EUROPEAN LANGUAGE EQUALITY

D2.9

Report from ELEN

Author	Davyth Hicks
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	D2.9
Deliverable title	Report from ELEN
Type	Report
Number of pages	41
Status and version	Final
Dissemination level	Public
Date of delivery	Contractual: 28-02-2022 – Actual: 28-02-2022
Work package	WP2: European Language Equality – The Future Situation in 2030
Task	Task 2.2 The perspective of European LT users and consumers
Author	Davyth Hicks
Reviewers	Teresa Lynn, Tereza Vojtěchová
EC project officers	Susan Fraser, Miklos Druskoczi
Contact	European Language Equality (ELE) ADAPT Centre, Dublin City University Glasnevin, Dublin 9, Ireland Prof. Dr. Andy Way – andy.way@adaptcentre.ie European Language Equality (ELE) DFKI GmbH Alt-Moabit 91c, 10559 Berlin, Germany Prof. Dr. Georg Rehm – georg.rehm@dfki.de http://www.european-language-equality.eu

Consortium

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

Contents

List of Figures

List of Tables

List of Acronyms

DL	Deep Learning
DLE	Digital Language Equality

DLDP	Digital Language Diversity Project
ELE	European Language Equality *(this project)*
ELE Programme	European Language Equality Programme *(the long-term, large-scale funding programme specified by the ELE project)*
ELEN	European Language Equality Network
EU	European Union
INI	Own-initiative Procedure
LT	Language Technology/Technologies
NGO	Non-governmental Organisation
RML	Regional and Minority Language
SRIA	Strategic Research and Innovation Agenda
STOA	Science and Technology Options Assessment
UK	United Kingdom
UN	United Nations

Abstract

The European Language Equality Network is the international civil society organisation for the protection, promotion and well-being of European minoritised, territorial and endangered languages (RMLs). ELEN comprises a broad range of member organisations (168) representing 47 languages in 24 European States. Member organisations include large cultural organisations such as Omnium Cultural in Catalonia, the Diwan Breton-medium school network in Brittany, and language activist organisations such as Cymdeithas Yr Iaith Gymraeg in Wales, and large umbrella organisations such as Kontseilua in the Basque Country, as well as several universities and research institutes. This report therefore represents the civil society and language community view of the current state of development for our languages with regard to Language Technology (LT). ELEN members work for the revitalisation, recovery and protection of their languages in all sociolinguistic domains of which digital development forms an increasingly important area. The views from ELEN members and followers reflect those of end-users of LT and not necessarily those of LT experts.

1. Introduction

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

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

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

Although the methodology and instruments utilised have been common to all ELE consortium members, this report covers and analyses the subset of responses of stakeholders contacted by the European Language Equality Network (ELEN).

1.1. About ELEN

ELEN is the international NGO for the protection, promotion and well-being of European territorial or minoritised languages (hereafter RMLs). ELEN brings together nearly all European RML civil society organisations giving the organisation a powerful mandate to represent the 50 million speakers of RMLs in Europe.

ELEN's work focuses on two main pillars. Firstly, advocacy work for our languages principally at the UN, Council of Europe and the EU, as well as with Member States, and in particular the development of language rights. Secondly, language-related project work to revitalise and recover our languages, in particular with the EU's Erasmus Plus and Horizon programmes.

From ELEN's perspective it has been clear from the outset that the ELE project, bringing together Europe's LT experts, would be vital for the development of digital provision – not

just for European minoritised and endangered languages – but for smaller-state languages as well. All of these languages are threatened by a digital divide. Poor digital provision already exacerbates the various threats to endangered languages especially in terms of lack of provision for young people. Research conducted under the EU-sponsored Digital Language Diversity Project (Soria et al., 2016)[1] (DLDP – comprising ELEN and several of its member organisations) had already identified many of these issues. Subsequently, several of the recommendations from the DLDP project are now in the process of being implemented via the ELE project. ELEN worked to support the European Parliament's 'Language Equality in the Digital Age' Science and Technology Options Assessment (STOA) Report. In order to build political support for the findings from this report Jill Evans, MEP, followed up with a European Parliament own initiative (INI) report, which ELEN helped to draft, as well as lobby for, so as to ensure that it had political support in the European Parliament. Following the INI Report's success in a Plenary vote, ELEN is happy to support and be part of the ELE project in order to provide a SRIA and roadmap for Digital Language Equality (DLE) by 2030.

ELEN comprises 168 member organisations representing 47 languages in 24 European States.

2. Methodology and Instruments

2.1. Online Survey

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

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

Table 1: Type of survey questions

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

- **Part A. Respondents' 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 in which respondents are based
  - Name of the organisation/representative body for which respondents work
  - Communities they represent (if applicable)
  - Type of organisation for which respondents work

[1] http://wp.dldp.eu

  - Sectors or domains in which respondents are active (if applicable)
  - Role of respondents in the organisation (if applicable)
  - Organisation's estimated revenue (if applicable)

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

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

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

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

The survey was designed, set up and published on the EU Survey platform.[2] The full survey, as published online, is presented in Appendix A (p. 13 ff.).

The survey was distributed by ELEN via e-mails to members of the organisation as well as to followers of the organisation on its Twitter and Facebook accounts. It was also advertised on the European Language Equality and European Language Technology websites, LinkedIn page and Twitter account.

The survey was opened on 21st June 2021 and closed on 18th October 2021. In total, 246 responses were collected, out of which 29 were respondents contacted by ELEN with a further four member organisations being separately interviewed. This subset of responses, representing the views of the stakeholders contacted by ELEN is analysed in this report.

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

2.2. Interviews

The ELEN members selected for interview were drawn from the member organisations and supporters that had completed the survey as well as being chosen for the type of language that they represented. Given the problems facing endangered languages and their lack of digital development we chose to interview representatives from several different endangered language communities – Cornish, Sardinian, Frisian and South Sámi.

Four interviews were conducted in November and December 2021 by e-mail. The interview questions were based on the survey with respondents being able to fully discuss their answers to the various questions.

3. Analysis of Responses

3.1. Survey Responses

3.1.1. Respondents' profile

In total, ELEN can account for 35 responses from seven Member-States, comprising 29 responses to the survey and six member organisations interviewed. Nearly all of the answers came from Western Europe which broadly reflects where ELEN's membership is concentrated with the majority of respondents coming from the UK (13) and Spain (12).

States covered via the survey include: UK, Spain, Sweden, Italy, Germany, Estonia.

Countries/ regions covered by the interviews include: Wales, Sardinia, Fryslan, and Sapmi (Norway). Figure 1 shows the breakdown of answers.

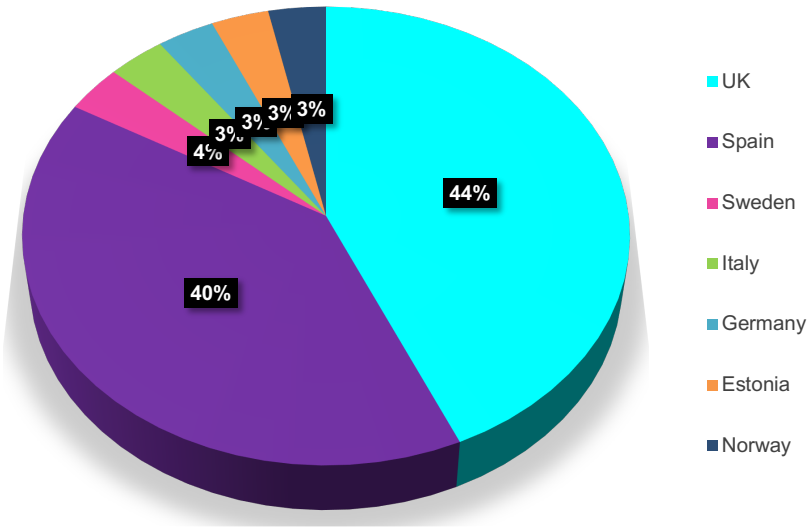


Figure 1: In which country are you based?

The majority of respondents came from the education and the local government sectors, with 11 respondents from education and 9 from government organisations. Other sectors are also represented with 5 NGOs, several from the private sector, research institutes and the self-employed. Those interviewed comprise one NGOs, two from education, and one from local government. Figure 2 shows the breakdown of sectors selected.

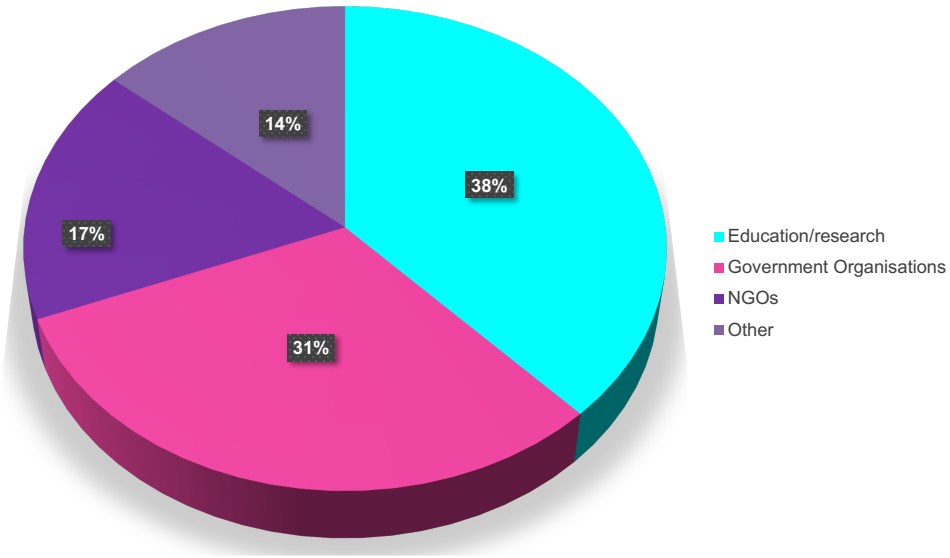


Figure 2: Which of the following best describes the type of organisation you work for?

Most of the organisations who responded comprise either university departments or local government departments that work on digital development for minoritised languages as well as translation. Only a few NGOs and private sector companies responded. Those who responded comprise senior level management as well as teachers, translators and researchers. Those interviewed comprise management, senior academics and a local government officer. Detailed statistics of the breakdown of organisation types and countries are provided in Appendix B Table 3 and Table 2.

