EUROPEAN
LANGUAGE
EQUALITY

# D2.5

# Report from META-NET

| | |
|---|---|
| Authors | Jan Hajič, Tea Vojtěchová, Maria Giagkou |
| 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 | D2.5 |
| Deliverable title | Report from META-NET |
| Type | Report |
| Number of pages | 42 |
| Status and version | Final |
| Dissemination level | Public |
| Date of delivery | Contractual: 28-02-2022 – Actual: 28-02-2022 |
| Work package | WP2: European Language Equality – The Future Situation in 2030 |
| Task | Task 2.1 The perspective of European LT developers (industry and research) |
| Authors | Jan Hajič, Tea Vojtěchová, Maria Giagkou |
| Reviewers | Maria Giagkou, Stefanie Hegele |
| EC project officers | Susan Fraser, Miklos Druskoczi |
| Contact | European Language Equality (ELE)<br>ADAPT Centre, Dublin City University<br>Glasnevin, Dublin 9, Ireland<br>Prof. Dr. Andy Way – andy.way@adaptcentre.ie<br>European Language Equality (ELE)<br>DFKI GmbH<br>Alt-Moabit 91c, 10559 Berlin, Germany<br>Prof. Dr. Georg Rehm – georg.rehm@dfki.de<br>http://www.european-language-equality.eu |
| |  |

# Consortium

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

# Contents

# List of Figures



# List of Tables

# List of Acronyms

| | |
|---|---|
| AI | Artificial Intelligence |
| CEF | Connecting Europe Facility |
| CRACKER | Cracking the Language Barrier (EU project, 2015–2017) |
| CULT | European Parliament's Committee on Culture and Education |
| DARIAH | Digital Research Infrastructure for the Arts and Humanities |
| DH | Digital Humanities |
| DSM | Digital Single Market |
| 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) |
| ELRA | European Language Resource Association |
| ELRC | European Language Resource Coordination |
| ELT | European Language Technology |
| ERC | European Research Council |
| GDPR | General Data Protection Regulation |
| HPC | High-Performance Computing |
| IoT | Internet of Things |
| IPR | Intellectual Property Rights |
| LT | Language Technology/Technologies |
| META-NET | EU Network of Excellence to foster META |
| ML | Machine Learning |
| NLP | Natural Language Processing |
| PSI | Public Sector Information |
| SRA | Strategic Research Agenda |
| SRIA | Strategic Research and Innovation Agenda |
| SSH | Social Sciences and the Humanities |
| STOA | Science and Technology Options Assessment |
| TDM | Text and Data Mining |

## Abstract

In 2012, META-NET published the META-NET White Paper series (Rehm and Uszkoreit, 2012) that described the situation of digital "well-being" of 31 European languages. In addition, the META-NET Strategic Research Agenda (Rehm and Uszkoreit, 2013) described the state-of-the-art in Language Technology (LT) and the necessary steps in order for the European languages to be treated equally and at a high technological level by 2020. The Strategic Research Agenda has been described in more detail and updated, with a proposition to establish a wide-reaching Human Language Project (Rehm et al., 2014; Rehm, 2015; Rehm et al., 2016; Rehm, 2016, 2017, 2018).

In the meantime, in 2018, the European Parliament published its resolution "Language equality in the digital age", which was passed by the European Parliament in a landslide vote with 592 votes in favour and only 45 against it. The report provides more than 40 recommendations, including the support for the Human Language Project, formulated as the "establishment of a large-scale, long-term coordinated funding programme for research, development and innovation in the field of Language Technology, at European, national and regional levels, tailored specifically to Europe's needs and demands as well as securing Europe's leadership in language-centric AI."

This survey follows a survey on Language Technology for Multilingual Europe conducted in 2018 (Rehm and Hegele, 2018), the results of which reiterated that the biggest challenge for the European Language Technology was the threat of digital extinction of languages with smaller numbers of speakers.

META-NET has conducted the present survey and interviews as part of the ELE project, targeting mainly research organizations and individuals from its membership base. The goal was to get, along with some partners that targeted similar constituencies and some that targeted different ones, a broad view on the status of Language Technology, Language Resources and the overall situation in Europe in language-centric AI, in order to support the conclusions of the 2018 European Parliament Resolution and to move the European LT area forward, eyeing the 2030 horizon.

The results of the present survey show that digital language equality has not been achieved yet. The strongest support as of yet was for a long-term, wide-ranging "European Language Equality Programme" that would support research, data collection and technology transfer to achieve such European language equality. Additional suggestions, especially in the interviews, stressed the importance of properly targeted EU support in terms of, e. g., regulations and procurement, and the importance of LT-oriented support at all levels of the educational system.


## 1. Introduction

This document reports on the results and findings of a consultation with representatives of the Language Technology (LT) community, i. e. industry and research/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, in order 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 collect and prepare language resource data, develop algorithms, pre-commercial LT prototypes, applications and systems (in this Deliverable, we describe most of the results as provided by this group, notably from the META-NET membership);

- 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 encompassing 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/Social Science 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 both research and industry. Specifically, this Deliverable provides results of the surveys and interviews with the META-NET community and membership.

Although the methodology and instruments utilized have been common to all stakeholders, this report covers and analyzes the subset of responses and input from members of the META-NET Network of Excellence.[1]

## 1.1. About META-NET

META-NET is a Network of Excellence dedicated to fostering the technological foundations of a multilingual European information society. These technological foundations cover language resources, general machine learning methods, basic language analysis algorithms and tools, as well as applications under the term “Language Technology(ies).” Language Technologies will:

- enable communication and cooperation across languages,
- secure users of any language equal access to information and knowledge,
- build upon and advance functionalities of networked information technology.

A concerted, substantial, continent-wide effort in LT research and engineering is needed for realising applications that enable automatic translation, multilingual information and knowledge management and content production across all European languages. This effort will also enhance the development of intuitive language-based interfaces to technology ranging from household electronics, machinery and vehicles to computers and robots.

To this end, META-NET is building the Multilingual Europe Technology Alliance (META), bringing together researchers, commercial technology providers, private and corporate language technology users, language professionals and other information society stakeholders.

META-NET consists of 60 research centres from 34 countries dedicated to building the technological foundations of a multilingual European information society. META-NET is forging META, the Multilingual Europe Technology Alliance, an open association with over 1300 members.

In 2012, after a two-years long effort, the final version of the META-NET Strategic Research Agenda for Multilingual Europe 2020 (SRA) was published (Rehm and Uszkoreit, 2012, 2013). This document was the result of a discussion between hundreds of experts from research

[1] Reports from other groups of ELE stakeholders will be published on the ELE website (https://european-language-equality.eu), as they become available.

and industry. At the same time, the White Paper Series about the status of the European Languages had been made available.[2]

The main purpose of the SRA and the White Papers was to raise awareness for the field of LT in Europe and attract the attention of and inform politicians and policy makers on the regional, national and international level in their decisions, especially with regard to the Horizon 2020 Programme and Connecting Europe Facility (CEF).

After the end of the initial project, META-NET has been supported by various projects, mostly CSAs, such as the Cracking the Language Barrier project (CRACKER) and recently its activities have been supported by the European Language Grid project (Rehm et al., 2020a) and, of course, the present European Language Equality Project. As a result of these activities, updates of the Strategic Agenda have been published, as was a Roadmap and additional suggestions, including the suggestion for a wide-ranging and wide-reaching Human Language Project (Rehm et al., 2014; Rehm, 2015; Rehm et al., 2016; Rehm, 2016, 2017, 2018). META-NET's activities have resonated well with the European Parliament, where the STOA (Scientific and Technology Options Assessment) committee[3] has organized several events to support the idea of language equality and the promotion of LT (STOA, 2017). As a result, the European Parliament has adopted, in 2018, the "Language equality in the digital age" resolution (European Parliament, 2018), which contains over 40 recommendations for fostering language equality and supporting the idea that it is only possible by enabling promoting, supporting and funding language technology and the necessary environment around it.

Most recently, Rehm et al. (2020b) described the situation in LT in 2020, in light of the latest developments in the recent ELG and ELE projects, and the perspectives towards 2030, to which the present deliverable contributes.

META-NET is co-organizing annual conferences under the name META-FORUM, where both research, public administrations and industry meet to discuss latest developments in Language Technology, its support from the European Commission, and innovation possibilities, use cases, and other achievements. Strategic documents are often first presented at these conferences. While some results were already presented at META-FORUM 2021, the final results of the ELG and ELE projects will be presented at the upcoming META-FORUM 2022 conference in June 2022.

# 2. Methodology and instruments

The views and opinions of META-NET members have been elicited by two main instruments: an online survey and a round of interviews.

## 2.1. Online survey

The survey addressed to LT researchers and developers, i. e. common to all the groups (and Deliverables) in Task 2.1, 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:

[2] http://www.meta-net.eu/whitepapers/overview
[3] https://www.europarl.europa.eu/stoa/en/home/highlights

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

Table 1: Types of survey questions

- **Part A. Respondents' profiling**: The first part of the survey included 13 questions for the demographic profiling of respondents with emphasis on characteristics relevant to the task at hand, i. e.
  - Country
  - Affiliation
  - Type of organisation
  - LT areas that the respondent is mainly active in
  - Participation/membership in networks/associations
  - Sectors/domains that the respondent is active in (if relevant)
- **Part B. Language coverage**: The second part investigated the degree of coverage of the European languages by the respondents' current research and development activities, i. e.
  - languages currently supported in research/products/services
  - languages planned to be supported in the short-/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.[4] The full survey, as published online, is presented in Appendix A (p. 23 ff.).

The survey was distributed through emails to all current META-NET members. It has additionally been advertised through the ELE, ELG and ELT websites.[5]

The survey was opened on 17 June 2021 and closed on 18 October 2021. In total, 333 responses have been collected, out of which 61 from respondents who indicated that they are members of the META-NET network.[6] This subset of responses, representing the views of META-NET, is analysed in this report.

## 2.2. Interviews

Since most members of META-NET asked to fill in the survey did so, the five interviewees were selected from influential persons from the area of LT, covering the following geographical areas and areas of expertise:

- Western Europe, incl. one with links to the U.S. and one with high-level managerial and government advisory roles, Southern Europe, former Eastern Europe, and Israel
- General NLP with a perspective from the public government and administration, speech applications, excellent basic and experimental research in machine and deep learning for all areas of NLP, language resources and infrastructures, and lexical resources and applications of dictionaries

The respondents covered industry (two out of five), academia (two) and one represented both academia and industry. One of the respondents represented also the political and government level in one of the large EU countries.

The interviews were conducted between December 1 and 16, 2021, remotely over the Zoom platform.

Given the selection of the interviewees, the questions were selected in a more general fashion than those in the questionnaire, to spark more open discussion, including open exchange of ideas to topics not scheduled originally. The questions were sent to the interviewees in advance. However, they were asked not to specifically prepare any answers, and to think about them more from the research perspective, which is the typical profile of a META-NET respondent. Three of the respondents had previously responded to the survey.

Three general questions were used as triggers for discussion:

- What do you think the Language Technologies in broad use will be in 2030?
- What are the biggest obstacles in general to get there by 2030? In research, infrastructure, development, innovations, human resources, education, funding, ...?
- What can EU/EC do to remove these obstacles in time and help EU to thrive in AI/LT?

The interviews were recorded for internal purposes only, and will be deleted, together with all internal notes, after this deliverable has been finished.

