EUROPEAN LANGUAGE EQUALITY

D2.6

Report from ELG

Authors	Stefanie Hegele, Katrin Marheinecke, Georg Rehm
Dissemination level	Public
Date	28-02-2022

About this document

Project	European Language Equality (ELE)
Grant agreement no.	LC-01641480 – 101018166 ELE
Coordinator	Prof. Dr. Andy Way (DCU)
Co-coordinator	Prof. Dr. Georg Rehm (DFKI)
Start date, duration	01-01-2021, 18 months
Deliverable number	D2.6
Deliverable title	Report from ELG
Type	Report
Number of pages	38
Status and version	Final
Dissemination level	Public
Date of delivery	Contractual: 28-02-2022 – Actual: 28-02-2022
Work package	WP2: European Language Equality – The Future Situation in 2030
Task	Task 2.1 The perspective of European LT developers (industry and research)
Authors	Stefanie Hegele, Katrin Marheinecke, Georg Rehm
Reviewers	Maria Giagkou, Jan Hajič
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

1. Introduction **1**
1.1. About the European Language Grid (ELG) 2

2. Methodology and instruments **2**
2.1. Online survey 2
2.2. Interviews 4

3. Analysis of responses to survey questions **4**
3.1. Respondents profiles 5
3.2. Language coverage 5
3.3. Evaluation of current situation 7
3.4. Predictions and visions for the future 9

4. Analysis of interviews **11**
4.1. Background of interview participants 12
4.2. European Language Technology R&D until 2030 12
4.3. Development and deployment of Language Technology for all European languages 13
4.4. Positioning of European Language Technology on a global level 14
4.5. Key challenges for a long-term Language Technology funding programme . . . 15
4.6. Additional suggestions with regard to European Language Equality 17

5. Conclusions **18**

6. Acknowledgements **19**

A. The LT researchers and developers full survey **22**

B. Additional tables and graphs **30**

List of Figures

1. Type of organisation . . . 5
2. LT areas in which the respondents conduct research or develop tools and services 6
3. Languages supported by the respondents' organisations in their research and development activities . . . 7
4. Languages not supported yet by the respondents, but included in their development plans for the next three years . . . 8
5. Full survey as published (page 1/9) . . . 22
6. Full survey as published (page 2/9) . . . 23
7. Full survey as published (page 3/9) . . . 24
8. Full survey as published (page 4/9) . . . 25
9. Full survey as published (page 5/9) . . . 26
10. Full survey as published (page 6/9) . . . 27
11. Full survey as published (page 7/9) . . . 28
12. Full survey as published (page 8/9) . . . 28
13. Full survey as published (page 9/9) . . . 29

List of Tables

1. Types of survey questions . . . 3
2. Number of interviews by country . . . 4
3. Mentions of the top drivers for the decision to support additional languages . 6
4. Interview participants . . . 19
5. Breakdown of answers to "Which of the following best describes the type of organisation you work for?" (mandatory closed question) . . . 30
6. Breakdown of answers to "Where is your organisation's headquarter based?" (mandatory closed question, plus "if other" as optional open-ended question) . 30
7. Sectors in which technologies, products or services are used . . . 31
8. Full list of answers to "Which LT areas do you mainly work in? . . . 31
9. 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") . . . 31
10. 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") . . . 32

List of Acronyms

ACL	Association for Computational Linguistics
AI	Artificial Intelligence
API	Application Programming Interface
CAT	Computer-Assisted Translation
CLAIRE	Confederation of Laboratories for AI Research in Europe
CLARIN	Common Language Resources and Technology Infrastructure
COLING	Conference on Computational Linguistics
COST	Funding organisation for science and technology research networks
DARPA	Defense Advanced Research Projects Agency
DH	Digital Humanities
EC	European Commission
ELE	European Language Equality *(this project)*
ELE Programme	European Language Equality Programme *(the long-term, large-scale funding programme specified by the ELE project)*
ELG	European Language Grid (EU project, 2019-2022)
ELRC	European Language Resource Coordination
EMNLP	Empirical Methods in Natural Language Processing Conference
EP	European Parliament
HPC	High-Performance Computing
ICT	Information and Communications Technology
ITS	Internationalization Tag Set
LR	Language Resources/Resources
LT	Language Technology/Technologies
ML	Machine Learning
MT	Machine Translation
NCC	National Competence Centre
NLP	Natural Language Processing
NLU	Natural Language Understanding
POS	Part-of-speech
SME	Small and medium-sized enterprise
SRA	Strategic Research Agenda
SRIA	Strategic Research and Innovation Agenda
SSH	Social Sciences and the Humanities
STOA	Science and Technology Options Assessment
STEM	Science, Technology, Engineering and Mathematics
TAPPICC	Translation API Cases and Classes

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 all voices are heard and taken into account for the ELE SRIA and roadmap. An online-survey, targeting in particular LT researchers and developers, was shared in June-October 2021. A total of 333 responses was collected. 54 members of the European Language Grid (ELG) community filled in the survey. This report documents the findings from this survey as well as the input shared in a round of follow-up expert interviews with other representatives of the ELG community. The survey was divided into four main parts and entailed 45 questions in total. The survey investigated 1) demographic information, 2) language coverage, 3) evaluation of the current situation and 4) predictions and visions for the future.

For the expert interviews a diverse set of 20 LT representatives from mainly industry, but also research and the European institutions was selected. Every interview lasted approximately 20-30 minutes. The interview questions dived into the interviewees' backgrounds, the domains they work in and the languages they cover. Further, their thoughts and ideas on European LT research and development until 2030 were evaluated. They elaborated 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.

1. Introduction

This document reports on the results and findings of a consultation with representatives of the Language Technologies (LT) community, i. e., industry as well as research and academia, conducted by the European Language Equality (ELE) project. The results documented in this report will serve as input for a strategic research and innovation agenda and roadmap to tackle the striking imbalance between Europe's languages in terms of the support they receive through language technologies by 2030.

The ELE project collected the views of European researchers and developers to consolidate their perspectives regarding the strengths and weaknesses of the field and also regarding the measures that need to be employed, so that all European languages are equally supported through technology by 2030. This diverse group of stakeholders comprises:

- academic and industrial researchers in the field of LT/NLP – beyond pure research, they develop algorithms, pre-commercial LT prototypes, applications and systems;
- innovators and entrepreneurs who commercialise LT to address the needs of digital content analysis and generation, pertinent content transformation and dissemination, as well as enhanced human-machine interaction.

Due to the multi- and interdisciplinary nature of Language Technology, which stands at the intersection of Linguistics and 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 more, the ELE stakeholders group of LT developers also includes neighbouring disciplines, especially AI and Digital Humanities as well as Social Sciences and Humanities (DH/SSH). To reach out to this diverse and extensive group of stakeholders, the partners of the ELE consortium mobilised various European networks, associations, initiatives and projects, covering research and industry.

Although the methodology and instruments utilised have been common to all stakeholders, the present report covers and analyses the subset of responses and input from members of the European Language Grid (ELG) community.

1.1. About the European Language Grid (ELG)

The European Language Grid (ELG)[1] cloud platform is targeted to evolve into the primary platform for Language Technology in Europe. Its aim is to provide one umbrella platform for all Language Technologies developed by the European LT community, including research and industry, addressing a gap that has been repeatedly raised by the by the European Parliament (STOA, 2017; European Parliament, 2018) and by the European LT community in a number of strategy papers throughout the years (Rehm and Uszkoreit, 2013; Rehm et al., 2016b; Rehm, 2017; Rehm and Hegele, 2018; Rehm et al., 2020b,a).

The ELG is meant to be a virtual home and marketplace for all products, services and organisations active in the LT space in Europe (Rehm et al., 2020a). It enables the European LT community to deposit and upload their technologies and data sets and to deploy them through the grid. The platform can be used by all stakeholders to showcase, share and distribute their products, services, tools and resources. At the point of writing, the ELG is still funded by the EU (2019-2022); it will establish a legal entity in the first half of 2022, so that the platform can continue to provide access to the commercial and non commercial tools and services as well as language resources it hosts.

In a wider context, the ELG is also meant to support digital language equality in Europe (STOA, 2017; European Parliament, 2018), i. e., to create a status quo in which all languages are supported through technologies equally well. The current imbalance is characterised by a stark predominance of LRTs for English, while almost all other languages are only marginally supported and, thus, in danger of digital language extinction (Rehm and Uszkoreit, 2012; Kornai, 2013; Rehm et al., 2014, 2016a; Berzins et al., 2019).

2. Methodology and instruments

The views of ELG members were elicited by utilising two main instruments: an online survey (Section 2.1) and a follow-up round of interviews (Section 2.2).

2.1. Online survey

The survey addressed to LT researchers and developers sought to elicit the respondents' views in a structured way that facilitates the analysis, consolidation and integration of the collected feedback into the ELE SRIA and roadmap.

It encompassed 45 questions in total, some of the questions depend upon previous answers. As a result, 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).

The survey was structured in four main parts:

- **Part A. Respondents' profiling**: The first part included 13 questions for the demographic profiling of respondents with emphasis on characteristics relevant to the task at hand, i. e.,
  - Country

[1] https://www.european-language-grid.eu

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

Table 1: Types of survey questions

  - 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-/middle-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 the 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 achieve 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 for an interview and, given an affirmative answer, his/her contact details.

The survey was designed, set up and published on the EU Survey platform.[2] The full survey, as published online, is presented in Appendix A (p. 22 ff.).

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

The survey was distributed through emails to all members of the European Language Grid (ELG). It was also advertised through the ELG, ELE and ELT websites,[3] as well as through the ELT social media accounts on Twitter and LinkedIn.

The survey was opened on 17 June 2021 and closed on 18 October 2021. In total, 333 responses have been collected, out of which 54 respondents indicated that they are members of the European Language Grid (ELG). This subset of responses, representing the views of participants in the European Language Grid (ELG) initiative, is analysed in this report.

2.2. Interviews

For the interviews a diverse shortlist of LT representatives from mainly industry, but also research and the European Institutions was compiled. Participants were invited to schedule an online interview or submit their interview in written form. For both versions of the interview, guidance and instructions were offered. In total, 20 people were interviewed, 14 of which have an industry background, four mainly work in research and two work for the EU and represent the administrative level. Finding the right balance between industry, research and administration was a decisive factor when choosing the final candidates. Moreover, we aimed for a good geographical coverage to ensure wide language coverage as well as diversity in opinion and experience. Table 2 shows the countries of the organisations the interview candidates work for. With regard to the specific language related questions, it allowed us to get perspectives on the bigger European languages such as English, Spanish and German but also some less supported languages such as Irish or Icelandic.

Country	Number of interviewees
Austria	2
Belgium	2
Denmark	1
Finland	2
Germany	6
Greece	1
Iceland	1
Ireland	1
Spain	1
Sweden	1
Switzerland	1
USA	1

Table 2: Number of interviews by country

Every interview lasted approximately 20-30 minutes. Each interviewee was asked the exact same set of questions with the options of adding questions or important remarks wherever they saw fit.

3. Analysis of responses to survey questions

This section analyses 54 survey responses that we received from participants associated with the ELG initiative. Section 3.1 outlines their profiles. Language coverage is reviewed in Section 3.2. Section 3.3 summarises the main opinions on the current situation. Predictions and visions for the future are discussed in Section 3.4.

