



EUROPEAN LANGUAGE EQUALITY

D2.1

Specification of the consultation process

Authors	Jan Hajič (CUNI), Maria Giagkou (ILSP), Stelios Piperidis (ILSP), Georg Rehm (DFKI), Natalia Resende (DCU)
Dissemination level	Public
Date	31-05-2021

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	D2.1
Deliverable title	Specification of the consultation process
Type	Report
Number of pages	37
Status and version	Final
Dissemination level	Public
Date of delivery	Contractual: 30-04-2021– Actual: 31-05-2021
Work package	WP2: European Language Equality – The Future Situation in 2030
Tasks	Task 2.1 The perspective of European LT developers (industry and research); Task 2.2 The perspective of European LT users and consumers; Task 2.3 Science – Technology – Society: Language Technology in 2030
Authors	Jan Hajič (CUNI), Maria Giagkou (ILSP), Stelios Piperidis (ILSP), Georg Rehm (DFKI), Natalia Resende (DCU)
Reviewers	Andy Way (DCU), German Rigau (UPV/EHU)
EC project officer	Aleksandra Wesolowska
Contact	<p>European Language Equality (ELE) ADAPT Centre, Dublin City University Glasnevin, Dublin 9, Ireland</p> <p>Prof. Dr. Andy Way – andy.way@adaptcentre.ie</p> <p>European Language Equality (ELE) DFKI GmbH Alt-Moabit 91c, 10559 Berlin, Germany</p> <p>Prof. Dr. Georg Rehm – georg.rehm@dfki.de http://www.european-language-equality.eu © 2021 ELE Consortium</p>

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 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	PANGAIC-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 Ülikool (University of Tartu)	UTART	EE
30	Helsingin Yliopisto (University of Helsinki)	UHEL	FI
31	Centre National de la Recherche Scientifique	CNRS	FR
32	Nyelvtudományi Kutatóközpont (Research Institute for Linguistics)	NYTK	HU
33	Stofnun Árna Magnússonar í íslenskum fræðum SAM (Árni Magnússon Inst. for Icelandic Studies)	SAM	IS
34	Fondazione Bruno Kessler	FBK	IT
35	Latvijas Universitātes Matemātikas un Informātikas institūts (Institute of Mathematics and Computer Science, University of Latvia)	IMCS	LV
36	Lietuvių Kalbos Institutas (Institute of the Lithuanian Language)	LKI	LT
37	Luxembourg Institute of Science and Technology	LIST	LU
38	Università ta Malta (University of Malta)	UM	MT
39	Stichting Instituut voor de Nederlandse Taal (Dutch Language Institute)	INT	NL
40	Språkrådet (Language Council of Norway)	LCNOR	NO
41	Instytut Podstaw Informatyki Polskiej Akademii Nauk (Polish Academy of Sciences)	IPIPAN	PL
42	Universidade de Lisboa, Faculdade de Ciências (University of Lisbon, Faculty of Science)	FCULisbon	PT
43	Institutul de Cercetări Pentru Inteligență Artificială (Romanian Academy)	ICIA	RO
44	University of Cyprus, French and European Studies	UCY	CY
45	Jazykovedný ústav Ľudovíta Štúra Slovenskej akadémie vied (Slovak Academy of Sciences)	JULS	SK
46	Institut Jožef Stefan (Jozef Stefan Institute)	JSI	SI
47	Centro Nacional de Supercomputación (Barcelona Supercomputing Center)	BSC	ES
48	Kungliga Tekniska högskolan (Royal Institute of Technology)	KTH	SE
49	Universität Zürich (University of Zurich)	UZH	CH
50	University of Sheffield	USFD	UK
51	Universidad de Vigo (University of Vigo)	UVIGO	ES
52	Bangor University	BNGR	UK