3.1.2. Language Coverage

The ELEN respondents all work with minoritised languages, usually in combination with the language of the state. In the UK they work with Welsh, Cornish and Scottish Gaelic, along with English, in Spain with Basque, Catalan and Galician, along with with Spanish, and so on. This is unsurprising as the ELEN membership and followers comprise most of the organisations that work for RMLs in Europe. For the ELEN responses the languages covered are: Welsh, Cornish, Scottish Gaelic, Irish, Breton, Catalan, Basque, Galician, Sámi, Frisian, and Sardinian, as well as Member-State languages such as French, Spanish, English, Swedish, Italian, Norwegian, and Dutch. Welsh responses comprise the majority (11) of the ELEN responses, followed by Catalan and Basque.

3.1.3. Evaluation of the Current Situation

The majority of ELEN respondents use their RML at work. As such, we do not discuss the usage by RML organisations of LT tools in EU official languages here, but only RML organisations usage of RML LT tools/ applications.

Nearly all of the ELEN survey respondents are users of general LT tools such as search tools (10), proofing tools (19), translation tools (19) and language learning tools (8). In addition, a few also use speech recognition and parsing tools (7). Interestingly, 3 respondents state that they do not use any tools in their language at all. Figure 3 shows the breakdown of tools used by respondents.

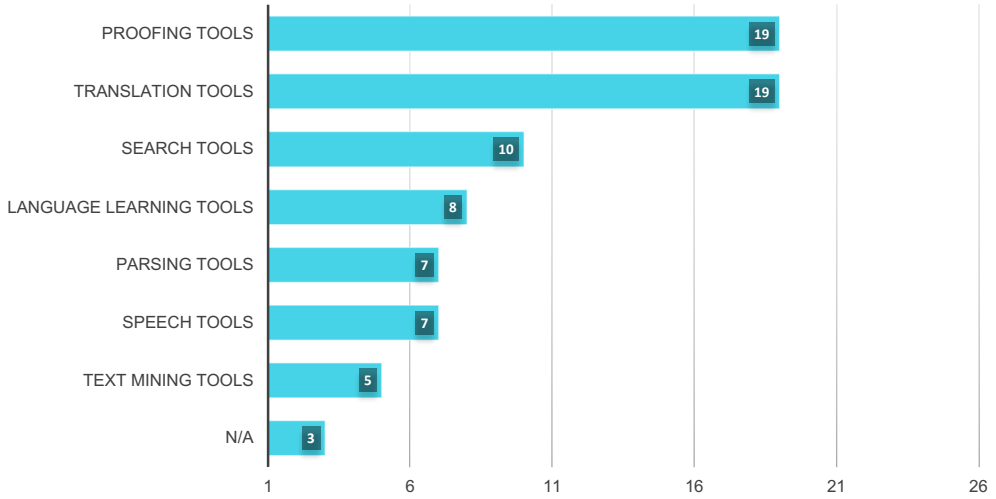


Figure 3: Which language technology tools/applications listed below do you or your organisation use with the official European language(s) you or your organisation work with?

Regarding proofing tools, most respondents used spell-checkers, grammar checkers and autocorrect. Spell checkers were the most used (18), closely followed by grammar checkers (17), and autocorrect (11). 9 respondents did not define which proofing tools they used. The vast majority of respondents (23) use online translation tools with Google Translate being the predominant tool of choice. Eight respondents named speech recognition tools, chiefly Siri, Alexa and Amazon Polly. Four respondents named parsing tools, in particular Stanford NLP. 15 respondents named Google as the search engine used. It is of note that in the speech recognition and parsing category there was an overall lack of answers (8 for speech recognition and only 4 four for parsing) suggesting a lack of tools for these categories. Regarding language learning tools, 11 respondents said that they used them with Duolingo, FluentU and SKELL mentioned most often along with web-based thesaurus tools such as thesaurus.com. There was one mention of the reverse proxy tool LinguaSkin. As with the previous sections the lack of answers is of note suggesting the lack of language learning tools for RMLs in this area.

All (28) except one respondent replied that there are gaps in technological support for RMLs. 20 respondents perceived gaps in the amount and variety of applications; in the quality of applications; in the variety of linguistic phenomena; and in the adaptability. Considering that the respondents languages are noted as Welsh, Catalan, Basque, Galician and Scottish Gaelic, which are comparatively better supported than other RMLs, it suggests that LT support for other less well-supported RMLs must be considerably worse.

There was a mixed range of answers to the question on how respondents evaluated the performance of LT tools. Respondents were asked to choose between 1-4 (where 1 = *very poor*, 2= *poor*, 3= *good*, 4= *excellent*). For proofing the majority found tools as *good* (12) or *excellent* (8). For translation tools a slight majority found these to be *poor* (9) or *very poor* (4) in terms of quality. For speech recognition tools, the majority found them *poor* (5) or *very poor* (8). Regarding parsing tools, the answers were mixed, with 6 finding their level of support to be *good* and 4 finding them to be *very poor*. For search tools the majority found them to be *good* (11) and *excellent* (2). Of those who answered with respect to sentiment analysis tools, most found them as *poor* (2) or *very poor* (4). Of note is the lack of responses provided overall in this section.

Answers varied on the question of the level of LT support for the RMLs worked with. Respondents were asked to choose between 1-4 to evaluate the level of technological support per language (where 1 = *very poor*, 2= *poor*, 3= *good*, 4= *excellent*). For proofing tools, the majority (18) found the level of support either *excellent* (8) or *good* (10). For translation tools the majority (15) saw the support as being *good* (12) or *excellent* (3). Regarding speech recognition, the majority perceived support to be either *very poor* (9) or *poor* (4). For parsing the majority saw the level of support as either *very poor* (5) or *poor* (3), with only 5 seeing support as being *good*. For search tools support, a slight majority saw support as *good* (8) or *excellent* (4), while 9 saw it as being *poor* (4) or *very poor* (5). Figure 4 shows the breakdown of average scores (1-4).

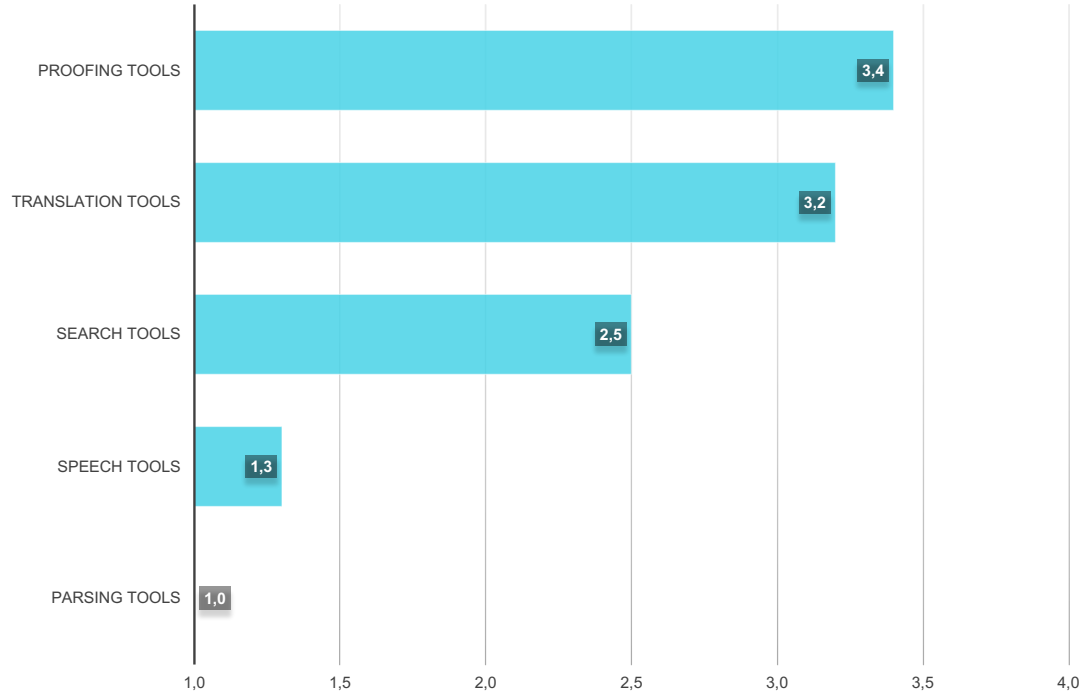


Figure 4: Please choose the option that best describes the level of language technology support for the official European language(s) you or your organisation work with.

On the question of frequency of use of LT tools for RMLs, the answers varied. For proofing, the majority (21) used proofing tools: 16 everyday, 3 frequently, and 2 sometimes. Similarly, with translation tools the majority (22) used them: 7 everyday, 8 frequently and 7 sometimes. However, for speech recognition tools there is far less usage, with 13 saying they never used them. Similarly with parsing tools, with 10 never using them, 3 rarely and 4 sometimes. For

search tools, the majority use them (17): with 12 everyday, 5 frequently, and 2 sometimes. For sentiment analysis tools the majority (15) never use them. Similarly, the majority state that they never use text summarisation tools (16) and text mining tools (14),. For language learning tools usage varies: 8 say they never use them, 8 use them sometimes, and 4 use them everyday.

3.1.4. Predictions and Visions for the Future

Regarding the question of what provision of resources would help to increase the use of LT tools, respondents were very clear in their feedback, with nearly all of them calling for higher quality of LT tools (26), as well as a wider range of tools (24). There was also strong support for more training of personnel who are able to deal with LT tools (11).

There was a wide variety of answers to the question relating to what the respondents would like to see in terms of new tools and apps that are not currently available. The chief call was for speech recognition tools for RMLs, or better speech recognition tools for those few that already have it. The need for spellcheckers, grammar checkers and predictive text were all frequently mentioned, and tools and apps that would help to build confidence in those using and learning endangered languages, for example, dictionary apps for phones. One respondent wrote that "voice recognition systems, such as Alexa or Siri, are not available at all in Catalan, our daily language which is spoken by 10 million Europeans. Also, the other types of technology are either unavailable, or of poor quality. This is a shame, and if Europe wants to protect non-State backed languages, there must be help for them to have accessible technology tools, as the digital one is the main way of communicating that new generations have, or they will be condemned to disappear". The full list of answers is provided in Appendix B Table 4.

On the multiple choice question on what would be the best option that describes the respondents vision for the next ten years, there was a broad range of opinion. With respect to a vision for higher quality language tools for all European languages, opinion was marginally in agreement (10 agree, 4 strongly agree) but with several (4) undecided. That there will be a wider range of language tools for all European languages found strong agreement, with 15 agreeing and 5 strongly agreeing, but with 7 undecided. A tentative majority(13) agreed that LT development will help prevent the further loss of European linguistic diversity with 8 agreeing and 5 strongly agreeing. However, 12 were undecided on the issue, with a further 6 disagreeing. Figure 5 shows the breakdown of answers on the visions for the future.