[3] https://www.european-language-grid.eu, https://www.european-language-equality.eu, https://www.european-language-technology.eu

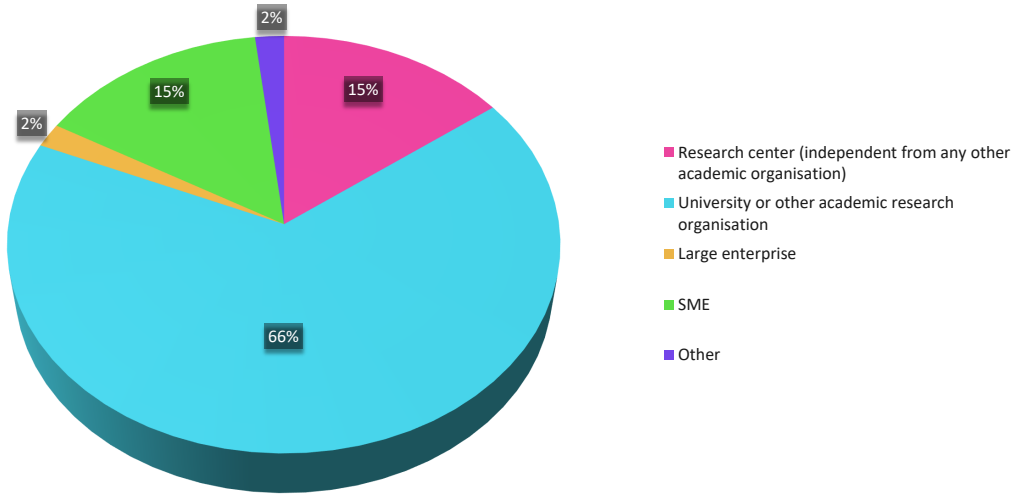


Figure 1: Type of organisation

3.1. Respondents profiles

One major goal of this survey was to bring the European LT community together and hence reach a wide and demographically distributed audience. The respondents represent 42 different organisations, out of which 67% are research or academic institutions, the remaining ones are industry practitioners or independent research centres (Figure 1). The headquarters of these organisations are located in 32 different countries, with most responses from 1) Germany, 2) Finland, 3) France, and 4) Romania. Detailed statistics of the breakdown of organisation types and countries are provided in Appendix B (Tables 5 and 6).

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) Language resources (data production, data aggregation), and 3) Translation technologies (Machine Translation, translation memories management, CAT tools) (Figure 2; see also Appendix B, Table 8).

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) Arts, culture and other services and 3) Health. See also Appendix B, Table 7 for a list of all sectors.

3.2. Language coverage

The 54 respondents listed a total of 32 different languages they actively include in their research and development work and for which they offer languages services, software, resources, models etc. The 5 most frequently mentioned languages are English, German, French, Spanish and Italian. Figure 3 displays all 32 languages and the number of times they were mentioned.

In addition, it was mentioned 13 times that other languages are also part of the portfolio. These other languages include, among others, languages spoken in the Middle East and Asia with Arabic being the most frequently mentioned one. Sign language was also mentioned.

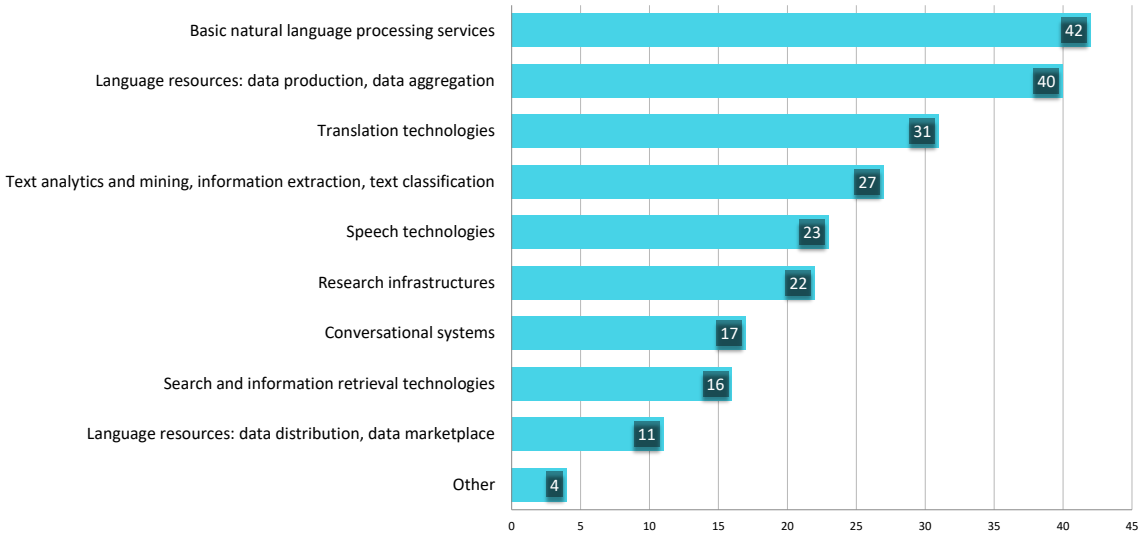


Figure 2: LT areas in which the respondents conduct research or develop tools and services

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. Respondents mentioned another 43 languages they would like to add to expand their portfolio of languages (Figure 4).

When considering the development and expansion plans with respect to language coverage the top three drivers for the decision to support additional languages are very different for research centres and universities as opposed to companies. Table 3 shows the detailed break down of answers, showing a clear tendency for universities and research centres to be driven by research and scientific interest. Available funding and investment also plays a pivotal role. The availability of language resources and human experts for these languages factor in, too. The latter is also true for companies. While there is not sufficient data to generalize the driving factors for companies, market interest and demand by users or consumers are what seems to influence decision-making. For SMEs, more than big organisations, funding and investment opportunities are also to be considered.

Drivers	Research center	University	Large enterprise	SME
Market interest/demand	4	9	1	8
Research/scientific interest	6	31	0	1
Available funding/investment	3	23	1	5
Availability of human experts	0	12	0	1
Availability of language resources	5	16	1	4
Availability of technologies/tools	2	4	0	3
Other	0	12	0	1

Table 3: Mentions of the top drivers for the decision to support additional languages

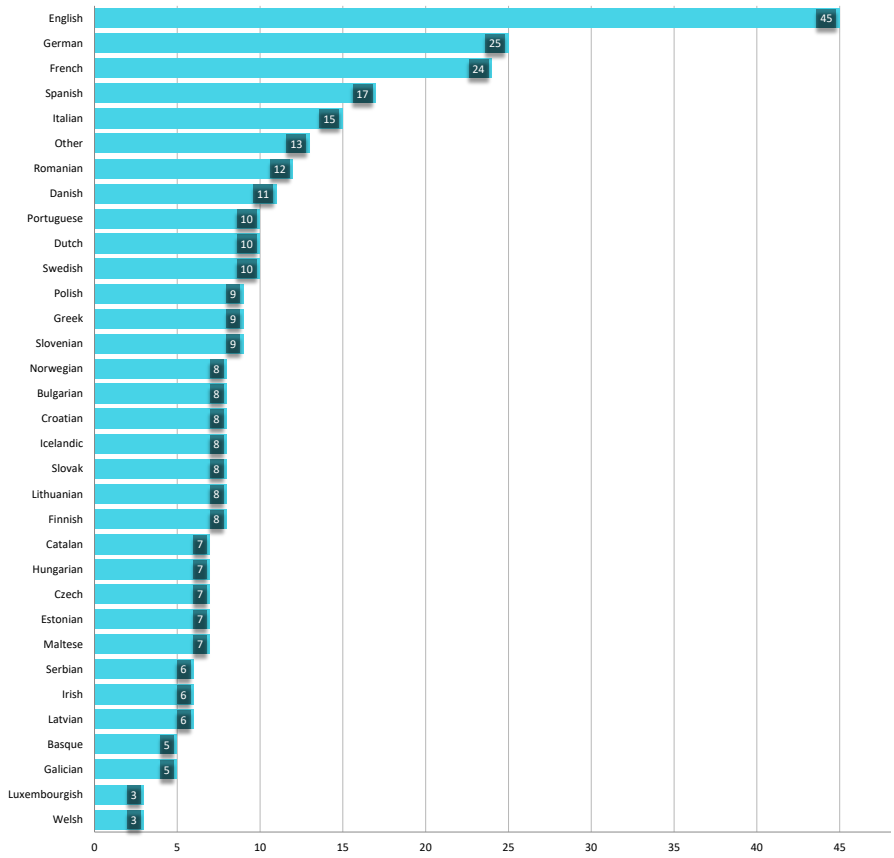


Figure 3: Languages supported by the respondents' organisations in their research and development activities

3.3. Evaluation of current situation

To evaluate the current situation participants were asked about their opinion on certain challenges and obstacles. They had the option to either strongly agree, agree, disagree, strongly disagree or express no opinion at all. A detailed numerical list of all answers can be found in Appendix B (Table 9). Respondents were also given the opportunity to elaborate on the obstacles and challenges indicated in the previous question and/or add any other obstacle/challenge not previously listed as part of a free text question.

Overall, we received almost uniform agreement that basic research is still necessarily needed. There was equally strong agreement on the inadequate recognition of the importance of multilingualism. This is reflected in the previously outlined fact that adequate language coverage only exists for a handful of European languages that have big speaker communities (see Section 3.2). The interviews we conducted (see Section 4) also reflect this view. In a large-scale survey on LT conducted in 2017 with more than 600 participants one of the biggest challenges identified was that European Language Technology is facing digital extinction for languages with smaller numbers of speakers. With the dominance of the English language, researchers are often given little incentive to focus on smaller or minority languages. Many of the respondents of the 2017 survey stressed the importance of keep-

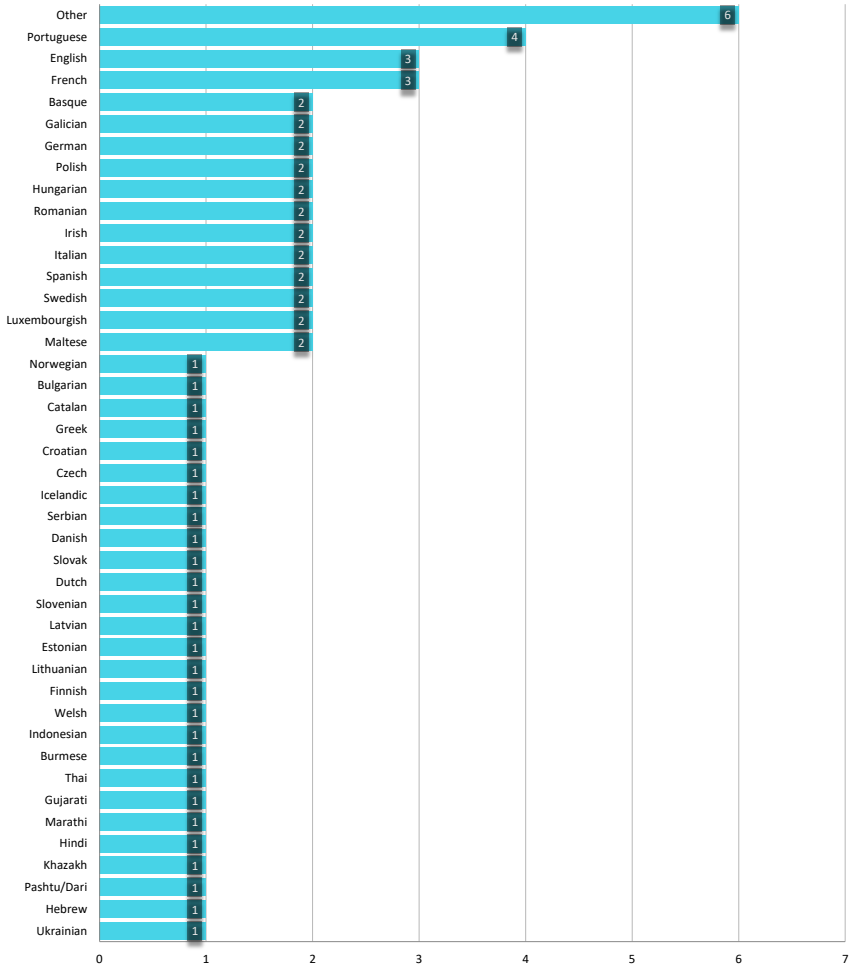


Figure 4: Languages not supported yet by the respondents, but included in their development plans for the next three years

ing multilingualism in Europe alive (Rehm and Hegele, 2018). Looking at the situation in 2021/2022 from an economic perspective, more than 90% lament the fact that there are insufficient markets to justify investments in LTs for smaller languages. There is also a lot of rising demand for sign languages, especially as accessibility rules and laws become stricter. However, currently there is little funding to develop technologies for sign languages.