Contents

1. Introduction	1
2. Methodology	1
3. The Perspective of European LT Developers (Supply Side)	4
3.1. Stakeholders	5
3.2. Instruments	9
3.2.1. Survey	9
3.2.2. Interviews and focused consultation meetings	10
4. The Perspective of European LT Users and Consumers (Demand Side)	10
4.1. Stakeholders	11
4.2. Instruments	13
4.2.1. Survey	13
4.2.2. Interviews and focused consultation meetings	14
5. Predicting LT in 2030: Technology Deep Dives	14
6. Collecting Additional Input and Feedback	16
6.1. Instrument: Website	16
6.2. Instrument: Social Media	17
6.2.1. LinkedIn	17
6.2.2. Twitter	17
6.3. Instrument: Email Surveys	18
7. Summary and Next Steps	18
A. LT Developers: Survey Questions (Draft)	21
B. LT Users and Consumers: Survey Questions (Draft)	24
C. LT Users and Consumers: Additional Stakeholders	29

List of Figures

1.	The five work packages of the ELE project	2
2.	Overall project timeline	3
3.	European LT developers communities (excerpt)	8

List of Tables

1.	Stakeholder groups and instruments relevant for the three tasks in WP2	4
2.	LT developer communities (core ELE stakeholders)	6
3.	LT developer communities (additional ELE stakeholders)	6
4.	LT users and consumers (core ELE stakeholders)	11
5.	Technological advances necessary to achieve DLE in 2030 documented in four technology deep dives (all deliverables due in M14)	16
6.	Consolidated WP2 timeline covering Task 2.1, Task 2.2 and Task 2.3	19
7.	LT developers survey (draft)	21
8.	LT users and consumers survey (draft)	24
9.	LT users and consumers survey – additional stakeholders	30

List of Acronyms

AI	Artificial Intelligence
AI4EU	AI4EU (EU project, 2019-2021)
CEF AT	Connecting Europe Facility, Automated Translation
CLAIRE	Confederation of Laboratories for AI Research in Europe
CLARIN	Common Language Resources and Technology Infrastructure
COST	European Cooperation in Science and Technology
DLE	Digital Language Equality
DH	Digital Humanities
EC	European Commission
ECSPM	European Civil Society Platform for Multilingualism
EFNIL	European Federation of National Institutions for Language
ELE	European Language Equality (<i>this project</i>)
ELE Programme	European Language Equality Programme (<i>the long-term, large-scale funding programme specified by the ELE project</i>)
ELEN	European Language Equality Network
ELEXIS	European Lexicographic Infrastructure
ELG	European Language Grid (EU project, 2019-2022)
EP	European Parliament
EU	European Union
euRobotics	euRobotics AISBL
FET	Future and Emerging Technologies
GDPR	General Data Protection Regulation
ICT	Information and Communications Technology
LT	Language Technology/Technologies
META	Multilingual Europe Technology Alliance
META-NET	EU Network of Excellence to foster META
ML	Machine Learning

MT	Machine Translation
NLP	Natural Language Processing
NGO	Non-governmental organisation
PPP	Public-Private Partnership
RML	Regional and minority languages
SME	Small or medium-sized enterprise
SSH	Social Sciences and the Humanities
WP	Work Package

Abstract

This report, ELE Deliverable D2.1, provides a thorough specification of the consultation process to be carried out under the umbrella of WP2 (European Language Equality – The Future Situation in 2030). We prepared the report in such a way so that the description of our project-internal processes and instruments to be followed and applied can be used as guidelines and instructions for all actively involved partners, especially with regard to Task 2.1 (The perspective of European LT developers – industry and research) and Task 2.2 (The perspective of European LT users and consumers), but also with regard to Task 2.3 (Science – Technology – Society: Language Technology in 2030).

After a brief introduction (Section 1), we provide a description of the overall methodology (Section 2). The following two sections specify how the consortium will conduct the consultations with the European LT developers (Section 3) and European LT users (Section 4). Afterwards, Section 5 describes the preparation process of the four technology deep dives. Section 6 explains the different instruments that will be used for the collection of additional input and additional feedback. Section 7 concludes the report with a consolidated timeline of the consultation process.

1. Introduction

ELE WP2 (European Language Equality – The Future Situation in 2030) collects a vast amount of input for the strategic agenda and produces several reports by a broad and diverse spectrum of stakeholders – from research through industry to users – about their views, needs and perspectives related to Language Technology, Language-centric AI and Digital Language Equality (DLE), while anticipating the expected developments over the next ten years. WP2 and, by extension, this report, puts a special focus upon ways and means of achieving DLE until 2030 through the development, implementation and use of Language Technology, in order to make Europeans of all regions and origins truly equal when accessing and interacting with education, business, governments and public services through their own language. A large part of the information eventually to be integrated into the strategic agenda and roadmap will be collected through carefully designed surveys to researchers, developers, innovators and users and their communities as well as through reports that will be produced by a number of ELE partners. All WP2 results, the production of which is specified in this report, will be described and summarised and used as one of the major sets of inputs to the development of the strategic research agenda and roadmap as carried out in WP3.