Finally, regarding the question of what the most relevant benefits are of improving technology for the selected languages, the vast majority of ELEN respondents supported the first three answers. The most relevant benefit is seen as *preventing minoritised territorial/ endangered languages from disappearing* (29), followed by the benefit of *increasing the number of speakers* (23), and *increasing people's exposure to RMLs* (21). Relatively few mentioned the other perceived benefits, such as *improved communications between native speakers* (7), *increased engagement with social, leisure and work activities* (6), *improved literacy rates for RMLs* (5). The perceived benefits for *those with disabilities* and for *online and offline trade* received little support.

3.2. Interview Responses

Interviews were conducted online with 4 member organisations (covering Sámi languages, Cornish, Frisian and Sardinian). In this section we provide a summary of the overall responses for each question. The responses are reprinted in full in Appendix C as they add the all important perspective of the language community to the current state of LT development and needs for the future.

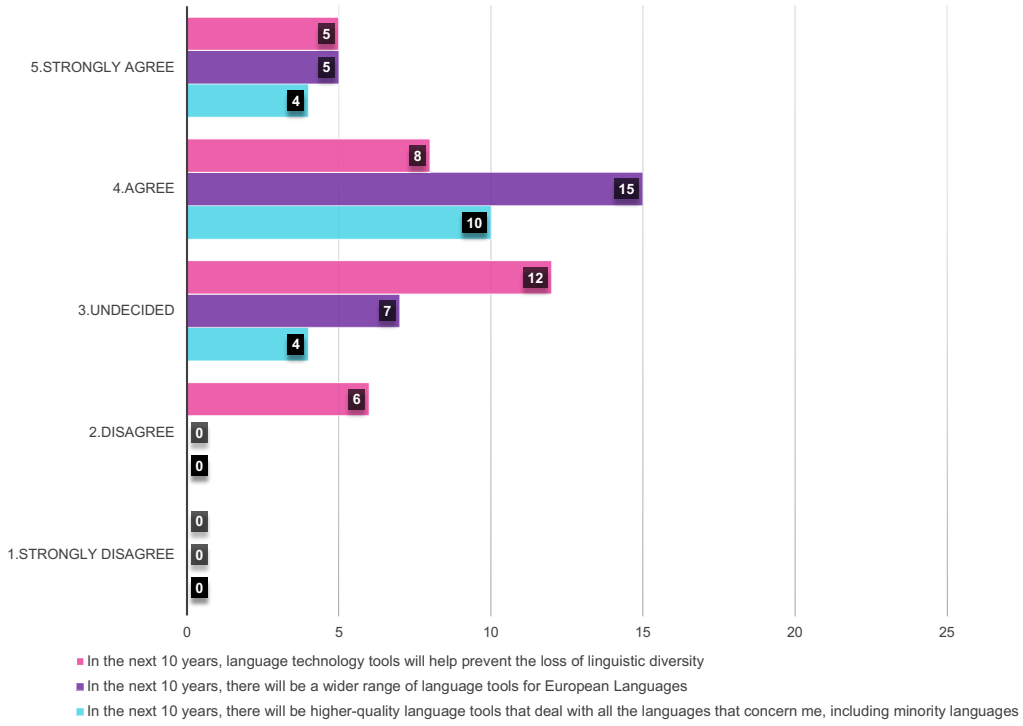


Figure 5: Please indicate the best option that describes your vision for the future of languages technology

3.2.1. Question 1: How would you evaluate the current levels of LT support for your language?

Based on answers from interviewees, overall there appears to be *poor support* for all lesser-used languages. Some respondents note the existence of basic tools and resources such as dictionaries and spell-checkers (all except Cornish), and in some cases translation tools (Frisian and Sardinian), yet acknowledge that this is nowhere near a sufficient level of digital support required. A general pattern can be observed in that the more endangered or the fewer users a language has, the less LT support it has.

3.2.2. Question 2: What gaps have you identified in terms of LT support?

Regarding gaps in LT support, one of the key phrases in the answers was the need to “lift current capacity”. Among the fundamental requests for lesser-used/endangered languages are spellcheckers, translation memory software, language learning apps, speech technologies. The understanding is that even the availability of basic tools would really help improve efficiency and add to capacity. Overall, the general opinion is that there is no need to reinvent the wheel but simply to ensure that each RML has a basic toolkit available so it can build its own digital capacity.

3.2.3. Question 3: With the existing LT support for your language, how would you evaluate its performance?

Regarding the performance of existing LT support, those interviewed mostly agreed that performance was good for the few tools that do exist. What is clear is that some advances that have been made are thanks to the efforts of wider research groups or open-source projects (Bangor University for Cornish, Mozilla/Apertium for Sardinian, Mozilla for Frisian). However, respondents noted the overall lack of tools for RMLs, and not knowing what to prioritise in terms of what tools should be developed next.

3.2.4. Question 4: What policies or instruments would speed up the aim of having digital language equality in Europe?

There were a wide range of answers to this question addressing the problem of the digital divide on the micro and macro level. At the micro level it was suggested that it would be "transformational" if existing tools be made available for smaller languages and an increase of availability of language resources in the public domain or with permissive licenses. As one respondent commented, "We can do the language bit – but we are having to do IT research for which we are not trained".

At the macro level it was proposed that there be a common lobby by RML communities towards large technology companies so as to ensure LT support for RMLs, for requirements in tenders to include RMLs in technology products, and more requirements on national governments to always include RMLs in their communications.

The overall observation is that state level and/ or EU level legislation will be needed to ensure such provision. For example, existing language legislation that protects RMLs could have clauses added to address the lack of provision in the digital domain. Or new legislation could be tabled to address the problem, c.f. the new legislation in Spain to ensure that Netflix produces a certain amount of content in Catalan, Basque and Galician.

3.2.5. Question 5: Do you recommend any particular research that would help speed up the provision of LT in your language?

There were specific suggestions for more open-source licensing, automatic subtitling technologies and machine translation for Sardinian. However, it is interesting to note that there were no answers provided by Sámi, Frisian or Cornish, as they clearly were not sure what research would help speed up provision of LT. This reinforces the message that the RML organisations are good at languages but not necessarily LT experts and therefore not able to recommend the research required that would help their language LT development the most. Clearly, this is a fundamental problem that needs to be addressed, where smaller RML communities would benefit from consultations with LT experts to fully understand which research is needed for LT development in their languages.

3.2.6. Question 6: What priorities should a future programme on digital European language equality address in order to help support minoritised and endangered languages?

Responses focused on various practical suggestions that a future programme on digital language equality should address. Chief among these was to provide default platforms particularly for Computer Aided Language Learning tools, along with translation memory software and spellcheckers as these would be the most beneficial as they provide building blocks for other tools – and they are basic established tech for other languages. Having access to LT, and for there to be an openness (in terms of licensing and sharing) and collaborative approach,

rather than having to research and make the case for access, would make a significant difference.

3.3. Summary

The interviews largely back up the data collected from the survey. The overarching challenge for all RMLs is the failure to keep pace with all the new developments for the widely-used languages. Such a challenge is not insurmountable. If a strategic development plan is put in place and financed the "bigger" RMLs would be well positioned to catch up with medium-sized EU official languages. However, of concern is the lack of answers and input from European endangered languages other than Sardinian, Sámi, Frisian and Cornish covered in the interviews. There is a danger of endangered languages slipping under the radar in European LT development, an issue which the ELE project is well-equipped to address.

At the European policy level the EU has come out in strong support for European digital language equality. It now needs to make this support tangible by proposing an EU Regulation or Directive that ensures digital language equality in its Member-States and that is particularly focused on ensuring that RMLs/ endangered languages are supported. Clearly, if we are to achieve the EU's stated aim of digital language equality, LT in lesser-used languages will need a lot more development and investment if they are to "catch-up" with the better supported languages.

4. Conclusions

From the survey of ELEN members and supporters regarding LT provision for RMLs a fairly uniform picture emerges. On the whole ELEN respondents use generic systems provided by online services for search, proofing, translation and language learning. Meanwhile, speech tools and tools for text summarising or sentiment analysis are rarely used. This matches the results found for the smaller and medium-sized EU official languages.

An overwhelming majority identified gaps in LT provision for RMLs. Most respondents perceived gaps in the amount and variety of applications; in the quality of applications; in the variety of linguistic phenomena covered by the tools; and in the adaptability to systems. Poor support and often very poor support is stated for all RMLs. According to the respondents the situation is better for generic proofing tools, search systems, translation tools and language learning applications, than for more advanced text and speech applications, such as text summarizing, sentiment analysis and Text-to-Speech systems.

On the topic of what would improve the situation the chief call was for speech recognition tools for RMLs, or better speech recognition tools for those few that already have it. The need for spellcheckers, grammar checkers and predictive text were all frequently mentioned, and tools and apps that would help to build confidence in those using and learning endangered languages, for example, dictionary apps for phones.

Regarding expectations for the future most respondents expect a wider range of tools as well as higher quality tools in all areas. Interestingly, a vast majority of respondents expected improvements in LT for RMLs to have a substantive effect in supporting endangered languages as well as increasing the number of speakers.

Of note is that the Digital Language Diversity Project's conclusions and roadmap[3] (Soria et al., 2018) are supported by the data collected in the ELE survey. It identified the pressing need for LT development for RMLs, backed with legislation as a matter of urgency in that RMLs face a digital time-bomb as society becomes increasingly digitised but only using a few

[3] http://wp.dldp.eu/the-dldp-roadmap/

of the more widely spoken languages. As society becomes more digitised so RMLs are increasingly excluded from various sociolinguistic domains that are vital for their continued usage and well being. For endangered languages, lack of LT development may exacerbate their endangerment unless they are supported and given the tools to close the digital language divide.

Finally, it is worth noting some feedback with respect to the survey design in the context of RMLs. Considering that the ELE project relates to language equality and that ELEN respondents work using their RML, many ELEN respondents questioned the logic in drawing the unnecessary distinction between EU official and EU RMLs in the survey when it comes to researching the use of LT tools and applications. The survey did ask about usage of LT tools and applications by RML users in a later section (starting on p12), but some NGOs pointed out that it would have been easier to have seen their language listed along with the EU official languages, and treated with some regard for the notion of equality.

References

Eneko Agirre, Izaskun Aldezabal, Iñaki Alegria, Xabier Arregi, Jose Mari Arriola, Xabier Artola, A Díaz de Ilarraza, N Ezeiza, K Gojenola, K Sarasola, et al. Developing language technology for a minority language: progress and strategy. *Elsnews*, 10:4–5, 2001.