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, but also more elaborate research conditions. However, the claim that brain drain is a very present problem, was not shared by the respondents of this recent survey. More than 50% disagreed with the statement which leads us to conclude that people just starting their career do consider Europe an option as a place to work. Nevertheless, almost all participants agree that there is competition with big non-European companies and market disruption caused by global players. An interesting point that was mentioned is the risk of talent/brain drain, if we equate Language Technology with Machine Learning (ML). Simplified, LT is just one of many hard challenges for ML. Conversely, ML is just one of many tools/methods in LT. But there is a risk that, if we stop educating LT and go for ML first, we'll

grind to a halt when the generic ML methods used now cannot take the state of the art in LT any further.

The fragmentation of the LT industry which has been the subject of various reports and research agendas of the last decade (Rehm and Hegele, 2018; Rehm, 2015, 2016, 2017), is still a non-solved obstacle. More than three quarters of the survey participants strongly agree or agree that it remains a challenge to this day. Individual answers emphasise that we need to establish conditions that create strength out of many small pieces and provide an 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. Funding application procedures for small companies need to be facilitated. A lack of time and money makes applying for most types of funding impossible. The insufficiency of public procurement is also considered an obstacle by a considerable number of respondents.

The answers outlined above are in line with some of the main findings from the 2017 survey (Rehm and Hegele, 2018), i. e., that 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 to compete with innovative approaches, systems and tools built overseas.

It was also encouraged that industry-oriented initiatives such as ELG need to be strengthened. The EU needs to support the implementation of Europe's own language processing infrastructure. Easy access to massive (quality) data and HPC, currently available almost exclusively to big tech companies, is crucial. Especially with the emergence of large-scale pre-trained models, access to HPC by small and medium-sized companies is critical in order for them to stay up-to-date with latest developments.

Also discussed by a few of the respondents are the challenges that smaller companies face. There aren't many easily accessible EU funding instruments that consider small companies with innovative research ideas. Receiving funding is a demanding and slow process that puts bigger companies at an advantage. The same is true for low-resourced communities. Often there is little awareness by the wider research community of the added value of providing solid solutions for low-resourced languages. The cost of developing LT for a specific language is inversely proportional to the number of its speakers, for languages with larger numbers of speakers the LRs can be collected in an easier manner since, e. g., 80 million speakers produce much more online text in a day, than two million speakers. Industry often finds a commercial interest in pre-competitive investments for "larger" languages, while this would rarely be the case for "smaller" ones. Given this situation, the role of additional investors for the development of LT for less-resourced languages should be played by public funding agencies, both at national and EU level.

3.4. 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. Similar to Section 3.3 on the evaluation of the current situation, survey participants had the option to rate a number of policies and instrument as either very effective, effective, slightly effective or not effective at all. A detailed numerical list of all answers can be found in Appendix B (Table 10). In addition, respondents were given the opportunity to elaborate on other policies/instruments not listed as part of the question, but which could be effective in speeding up the development and deployment of LT in Europe equally for all languages. The responses were provided as free text.

A large-scale, long-term funding programme until 2030

To dive a bit deeper into the opportunities that a large-scale, long-term funding programme dedicated to European Language Technology would offer, we asked participants what they consider as key challenges with regard to basic and applied research and innovation.

As for basic and applied research the most frequently mentioned areas and tasks where research is needed include language data collection (text, dialog, other forms of interactions), speech analysis, AI, human-computer interaction, machine learning, robotics, natural language understanding and processing tasks such as machine reading, text analysis, machine translation, chatbots, virtual assistants, summarisation, etc. Innovation includes concentrating on multidisciplinary approaches, developing actual language understanding with alternative methods. It was argued that statistical methods have reached their limits, as they only work for some very narrow tasks and specialised purposes and that, in order to achieve deep language understanding, we need to finance and investigate fields, such as cognitive AI, symbolic AI, pattern-based AI further. On a technological level that includes investing in the development of new (scientific/technological) methodologies for transfer/adaptation of resources/technologies to other domains and languages.

The current fragmentation of the LT industry poses a challenge to basic and applied research (see Section 3.3). It is necessary to start building common data and product spaces (such as ELG and ELRC), LT business network spaces, foster small and medium LT businesses through public procurement, strengthen SME-research ties, create conditions for growth and scaling of SMEs beyond their often small national or language foci. Not only is it a challenge to bring together all European countries under the same umbrella programme, but agreeing upon a number of strategic objectives along a small number of axes and dimensions and making sure that the R&D activities are synchronised is difficult, too. Providing sufficient compute resources is challenging as well as the provision of digital infrastructure to support the umbrella programme. The ELG is a suitable candidate that such a long-term funding umbrella programme could be built upon. The infrastructure has the potential to become the de facto standard that the LT community has been discussing for decades.

Challenges with regard to innovation and the LT industry include filling the gaps in LT support and language resource availability for smaller European languages. The scaling of language models turned out to be critical to success and only well-resourced organisations are able to produce massive models of competitive size. Research on building models is done almost exclusively by the big tech companies (Google, Facebook, Microsoft, OpenAI, Amazon, Apple, Baidu etc.). We need a strong public research initiative in the EU. More support for small and medium EU languages, at least a number of them, is necessary as well as massive multilingual European models/data. However, for very large models, the compute, data and memory requirements are so extreme that they are difficult to obtain.

In this context, retention of researchers in the face of much higher income possibilities with U.S. and Chinese companies/labs is crucial. Since everything starts with education, we need to empower linguists to better understand computational processes and actually lead the development of language technologies. Study programs in Computational Linguistics, also focusing on the emulation of cognitive processes in the human brain, would pave the way into the right direction.

Policies and instruments

To regain European global excellence and leadership in LT, massive investments in industry and research collaboration for the creation and deployment of innovative LT solutions are needed. A long-term project of 10 or more years can potentially lead to breakthrough research and subsequently to the desired leap from simple language processing to deep language understanding.

At the administrative level, a commitment on the side of the EU and national administrations to use European LT is considered crucial. As a starting point, this could be materialised by making all public services and websites in Europe multilingual, in at least the 24 official EU languages. Overall, there is a strong tendency to require to strengthen multilingualism with developing LTs for smaller and less supported languages. Equivalently, it is seen as effective to raise awareness of the benefits for companies, public bodies, and citizens about the availability of online services, contents and products in multiple languages. Development and raising awareness would go hand in hand. It could be also beneficial to impose content accessibility regulations, e. g., for multimedia subtitling, readability, dubbing, availability of content in multiple languages in some contexts.

In terms of organisational structures, the creation of national bodies or associations would be rendered useful that can amalgamate the voices of local players at national level and increase industry visibility at national and European level. Limited visibility is identified as a key challenges the LT community faces, which in turn affects the LT market size. National associations, representing the critical mass of the LT industry and supported by current initiatives, could act as further leverage for awareness raising towards policy makers.

The initiation of investment instruments and accelerator programs targeting LT start-ups was appreciated, but only rated as moderately effective by more than 20% of respondents. As far as funding is concerned, an adequate venture capital landscape would help to fund start-ups. In general, substantial funding approaches for language processing are needed.

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 90% 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. Multilingualism and language diversity should also be a focus in schools. Children should get familiar with multilingual technologies from an early age. In general, language and cultural diversity (dialects, minority languages) should be supported within a broader context.

The current academic model requires a number of publications at top tier conferences such as ACL/EMNLP etc. This discourages researchers from spending time on much needed resource creation because the return on investment is not readily apparent. More venues like LREC should be valued/promoted to encourage more work in this area. In general, more awareness in big LT research communities needs to be raised on the issue of digital equality, e. g., at ACL, EMNLP, COLING, etc. to make results of such research also publishable.

Another valid suggestion was to modernise EU copyright regulations, especially the Directive on Copyright in the Digital Single Market (2019), to enable the use of copyrighted language data for R&D, without infringing the rights of authors, similar to the fair-use principle in the United States.

4. Analysis of interviews

This section summarises the 20 interviews we led with experts from industry, research and the European Union. Section 4.1 gives a brief overview of the interviewees' backgrounds, the domains they work in and the languages they cover. We then summarise their input on European Language Technology research and development until 2030, how to do justice to all European languages, ways how to position European LT on a global level and the key challenges for a long-term funded LT programme.

4.1. Background of interview participants

When selecting participants for the interviews we were interested in bringing in individuals from a wide variety of organisations, representing the perspective of both industry and research but also the more political level. The majority of the participants have expertise in various LT areas, such as Speech Processing, Natural Language Processing, Machine Translation, Dialogue Systems etc. Companies represented in these interviews usually have an even more fine-grained focus such as voice biometrics or sign language applications. While most of the research representatives usually list all disciplines and applications in their curriculum, they often excel in one area such as digital humanities or concentrate on a very specific under-resourced language. Besides, we also had the chance to talk to a project lead working on a national LT plan and strategy that was set out to help a less resourced language to get more visibility. On the level of the European Union we interviewed an expert user and also trainer with vast experience in the conference interpreting and communication space.

The sectors and domains also cover a wide spectrum. We interviewed representatives from the media, news and broadcasting sector including institutions such as Deutsche Welle, the Cultural Broadcasting Archive and the German Federal Association for AI (KI Bundesverband e. V.). On the industry side we have companies such as Retresco and Slator which have been active in the Language Technology area for many years and count as leading companies in their respective fields. Lingsoft and Vicomtech are good examples for local/regional companies that develop technologies for local/regional languages. With respect to universities and research centres we were particularly interested in those located in countries where the official language is a lesser spoken one, such as the University of Helsinki and the Insight Centre in Ireland. This resulted in a broad coverage of sectors and domains including the manufacturing and process industry, the healthcare and medicine sector, digital security, ICT, media, industry, public administration, bio sciences, computer games, marketing, risk management, banking, finance, insurance, text analysis and media monitoring and many subcategories of the above mentioned ones.

All organisations support more than one language. English is usually one of the standard languages often combined with other big European languages. In addition, many work with languages spoken in the country or region of residence. For instance Spanish, Basque, Catalan, Galician are mentioned together. The same applies to Scandinavian or Eastern European languages. Deutsche Welle as a public broadcaster, producing in over 30 languages, covers more than 100 languages. Some organisations such as the Insight Centre in Galway have made under-resourced languages such as Irish and various Indian languages their focus.

4.2. European Language Technology R&D until 2030

The interview participants were asked which 3-5 main topics European Language Technology research and development (incl. NLP, AI etc.) should concentrate upon until 2030.

One of the most frequently mentioned topics was the availability of large and efficient multimodal language models. Publicly available language models (for fine-tuning and downstream tasks) and resources (datasets for model training and testing) are needed. Large open-source language models that work for all EU languages would be ideal. Currently, these large language models still need too much data or cost too much (in terms of compute time). Model optimisation should focus on neural model compression for low-latency decoding, efficient CPU-bound neural models, and quality embedded neural models.