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

[5] https://european-language-equality.eu, https://www.european-language-grid.eu, https://www.european-language-technology.eu as well as through the ELT social media accounts on Twitter and LinkedIn.

[6] These respondents represented 44 unique organizations from 24 countries across the EU and Israel, UK and Brazil.

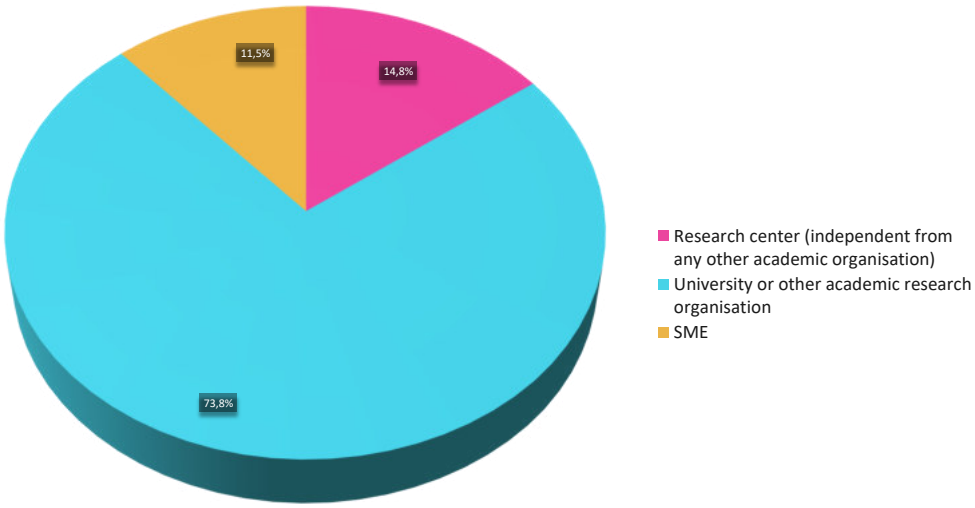



Figure 1: Type of organisation

# 3. Analysis of responses to survey questions

## 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 44 different organisations, out of which 89% are academic institutions or university-affiliated research centers, the rest are industrial research institutions and industry practitioners (Figure 1). The headquarters of these organisations are located in 28 different countries, with most responses from 1) Spain, 2) France, 3) Denmark, 4) Lithuania and 5) Romania. Detailed statistics of the breakdown of organisation types and countries are provided in Appendix B (Tables 2 and 3).

The respondents are mainly active in the following LT areas, by order of frequency: 1) Language Resources (production and aggregation) 2) Basic NLP technologies, 3) Translation technologies 4) Speech Technologies and 5) Text analytics and mining (Figure 2; see also Appendix B, Table 4).

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 Information and Communication Technologies, Digital Humanities, Arts, Culture and Other Services and Education.

## 3.2. Language coverage

A total of 33+ languages are currently supported by survey participants. The strongest support is first and foremost represented by English (49 out of 61 respondents), followed by Spanish, German and French. (For a more detailed overview, see also Appendix, Table 5). Lan-

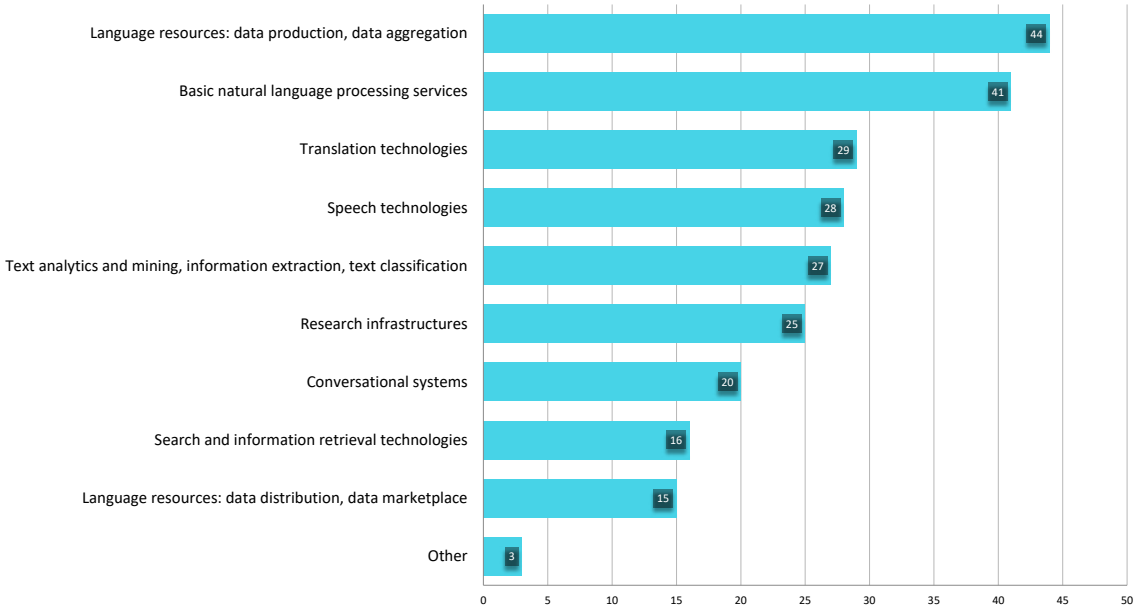



Figure 2: LT areas in which the respondents conduct research or develop tools and services

guages participants indicated as "Other" a surprisingly large variety of languages: Afrikaans, Akkadian, Amharic, Ancient Greek, Arabic, Armenian, Assyrian Neo-Aramaic, Bambara, Belarussian, Bhojpuri, Brasilian Portuguese, Breton, Buriat, Chinese, Church Slavic, Classical Armenian, Coptic, Cornish, Erzya, Faroese, Fenno-Ugric languages, Frisian, Georgian, Gothic, Hebrew, Hindi, Ibibio, Indonesian, Japanese, Karelian, Kazakh, Komi, Komi-Permyak, Korean, Latgalian, Livvi, Malay, Marathi, Mbyá Guaraní, Moksha, Nigerian Pidgin, Northern Kurdish, Northern Sami, Norwegian Bokmål, Norwegian Nynorsk, Old English, Old French, Old Norse, Old Russian, Persian, Russian, Sanskrit, Scottish Gaelic, Skolt Sami, Swedish Sign Language, Swiss German, Tagalog, Tamil, Telugu, Thai, Turkish, Uighur, Ukrainian, Upper Sorbian, Urdu, Vietnamese, Warlpiri, Wolof, Yoruba and Yue Chinese.

Not surprisingly, with most respondents covering English, a possible confirmation can be seen in the statement that the technological support for other European languages are characterised by a strong imbalance and many resources and technologies exist for English while other languages experience a lack of technological support. The results also still confirm the key results from the 2012 META-NET White Paper series "Europe's Languages in the Digital Age" that languages with more speakers had better support through Language Technology.[7]

Eighteen organisations indicated various languages that they plan to support in the short and medium term, including support for sign languages. For a more detailed overview, see also Appendix B, Table 6. The participants were asked to select the main factors that influence their decision to support additional languages. The respondents' top three drivers were research/scientific interest, available funding/investment and availability of language resources, with market interest and demand as distant fourth. (see also Appendix B, Table 7).

[7] http://www.meta-net.eu/whitepapers/key-results-and-cross-language-comparison

## 3.3. Evaluation of current situation

The survey asked the respondents to mark their perception of the challenges and obstacles (if any) in the present time and current situation as they know it from their institutions and their own research and development practice. There were 9 questions related to this, one of them open-ended (free text, see below).

The strongest support from all the 61 participants was, not surprisingly, that basic research in Language Technology is still needed (average 3.4 on a 1-4 scale) – 57 of the 61 agreed or strongly agreed with this proposition (and of the remaining four, two disagreed and two abstained from an answer); see Fig. 3.

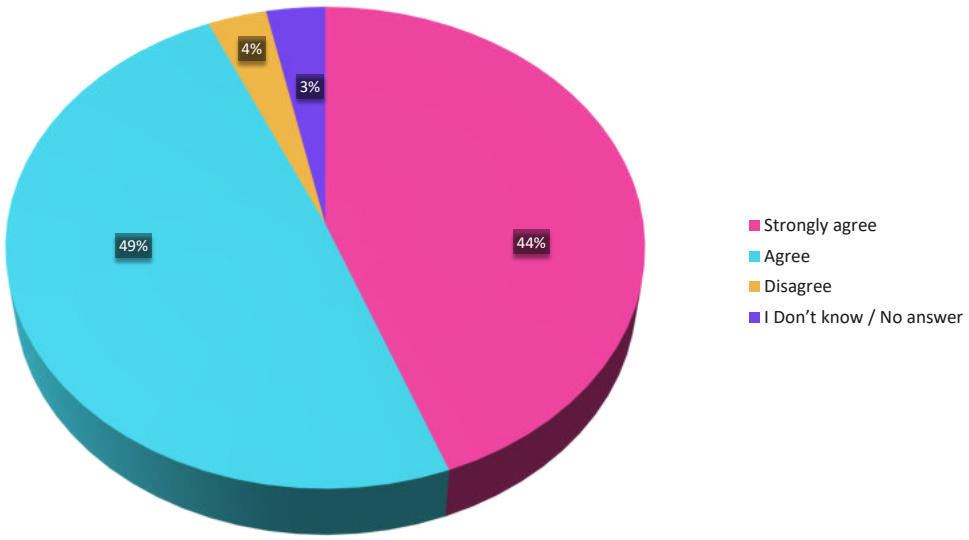



Figure 3: Need for basic research in Language Technology

Similarly strong support was found in the area of challenges caused by the competition with non-European companies and the related market disruption: 51 respondents agreed or strongly agreed, with 5 abstentions and 5 votes to the contrary (but no strong disagreement).

Somewhat smaller but still strong support (average 3.3 on the 1-4 scale) was given to lack of procurement, with 15 abstentions; this was the highest number of abstentions in this section of the survey, most probably due to the impossibility of some research organizations and universities to effectively compete in procurement calls, mostly due to co-funding and other conditions, otherwise suitable for companies.

Inadequate recognition of the importance of multilinguality was another obstacle widely perceived as a negative factor, with 46 in favor and 12 disagreeing, and only 3 abstaining (avg. of 3.2 on the 1-4 scale). While this obstacle might be seen as subjective, it is certainly not the case for the insufficient size of the markets to justify investment in LT for smaller languages, which was judged as an important obstacle in digital language equality (also avg. 3.2 on the 1-4 scale).

While some felt that the fragmentation of the LT industry in Europe is a problem, it was interesting that 11 respondents abstained – the second highest abstentions (after the procurement question); perhaps the question was not clear to them and the absentees wanted to avoid the perception that they are in favor of large monopolies in Europe, something that

is seen as a general obstacle on the global scene. A similar situation is sketched by the answers to the question about the lack of coordination and missing links between research, vendors, integrators and customers: while the average was somewhat lower than for the fragmentation question, slightly more respondents disagreed (8 vs. 5) that this is a problem.