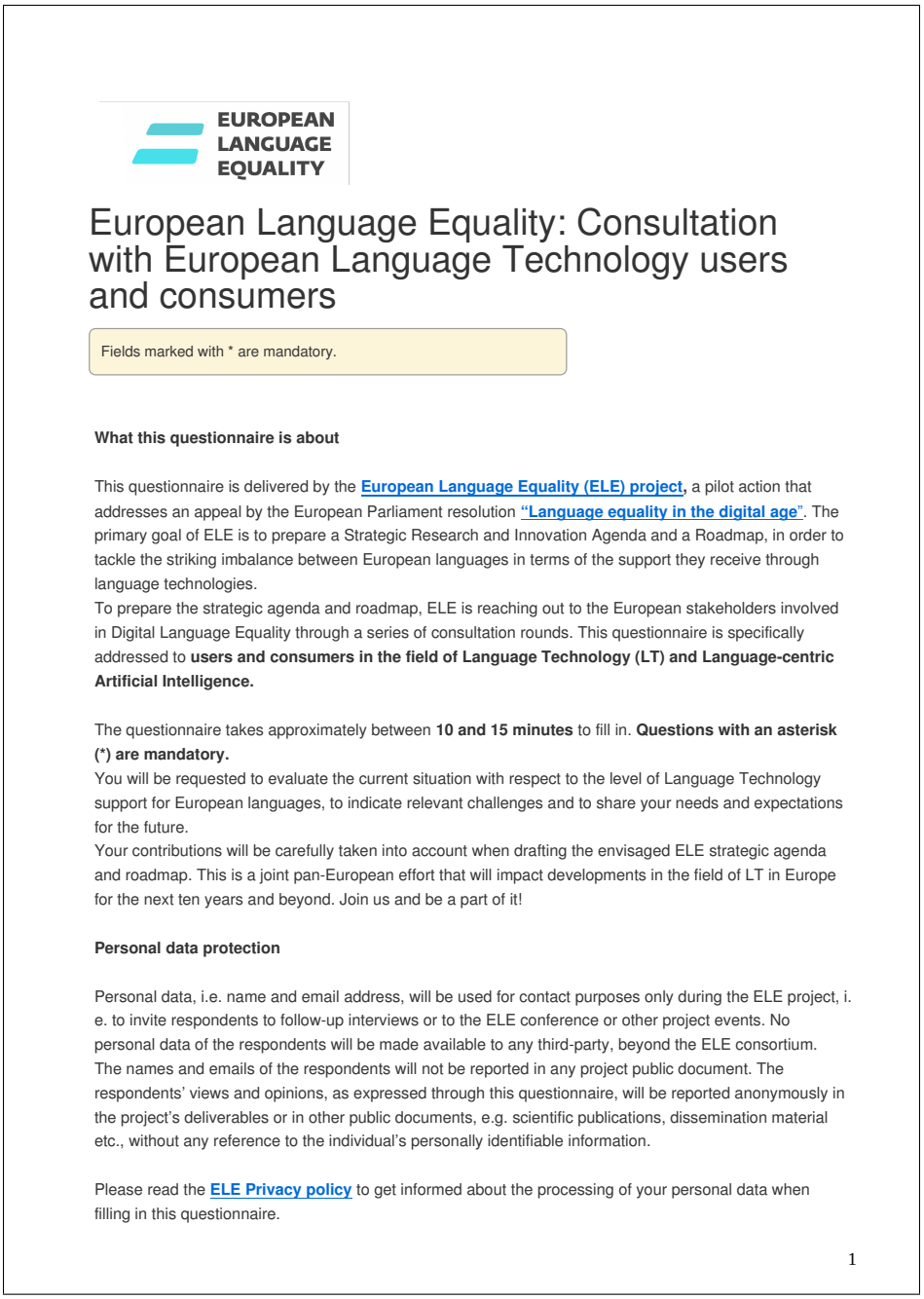
Kepa Sarasola. Strategic priorities for the development of language technology in minority languages. In *Proceedings of the Workshop on Developing language resources for minority languages: reuseability and strategic priorities. Second International Conference on Language Resources and Evaluation (LREC 2000) (LREC'16)*, pages 106–109, Athens, Greece, 2000. European Language Resources Association (ELRA).

Claudia Soria, Irene Russo, Valeria Quochi, Davyth Hicks, Antton Gurrutxaga, Anneli Sarhimaa, and Matti Tuomisto. Fostering digital representation of EU regional and minority languages: the digital language diversity project. In *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*, pages 3256–3260, Portorož, Slovenia, May 2016. European Language Resources Association (ELRA). URL https://aclanthology.org/L16-1518.

Claudia Soria, Valeria Quochi, and Irene Russo. The DLDP Survey on Digital Use and Usability of EU Regional and Minority Languages. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan, May 7-12, 2018 2018. European Language Resources Association (ELRA). ISBN 979-10-95546-00-9.

A. LT users and consumers survey

Figures 6 to 23 show the complete LT research and developers survey.



EUROPEAN LANGUAGE EQUALITY

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

Fields marked with * are mandatory.

What this questionnaire is about

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

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

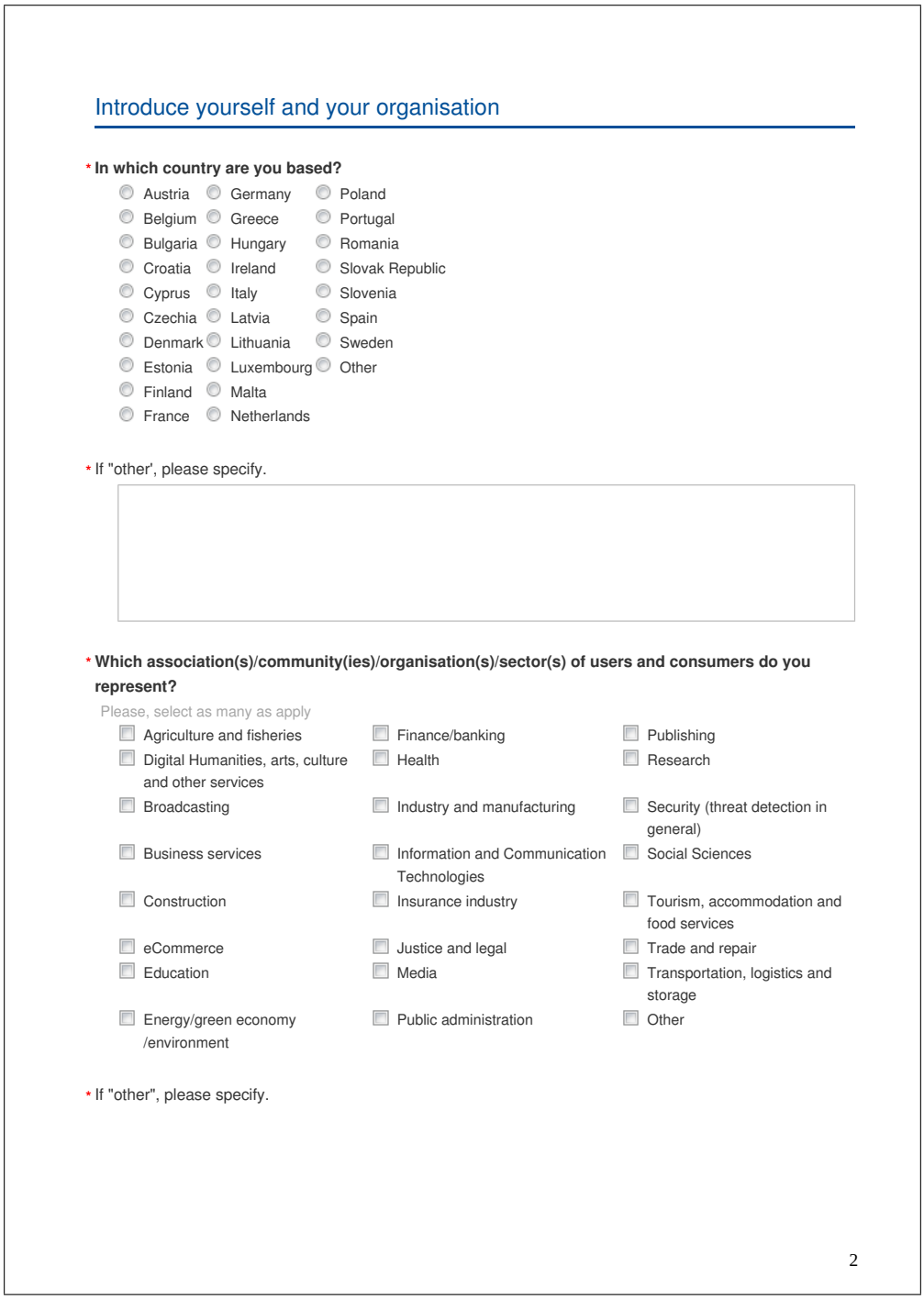
Personal data protection

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

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

1

Figure 6: Full survey as published (page 1/18)



Introduce yourself and your organisation

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

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

* If "other', please specify.