While text technologies have long been the focus of NLP, speech technology and speech resources need more attention. Once speech processing becomes more prevalent, users will expect human-like understanding of available services. This human capability to accumulate accurate context-specific knowledge through conversation needs to be a priority when developing language-centric AI for speech. Progress in the area of speech applications has

been made, including the research areas of spontaneous and noise-robust speech recognition, emotional speech synthesis, speech translation. Also important is the development of good, natural synthetic voices, allowing users to obtain content in their own spoken language. Automatic speech recognition, machine translation and text-to-speech should focus on a wide variety of languages, including variants and dialects, to keep those alive. It was also implied to create a standardised audiovisual data space for SMEs and EU research, to create acoustic models for most EU languages and focus on developing speech recognition applications such as automated summarisation of speech.

The interviewees also emphasise that it is crucial to focus on language equality and the needed provisioning of basic technologies and services for languages outside of the often preferred languages such as English, German, Spanish, and French, ideally through open models and frameworks. Particular attention should be paid to the many low-resource languages. Key research questions are, among others, techniques and technologies for how to leverage multi-modal and multilingual resources to support the development of applications for languages and language varieties with scarce resources. This would also benefit specialised application domains that have scarce resources even in a well-resourced language.

Further, question answering and dialogue systems were brought up as areas with immense potential, as there exists a massive industry demand for these kinds of applications. However, they are still not well addressed and there is a scarcity of available data, including multi-domain dialogue datasets with anaphora, ellipsis, dialogue acts among other types of linguistic annotation.

Machine translation is an area still far from being solved. Direct and near-real time speech-to-speech MT and adaptive MT, where the system learns from linguists' input, would be desirable as well as low resource MT.

For all work in NLP, the provisioning of high-quality, human-generated open and free datasets for various languages is a prerequisite for progress in the field. Access to resources regarding licenses for research needs to be facilitated. In general, more text corpora with language data need to be made available for research. Shared formats for representing NLP data and results such as the already existing (ITS 2.0, TAPPICC, etc.) are much needed, since the heterogeneity of formats remains an issue, hindering interoperability.

Cooperation between industry and research is necessary to utilising the available funding in the best possible way. But also the connection to everyday usage should not be lost. Making language technology more accessible to a wide variety of people (e. g., language professionals wanting to use it for their work) is key here. We need to raise awareness among users who do not think they "need" multilingualism. Accessibility, through sign language, avatar technology, incl. motion capturing and sign language recognition are equally important when trying to make LT available and useful to everyone.

4.3. Development and deployment of Language Technology for all European languages

The interview participants were asked about their ideas (e. g., specific policies or regulations) on how to speed up the development and deployment of Language Technology in Europe equally for all European languages.

One suggestion was to require all EC-funded projects to have a language diversity plan and show how they will ensure that their results are applicable to people who speak languages other than English. Regulations to strengthen support for minority or low resource languages and the allocation of funds for their preservation and development of language technologies could make a real difference.

Also regarding EU projects, it was suggested to make it a requirement that in addition to having partners from at least three countries, at least three (or more) languages are to be

supported in the interface of the software application with one of them from a less-widely spoken language. As has become customary, the partners are from many different European countries, and with time, the projects will most likely be supporting all EU languages by default using open source NLP libraries with standard APIs. This will create a real demand for a marketplace with NLP components like the European Language Grid. Having this as a required section in applications, and serious consideration of it in the funding phase, could really change things, generally at little additional cost. Following this approach would also support the creation of competitive large multilingual language models covering all European languages and other languages with importance to European industry and society.

In general, it is important to position language technology as a field with a long-term demand. For instance, the EU made it a requirement to translate into Irish which has increased demand for new NLP applications. Unless it is already the case, the EU institutions should demand an open API to support additional languages when buying software to enable SMEs to provide support for less-widely spoken languages and to avoid dependence on large international software providers with their proprietary in-house palettes of languages.

Regulations to require software companies selling more than a certain amount in the EU to provide support for the EU official and regional languages could also be a solution, knowing that regulatory requirements are a major driver for localisation – bigger than consumer demand – in certain sectors.

It was also suggested to set up a strategy to construct a multilingual language technology benchmark representing common problems for the European industry with all or at least most of the European languages to achieve technological sovereignty. A European "SuperGLUE"-style shared benchmark and leaderboard would provide a central target for the community to work on.

A tool to help speed up the development and deployment would be to regulate the exchange of datasets, as that is needed to improve technologies. This could be the provision of open source language data by the European Commission or Member States to enable a broader range of developers to train their own models. It is crucial to facilitate making existing technologies, models and tools (open source and other types) available and enabling exchange. Too much is not being used that has been developed. If necessary, the use of language data for language models from (inappropriate interpretations of) ancillary copyright law should be exempted.

4.4. Positioning of European Language Technology on a global level

The interviewees were asked for suggestions how to position European Language Technology Research and Development strategically on the global level in the long run.

The answers stress that 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 EU languages. There needs to be a focus on diversity. Languages are valuable for Europe as they contain our culture and Europe has vast experience in multilingualism (compared to, say, the U.S.), which is an asset that needs to leveraged. Europe is unique in that there is a large number of languages, large and small in numbers of speakers, that have high prestige and value in their local markets. Whereas elsewhere in the world it is easier to disregard minority languages, in Europe, these languages typically are standard governmental and educational languages in their local markets. There is lots of diversity in emerging markets such as India and Africa where NLP could be more quickly adapted if a strong and leading European example existed.

Competing with the budgets of big companies and globally renowned universities is an im-

possible endeavour at this point. In many application areas of AI including LT, the U.S. and China are leading, especially in online applications and other end-consumer products and services. The EC should not focus on competing in these areas until there are industries that actually stand the chance of competition. However, European AI still has huge opportunities in Enterprise AI, AI solutions for sustainability, new forms of education, personalised medicine, intercultural business or citizen participation in political and social processes. There is still huge potential in the next generation of AI that combines neural learning and knowledge processing including common sense knowledge. Moreover, Europe has a good chance to become the leader in trustworthy, responsible, ethical AI but it can only reach this goal by creating powerful AI/LT applications exhibiting the right properties, not by fabricating restrictions without being able to demonstrate and follow these in a competitive way.

A particular focus should be put on open ecosystems, fostering publicly available resources that facilitate innovation and research for both commercial and non-commercial actors – in contrast to the "locked-in" and not very transparent approach of some of the larger international providers. Supporting the development of European open-source solutions should be a key focus including the development of large AI models and providing them as open-source to build an ecosystem of data, models and algorithms. There needs to be more delivery of open-source infrastructure and tools (such as in Stanza or spaCy) with standardised high quality levels for all European languages, deployed through the European Language Grid. Making sure that speech and text processing modules for all EU languages with standardised APIs are available through a marketplace like ELG can counteract the dominance of big international companies that consider only their own standardised language palettes, neglecting opportunities for less-widely spoken languages.

Moreover, the EU could enable its Member States to acquire LT for their local industries without depending on non-European technology providers. This stimulates growth in the LT industry as well as in classical industries. Providing local industries with LT capabilities, supporting their native languages, and with actionable insights is an excellent opportunity to strengthen the European market and create technological sovereignty.

In this context, it has been noted that keeping skilled professionals in Europe is a serious problem. Scenarios where MT or other LT is developed in Europe and European engineers then get hired overseas with higher salaries and better funded career opportunities have to be avoided. To support the creation of several core centers of competence for Language AI and NLP at university-level would be a step in the right direction to keep talent in Europe. Programmes have to be developed that position language as a STEM subject (science, technology, engineering, and mathematics) early, rather than keeping it almost exclusively in its current perception as a subject of the humanities.

Currently, there is potential to do a lot better with exploitation. Streamlining routes to funding for industrial and academic collaborations would help. Targeted, medium-length projects with low application overhead and relatively rapid turnaround (like the extremely well received ELG Open Calls and Pilot Projects) would support this even further. LT R&D should get a prominent place in EU research and innovation projects, both as a separate focus as well as part of other AI and related innovation programs, as it is an essential factor in many areas and applications and it is the first entry point to the citizen.

4.5. Key challenges for a long-term Language Technology funding programme

The interviewees were asked about the key challenges associated with a large-scale, long-term funding programme dedicated to European LT research, development and innovation.

One of the key challenges lies in targeting European language equality. Linguistic diversity and focusing on preserving our culture and history are key pillars. Europe has a responsi-

bility for the many minority and threatened languages as they are part of our cultural heritage. However, current EU projects tend to concentrate on a limited, non-threatened subset of languages, allowing the technology gap between less-resourced and well-resourced European languages to grow. A mechanism for encouraging work on non-majority languages and dialects should be in place (similar to country-level diversity through Inclusiveness Target Countries in COST actions). Supporting low-resourced, local language variants should be made more attractive as there is an obvious lack of viable economic opportunity for some languages. Further, multilingual technologies made in Europe should be demonstrated and exported in order to help other areas of the world such as regions in Africa or Southeast Asia to adopt the European model of supranational integration with the preservation of multilingualism, diversity and equality.

Building stronger active co-operations on the strategic and public policy front as well as co-operations with businesses is critical. Alliances across European R&D and LT providers could be fruitful, encouraging industry to take more initiative to give ideas and define what consumers need. There are many mid-sized stakeholders and many different contexts in which to share technology. Connecting tech providers and consumers is key in this intrinsically diverse environment and the EU should fund this kind of knowledge transfer. Furthermore, with 24 official languages plus regional and minority languages, providing actual localised support in projects would add significant effort to any project. Many organisations struggle with a few languages already, but moving to 25–30 languages would require small companies to implement the kind of support that typically only large corporations are capable of providing. Therefore, requiring only internationalisation enablement is more realistic.

Another challenging task is the compilation of data, ensuring data and annotation quality across languages and application domains. Not only is there a low amount of non-English labeled data, but not enough data is free to use. Enabling sharing and better use of resources is absolutely essential. Historical material could also be included more frequently. Most languages in Europe have long written histories which can be leveraged to increase the historical depth of collections and language models. Competitive multilingual European language models and other base technologies that cover at least all EU languages and that can be used by industry to provide their products and services to all European citizens are much needed. Many member states are excluded from high-performing language technology due to a lack of large open-source language models. When working with data, issues of quality related to gender and race bias need to be taken into account. European LT should work for everyone and inclusiveness and accessibility are important factors.

Utilising AI with a strategy that is thought through and well funded is another challenge. Europe should develop large language models. Equal access to LT and AI should be granted and the playing field should be leveled so that commercial and non-commercial stakeholders alike are not shut out of innovation due to a lack of access and affordability to the (computational) resources necessary to train large language models. Integrating explainable AI into NLP is still an open challenge. Machine learning models with estimates that are more transparent to the user are needed. Combining interactive LT (conversational AI) with text, knowledge and multimedia technologies for a new generation of applications is also an unsolved challenge. One prerequisite is an accessible European compute infrastructure.

Also, tested and certified reliable accurate context-aware specialised machine translation for a number of critical use cases such as medical use cases (anamnesis, emergency room, teleconsulting), security related situations (e. g., immigration, asylum seeking, threats) contract negotiations or commercial disputes is an open challenge to be tackled.

In the context of research and development, sustainability also plays a significant role. While many projects achieve good results, they are too often not usable after the end of the project. The LT field is in a well-known reproducibility crisis. Results are published or used in one context, but cannot be reproduced or replicated elsewhere. Sustainability goes hand in hand with public awareness. Delivering real results outside the lab is crucial.

Implementation of incubators could be a solution to promote results made in EU projects.

Further, developing incentive structures that make staying in Europe more attractive and prevent brain drain to the US should be a priority. Last but not least, there are societal and system-inherent factors that hinder the progress in the field of European LT. The pervasive European bureaucracy hampers fast and agile processes. There is a need to react more quickly to new developments and processes. At the moment, Europe is too slow in adaption and mindset. As a result, regulations often stand in the way of speedy implementations in urgently needed LT solutions and applications in real-life scenarios.