Clearly, the least concern has been about the access to computing infrastructure and lack of talent. For the cost of access to computing infrastructure (HPC and other compute), only 7 out of the 61 respondents strongly agreed that this is a problem, while 15 disagreed (even if not strongly). This was an interesting result, since (anecdotally) a lot of complaints are floating among researchers with regard to the resources available to large multinational companies as opposed to universities and research centers, but apparently in practice it might not be the biggest problem of all. Finally, the least concern of all the nine questions was in terms of lack of talent: apparently, Europe is producing a lot of talent which is known to the respondents, but given that the respondents are mostly form research institutions, it might not correspond to what LT companies in Europe see as a serious problem – namely the difficulty of getting well-trained specialists in machine learning and LT (at a reasonable costs). We should also stress that even if this question was "numerically" of least concern, it was still above average (2.6 on the scale of 1-4, i. e. over the 2.5 average – see Fig. 4), and a similar future-oriented question about increased availability of personnel and incentives for researcher retention was judged by 58 out of the 61 respondents as moderately to very effective, i. e. positively.

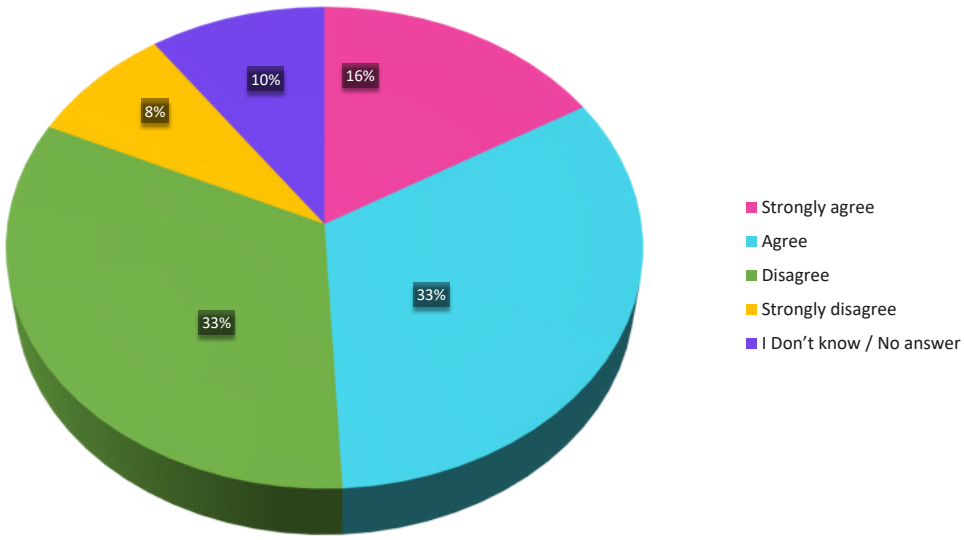



Figure 4: Need for retaining talent and avoiding brain drain

Thus as META-NET, we strongly support the push for establishing Europe-wide talent-retaining programs in AI and specifically in LT. Such support might be demonstrated by measures ranging from establishing a special European Research Council (ERC) panel on AI/LT, to support for LT-specific Horizon Europe calls. Both these instruments are invaluable for graduate students and postdocs in order to gain experience in international settings, get connections across Europe and across academia as well as industry, making it more likely they will stay in Europe afterwards.

The last question of the current situation section was a free-text (open-ended) question requesting additional views on the previous questions in topics (i. e. obstacles or challenges

for the LT field) not covered in the preceding (closed, scaled) questions. Six respondents grabbed the opportunity and responded, some in quite a length. We can summarize these responses as follows:

- half of the responses mentioned the "small" languages as being disadvantaged and the difficulties that companies as well as researchers have in tackling the problems of equal quality tools for those languages, from data to methods specifically tailored to low resourced languages.
- on a related note, two of the respondents stressed basic research as not being an obstacle, but a necessity; in particular, we need basic research to advance the methods along with data collection and annotation, and we need "LT-first" researchers educated secondarily in Machine Learning / Deep Learning, not the other way round.
- fragmentation of the landscape at all levels and/or too-small-scale efforts have also been mentioned, related to various areas of LT: in tackling low-resource languages, in language processing infrastructure and in education.

A detailed list and more exhaustive summary of all answers can be found in Appendix B (Table 8).

## 3.4. Predictions and visions for the future

This subsection describes the respondents' views with regard to the measures and instruments that are deemed effective and to the key challenges that a future ELE programme should address. More specifically, the participants were asked to rate, in their opinion, the effectiveness of selected policies and instruments to speed up the development and deployment of LT in Europe equally for all languages. Almost all of the suggested measures were considered effective.

- Initiate a large-scale and long-term funding programme for European LT development,
- Continuous investment in the research infrastructures that support LT,
- Investment in the development of new methodologies for transfer/adaptation of resources/technologies to other domains and languages, and
- Reinforce training and education initiatives, including undergraduate and masters programs and vocational training in LT

The support for the large-scale and long-term funding programme for European LT development was overwhelming, with an average support of 4.38 on a 1-5 scale (Fig. 5).

This result is not unexpected; it is in line with previous META-NET findings, as formulated in the Strategic Research Agenda and its updates, especially in Rehm (2017).

Perhaps more surprising is the even more convincing result regarding continuous investment in Research Infrastructures that support LT (Fig. 6). Out of the 61 respondents, 60 perceive the investment to Research Infrastructures as at least moderately effective, and 57 (still a whopping 93%) consider it either effective or very effective. While the project funding for Research Infrastructures comes from the EC, it is to be noted that Member States are contributing a large amount of stable funding to the ERICs, including those providing Language Resources. Adding ELRA, which is mostly self-sustained, this shows that this relatively recent combined investment into the data foundations for today's prevailing algorithms in Language Technology is already perceived as the most effective for the development and future success of Language Technology.

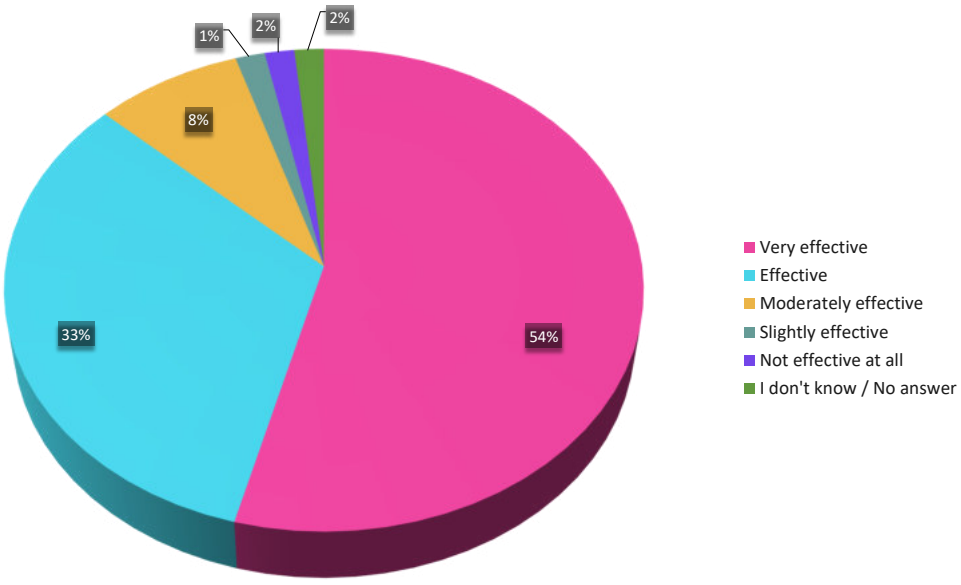



Figure 5: Support for a long-term, large-scale funding programme for Language Technology

Investment into new scientific methods and algorithms ranked third most effective: with three abstentions, no one perceives it as not effective, and only four respondents of the 61 marked it as only "slightly" effective. The remaining 54 (88%) consider this type of instrument at least moderately effective, with 77% as effective or very effective. The overall average on the 1-5 effectiveness scale was 4.14.

Similar average (4.13 on the 1-5 scale) was then reached in the area of education and training (including vocational, undergraduate and Masters programmes). More than 76% of respondents consider education as effective or very effective, and no one doubts effectiveness (with only one abstention). It should be stressed that this question aimed at (at most) undergraduate level, while graduate level is covered by the research programmes themselves due to high participation of graduate students and postdocs; funding to the lower levels should thus be considered as one of the priorities should a LT programme be established, which would be a new type of support traditionally confined to Member States (or regions) and their general educational support.

As has been mentioned in the previous section on current obstacles and challenges, retention of talent has been seen mostly as a possible target of new policies and instruments. It also reached an average of over 4 (more precisely, 4.05) on the 1-5 scale. Very high percentage of respondents, a whole 95%, believe that support for "increased availability of qualified personnel on LT and incentives for talent retention" would be at least moderately effective for this purpose (with 82% convinced that it is effective or very effective).

The next instrument, "Raise awareness of the benefits for companies, public bodies, and citizens of the availability of on-line services, contents and products in multiple languages" have been rated at the average score of 3.97, still a high score on the 1-5 scale. While this awareness will come naturally to companies through the emerging network of market survey providers and their own experience in localizing their products and offering the various aspects of their products in many languages, it is important to work on the awareness of public bodies and – to an extent – of citizens, even though they are aware of the problem first-hand (cf. the complaints on social media of, e. g. some of the voice assistants of the large

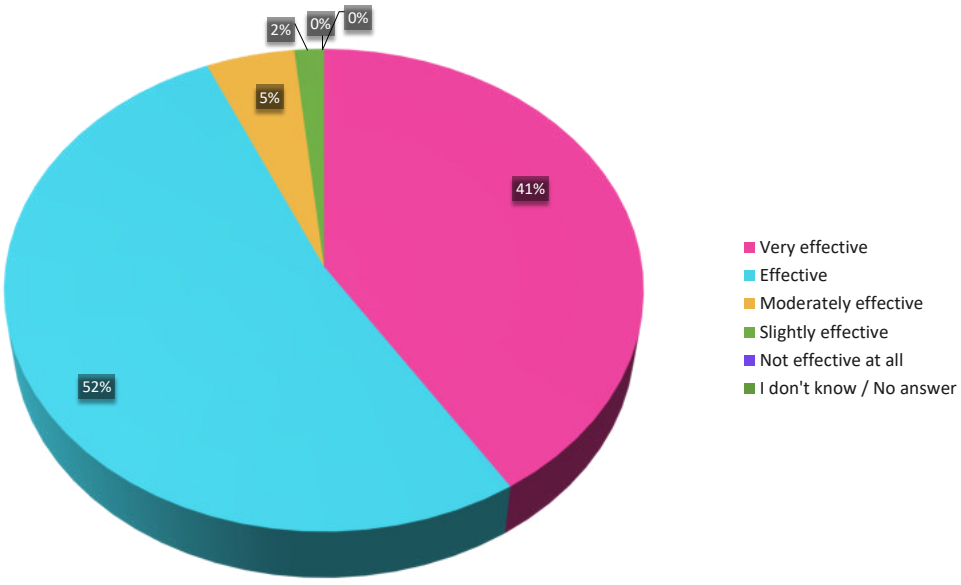



Figure 6: Support for continuous investment in Research Infrastructures that support LT

global companies not being available in their native tongue). For public administrations, we believe that now only campaigns, but also regulation might be a solution (similar to the AV Media directive which has been aimed at the hearing and vision impaired): we need a Single Market supporting directive that will require for public government and administration texts, on all levels, to be available in all EU languages (both official and those widely spoken in certain regions). There is no reasons not to: the previous efforts of the EC, primarily through the CEF-funded ELRC, have produced resources and the EC own efforts have led to the existence of eTranslation which now all public administrations can use. However, continued support is needed to still keep improving its quality (which mostly means to enlarge the available datasets) so that eventually no human intervention is needed in the process, e. g. publishing all web-based public administration information automatically in all the aforementioned languages. This might be a distant goal for the smaller and less-resourced languages, but there are some proofs already that this is indeed a possible and reachable goal (Popel et al., 2020). The regulation question (requirement to include multilanguage content, subtitles etc.) has supported this view, but perhaps not so strongly: the average score of that question was 3.84 on the 1-5 scale, with almost 15% abstaining or considering the regulation tool(s) as only slightly or not effective. This result perhaps reflects the fact that the question did not distinguish the private and public sector, and for some, overregulation of the private sector is a real danger. Still, the 85% support is convincing, and we believe that it would be higher if there were a separate question regarding the regulation of primarily the public sector in this area.

