage

The respondents listed a wide range of different 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, regional and/or co-official European languages. The 5 most frequently mentioned languages are English, German, Spanish, French and Italian. Figure 4 shows some of the European languages addressed by ELE and the number of times they were mentioned by survey respondents.

In addition to the closed list of European languages, 80 respondents indicated “Other” languages they support in their products or research. To name but a few, these other languages include 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.

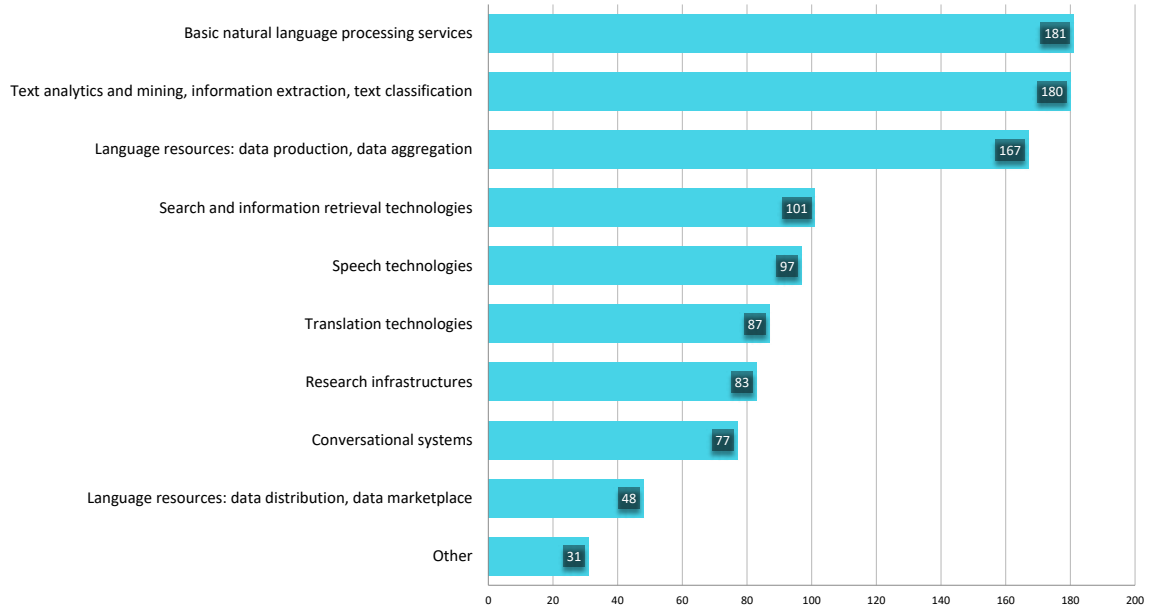


Figure 3: LT areas in which the respondents conduct research or develop tools and services

To get an idea of the current gaps and the focus of future work, 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 or minority languages, such as Basque, Catalan, Breton, Mirandese, Romani, Aromanian, etc. 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, Spanish varieties, etc. (see Appendix B, Table 12 for the full list of languages).

When considering the top three drivers for the decision to support additional languages, the most frequently selected factor is research interest (212 mentions), followed by the availability of language resources (144) and market interest (138). As expected, the prioritisation of these factors is different when the type of organisation the respondent represents is taken into account. For the industry (including large enterprises and SMEs) market interest and demand by users or consumers play a pivotal role, while the availability of language resources follows at a distance. For research organisations and SMEs, more than big organisations, funding and investment opportunities are also to be considered. When the "Other" value has been selected, this was in most cases specified in the subsequent open-ended question 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 underrepresented language communities to work towards the knowledge equity goals*".

4.1.3 Evaluation of Current Situation

In order to evaluate the current situation and to further grasp the main challenges and obstacles the European LT community faces, the survey participants were asked to indicate their level of agreement with a set of statements. Respondents were also given the opportunity to elaborate on the obstacles and challenges indicated in the questions and/or add any other obstacle/challenge not previously listed as part of a free text question. A detailed list of all

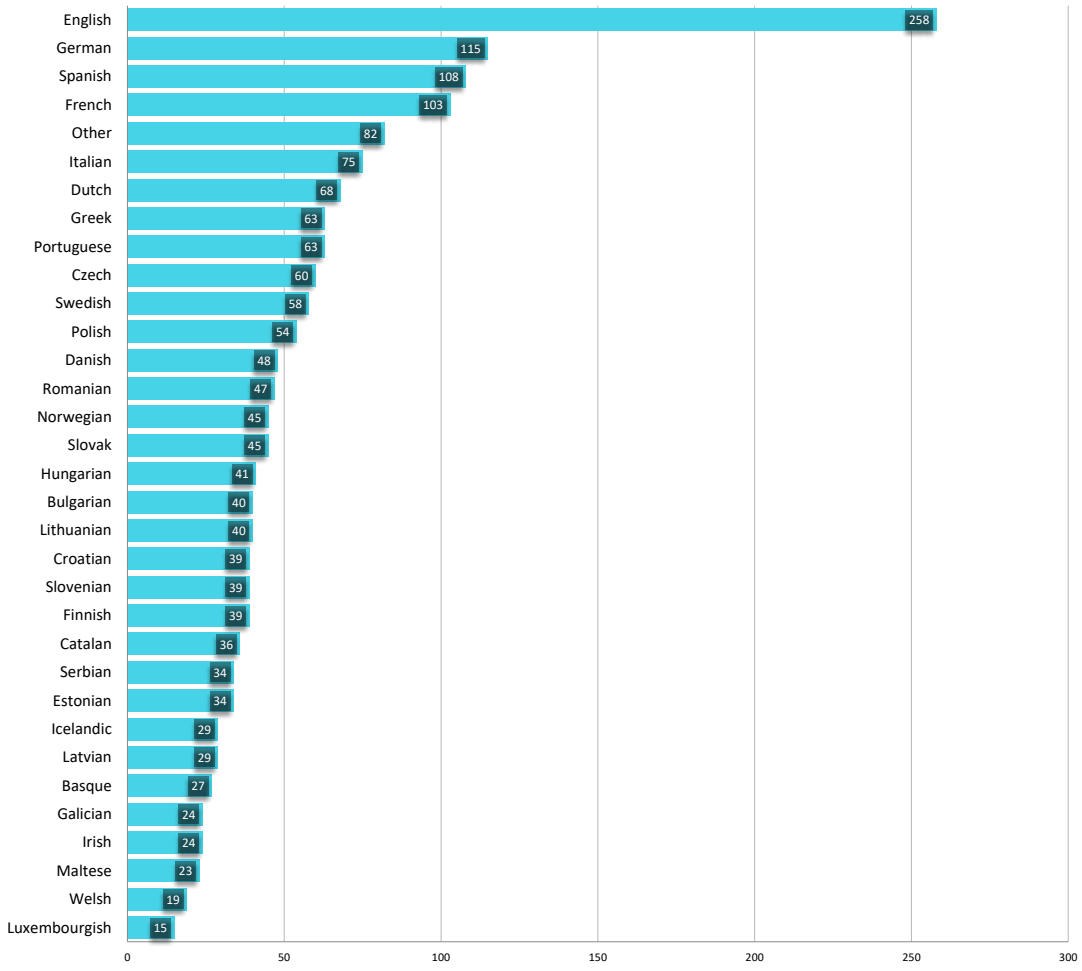


Figure 4: Languages supported by the respondents' organisations in their research and development activities

answers can be found in Appendix B (Table 13).

Overall, there was almost uniform agreement that basic research is still necessarily needed: 88% or the respondents agree or strongly agree with this statement (Figure 5). According to the participants, the focus on the development of new technologies shifts attention from fundamental research, even though this type of research can provide a solid base for future technology development. While in recent years some advanced technology opened by mostly non-European commercial players such as Google, Microsoft, Amazon and Facebook propelled LT research, and fuelled advancements across the NLP and AI fields, the next steps of more complex tasks cannot rely only on a "throw more data at it" approach and they certainly cannot rely on research in algorithms and machine learning. Support for basic research in Linguistics and in other DH/SSH disciplines is also urgently needed.

The next two statements, each perceived as an obstacle by 82% of the respondents, concerned competition with non-European big companies and market disruption by global players as well as the fragmentation of the European LT industry. The fragmentation of the LT industry which has been the subject of various reports and research agendas of the last decade

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

Table 5: Mentions (by organisation type) of the top drivers for the decision to support additional languages

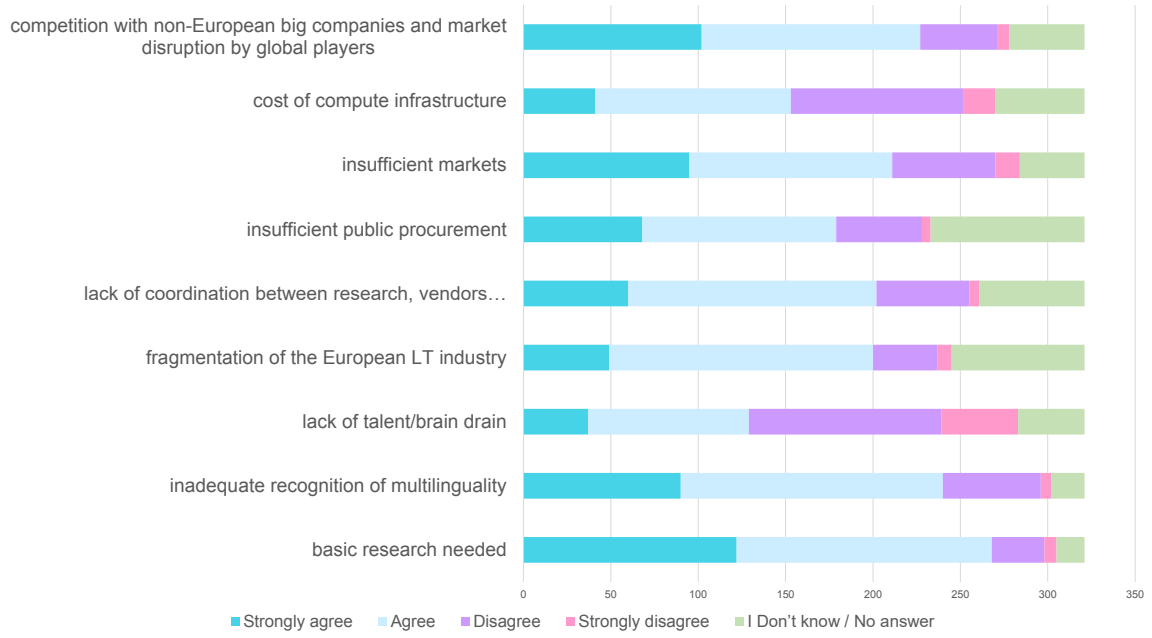


Figure 5: Perceived challenges the European LT community currently faces

(Rehm and Hegele, 2018), is still a non-solved obstacle. Individual answers emphasise that we need to establish conditions that create strength out of many small players and provide a dynamic ecosystem for growth. In this context, a lack of coordination and missing links between research, LT vendors, integrators and customers is also visible for the large majority of participants.

At the same time, the importance of multilinguality in the European landscape does not always get adequate recognition (79% agree or strongly agree in this regard), and the smaller languages appear not to be attractive enough for the industry and investors (74% agree or strongly agree on this point). Very often one can rely on public funding only (practically) to work on the smaller languages, as there is no market for such LTs. These public investments for small languages are necessary on a larger scale to really make them available to the wider community. The cost of developing LT for a language is usually constant, regardless of the number of speakers of that language. Even more, for languages with larger numbers of speakers, the LRs could be collected in an easier manner: for instance, the larger the number of speakers, the more online-collectable text is produced in a day. Industrial players can find a commercial interest in pre-competitive investments for "larger" languages, while this will rarely be the case for "smaller" ones. In that situation, the role of additional investors for the development of LT for "smaller" languages should be played by bodies either at national or EU level. This situation is even worse for non-standard languages: local dialects, non-standard written language on social media platforms, non-standard language for speech recognition, and non-standard language as used by migrants or citizens with a migration background. There are almost no calls to fund work on creating language resources for training models or the research into these languages. There is equally little incentive for researchers to publish their work on small languages, resulting in the dominance of the English language in scientific literature.

Frequently discussed when it comes to research in Europe is the observation that many early stage researchers are attracted by non-European companies which offer higher compensation. However, the claim that brain drain is a very pressing problem was not shared by the respondents of this survey. More than 50% disagreed or strongly disagreed with the respective statement. An interesting point that was mentioned is that the risk of lack of talent in LT will be evident if we stop educating new generations in LT and go for ML first, i. e. if we equate LT with Machine Learning (ML). LT is just one of many hard challenges for ML. Conversely, ML is just one of many tools/methods in LT. But if it is considered that LT development can be achieved with ML only, we'll reach to a point where the generic ML methods used now cannot take the state of the art in LT any further.

The cost of access to computing infrastructure also received one of the lowest agreement scores (only 57% agreed that this is an obstacle), although in some comments the limited computing resources of universities and research centers as opposed to those of large companies have been highlighted. Easy access to massive (quality) data and HPC, currently available almost exclusively to big tech companies, was considered as crucial by some respondents who argued that, especially with the emergence of large-scale pre-trained models, access to HPC by research organisation and SMEs is critical in order for them to stay up-to-date with the latest developments.

The insufficiency of public procurement is also considered an obstacle by a considerable number of respondents (77% agree or strongly agree). It was specifically underlined in this context that funding application procedures for small companies need to be facilitated.

4.1.4 Predictions and Visions 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. A detailed list of all answers can be found in Appendix B (Table 14). In addition, respondents were given the opportunity to elaborate on other policies/instruments not listed as part of the question, but 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 digital language equality, as brought up in multiple answers, is the availability of resources. By 2030 all European languages should have developed the critical mass of resources that are needed for developing LTs. These include not only raw data, but also massive 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 in the coming years 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 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 LT where the leap from 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, as presented in Appendix B, Table 14. A long-term programme of 10 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/technological) methodologies for transfer/adaptation of resources/technologies to other domains and languages is considered an effective measure to boost digital readiness of less supported languages (average score 4.0). Given the importance of a strong foundation in basic research as discussed above, it does not come as a surprise that a large majority of over 86% of respondents welcomes an increase of the availability of qualified LT personnel and incentives for talent retention. That also includes reinforcing training and education initiatives, including undergraduate and masters programmes and vocational training in LT.

A number of elaborate open 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 a constant returning issue among the answers. With respect to the beneficiaries of funding, a number of survey 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, and special emphasis is put on the limitation of bureaucracy in application procedures which introduces considerable overheads for small companies.

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 a 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 employing a 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/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 criticized for its bias towards publications on a small number of the world's languages. Raising awareness of digital equality issues in the 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 multi-language 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 factor 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 in Europe. By supporting its linguistic diversity, Europe can achieve digital self-determination and sovereignty.

4.2 Users of Language Technologies

The LT users and consumers survey brings together diverse groups of stakeholders including representatives of communities of LT users and consumers, academic stakeholders, 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 emails by all six relevant ELE partners, namely, ELEN, LIBER, ECSPM, NEM, EFNIL and Wikipedia, but dissemination through social networks was also carried out. The leaders from the six European initiatives promoted the survey within their networks targeting representatives of organisations and communities of users and consumers. Through this campaign, 125 responses were collected. In addition, as a partner and coordinator of the project, the DCU team also worked on the dissemination which

promoted the survey within its networks as well as through regular posts on social media disseminated from August to September 2021. DCU collected a total of 14 responses from researchers affiliated to its School of Computing in addition to research staff based in Trinity College Dublin and University College Dublin, while the 107 remaining responses were collected through social media and word of mouth.

Table 6 shows the breakdown of responses collected through the survey by all the European initiatives and DCU along with the number of interviewed stakeholders per ELE partner.

Initiative	Answers counts	Interviews
ECSMP	10	2
EFNIL	28	6
ELEN	7	19
LIBER	29	3
NEM	29	6
Wikipedia	22	3
DCU	14	0
Other (e. g. social media)	107	–
Total	246	39

Table 6: Number of survey responses collected through the European Initiatives and interviewed stakeholders

4.2.1 Respondents' Profiles

The survey obtained a total of 246 responses. The results show that contributions came from a diverse range of economic sectors and professional activities, but most of the respondents 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 NGOs, large enterprises, SMEs, government departments and independent contractors and consultants in diverse economic sectors. The 15 (6%) respondents who selected the option "other" represented non-governmental bodies, non-profit organisations, public sector organisations, social organisations and independent government departments.

Figure 6 displays the breakdown of types of organisations that were represented in the survey responses.

Contributions to the survey came from all over Europe and, due to social media sharing, some responses from people based outside European countries such as the United States, Democratic Republic of Congo and Russian Federation. In Europe, the most represented countries were Croatia (33 responses) , Spain (23 responses), UK (23 responses), Ireland (17 responses), Germany (16 responses) and France (14 responses). Table 16 presents the complete summary and detailed statistics of the countries represented in the survey. Figure 7 illustrates the all the countries in which respondents were based.

4.2.2 Language Coverage

The survey indicated that 74% of the respondents work with English, which is the dominant language followed by a well balanced group of languages composed by German with 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

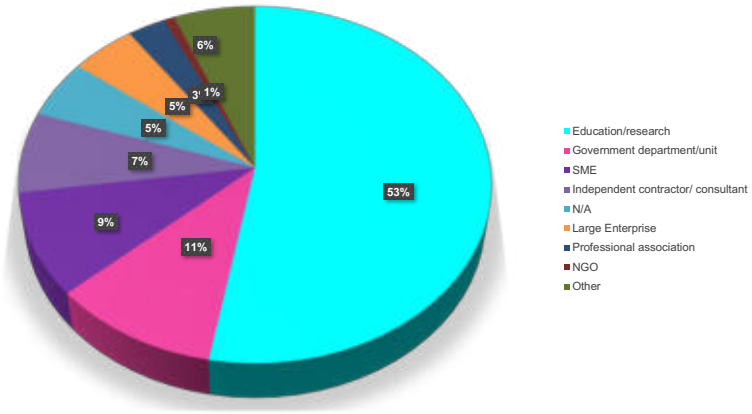


Figure 6: Type of organisation

as few respondents (between 1 and 3) indicated that they work with them. In relation to respondents that selected the option "other", they 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 8 shows the breakdown of European languages the respondents work with.

Table 17 in Appendix D presents the complete summary and detailed statistics of all the languages represented in the survey.

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. Thus, again, the survey shows the English predominance over all languages followed by German, Spanish and French. Other official European languages were mentioned by few respondents (between 2 and 3 respondents only) such as Italian, Portuguese and Greek as well as some minority, regional, lesser-used languages such as Breton, Catalan, Faroese, Valencian dialect but only one respondent each. Taken together, 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 and engagement.

4.2.3 Evaluation of the Current Situation

Respondents had the chance to evaluate, based on their experience, the level of technological support for the languages they work with, through on a 4-point Likert scale (where 1 indicated very poor support, 2 poor support, 3 good support and 4 excellent support). The list of language technologies evaluated can be seen in page 5 of the full survey presented in Appendix C. The results show striking differences in technological support between European languages.

Unsurprisingly, that English is very well supported with a mean score of 3.4, while the group formed by German, French and Spanish follows with a mean score between 2.4 and 2.5. All other European languages were considered to have either poor support (mean scores ranging from 1 to 1.3), very poor support or no support at all with scores below 1. Figure 9 shows the average score for each of the European languages evaluated.

When respondents were asked to select the LT tools used in the EU official languages they work with, the results show that, in EU official languages, translation tools account for 68% of the total responses, proofing tools 56%, search engines 55% of responses and language

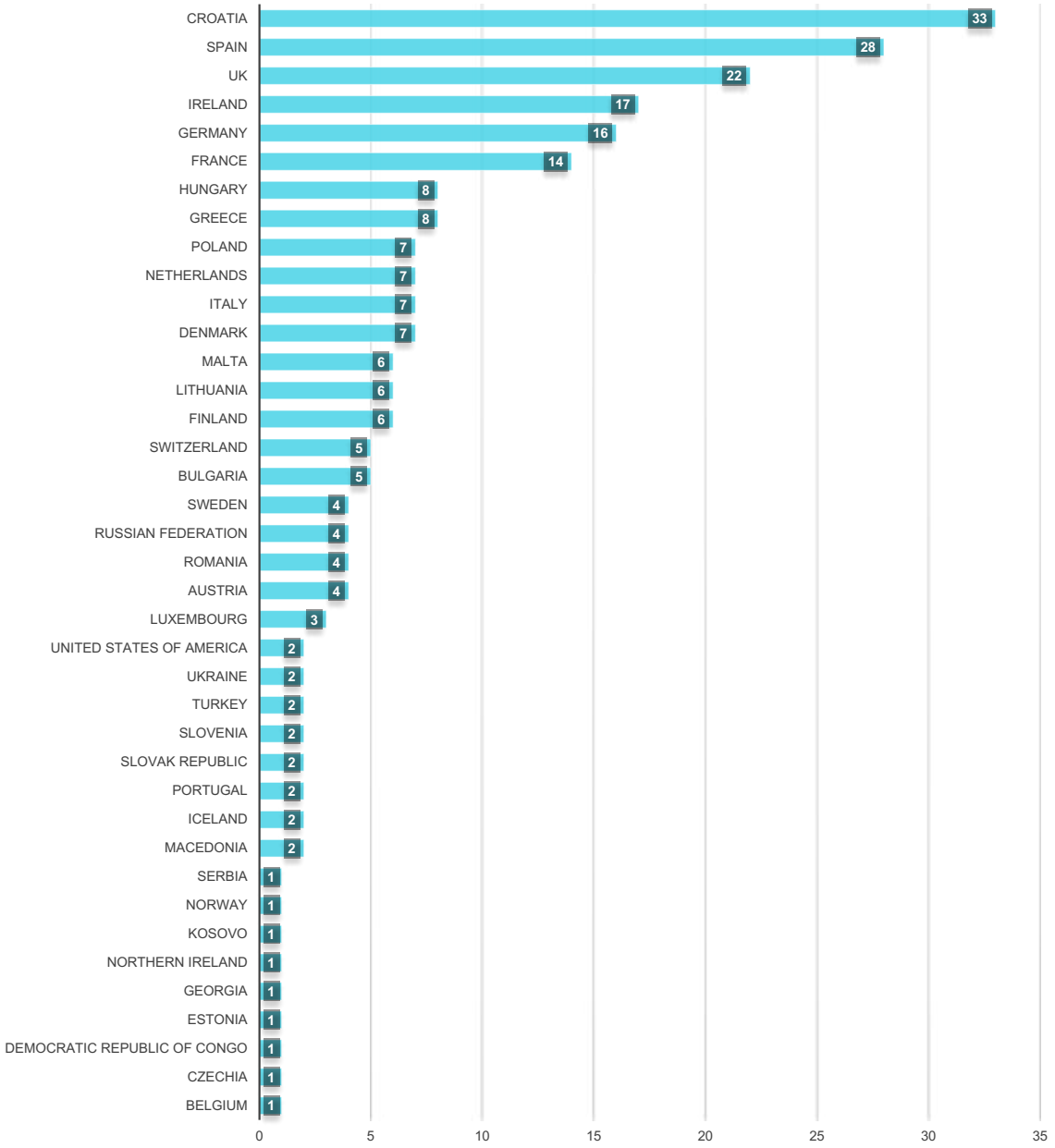


Figure 7: Countries in which respondents were based

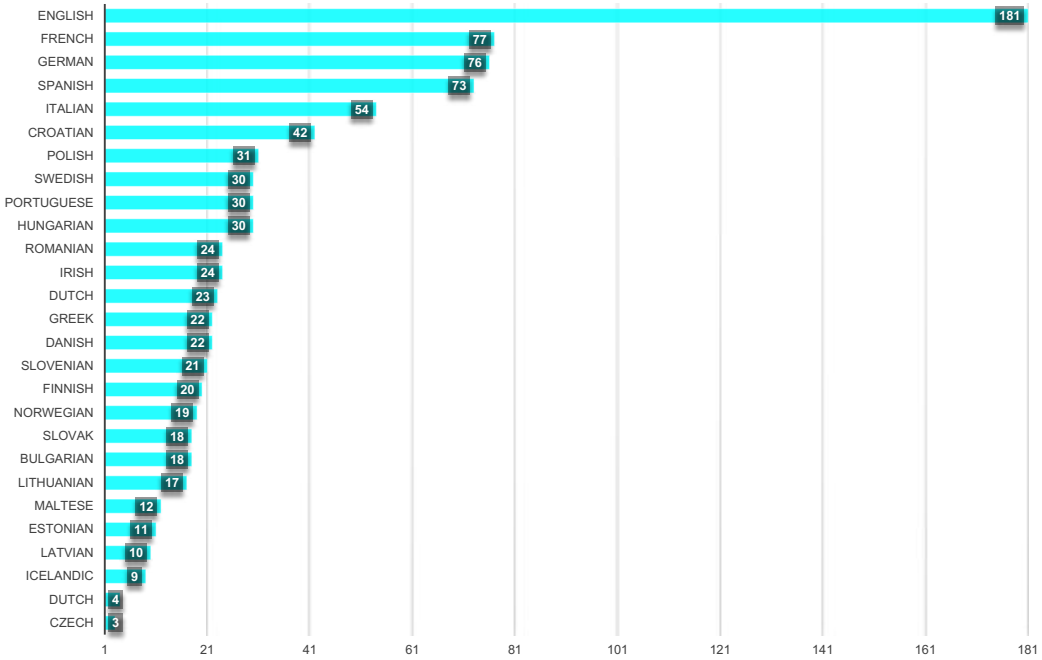


Figure 8: European languages respondents work with

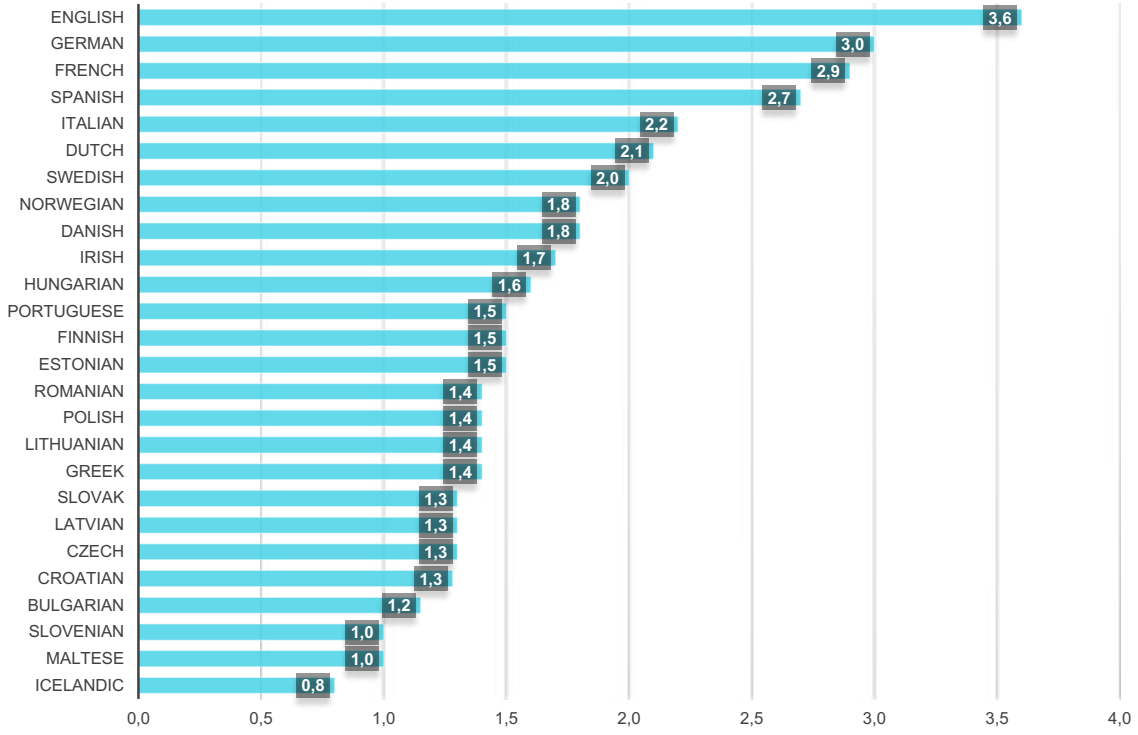


Figure 9: Level of technological support average scores for the European language(s) respondents work with

learning tools 39%. Similarly, the most used LT tools in minority, regional, lesser-used languages are translation tools representing 88% of the LTs selected by the 49 respondents that work with these languages, followed by proofing tools with 74% of the answers selected and search engines with 63%. These results suggest that, although LT tools are available for use in minority, regional, lesser-used languages such as Welsh, Basque, Scottish Gaelic, search engines are less likely to be used with them. This is possibly due to the poor performance of search engines in these languages. Based on a 4-point Likert scale (where 1 indicated very poor performance, 2 poor, 3 good and 4 excellent) the performance of search engines was considered by respondents good (average score is 3.08) in EU official languages while is considered poorer in minority, regional, lesser-used languages (average score is 2.3). Concerning the performance of other LT tools, again, there are clear differences between scores given to official EU languages on the one hand, and minority, regional, lesser-used languages on the other. The average score for proofing tools and translation tools indicates that these types of LTs are poor in general for minority, regional, lesser-used languages with scores below 2.5 (2.2 and 2.0 respectively) while for EU official languages the scores are above 2.5, on average, 2.8 and 2.6 respectively. For speech technologies the performance is evaluated as very poor on average for minority, regional, lesser-used languages (0.75), but the average score increases for official EU official languages (1.75).