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

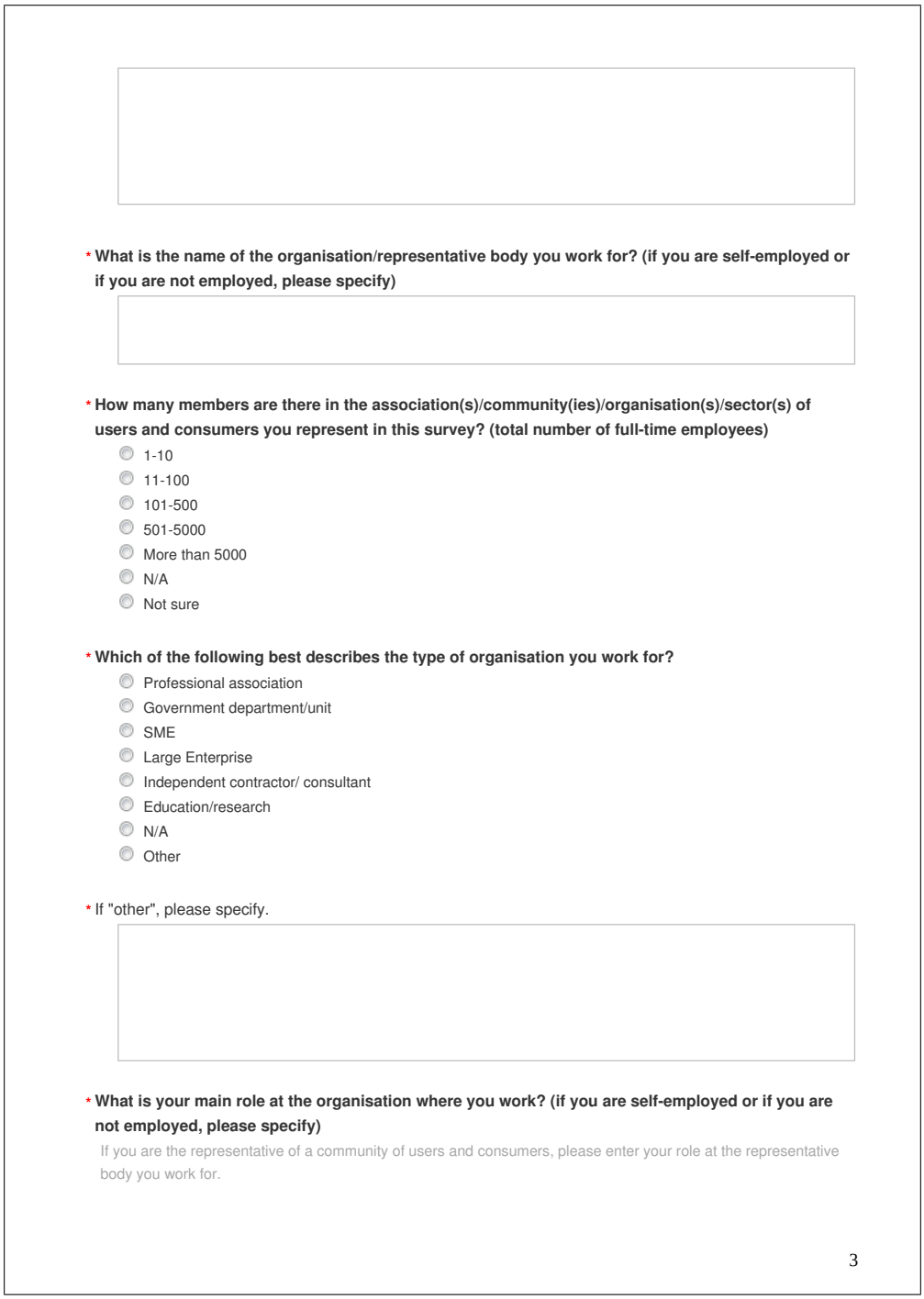
Please, select as many as apply

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

* If "other", please specify.

2

Figure 7: Full survey as published (page 2/18)



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

* **How many members are there in the association(s)/community(ies)/organisation(s)/sector(s) of users and consumers you represent in this survey? (total number of full-time employees)**

- 1-10
- 11-100
- 101-500
- 501-5000
- More than 5000
- N/A
- Not sure

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

- Professional association
- Government department/unit
- SME
- Large Enterprise
- Independent contractor/ consultant
- Education/research
- N/A
- Other

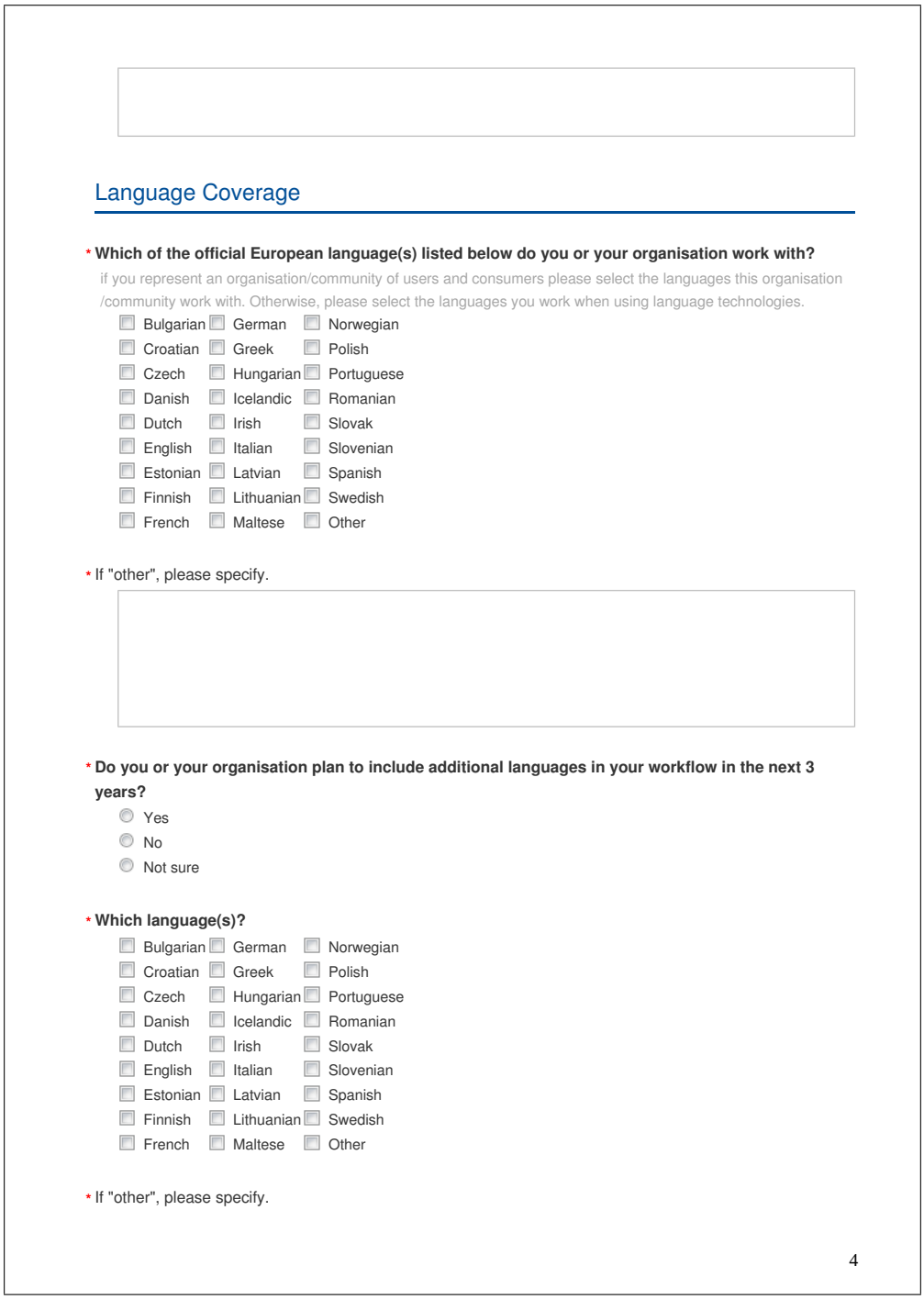
* If "other", please specify.

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

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

3

Figure 8: Full survey as published (page 3/18)



Language Coverage

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

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

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

* If "other", please specify.

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

- Yes
- No
- Not sure

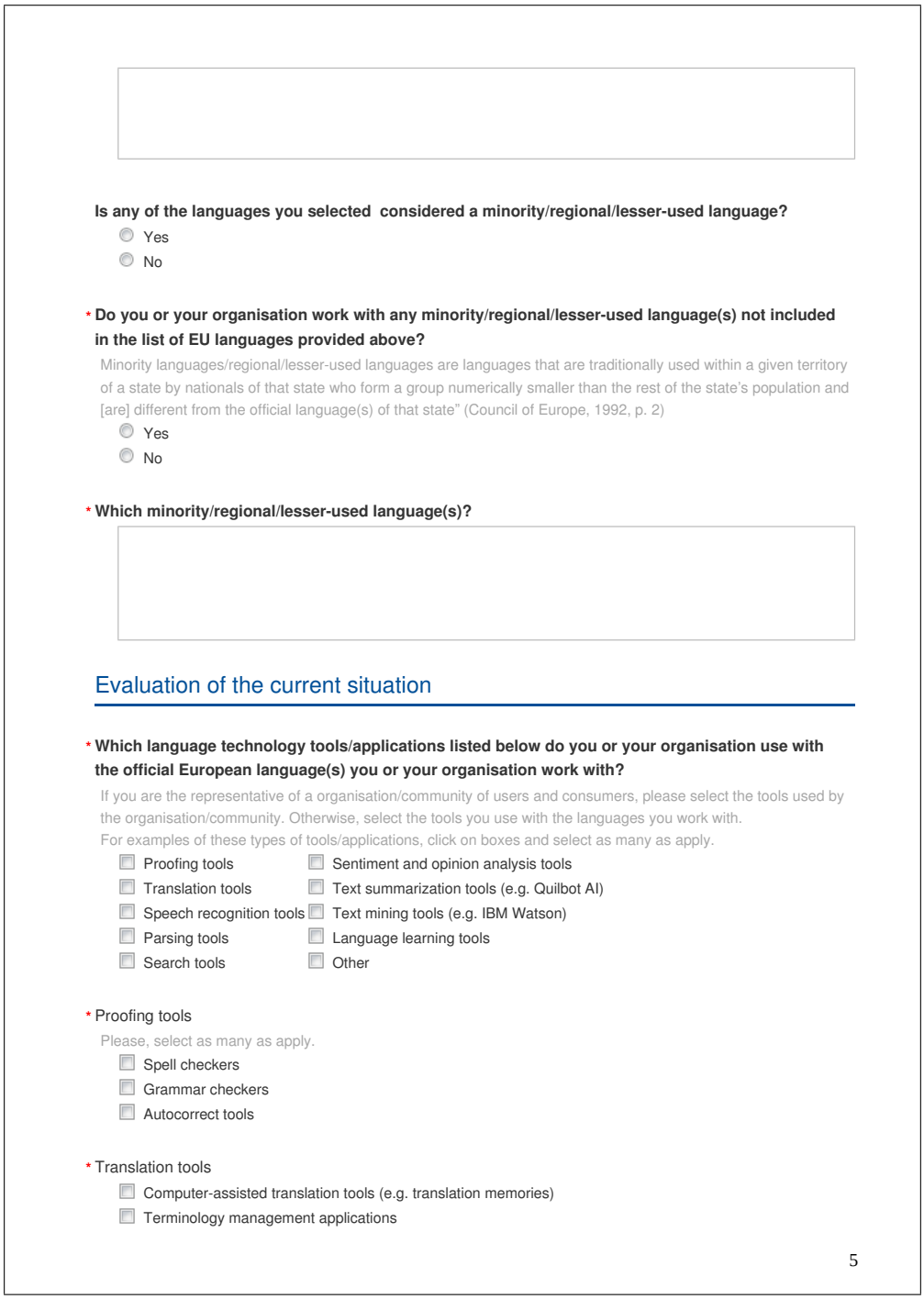
* **Which language(s)?**

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

* If "other", please specify.