Even though most respondents in the META-NET survey come from academia, the investment instrument question (aimed at support for startups: “initiate investment instruments and accelerator programs targeting LT start-ups”) has also been met with over average enthusiasm (average score 3.88 on the 1-5 scale). With four abstentions, understandable in academia, and one exception (“Not effective”), all others (almost 92%) believe such support is at least moderately effective to support the uptake of LT.

Finally, the “Public procurement of innovative technology and pre-commercial public pro-

curement" question also received positive feedback even though the least from the set of policies/instrument questions (average score 3.81 on the 1-5 scale, still however in the positive side). Over 62% of respondents considered the procurement support effective of very effective.

In addition to the closed, scaled questions the participants have been presented with four open, free-text questions that we summarize here.

The first question was formulated as a request to describe what might have been missing in the previous closed questions ("Are there any other policies/instruments not listed in the previous questions, which in your opinion can be effective be in speeding up the development and deployment of LT in Europe equally for all languages?"). Ten respondents used the opportunity to express a view. These ten responses can be summarized as follows:

- Several of the responses focused on data availability, collection and annotation, including legal issues ([non-]availability of public data, further steps in copyright reform) and publication practices/rewards for creating data and/or making data openly available; this included repeating the demands to specifically focus on low resource / minority languages.
- The need for basic research has been mentioned again (in the context that no question from the set contained a specific reference to basic research).
- (The lack of) awareness of what LT actually is among citizens have been mentioned too, as well educational issues regarding multilinguality and the use of such technologies in school curricula.

The next two open ended questions asked for possible key elements of a long-term European Language Technology research, development and innovation programme: "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?" — the first focused on basic and applied research, and the second one on innovation and the LT industry. Since the respondents in some cases mixed the two, or provided an unrelated answer, here we summarise the responses according to the original focus of the questions:

In the basic and applied research area, the respondents' suggestions for the key challenges can be listed as follows:

- multilinguality and (same) quality for all languages, and (continued) collection of data
- focus on speech technology(ies)
- focus on language understanding (as a natural, more advanced followup to the current natural language processing technology(ies))
- interdisciplinarity in terms of multimodality and bridges to AI, human-computer interaction, vision, sensors (perception), knowledge as well as robotics
- provision of enough computing power for LT research
- infrastructural support and "data spaces" (for research)

In addition, recommendations for focusing on larger, top-tier teams with enough funding for long-term research and continuity, as well as for focusing on a smaller number of strategic but well-funded issues, appeared in this section, too.

In the innovation and LT industry area, the suggestions for key components of such a long-term programme can be listed as follows:

- ubiquitous natural language interfaces as a commodity rather than “nice-to-have” feature
- create European, large-scale alternatives to key text and speech technologies now offered by global players
- seamless human-like interactivity and behaviour, discourse interpretation
- business data spaces and data sharing
- high-performance applications (in terms of speed and quality), for all languages
- combination of support and regulations for public institutions to become truly multilingual

Two more interesting ideas have been proposed: first, to allow for non-consortial support (i. e. even for a single company), perhaps with a modified funding mechanism to pay “after the fact” (read: after at least some success). Another idea was to support strong national bodies or tight consortia, such as DFKI in Germany or the ADAPT centre in Ireland, which, by their national gravity, will also become important players and technology transfer hubs at the European level.

Finally, a catch-all open-ended question has been provided at the end of the survey. Seven respondents provided us with input. This input reiterated the requirement for free access (for research) to powerful computers and HPCs, the need to support low resource languages to catch up with major languages, creation of high-quality basic tools for all languages, the lack of multimodality (including in the questions), and a push for European Data Strategy in LT. One response asked for reinforcing the perceived connection between AI and LT. Finally, one response simply stated “Make it happen!”, which is indeed an appropriate message to close this section on predictions and visions for the future from the members of META-NET.

# 4. Analysis of interviews

As already described in Sect. 2.2, the five interviewees were selected as influential figures to represent various (sub)fields of Language Technology and various backgrounds, from basic research to industry and high-level public administration.

The questions asked were quite general, to elicit true opinions of those interviewed, and also give them some leeway to introduce topics and opinions not directly solicited in the more specific survey questions. We repeat here the questions already mentioned in Sect. 2.2:

- What do you think the Language Technologies in broad use will be in 2030?
- What are the biggest obstacles in general to get there by 2030? In research, infrastructure, development, innovations, human resources, education, funding, ...?
- What can EU/EC do to remove these obstacles in time and help EU to thrive in AI/LT?

## 4.1. Language Technologies in 2030

All of the interviewees cite the ubiquity of LT in products and services, using phrases like “LT everywhere,” “MT at our fingertips” or “language-transparent society”. In other words, any text, speech or sign in the digital sphere will be possible to be read, listened to or visually processed in any language, in an easy way, as we are already experiencing today in some browsers, but more user-friendly (“seamless”) and technically better integrated, faster and

especially more accurate in many (but perhaps not yet all) languages. Input will be possible in text or speech regardless of device, application or communication channel. Examples covered video and film dubbing, any media streams, social media communication, the web (as a matter of course, whatever it will be like in 2030), and in fact anything else.

LT products will be socially aware, free of cultural or social bias and respect the non-verbal clues (if available). More non-speech "signal" will be integrated to speech and text processing, such as video, images, and possibly other inputs (motion, gesture, haptics), on both the input and output side. In robotics, more communication with robots will be in natural language, in terms of solving exceptional situations (e.g., in natural disasters), (re)programming for different tasks, or simply information exchange. Knowledge systems (combining long-term knowledge bases and facts and immediate language-based communication) will be able to assist in decision making, learning, deduction systems, security and safety applications, in various domains (such as medicine or science).

On the technical side, a tremendous progress in basic and experimental research will bring qualitative advances in many areas. A lot less data will be needed for training systems by then-available machine learning methods, thus improving availability and quality of systems and applications in languages traditionally described as low-resourced. Large language models and transfer learning will contribute to digital language equality in applications, even if, on the data side, inequalities will still be present. The same will be true for dialects and accented speech. In addition, practical aspects of incremental learning will be solved, so that systems will be able to adapt, in non-trivial ways, to user's behavior, customs, habits, and experience.

The progress on all these fronts will enable new applications and devices, including very small devices all the way to embedded IoT or wearable devices. But miniaturization is not the only vision here; general AI has been cited as an enabler of applications like human assistants, multipurpose chatbots, and similar, mainly conversational systems, devices, and appliances.

## 4.2. Obstacles to get there

Comparing the current state to the predictions and visions expressed by the five interviewed experts, as described in the previous section, gives already some clear list of obstacles that stand in the way of implementing these visions in reality.

For example, the unavailability and inequality of data for all languages, all applications and all domains, not to mention novel multimodal applications, is a major obstacle. Alternatively, basic and experimental research will have to come up with methods and learning architectures, unknown as of yet, to overcome the highly unequal data availability, by focusing on, for example, annotation transfer, synthetic data and their proper use in machine learning, multilingual models preserving quality and coverage, few-shot or zero-shot learning. However, it has been mentioned that this is not either-or, at least for some time: investment in both data and methods is needed, since it is still unclear which approach or combination of approaches will achieve the most language-neutral (= language-equal) results.

Another obstacle mentioned in the interviews was about the combination of long-term "encyclopedic" knowledge with a short-term, learned "knowledge representation", as in cooperative dialogues, serving a collaborative purpose (such as solving a particular task between a human and machine). In this respect, progress in reinforcement-based learning, novel dialogue management strategies and situation-aware, natural language generation has been mentioned. In addition, knowledge databases (in the broad sense, i. e. including ontologies, semantic lexicons, terminology databases) in both human- and computer-understandable form are necessary to be extended in many languages and domains (possibly with the help of automatic processes, but in some critical domains, e. g. health, verified by humans).

Perhaps related to the previous problem, namely of situational dialogues, is the issue of personal data protection. Given the current regulations, especially the GDPR (moreover, transposed to member states' legal systems in very different ways), it is hard to record data in many situations of interest to the NLP/AI community. It gets even harder when it comes to situations which are in a medical environment or when minors are involved – which are exactly the situations which might have very interesting business applications with high positive societal impact.

The reference to GDPR also brought into focus other legal obstacles relevant for the field of LT (and AI as well), such as copyright and proper handling of IP rights. Unfortunately, even the Text and Data Mining (TDM) exception as defined in the 2019 Copyright Directive of the EC is not a solution that could help put Europe on the same footing as many other countries in the world. First, it is still weaker than, e. g. the U.S. Fair Use approach, and second, it has (so far) been accepted only by a minority of EU countries, despite the fact that the deadline for full transposition has passed mid-2021. Thus a concerted push and effort will be needed, when the current Copyright Directive from 2019 is to undergo revision at the EC level in 2024, to argue unanimously for a much stronger version in the TDM area, especially for businesses, based on international practice and taking into account the technologies available at the time. At the same time, the business-developed IPR must be adequately protected: smaller companies must not fear loss of IPR when developing LT applications. On the industrial side, a concern has been voiced if the very broad EU's AI Regulation and AI Act now under discussion is a hindrance to LT development (and even academic research, despite the current wording excluding research from it). A more positive view of the AI Regulation has been voiced by research representatives, but both sides agreed that the development of the AI Regulation should be very closely monitored, due to the usual "devil in the details".

On the technology transfer and academia-industry cooperation, various opinions have been voiced (see also below regarding the recommendations and ideas for public support and where it can be most efficient). The current situation has been, however, in general seen as a hindrance to faster adoption of language technology in industry and private and public services. The RIA calls from the H2020 programme have been praised for advancing the field, including contacts to companies active in the field; concern has been voiced whether similar calls will be seen in the in the 2023-4 and later Workprogrammes, given that the current Workprogramme is more development-oriented. Also, the adoption of commercial, Europe-made language technology by public institutions in Europe have been seen as weak, citing cases where actually equal-quality non-European solutions have been acquired by public institutions in procurement processes.

Some views of the interviewees confirmed the survey results that talent is plentiful in Europe, including third-country graduate students being interested in LT-related graduate education offered by European universities (such as the LCT programme, which was very successful in this respect). At the same time, some still see more space for more graduates in this area. However, there is still a high number of departures to non-EU regions, or to non-EU companies. While it is clear this is not an inherently LT question, but rather a more general business environment question, it is a very sensitive issue especially in the LT (and AI) area due to several reasons: the language equality we are after in this project, its potential dual use, and most importantly, the potential of these technologies to contribute to both economical and social well/being in Europe.