A closer analysis of the responses collected via partner DCU revealed that most of the respondents, i. e. 71 out of 122 (58%), perceive gaps in the LT tools they use. The most common gaps perceived are in relation to the amount and variety of applications available. Within this group of responses, this gap was more frequently perceived by respondents working with LTs in Estonian (100% of respondents), Maltese (86% of respondents), Latvian (83% of respondents), Bulgarian (72% of respondents), Czech (67% of respondents), Slovak (58% of respondents), Irish (56% of respondents) and Romanian (50% of respondents). In contrast, for English, this gap is only perceived by 4% of respondents, German 10%, French 10%, Spanish 11% and Italian 14%. Gaps in the quality of available applications were more frequently perceived by respondents using LT tools in Icelandic (50% of respondents), Maltese (43% of respondents), Croatian (38% of respondents) and Bulgarian (36% of respondents). Gaps in quality of the tools are less perceived by respondents using LT tools in Italian (3%) and English (5%). Gaps in the variety of linguistic phenomena covered by the tools were perceived by 50% of respondents using them in Icelandic, 43% in Maltese and 39% in Irish, but this gap was only perceived by 1.9% of respondents for English, 6% Swedish and 8% Italian.

It is clear from the responses to the open-ended question that the LT users and consumers community wish to increase the variety of tools and resources available for minority, regional, lesser-used languages. Responses show, for instance, a desire for localised social media such as Twitter and the availability of personal assistant tools (e. g., Alexa or Siri) in regional languages such as Basque and Catalan. Another crucial gap in LTs that was pointed out is the limited adaptability of speech technology tools to the most common operating systems such as Android and iOS, which only allow users to use devices developed by these companies. Thus, software that has been developed by other companies and that supports languages not served by Android or iOS cannot be technically integrated with them. In addition, some responses also reveal that improved LT support for disabled people is required. On this topic, survey results also show the social dimension of LTs that developers should be aware of, and sensitive to, when developing tools and services.

The answers to this survey show that raising awareness for the LT potential in Europe on a political level is more important than ever before. The European LT community is in a place where change is needed in order to compete with innovative systems and tools built elsewhere in the world. On a political level, this involves more commitment from the European institutions as well as those of the Member States. This observation raises the need for legal measures to ensure the open and flexible integration of LT services and tools with the most widely used operating systems.

4.2.4 Predictions and Visions for the Future

Respondents also had the chance to express their opinions in relation to the provision of resources that would increase the use of LTs. The answers option provided were *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* and *Other*. Results show that 79% of respondents agree that higher quality of tools is crucial to increase the use of LTs as well as a wider range of tools for the languages they work with (61% of the selected answers). 48% of respondents think that more training of personnel dealing with such tools would help.

In relation to their predictions for the future, the range of opinions was broad. In general, most respondents (68%) are confident that in the next 10 years there will be higher-quality tools for all European languages including minority, regional, lesser-used languages and that there will also be a wider range of language tools for 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 minority/regional languages 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. Other benefits were also mentioned in the open-ended questions. Among these benefits, 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 enhance its visibility. Another respondent pointed out that LTs can improve online and print publishing in minority, regional, lesser-used languages, including academic publications and works of fiction.

In addition, the survey looked into the respondents' wishes 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 very interesting responses and, 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 other languages such as "speech recognition for Welsh", "speech recognition for Catalan, better grammar checking 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 language". Other respondents described that they would like to see some of the existing LT tools 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".

Based on the results, 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 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 digital language equality 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.

4.3 EU Citizens: Preliminary results

As described in Section E, an additional survey was carried out targeting EU citizens with the aim of taking into account the opinions, individual needs, wishes, and general demands, and more importantly to make sure that their voices play a decisive role in the pursuit of full digital language equality supported by LTs. The survey was disseminated in 28 countries through Lucid's services (see Section E). Additional dissemination was also carried out with the help of ELE partners who promoted the survey on social media, within their networks and through the ELE project website. [11]

In this section, we only present preliminary results, due the fact that at the time of writing the survey has not yet closed. Therefore we present only the main trends observed from the data collected from this larger cohort of respondents in order to obtain a better picture of the current scenario in terms of language technology support across European languages. For these preliminary results, we present the data collected in the 28 European countries covered by Lucid market services and 28 languages[12] from January 2022 to March 2022, along with additional data collected through the ELE consortium to date, i. e., as of March 29th 2022, totalling 17,851 responses.

4.3.1 Preliminary Results

The first question of the survey aimed at investigating which terms used in the domain of language technologies respondents were familiar with and could immediately recognise. The preliminary results show that the terms "Machine Translation" and "Chatbot" are the most widely known among respondents. Although the term "Chatbot" appears as one of three most known terms in 22 out of 28 countries, the most selected term is Machine Translation and it appears among the 3 most known terms in 18 countries. The third most known term is "smart personal assistants" and it is among the three top most selected in 10 countries, followed by "information retrieval" in 9 countries, "language technology" in 7 countries , "speech processing" in 5 countries and "natural language processing" which appears in 2 countries only, despite being a widely used term in academia and in the industry.

4.3.2 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. As mentioned above, we collected responses from 28 countries in the first round of data collection via Lucid's services. Figure 11 shows the breakdown of contributions per country to date (March 29th 2022).

The most represented age groups were 25-34 (30% of the total sample), 18-24 (21%) and 35-44 (13%). The remaining 36% belonged to the age group from 45 to 64 years old, were over 65 years old or preferred not to respond. The data also showed that 41.2% of the sample has a high school education level, followed by 24.4% with a bachelor's degree.

4.3.3 Language Coverage

In this section, we asked respondents to selected all the languages they use both socially and professionally. Overall, results show that many respondents use their native language

[11] https://european-language-equality.eu

[12] Here we report data collected from multilingual European countries in the most widely spoken language of that country. Thus, in Spain, we report data collected in Spanish, in Switzerland we report data collected in German, in Finland we report data collected in Finnish and in Belgium we report data collected in Dutch.

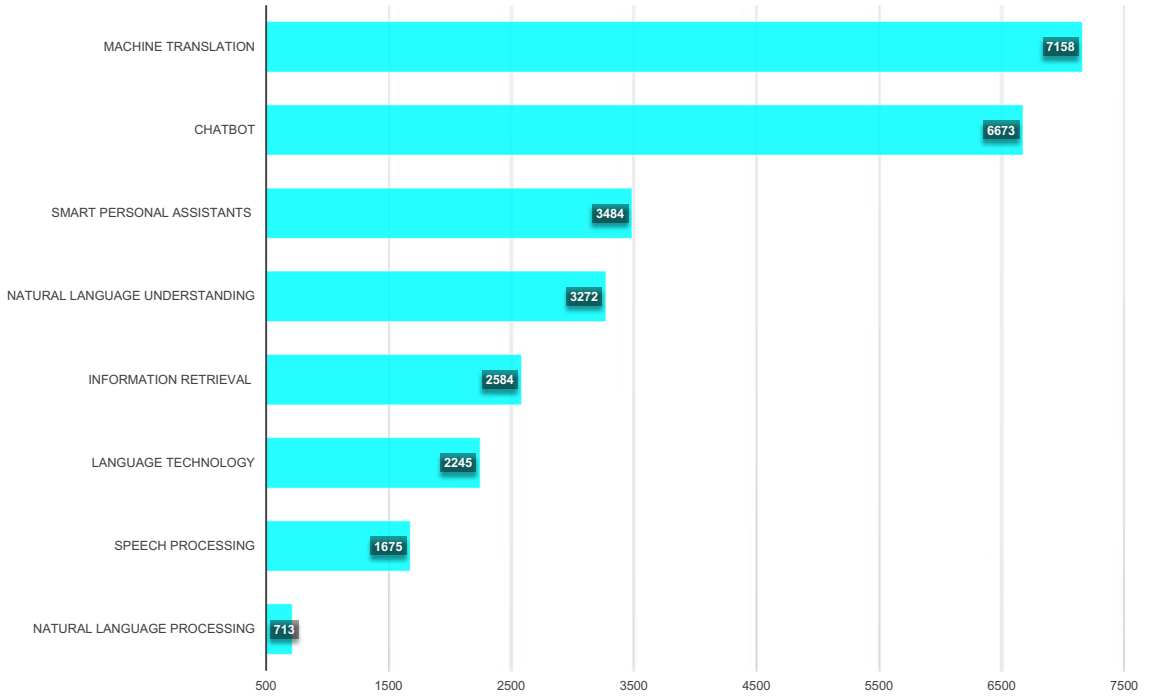


Figure 10: Terms that respondents are more familiar with or can immediately recognise

in addition to English even if they are not based in English-speaking countries. Therefore, we see the 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 12 illustrates the comparison of the most represented languages in the survey.

4.3.4 Evaluation of the Current Situation

Respondents were requested to select and, at the same time, rate the tools they use in each of the languages that apply to them. The preliminary results show that the top three most used tools are machine translation, search tools and proofing tools. This result is consistent across all countries covered in this analysis and all languages. Automatic subtitling tools are also among the most well evaluated tools in four languages, namely, Hungarian, Serbian, Lithuanian and Dutch.

The results also show that these three most used tools are also the top rated tools in a 5-point scale. However, the mean scores within the top rated tools vary across languages. The mean scores for search tools range from 4.5 to 3.8, translation tools from 4.3 to 3.5 and proofing tools from 4.1 to 3.3. Thus, while search tools are rated with a mean score above 4 in certain languages (e. g. Bulgarian, Romanian, Polish), they are also among the top rated in other languages (e. g. Portuguese and Dutch), but with lower rating mean scores (3.8 and 3.9 respectively). Similarly, while translation tools are among the three most well evaluated tools with mean scores above 4 for Finnish, Romanian and Bulgarian, for Danish and Norwegian the mean scores are lower ranging between 3.7-3.5. Thus, while the tools with the best performance in some languages would be between 3 and 4 star level (indicating that they are perceived as good), the performance of the same tools in other languages is rated between 4 and 5, suggesting they are perceived with excellent performance. Interestingly,

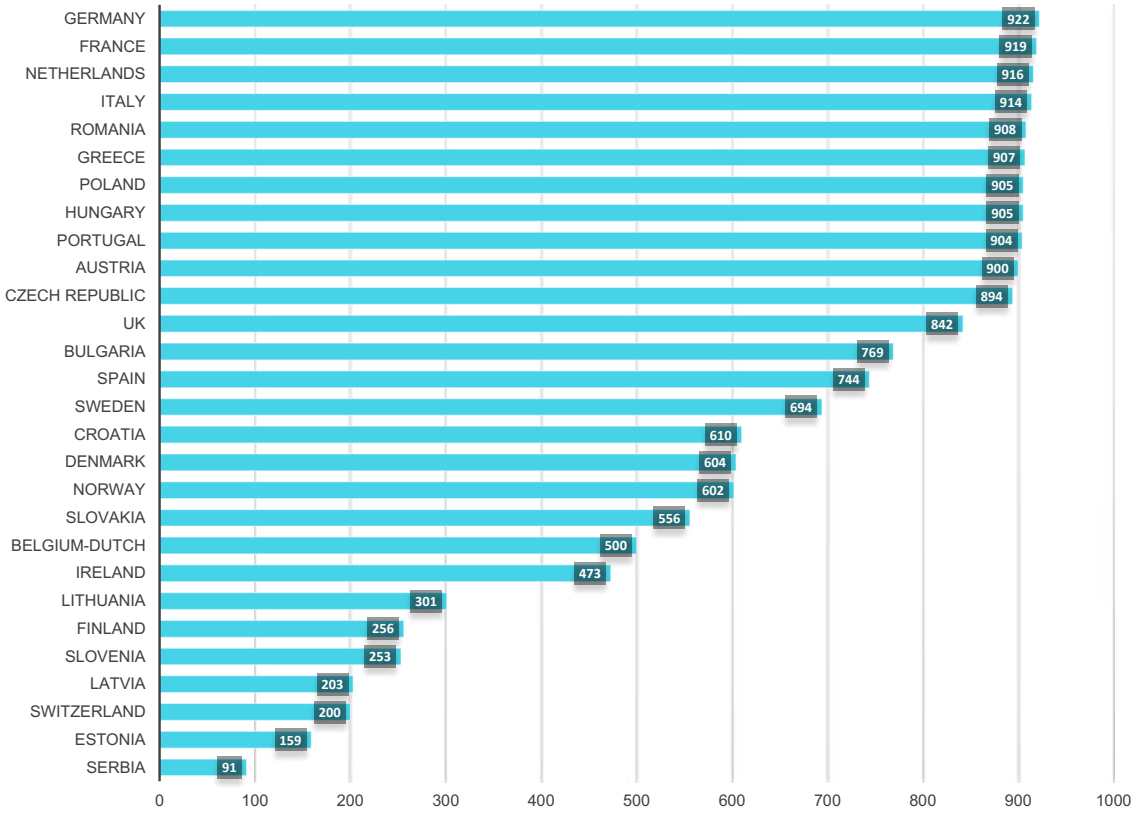


Figure 11: Number of responses collected in 28 European countries

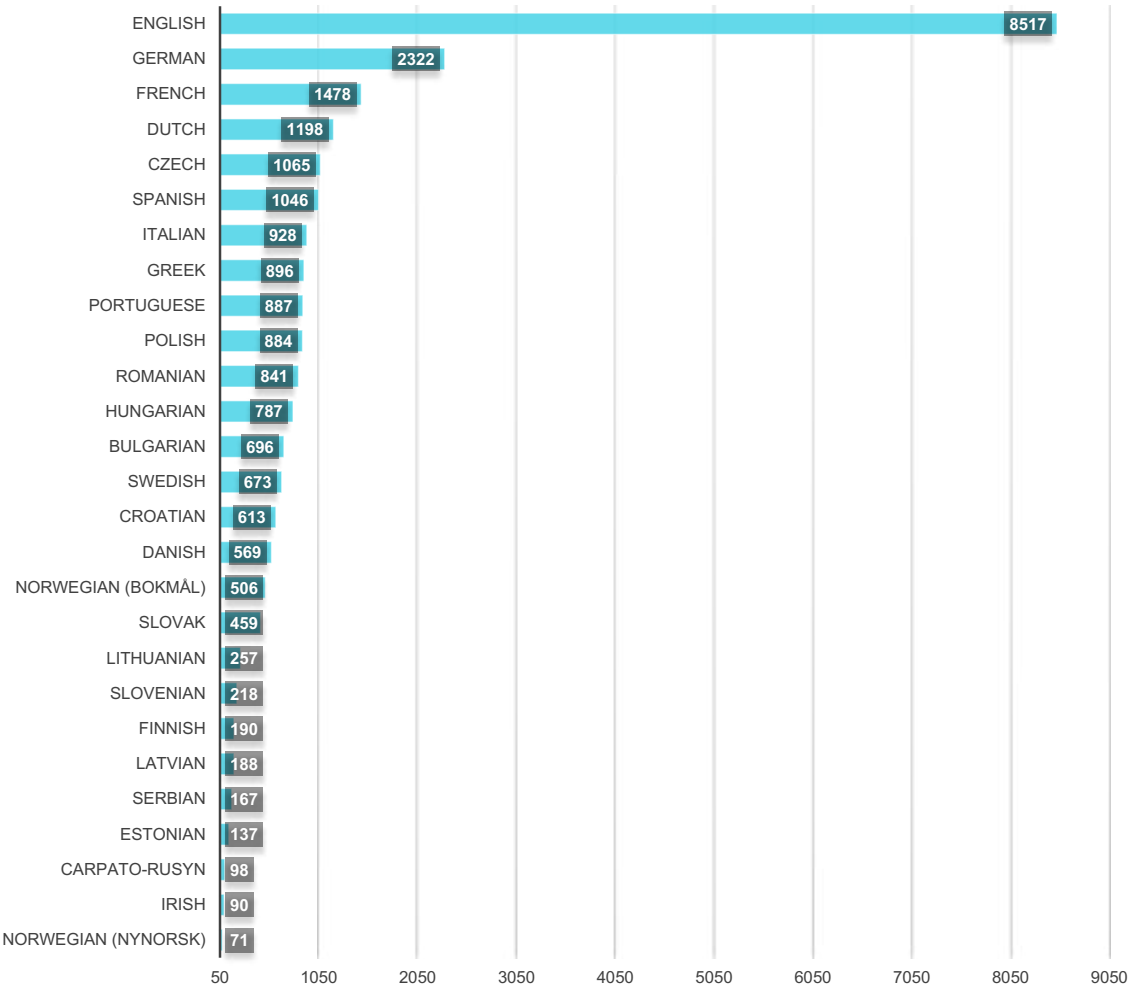


Figure 12: Languages most represented in the survey

for English, in line with results for other European languages, search tools are the top-rated technology, but responses collected from English-speaking countries (UK, Ireland) show that personal assistants are among the three best-rated tools, suggesting that this tool is widely used in English and is perceived by English speakers as performing well. This is not the case for other languages. The complete breakdown of rating scores per country and tools is presented in Table 18, Table 19, Table 20 and Table 21 in Appendix F.

When asked the question ”In general, what holds you back from using some of these apps or tools in your languages?”, the most selected response was ”I don’t need to use any apps or tools for this language”. However, it is interesting to see that, for certain languages, tools are not used because they are not available. The option ”lack of available tools” was selected by respondents that use LTs in Valencian/Catalan, Czech, Bulgarian, Slovenian, Polish to name a few languages. Regarding future demands, the survey shows that personal assistant tools are the most desired tool for the future in many European languages (e. g. Bulgarian, Croatian, Czech, Hungarian and Lithuanian). These results suggest that personal assistant tools such as Alexa or Siri are not available in certain languages or, if available, they are not yet frequently used in many European languages. Thus, the survey shows that there is currently a high demand for the use of these tools in the (near) future.

4.3.5 Predictions for the Future

Through a multiple-choice question, respondents were requested to select the top 3 advantages of improving apps and tools for all languages. As illustrated in Figure 13, the preliminary results show that the top 3 advantages in respondents’ opinions are ”to increase peoples’ exposure to these languages” with 8828 responses (50% of the total sample); the second advantage in respondents’ opinions is ”to improve communication between speakers of different languages” with 8538 responses (48%); and, finally, ”increase the number of speakers of languages, including minority and regional languages” with 7112 responses (40%).

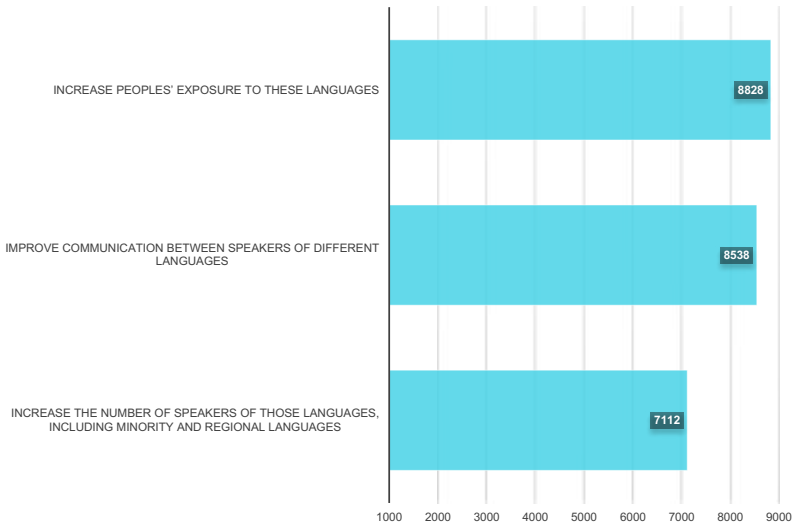


Figure 13: Top 3 advantages of improving apps and tools for all languages in respondents’ opinions

5 Deep Dives

This section summarises the results of the four Deep Dives into the topics of Machine Translation, Speech-related Technologies, Text Analytics, and a horizontal Data (Language Resources) topic. The full reports are available as specific Deliverables – Bērziņš et al. (2022); Backfried et al. (2022); Gomez-Perez et al. (2022) and Kaltenboeck et al. (2022), respectively.

5.1 Machine Translation

This deep dive has provided a condensed summary of the current state of the art in the field of MT and has suggested recommendations and directions for expected and desirable developments going towards 2030, especially to ensure that MT contributes to achieving DLE for all the languages of Europe.

From the beginning, the main goal of MT has been to provide high-quality, robust translation between any language pair. Today translation technologies are widely used the by general public, public sector and government agencies, SMEs, LSPs and many other industries where multilingual content is indispensable. The use of translation technology will definitely continue growing, covering new application areas (e. g., Internet of Things, smart homes and other smart devices), markets, supporting Europe's digital single market and language equality. When looking forward to 2030, we expect the movement towards Deep Natural Language Understanding enabling efficient and real-time translation to support human-to-human and human-to-machine communication.

Despite the widespread celebration of multilingualism in the EU, **there is no common EU-wide policy specific addressing language barriers**. With the exception of EC-developed and run *eTranslation* service, which now allows for a broader use than just public administrations, there is also a gap in publicly available MT services which cater specifically to the needs of people in Europe. This is slightly mitigated with the recent addition of open and free services offered by various developers (at or through ELG and/or provided by the Opus MT project, LINDAT/CLARIAH-CZ Clarin node etc.), but by no means complete or ready for massive use. Users around the world avail of free-of-charge MT services provided by Google[13], Microsoft[14], Baidu[15], etc. The risk is that what is freely available now could (easily) be taken away more or less unexpectedly for a variety of reasons, e. g. if those companies find a way to increase revenue in other directions, as has happened with other services provided by these large corporations. The absence of a clear roadmap and support for LT at European level translates into an incohesive, fragmented European market with disparate language support for the language communities of Europe. The future **publicly available MT systems** should not depend on large companies, especially those which are not **European**.

With the help of neural networks, MT has recently improved significantly in its quality, consistency and productivity. However, in many cases the focus of new technologies is still on big, fully-resourced languages, in particular English, thus limiting diversity and reinforcing already-existing disparities. At the same time, the general progress in AI and neural network techniques have opened the path to developing a universal translation engine aiming to translate between any language pair with help of a single model. The application of neural networks to MT allows also to forego the independence constraints and move towards context-aware methodologies in MT. A novel approach attracting the attention of many researchers is unsupervised MT, where (every time less) monolingual data suffices to build a working system. While much research (including fundamental research) work remains to

[13] https://translate.google.com
[14] https://www.bing.com/translator
[15] https://fanyi.baidu.com

be done in this area, together with "universal" MT, it emerges as one of the key pillars to **drive language equality**.

Another challenge of the current systems are various **biases in the models**, such as gender, racial and ethnic bias. In the future, ethical and fair MT should not further propagate notions of inequality or exclusion, but rather foster **an inclusive society**.

Explainable and interpretable machine learning is attracting more and more attention in the research community. A fundamental breakthrough is needed in the understanding of how current MT algorithms work, in an attempt to make them more transparent and to improve the accountability of the systems that incorporate them.

Another field in which a breakthrough is needed is quantum computing. Promising first theoretical steps towards reformulating MT and NLP as quantum computing problems have finally been successful, thus more research on how MT and NLP in general can be re-framed as a quantum computing problem is necessary.

The increasing quality of MT and the expanding preference (especially among younger users) for voice-based interaction with devices points to the need for more and improved applications for **speech-to-speech translation and multi-modal machine translation**. Speech translation is a key area to break the language barrier for human communication, to facilitate real inter-linguistic and cross-cultural understanding. In order to achieve human-like language processing capabilities, machines should be able to jointly process multimodal data, and not just text, images, or speech in isolation. There is a growing need for the translation of audiovisual content and development of multilingual and cross-lingual text-to-speech and speech-to-text applications that can support the meaningful integration of the written and spoken word and images. There is also a need for accessible content in the form of subtitles and audio descriptions, particularly – but not exclusively – to include disabled people and viewers with sensory impairments; this is particularly important with the increasingly mobile aging population across Europe. One step in this direction is represented by the recommendations of the New European Media Strategic Research Agenda to develop tools for automatic translation from speech to subtitles, from text to Sign Language, and from Sign Language to text (New European Media Initiative, 2020).

The **collection of usable language data** is particularly important: while the intensive use of MT systems developed by large global, non-EU, companies may put them in a position to collect and re-use user data, similar or related services in Europe would not be able to re-use user data in this way due to GDPR (Aldabe et al., 2021). In addition, the current European environment surrounding the **copyright laws** may pose a further barrier in Europe: while copyright law is subject to fair-use exceptions in countries such as the US, European law is far less flexible.[16] If lawmakers could agree that using aligned translations of copyrighted data constitutes fair use, as long as it does not impair the value of the materials and does not curtail the profits reasonably expected by the owner, LT stakeholders could avail of ever-growing high-quality language data to develop and offer a very wide range of much-needed tools and technologies, especially supporting underserved languages in the interest of digital language equality.

There is also a disparity between publicly available and proprietary bilingual corpora. Although the EU Coordinated Plan on Artificial Intelligence has foreseen a framework for the free flow of non-personal data in the European Union, the data resulting from public procurement procedures has the tendency to remain locked up in privately-owned data silos, while the research community and LT industry struggle to find, identify and reconstruct the public part of this data. A crucial breakthrough could be achieved if existing policy frameworks were adapted to make it mandatory for EU Member States to **make all data** in natural language-related workflows **publicly available**, at least those produced and directly man-

[16] It does not help that many Member States have not yet adopted even the 2019 Copyright Directive (however still not ideal) in their legal systems.

aged by Public Administrations and governments.

Finally, in general, the availability and quality of training and test data should be increased. Publicly available multilingual data should include a greater **diversity of domains and languages**, so that building high-quality MT systems becomes an option for all, and is not effectively restricted to popular languages or combinations with commercial interest. Future systems should be able to cover all European languages across all possible combinations with very similar, if not identical, quality expectations, and be trained on many different domains and genres.

In addition, **current evaluation metrics do not necessarily reflect actual translation quality** in the real world: their accuracy and reliability have been repeatedly called into question. Future systems should be evaluated by new automatic metrics which represent better approximations of human judgments and also ideally abandon the dependence on human reference translations. Moreover, evaluation should not be carried out on isolated sentences/segments, but be performed taking into account broader units, ideally considering entire texts, so as to adequately assess a range of supra-sentential phenomena. Adopting a single metric as a standard for measuring MT would possibly allow for a **widespread benchmarking** of (multilingual) LT across Europe. Increased attention should be paid to the human judgments used for tailoring the automatic metrics, as well as to manual evaluation in general.

There is also **a lack of necessary resources (experts, HPC capabilities, etc.) in Europe** compared to large US and Chinese IT corporations (e. g., Google, OpenAI, Facebook, Baidu, etc.). While in North America and Asia public and private resources can be allocated to only a limited number of languages, to effectively honour the well-entrenched commitment to promoting multilingualism in Europe resources must be distributed across a large number of official and unofficial EU languages, so that the respective language communities are treated fairly. There is also an uneven distribution of resources across countries, regions and languages (Aldabe et al., 2021). Considering the massive infrastructure that is required to train very large state-of-the-art LT systems, Europe starts with a systemic handicap. Europe's strong foundation in research and innovation can compensate for the disadvantage European organisations have with respect to infrastructure, provided that a concerted effort is undertaken in researching the development of **new hardware platforms** and respective AI training paradigms.

Finally, **the hardware on which MT runs must be scaled down**. Several approaches to replace GPU-based computing are already under investigation. By ensuring that the capabilities of the hardware are aligned with the needs of MT training and inference models, smaller models would be easy to integrate and use on any device and also be greener by requiring fewer resources, since training neural MT engines is resource-intensive and has a heavy carbon footprint; this not only has undesirable environmental consequences, but it also puts smaller and not well-resourced institutions and companies at a disadvantage in this area. The EU has the opportunity to be a pioneer in training and developing green LT by developing efficient models and hardware, as recommended by Strubell et al. (2019).

Since more and more members of the general public of all generations, many of whom are not trained linguists or language professionals, employ MT daily, there is a need for adequate literacy which includes a real understanding of LT and its capabilities. The growing need for **MT literacy** is already widely felt in the language industry, where linguists who are less familiar with MT are taught to use MT critically (Bowker, 2021).

From the end-user/localisation service provider perspective, the pricing pressure often arises as a consequence of not taking into account extra factors which make MT post-editing a more complex task than someone unfamiliar with MT might initially think. This also (still) leads to negative dispositions to MT and CAT tools by translators and post-editors.

At the level of policies/instruments, much more synchronisation of activities and alignment of best practices between regional, national and international bodies is necessary. An

instrument for efficient and homogeneous implementation of **policies towards DLE** would involve more equal support for all EU languages, including similar direct involvement of the relevant regional and/or national research communities.

5.2 Speech

The substantial advances made in the field of Speech Technologies (ST) over the past decades hold the potential for disruptive innovation in many areas and application domains that are of particular importance in Europe. Combined with the progress of related fields such as AI, NLU, NLP and ML, they provide the basis for broad adoption of speech and voice as the primary modality to interact with computer systems. Interaction can be expected to take place as part of larger and more complex systems modelling human-like communication and thus allowing a wider adoption of ST, NLP and NLU. In parallel, the individual technologies and components will continue to be improved, both in terms of accuracy and of coverage (of languages and dialects, as well as domains). All of these strands of advancements can support the overarching goal of achieving digital language equality in Europe within a decade by providing services to larger audiences on increased (and similar) levels of scope and performance.