4.6. Additional suggestions with regard to European Language Equality

The interviewees were asked about additional suggestions or topics they want to emphasise with regard to European language equality and our strategic recommendations to the EU.

One respondent stressed that ELE is an extremely important project for smaller languages, which is why digital equality efforts need to be enforced. There is a need for continuous support for language-focused projects bringing LT providers and users together in innovation projects to jointly develop and improve platforms and tools. An adequate, accessible and interoperable infrastructure with resources and tools for all European and other important languages such as ELG is needed and must be reinforced. Promoting minority languages and sign languages and supporting other information accessibility measures, e. g., the use of simple and easy language is a key takeaway to ensure language equality.

Looking at the current EU funding programmes, suggestions for alternative administration processes were made. More two-stage calls and simplified EU bureaucracy are desirable. Closing the European gap between application-oriented basic research and commercially focused technology requires new forms of funding and organisation. Almost all IT breakthroughs in the U.S. were achieved as part of the DARPA (Defense Advanced Research Projects Agency) programs, often based on European basic research results. The decisive factor is not only the different funding sources but also the well-organised competitive scheme. This entails strong leadership by programme managers and the planning of goals and objective success criteria by top experts under the leadership of the programme manager. Typically, there is parallel funding of several competing consortia and regular technology competitions on especially prepared benchmarks. The European LT landscape and language equality in general would benefit immensely from such an approach.

To promote speech processing, the recordings of speeches in the European Parliament are available online for speech recognition training, but similar recordings for colloquial speech collected on a national level and transcribed for speech recognition training would be valuable to make sure that everyday speech is available as well.

Recently gained new insights on the power and limitations of large language models suggest a revised approach to dealing with the problem of under-resourced languages. The development of cross-lingual applications becomes much easier with a large multilingual language model, even if some languages are only represented by rather small data volumes in pre-training. Even for monolingual applications, languages with smaller numbers of speakers benefit tremendously from multilingual language models. Adding explicit knowledge to language models boosts the performance for larger and smaller languages with respect to the proportions of pre-training data. Instead of censoring the data that go into the language model (as in, for example, Clean Common Crawl Corpus) post-training can be used for the reduction of bias and toxicity.

5. Conclusions

The survey addressed to LT researchers and developers sought to elicit the respondents' views in a structured way that facilitates the analysis, consolidation and integration of the collected feedback into the ELE SRIA and roadmap. Both the survey and interviews brought together a diverse and demographically distributed audience, proving that there is profound common interest and passion not only with regard to Multilingual Europe but also in making a large-scale, long-term funding programme for European LT development a reality.

The survey respondents represent 42 different organisations, out of which 67% are research or academic institutions, the remaining ones are industry practitioners and independent research centres, headquartered in 32 countries. Most have expertise in various natural language processing services and offer their products and services for multiple domains. While the respondents cover 32 languages actively in their research or portfolio of services, software, resources, models etc., the majority has a strong focus on English, German, French, Spanish and Italian. Even though expansion to more languages is desired and met with a great interest and goodwill by both academic and industry players, limited funding, demand and obstacles with regard to available resources, make it a challenging endeavour.

Almost all participants agree uniformly that basic research is still urgently needed. They also recognise was the fact that many early stage researchers are attracted by non-European companies which offer higher compensation, but also more elaborated research conditions.

Also, more than three quarters either strongly agree or agree that the fragmentation of the LT industry remains a challenge to this day. As a matter of fact, this obstacle makes the efforts of the ELG initiative even more important and valuable. To counteract the still existing fragmentation, it is necessary to start building common data and product spaces, LT business network spaces, foster small and medium LT businesses through public procurement, strengthen SME-research ties and create conditions for growth and scaling of SMEs beyond their often rather small national or language focus. It was frequently mentioned that industry-oriented initiatives such as ELG have to be strengthened and expanded.

With regard to visions for the future, it was stressed that basic and applied research must not be neglected as it is still needed in many areas. To regain European global excellence and leadership in LT, massive investments in industry and research collaboration for the creation and deployment of innovative LT solutions are needed. A long-term project of ten or more years can potentially lead to breakthrough research and subsequently to the desired leap from simple and superficial language processing to deep language understanding.

The expert interviews echoed many of the points already mentioned. The interviews represent the perspective of both industry and academia but also the more political level. Nearly all participants have longstanding expertise in various LT areas.

For the main topics of research and development until 2030, there was a huge consensus on developing large and efficient language models. While text technologies have long been the main interest of NLP, speech technology and speech resources need more attention. The interviewees also emphasise that a crucial topic to concentrate on until 2030 is language diversity and equality and the accompanying provisioning of basic technologies and services for languages outside of the often preferred group of languages such as English, German, Spanish, and French. To speed up the development and deployment of Language Technology in Europe equally for all European languages, a restructuring of the current EU project schemes and new regulations could be helpful.

The interviews have emphasised that 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.

6. Acknowledgements

We would like to thank the following interviewees for their time to conduct interviews with us, providing insightful feedback and food for thought.

Alexander Baratsits	Cultural Broadcasting Archive	Austria
Marc Berthiaume	European Commission	Belgium
Jörg Bienert	KI Bundesverband e. V.	Germany
Patrick Bunk	Ubermetrics Technologies GmbH	Germany
Leon Derczynski	IT University of Copenhagen	Denmark
Alexander Drechsel	European Parliament	Belgium
Florian Faes	Slator	Switzerland
Jóhanna Vigdís Guðmundsdóttir	Almannarómur	Iceland
Jussi Karlgren	Gavagai and KTH Royal Institute of Technology	Sweden
Peggy van der Kreeft	Deutsche Welle	Germany
Krister Lindén	University of Helsinki	Finland
Arle Lommel	CSA Research	USA
John McCrae	Insight Centre for Data Analytics	Ireland
Arantza Del Pozo	Vicomtech	Spain
Alexander Siebert	Retresco GmbH	Germany
Themos Stafylakis	Omilia Ltd.	Greece
Michael Stormbom	Lingsoft Oy	Finland
Georg Tschare	Sign Time GmbH	Austria
Hans Uszkoreit	DFKI GmbH and Giance GmbH	Germany
Sebastian Wohlrapp	Field 33 GmbH	Germany

Table 4: Interview participants

References

Aivars Berzins, Khalid Choukri, Maria Giagkou, Andrea Lösch, Helene Mazo, Stelios Piperidis, Mickaël Rigault, Eileen Schnur, Lilli Small, Josef van Genabith, Andrejs Vasiljevs, Andero Adamson, Dimitra Anastasiou, Natassa Avraamides-Haratsi, Núria Bel, Zoltán Bódi, António Branco, Gerhard Budin, Virginijus Dadurkevicius, Stijn de Smeytere, Hrístina Dobreva, Rickard Domeij, Jane Dunne, Kristine Eide, Claudia Foti, Maria Gavriilidou, Thibault Grouas, Normund Gruzitis, Jan Hajic, Barbara Heinisch, Verónique Hoste, Arne Jönsson, Fryni Kakoyianni-Doa, Sabine Kirchmeier, Svetla Koeva, Lucia Konturová, Jürgen Kotzian, Simon Krek, Gauti Kristmannsson, Kaisamari Kuhmonen, Krister Lindén, Teresa Lynn, Armands Magone, Hélène Mazo, Maite Melero, Laura Mihailescu, Simonetta Montemagni, Micheál Õ Conaire, Jan Odijk, Maciej Ogrodniczuk, Pavel Pecina, Jon Arild Olsen, Bolette Sandford Pedersen, David Perez, Andras Repar, Ayla Rigouts Terryn, Eirikur Rögnvaldsson, Mike Rosner, Nancy Routzouni, Claudia Soria, Alexandra Soska, Donatienne Spiteri, Marko Tadic, Carole Tiberius, Dan Tufis, Andrius Utka, Paolo Vale, Piet van den Berg, Tamás Váradi, Kadri Vare, Andreas Witt, Francois Yvon, Janis Ziedins, and Miroslav Zumrik. *Sustainable Language Data Sharing to Support Language Equality in Multilingual Europe - Why Language Data Matters: ELRC White Paper*. ELRC Consortium, 2 edition, 2019. ISBN 978-3-943853-05-6.

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

András Kornai. Digital language death. *PloS one*, 8(10):e77056, 2013.

Georg Rehm. Strategic Agenda for the Multilingual Digital Single Market – Technologies for Overcoming Language Barriers towards a truly integrated European Online Market, April 2015. Version 0.5. April 22, 2015. Prepared by the EU-funded projects CRACKER and LT_Observatory.

Georg Rehm. Language as a Data Type and Key Challenge for Big Data. Strategic Research and Innovation Agenda for the Multilingual Digital Single Market. Enabling the Multilingual Digital Single Market through technologies for translating, analysing, processing and curating natural language content, July 2016. URL http://cracker-project.eu/sria/. Version 0.9. July 04, 2016. Prepared by the Cracking the Language Barrier federation, supported by the EU-funded projects CRACKER and LT_Observatory.

Georg Rehm. Language Technologies for Multilingual Europe: Towards a Human Language Project. Strategic Research and Innovation Agenda, November 2017. URL http://cracker-project.eu/sria/. Version 1.0. Unveiled at META-FORUM 2017 in Brussels, Belgium, on November 13/14, 2017. Prepared by the Cracking the Language Barrier federation, supported by the EU-funded project CRACKER.

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

Georg Rehm and Hans Uszkoreit, editors. *The META-NET Strategic Research Agenda for Multilingual Europe 2020*. Springer, Heidelberg, New York, Dordrecht, London, 2013. URL http://www.meta-net.eu/sra. More than 200 contributors from research and industry. This book at springer.com or amazon.de.

Georg Rehm, Hans Uszkoreit, Ido Dagan, Vartkes Goetcherian, Mehmet Ugur Dogan, Coskun Mermer, Tamás Váradi, Sabine Kirchmeier-Andersen, Gerhard Stickel, Meirion Prys Jones, Stefan Oeter, and

Sigve Gramstad. An Update and Extension of the META-NET Study "Europe's Languages in the Digital Age". In Laurette Pretorius, Claudia Soria, and Paola Baroni, editors, *Proceedings of the Workshop on Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era (CCURL 2014)*, pages 30–37, Reykjavik, Iceland, May 2014.

Georg Rehm, Jan Hajic, Josef van Genabith, and Andrejs Vasiļjevs. Fostering the Next Generation of European Language Technology: Recent Developments – Emerging Initiatives – Challenges and Opportunities. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Marko Grobelnik, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the 10th Language Resources and Evaluation Conference (LREC 2016)*, pages 1586–1592, Portorož, Slovenia, May 2016a. European Language Resources Association (ELRA).

Georg Rehm, Hans Uszkoreit, Sophia Ananiadou, Núria Bel, Audronė Bielevičienė, Lars Borin, António Branco, Gerhard Budin, Nicoletta Calzolari, Walter Daelemans, Radovan Garabík, Marko Grobelnik, Carmen García-Mateo, Josef van Genabith, Jan Hajič, Inma Hernáez, John Judge, Svetla Koeva, Simon Krek, Cvetana Krstev, Krister Lindén, Bernardo Magnini, Joseph Mariani, John McNaught, Maite Melero, Monica Monachini, Asunción Moreno, Jan Odjik, Maciej Ogrodniczuk, Piotr Pęzik, Stelios Piperidis, Adam Przepiórkowski, Eiríkur Rögnvaldsson, Mike Rosner, Bolette Sandford Pedersen, Inguna Skadiņa, Koenraad De Smedt, Marko Tadić, Paul Thompson, Dan Tufiş, Tamás Váradi, Andrejs Vasiļjevs, Kadri Vider, and Jolanta Zabarskaite. The Strategic Impact of META-NET on the Regional, National and International Level. *Language Resources and Evaluation Journal*, 50(2):351–374, 2016b. 10.1007/s10579-015-9333-4.