## 4.3. How can LT be supported in EU to excel

Starting from the technology and data side, in all cases, support to data acquisition and development has been voiced as one of the areas where public support is still heavily needed. ELRC has been mentioned as a good example, but concern has been voiced if this is suffi-

cient, especially since ELRC is focused on public (PSI) data and the funding available can support data verification and delivery, but not really larger-scale data acquisition, cleaning, annotation or other necessary processing. Especially today, when the use of so-called Large Language Models seems to be an important prerequisite to interesting LT applications, the support for petabyte-scale language resources (which is beyond the reach of any individual institution, both space- and compute-wise) is absolutely necessary.

Related to language resources, one interviewee would like to see public support for the "database type", i. e. ontologies, lexicons, and in general any knowledge sources linked to language, in a human-understandable and computer-readable standard form. The goal of general AI (including human language communication) has been cited, with the assumption that without such resources general AI is not attainable.

Before coming to methods, models and algorithms, it must be noted that the availability of HPC resources has been mentioned, citing the current structure of HPC centers as not meeting the needs of machine learning, which is the prevailing technology in both LT and AI. This is a problem of academic institutions not only in Europe. It was recommended that AI and LT representatives approach the EuroHPC and similar programmes and make an effort to have the right hardware and software support and flexible access to such machine-learning-capable facilities across Europe, for European institutions. This is related to the large data acquisition efforts and basic text analysis and model building mentioned above.

On the methods and algorithms front, support for both basic and experimental research has been voiced in multiple cases. It has been argued that data collection efforts are still worthwhile, specifically considering digital language equality, as there will never be an equal amount of data available for all languages. In addition, further research is necessary on the algorithmic front.

Some of the interviewees voiced support for the business development of LT applications and services, ranging from startups/spinoffs to larger, successful companies facing the well-known "death valley" on the trajectory to becoming a large, global player. It has been cited that often the successful companies that made the leap from startup or even university spinoff to a successful, profitable company with revenue in the tens of millions of EUR (or more) are acquired by non-European companies or non-European holdings. These suggestions did not go as far as suggesting "public venture capital", but other possibilities were mentioned, such as acquiring European-based LT/AI services and software solutions by public bodies, in order to help a company grow over the "death valley" period.

# 5. Conclusions

## 5.1. Starting point and background

The survey and interviews followed a series of White Papers, the SRA and a set of followup publications (Rehm and Uszkoreit, 2012, 2013; Rehm et al., 2014; Rehm, 2015; Rehm et al., 2016; Rehm, 2016, 2017, 2018) and a survey on Language Technology for Multilingual Europe conducted in 2018 (Rehm and Hegele, 2018). Especially the latter reiterated that the biggest challenge for the European LT was the threat of digital extinction of languages with smaller numbers of speakers. The conclusion is drawn that, with the dominance of the English language, researchers are often given little incentive to focus on smaller or minority languages. For instance, when it comes to publishing there is a strong bias towards incorporating results for English. The same appeared to be true for LT funding available for projects focusing on languages other than English. According to the 2018 survey, as a consequence of the above-mentioned bias, many participants commented on the lack of available data resources for smaller languages which are especially needed to further improve the quality

of current LT systems. There was also a very strong agreement for the need to support European basic research in language-centric AI in a variety of research areas, where research and development fall behind USA and Asia.

## 5.2. Main result

The present survey and interviews, performed by META-NET among its members and several prominent figures from the field of LT, largely confirmed the previous results. Despite the fact that the technology has made tremendous advances (in terms of accuracy and usability as well as actual use), the inequalities of EU official and regional/minority languages alike remain.

Details of the surveys and their quantitative and qualitative results can be found in Section 3 and the contents, focus points, views, recommendations and conclusions of the interviews in Section 4.

## 5.3. Current situation, challenges and obstacles

In both the survey answers and the interviews, the following points have strong support and/or have been mentioned multiple times as a current challenge/obstacle (in descending order of support level):

- insufficient support for basic and experimental (oriented, applied) research, including appropriate data collection efforts
- competition with global companies and the related market disruption caused by the fact that LT is often not their core business
- lack of procurement calls, possibly with specific priority to EU companies
- lack of recognition of the importance of multilinguality in all areas of business and public life

It should also mentioned that lack of talent has not been considered a severe problem, but concerns about the **loss of that talent** have been voiced from both academia and industry, in some cases citing not only loss to global companies, but also to non-LT fields of AI due to the common machine learning (ML) skills that students and young researchers now obtain when studying LT.

Fragmentation of the LT landscape and environment has also been mentioned, but perhaps not so strongly as at other occasions and in earlier surveys. It has been mentioned that fragmentation should not be confused with free competition in ideas, education as well as research.

## 5.4. Predictions and visions for the future

Almost all the measures mentioned in the survey have been strongly supported (again, in descending order support):

- initiate a large-scale and long-term funding programme for European LT development
- continuous support (with more EU-funding) for research infrastructures that support LT

- support for the development of methodologies and technologies for achieving language equality, i.e., for transfer/adaptation of resources/technologies to other languages and domains
- reinforce training and education support at all levels, including undergraduate (bachelor) and masters programmes as well as vocational training in LT

The support for the first point above was overwhelming and reached the highest average of all survey questions (4.38 on a 1-5 scale, Fig. 5).

In the open-ended questions and in the interviews, **basic and experimental research** has been mentioned again as an efficient tool to overcome some of the inequalities among languages and/or domains of use. Some of these responses reiterated that **data availability** is crucial, stressing that the EU environment, incl. the legal framework, is not yet as open to text, audio and video data use as in other parts of the world.

In terms of what a long-term focus should be, the participants mentioned data (again), **speech and "language understanding"** in general, interdisciplinarity, **support for infrastructures** and access to **computing power**.

The answers to open-ended questions featured several additional suggestions in the area of innovation, future applications and suggestions for EU support. While details are provided in the respective section, let us stress a few: create a EU alternative for the basic text and speech technologies now owned by large multinational companies; use a mix of regulation and support to make public administrations and institutions fully multilingual; special programmes or rules for national, single-entity projects funded by the EU as "points of attraction" to create critical-mass institutions bringing in talent and companies; access to large HPCs with appropriate computing hardware and structure.

## 5.5. Additional results: the interviews

In the META-NET case, the survey results have been deemed representative enough of the constituency size, and thus the five interviews served as additional source of information, based on three broad questions: (1) What LT will look like in 2030, (2) How to get there, overcoming the present obstacles and (3) What the EU/EC can do for it.

The answers were broad and understandably focused on the area of expertise of the individual interviewees. We bring here, in our view, the most important ones:

1. LT in 2030
   - ubiquitous LT in products and services
   - near-full "language understanding"
   - socially aware, seamless, high-accuracy LT
   - combining knowledge and LT
   - progress in technology (less data, multilingual models, etc.)
2. How to get there over current obstacles
   - more data, but at the same time, more basic and experimental research to achieve methodological and algorithmic advances that can work with less data
   - basic and experimental research on algorithmic progress in combining data sources (databases, ontologies, LT data, streamed data, multimodal data)
   - broader focus on conversational and situational dialogue-based systems, both textual and spoken
   - education in LT

3. EU/EC support to achieve the aforementioned goals
    - continued long-term, regular and uninterrupted support for research and data collection, including large open data and knowledge-like resources connected to language use
    - support for the right HPC infrastructure availability for research
    - changes in IPR legal basis and proper regulatory ecosystem to foster broader use of data on one side, and true multilinguality in products and services on the other
    - on the innovation side, apart from startup/spinoff support, also support of an appropriate environment that allows for SMEs to grow

Overall, the results confirmed previous findings especially with respect to language (in)equality. The results confirmed the need for continuous public funding support, with some interesting new ideas and suggestions on how such support for research, SMEs, and education can be instantiated. In addition, continued legal obstacles have been mentioned, perhaps even more intensely than in previous surveys, and suggestions on how the EC/EU can make a positive impact have been put forward.

## References

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.

Martin Popel, Marketa Tomkova, Jakub Tomek, Łukasz Kaiser, Jakob Uszkoreit, Ondřej Bojar, and Zdeněk Žabokrtský. Transforming machine translation: a deep learning system reaches news translation quality comparable to human professionals. *Nature Communications*, 11(4381):1–15, 2020. ISSN 2041-1723.

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. The META-NET Strategic Research Agenda for Language Technology in Europe. In Georg Rehm, Felix Sasaki, Daniel Stein, and Andreas Witt, editors, *Language Technologies for a Multilingual Europe*, Translation and Multilingual Natural Language Processing, pages 19–41. Language Science Press, Berlin, June 2018. Reissue of the special volume 3(1) of Translation: Computation, Corpora, Cognition (2013), which is no longer available. More information on the Language Science Press website.

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.

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. In Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the 9th Language Resources and Evaluation Conference (LREC 2014)*, pages 1517–1524, Reykjavik, Iceland, May 2014. 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, 2016. 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).

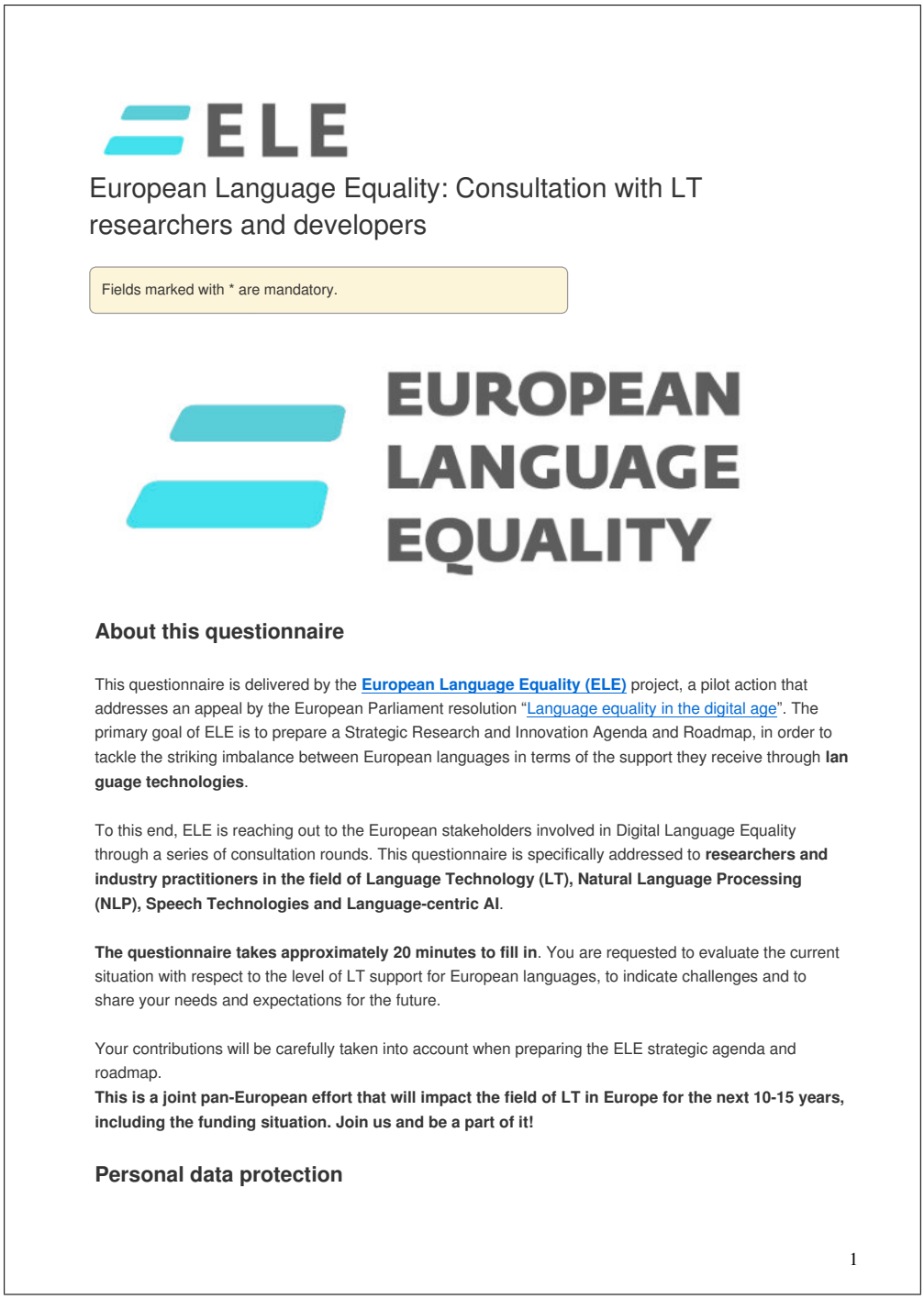
Georg Rehm, Katrin Marheinecke, Stefanie Hegele, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Khalid Choukri, Andrejs Vasiļjevs, Gerhard Backfried, Christoph Prinz, José Manuel Gómez Pérez, Luc Meertens, Paul Lukowicz, Josef van Genabith, Andrea Lösch, Philipp Slusallek, Morten Irgens, Patrick Gatellier, 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 7 to 15 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 7: 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.