The **main trends** of ST regarding a foresight towards 2030 which have been identified in the ST Deep Dive include the following:

- a continued trend towards integration of ST into richer user and application contexts
- integration of ST into appliances and services and thus wide adoption (**ST as a commodity**)
- a continued trend towards ever larger models, requiring more data and resources (limiting the set of actors able to produce them) and in parallel a **trend towards methods requiring less data** and consuming less resources
- an increase in **scope of end-to-end systems**, incorporating further elements and knowledge sources
- a trend towards **open-sourcing of models** paired with the fact that the best models will not be open-sourced (at first) and function as a competitive advantage for those able to produce such models
- a further increase in the development pace supported by **pre-train/fine-tune paradigm**
- increased presence of **multilingual and language-agnostic methods and models**
- the **confluence of approaches and models** of ST and AI in general as levels of abstraction become more complex (and require deeper and broader knowledge sources and/or “common sense”)
- continued work on **filling existing technological gaps** and improving individual ST technologies
- a growing prominence of **multimodal models**, combining ST with other modalities
- extension (in complexity, diversity, multifaceted character) and **standardisation of training sets and evaluation criteria**
- further **increase in the capacities** of ST related **hardware and infrastructure**
- an increased interest in **de-biasing** as well as **ethical aspects** of ST

- the **combination of statistical and symbolic models** (e. g. knowledge-graphs)
- **more inter-disciplinary work** and approaches with fields such as cognitive science, psychology, etc. to yield more human-like language-learning and language understanding, leading to novel models of learning including language-learning (and hence ST)

From a technical point of view, the continuation of the pre-train/fine-tune paradigm will strengthen players who possess the capacities to produce such models. **Availability of data and hardware will remain determining factors for short- and mid-term progress**. Open-sourcing allows the widening of scope and participants, however, the best-of-breed models will remain the privilege of a few selected players. In parallel, **ideas concerning sustainability, green AI** and the mere desire to not depend on the good-will of large players will likely cause activities to emerge focusing on **less-data, reuse, augmentation** etc. **Inter-disciplinary approaches** and cooperation may produce approaches mimicking more human-like learning - which e. g. does not depend on billions of examples to learn concepts, uses a variety of knowledge-sources, builds on expectations and common-sense etc. As long as current 'larger-is-better' approaches dominate, the availability and control of datasets and (computing) resources will remain key, both in technical as well as in commercial terms.

The **future of ST** will be **strongly influenced by the regulations governing the collection, storage, transmission, and use of personal data**. These relate to the users' concerns and expectations, the influence of the groups of interest, both at the national and transnational levels, and the future developments of the ST themselves, their growing scope of application, functionalities and performance improvements. In the context of European AI companies and research institutes, the development pace appears to be particularly strongly influenced by the current and future regulation schemes. **Lawmakers' decisions will thus have to consider the wide and profound impact of their regulations** – on the protection of citizens' personal data and privacy on the one hand, and on the wider field of AI technologies (research, development and application) and the comparative economic advantages and disadvantages vis-a-vis other geopolitical regions on the other hand. Extrapolating from the current regulations concerning user privacy, **the differences in data collection and use, the divide between the EU and non-EU countries is likely to continue to grow**. As AI technologies in the future will play a crucial role in defining competitive advantages across the different fields of human activities, including the commercial, social, military and intelligence, it is unlikely that a wider and far-reaching consensus between the competing countries and regions will be found, which would lead to a **standardising set of regulations across the regions**.

With the growing presence of ST, ML and AI in general, concerns are growing regarding the **hidden flaws, shortcomings and baked-in biases of such systems**. This is certainly true from the citizens' perspectives, but also from the points of view of academia and the industry. **Explainability and accountability of systems performance**, i. e., results and decisions, will remain hot-topics, both from ethical as well as technical viewpoints.

Users will neither be able nor want to distinguish between AI, NLP or NLU, between a platform and a particular application or part thereof. To them, the overall system will be what they interact with and potentially what they will perceive as being biased, unfair or harming them in any way. In a sense, **ST will become a hidden commodity within much larger intelligent systems which will be ubiquitous in a transparent way**, e. g. embedded into devices, wearables or part of Augmented Reality or Virtual Reality.

5.3 Text Analytics

Text analytics (TA) and natural language understanding (NLU) deal with extracting meaningful information and insights from text documents, as well as enabling machines to under-

stand such content in depth, similar to how a human would read a document. These types of tools have been on the market for several years and have successfully found applications in many sectors including health, education, legal, security, defense, insurance, and finance, amongst others. Conventional text analytics services available in the market include syntactic analysis, extractive summarisation, key phrase extraction, entity detection and linking, relation extraction, sentiment analysis, extraction of personal identifiable information, language detection, text classification, categorisation, and topic modeling, to name but a few. In addition, conversational AI services and tools, including chatbots and virtual agents, are frequently offered under the umbrella of text analytics. More recent additions to the text analytics catalogue are machine reading comprehension services based on tasks such as extractive question answering, which are usually marketed as part of both virtual agents and intelligent search engines to provide exact answers to user questions.

The success of machine and deep learning has caused a noticeable shift from knowledge-based and human-engineered methods to data-driven architectures in text processing. The text analytics industry has embraced this technology and hybrid tools are incipiently emerging nowadays, combining or replacing robust rule-based systems that have been the norm in the market until now with machine learning methods.

Recent breakthroughs in deep learning have made impressive progress in NLP. Neural language models like BERT and GPT-3, to name some of the more widely-used, are able to infer linguistic knowledge from large collections of text that can then be transferred to deal effectively with NLP tasks without requiring too much additional effort. While the progress made in the last few years is undeniably impressive, there are still many gaps and shortcomings that need to be addressed to make these tools and technologies fully operational and especially to benefit all European Language, which we summarise here.

Due to the data-driven nature of these deep learning paradigms, the availability of suitable data for use in both training and evaluating today's state-of-the-art NLP tools is crucial. As it stands, beyond general purpose datasets, **labelled data is scarce**, labour-intensive and therefore expensive to generate. In particular, the **lack of domain-specific labelled data** (e. g., insurance, legal, financial, medical, etc.) is a major bottleneck to data-driven approaches in business applications. **Poor language coverage** is another concerning issue as the majority of datasets being produced that are relevant to Europe are based on the major languages such as English, German, Spanish and French. Moreover, **low-quality datasets** negatively influence the trained models. High-quality datasets contain reliable content (i. e. no fake news), balanced content (e. g. unbiased) and clean content (i. e. non-toxic/hate-speech).

From a legal point of view, **data protection and privacy** (DPP) policies can put limits on the type of data that can be made available for text analytics. **GDPR** (the EU's General Data Protection Regulation), while important for EU citizens' protection, significantly **hampers the extent to which language data can be sourced and reused** for machine learning based tools in Europe. European-based researchers and LT developers cannot therefore use, share, modify or build upon many of these datasets – which sets DPP-compliant players in this field at a competitive disadvantage.

In terms of language coverage, it is worth noting that most of **today's text analytics solutions are language specific**. Various challenges arise in many contexts (business, personal, governmental), where the multilingual needs of customers and users from across Europe and around the globe need to be met. In addition, **text analytics tools are typically unaware of other modalities** of information often used alongside text, such as images, audio, and video, which can enrich the analytics process. Furthermore, data-driven text analytics work as **black boxes** that are hard to interpret. This lack of transparency makes it difficult to build trust between human users and system decisions. It also makes it difficult to bring such technology to domains where regulations can demand that systems explain every decision. Moreover, many current text analytics **evaluation benchmarks are unreliable and**

biased, rewarding leaderboard positioning over other important features including carbon footprint.

From a business-oriented perspective, **the absence of standards** for text analytics tools hampers **interoperability at an enterprise level** and can potentially lead to vendor lock-in situations. Low interoperability can put significant investments in LT at risk. In addition, **certification of text analytics tools** is a pending task important to guarantee compliance with standard conformance criteria. Furthermore, current text analytic tools require expert level skills, rendering it difficult for domain experts and other users to contribute to the analysis process.

To address such challenges in text analytics tools not only technical, but regulatory and societal advances are required. In the following we summarise our technology visions in a 10 years horizon that will lay the foundations of the next generation of text analytics tools (Gomez-Perez et al., 2022).

Neural language models have a prominent role in the technology visions as a key data-driven emergent technology in text analytics, with the potential to revolutionise the offer of text understanding functionalities and to increase the coverage of such tools for less widely spoken languages. **Therefore, the ability to build neural language models for target languages to the same standards as English is key for language equality**. Nevertheless, we should not take for granted that large amounts of publicly available corpora of good quality can be readily obtained for all European languages, rather the contrary. **The effort to ensure that all languages have large amounts of publicly available corpora of good quality, taking fairness issues into account, should be at the centre of any future efforts for digital language equality.** Data sharing policies both at national and European level, along with widespread awareness of the benefits of open-data will play an important role here. In addition, there is an immediate need for more sophisticated tools that can assist with more accurate data curation, such as domain filtering, bias and toxic content detection and elimination, and so on.

Research on unsupervised and zero-shot learning opens new possibilities to increase the coverage of minority and under-resourced languages in the text analytics industry. It is hoped that the language coverage of text analytics tools will be enhanced thanks to a mixture of research in language models, language agnostic models, and neural MT. Research on these subjects is underway and shows promising results, even for under-resourced languages, paving the way for truly **multilingual language technologies**.

Deep learning systems need to coexist with knowledge-based systems, also referred to as Symbolic AI, for natural language processing that have existed and been used in real-life applications for many years. One of the main development challenges for NLP is employing hybrid approaches, where it is possible **to get the best from deep learning and symbolic systems**, while minimising their respective drawbacks. The aim therefore is not only to make NLU models aware of the entities contained in a knowledge base and the relations between them from a general-purpose point of view (as provided by resources like Wikipedia), but to also quickly incorporate pre-existing resources from vertical domains and custom organisations into our models in a cheap and scalable way.

It has been shown that different modalities can be combined to provide complementary information that may be redundant, yet convey information more effectively. Thus, there is a need for analytic tools that can jointly analyze different modalities and extract information from them to carry out a more comprehensive analysis. This **convergence across modalities** requires synergies from AI research fields that until now have been conducted separately, such as NLP, automatic speech recognition and computer vision.

With respect to addressing issues related to the blackbox nature of AI-driven text analytics, there has been a notable increase in attention given to the area of **Explainable AI**. In cases where decisions are made based on AI model prediction, it is important that businesses can assess these models' level of accuracy, fairness and transparency.

We advocate for a next generation of language processing tools that consider end users' needs and expectations, making them part of the design and learning process. These tools should be **human-aware and trustworthy, avoiding bias, offering explanations, and respecting user privacy**. Moreover, **human intelligence should be used alongside machine learning techniques** to produce better language technologies. Human feedback can serve as a guide in the learning process, informing the machine on what users do and do not want. Reinforcement learning from human feedback is a promising research avenue which uses human intelligence to improve language processing tools. Also, improved interactivity with domain experts and users is a key area for further advances beyond the usual supervised paradigm.

To conclude, within AI the field of text analytics and natural language understanding has an enormous potential to impact the development of businesses and societies across the EU. At the same time, it also has the potential to address the challenges related to linguistic discrimination and language barriers to communication and the free flow of information, an utmost priority for Europe. Text analytics and NLU have a decisive role to play in establishing a fair, inclusive and sustainable Multilingual Digital Single Market that is based on equality, and acts as a multiplier of opportunities and collaboration among key European stakeholders, including academia, industry, public administration, and citizens.

5.4 Data and Knowledge

The Data and Knowledge Deep Dive is a horizontal report covering **data, language resources, and Knowledge Graphs** as a support to LT, as data forms the basis and backbone for technologies and solutions in the area of Language Technology and thereby for Digital Language Equality in Europe.

The report provided the current state of the art in Data and Knowledge sources, identified main gaps, analyzed the situation with regard to breakthroughs needed, and elaborated on the vision for 2030.

In order to do so, and on top of desk research, two virtual workshops with ELE consortium members have been organized on state of the art and main gaps as well as on future use cases and requirements regarding data and future technology visions, It was followed by discussions with industry representatives about the topics discussed at the workshops.

The main issues concerning Data, Language Resources, Knowledge Graphs were identified as follows:

- Availability of data and metadata
- Accessibility of data
- Quality of data
- Data Interoperability
- Licenses and data related regulations
- Data and ethics
- Data literacy

All of these issues need to be tackled in the future to allow for data collection and provision with fair conditions and costs for all relevant stakeholders to achieve Digital Language Equality.

Related to these issues are the following horizontal and vertical areas which will be part of the technology vision for 2030:

- Data infrastructures, data spaces and datamarkets
- Knowledge Graphs
- Semantic AI: statistical and symbolic AI in combination
- Innovative data and metadata management tools.

As an add-on component, the topic of *data-related business models* was tackled, where the importance of working and sustainable data-related business models was identified as a prerequisite for a working data economy and ecosystem that thereby stimulates and fosters the above listed data related components, and thereby a well functioning language technology landscape that provides the basis for a digital European language equality.

A list of future use cases have been collected and described, the related data requirements specified, and the main technology areas have been identified as (i) conversational AI, and (ii) insight engines.

Beside technology, interoperability or data related attributes, there must be a strong focus established on applying all these mechanisms and methodologies to **the widest range of languages possible** – at least to EU languages but also local and regional dialects of these languages, as well as to non-EU languages that are widespread across Europe. Without such data and language resources in place, digital language equality cannot be reached.

To fill the identified gaps in data, language resources, and Knowledge Graphs we recommend and suggest a future path for Europe towards **comprehensive and interlinked data infrastructures**. These infrastructures have to provide interoperability out-of-the-box by following harmonised and well-proven standards, regarding (i) **data (semantic data) interoperability** as well as (ii) services and (iii) innovative metadata and data management tools that are available along all steps of the data life cycle.

Metadata, data, data-driven services and data-driven tools must be an integral part of these data infrastructures, without today's huge efforts in data cleaning and data integration, or service- and tool integration. This future technology vision of integrated and interoperable data infrastructures shall follow the idea of a Semantic Data Factory including rich semantics, and thereby providing context and meaning, as well as including dynamic metadata and augmented metadata and data management. By this approach a **federated network and infrastructure of interlinked data spaces** for LT can be realised. Existing data spaces as well as newly developed ones should be integrated, where appropriate and possible.

In such a federated ecosystem, **relevant data regarding a domain and/or language can easily be identified, loaded, and evaluated for specific use cases**. Data driven services are provided and can be used according to end users requirements. **Integrated crowd-sourcing and/or citizen science mechanisms** allow human-machine interaction to foster **data acquisition, cleaning and enrichment** (e. g., annotation, classification, quality validation and repair, domain specific model creation, etc.). Raw data can be loaded into available tools to train algorithms or create translation memories and/or (language) models for specific use cases, but also existing algorithms, models or vocabularies are available and can be easily loaded and re-used to avoid unnecessary energy consumption of computing power to foster the idea of **energy efficient data management**.

In addition, high importance needs to be put on **privacy protection** (related to personal identifiable information and beyond), **the avoidance of bias** (for example, but not exclusively, on gender), and on data sovereignty.

The approach of such data infrastructures requires working and sustainable **business models that allow data trading, data sharing and collaboration**. It also requires supporting policies, as well as **sustainable data governance models around data creation, data provision and data sharing**. Well targeted **publicly funded/supported programmes**

and activities in the area of data literacy are required from early education onward, to ensure that sufficiently skilled resources in the field are available in the future.

In addition, **an action plan for the collection and the development of data and language resources** that are relevant for LT, including Knowledge Graphs, is needed to ensure the availability of sufficient data in the EU languages, as well as in dialects and important non-EU languages. The recommendation for this is to look into three areas, i. e.: (i) **Language Equality Action Plan** by means of targeted national and European funding along a matrix of relevant resources and languages, combined with (ii) **more measures in the fields of crowd-sourcing and citizen science**, and (iii) the development of functioning **data-related business models**.

The **availability of high quality data, language resources and knowledge graphs in at least the EU official 24 languages**, but extending to as many languages as possible, that are **easily accessible with fair conditions and costs in a clearly specified legal environment** providing transparent rules and regulations can support clear benefits and competitive advantage for the stakeholders.

This would encourage the European research community to foster innovations in the field, the industry to successfully compete in a global market, and thereby the European citizens and their societies, that are constantly growing in regard to their diversity and a wide and increasing variety of languages. **Data, language resources, and Knowledge Graphs are thereby a crucial factor on our way to European DLE**.

6 Summary and Conclusions

This report describes the process of collecting material evidence from and opinions of multiple stakeholders groups relevant for Language Technology research, development and use in Europe. It has been extracted from the deliverables describing the results by stakeholder group (D2.2 to D2.12) or topic (D2.13 to D2.16). The material, facts, figures and views are to be further distilled to support the **strategic research, innovation and implementation agenda (SRIA) and roadmap** to be produced in WP3 and its deliverables.

The **LT developers survey** addressed the European LT community together, reaching a wide and demographically distributed audience. It was answered by 321 different respondents who represent 223 different organisations located in 32 different countries. The respondents have been recruited by the established research networks in LT/AI, i. e., META-NET, CLARIN and CLAIRE, projects like ELG (ELE's sister project) and other related projects and networks focusing on LT or on neighbouring fields, such as AI4EU, ELISE, ELEXIS, and Nexus Linguarum. Additional networks, associations and projects indicated by the respondents include the ELRC initiative, ELRA, ACL, EAMT, DARIAH and many more. The areas in which the respondents are active covered the full range of LT areas both in terms of research and applications. 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 in this survey are the insufficient support for basic (fundamental) 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 the 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 (or inadequacies, or lack of regulations, depending on the topic and area), 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, raising awareness, managing expectations, etc.).

The **LT Users and Consumers survey** brought together diverse groups of stakeholders including representatives of communities of LT users and consumers, academic and commercial stakeholders, 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, public administration, language services). The survey was disseminated mainly via emails by all six relevant ELE partners, namely, ELEN, LIBER, ECSPM, NEM, EFNIL and Wikipedia, and additional dissemination through social networks was also carried out. The leaders from the six European initiatives promoted the survey in their networks targeting representatives of organisations and communities of users and consumers. In addition, as a partner and coordinator of the project, the DCU team also worked on the dissemination which promoted the survey within its networks as well as through regular posts on social media from August to September 2021. DCU collected a total of 14 responses from researchers affiliated to the various offices that belong to the ADAPT Centre (in addition to DCU, Trinity College Dublin and University College Dublin), while the 108 remaining responses were collected through social media and word of mouth. Based on the results, it can be concluded that the most important finding 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 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. In order to achieve full digital language equality as a crucial step to maintain and promote 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 that are relevant in the digital age.

The four **Deep Dives** covering the areas of machine translation, text analytics, speech and data analysed the current situation horizontally in these very important areas of LT. They provided a detailed description of the state-of-the-art, analysing in particular current gaps and outlining the vision for future developments. Each Deep Dive presented an informed view on the future situation in these fields by 2030 from the perspective of digital language equality, and the key findings of these in-depth analysis are summarised in this deliverable, also to allow a comparison of priorities and the identification of common cross-cutting themes, concerns and opportunities for potential converging developments (Section 5).

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 digital language equality supported by LTs. The survey was disseminated in 28 countries thorough Lucid's services. Additional dissemination was also 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 (LT Developers and LT Users and Consumers, as described above), there are several similarities not only in terms of scope of the analysis, but also of the key results that were obtained: languages other than English are poorly supported (with some rather randomly distributed exceptions) - something evident even from the distribution of languages that the respondents considered in their responses. These (preliminary) 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.

To summarise, the **European Language Equality Programme**, in the form of the SRIA as well as a roadmap, is seen as the main output of the project which will serve as a blueprint for achieving full digital language equality in Europe by 2030. The surveys and expert in-

terviews discussed here targeted LT developers, users and – equally importantly – the EU citizens. The surveys investigated language coverage, evaluated the current situation of LT in Europe and encouraged participants to share their predictions and visions for the future. More than 450 survey responses were collected and dozens of expert interviews were conducted. In addition, the EU citizen survey is, at the time of writing, still open so that citizens can provide their opinions regarding digital support for their languages. Further, four ELE industry partners compiled deep dive reports for the fields of Machine Translation, Speech, Text Analytics and Data and Knowledge. All this data and opinions collected through the open questions in the surveys and presented during the personal interviews are a broad, independent and solid evidence base for a trusted, bias-free and targeted SRIA and roadmap, leading, if implemented by the political authorities, to true digital language equality in Europe by 2030.

References

Itziar Aldabe, Georg Rehm, German Rigau, and Andy Way. Deliverable D3.1 Report on existing strategic documents and projects in LT/AI, 2021. URL https://european-language-equality.eu/wp-content/uploads/2021/12/ELE___Deliverable_D3_1__revised_.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Gerhard Backfried, Marcin Skowron, Eva Navas, Aivars Bērziņš, Joachim Van den Bogaert, Franciska de Jong, Andrea DeMarco, Inma Hernaez, Marek Kováč, Peter Polák, Johan Rohdin, Michael Rosner, Jon Sanchez, Ibon Saratxaga, and Petr Schwarz. Deliverable D2.14 Technology Deep Dive – Speech Technologies, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D2_14__Speech__Technologies.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Lynne Bowker. Beyond the language industry: Helping others to develop machine translation literacy skills, 2021. URL https://www.gala-global.org/knowledge-center/professional-development/blogs/beyond-language-industry-helping-others-develop.

Aivars Bērziņš, Mārcis Pinnis, Inguna Skadiņa, Andrejs Vasiļjevs, Nora Aranberri, Joachim Van den Bogaert, Sally O'Connor, Mercedes García–Martínez, Iakes Goenaga, Jan Hajič, Manuel Herranz, Christian Lieske, Martin Popel, Maja Popović, Sheila Castilho, Federico Gaspari, Rudolf Rosa, Riccardo Superbo, and Andy Way. Deliverable D2.13 Technology Deep Dive – Machine Translation, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D2_13__Machine_Translation_.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Maria Eskevich and Franciska de Jong. Deliverable D2.3 Report from CLARIN, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D2_3_ _Report_from_CLARIN_.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Jose Manuel Gomez-Perez, Andres Garcia-Silva, Cristian Berrio, German Rigau, Aitor Soroa, Christian Lieske, Johannes Hoffart, Felix Sasaki, Daniel Dahlmeier, Inguna Skadiņa, Aivars Bērziņš, Andrejs Vasiļjevs, and Teresa Lynn. Deliverable D2.15 Technology Deep Dive – Text Analytics, Text and Data Mining, NLU, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D2_15__Text_Analytics_.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Jan Hajič, Maria Giagkou, Stelios Piperidis, Georg Rehm, and Natalia Resende. Deliverable D2.1 Specification of the consultation process, 2021. URL https://european-language-equality.eu/wp-content/uploads/2021/05/ELE___Deliverable_D2_1.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Jan Hajič, Tea Vojtěchová, and Maria Giagkou. Deliverable D2.5 Report from META-NET, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D2_5__Report_from_META_NET_.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Stefanie Hegele, Katrin Marheinecke, and Georg Rehm. Deliverable D2.6 Report from ELG, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D2_6__Report_from_ELG_.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Martin Kaltenboeck, Artem Revenko, Khalid Choukri, Svetla Boytcheva, Christian Lieske, Teresa Lynn, German Rigau, Maria Heuschkel, Aritz Farwell, Gareth Jones, Itziar Aldabe, Ainara Estarrona, Katrin Marheinecke, Stelios Piperidis, Victoria Arranz, Vincent Vandeghinste, and Claudia Borg. Deliverable D2.16 Technology Deep Dive – Data, Language Resources, Knowledge Graphs, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D2_16__Data_and_Knowledge_.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

LT-Innovate. The lt-innovate innovation agenda, 2016. URL http://www.lt-innovate.org/sites/default/files/2904-LTi_Innovation_Agenda.pdf.

New European Media Initiative. Stategic research and innovation agenda 2020, 2020. URL https://nem-initiative.org/wp-content/uploads/2020/06/nem-strategic-research-and-innovation-agenda-2020.pdf?x98588.

Georg Rehm and Stefanie Hegele. Language Technology for Multilingual Europe: An Analysis of a Large-Scale Survey regarding Challenges, Demands, Gaps and Needs. In Nicoletta Calzolari, Khalid Choukri, Christopher Cieri, Thierry Declerck, Sara Goggi, Koiti Hasida, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Helene Mazo, Asuncion Moreno, Jan Odijk, Stelios Piperidis, and Takenobu Tokunaga, editors, *Proceedings of the 11th Language Resources and Evaluation Conference (LREC 2018)*, pages 3282–3289, Miyazaki, Japan, May 2018. European Language Resources Association (ELRA).

Georg Rehm, Maria Berger, Ela Elsholz, Stefanie Hegele, Florian Kintzel, Katrin Marheinecke, Stelios Piperidis, Miltos Deligiannis, Dimitris Galanis, Katerina Gkirtzou, Penny Labropoulou, Kalina Bontcheva, David Jones, Ian Roberts, Jan Hajic, Jana Hamrlová, Lukáš Kačena, Khalid Choukri, Victoria Arranz, Andrejs Vasiļjevs, Orians Anvari, Andis Lagzdiņš, Jūlija Meļņika, Gerhard Backfried, Erinç Dikici, Miroslav Janosik, Katja Prinz, Christoph Prinz, Severin Stampler, Dorothea Thomas-Aniola, José Manuel Gómez Pérez, Andres Garcia Silva, Christian Berrío, Ulrich Germann, Steve Renals, and Ondrej Klejch. European Language Grid: An Overview. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Christopher Cieri, Khalid Choukri, Thierry Declerck, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, pages 3359–3373, Marseille, France, 5 2020. European Language Resources Association (ELRA).

Georg Rehm, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Victoria Arranz, Andrejs Vasiļjevs, Gerhard Backfried, José Manuel Gómez Pérez, Ulrich Germann, Rémi Calizzano, Nils Feldhus, Stefanie Hegele, Florian Kintzel, Katrin Marheinecke, Julian Moreno-Schneider, Dimitris Galanis, Penny Labropoulou, Miltos Deligiannis, Katerina Gkirtzou, Athanasia Kolovou, Dimitris Gkoumas, Leon Voukoutis, Ian Roberts, Jana Hamrlová, Dusan Varis, Lukáš Kačena, Khalid Choukri, Valérie Mapelli, Mickaël Rigault, Jūlija Meļņika, Miro Janosik, Katja Prinz, Andres Garcia-Silva, Cristian Berrio, Ondrej Klejch, and Steve Renals. European Language Grid: A Joint Platform for the European Language Technology Community. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations (EACL 2021)*, pages 221–230, Kyiv, Ukraine, 4 2021. Association for Computational Linguistics (ACL).

Andrew Rufener and Philippe Wacker. Deliverable D2.4 Report from LT-innovate, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D2_4__Report_from_LT_Innovate_.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Emma Strubell, Ananya Ganesh, and Andrew McCallum. Energy and policy considerations for deep learning in nlp. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, Florence, Italy, 2019.

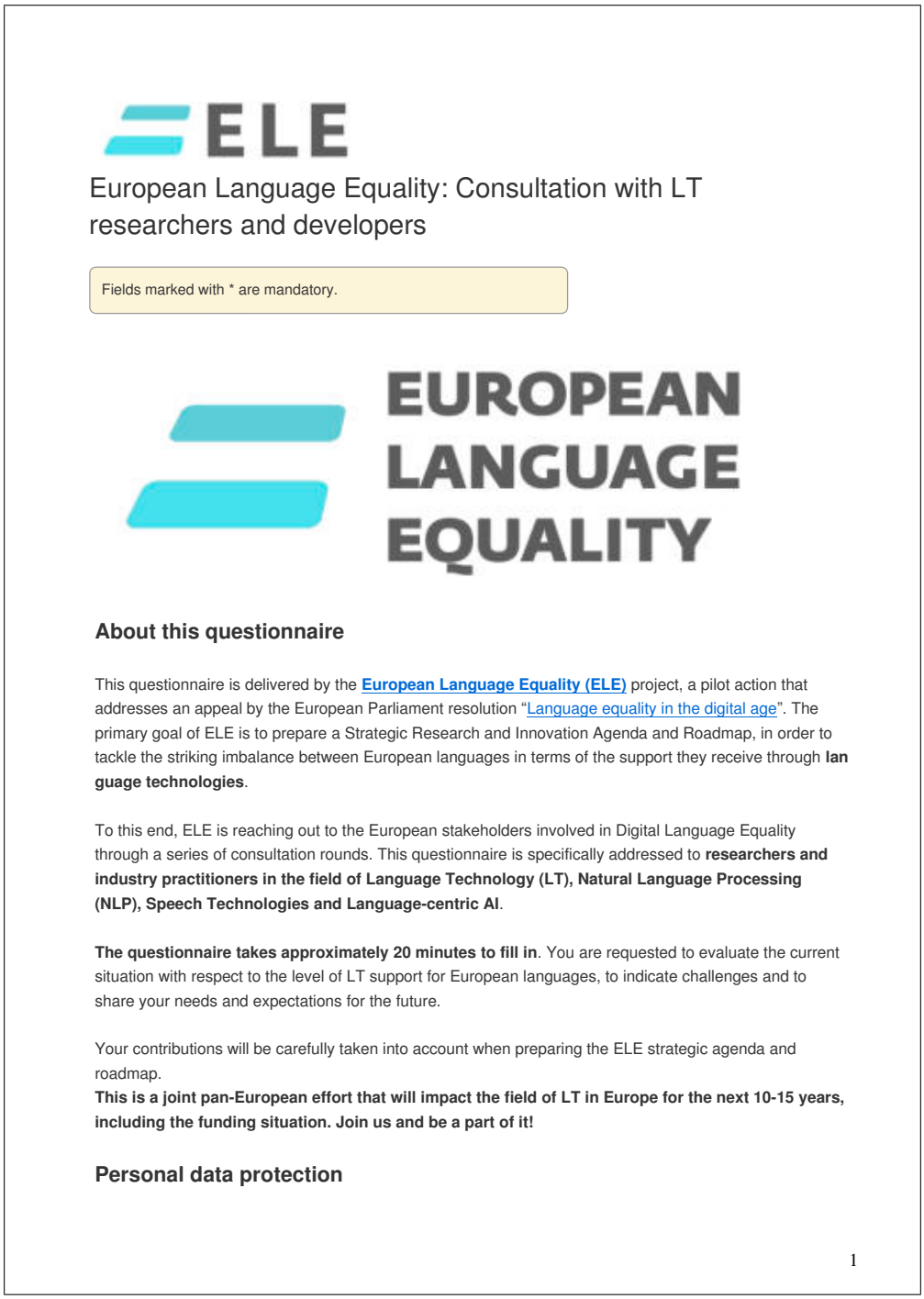
Marlies Thönnissen. Deliverable D2.2 Report from CLAIRE, 2022. URL https://european-language-equality.eu/wp-content/uploads/2022/03/ELE___Deliverable_D2_2__Report_from_CLAIRE_.pdf. Project deliverable; EU project European Language Equality (ELE); Grant Agreement no. LC-01641480 – 101018166 ELE.