We prepared the report in such a way so that the description of our project-internal processes and instruments to be followed and applied can be used as guidelines and instructions for all actively involved partners, especially with regard to Task 2.1 (The perspective of European LT developers – industry and research) and Task 2.2 (The perspective of European LT users and consumers), but also with regard to Task 2.3 (Science – Technology – Society: Language Technology in 2030).

2. Methodology

Our objective is the preparation of the strategic research, innovation and implementation agenda and roadmap for achieving full DLE in Europe by 2030. Since the goal of achieving DLE involves a large number of stakeholders, the process of preparing, discussing and finalising the different parts of the strategic agenda and roadmap is carried out together with all

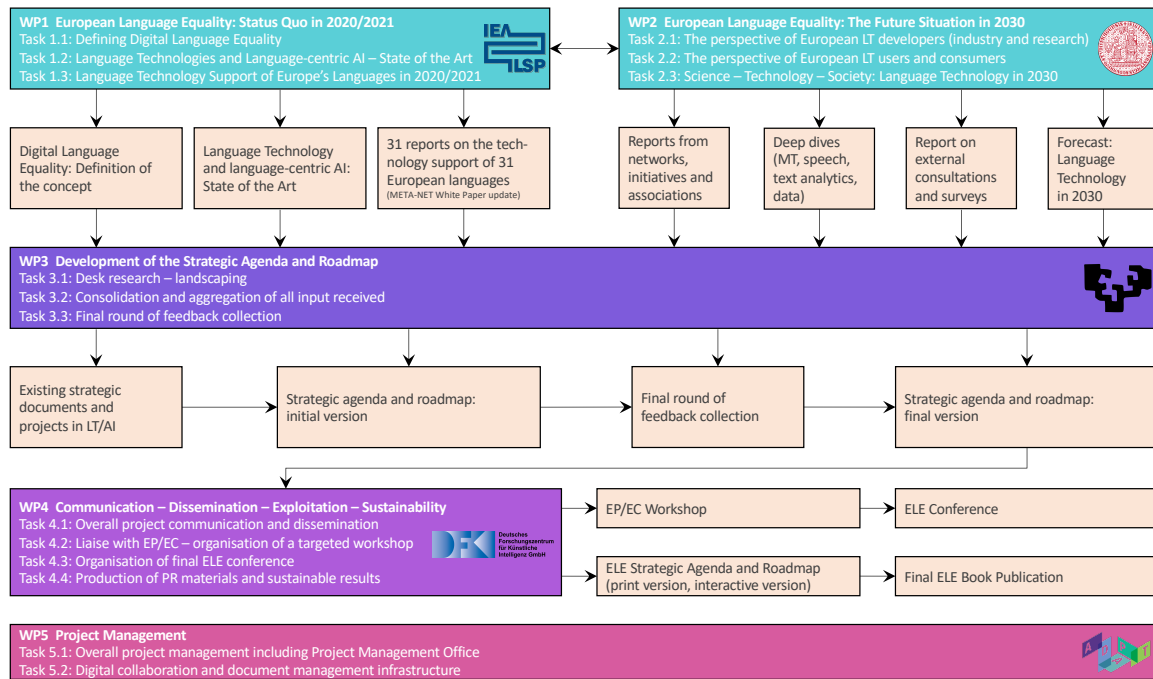


Figure 1: The five work packages of the ELE project

52 partners of the consortium and the wider European LT community, which we communicate with via the consortium partners' many strong networks and connections.