4

Figure 9: Full survey as published (page 4/18)



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

- Yes
- No

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

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

- Yes
- No

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

Evaluation of the current situation

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

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

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

* Proofing tools

Please, select as many as apply.

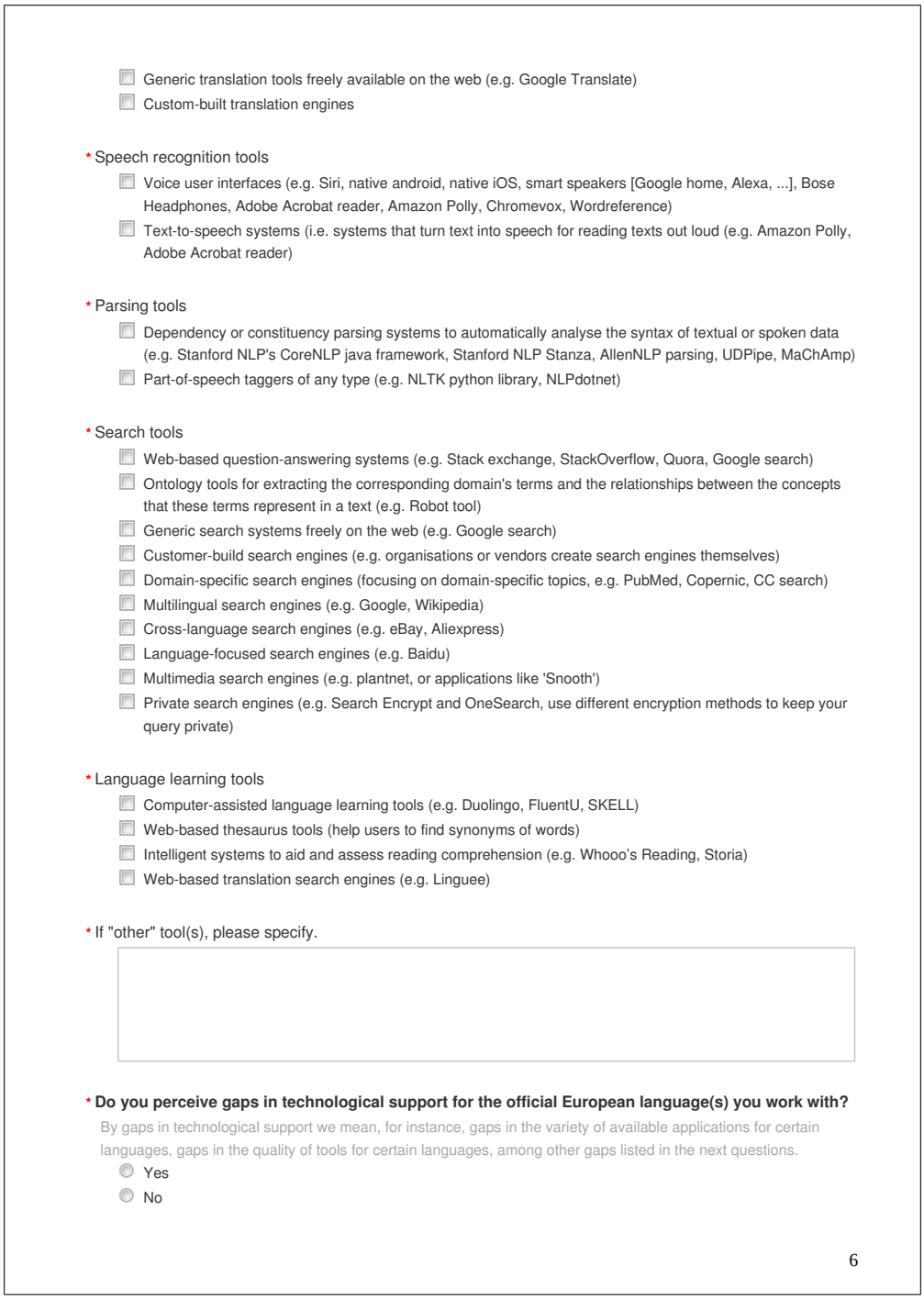
- Spell checkers
- Grammar checkers
- Autocorrect tools

* Translation tools

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

5

Figure 10: Full survey as published (page 5/18)



- [ ] Generic translation tools freely available on the web (e.g. Google Translate)
- [ ] Custom-built translation engines

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

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

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

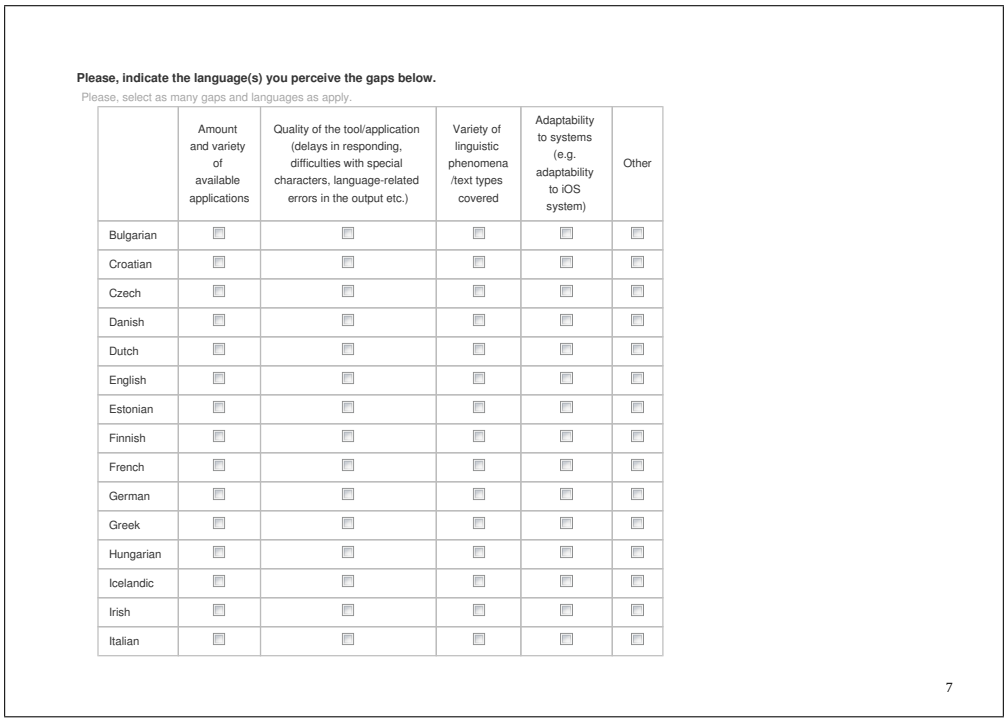
* Language learning tools
  - [ ] Computer-assisted language learning tools (e.g. Duolingo, FluentU, SKELL)
  - [ ] Web-based thesaurus tools (help users to find synonyms of words)
  - [ ] Intelligent systems to aid and assess reading comprehension (e.g. Whooo's Reading, Storia)
  - [ ] Web-based translation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

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

* **Do you perceive gaps in technological support for the official European language(s) you work with?**
  By gaps in technological support we mean, for instance, gaps in the variety of available applications for certain languages, gaps in the quality of tools for certain languages, among other gaps listed in the next questions.
  - ( ) Yes
  - ( ) No

6

Figure 11: Full survey as published (page 6/18)



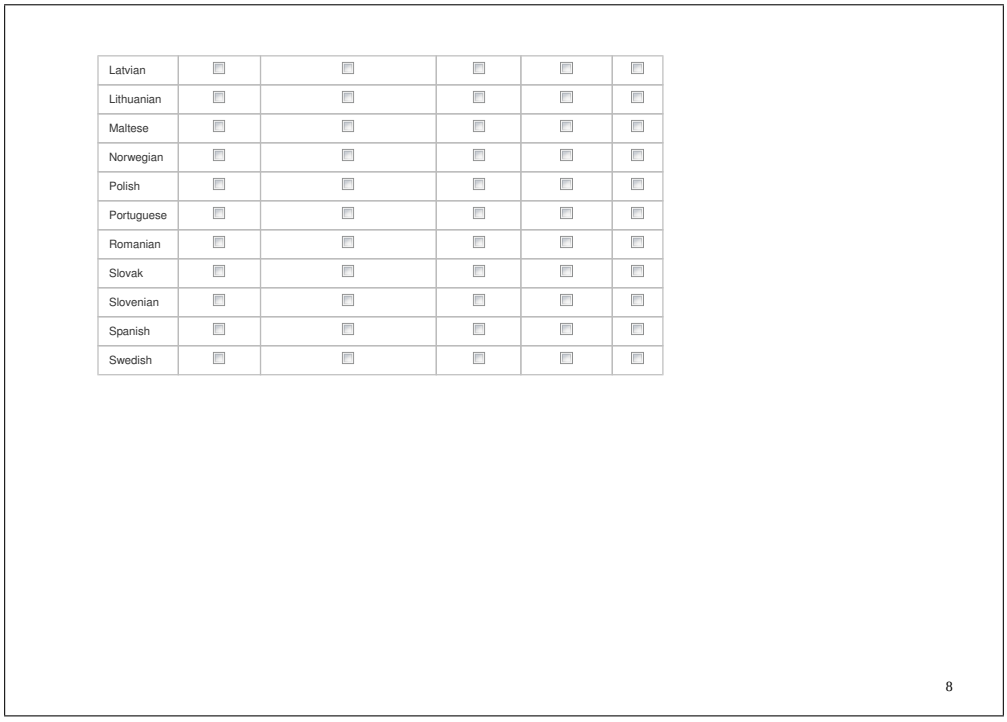
Please, indicate the language(s) you perceive the gaps below.

Please, select as many gaps and languages as apply.

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

7

Figure 12: Full survey as published (page 7/18)



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

8

Figure 13: Full survey as published (page 8/18)

* If "other", please specify.

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

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

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

* If "other", please specify.

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

Please, choose as many languages as apply.

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

9

Figure 14: Full survey as published (page 9/18)

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

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

Please, select as many tools as apply.

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

10

Figure 15: Full survey as published (page 10/18)

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

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

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

Please, select as many tools and languages as apply.

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

11

Figure 16: Full survey as published (page 11/18)

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

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

Are there language technology tools/applications available for the minority/regional/lesser-used language(s) you or your organisation work with?