Andrejs Vasiljevs, Khalid Choukri, Luc Meertens, and Stefania Aguzzi. Final study report on CEF Automated Translation value proposition in the context of the European LT market/ecosystem, 2019. DOI 10.2759/142151. A study prepared for the European Commission, DG Communications Networks, Content & Technology by Crosslang, Tilde, ELDA, IDC.

Appendix

A LT Developers Survey

Figures 14 to 22 show the complete LT research and developers survey.



ELE

European Language Equality: Consultation with LT researchers and developers

Fields marked with * are mandatory.

EUROPEAN LANGUAGE EQUALITY

About this questionnaire

This questionnaire is delivered by the **European Language Equality (ELE)** project, a pilot action that addresses an appeal by the European Parliament resolution "Language equality in the digital age". The primary goal of ELE is to prepare a Strategic Research and Innovation Agenda and Roadmap, in order to tackle the striking imbalance between European languages in terms of the support they receive through **lan guage technologies**.

To this end, ELE is reaching out to the European stakeholders involved in Digital Language Equality through a series of consultation rounds. This questionnaire is specifically addressed to **researchers and industry practitioners in the field of Language Technology (LT), Natural Language Processing (NLP), Speech Technologies and Language-centric AI**.

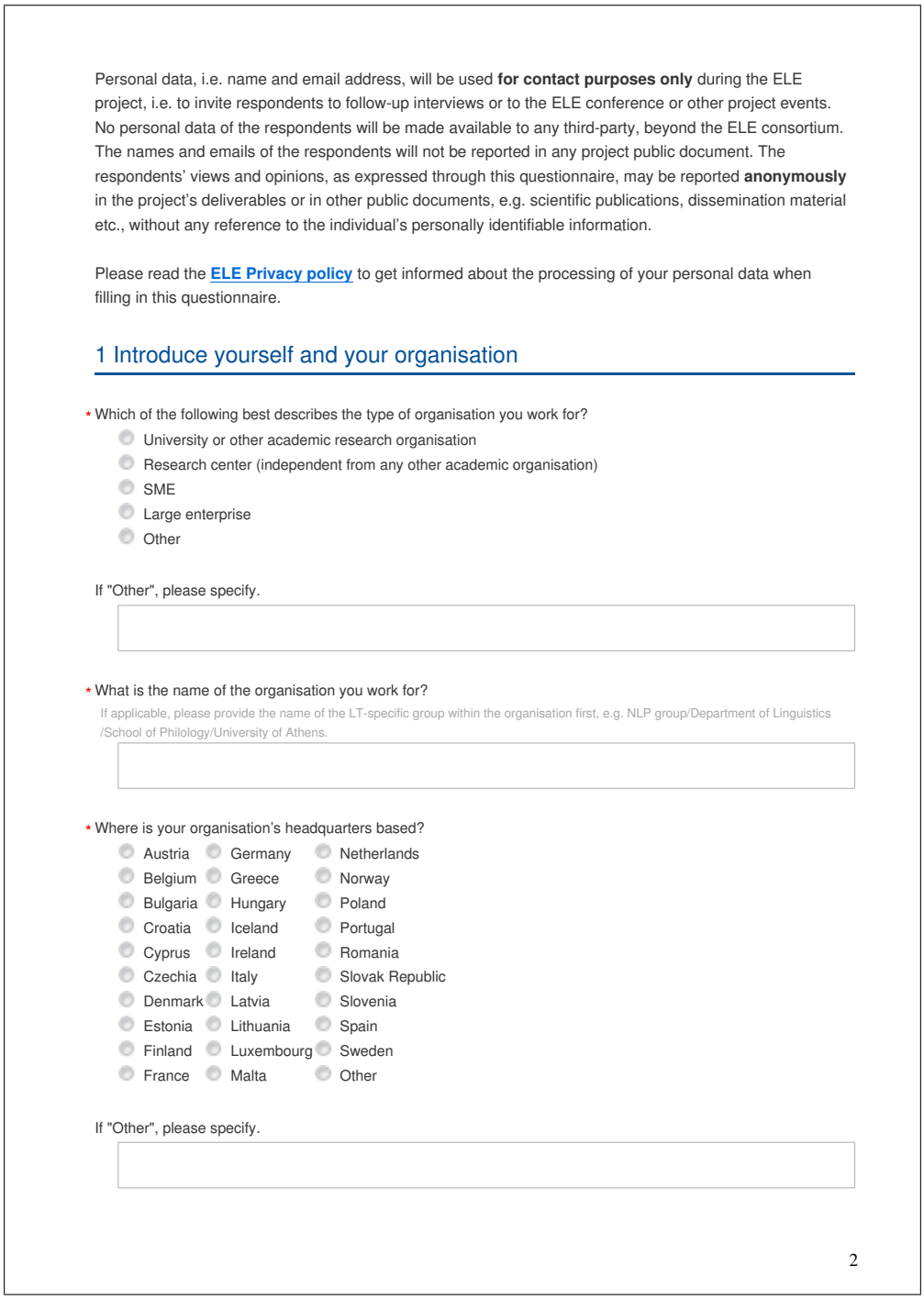
The questionnaire takes approximately 20 minutes to fill in. You are requested to evaluate the current situation with respect to the level of LT support for European languages, to indicate challenges and to share your needs and expectations for the future.

Your contributions will be carefully taken into account when preparing the ELE strategic agenda and roadmap.
This is a joint pan-European effort that will impact the field of LT in Europe for the next 10-15 years, including the funding situation. Join us and be a part of it!

Personal data protection

1

Figure 14: LT developers: full survey as published (page 1/9)



Personal data, i.e. name and email address, will be used **for contact purposes only** during the ELE project, i.e. to invite respondents to follow-up interviews or to the ELE conference or other project events. No personal data of the respondents will be made available to any third-party, beyond the ELE consortium. The names and emails of the respondents will not be reported in any project public document. The respondents' views and opinions, as expressed through this questionnaire, may be reported **anonymously** in the project's deliverables or in other public documents, e.g. scientific publications, dissemination material etc., without any reference to the individual's personally identifiable information.

Please read the **ELE Privacy policy** to get informed about the processing of your personal data when filling in this questionnaire.

1 Introduce yourself and your organisation

* Which of the following best describes the type of organisation you work for?

- University or other academic research organisation
- Research center (independent from any other academic organisation)
- SME
- Large enterprise
- Other

If "Other", please specify.

* What is the name of the organisation you work for?

If applicable, please provide the name of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics /School of Philology/University of Athens.

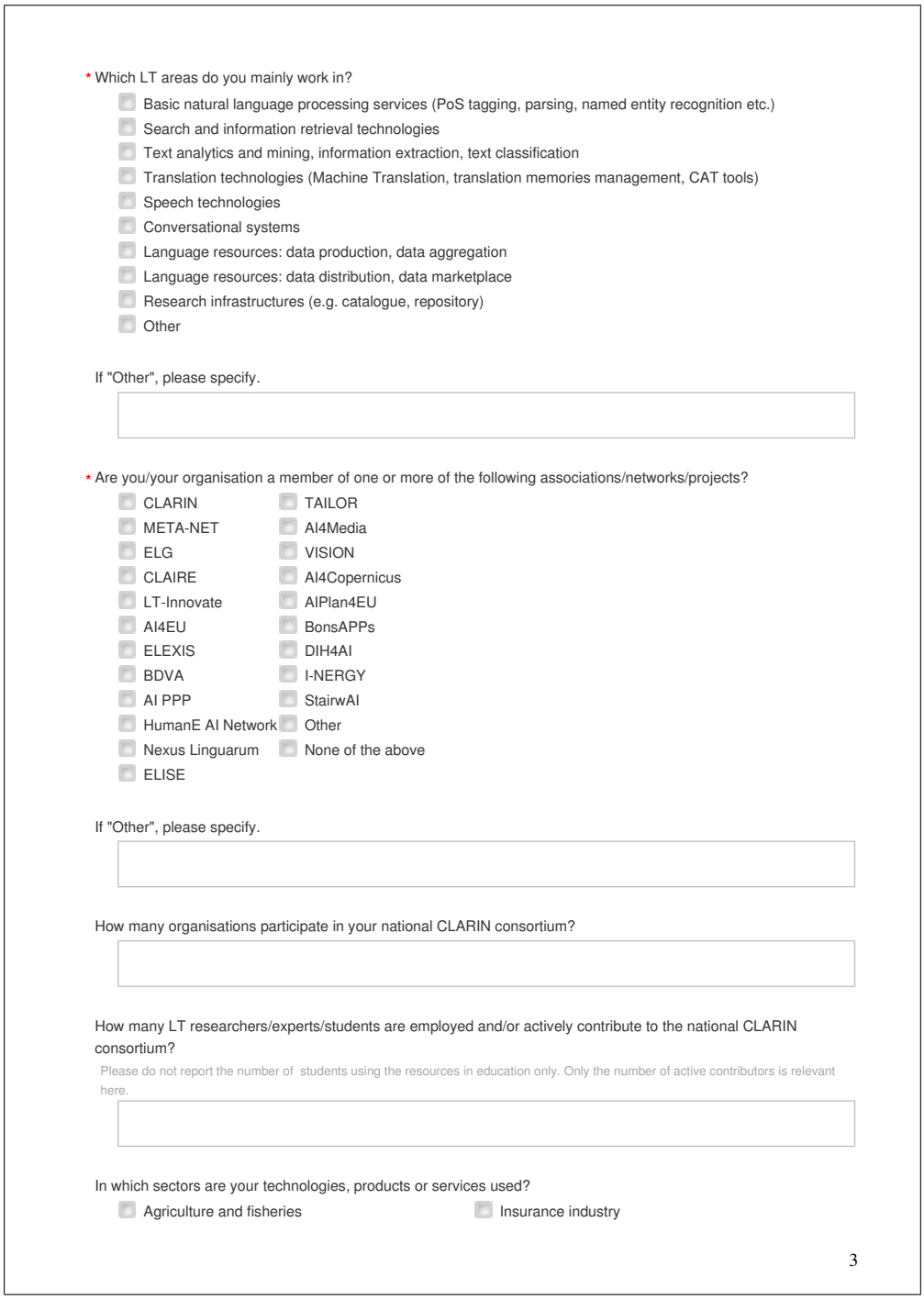
* Where is your organisation's headquarters based?

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

If "Other", please specify.

2

Figure 15: LT developers: full survey as published (page 2/9)



* Which LT areas do you mainly work in?

- Basic natural language processing services (PoS tagging, parsing, named entity recognition etc.)
- Search and information retrieval technologies
- Text analytics and mining, information extraction, text classification
- Translation technologies (Machine Translation, translation memories management, CAT tools)
- Speech technologies
- Conversational systems
- Language resources: data production, data aggregation
- Language resources: data distribution, data marketplace
- Research infrastructures (e.g. catalogue, repository)
- Other

If "Other", please specify.

* Are you/your organisation a member of one or more of the following associations/networks/projects?

- CLARIN
- META-NET
- ELG
- CLAIRE
- LT-Innovate
- AI4EU
- ELEXIS
- BDVA
- AI PPP
- HumanE AI Network
- Nexus Linguarum
- ELISE
- TAILOR
- AI4Media
- VISION
- AI4Copernicus
- AIPlan4EU
- BonsAPPs
- DIH4AI
- I-NERGY
- StairwAI
- Other
- None of the above

If "Other", please specify.

How many organisations participate in your national CLARIN consortium?

How many LT researchers/experts/students are employed and/or actively contribute to the national CLARIN consortium?

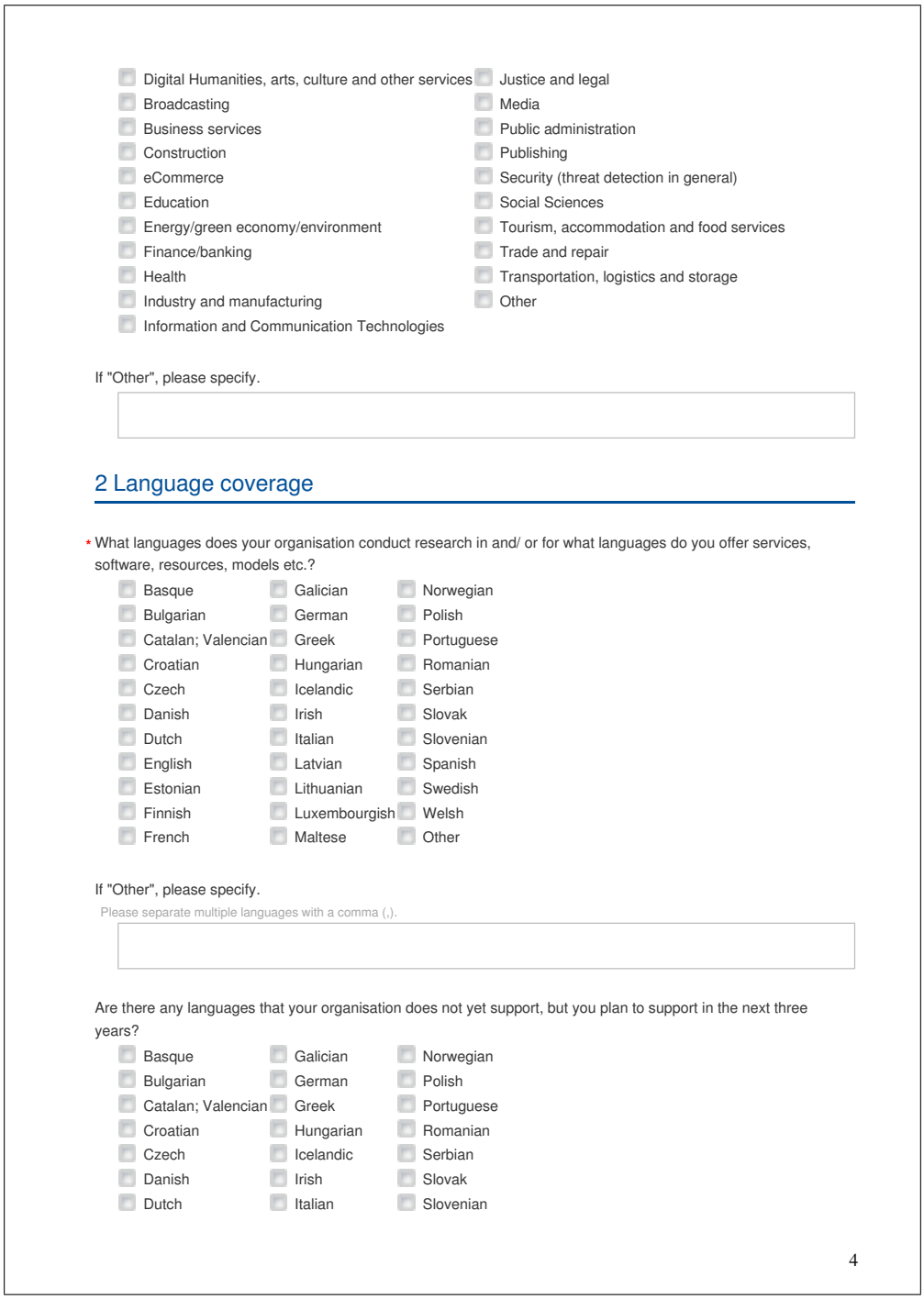
Please do not report the number of students using the resources in education only. Only the number of active contributors is relevant here.

In which sectors are your technologies, products or services used?

- Agriculture and fisheries
- Insurance industry

3

Figure 16: LT developers: full survey as published (page 3/9)



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

If "Other", please specify.

2 Language coverage

* What languages does your organisation conduct research in and/ or for what languages do you offer services, software, resources, models etc.?

- [ ] Basque
- [ ] Bulgarian
- [ ] Catalan; Valencian
- [ ] Croatian
- [ ] Czech
- [ ] Danish
- [ ] Dutch
- [ ] English
- [ ] Estonian
- [ ] Finnish
- [ ] French
- [ ] Galician
- [ ] German
- [ ] Greek
- [ ] Hungarian
- [ ] Icelandic
- [ ] Irish
- [ ] Italian
- [ ] Latvian
- [ ] Lithuanian
- [ ] Luxembourgish
- [ ] Maltese
- [ ] Norwegian
- [ ] Polish
- [ ] Portuguese
- [ ] Romanian
- [ ] Serbian
- [ ] Slovak
- [ ] Slovenian
- [ ] Spanish
- [ ] Swedish
- [ ] Welsh
- [ ] Other

If "Other", please specify.

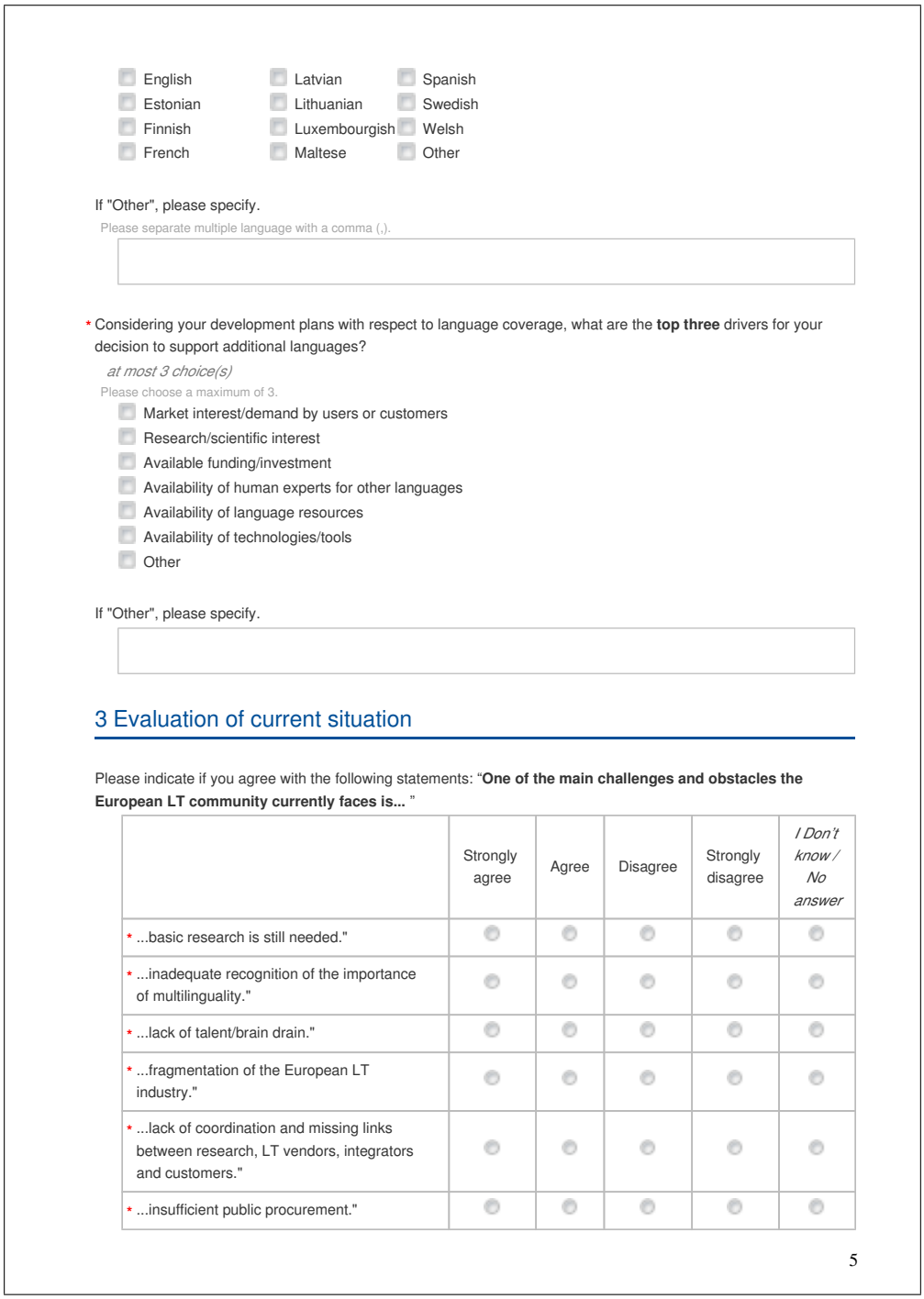
Please separate multiple languages with a comma (,).

Are there any languages that your organisation does not yet support, but you plan to support in the next three years?

- [ ] Basque
- [ ] Bulgarian
- [ ] Catalan; Valencian
- [ ] Croatian
- [ ] Czech
- [ ] Danish
- [ ] Dutch
- [ ] Galician
- [ ] German
- [ ] Greek
- [ ] Hungarian
- [ ] Icelandic
- [ ] Irish
- [ ] Italian
- [ ] Norwegian
- [ ] Polish
- [ ] Portuguese
- [ ] Romanian
- [ ] Serbian
- [ ] Slovak
- [ ] Slovenian

4

Figure 17: LT developers: full survey as published (page 4/9)



English
Estonian
Finnish
French
Latvian
Lithuanian
Luxembourgish
Maltese
Spanish
Swedish
Welsh
Other

If "Other", please specify.

Please separate multiple language with a comma (,).

* Considering your development plans with respect to language coverage, what are the **top three** drivers for your decision to support additional languages?

at most 3 choice(s)

Please choose a maximum of 3.

- Market interest/demand by users or customers
- Research/scientific interest
- Available funding/investment
- Availability of human experts for other languages
- Availability of language resources
- Availability of technologies/tools
- Other

If "Other", please specify.

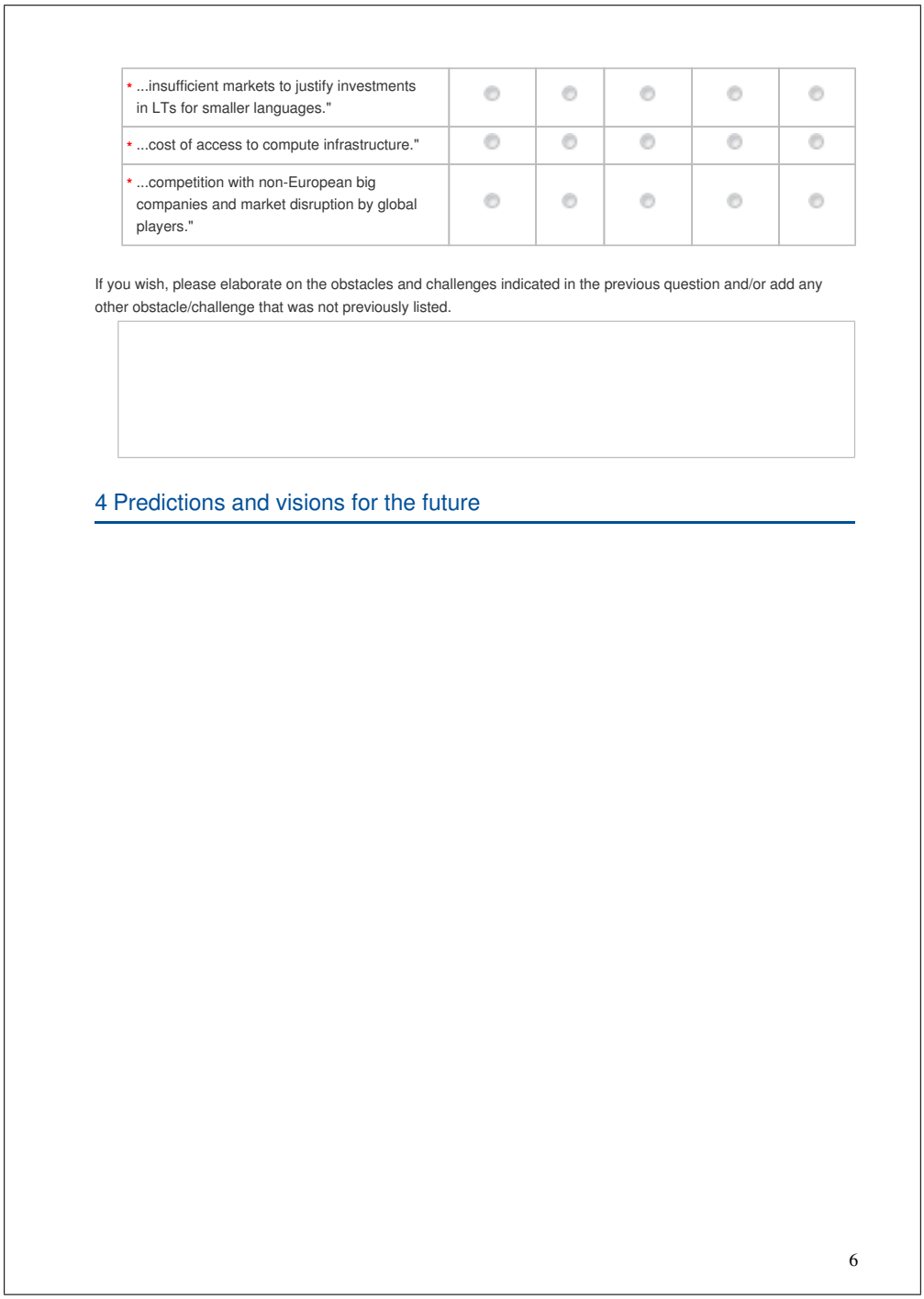
3 Evaluation of current situation

Please indicate if you agree with the following statements: "**One of the main challenges and obstacles the European LT community currently faces is...** "

	Strongly agree	Agree	Disagree	Strongly disagree	*I Don't know / No answer*
* ...basic research is still needed."					
* ...inadequate recognition of the importance of multilinguality."					
* ...lack of talent/brain drain."					
* ...fragmentation of the European LT industry."					
* ...lack of coordination and missing links between research, LT vendors, integrators and customers."					
* ...insufficient public procurement."					

5

Figure 18: LT developers: full survey as published (page 5/9)



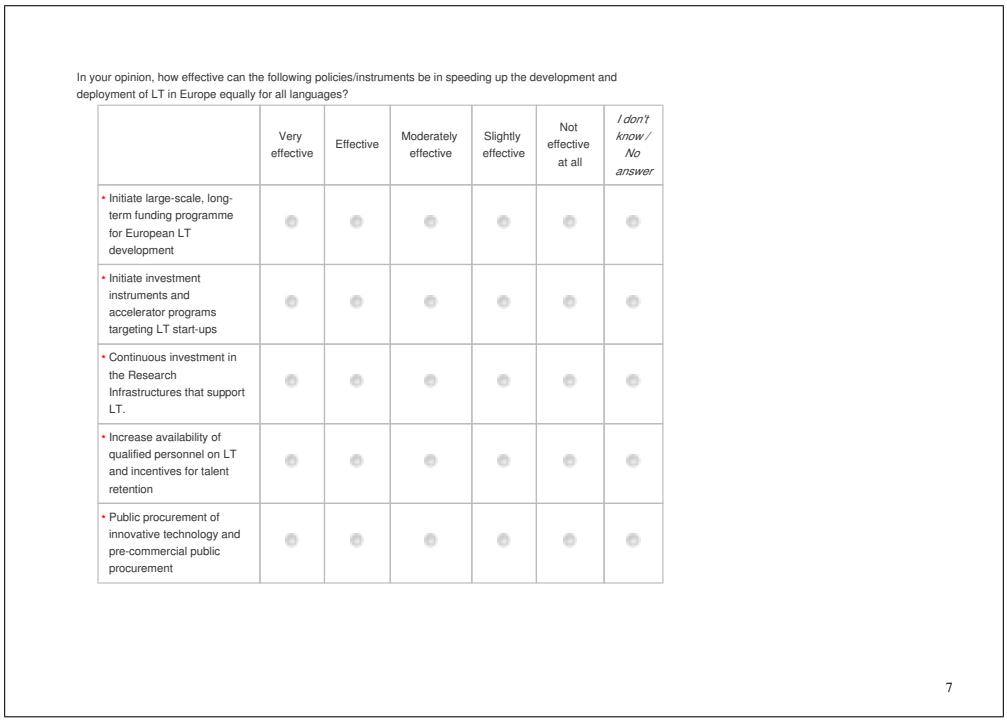
* ...insufficient markets to justify investments in LTs for smaller languages."					
* ...cost of access to compute infrastructure."					
* ...competition with non-European big companies and market disruption by global players."					

If you wish, please elaborate on the obstacles and challenges indicated in the previous question and/or add any other obstacle/challenge that was not previously listed.