Only with the strong support of the consolidated European LT research and industry' community can the project be successful and produce a convincing, sustainable and evidence-based agenda and roadmap. Only with the input and feedback from experts working in different areas of our core field of Computational Linguistics and Language Technology and also on the borders to other fields such as, among others, Cognitive Science, AI, Machine Learning, Data Science and Knowledge Technologies, can the agenda and roadmap be prepared in a way that is goal-oriented, all-encompassing, realistic, supported and overall meaningful. Only with the inclusion of representatives from the many different companies active in the field can the involvement of industry make sense in the grander scheme of things, especially regarding the inclusion of their needs and wishes and goals towards growth. The same holds for the non-industrial, but strong stakeholders as users and consumers, in the areas such as Digital Humanities/Social Science and Humanities (DH/SSH) research, policymaking, normative care of languages (including minority ones), education and others.

At the most abstract level, our main approach is two-fold: we distinguish between input for the agenda and roadmap generated within the consortium, and input generated by organisations not participating as partners in the ELE project (through surveys, interviews, external consultation meetings, etc.). When composing the consortium, we opted for a large number of partners that cover many relevant areas that need to be taken into account for the development of the agenda and roadmap. The consortium-internal or consortium-external stakeholders' input and feedback will be systematically collected, structured and included in the agenda and roadmap development process so that we can deliver, at the end of the project, a coherent and convincing strategic roadmap with agreed upon research questions and research goals, realistic timing, and a meaningful plan.

Figure 1 shows the five work packages of the ELE project and their main tangible outcomes in terms of reports and deliverables and how they are used as input in later stages towards

the development of the final strategic agenda and roadmap. While WP1 describes the current state of play in terms of European Language Equality in 2020/2021, WP2 describes the future situation in 2030. WP3 develops the strategic agenda and roadmap and WP4 takes care of communication, dissemination, exploitation and sustainability. Project management and the project management office are handled by WP5.