- ○ Yes
- ○ No
- ○ I do not know

* **Which tools/applications do you use with these minority/regional/lesser-used languages?**

For more examples of these types of tools, click on the boxes and select as many tools as apply.

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

* Proofing tools

Select as many as apply.

- ☐ Spell checkers
- ☐ Grammar checkers
- ☐ Autocorrect

12

Figure 17: Full survey as published (page 12/18)

* Translation tools

Select as many as apply.

- Computer-assisted translation tools (e.g. translation memories)
- Terminology management applications
- Generic translation tools freely available on the web (e.g. Google Translate)
- Custom-built translation engines

* Speech recognition/synthesis tools

Select as many as apply.

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

* Parsing tools

Please, select as many as apply.

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

* Search tools

Please, select as many as apply.

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

* Language learning tools

Please, select as many as apply.

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

* If "other", please specify.

13

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

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

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

- Yes
- No

Please, indicate the gap(s) you perceive.

Please, select as many as apply.

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

* If "other", please specify.

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

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

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

14

Figure 19: Full survey as published (page 14/18)

Text summarization (e.g. Quillbot)					
Text mining (e.g. IBM Watson)					
Language learning (e.g. Duolingo, thesaurus, bilingual dictionaries)					
Other					

If "other", please specify.

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

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

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

* If "other", please specify.

15

Figure 20: Full survey as published (page 15/18)

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

Please, select as many tools as apply.

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

* If "other" tool, please specify.

Predictions and visions for future

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

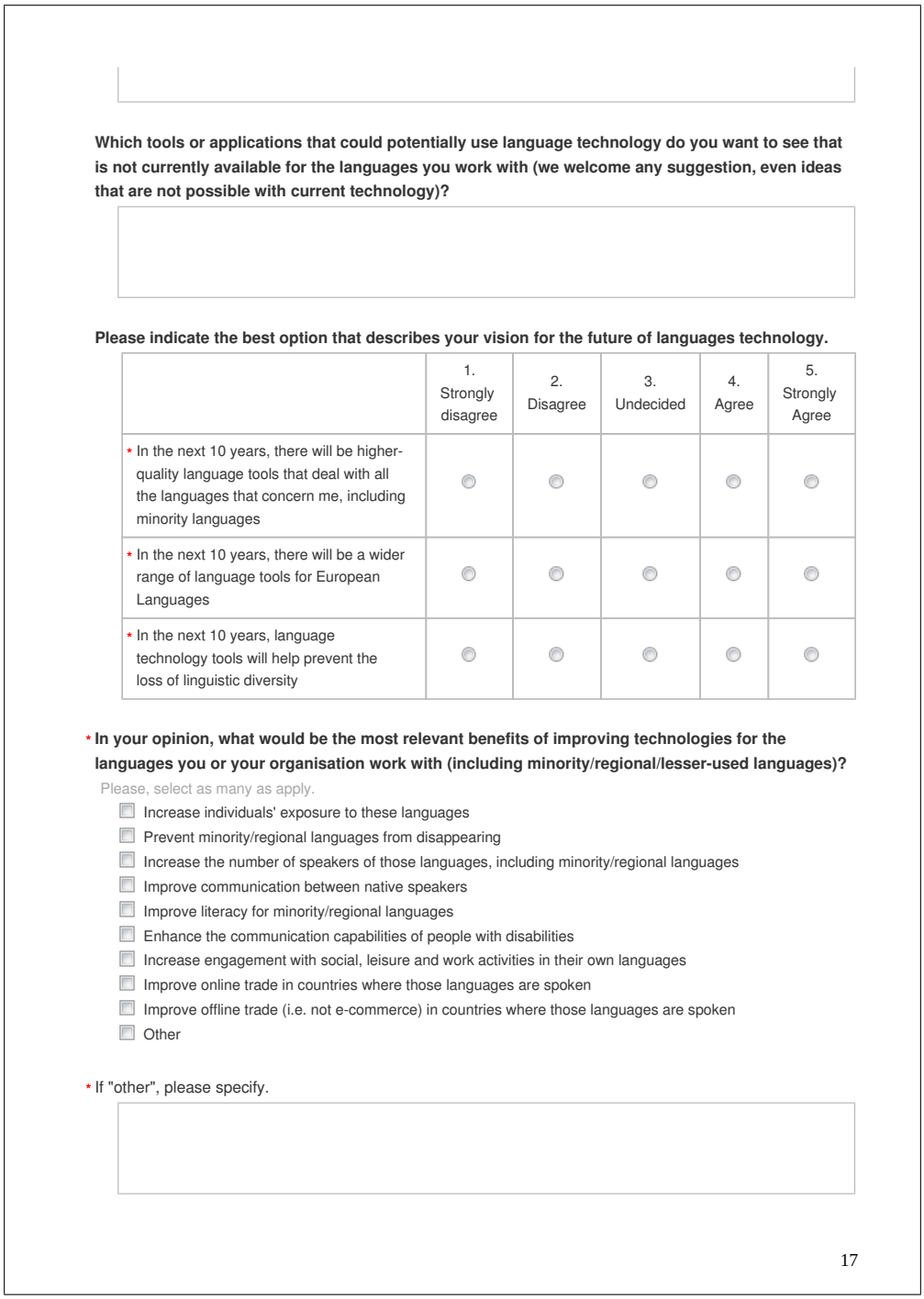
Please, select as many as apply.

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

* If "other", please specify.

16

Figure 21: Full survey as published (page 16/18)



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

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

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

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

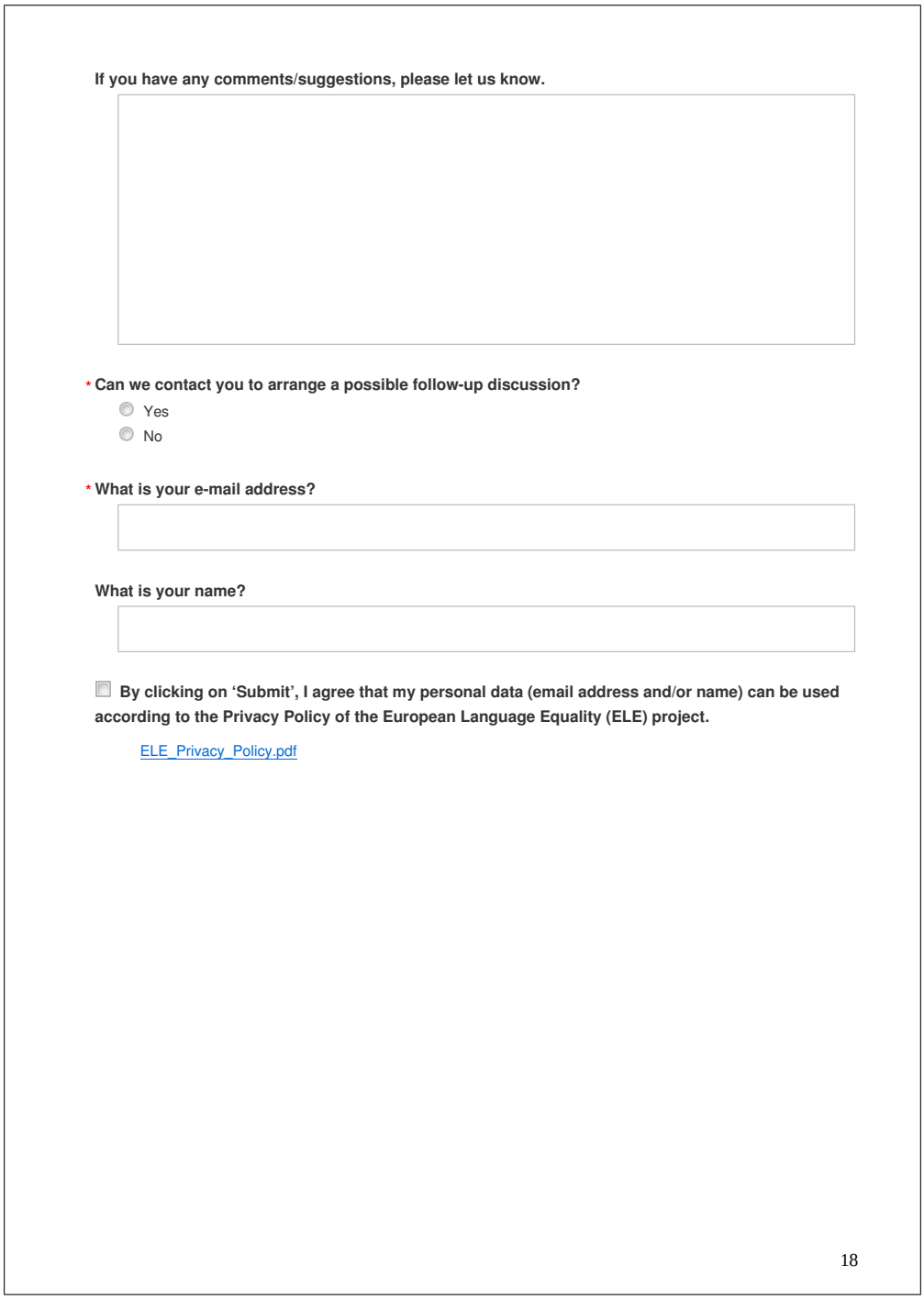
Please, select as many as apply.

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

* If "other", please specify.

17

Figure 22: Full survey as published (page 17/18)



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

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

- Yes
- No

* What is your e-mail address?

What is your name?

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

ELE_Privacy_Policy.pdf

18

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

B. Additional tables and graphs

Types of organisations	Answers count	%
Education/research	11	38
Government organisation	9	31
NGO	5	17.2
Other	4	13.8

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

Country	Respondents (%)	
UK	13	44.8
Spain	12	41.4
Sweden	1	3.4
Italy	1	3.4
Germany	1	3.4
Estonia	1	3.4

Table 3: Breakdown of answers to the question “Where are you based in?” (Example of mandatory closed question, plus “if other” as optional open-ended question)

Speech recognition tools for RMLs
Better speech recognition tools for those who already have it
Spell checkers
Grammar checkers
Predictive text
Tools and apps that would help to build confidence in learning endangered languages (e.g.. Dictionary apps)
voice recognition for Catalan

Table 4: Full list of answers to “Which tools or applications that could potentially use language technology do you want to see that is not currently available for the languages you work with (we welcome any suggestion, even ideas that are not possible with current technology)?” (Example of optional open-ended question)

C. Content of Interviews

C.1. Interview 1. Sámi languages. Academic

1) How would you evaluate the current levels of LT support for South Sámi? I would say that the current levels of LT support for the Sámi languages are relatively high. All Sámi languages have digital dictionaries, spell checkers, paradigm generators, text analysis, and keyboards with the characters in the Sámi alphabets that are not found in Norwegian keyboards. There are also text corpora and different levels of machine translation between the Sámi languages (at beta level). There are some more programs for North Sámi, such as machine translation from North Sámi to Norwegian, a text-to-speech program and grammar checker and a couple language learning programs. In addition, there Sámi keyboards for cellphones and tablets for all Sámi languages.