4 Predictions and visions for the future

6

Figure 19: LT developers: full survey as published (page 6/9)

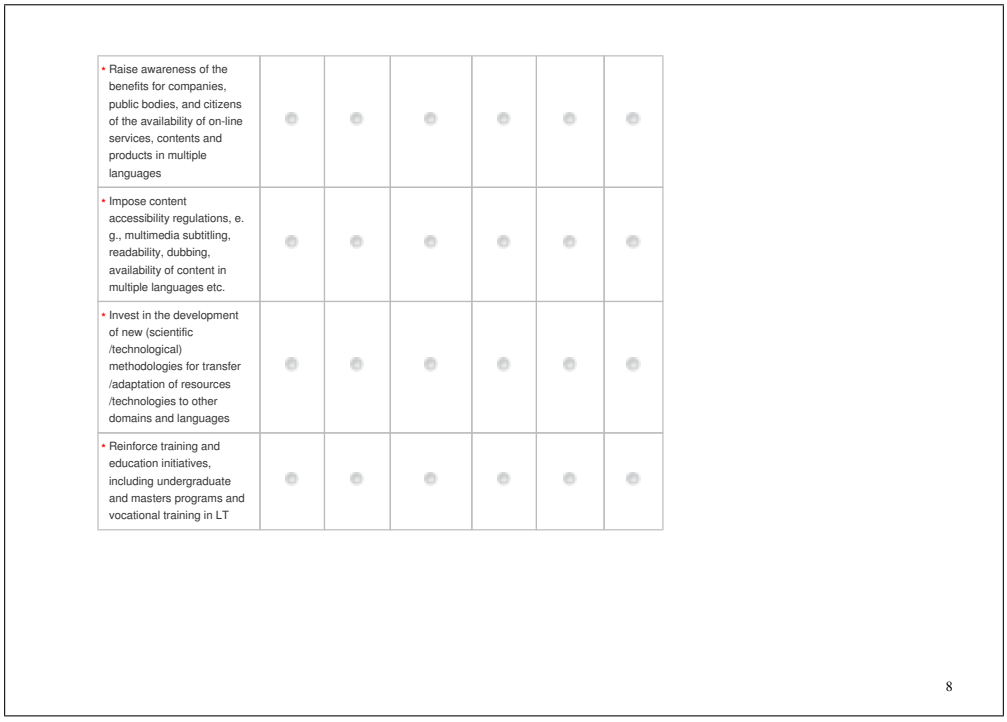


In your opinion, how effective can the following policies/instruments be in speeding up the development and deployment of LT in Europe equally for all languages?

	Very effective	Effective	Moderately effective	Slightly effective	Not effective at all	*I don't know / No answer*
• Initiate large-scale, long-term funding programme for European LT development						
• Initiate investment instruments and accelerator programs targeting LT start-ups						
• Continuous investment in the Research Infrastructures that support LT.						
• Increase availability of qualified personnel on LT and incentives for talent retention						
• Public procurement of innovative technology and pre-commercial public procurement						

7

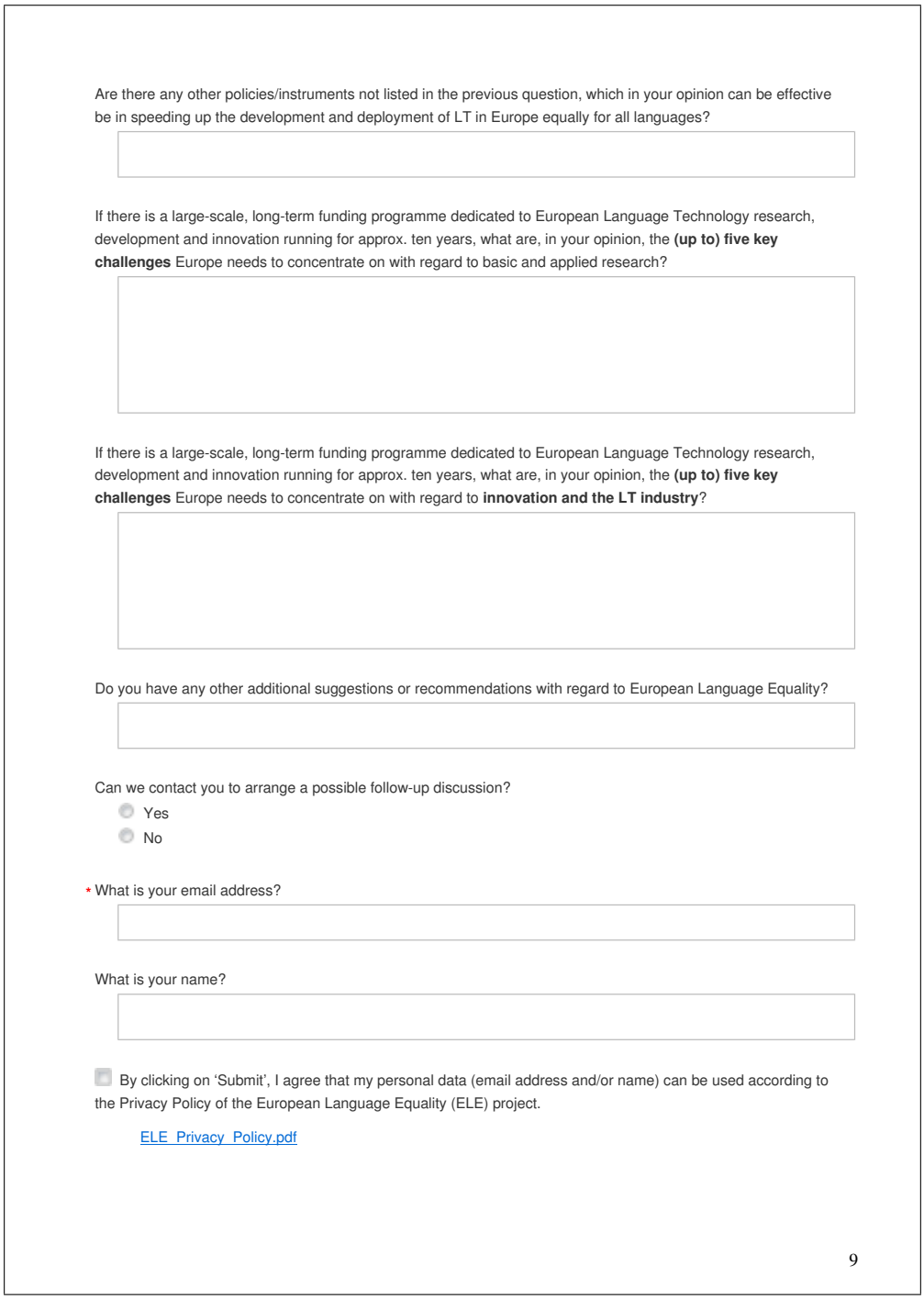
Figure 20: LT developers: full survey as published (page 7/9)



• Raise awareness of the benefits for companies, public bodies, and citizens of the availability of on-line services, contents and products in multiple languages						
• Impose content accessibility regulations, e. g., multimedia subtitling, readability, dubbing, availability of content in multiple languages etc.						
• Invest in the development of new (scientific /technological) methodologies for transfer /adaptation of resources /technologies to other domains and languages						
• Reinforce training and education initiatives, including undergraduate and masters programs and vocational training in LT						

8

Figure 21: LT developers: full survey as published (page 8/9)



Are there any other policies/instruments not listed in the previous question, which in your opinion can be effective be in speeding up the development and deployment of LT in Europe equally for all languages?

If there is a large-scale, long-term funding programme dedicated to European Language Technology research, development and innovation running for approx. ten years, what are, in your opinion, the **(up to) five key challenges** Europe needs to concentrate on with regard to basic and applied research?

If there is a large-scale, long-term funding programme dedicated to European Language Technology research, development and innovation running for approx. ten years, what are, in your opinion, the **(up to) five key challenges** Europe needs to concentrate on with regard to **innovation and the LT industry**?

Do you have any other additional suggestions or recommendations with regard to European Language Equality?

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

- Yes
- No

* What is your email 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

9

Figure 22: LT developers: full survey as published (page 9/9)

B LT Developers Survey: Additional Material

Type of organisation	Answers	%
Research centre (independent)	31	10%
University or academic research	203	63%
Large enterprise	17	5%
SME	55	17%
Other	15	5%
Total	321	100%

Table 7: Breakdown of answers to "Which of the following best describes the type of organisation you work for?" (mandatory closed question)

Table 8: All organisations represented by the LT Developers survey respondents (Departments. faculties. institutes or other units of the same organisation are not listed as different entities.)

LT Developers – Organisations		
4i intelligent insights	KU Leuven	Trust Stamp
A Data Pro	LAB University of Applied Sciences	UAB "Proit"
Accademia della Crusca	Laboratoire Hubert Curien	Umeå university
Adam Mickiewicz University	Le français des affaires / CCI Paris Ile-de-France	undisclosed
AGI - Information Management Consultants	Lexical Computing	Universidad de Alicante
Ai4Value	Lingsoft	Universidad de Jaén
ALAN Analytics s.r.o	Linköping University	Universidad de Murcia
AlfaNum	LT3. Ghent University	Università Cattolica del Sacro Cuore
Almannarómur / The Voice of the People	Lucid	Università degli studi di Torino
Amu	Lund University Humanities Lab	Universität Hamburg
Analyse & Tal	Luxembourg Institute of Science and Technology	Universitat Jaume I
Aristotle University of Thessaloniki	Maastricht University	Universitat Politècnica de Catalunya
Athena Research Center	Macedonian Academy of Sciences and Arts	Universitat Pompeu Fabra
Athens University of Economics and Business	magiquo data live s.a	Université Parus-Saclay
Audio-Visual Machine Perception Limited	Masaryk University	University "Politehnica" of Bucharest
Austrian Research Institute for Artificial Intelligence	Massey University	University of Alcalá

Continued on next page

Table 8 – *Continued from previous page*

LT Developers – Organisations		
Autonomous University of Barcelona	Meddal.com	University of Amsterdam
Bangor University	Medical University of Vienna	University of Antwerp
Barcelona SUpercomputing Center	Meltwater Group	University of Belgrade
Bulgarian Academy of Sciences	Memsource a.s.	University of Bergen
Center for Cultural Protection and Technological Development of Georgian State Languages	Moravská zemská knihovna v Brně	University of Brasília
Center for the Greek Language	Morningsun Technology GmbH	University of Bristol
Centre for Aromanian Language and Culture in Bulgaria - CALCB	Mozaika	University of Coimbra
Cerence	Multilingues21. Lda.	University of Copenhagen
CERTH	Národní filmový archiv. Prague	University of Edinburgh
Charles University	National and Kapodistrian University of Athens	University of Essex
Ciklopea d.o.o.	National University of Ireland Galway	University of Gothenburg
CIP4N GmbH Deutschland	NCSR “Demokritos”	University of Groningen
Cloudwise	Netherlands eScience Center	University of Haifa
CNRS	nettle.ai	University of Helsinki
Consiglio Nazionale delle Ricerche (CNR)	New York University	University of Jaén
Convforth SRL	Nico van de Water Linguistic Services	University of Library Studies and Information Technologies
Cornelistools B.V.	Omilia	University of Lisbon
Cyprus University of Technology / Unesco Chair on Digital Cultural heritage	Pangeanic	University of Ljubljana
Czech Academy of Sciences	Phonexia s.r.o.	University of Luxembourg
Dalle Molle Institute for Artificial Intelligence	Polish Academy of Sciences	University of Malta
Danish Language Council	Polish-Japanese Academy of Information Technology	University of Manchester
Darmstadt University of Applied Sciences	Research Institute for Artificial Intelligence “Mihai Draganescu”. Romanian Academy	University of Maribor
Deloitte	Royal Netherlands Academy of Arts and Sciences	University of Nova Gorica
DFKI	RTL	University of Patras
Dublin City University	Ruhr-Universität Bochum	University of Pécs
E4 Computer Engineering SpA	Russian Academy of Sciences	University of Porto

Continued on next page

Table 8 – *Continued from previous page*

LT Developers – Organisations		
EDIA	RWS	University of Primorska
emagine GmbH	Samsung Electronics	University of Santiago de Compostela
EML Speech Technology GmbH	Sberbank	University of Sheffield
Ensoul	SciFY PNPC	University of St-Etienne
Entefy	Scriptix	University of Stuttgart
EPFL / Idiap Research Institute	SEMLAB	University of Szeged
Eurac Research	Serbian Academy of Sciences and Arts	University of Tartu
Fondazione Bruno Kessler	Sign Time GmbH	University of the Aegean
FORTH	Sinequa	University of the Basque Country (UPV/EHU)
Fraunhofer Gesellschaft	Sirma AI (Ontotext)	University of Twente
Free University of Bozen-Bolzano	Slovak Academy of Sciences	University of Vienna
Furtwangen University	Slovenian Academy of Sciences and Arts	University of Vigo
Globalese	Spanish Society for Natural Language Processing (SEPLN)	University of Warsaw
Goethe-University Frankfurt	SpeechTech	University of West Bohemia
Grammatek	Stockholm University	University of Zagreb
HENSOLDT Analytics	Sunda Systems Oy	University of Zurich
Heriot-Watt University	Syllabs	University Politehnica Bucharest
HiTZ Basque Center for Language Technology	Talkie.ai	University Ss. Cyril and Methodius
Hof University of Applied Sciences	Tallinn University of Technology	Uppsala University
Human Centered Data Analytics. Centrum Wiskunde & Informatica	Technische Universität Dresden	Utrecht University
Hungarian Research Centre for Linguistics	Text Technology Lab / Department of Computer Science and Mathematics / Goethe University Frankfurt	Vicomtech
Ilia State University	The Árni Magnússon Institute for Icelandic Studies	Vilnius university
Institute of Philosophy. CAS	The Citizens' Association for the Promotion of Roma Education "Otaharin"	Visma
Institute of the Lithuanian Language	The Language Council of Sweden at the Institute for Language and Folklore (ISOF)	VÓCALI Sistemas Inteligentes
Intelartes Sprl	The MAMA AI. SE	Vocapia Research
Ionian University	The National Library of the Czech Republic	Vytautas Magnus University

Continued on next page

Table 8 – *Continued from previous page*

LT Developers – Organisations		
Jožef Stefan Institute	The Welsh Government	Wikimedia Deutschland
JSC I-Teco	Tilburg University	WordFinder Software International AB
K Dictionaries - Lexicala	TILDE	Worldwide Bildungswerk
KantanAI	TMServe	Wrocław University of Science and Technology
Kempelen Institute of Intelligent Technologies	Toros University. Turkey	WWU Münster
Kielikone Oy	Trinity College Dublin	Zurich University of Applied Sciences
KTH Royal Institute of Technology		

Country	Respondents
Spain	35
Germany	28
Greece	22
Czechia	19
Netherlands	18
France	17
Lithuania	13
Italy	12
Poland	11
United Kingdom	11
Sweden	10
Switzerland	9
Bulgaria	8
Denmark	8
Ireland	8
Finland	7
Portugal	7
Romania	7
Slovenia	7
Austria	6
Latvia	5
Slovakia	5
United States of America	5
Belgium	4
Croatia	3
Estonia	3
Hungary	3
Iceland	3
Luxembourg	3
Malta	3
Russia	3
Serbia	3
Cyprus	2
Georgia	2
Israel	2
North Macedonia	2
Bosnia and Herzegovina	1
Brazil	1
New Zealand	1
Norway	1
Turkey	1

Table 9: Breakdown of answers to “Where is your organisation’s headquarter based?” (mandatory closed question. plus “if other” as optional open-ended question)

Network, association, project
4CH
ARIADNEplus
Association for the Advancement of Artificial Intelligence (AAAI)
Asscition of Computational Linguistics (ACL)
British Irish Council's Indigenous Minority Languages Group
Cultivating Research & Equity in Sign-related Technology (CREST Network) (by Gallaudet university)
Digital Research Infrastructure for the Arts and Humanities (DARIAH)
ESPERANTO
EU SME Focus Group on AI (DigitalSME Network)
European Association for Machine Translation (EAMT)
European Association of Terminology
European Holocaust Research Infrastructure (EHRI)
European Language Resource Coordination (ELRC)
European Language Resources Association (ELRA)
IEEE
Infobalt Association
Interactive Natural Language Technology for Explainable Artificial Intelligence (NL4XAI. a Marie Curie ITN)
International Speech Communication Association (ISCA)
Language In The Human-Machine Era (LITHME)
Multi-task. Multilingual. Multi-modal Language Generation (Multi3Generation)
Network to Promote Linguistic Diversity (NPLD)
Neural Translation for the EU (NTEU)
NOTaS
Red Temática en Tecnologías del Habla (RTTH)
Sociedad Española para el Procesamiento del Lenguaje Natural (SEPLN)
TimeMachine
UNESCO's year of indigenous languages
WOSC

Table 10: Additional networks, projects and association the survey respondents participate to.

Sectors	Number of mentions
Information and Communication Technologies	180
Digital Humanities, arts, culture and other services	171
Education	154
Health	101
Media	98
Social Sciences	89
Public administration	75
Business services	71
Justice and legal	70
Finance/banking	57
Broadcasting	50
Publishing	49
eCommerce	42
Industry and manufacturing	42
Tourism, accommodation and food services	35
Insurance industry	28
Security (threat detection in general)	28
Transportation, logistics and storage	24
Other	21
Energy/green economy/environment	16
Agriculture and fisheries	13
Construction	13
Trade and repair	7

Table 11: Sectors in which technologies, products or services are used

Languages	Number of mentions
Basque	5
Galician	4
Norwegian	5
Bulgarian	6
German	18
Polish	13
Catalan	7
Greek	4
Portuguese	9
Croatian	6
Hungarian	8
Romanian	9
Czech	8
Icelandic	5
Serbian	6
Danish	5
Irish	3
Slovak	5
Dutch	4
Italian	9
Slovenian	4
English	13
Latvian	4
Spanish	10
Estonian	5
Lithuanian	7
Swedish	5
Finnish	4
Luxembourgish	4
Welsh	3
French	20
Maltese	3
Other	38

Table 12: Breakdown of answers to “Are there any languages that your organisation does not yet support, but you plan to support in the next three years? If Other, please specify.”

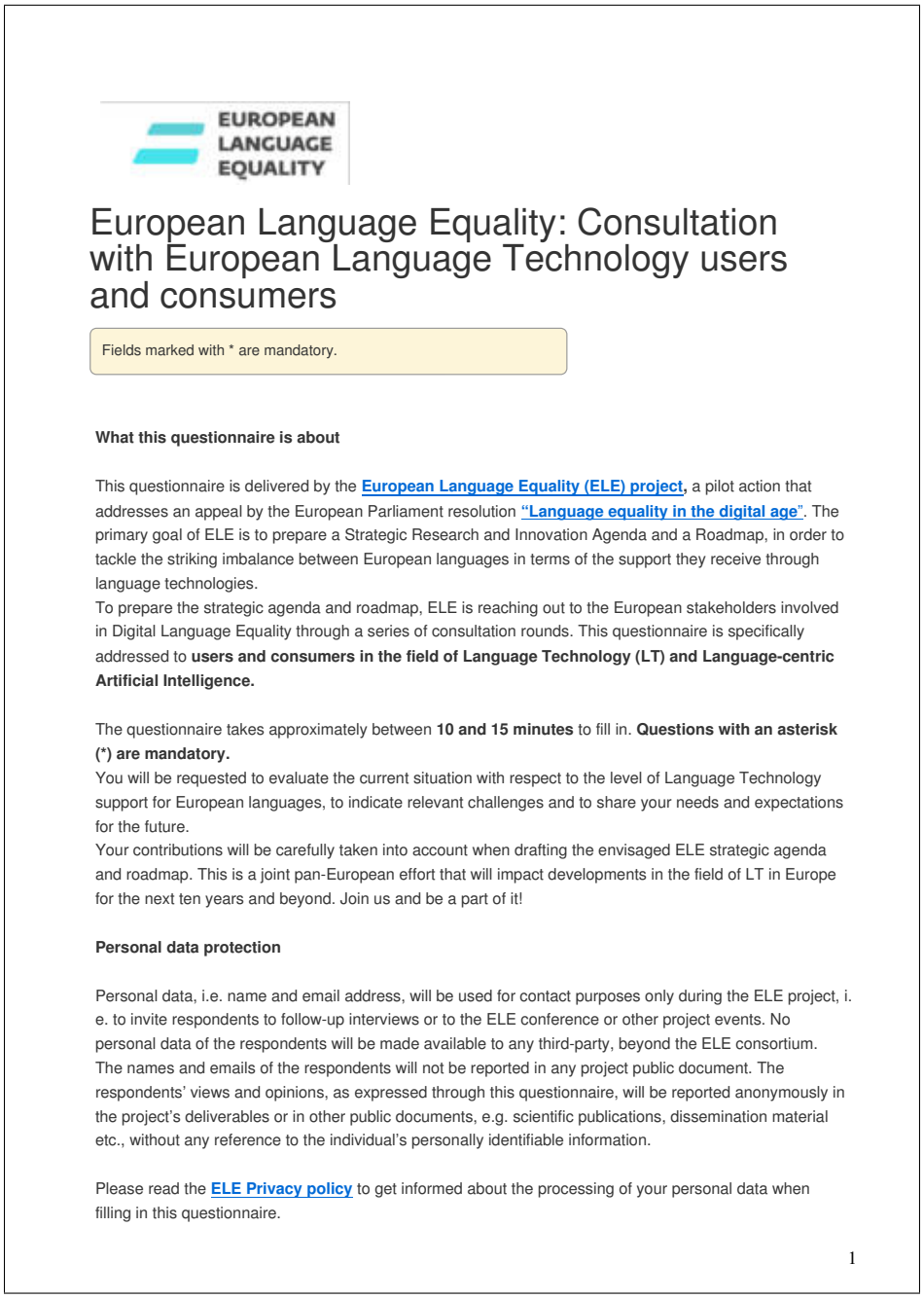
	Strongly agree	Agree	Disagree	Strongly disagree	N/A	Av. score (4-1)
basic research is still needed	122	146	30	7	16	3.3
inadequate recognition of the importance of multilinguality	90	150	56	6	19	3.1
lack of talent/brain drain	37	92	110	44	38	2.4
fragmentation of the European LT industry	49	151	37	8	76	3.0
lack of coordination and missing links between research, LT vendors, integrators and customers	60	142	53	6	60	3.0
insufficient public procurement	68	111	49	5	88	3.0
insufficient markets to justify investments in LTs for smaller languages	95	116	59	14	37	3.0
cost of access to compute infrastructure	41	112	99	18	51	2.7
competition with non-European big companies and market disruption by global players	102	125	44	7	43	3.2

Table 13: Answers to the question: “Please indicate if you agree with the following statements: “One of the main challenges and obstacles the European LT community currently faces is…” (mandatory closed question, answers provided on a four-point scale. plus “I don’t know/No answer”)

	Very effective	Effective	Moderately effective	Slightly effective	Not effective at all	N/A	Av. score (5-1)
Initiate large-scale, long-term funding programme for European LT development	148	110	32	18	4	9	4.2
Initiate investment instruments and accelerator programs targeting LT start-ups	79	118	77	17	8	22	3.8
Continuous investment in the Research Infrastructures that support LT	129	144	21	19	1	7	4.21
Increase availability of qualified personnel on LT and incentives for talent retention	97	139	43	17	5	20	4.0
Public procurement of innovative technology and pre-commercial public procurement	65	110	60	26	5	55	3.8
Raise awareness of the benefits for companies, public bodies, and citizens of the availability of on-line services, contents and products in multiple languages	82	105	78	27	11	18	3.7
Impose content accessibility regulations, e. g. multimedia subtitling, readability, dubbing, availability of content in multiple languages etc.	83	95	71	36	10	26	3.7
Invest in the development of new (scientific/technological) methodologies for transfer/adaptation of resources/technologies to other domains and languages	105	133	48	17	4	14	4.0
Reinforce training and education initiatives, including undergraduate and masters programs and vocational training in LT	110	116	62	19	3	11	4.0

Table 14: Answers to the question: “In your opinion. how effective can the following policies or instruments be in speeding up the development and deployment of LT in Europe equally for all languages?” (mandatory closed question, answers provided on a five-point scale, plus “I don’t know/No answer”)

C LT Users and Consumers Survey



EUROPEAN LANGUAGE EQUALITY

European Language Equality: Consultation with European Language Technology users and consumers

Fields marked with * are mandatory.

What this questionnaire is about

This questionnaire is delivered by the **European Language Equality (ELE) project,** a pilot action that addresses an appeal by the European Parliament resolution **"Language equality in the digital age"**. The primary goal of ELE is to prepare a Strategic Research and Innovation Agenda and a Roadmap, in order to tackle the striking imbalance between European languages in terms of the support they receive through language technologies.
To prepare the strategic agenda and roadmap, ELE is reaching out to the European stakeholders involved in Digital Language Equality through a series of consultation rounds. This questionnaire is specifically addressed to **users and consumers in the field of Language Technology (LT) and Language-centric Artificial Intelligence.**

The questionnaire takes approximately between **10 and 15 minutes** to fill in. **Questions with an asterisk (*) are mandatory.**
You will be requested to evaluate the current situation with respect to the level of Language Technology support for European languages, to indicate relevant challenges and to share your needs and expectations for the future.
Your contributions will be carefully taken into account when drafting the envisaged ELE strategic agenda and roadmap. This is a joint pan-European effort that will impact developments in the field of LT in Europe for the next ten years and beyond. Join us and be a part of it!

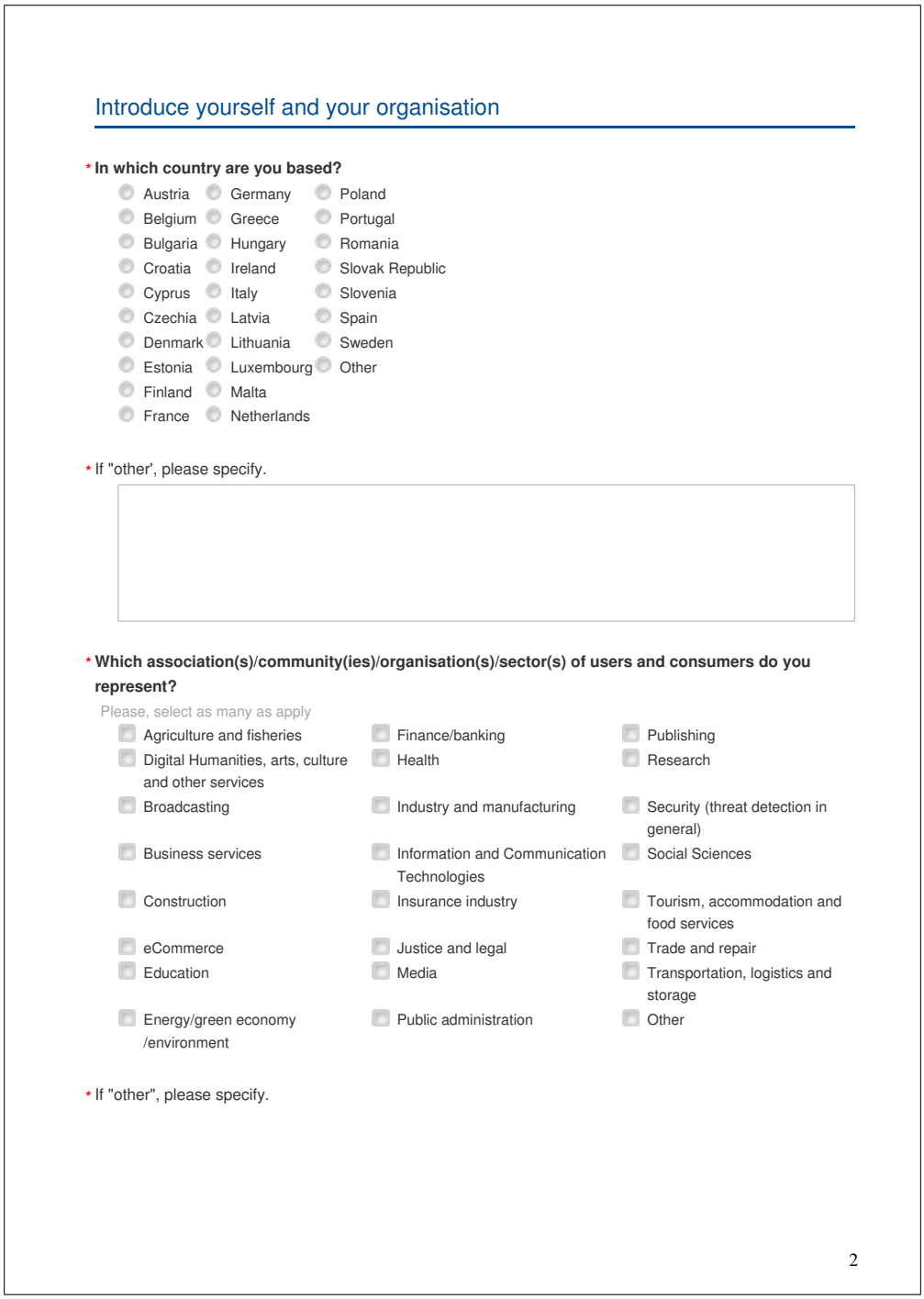
Personal data protection

Personal data, i.e. name and email address, will be used for contact purposes only during the ELE project, i. e. to invite respondents to follow-up interviews or to the ELE conference or other project events. No personal data of the respondents will be made available to any third-party, beyond the ELE consortium. The names and emails of the respondents will not be reported in any project public document. The respondents' views and opinions, as expressed through this questionnaire, will be reported anonymously in the project's deliverables or in other public documents, e.g. scientific publications, dissemination material etc., without any reference to the individual's personally identifiable information.

Please read the **ELE Privacy policy** to get informed about the processing of your personal data when filling in this questionnaire.

1

Figure 23: LT users: full survey as published (page 1/18)



Introduce yourself and your organisation

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

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

* If "other', please specify.