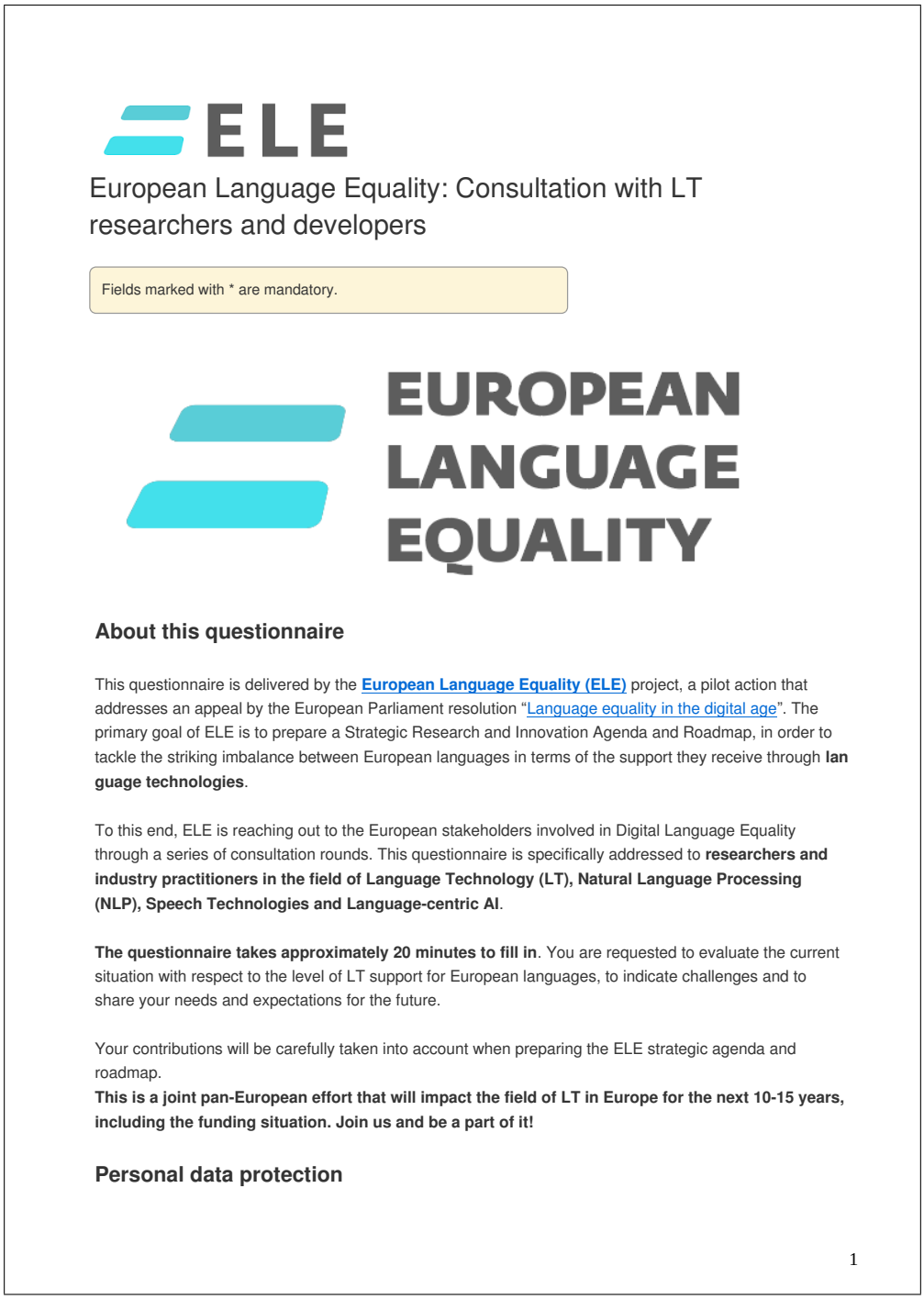
Georg Rehm, Maria Berger, Ela Elsholz, Stefanie Hegele, Florian Kintzel, Katrin Marheinecke, Stelios Piperidis, Miltos Deligiannis, Dimitris Galanis, Katerina Gkirtzou, Penny Labropoulou, Kalina Bontcheva, David Jones, Ian Roberts, Jan Hajic, Jana Hamrlová, Lukáš Kačena, Khalid Choukri, Victoria Arranz, 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 2020a. European Language Resources Association (ELRA).

Georg Rehm, Katrin Marheinecke, Stefanie Hegele, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Khalid Choukri, Andrejs Vasiļjevs, Gerhard Backfried, Christoph Prinz, José Manuel Gómez Pérez, Luc Meertens, Paul Lukowicz, Josef van Genabith, Andrea Lösch, Philipp Slusallek, Morten Irgens, Patrick Gatellier, Joachim Köhler, Laure Le Bars, Dimitra Anastasiou, Albina Auksoriūtė, Núria Bel, António Branco, Gerhard Budin, Walter Daelemans, Koenraad De Smedt, Radovan Garabík, Maria Gavriilidou, Dagmar Gromann, Svetla Koeva, Simon Krek, Cvetana Krstev, Krister Lindén, Bernardo Magnini, Jan Odijk, Maciej Ogrodniczuk, Eiríkur Rögnvaldsson, Mike Rosner, Bolette Pedersen, Inguna Skadina, Marko Tadić, Dan Tufiș, Tamás Váradi, Kadri Vider, Andy Way, and François Yvon. The European Language Technology Landscape in 2020: Language-Centric and Human-Centric AI for Cross-Cultural Communication in Multilingual Europe. 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 3315–3325, Marseille, France, 5 2020b. European Language Resources Association (ELRA).

STOA. Language equality in the digital age – Towards a Human Language Project. STOA study (PE 598.621), IP/G/STOA/FWC/2013-001/Lot4/C2, March 2017. Carried out by Iclaves SL (Spain) at the request of the Science and Technology Options Assessment (STOA) Panel, managed by the Scientific Foresight Unit (STOA), within the Directorate-General for Parliamentary Research Services (DG EPRS) of the European Parliament, March 2017. http://www.europarl.europa.eu/stoa/.

A. The LT researchers and developers full survey

Figures 5 to 13 show the complete LT research and developers survey.



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 5: 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.

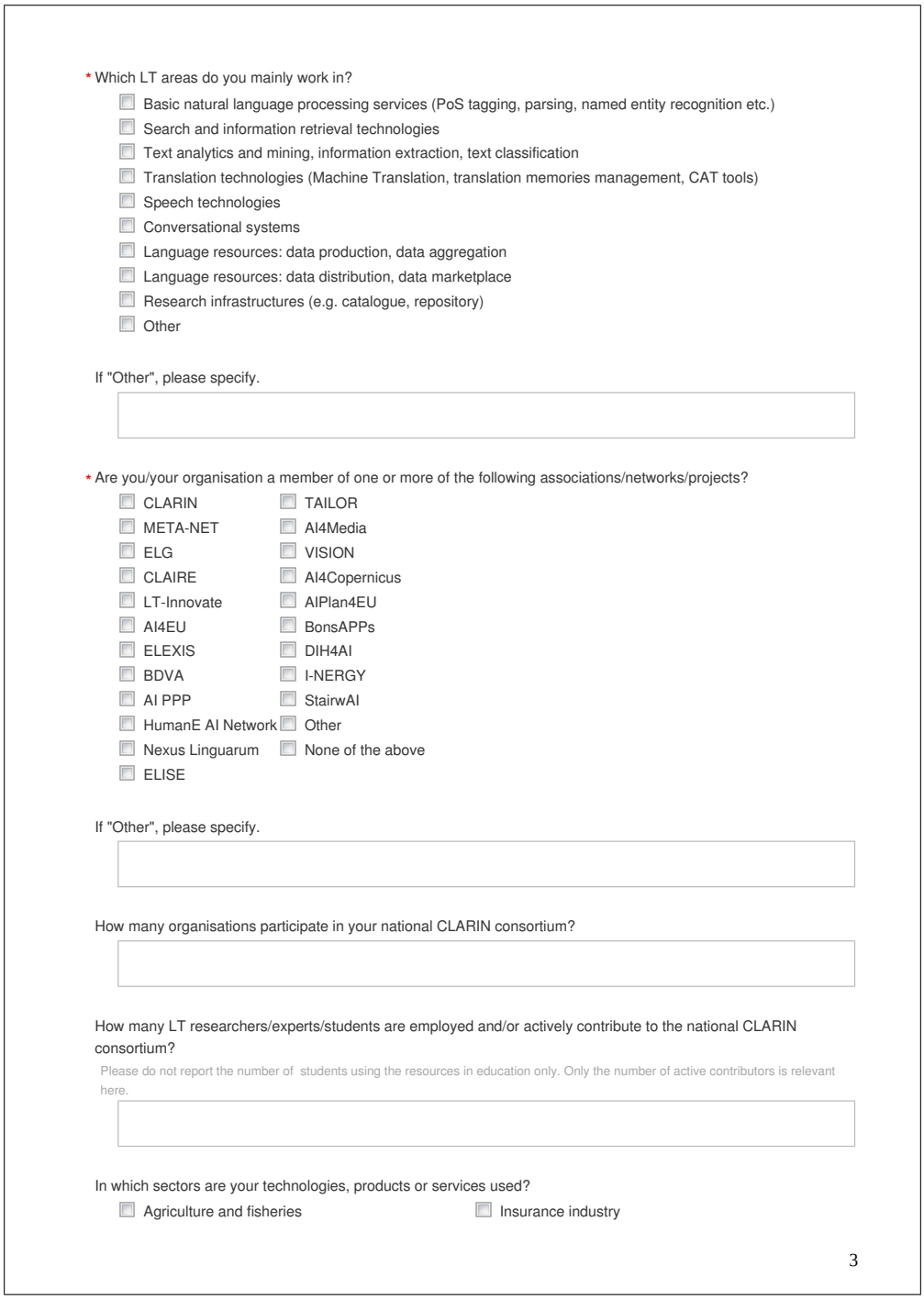
* Where is your organisation's headquarters based?

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 6: 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?