2) What gaps have you identified in terms of LT support? There is obviously a gap between the different levels of LT support in the Sámi languages, with North Sámi having the most support. This is due to human resources as there are more North Sámi speakers than South- and Lule Sámi.

3) With the existing LT support for South Sámi, how would you evaluate its performance? I have not tried all the programs for Lule Sámi (and none of the South or North Sámi programs), but the Lule Sámi programs are good. There are some difficulties with running the newest spell checker at Apple computers.

4) What policies or instruments would speed up the aim of having digital language equality in Europe? Resources to develop for instance digital language courses.

5) Do you recommend any particular research that would help speed up the provision of LT in Sámi? No answer.

6) What priorities should a future programme on digital European language equality address in order to help support minoritised and endangered languages? Digital language courses, as apps, easily downloadable on your phone (think duolingo, etc). I think easy access to language learning products on one´s phones would make it easier to learn and practice the language.

C.2. Interview 2. Cornish. Government Officer

1) How would you evaluate the current levels of LT support for Cornish? Poor/low support. While for major languages, LT providers will be developing products and services to offer to that market, we have to research tools ourselves to find what could work, make the case and find funding for each tool.

2) What gaps have you identified in terms of LT support? Having developed our dictionary database, the two priority gaps for us are a) translation memory software, and b) spellchecker. These are basic tools that can really help improve our efficiency and add to our capacity. We use SmartCat translation memory software which is good, but it does not properly support Cornish as a language as we have to do additional admin. We don't need advance functionality – a similar open source, basic translation memory software platform

would help our efficiency and be the basis for building a digital corpus to serve other tools. The spellchecker would also be a good way for us to reinforce standardisation and enable more of our less confident speakers to write and publish – so again lifting our capacity.

3) With the existing LT support for Cornish, how would you evaluate its performance? LT tools where we have them are extremely important. With a small number of speakers, we need to make sure each speaker has the tools to do all sorts of tasks like translating, teaching, writing articles – so having our online dictionary and corpus search given them support to do these tasks more quickly. We have worked with Bangor University to put in place these tools and they work very well for us and indeed, having template tools is really empowering and galvanising. We have really relied on Bangor Unis evaluation of the technology itself and have then monitored use of the tools such as the online dictionary – but we are starting from such a low base and we don't really have a comparator. Being able to compare user numbers with other small languages in a similar position would be useful as we could see whether our LT usage is similar, better or where we have gaps. We don't really know how to prioritise – with limited resources, which tool is the most important next step.

4) What policies or instruments would speed up the aim of having digital language equality in Europe? We spend a lot of time researching what tools we should be using and how to develop them. We are small in number as it is and having established tools opened out for smaller languages to plug in their languages would be transformational for us. We can do the language bit – but we are having to do IT research for which we are not trained.

5) Do you recommend any particular research that would help speed up the provision of LT in Cornish? Not sure

6) What priorities should a future programme on digital European language equality address in order to help support minoritised and endangered languages? Providing default platforms particularly for translation memory software and spellcheckers would be the most beneficial to Cornish as these provide building blocks for other tools – and they are basic established tech for other languages. Having access to LT, rather than having to research and make the case for access, would make a significant difference.

C.3. Interview 3. Frisian. NGO

1) How would you evaluate the current levels of LT support for Frisian? It is on an intermediate level. On the one hand, several applications have been developed. On the other hand, it are usually not the big tech companies that automatically develop their applications in Frisian. The following applications are now available: Google Translate through community development Gboard and Swiftkey as automatic text predictors Frisian as a language in Microsoft Office with built-in spelling and grammatical corrector on Windows operating systems (however, this has to be downloaded separately, and is not automatically integrated in updates and when people buy Microsoft Office software). That this is now available is a result of provincial funding and cooperation with the Fryske Akademy and Microsoft. Navigation in Frisian through Waze Online Frisian-Dutch dictionaries developed by Fryske Akademy and based on those data popular language applications In development: A PhD from RUG Campus Fryslân is developing a synthetic voice in Frisian. Several parties are working on POS-taggers. An automatic bilingual subtitle is in development. The performance is still not adequate.

2) What gaps have you identified in terms of LT support?

- Frisian as a language in Apple products
- Text to speech and speech to text applications. We are working together with Mozilla Common Voice to build up a database. This slowly progresses. We are also exploring the possibilities to start pilots with this technology, e.g. in healthcare with social robots.
- Artificial Intelligence in Frisian
- Language learning apps

3) With the existing LT support for Frisian, how would you evaluate its performance? Intermediate

4) What policies or instruments would speed up the aim of having digital language equality in Europe? Common lobby towards big tech companies with the minority language communities together. Requirements in tenders to include the (recognised) minority languages in technology products as well. More requirements to national governments to always include the minority languages in their communications.

5) Do you recommend any particular research that would help speed up the provision of LT in Frisian? No answer.

6) What priorities should a future programme on digital European language equality address in order to help support minoritised and endangered languages? Text to speech and speech to text applications in combination with artificial intelligence. We are particularly interested in technology that is applied in the home as the home domain is still very strong when it concerns the use of Frisian. Through technology other languages are however introduced in the home, think of smart home appliances, social care robots, etc.

C.4. Interview 4. Sardinian. LT academic.

1) How would you evaluate the current levels of LT support for Sardinian? Poor. While acknowledging that, as Russo and Soria state in their 2017 report[4], "the existence of a considerable number of language resources such as dictionaries, spell checkers, and even an automatic translation system, is a good sign of the potential for this language to become a fully digital language", the limited support is clear when looking at roadmaps for the digitization of minoritized languages, such as the one by the IXA group (Sarasola, 2000; Agirre et al., 2001). A quick count of the number of available technologies in Sardinian for each of the five categories established by the authors, none are found in levels 3, 4 and 5, and just a few are available among the first phase (laying foundations) and the second phase (basic tools).

2) What gaps have you identified in terms of LT support? The lack of consensus on the language model partially explains this poor support, but not only. Other languages in similar scenarios (in terms of number of speakers or internal variation) have a much wider range of available tools, be it because of institutional support, academic involvement or committed communities. The robustness principle, applied by LTs in other languages, represents a good

[4] http://wp.dldp.eu/download/report-sardinian-a-digital-language/

trade-off in the standardization debate: ”be conservative in what you send, be liberal in what you accept”. See also comments on the limited available of resources and permissive licenses below.

3) With the existing LT support for Sardinian how would you evaluate its performance? The performance is remarkable, especially when it comes to specific tools like spell-checkers (CROS), MT engines (Apertium) and electronic dictionaries (Ditzionàriu). Other important initiatives have been incomprehensibly discontinued, such as the Sintesa speech synthesizer. In the field of dictionaries, several initiatives boast a notable quality (often times, thanks to the commitment of single individuals), although they are only offered in PDF or with limited search engines.

4) What policies or instruments would speed up the aim of having digital language equality in Europe? At the political level, regional, national and European support would have a huge impact, especially in the following fields:

- Promotion of multilingualism: Sardinia is an incredible example of a living ecosystem with several endangered languages: all of them deserve protection and promotion. According to experts (Junyent 1998), promoting multilingualism might be even more effective than some classical language planning policies.
- Promotion of Free, Libre, Open-Source Software (FLOSS): In line with the Manifesto for Open Language Technology (https://openlt.org/), if all software vendors allowed access to the source code, endangered language communities would have better chances of getting organized and providing their own localizations.
- Increased availability of language resources in the public domain or with permissive licenses. A machine translation engine or a speech recognition system can be trained in a matter of hours with the present state of technology. But the lack of publicly available resources hinders achieving this: multimedia creators, editors in the publishing domain and, especially, the institutions should be encouraged to release all texts with Creative Commons or similar.
- What the case of Sardinia shows clearly is that endangered languages can hardly be protected without the direct implication of their speakers (Tabarkin Ligurian, a Sardinian language, has around 4,000 speakers; 86% of the total population; approx. figures off the top of my head). In this sense, political actors in Europe and Sardinia should not disregard the transformative potential of initiatives such as the Universal Basic Income (UBI) when it comes to ensuring the active involvement by the citizenship in the preservation of the cultural heritage.
- At the regional level, a lack of interest can be observed on the institutional side, as the discontinuation of publicly funded projects (such as Sintesa) shows. Still more striking is how decision-makers disregard the communitarian approach: adopting free products developed by the community (such as, for instance, offering a machine-translated version of institutional websites through Apertium) would have a highly symbolic value at no cost. This is something that other language communities have done in the past (Catalan, Occitan, etc.).
- Increased funds for publicly-funded projects. 2020 was exceptionally generous in the amount of money devoted by the Sardinian government to language policies. However, when it comes to the development of language technologies, few applications made it to the last step, and the budget was ultimately trimmed. Easing the bureaucratic burden would surely raise the amount of successful applications.

- The university as an institution should lead or at least participate in any effort involving the languages of Sardinia. I can hardly see any way of increasing its involvement, probably due to the specificities of the Italian university system.
- At the European level, maybe joint efforts such as yours could provide basic online infrastructure (hardware, technical support, roadmaps) for language communities, pretty much as living labs do in the physical world.

5) Do you recommend any particular research that would help speed up the provision of LT in Sardinian? We at Sardware have been working on MT and hope that our collateral outputs will help bringing Sardinian to a higher level: in the context of an English-Sardinian translator, we will be releasing a tokenizer (the first one to the best of our knowledge for Sardinian) and a corpus (for which we won't have reproduction rights). We have been working on speech recognition in the context of the Common Voice platform; the Catalan community is achieving very positive results (even in the field of home assistants) thanks to the project. In the Sardinian case, the need for texts in the public domain is urgent. Multimedia initiatives in Sardinian (EjaTV, Sardegna1, to some extent even RAI) should be reminded of the crucial role of subtitling all their products and releasing the subtitles under permissive licenses, not only because of a pragmatic need, but also as part of a broader effort towards the hearing-impaired.

Sardinian Sign Language is something that has been mentioned lately by the regional government, but I could never identify existing efforts in this sense.

6) What priorities should a future programme on digital European language equality address in order to help support endangered languages? Openness (permissive licenses), collaborative approach (bringing grassroots LT activists in the discussion).