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

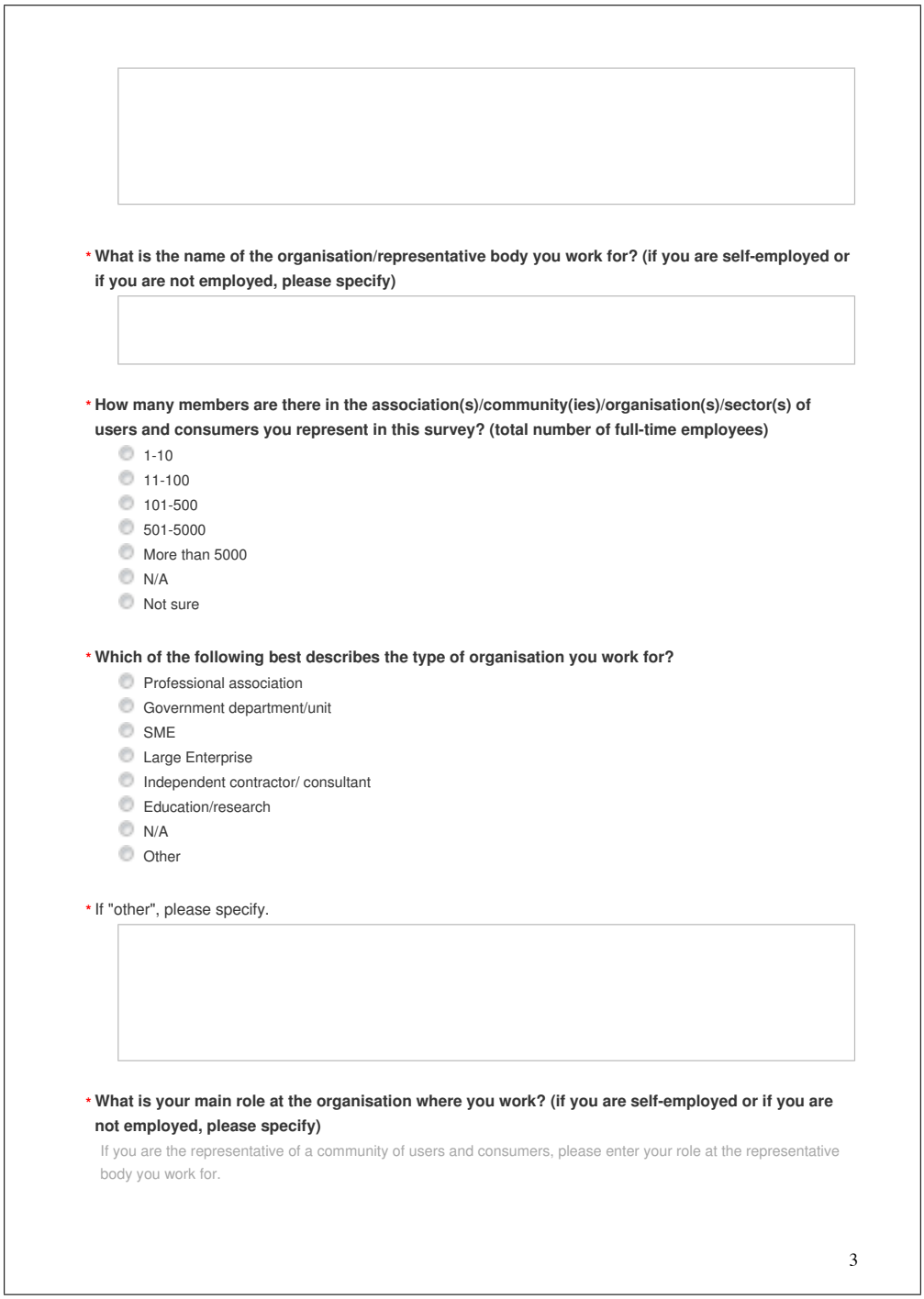
Please, select as many as apply

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

* If "other", please specify.

2

Figure 24: LT users: 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

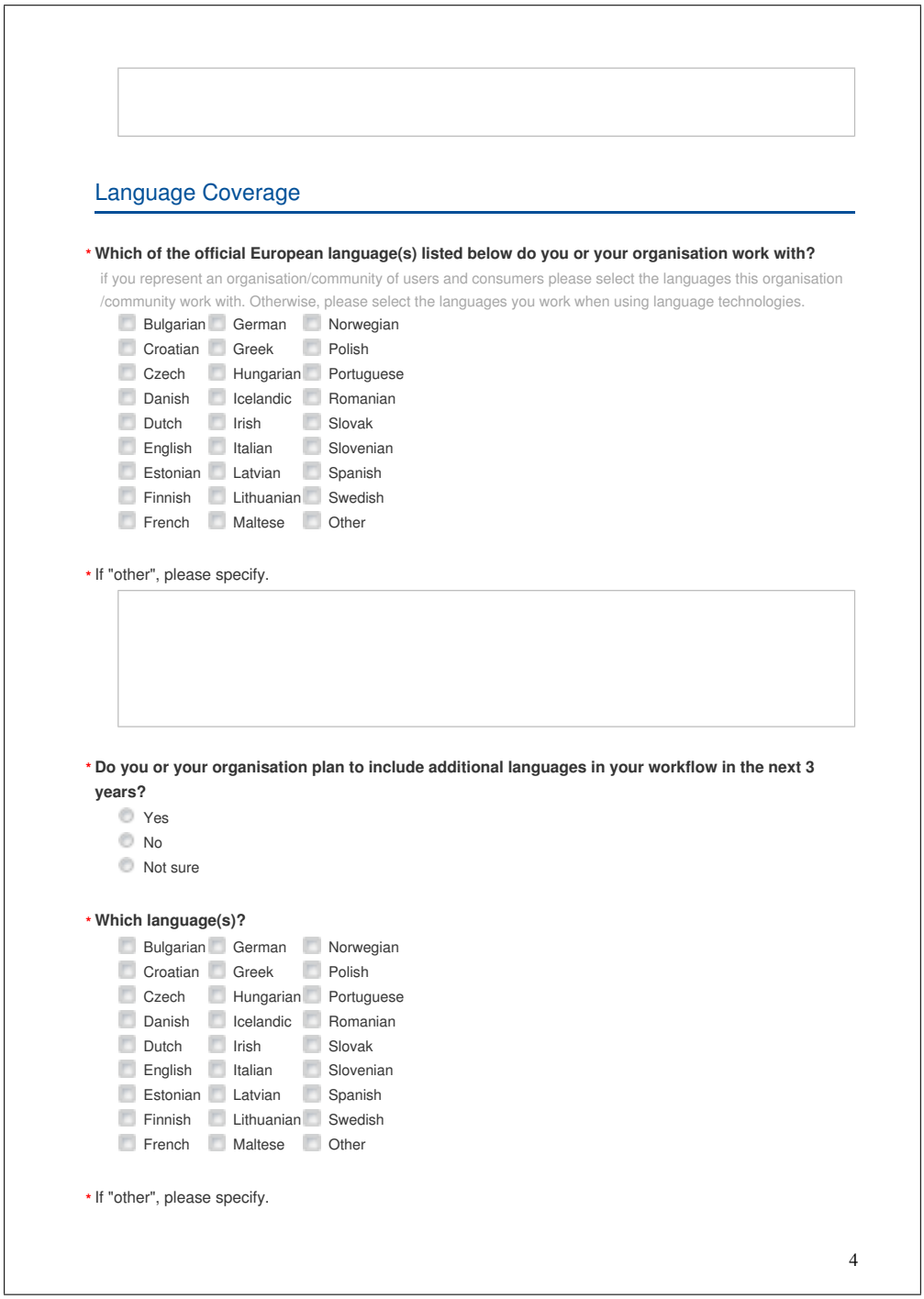
* If "other", please specify.

* **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 25: LT users: 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
- Croatian
- Czech
- Danish
- Dutch
- English
- Estonian
- Finnish
- French
- German
- Greek
- Hungarian
- Icelandic
- Irish
- Italian
- Latvian
- Lithuanian
- Maltese
- Norwegian
- Polish
- Portuguese
- Romanian
- Slovak
- Slovenian
- Spanish
- Swedish
- 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

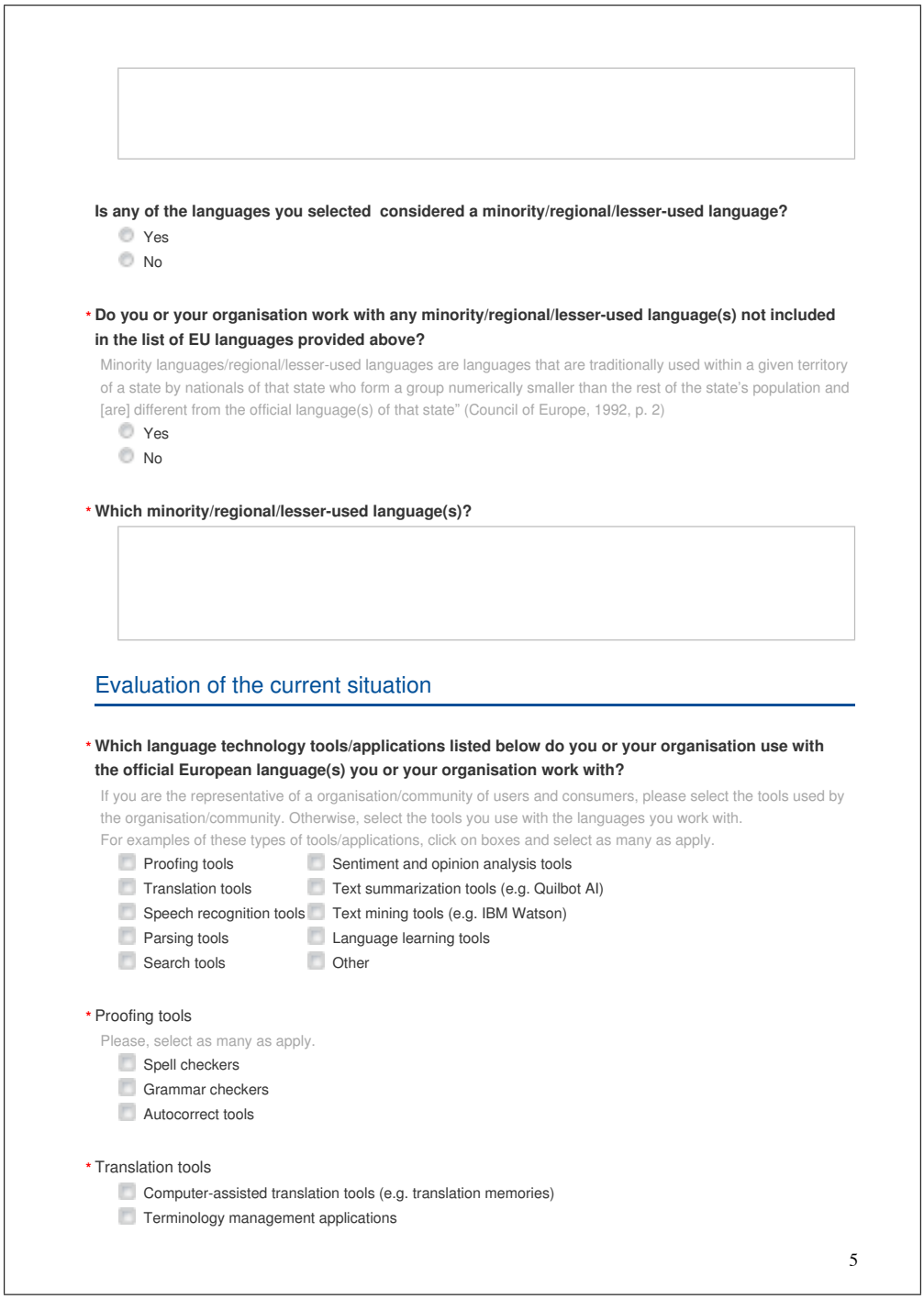
* **Which language(s)?**

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

* If "other", please specify.

4

Figure 26: LT users: 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
- 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

Please, select as many as apply.

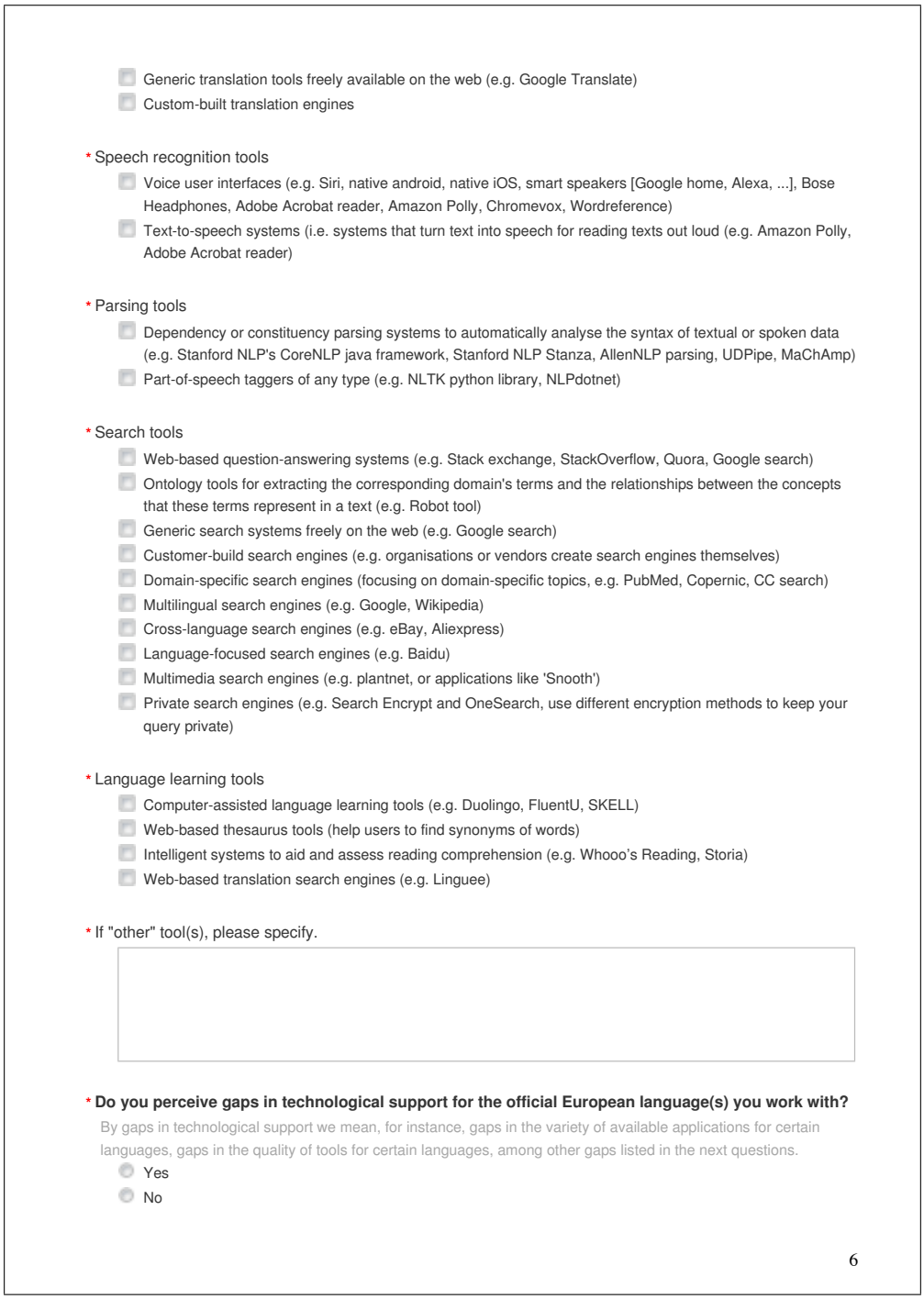
- Spell checkers
- Grammar checkers
- Autocorrect tools

* Translation tools

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

5

Figure 27: LT users: 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)

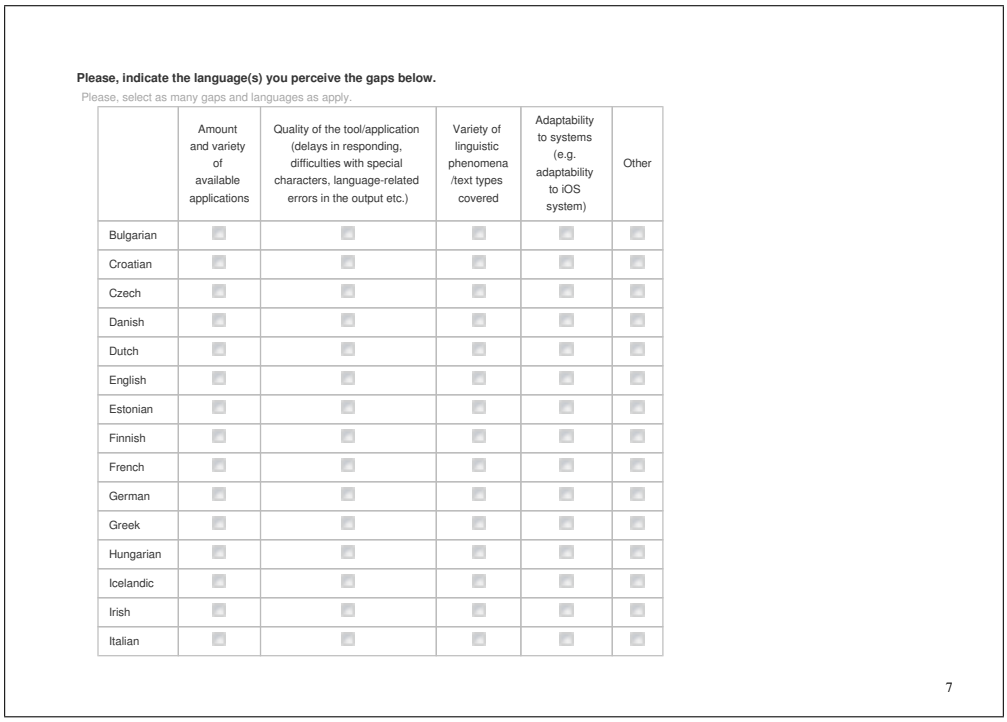
* Language learning tools
  - Computer-assisted language learning tools (e.g. Duolingo, FluentU, SKELL)
  - Web-based thesaurus tools (help users to find synonyms of words)
  - Intelligent systems to aid and assess reading comprehension (e.g. Whooo's Reading, Storia)
  - Web-based translation search 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 28: LT users: full survey as published (page 6/18)



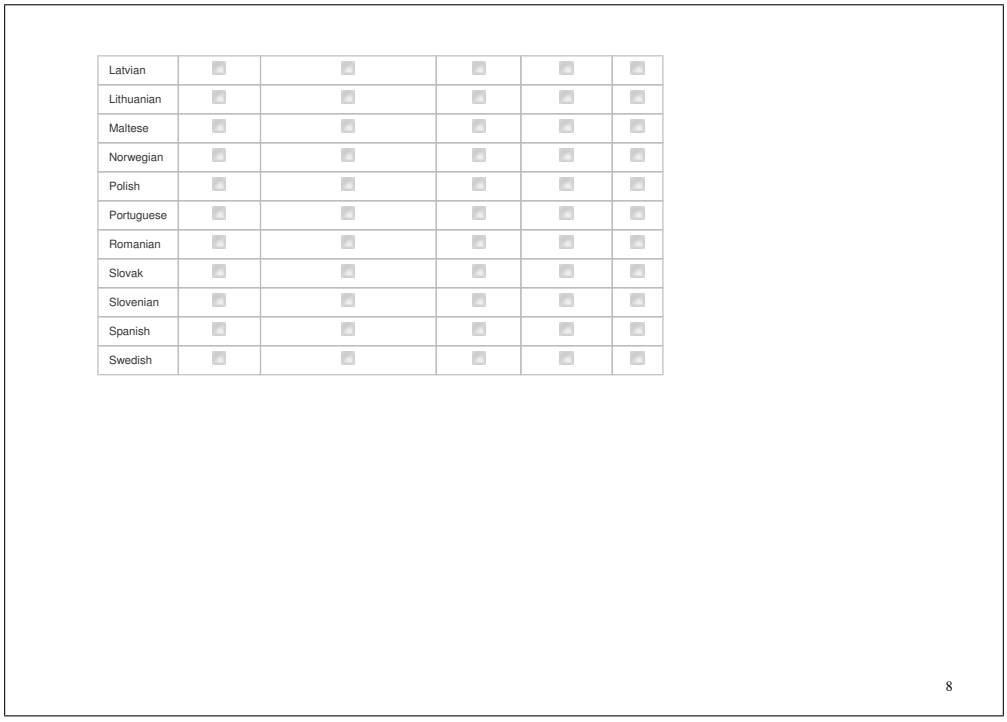
Please, indicate the language(s) you perceive the gaps below.

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 29: LT users: full survey as published (page 7/18)



Latvian					
Lithuanian					
Maltese					
Norwegian					
Polish					
Portuguese					
Romanian					
Slovak					
Slovenian					
Spanish					
Swedish					

8

Figure 30: LT users: 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 31: LT users: full survey as published (page 9/18)

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 32: LT users: 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.

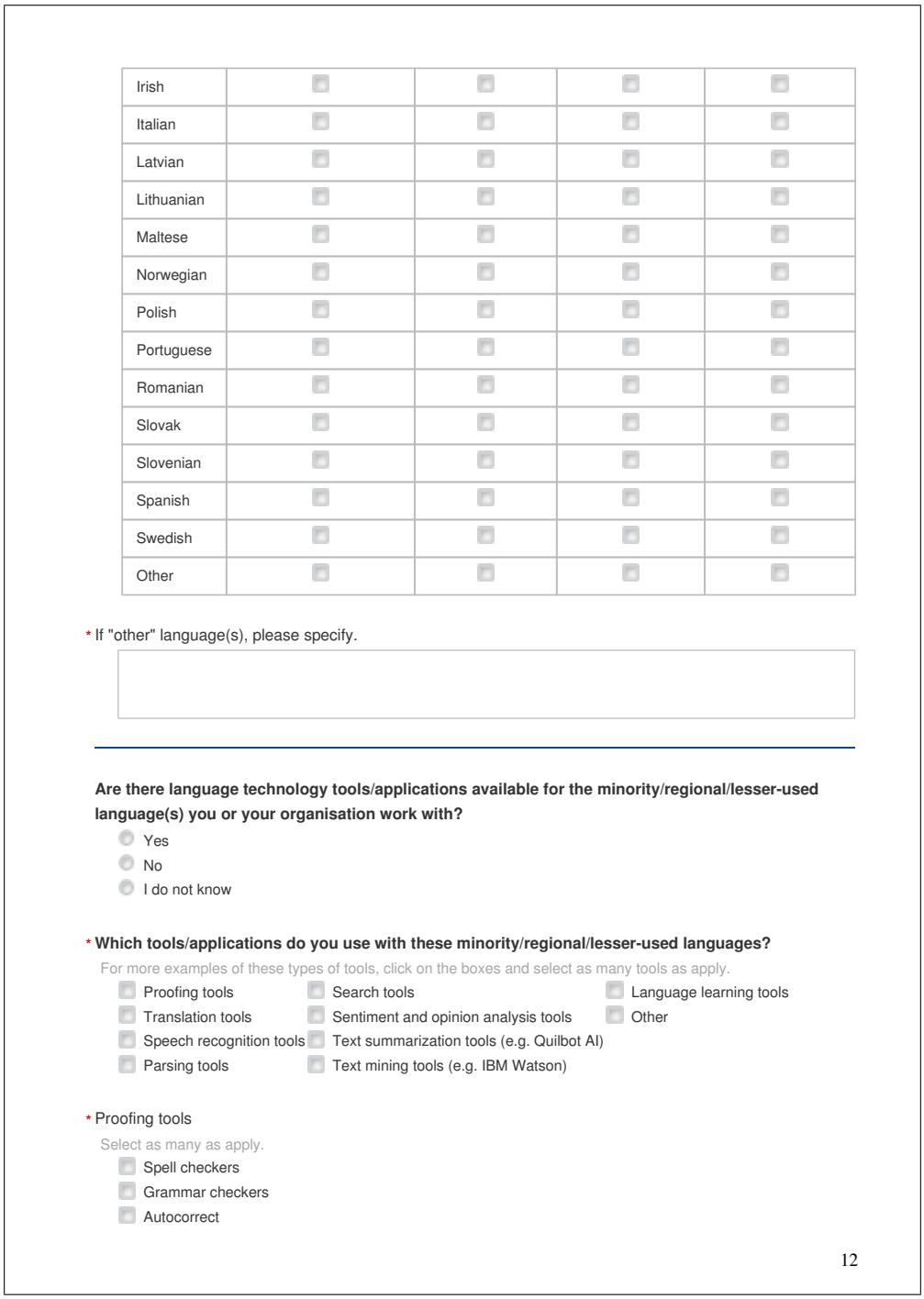
Please indicate for which language(s) you or your organisation use the language technology tools /applications listed below.

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 33: LT users: 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 34: LT users: 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 35: LT users: 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 36: LT users: full survey as published (page 14/18)

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 37: LT users: 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?**

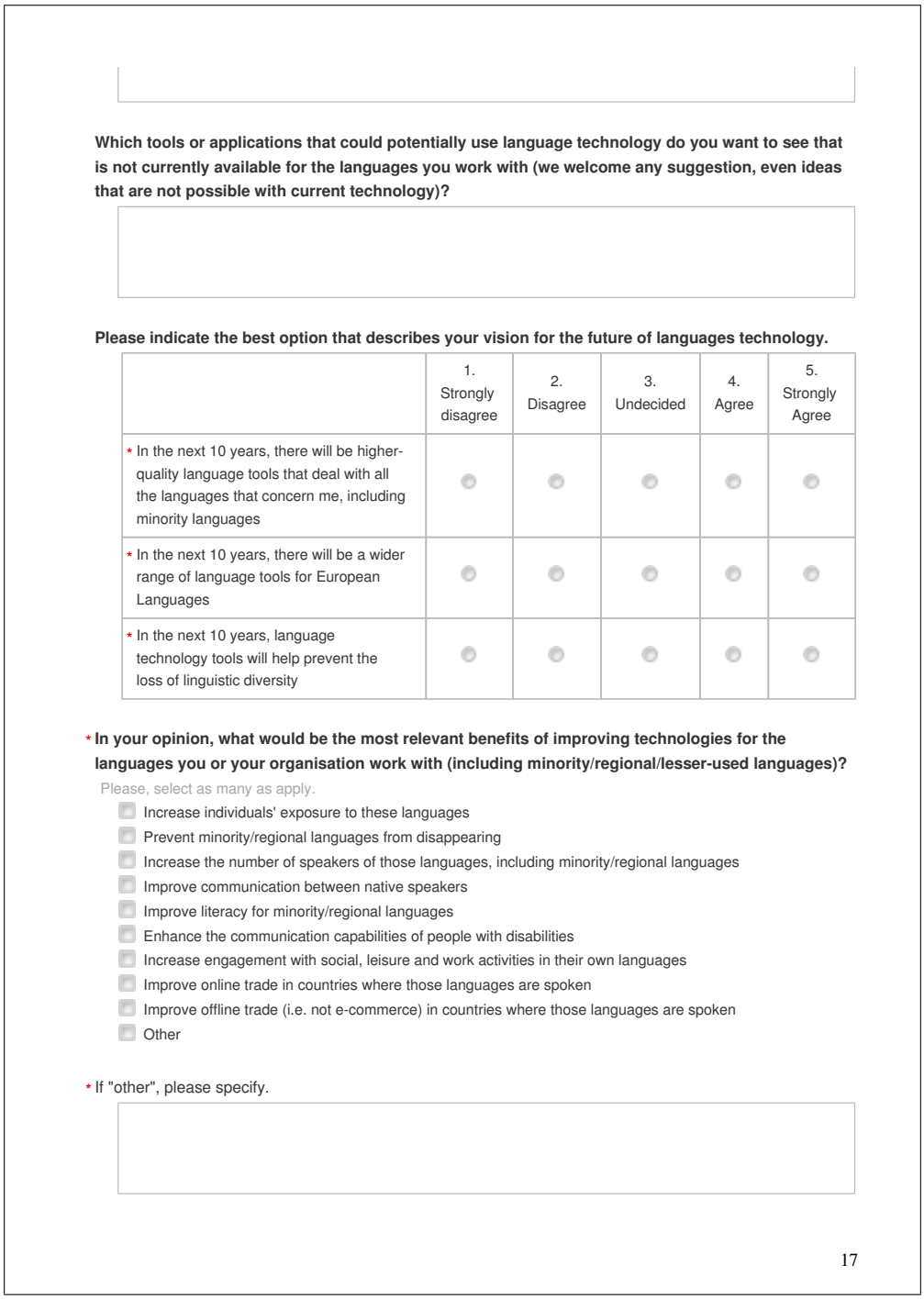
Please, select as many as apply.