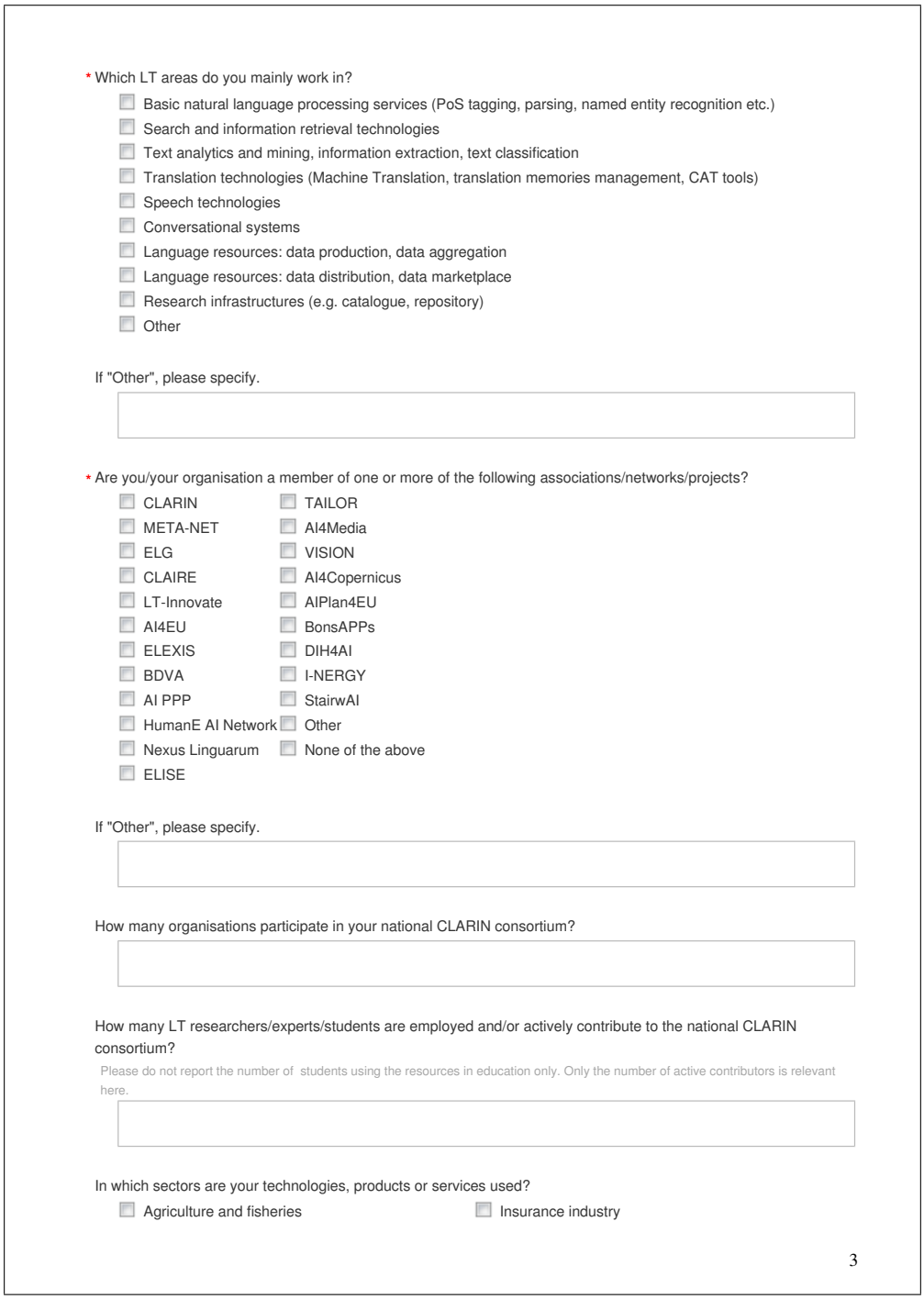
* Where is your organisation's headquarters based?

| | | |
|---|---|---|
| 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 8: 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?

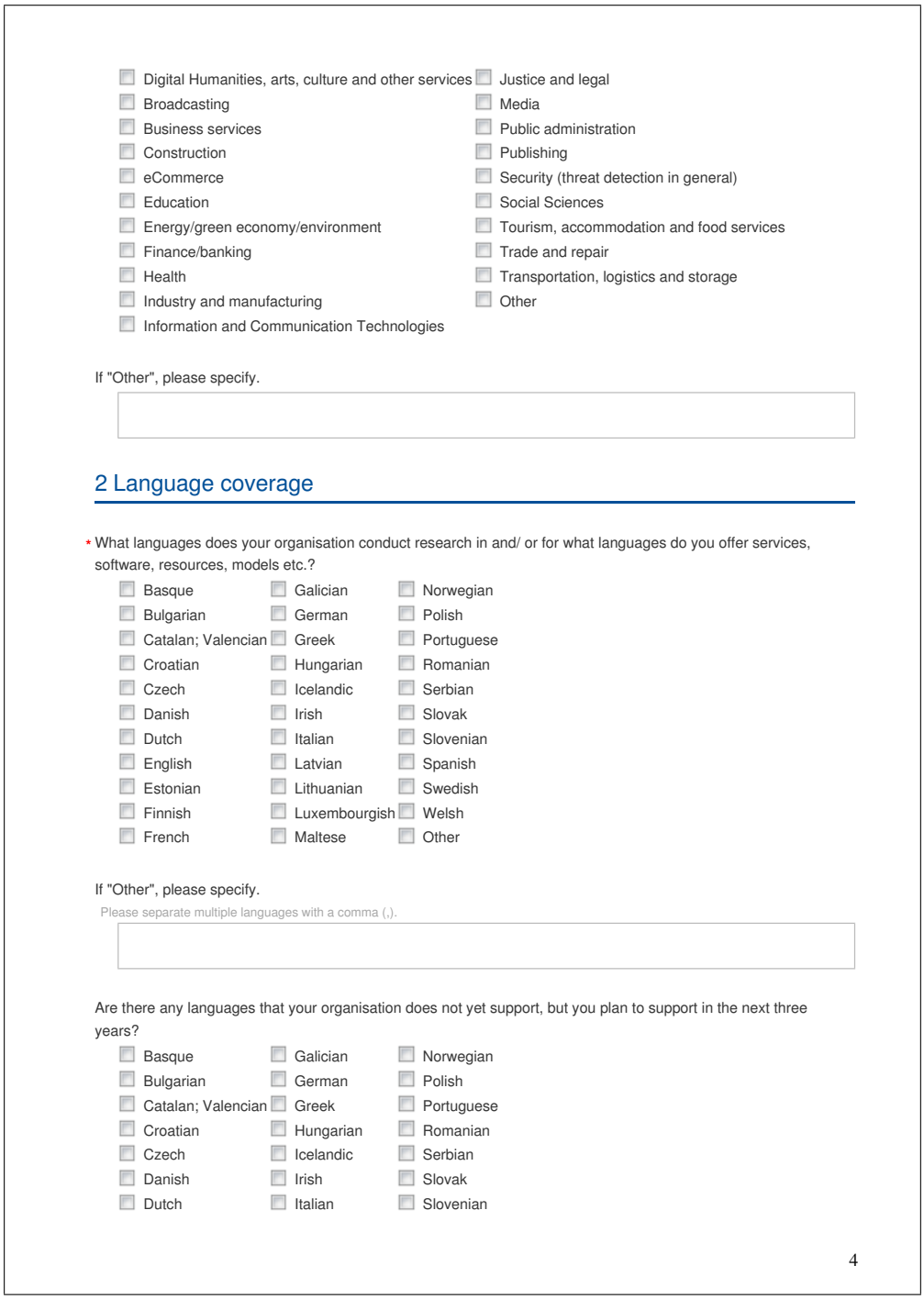
Please do not report the number of students using the resources in education only. Only the number of active contributors is relevant here.

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

- [ ] Agriculture and fisheries
- [ ] Insurance industry

3

Figure 9: 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.