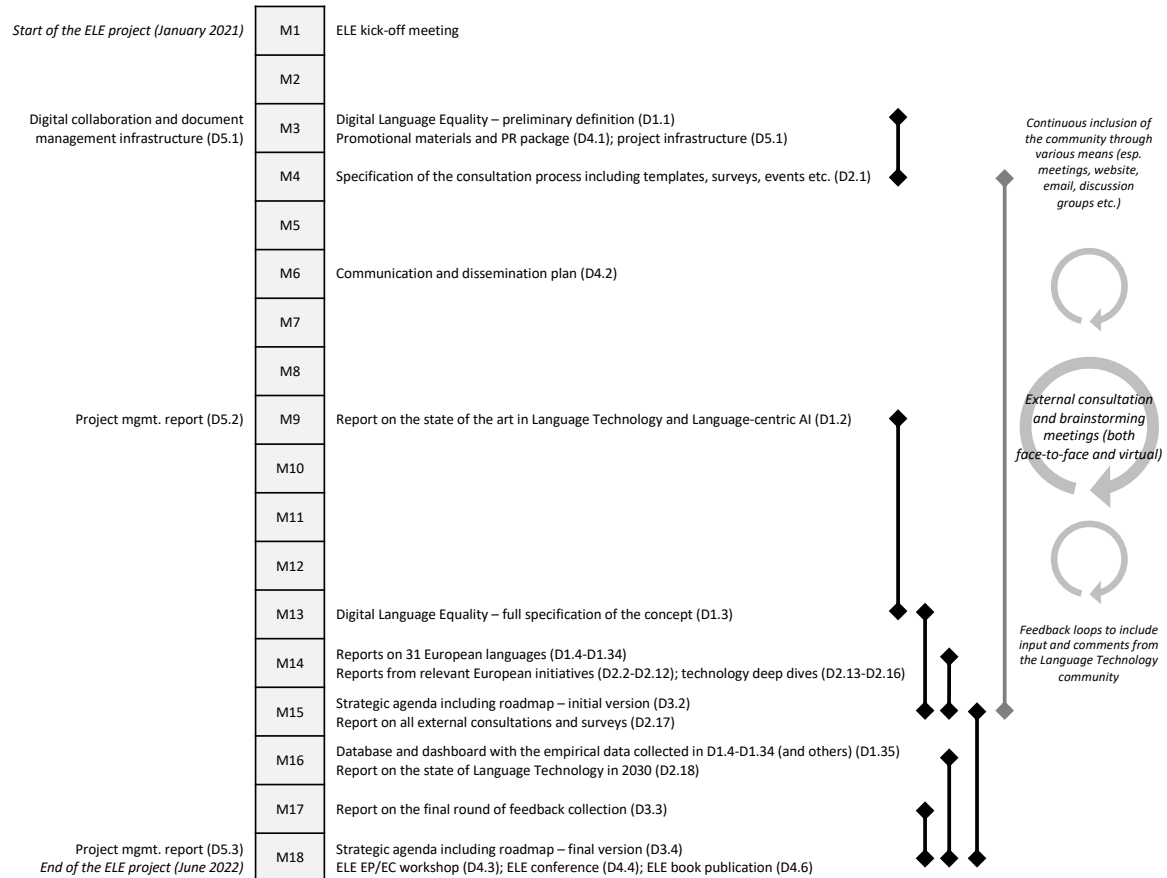


Figure 2: Overall project timeline

WP2 (European Language Equality: The Future Situation in 2030) describes the future situation, i. e., DLE in 2030. We distinguish between two main stakeholder groups: LT developers (industry and research) and LT users as well as consumers. Both groups are represented in ELE with several networks, initiatives and associations who produce one report each, together with their respective constituencies, highlighting their own individual needs, wishes and demands towards DLE. The industry partners of the consortium generate, in various tandem groups, four technology deep dives to provide the needs, wishes, demands and visions of European industry, structured into Machine Translation, Speech, Text Analytics and Data. WP2 carries out a large number of additional surveys and consultation meetings with stakeholders who are not represented in the consortium. These will be documented in a report. The last deliverable of WP2 is a technology forecast that attempts to predict what Language Technology will look like in 2030 (including capabilities, deployment, etc.).

The methodology applied in WP2 is, thus, based on a number of stakeholder-specific surveys (inspired by Rehm and Hegele, 2018) as well as collaborative document preparation (“deep dives”) that also involves technology foresighting. Both approaches are complemented

Nature of Stakeholder Group	Instruments
<i>Task 2.1 The perspective of European LT developers (industry and research)</i>	
European LT developers (industry and research): <i>Closed set</i> that is well represented through relevant organisations, networks and initiatives in the ELE consortium	Surveys, Interviews
⇒ Approach further detailed in Section 3	
⇒ Results to be reported in deliverables D2.2-D2.6	
<i>Task 2.2 The perspective of European LT users and consumers</i>	
All potential European LT users: <i>Open set</i> that is only partially represented through relevant organisations, networks and initiatives in the ELE consortium	Surveys, Interviews
⇒ Approach further detailed in Section 4	
⇒ Results to be reported in deliverables D2.7-D2.12	
<i>Task 2.3 Science – Technology – Society: Language Technology in 2030</i>	
Prominent companies of the European LT developer landscape, all represented in the ELE consortium: <i>Closed set</i>	Collaboratively created technology deep dives
⇒ Approach further detailed in Section 5	
⇒ Results to be reported in deliverables D2.13-D2.16	

Table 1: Stakeholder groups and instruments relevant for the three tasks in WP2

through the collection of additional input and feedback through various online channels. As Table 1 illustrates, the two stakeholder groups addressed in Task 2.1 and Task 2.2, respectively, differ in one substantial way: while the group of commercial or academic LT developers (Task 2.1) is, in a certain way, closed and well represented through relevant organisations, networks and initiatives in the ELE consortium, the group of LT users (Task 2.2) is an open set of stakeholders that is only partially represented through relevant organisations, networks and initiatives in our consortium. Both stakeholder groups will be addressed with targeted and stakeholder-specific surveys that will be distributed to the relevant stakeholders through the responsible ELE partners (Task 2.1 and, partially, Task 2.2). In addition, we will communicate with a large number of additional stakeholders under the umbrella of Task 2.2. For this purpose, we will use shorter surveys and targeted online meetings.

3. The Perspective of European LT Developers (Supply Side)

One crucial and mission-critical aspect of WP2 is the collection of input, demands, needs and visions with regard to the wider topic of digital language equality from the community of European LT developers and also the highly diverse group of European LT users. This section, which corresponds to Task 2.1, describes our process for engaging with the LT developers (supply side) while Section 4, corresponding to Task 2.2, describes how we will collaborate and communicate with the LT users (demand side).