- 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 38: LT users: 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)?**

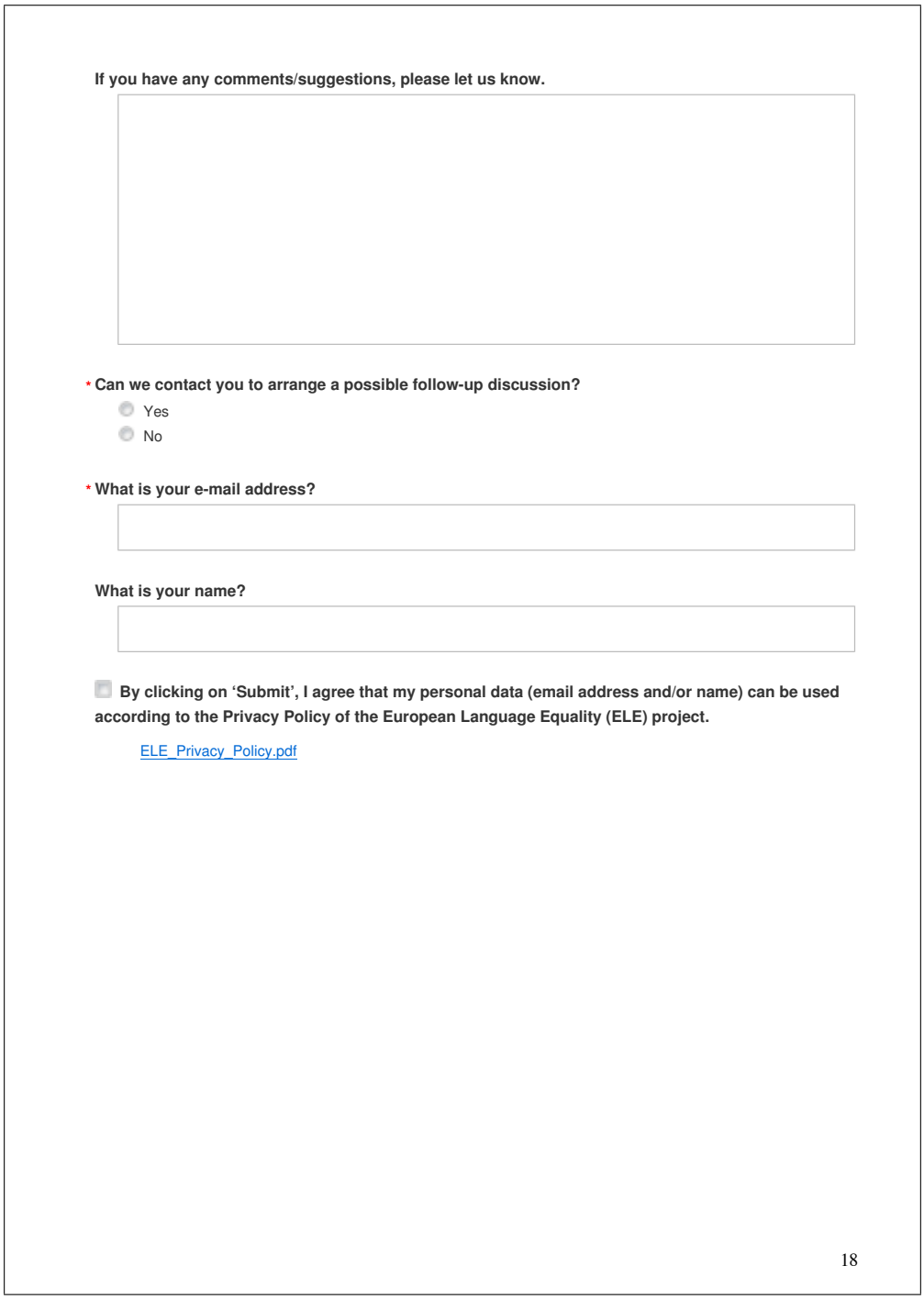
Please, select as many as apply.

- 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 39: LT users: 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 40: LT users: full survey as published (page 18/18)

D LT Users and Consumers: Additional Material

Table 15: All types of stakeholders represented by the LT users survey (Departments, faculties, institutes, organisations, companies, independent professionals, retired professionals, students.)

stakeholders-users
The Árni Magnússon Institute for Icelandic Studies
ADAPT Centre
Foras na Gaeilge
Research Centre of the Slovenian Academy of Sciences and Arts, Fran Ramovš Institute of the Slovenian Language
Staroslavenski institut (Old Church Slavonic Institute), Zagreb
Wikimedia Deutschland e.V.
self-employed, my own company Euroglossa d.o.o.
Mirara Translations
I am self-employed.
Self employed
freelance self-employed
Institute for Social Research in Zagreb
self imployed
Intellectual Property Office
HGK
Retired but owe own company emgaged in translstion services -Max fin doo Zagreb
SPES d.o.o.
Projectus grupa
Globe
self employed translator
SELF-EMPLOYED
Global Link d.o.o.
The Finnish Social Insurance Institution
Institute of Croatian Language and Linguistics
Vytautas Magnus University
Central State Office for the Development of the Digital Society
The Institute of the Lithuanian Language
University of Zagreb, Faculty of Electrical Engineering and Computing
Zagreb School of Economics and Management
LTU, Canolfan Bedwyr, Bangor University
Self employed
Self-employed
Croatian Parliament
interpreter and translator, self-employed
Croatian Association of Scientific and Technical Translators
Faculty of Science, University of Split
University
Croatian Academy of Sciences and Arts
Megabyte Ltd
University of Rijeka
University of Zagreb, Faculty of Humanities and Social Sciences

Continued on next page

Table 15 – *Continued from previous page*

LT users –stakeholders
Universidade de Santiago de Compostela
Bangor University
ART G.E.I.E.
Federal Lezghin National and Cultural Autonomy
University of The Basque Country
Babeş-Bolyai University
University of the Basque Country (UPV/EHU)
CENTRE FOR AROMANIAN LANGUAGE AND CULTURE IN BULGARIA - CALCB
ministère de l'éducation nationale
CBAC-WJEC
Council for the Maltese Language
Hitz Center (Ixa Research Group)
Departament d'Educació
Instituto da Lingua Galega (Universidade de Santiago de Compostela)
University of Vigo
Euskal Irrati Telebista (Basque Broadcasting Company)
EITB - Basque Radio Television Public Group (PSM)
Self-employed
University of Economics, Bratislava
National Research Council of Italy
University of the Basque Country
Y Coleg Cymraeg Cenedlaethol
Universitat Oberta de Catalunya
Athena Research Centre
Institute of the Lithuanian Language
Self-employed
self
IURIDICO Legal & Financial Translation sp. z o.o.
Archil Eliashvili Institute of Control Systems of Georgian Technical University
Self-employed
Self -Employed
Institute of the Lithuanian Language
Vilnius university
Open University of Catalonia (UOC)
Language Technologies Unit Bangor University
Librezale
Debagoieneko Mankomunitatea
University of Copenhagen
self-employed, translation and interpreting associations APTIC, BDÜ, VKD
The National Library of Wales
Pázmany Péter Catholic University
University of Hamburg
Food Standards Agency
not replying on behalf of an organisation
University of Osijek, FFOS
university college dublin
Hilfsgemeinschaft der Blinden und Sehschwachen Österreichs

Continued on next page

Table 15 – *Continued from previous page*

LT users –stakeholders
Hungarian Research Centre for Linguistics, Budapest University of Technology and Economics
Institute for Language and Folklore
Cymdeithas Cyfieithwyr Cymru
Local government
self-employed
Hse
Menai Science Park Ltd
Teaching council of Ireland
Maynooth university
University of Luxembourg
Atercin
Spencer Stuart
Tampere University
NHS
Self-employed
ADAPT
Eurescom GmbH
Zurich University of Applied Sciences
Wrocław University of Science and Technology
Research Center for Linguistics
University of Athens
MITA
LIBER
University of Padua
National Youth Service - Ministry of Education, Children and Youth
NPLD
Catholic University Eichstätt-Ingolstadt
ICC-Languages
National Research, Development and Innovation Office
University
BEIA
Grow Coaching Alliance
Educational & Training Concepts
Trinity College Dublin (employer); Patron of national association of deaf women in Ireland - NDWI
Library and Information Centre, Hungarian Academy of Sciences
Self-employed and Interactions LLC
Mercell
University of Bristol
National and Kapodistrian University of Athens
All Ukrainian National Culteral Moldovan Association
Òmnium Cultural
RTL
Polytechnic University of Valencia
Regione autonoma Valle d'Aosta
Washington Metropolitan University
Directorate-General for Language Policy. Government of the Balearic Islands

Continued on next page

Table 15 – *Continued from previous page*

LT users –stakeholders
Institutul de Filologie Română “A. Philippide”, Institutul de Lingvistică „Iorgu Iordan – Al. Rosetti”, Institutul de Lingvistică şi Istorie Literară „Sextil Puşcariu”, Academia Română , Romanian Academy
University of Malta
University of Malta
Universitat autonoma de barcelona
The Institute for the Languages of Finland
Språkrådet (The Language Council of Norway)
FP CGIL, Spaciada sa bregùngia, RAS
Haute Ecole pédagogique Vaud
Food industry employee. Private business owner (art and education oriented)
University of Győr and Edilic Association
EPE - Ensino Português no Estrangeiro
Centre for the Greek Language
University of Cambridge
Dublin City University
ADAPT Centre, Dublin City University
ADAPT Centre, Trinity College Dublin
ADAPT
Trinity College Dublin
European Culture and Technology Lab+, Technological University Dublin
Dublin City University
University of Strasbourg
University of Strasbourg (Faculty of Education and Life Long Learning)
Interregional public association of Meskhetian Turks “Vatan”
Carpatho-Rusyns
STOCKHOLM UNIVERSITY
University of York
University of Thessaly
University of Edinburgh
Gimara Ltd (own company) and Lab University of Applied Sciences
LIDILE/Université Rennes 2
Ministry of Culture
Self-employed
CCG - Centro de Computação Gráfica (CCG - Center for Computer Graphics)
Ministry of Education
Leibniz Institute for the Ferman Language
Leibniz-Institut für Deutsche Sprache
Centre of the Greek Language
Nuance Communication
CNRS
LAB University of Applied Sciences
German Research Center for Artificial Intelligence (DFKI)
Cornwall Council
Inuits
university
University of Eastern Finland
Macedonian Academy of Sciences and Arts

Continued on next page

Table 15 – *Continued from previous page*

LT users –stakeholders
Nara Educational Technologies
Shell
ACAPELA GROUP
University of Extremadura
Universitat Politecnica de Valencia
ATI and storyfact.
I am a university student and I am self-wmployed as well
ECSPM
CNRS, French research operating organisation
Universtity rennes 2
Institute of the Estonian Language
Smith& Nephew
Toros University, Turkey
University of Belgrade, Faculty of Mining and Geology (Language Technology group)
University of Porto
Top Communica
Bibliothèque universitaire des langues et civilisations (GIP BULAC)
Université Rennes 2
university of Lyon
Non profit ISSA Polska
Self-employed
University of Food Technologies - Plovdiv
JOANNEUM RESEARCH
Longbrook Translation
Not employed - volunteer
Institute of the Lithuanian Language
Self-employed (Nico van de Water Linguistic Services)
retired, self-employed
Telecats BV.
Retired
Accademia della Crusca
Institute of Multilingualism at the University of Fribourg i.Ü.
University of Bamberg
self-employed translator and interpreter
Self-employed
Institute of Croatian Language and Linguitics
ALTE - Association of Language Testers in Europe
Danish Language Council
Bulgarian Academy of Sciences, Institute for Bulgarian Language
Universitat Autònoma de Barcelona
Linköping university
Media Perspectives
International Medical Informatics Association (IMIA)
Basque Wikipedia
Wikimedians of Slovakia
French Wikipedia, Tacawit Wiktionary, Commons
English Wiktionary
Amical Wikimedia

Continued on next page

Table 15 – *Continued from previous page*

LT users –stakeholders
Macedonian Wikipedia
Wikimedia Foundation Search Platform Team
Aragonese Wikipedia, Universidad de Zaragoza
Wikimedia Community Ireland
CEE Spring
Wikipedia
Research center
Technical University of Denmark, Wikimedia Denmark
Wikidata Lexemes/Lexicographical data
Bulgarian Wikipedia
French Wiktionary
Wikimedia UK
Lingua Libre
volunteer at Wikimédia France
Wikimedia Hungary
Wikimedia Community User Group Malta

Countries	Counts	%
Croatia	33	13.4
Spain	28	11.4
UK	22	8.9
Germany	16	6.5
Ireland	17	6.9
France	14	5.7
Greece	8	3.3
Hungary	8	3.3
Denmark	7	2.8
Italy	7	2.8
Netherlands	7	2.8
Poland	7	2.8
Finland	6	2.4
Lithuania	6	2.4
Malta	6	2.4
Bulgaria	5	2.0
Switzerland	5	2.0
Austria	4	1.6
Romania	4	1.6
Russian Federation	4	1.6
Sweden	4	1.6
Luxembourg	3	1.2
Iceland	2	0.8
Portugal	2	0.8
Slovak Republic	2	0.8
Slovenia	2	0.8
Turkey	2	0.8
Ukraine	2	0.8
United States of America	2	0.8
Belgium	1	0.4
Czechia	1	0.4
Democratic Republic of Congo	1	0.4
Estonia	1	0.4
Georgia	1	0.4
Northern Ireland	1	0.4
Kosovo	1	0.4
Macedonia	1	0.4
North Macedonia	1	0.4
Norway	1	0.4
Serbia	1	0.4
Total	246	100%

Table 16: Answers to the question “In which country are you based?” if “other”, please specify.

Language	counts	%
English	181	73.6
French	77	31.3
German	76	30.9
Spanish	73	29.7
Italian	54	22.0
Croatian	42	17.1
Polish	31	12.6
Hungarian	30	12.2
Portuguese	30	12.2
Swedish	30	12.2
Irish	24	9.8
Romanian	24	9.8
Dutch	23	9.3
Danish	22	8.9
Greek	22	8.9
Slovenian	21	8.5
Finnish	20	8.1
Norwegian	19	7.7
Bulgarian	18	7.3
Slovak	18	7.3
Lithuanian	17	6.9
Maltese	12	4.9
Estonian	11	4.5
Latvian	10	4.1
Icelandic	9	3.7
Dutch	4	1.6
Czech	3	1.2
Russian	2	0.8
Basque	1	0.4
Catalan	1	0.4
Chinese	1	0.4
Macedonian	1	0.4
Mandarin	1	0.4
Welsh	1	0.4

Table 17: Answers to the question “Which of the official European language(s) listed below do you or your organisation work with?” if “other”, please specify.

E EU Citizen Survey

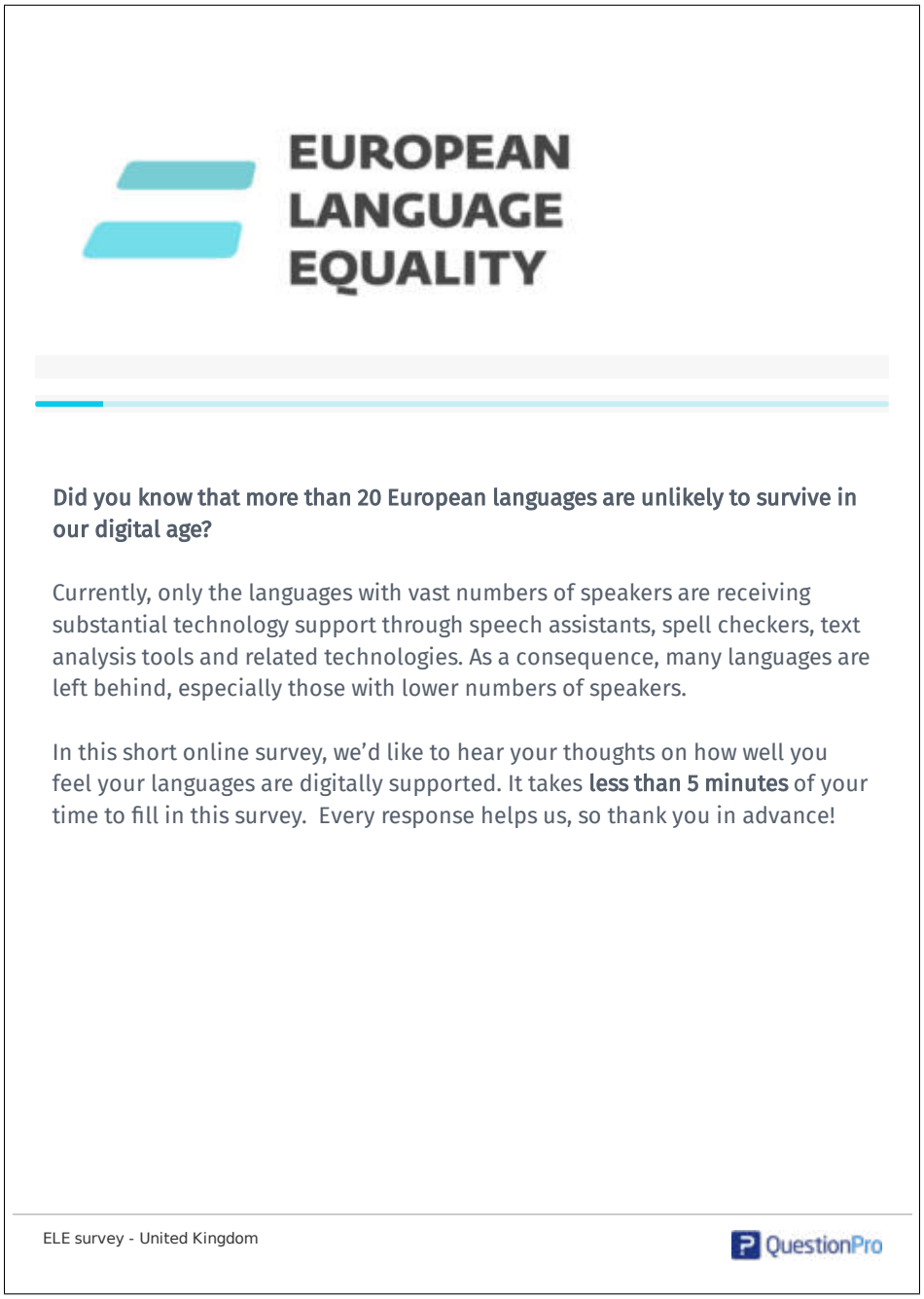
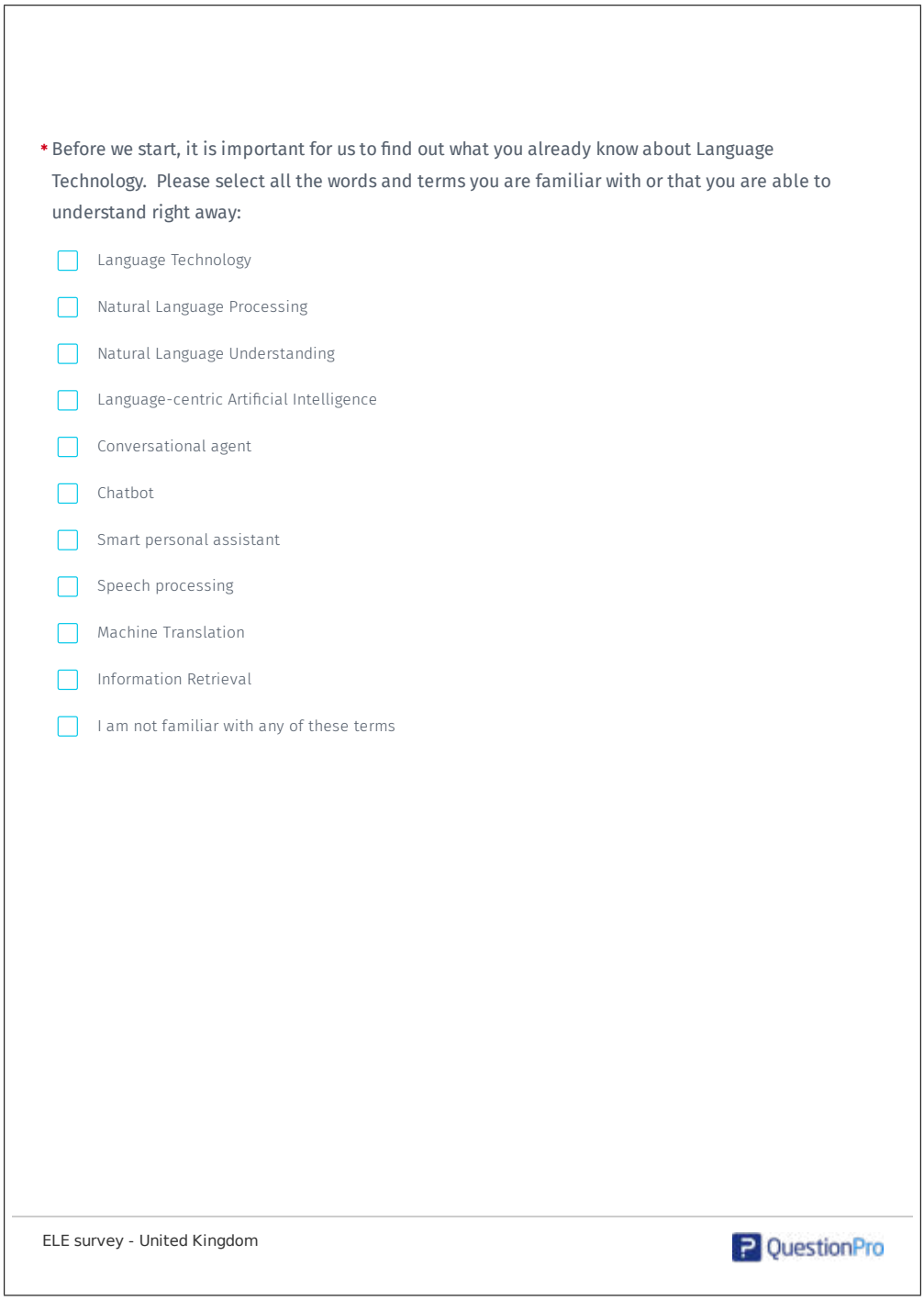


Figure 41: EU citizens: full survey as published (page 1/18)



* Before we start, it is important for us to find out what you already know about Language Technology. Please select all the words and terms you are familiar with or that you are able to understand right away:

- [ ] Language Technology
- [ ] Natural Language Processing
- [ ] Natural Language Understanding
- [ ] Language-centric Artificial Intelligence
- [ ] Conversational agent
- [ ] Chatbot
- [ ] Smart personal assistant
- [ ] Speech processing
- [ ] Machine Translation
- [ ] Information Retrieval
- [ ] I am not familiar with any of these terms

ELE survey - United Kingdom

QuestionPro

Figure 42: EU citizens: full survey as published (page 2/18)

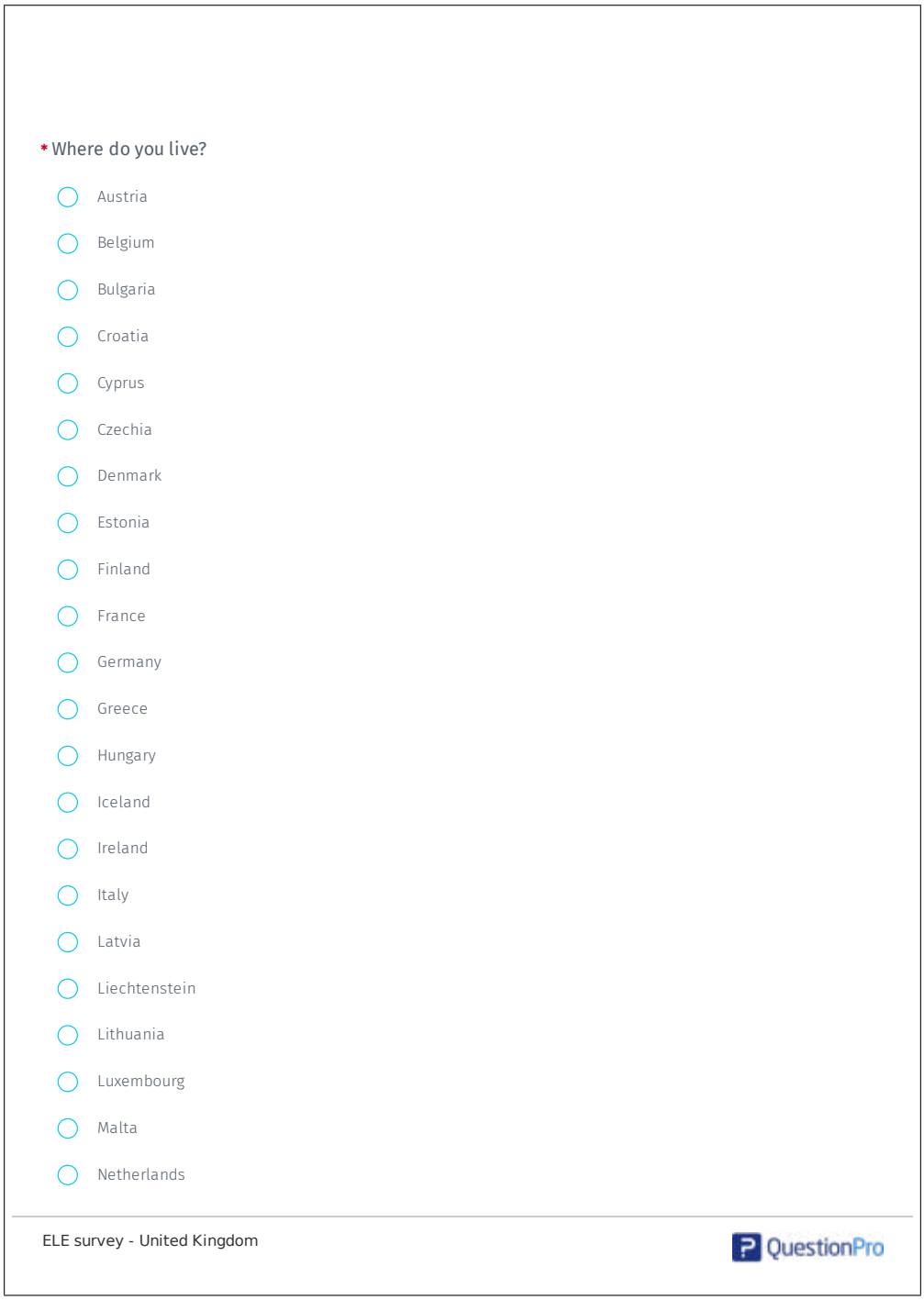


Figure 43: EU citizens: full survey as published (page 3/18)

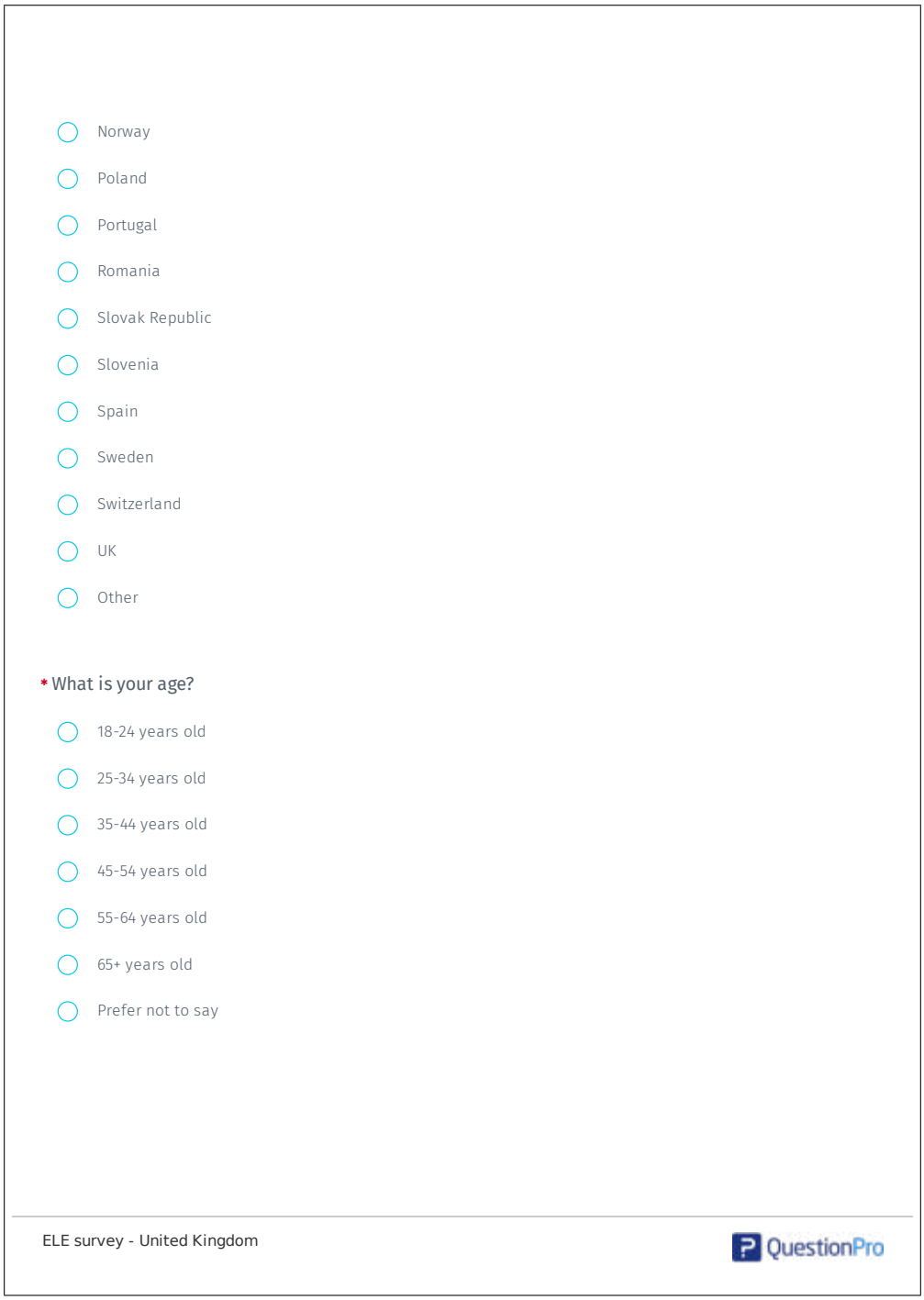
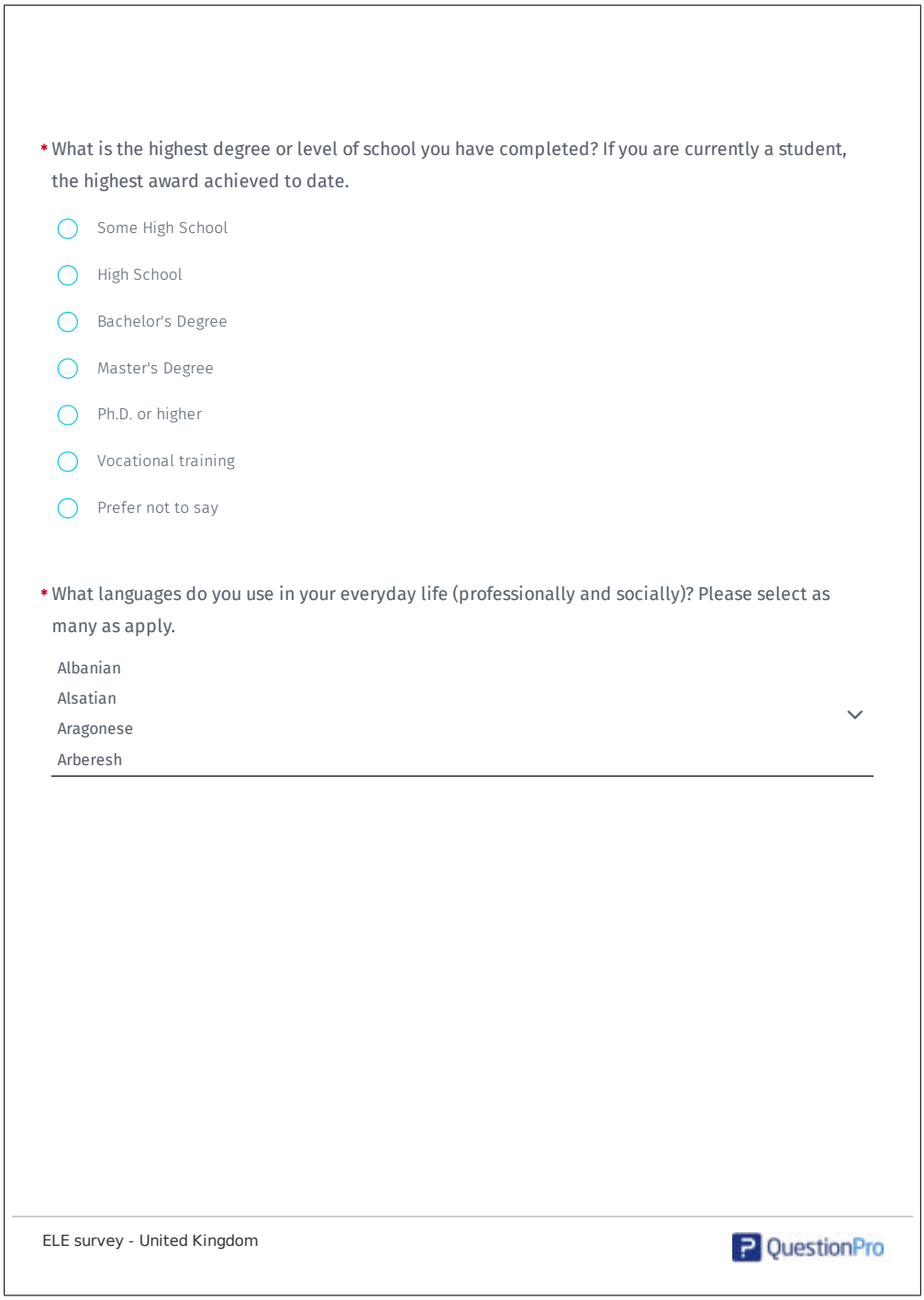


Figure 44: EU citizens: full survey as published (page 4/18)



* What is the highest degree or level of school you have completed? If you are currently a student, the highest award achieved to date.