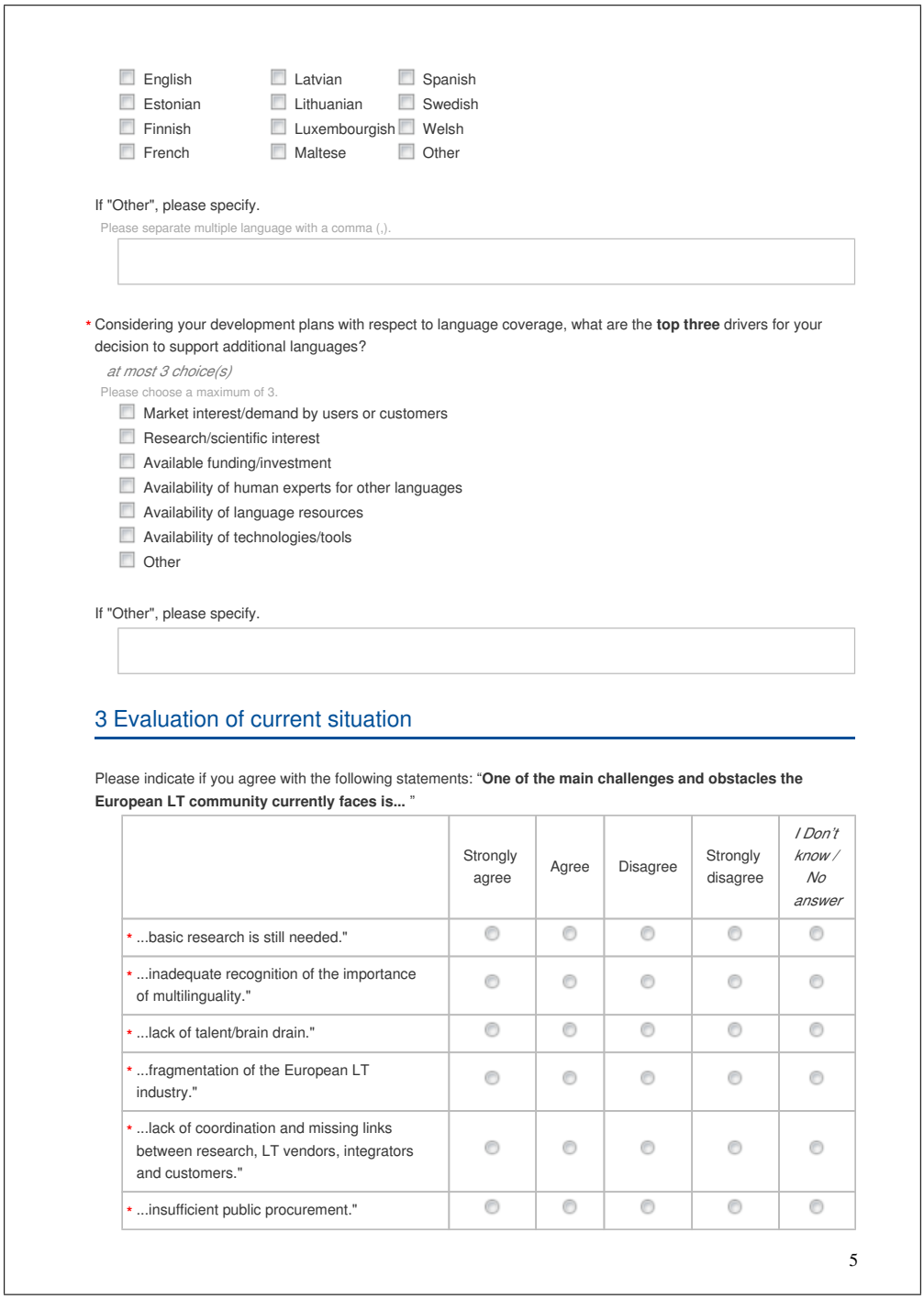
Please separate multiple languages with a comma (,).

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 10: 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.

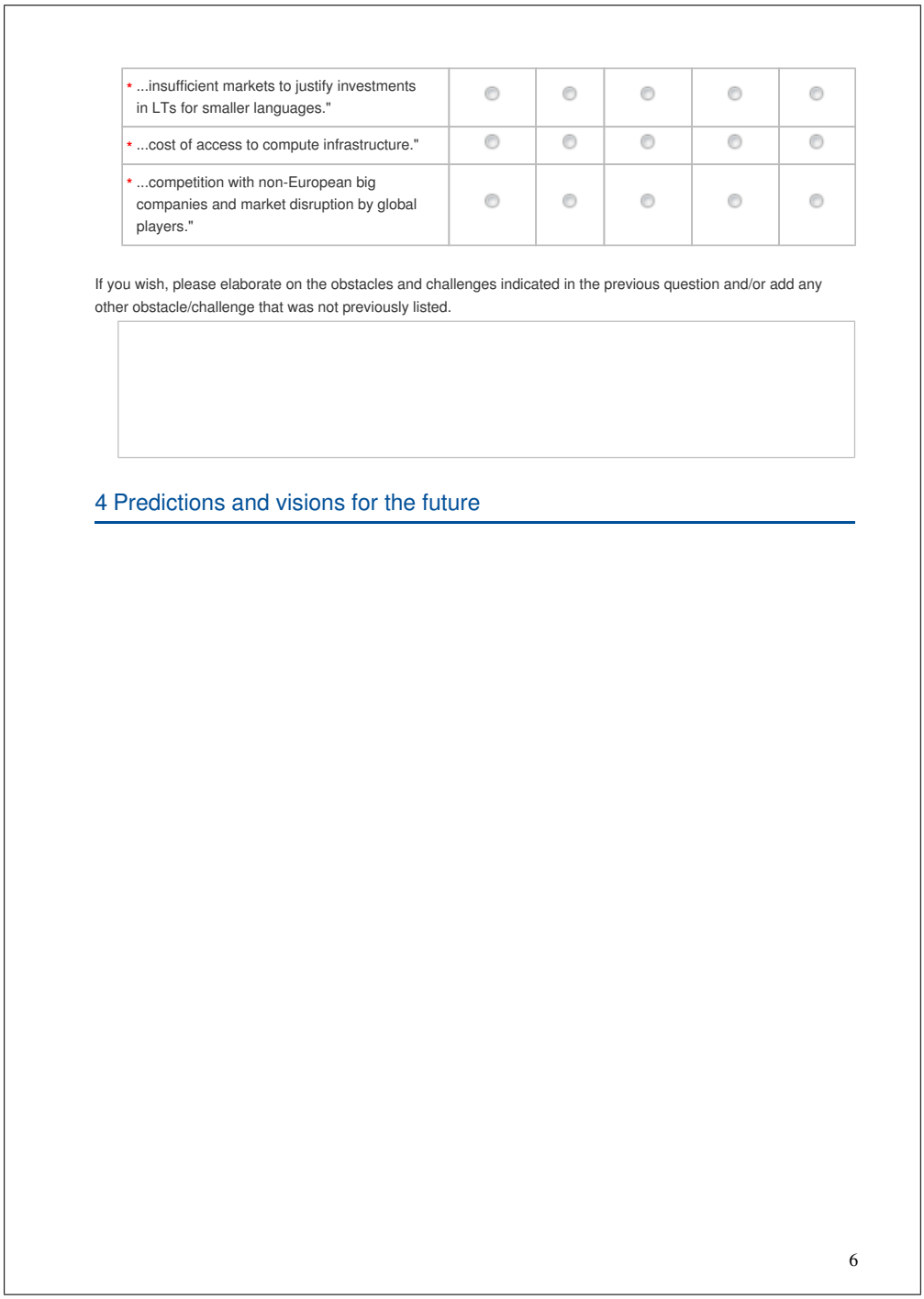
## 3 Evaluation of current situation

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



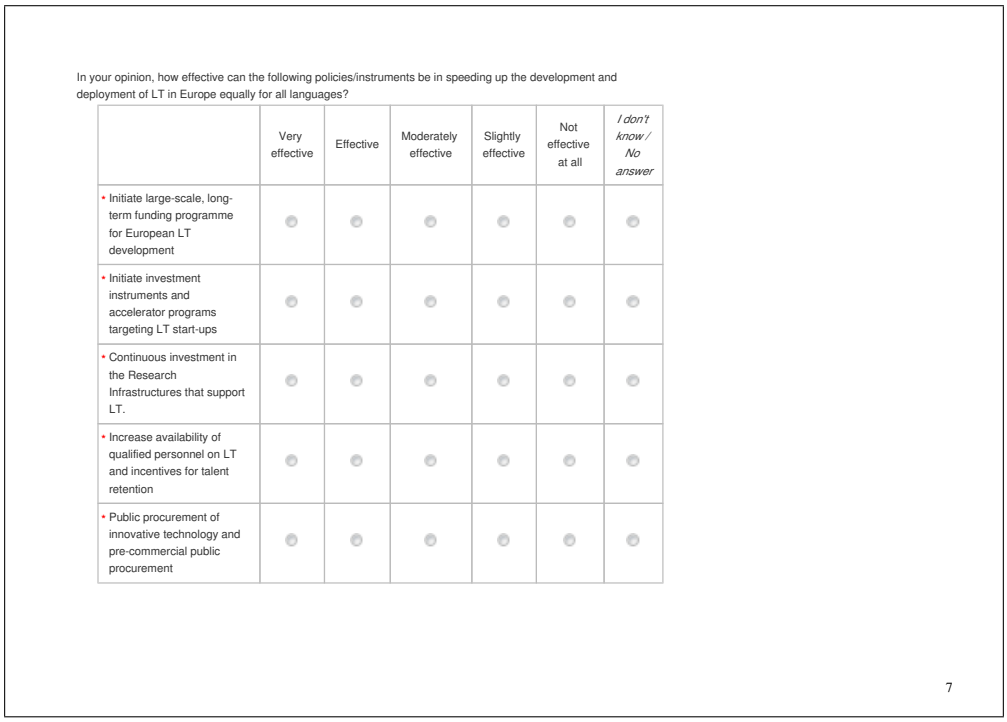
| | | | | | |
|---|---|---|---|---|---|
| * ...insufficient markets to justify investments in LTs for smaller languages." | | | | | |
| * ...cost of access to compute infrastructure." | | | | | |
| * ...competition with non-European big companies and market disruption by global players." | | | | | |

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

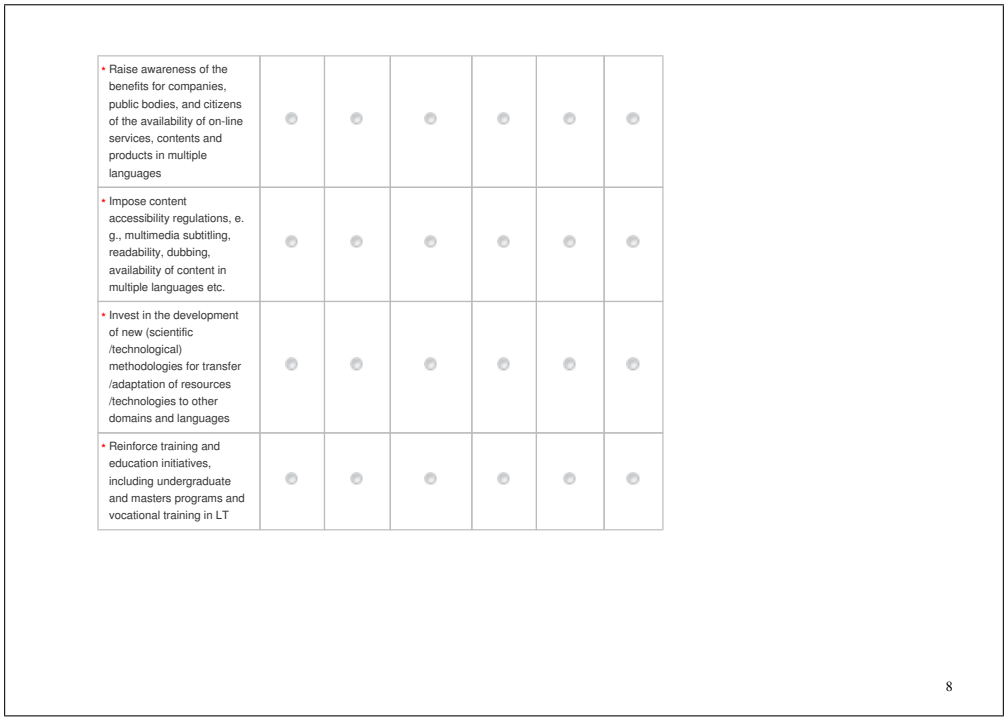


In your opinion, how effective can the following policies/instruments be in speeding up the development and deployment of LT in Europe equally for all languages?