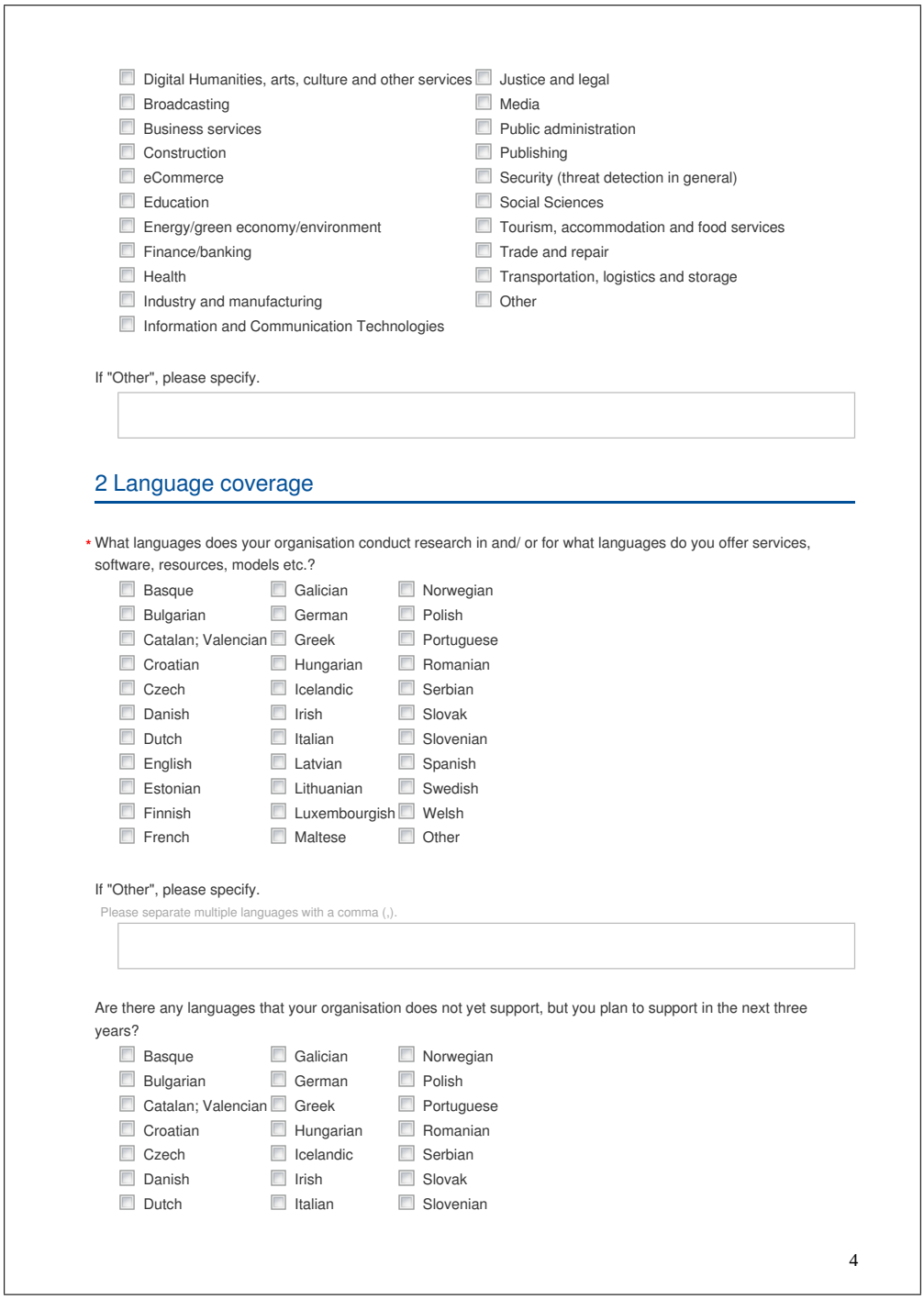
Please do not report the number of students using the resources in education only. Only the number of active contributors is relevant here.

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

- [ ] Agriculture and fisheries
- [ ] Insurance industry

3

Figure 7: 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.

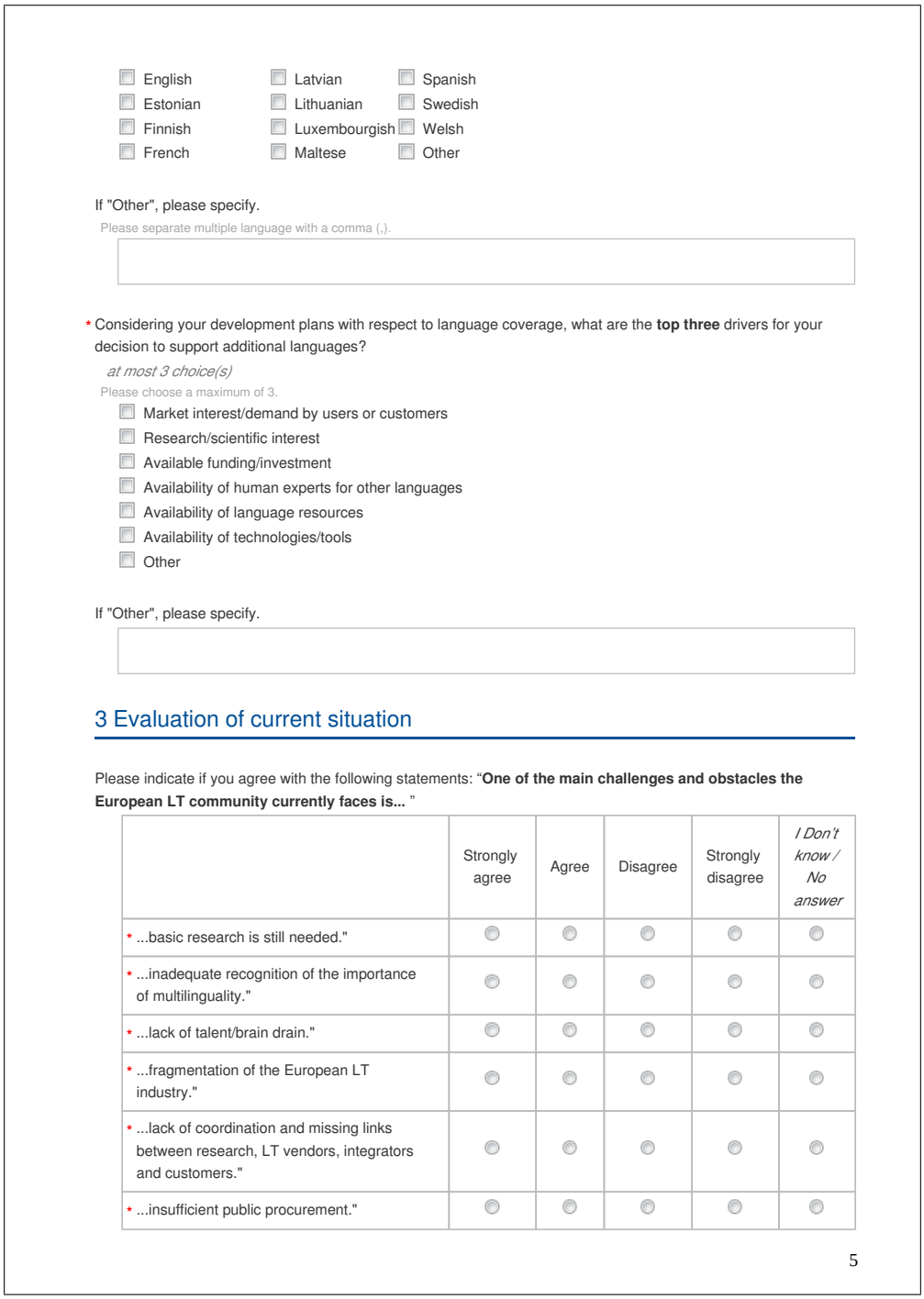
Please separate multiple languages with a comma (,).

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 8: 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.

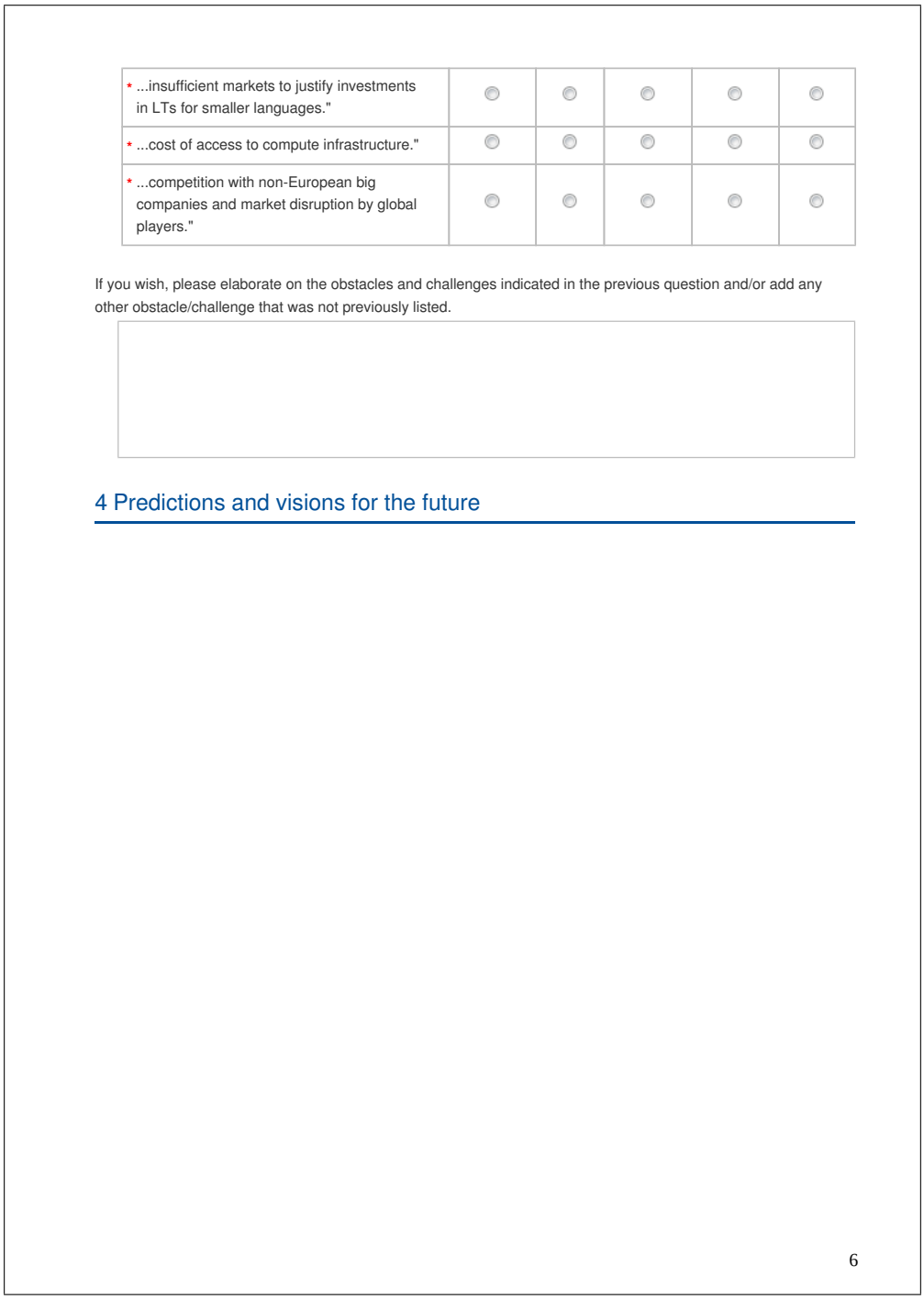
3 Evaluation of current situation

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 9: Full survey as published (page 5/9)



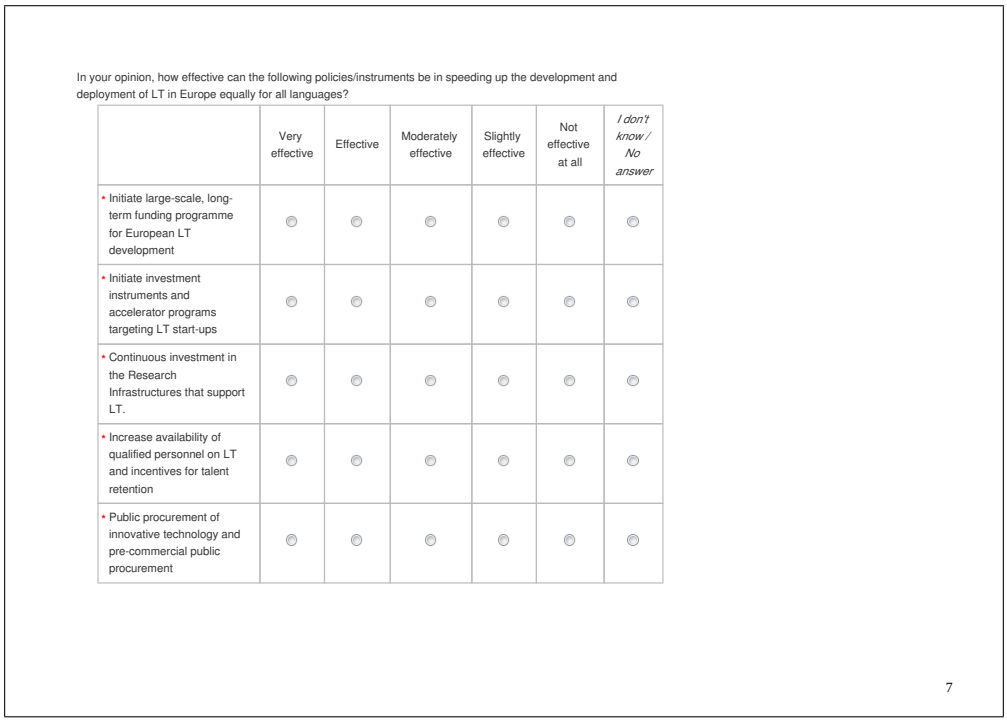
* ...insufficient markets to justify investments in LTs for smaller languages."					
* ...cost of access to compute infrastructure."					
* ...competition with non-European big companies and market disruption by global players."					