- Some High School
- High School
- Bachelor's Degree
- Master's Degree
- Ph.D. or higher
- Vocational training
- Prefer not to say

* What languages do you use in your everyday life (professionally and socially)? Please select as many as apply.

Albanian
Alsatian
Aragonese
Arberesh

ELE survey - United Kingdom

QuestionPro

Figure 45: EU citizens: full survey as published (page 5/18)

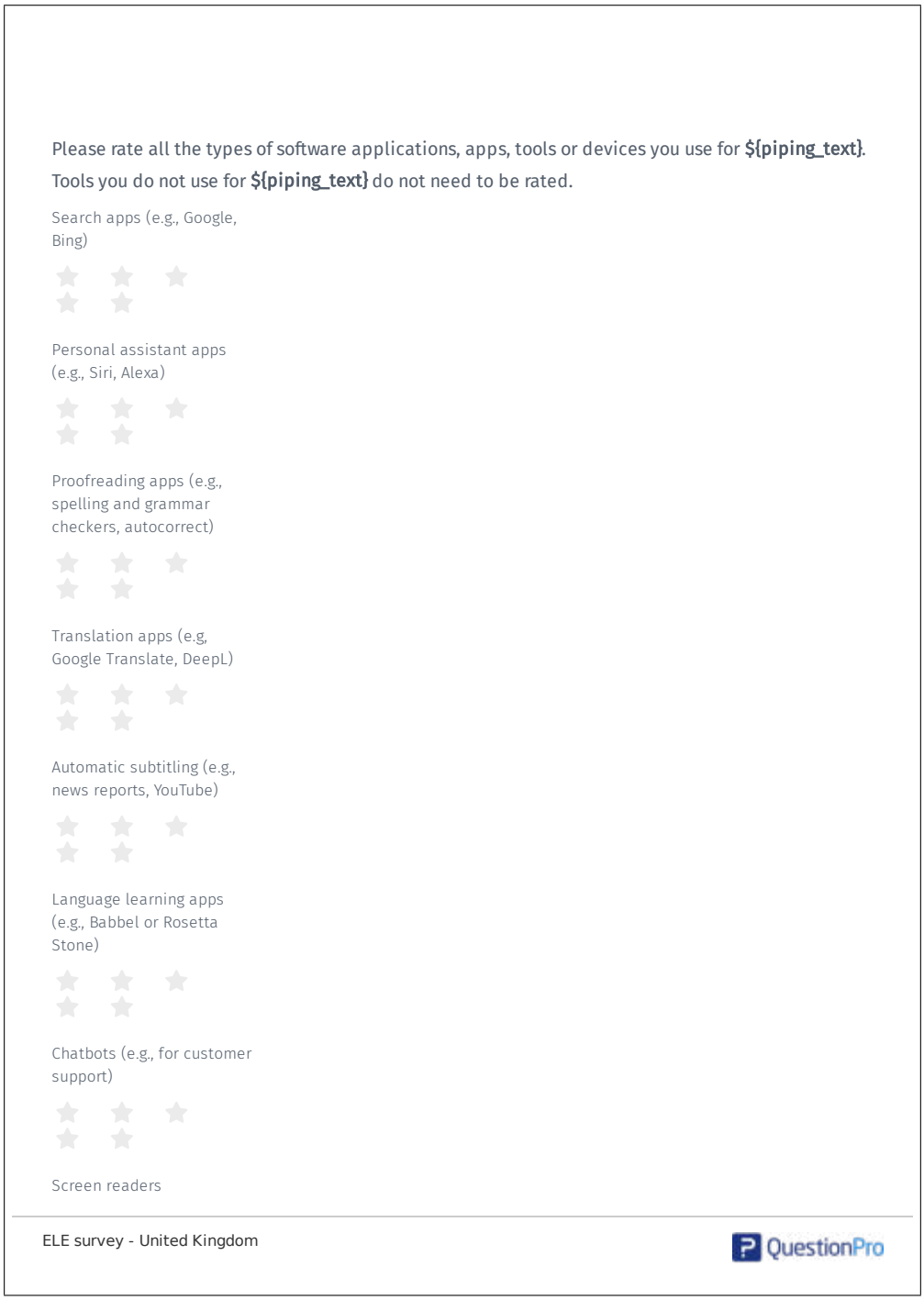


Figure 46: EU citizens: full survey as published (page 6/18)

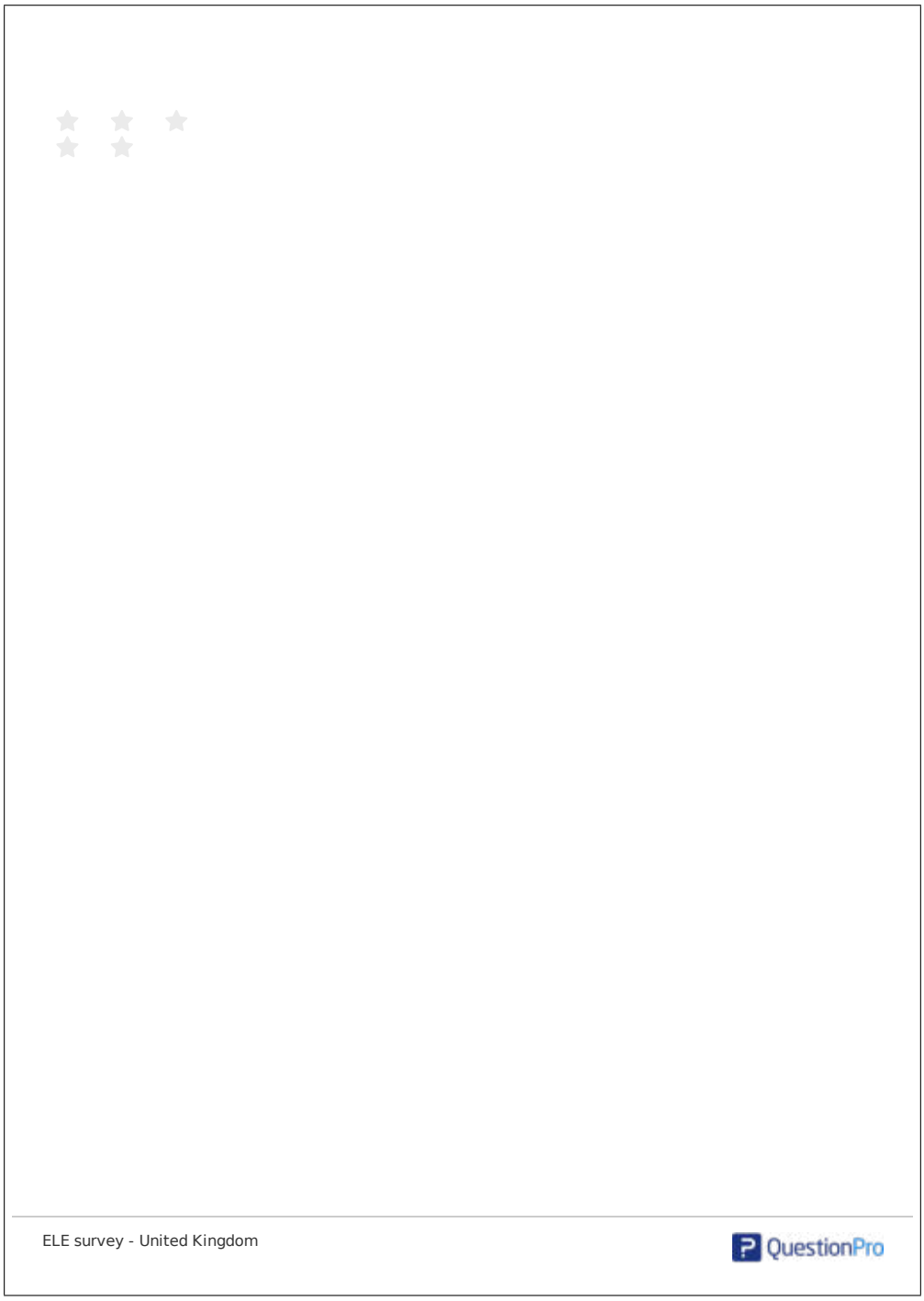


Figure 47: EU citizens: full survey as published (page 7/18)

In general, what holds you back from using some of these apps or tools in your languages?

	Lack of available apps or tools	Issues with the quality of the available apps or tools	My keyboard doesn't work for this language	Very few people understand this language	This language is usually only spoken, not written.	I don't need to use any apps or tools for this language	I don't know any apps or tools for this language	I don't know
Albanian	◯	◯	◯	◯	◯	◯	◯	◯
Alsatian	◯	◯	◯	◯	◯	◯	◯	◯
Aragonese	◯	◯	◯	◯	◯	◯	◯	◯
Arberesh	◯	◯	◯	◯	◯	◯	◯	◯
Aromanian/Arman	◯	◯	◯	◯	◯	◯	◯	◯
Asturian	◯	◯	◯	◯	◯	◯	◯	◯
Basque	◯	◯	◯	◯	◯	◯	◯	◯
Bosnian	◯	◯	◯	◯	◯	◯	◯	◯
Breton	◯	◯	◯	◯	◯	◯	◯	◯
Bulgarian	◯	◯	◯	◯	◯	◯	◯	◯

	Lack of available apps or tools	Issues with the quality of the available apps or tools	My keyboard doesn't work for this language	Very few people understand this language	This language is usually only spoken, not written.	I don't need to use any apps or tools for this language	I don't know any apps or tools for this language	I don't know
Carpathian-German	◯	◯	◯	◯	◯	◯	◯	◯
Carpato-Rusyn	◯	◯	◯	◯	◯	◯	◯	◯
Catalan, Valencian	◯	◯	◯	◯	◯	◯	◯	◯
Cimbrian	◯	◯	◯	◯	◯	◯	◯	◯
Cornish	◯	◯	◯	◯	◯	◯	◯	◯
Croatian	◯	◯	◯	◯	◯	◯	◯	◯
Czech	◯	◯	◯	◯	◯	◯	◯	◯

ELE survey - United Kingdom

QuestionPro

Figure 48: EU citizens: full survey as published (page 8/18)

Figure 49: EU citizens: full survey as published (page 9/18)

	Lack of available apps or tools	Issues with the quality of the available apps or tools	My keyboard doesn't work for this language	Very few people understand this language	This language is usually only spoken, not written.	I don't need to use any apps or tools for this language	I don't know any apps or tools for this language	I don't know
Indigenous Languages in Guyane	○	○	○	○	○	○	○	○
Irish	○	○	○	○	○	○	○	○
Italian	○	○	○	○	○	○	○	○
Jerriais	○	○	○	○	○	○	○	○
Karelian	○	○	○	○	○	○	○	○
Kashubian	○	○	○	○	○	○	○	○
Ladin	○	○	○	○	○	○	○	○
Latgalian	○	○	○	○	○	○	○	○
Latvian	○	○	○	○	○	○	○	○
Lezghin	○	○	○	○	○	○	○	○
Ligurian	○	○	○	○	○	○	○	○
Lithuanian	○	○	○	○	○	○	○	○
Lombard	○	○	○	○	○	○	○	○
Luxembourgish	○	○	○	○	○	○	○	○
Macedonian	○	○	○	○	○	○	○	○
Maltese	○	○	○	○	○	○	○	○
Manx	○	○	○	○	○	○	○	○
Meskhetian	○	○	○	○	○	○	○	○

Figure 50: EU citizens: full survey as published (page 10/18)

	Lack of available apps or tools	Issues with the quality of the available apps or tools	My keyboard doesn't work for this language	Very few people understand this language	This language is usually only spoken, not written.	I don't need to use any apps or tools for this language	I don't know any apps or tools for this language	I don't know
Mirandese	○	○	○	○	○	○	○	○
Mocheno	○	○	○	○	○	○	○	○
Moldovian	○	○	○	○	○	○	○	○
Norwegian (Bokmål, Nynorsk)	○	○	○	○	○	○	○	○
Occitan	○	○	○	○	○	○	○	○
Picard	○	○	○	○	○	○	○	○
Piedmontese	○	○	○	○	○	○	○	○
Plattdeutsch	○	○	○	○	○	○	○	○
Polish	○	○	○	○	○	○	○	○
Portuguese	○	○	○	○	○	○	○	○
Réunion Creole	○	○	○	○	○	○	○	○
Romagnol	○	○	○	○	○	○	○	○
Romani	○	○	○	○	○	○	○	○
Romanian	○	○	○	○	○	○	○	○
Saami (North, South, Inari, Skolt, Kildin, Pite, Lule)	○	○	○	○	○	○	○	○
Sardinian	○	○	○	○	○	○	○	○
Scottish Gaelic	○	○	○	○	○	○	○	○

	Lack of available apps or tools	Issues with the quality of the available apps or tools	My keyboard doesn't work for this language	Very few people understand this language	This language is usually only spoken, not written.	I don't need to use any apps or tools for this language	I don't know any apps or tools for this language	I don't know

Figure 51: EU citizens: full survey as published (page 11/18)

	Lack of available apps or tools	Issues with the quality of the available apps or tools	My keyboard doesn't work for this language	Very few people understand this language	This language is usually only spoken, not written.	I don't need to use any apps or tools for this language	I don't know any apps or tools for this language	I don't know
Serbian	○	○	○	○	○	○	○	○
Sicilian	○	○	○	○	○	○	○	○
Slovak	○	○	○	○	○	○	○	○
Slovene	○	○	○	○	○	○	○	○
Sorbian (Lower, Upper)	○	○	○	○	○	○	○	○
Southern Italian	○	○	○	○	○	○	○	○
Spanish	○	○	○	○	○	○	○	○
Swedish	○	○	○	○	○	○	○	○
Tatar	○	○	○	○	○	○	○	○
Tornedalian Finnish	○	○	○	○	○	○	○	○
Turkish	○	○	○	○	○	○	○	○
Venetian	○	○	○	○	○	○	○	○
Võro (Seto, Mulgi, Tarto)	○	○	○	○	○	○	○	○
Walser	○	○	○	○	○	○	○	○
Welsh	○	○	○	○	○	○	○	○
Yiddish	○	○	○	○	○	○	○	○
Other	○	○	○	○	○	○	○	○

Are you aware of any language apps or tools for other languages that you would also like to use for your own languages?

○ Yes

○ No

ELE survey - United Kingdom

QuestionPro

Figure 52: EU citizens: full survey as published (page 12/18)

Please select the tools that you currently do not use but would like to use in the future.

	Search apps (e.g., Google, Bing)	Personal assistant apps (e.g., Siri, Alexa)	Proofreading apps (e.g., spelling and grammar checkers, autocorrect)	Translation apps (e.g., Google Translate, DeepL)	Automatic subtitling (e.g., news reports, YouTube)	Language learning apps (e.g., Babbel or Rosetta Stone)	Chatbots (e.g., for customer support)	Screen readers
Albanian	☐	☐	☐	☐	☐	☐	☐	☐
Alsatian	☐	☐	☐	☐	☐	☐	☐	☐
Aragonese	☐	☐	☐	☐	☐	☐	☐	☐
Arberesh	☐	☐	☐	☐	☐	☐	☐	☐
Aromanian/Arman	☐	☐	☐	☐	☐	☐	☐	☐
Asturian	☐	☐	☐	☐	☐	☐	☐	☐
Basque	☐	☐	☐	☐	☐	☐	☐	☐
Bosnian	☐	☐	☐	☐	☐	☐	☐	☐
Breton	☐	☐	☐	☐	☐	☐	☐	☐
Bulgarian	☐	☐	☐	☐	☐	☐	☐	☐

	Search apps (e.g., Google, Bing)	Personal assistant apps (e.g., Siri, Alexa)	Proofreading apps (e.g., spelling and grammar checkers, autocorrect)	Translation apps (e.g., Google Translate, DeepL)	Automatic subtitling (e.g., news reports, YouTube)	Language learning apps (e.g., Babbel or Rosetta Stone)	Chatbots (e.g., for customer support)	Screen readers
Carpathian-German	☐	☐	☐	☐	☐	☐	☐	☐
Carpato-Rusyn	☐	☐	☐	☐	☐	☐	☐	☐
Catalan, Valencian	☐	☐	☐	☐	☐	☐	☐	☐
Cimbrian	☐	☐	☐	☐	☐	☐	☐	☐
Cornish	☐	☐	☐	☐	☐	☐	☐	☐
Croatian	☐	☐	☐	☐	☐	☐	☐	☐
Czech	☐	☐	☐	☐	☐	☐	☐	☐

ELE survey - United Kingdom

QuestionPro

Figure 53: EU citizens: full survey as published (page 13/18)

Figure 54: EU citizens: full survey as published (page 14/18)

Figure 55: EU citizens: full survey as published (page 15/18)

	Search apps (e.g., Google, Bing)	Personal assistant apps (e.g., Siri, Alexa)	Proofreading apps (e.g., spelling and grammar checkers, autocorrect)	Translation apps (e.g., Google Translate, DeepL)	Automatic subtitling (e.g., news reports, YouTube)	Language learning apps (e.g., Babbel or Rosetta Stone)	Chatbots (e.g., for customer support)	Screen readers
Mirandese	☐	☐	☐	☐	☐	☐	☐	☐
Mocheno	☐	☐	☐	☐	☐	☐	☐	☐
Moldovian	☐	☐	☐	☐	☐	☐	☐	☐
Norwegian (Bokmål, Nynorsk)	☐	☐	☐	☐	☐	☐	☐	☐
Occitan	☐	☐	☐	☐	☐	☐	☐	☐
Picard	☐	☐	☐	☐	☐	☐	☐	☐
Piedmontese	☐	☐	☐	☐	☐	☐	☐	☐

	Search apps (e.g., Google, Bing)	Personal assistant apps (e.g., Siri, Alexa)	Proofreading apps (e.g., spelling and grammar checkers, autocorrect)	Translation apps (e.g., Google Translate, DeepL)	Automatic subtitling (e.g., news reports, YouTube)	Language learning apps (e.g., Babbel or Rosetta Stone)	Chatbots (e.g., for customer support)	Screen readers
Plattdeutsch	☐	☐	☐	☐	☐	☐	☐	☐
Polish	☐	☐	☐	☐	☐	☐	☐	☐
Portuguese	☐	☐	☐	☐	☐	☐	☐	☐
Réunion Creole	☐	☐	☐	☐	☐	☐	☐	☐
Romagnol	☐	☐	☐	☐	☐	☐	☐	☐
Romani	☐	☐	☐	☐	☐	☐	☐	☐
Romanian	☐	☐	☐	☐	☐	☐	☐	☐
Saami (North, South, Inari, Skolt, Kildin, Pite, Lule)	☐	☐	☐	☐	☐	☐	☐	☐
Sardinian	☐	☐	☐	☐	☐	☐	☐	☐
Scottish Gaelic	☐	☐	☐	☐	☐	☐	☐	☐

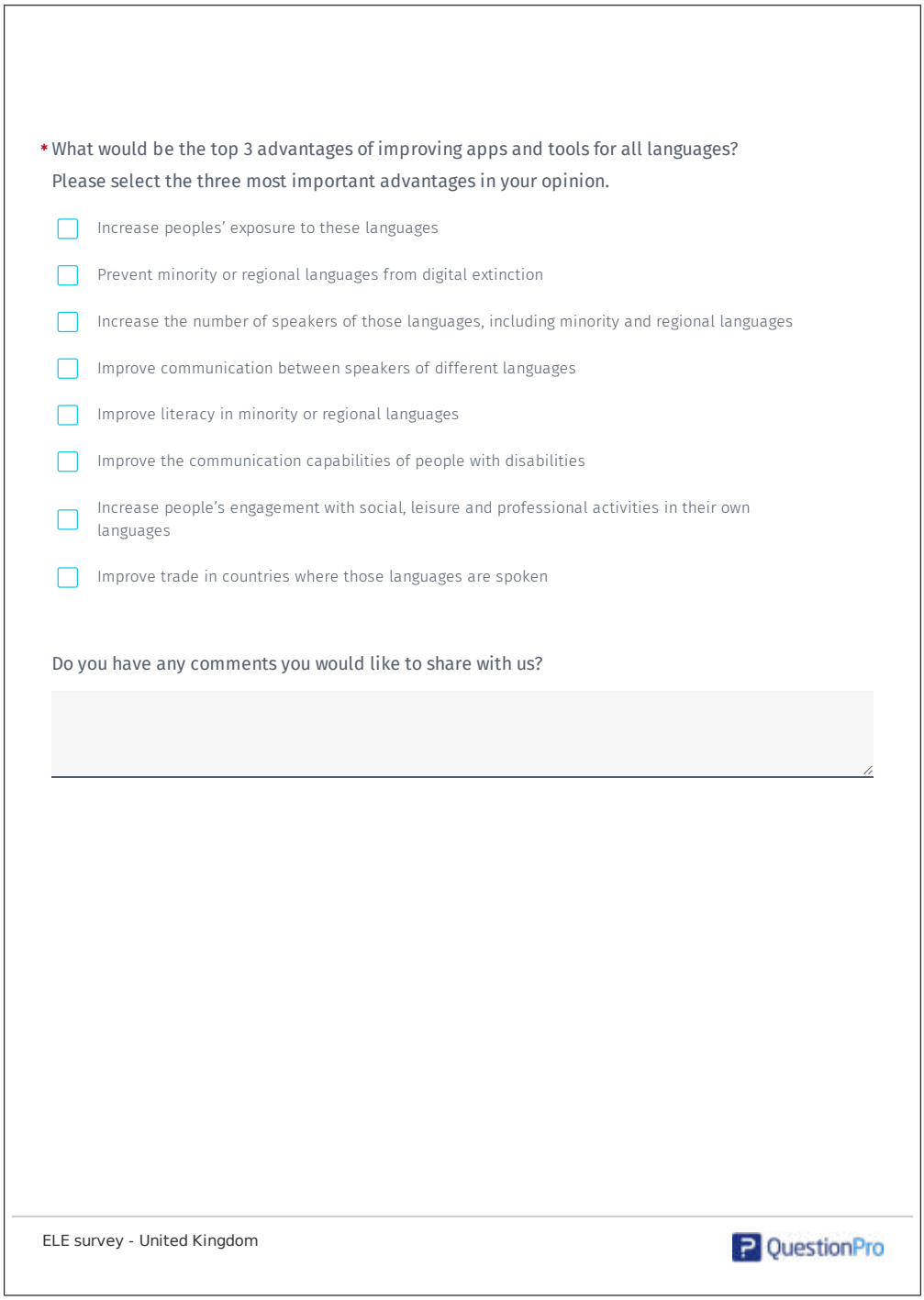
	Search apps (e.g., Google, Bing)	Personal assistant apps (e.g., Siri, Alexa)	Proofreading apps (e.g., spelling and grammar checkers, autocorrect)	Translation apps (e.g., Google Translate, DeepL)	Automatic subtitling (e.g., news reports, YouTube)	Language learning apps (e.g., Babbel or Rosetta Stone)	Chatbots (e.g., for customer support)	Screen readers

ELE survey - United Kingdom

QuestionPro

Figure 56: EU citizens: full survey as published (page 16/18)

Figure 57: EU citizens: full survey as published (page 17/18)



* 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.

- [ ] Increase peoples' exposure to these languages
- [ ] Prevent minority or regional languages from digital extinction
- [ ] Increase the number of speakers of those languages, including minority and regional languages
- [ ] Improve communication between speakers of different languages
- [ ] Improve literacy in minority or regional languages
- [ ] Improve the communication capabilities of people with disabilities
- [ ] Increase people's engagement with social, leisure and professional activities in their own languages
- [ ] Improve trade in countries where those languages are spoken

Do you have any comments you would like to share with us?

ELE survey - United Kingdom

QuestionPro

Figure 58: EU citizens: full survey as published (page 18/18)

F EU Citizen Survey: Additional Material

Countries	Tool	mean scores%
Bulgaria	Search tools	4.5
Romania	Search tools	4.4
Poland	Search tools	4.4
Ireland	Search tools	4.4
Lithuania	Search tools	4.4
Austria	Search tools	4.4
Finland	Search tools	4.4
Estonia	Search tools	4.4
Hungary	Search tools	4.3
Serbia	Search tools	4.3
Germany	Search tools	4.3
Croatia	Search tools	4.2
Uk	Search tools	4.2
France	Search tools	4.2
Italy	Search tools	4.2
Czechia	Search tools	4.2
Slovakia	Search tools	4.2
Spain	Search tools	4.2
Latvia	Search tools	4.1
Norway	Search tools	4.1
Switzerland (German)	Search tools	4.1
Slovenia	Search tools	4.1
Sweden	Search tools	4.0
Belgium (Dutch)	Search tools	4.0
Denmark	Search tools	4.0
Netherlands	Search tools	3.9
Portugal	Search tools	3.8
Greece	Search tools	3.4

Table 18: Search tools mean score performance per country

Countries	Tool	mean scores%
Finland	Translation tools	4.28
Romania	Translation tools	4.16
Bulgaria	Translation tools	4.03
Poland	Translation tools	4.0
Czechia	Translation tools	3.93
Serbia	Translation tools	3.9
Hungary	Translation tools	3.89
Estonia	Translation tools	3.86
Spain	Translation tools	3.85
France	Translation tools	3.83
Slovakia	Translation tools	3.81
Latvia	Translation tools	3.8
Italy	Translation tools	3.79
Slovenia	Translation tools	3.77
Germany	Translation tools	3.76
Belgium-Dutch	Translation tools	3.76
Ireland	Translation tools	3.72
Austria	Translation tools	3.72
Lithuania	Translation tools	3.71
Portugal	Translation tools	3.7
Croatia	Translation tools	3.68
Netherlands	Translation tools	3.65
Sweden	Translation tools	3.56
Norway	Translation tools	3.5
Denmark	Translation tools	3.49
Switzerland-German	Translation tools	3.45
Greece	Translation tools	3.35

Table 19: Translation tools mean score performance per country

Countries	Tool	mean scores%
Finland	Proofing tools	4.11
Romania	Proofing tools	3.98
Bulgaria	Proofing tools	3.89
Poland	Proofing tools	3.85
Spain	Proofing tools	3.83
France	Proofing tools	3.83
Italy	Proofing tools	3.78
Ireland	Proofing tools	3.78
Switzerland (German)	Proofing tools	3.75
Croatia	Proofing tools	3.72
Austria	Proofing tools	3.72
Germany	Proofing tools	3.71
Slovakia	Proofing tools	3.67
Czechia	Proofing tools	3.6
UK	Proofing tools	3.6
Slovenia	Proofing tools	3.54
Estonia	Proofing tools	3.53
Norway	Proofing tools	3.52
Latvia	Proofing tools	3.5
Belgium (Dutch)	Proofing tools	3.47
Sweden	Proofing tools	3.46
Portugal	Proofing tools	3.4
Denmark	Proofing tools	3.33
Greece	Proofing tools	3.26

Table 20: Proofing tools mean score performance per country

Countries	Tool	mean scores%
Hungary	Automatic subtitling	3.77
Serbia	Automatic subtitling	3.6
Lithuania	Automatic subtitling	3.5
Netherlands	Automatic subtitling	3.41

Table 21: Automatic subtitling tools mean score performance per country