| | 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

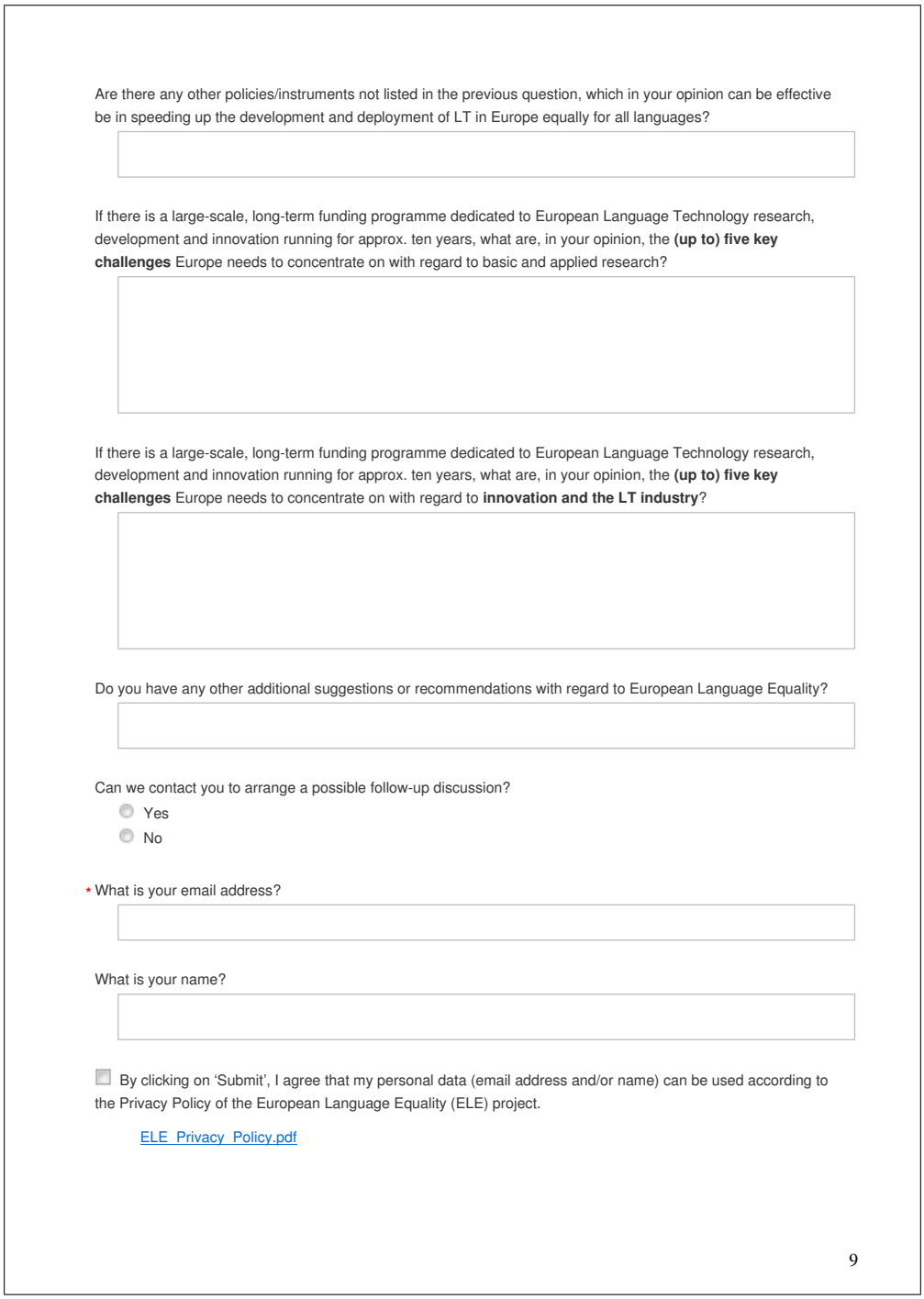
Figure 13: Full survey as published (page 7/9)



| | | | | | | |
|---|---|---|---|---|---|---|
| * 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 14: 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 15: Full survey as published (page 9/9)

# B. Additional tables and graphs

| Type of organisation | Answers | (Perc.) |
|---|---|---|
| Research center (independent) | 9 | 15% |
| University or academic research | 45 | 74% |
| Large enterprise | 0 | 0% |
| SME | 7 | 11% |
| Other | 0 | 0% |
| Total | 61 | |

Table 2: Breakdown of answers to “Which of the following best describes the type of organisation you work for?” (mandatory closed question)

| Country | Respondents | (Perc.) |
|---|---|---|
| Belgium | 1 | 2% |
| Bulgaria | 2 | 3% |
| Croatia | 2 | 3% |
| Czechia | 1 | 2% |
| Denmark | 4 | 7% |
| Estonia | 2 | 3% |
| Finland | 2 | 3% |
| France | 5 | 8% |
| Germany | 3 | 5% |
| Greece | 1 | 2% |
| Hungary | 1 | 2% |
| Iceland | 1 | 2% |
| Ireland | 2 | 3% |
| Italy | 3 | 5% |
| Latvia | 3 | 5% |
| Lithuania | 4 | 7% |
| Malta | 2 | 3% |
| Norway | 1 | 2% |
| Poland | 1 | 2% |
| Portugal | 1 | 2% |
| Romania | 4 | 7% |
| Slovak Republic | 1 | 2% |
| Spain | 7 | 11% |
| Sweden | 1 | 2% |
| Other | 6 | 10% |
| Total | 61 | |

Table 3: Breakdown of answers to "Where is your organisation's headquarter based?" (mandatory closed question, plus "if other" as optional open-ended question)

| | |
|---|---|
| Basic natural language processing services (PoS tagging, parsing, named entity recognition etc.) | 41 |
| Search and information retrieval technologies | 16 |
| Text analytics and mining, information extraction, text classification | 27 |
| Translation technologies (Machine Translation, translation memories management, CAT tools) | 29 |
| Speech technologies | 28 |
| Conversational systems | 20 |
| Language resources: data production, data aggregation | 44 |
| Language resources: data distribution, data marketplace | 15 |
| Research infrastructures (e.g. catalogue, repository) | 25 |
| Other | 3 |

Table 4: Full list of answers to "Which LT areas do you mainly work in? If "Other", please specify." (optional open-ended question)

| Language | Mentions (count) |
|---|---|
| Basque | 4 |
| Galician | 7 |
| Norwegian | 6 |
| Bulgarian | 8 |
| German | 19 |
| Polish | 9 |
| Catalan; Valencian | 7 |
| Greek | 8 |
| Portuguese | 14 |
| Croatian | 8 |
| Hungarian | 8 |
| Romanian | 13 |
| Czech | 9 |
| Icelandic | 8 |
| Serbian | 6 |
| Danish | 11 |
| Irish | 5 |
| Slovak | 9 |
| Dutch | 10 |
| Italian | 13 |
| Slovenian | 10 |
| English | 49 |
| Latvian | 8 |
| Spanish | 22 |
| Estonian | 10 |
| Lithuanian | 10 |
| Swedish | 9 |
| Finnish | 7 |
| Luxembourgish | 1 |
| Welsh | 4 |
| French | 19 |
| Maltese | 6 |
| Other | 13 |

Table 5: Breakdown of answers to “What languages does your organisation conduct research in and/ or for what languages do you offer services, software, resources, models etc.?” (mandatory multiple choice question, plus “if other” as optional open-ended question)

| Language | Mentions (count) |
|---|---|
| Basque | 2 |
| Galician | 1 |
| Norwegian | 2 |
| Bulgarian | 1 |
| German | 4 |
| Polish | 1 |
| Catalan; Valencian | 1 |
| Greek | 2 |
| Portuguese | 2 |
| Croatian | 1 |
| Hungarian | 2 |
| Romanian | 1 |
| Czech | 1 |
| Icelandic | 1 |
| Serbian | 1 |
| Danish | 2 |
| Irish | 2 |
| Slovak | 1 |
| Dutch | 2 |
| Italian | 2 |
| Slovenian | 1 |
| English | 2 |
| Latvian | 1 |
| Spanish | 1 |
| Estonian | 1 |
| Lithuanian | 2 |
| Swedish | 3 |
| Finnish | 1 |
| Luxembourgish | 1 |
| Welsh | 1 |
| French | 3 |
| Maltese | 1 |
| Other | 8 |

Table 6: Breakdown of answers to “Are there any languages that your organisation does not yet support, but you plan to support in the next three years?” (mandatory multiple choice question, plus “if other” as optional open-ended question)

| Drivers | Mentions (count) |
|---|---|
| Market interest/demand by users or customers | 22 |
| Research/scientific interest | 48 |
| Available funding/investment | 31 |
| Availability of human experts for other languages | 13 |
| Availability of language resources | 31 |
| Availability of technologies/tools | 10 |
| Other | 15 |

Table 7: Mentions of the top drivers for the decision to support additional languages

| | Strongly agree | Agree | Disagree | Strongly disagree | I don't know / No answer |
|---|---|---|---|---|---|
| basic research is still needed | 27 | 30 | 2 | 0 | 2 |
| inadequate recognition of the importance of multilinguality | 21 | 25 | 12 | 0 | 3 |
| lack of talent/brain drain | 10 | 20 | 20 | 5 | 6 |
| fragmentation of the European LT industry | 16 | 28 | 5 | 1 | 11 |
| lack of coordination and missing links between research, LT vendors, integrators and customers | 21 | 22 | 8 | 2 | 8 |
| insufficient public procurement | 21 | 18 | 5 | 2 | 15 |
| insufficient markets to justify investments in LTs for smaller languages | 22 | 24 | 11 | 0 | 4 |
| cost of access to compute infrastructure | 7 | 31 | 15 | 0 | 8 |
| competition with non-European big companies and market disruption by global players | 27 | 24 | 5 | 0 | 5 |

Table 8: Answers to the question: "Please indicate if you agree with the following statements: "One of the main challenges and obstacles the European LT community currently faces is…" (mandatory closed question, answers provided on a four-point scale, plus "I don't know/No answer")

| | Very effective | Effective | Moderately effective | Slightly effective | Not effective at all | I don't know / No answer |
|---|---|---|---|---|---|---|
| Initiate long-scale, long-term funding programme for European LT development | 33 | 20 | 5 | 1 | 1 | 1 |
| Initiate investment instruments and accelerator programs targeting LT start-ups | 13 | 26 | 17 | 0 | 1 | 4 |
| Continuous investment in the Research Infrastructures that support LT | 25 | 32 | 3 | 1 | 0 | 0 |
| Increase availability of qualified personnel on LT and incentives for talent retention | 16 | 34 | 8 | 1 | 1 | 1 |
| Public procurement of innovative technology and pre-commercial public procurement | 14 | 24 | 9 | 6 | 1 | 7 |
| Raise awareness of the benefits for companies, public bodies, and citizens of the availability of online services, contents and products in multiple languages | 20 | 21 | 15 | 2 | 1 | 2 |
| Impose content accessibility regulations, e.g., multimedia subtitling, readability, dubbing, availability of content in multiple languages etc. | 19 | 16 | 17 | 4 | 1 | 4 |
| Invest in the development of new (scientific/technological) methodologies for transfer / adaptation of resources / technologies to other domains and languages | 23 | 24 | 7 | 4 | 0 | 3 |
| Reinforce training and education initiatives, including undergraduate and masters programs and vocational training in LT | 23 | 23 | 13 | 1 | 0 | 1 |

Table 9: Answers to the question: "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?" (mandatory closed question, answers provided on a five-point scale, plus "I don't know/No answer")