We will analyse the views of European LT developers and providers, i. e., representatives both from research and industry to investigate the ideas, demands, visions and predictions of this group of stakeholders with regard to digital language equality. We will explore the

factors that drive their development plans and investments (e.g., market demand, number of speakers, available funds etc.) and the perceived obstacles that should be overcome to achieve language equality. The main instrument for collecting the LT developers' views are surveys, which will be distributed through the established research and industry networks of the ELE consortium to their members. In addition to these networks, the survey will be forwarded to other pan-European initiatives, thus covering the widest possible range from generic AI to media and language-related infrastructures. The data collection activity will, if deemed necessary, be supplemented by focused meetings and consultations (virtual or face-to-face) with targeted informants to be selected based on either the quality of their input to the survey or their prominence in and impact for the European LT landscape. The collected feedback and views of the European LT developers will be augmented with additional input produced by the networks, analysed and consolidated in five different reports (D2.2-D2.6).

3.1. Stakeholders

The European LT developers are a diverse but still, especially compared to the LT users, rather clear-cut group of stakeholders, comprising *academic and industrial researchers* in the field of LT/NLP – beyond research, they develop pre-commercial prototypes, algorithms, applications and systems as well as *innovators and entrepreneurs* who productise and commercialise LT to address, among others, the needs for digital content analysis and generation, as well as for pertinent content transformation and dissemination. An initial grouping is, thus, *LT research* and *LT industry*.

Europe has a long-standing research, development and innovation tradition in LT with over 800 centres performing excellent, highly visible and internationally recognised research on all European and many non-European languages. The European LT industry has been estimated to comprise 435 companies, according to LT-Innovate (2016) or 473 LT vendors in EU26 plus Iceland and Norway in 2017 (Vasiljevs et al., 2019). In May 2021, the ELG catalogue comprises more than 800 commercial entities, also including integrators and a certain number of user companies (Rehm et al., 2020, 2021).

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

With the aim of informing the ELE strategic agenda and roadmap with the opinions, views and demands of the widest possible, directly or indirectly involved group of stakeholders, in the consultation process we will engage with research and industry and, with regard to their field, with LT proper, AI and DH/SSH communities. We will mobilise existing European networks, associations, initiatives and projects. Some of the well-established and long-standing pan-European LT networks are represented in the ELE consortium and they constitute the core ELE LT developers stakeholders groups (Table 2). The ELE partners that represent these initiatives will contribute their views to the project and also facilitate access to and elicitation of the views of their constituency and members with regard to DLE. In particular, they will coordinate the distribution of a questionnaire to their members, conduct interviews and focused consultation meetings, where needed and appropriate (see Section 3.2), and consolidate and report back on their feedback in D2.2-D2.6.

While these stakeholders already represent a significant part of the European LT community, we will engage additional initiatives and networks in the consultation process (see Ta-

Initiative	Description	Stakeholder group	Represented by ELE consortium partner(s)
META-NET	The META-NET Network of Excellence consists of 60 research centres in 34 European countries. It develops the technical foundations of a multilingual, inclusive and innovative European society, supporting all European languages.	European LT community (esp. research)	DFKI, CUNI, ILSP, TILDE and others
ELG	The European Language Grid (ELG) project develops a cloud platform and marketplace for the whole European LT community. The shared platform includes language resources, datasets and services to benefit European society and industry. It addresses the fragmentation of the European LT landscape.	European LT community	DFKI, CUNI, ILSP and others
LT-Innovate	LT-Innovate is the European LT industry association with more than 200 members. It supports its members by promoting the industry as a whole in the most promising target markets.	European LT industry	CRSLNG (HENS, EX-PSYS, TILDE are members)
CLARIN	The European Research Infrastructure for Language Resources and Technology consists of more than 20 national consortia, which themselves consist of multiple partners. CLARIN makes language resources available to researchers and students from all disciplines, especially in the humanities and social sciences, through single sign-on access.	European DH/NLP/SSH community	CLARIN ERIC (CUNI and ILSP are members)
CLAIRE	The Confederation of Laboratories for AI Research in Europe has 394 members in 36 countries. CLAIRE seeks to strengthen European excellence in AI research and innovation across all of AI, for all of Europe, with a human-centred focus. It is now supported by nine EU Member State governments.	European AI community	ULEID

Table 2: LT developer communities (core ELE stakeholders)

ble 3). Some of them are represented in the consortium while others are not. In all cases we will utilise existing connections or other means (see Section 6) to reach out to these networks, initiatives or projects. The feedback will be collectively reported in D2.17.