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 10: Full survey as published (page 6/9)

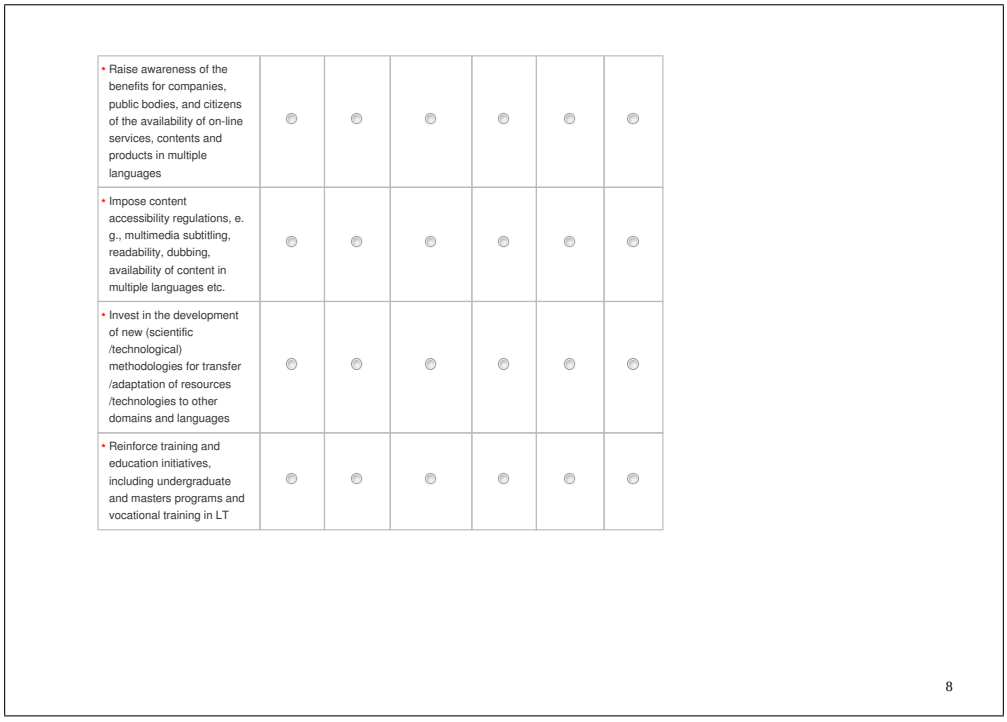


In your opinion, how effective can the following policies/instruments be in speeding up the development and deployment of LT in Europe equally for all languages?

	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

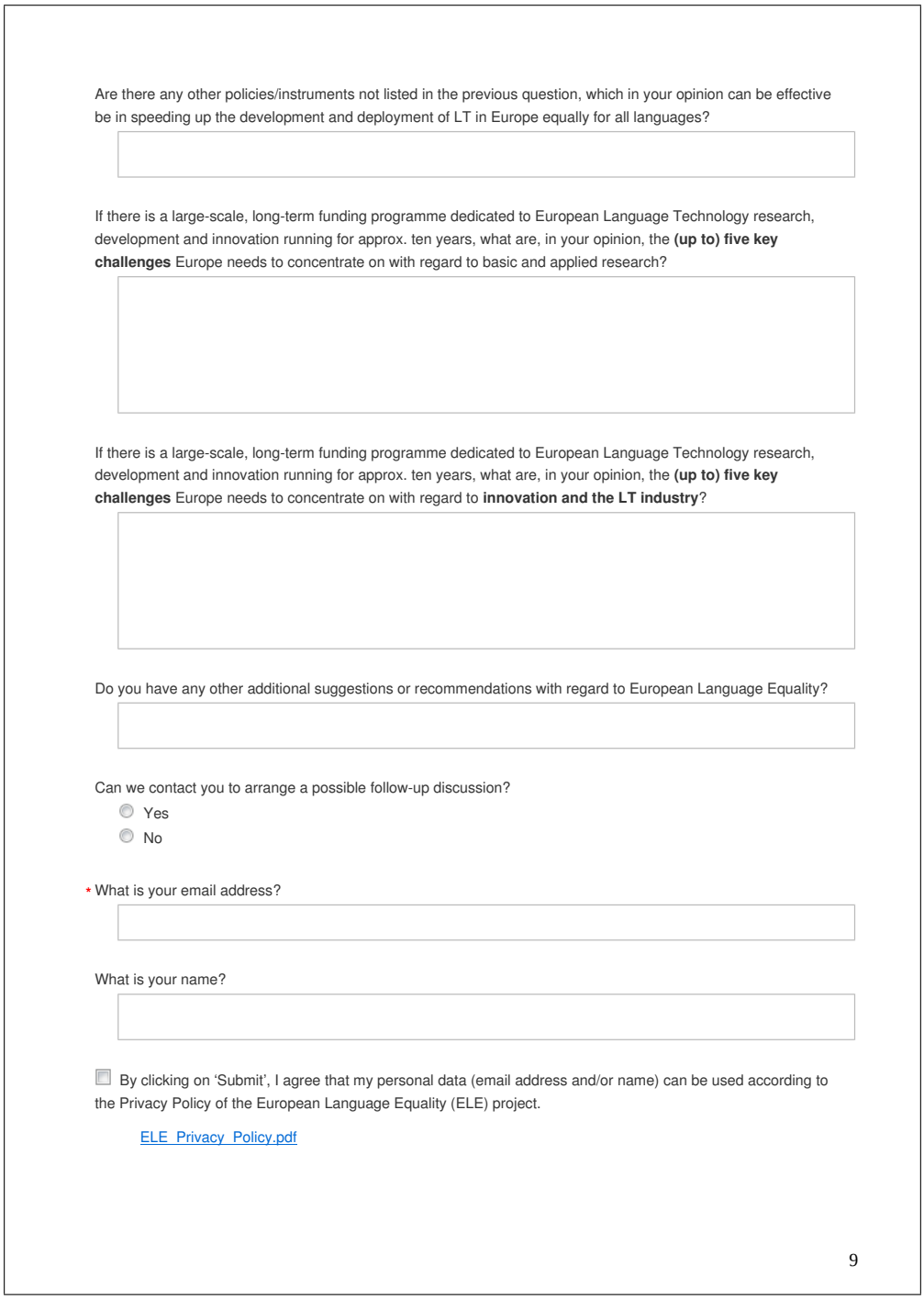
Figure 11: Full survey as published (page 7/9)



* 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 12: 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 13: Full survey as published (page 9/9)

B. Additional tables and graphs

Type of organisation	Answers (Perc.)	
Research center (independent)	8	15%
University or academic research	36	67%
Large enterprise	1	2%
SME	6	15%
Other	1	2%
Total	54	100%

Table 5: Breakdown of answers to “Which of the following best describes the type of organisation you work for?” (mandatory closed question)

Country	Respondents (Percentage)	
Austria	3	6%
Belgium	1	2%
Bulgaria	1	2%
Croatia	1	2%
Cyprus	0	0%
Czechia	0	0%
Denmark	3	6%
Estonia	0	0%
Finland	4	7%
France	4	7%
Germany	6	11%
Greece	2	4%
Hungary	0	0%
Iceland	2	4%
Ireland	2	4%
Italy	2	4%
Latvia	3	6%
Lithuania	3	6%
Luxembourg	1	2%
Malta	2	4%
Netherlands	0	0%
Norway	0	0%
Poland	1	2%
Portugal	1	2%
Romania	4	7%
Serbia	1	2%
Slovak Republic	1	2%
Slovenia	0	0%
Spain	3	6%
Sweden	1	2%
Switzerland	1	2%
United Kingdom	1	2%
Total	54	100%

Table 6: Breakdown of answers to “Where is your organisation’s headquarter based?” (mandatory closed question, plus “if other” as optional open-ended question)

Sectors	Number of mentions
Information and Communication Technologies	39
Other	38
Arts, culture and other services	35
Education	32
Public administration	28
Bussiness services	20
Health	20
Finance/banking	13
Justice and legal	12
Publishing	12
eCommerce	10
Security (threat detection in general)	7
Transportation, logistics and storage	6
Energy/green economy/environment	5
Insurance industry	5
Construction	3
Agriculture and fisheries	2
Trade and repair	1
Industry and Manufacturing	0
Tourism, accomodation and food services	0

Table 7: Sectors in which technologies, products or services are used

LT areas	Number of mentions
Basic natural language processing services (PoS tagging, parsing, named entity recognition etc.)	42
Language resources: data production, data aggregation	40
Translation technologies (Machine Translation, translation memories management, CAT tools)	31
Text analytics and mining, information extraction, text classification	27
Speech technologies	23
Research infrastructures (e.g. catalogue, repository)	22
Conversational systems	17
Search and information retrieval technologies	16
Language resources: data distribution, data marketplace	11
Other	4

Table 8: Full list of answers to “Which LT areas do you mainly work in?

	Strongly agree	**Agree**	**Disagree**	**Strongly disagree**	**I don’t know / No answer**
basic research is still needed	24	28	2	0	0
inadequate recognition of the importance of multilinguality	21	24	7	0	2
lack of talent/brain drain	6	13	23	7	5
fragmentation of the European LT industry	15	27	5	1	6
lack of coordination and missing links between research, LT vendors, integrators and customers	20	21	6	2	5
insufficient public procurement	21	17	2	2	12
insufficient markets to justify investments in LTs for smaller languages	19	22	9	0	4
cost of access to compute infrastructure	9	24	14	1	6
competition with non-European big companies and market disruption by global players	27	22	3	1	1

Table 9: 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 et all	I don't know / No answer
Initiate large-scale, long-term funding programme for European LT development	33	15	4	0	2	0
Initiate investment instruments and accelerator programs targeting LT start-ups	16	20	14	2	1	1
Continuous investment in the Research Infrastructures that support LT	25	24	4	1	0	0
Increase availability of qualified personnel on LT and incentives for talent retention	15	28	7	3	1	0
Public procurement of innovative technology and pre-commercial public procurement	16	20	6	5	2	5
Raise awareness of the benefits for companies, public bodies, and citizens of the availability of on-line services, contents and products in multiple languages	20	17	11	4	1	1
Impose content accessibility regulations, e.g., multimedia subtitling, readability, dubbing, availability of content in multiple languages etc.	17	16	13	4	1	3
Invest in the development of new (scientific/technological) methodologies for transfer/adaptation of resources/technologies to other domains and languages	19	26	5	4	0	0
Reinforce training and education initiatives, including undergraduate and masters programs and vocational training in LT	20	20	12	2	0	0

Table 10: 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")