Table 3: LT developer communities (additional ELE stakeholders)

Initiative	Short description	Stakeholder group	Represented by ELE consortium partner(s)
ELEXIS	The European Lexicographic Infrastructure is a H2020 project (2018-2022) dedicated to lexicography. It creates an infrastructure which will enable efficient access to high quality lexicographic data, and bridge the gap between more advanced and less-resourced scholarly communities working on lexicographic resources.	European DH/NLP/SSH community	JSI, UCPH, IBL, NYTK

Continued on next page

Table 3 – Continued from previous page

Initiative	Short description	Stakeholder group	Represented by ELE consortium partner(s)
BDVA	The Big Data Value Association is an industry-driven international not-for-profit organisation with more than 200 members from all over Europe and a well-balanced composition of large, small, and medium-sized industries as well as research and user organizations. The EC and BDVA implement the Big Data Value PPP.	European AI community, esp. data	SAP, BSC, DCU, DFKI
AI PPP	The Public-Private Partnership on AI is an initiative of two industry-driven associations, BDVA and euRobotics, who have recently joined forces and started a collaboration towards their vision of establishing a European PPP on AI, Data and Robotics.	European AI community	SAP, Tilde
AI4EU	The European AI on Demand Platform is an EU-funded project in which 81 partners from 21 countries develop a platform that bundles and connects European AI resources with the goal of ensuring European independence and leadership in AI.	European AI community	BSC, CNRS, DFKI, FBK, JSI, SAP, Tilde
HAI-Net	The HumanE AI Network is an EU-funded project that brings together leading European research centres, universities and industrial enterprises into a network of centres of excellence.	European AI community	DFKI, Athena RC, CNRS, FBK and others
ELISE	The European Learning and Intelligent Systems Excellence is an EU-funded project that aims to increase Europe's competitiveness in ML and AI.	European AI community	–
AI4Media	The European Excellence Centre for Media, Society and Democracy is an EU-funded project that focuses upon core AI advances to serve the media sector.	European AI community	UM, BSC
TAILOR	Foundations of Trustworthy AI – Integrating Reasoning, Learning and Optimization is an EU-funded project that establishes a scientific network that focuses upon trustworthy AI, reducing the fragmentation and increasing the joint AI research capacity of Europe.	European AI community	ULEI, CNRS, DFKI, FBK
VISION	The Value and Impact through Synergy, Interaction and cooperation of Networks of AI Excellence Centres is an EU-funded project which is connecting and strengthening AI research centres across Europe.	European AI community	ULEI, FBK
AI4Copernicus	This EU project reinforces the AI4EU Platform by Advancing Earth Observation Intelligence, Innovation and Adoption. It makes use of the AI4EU platform for users of Copernicus data (scientists, SMEs, non-tech sector).	European AI community	–
AIPlan4EU	The EU project Bringing AI Planning to the European AI On-Demand Platform integrates AI planning as a first-class citizen into the AI4EU platform.	European AI community	FBK, CNRS, DFKI
BonsAPPs	This EU project develops a scalable AI-as-a-Service layer that will be interoperable with the AI4EU platform as an external service, enhancing an existing AI platform (Bonseyes Marketplace).	European AI community	–
DIH4AI	The AI on-demand platform for regional interoperable Digital Innovation Hubs Network project builds a network of AI innovation and collaboration platforms.	European AI community	–

Continued on next page

Table 3 – Continued from previous page

Initiative	Short description	Stakeholder group	Represented by ELE consortium partner(s)
I-ENERGY	The AI for Next Generation Energy project aims at evolving, scaling up and demonstrating innovative AI energy analytics applications and digital twins services.	European AI community	DFKI
StairwAI	The Ease the Engagement of Low-Tech users to the AI-on-Demand platform through AI project targets low-tech users with the goal of facilitating their engagement on the AI4EU Platform through a new service layer enriching the functionalities of the platform.	European AI community	Tilde
Nexus Linguarum	This COST action aims at promoting synergies across Europe between linguists, computer scientists and terminologists in industry and society to investigate and extend the area of linguistic data science.	European DH/NLP/SSH community, esp. data	ILSP, UVIE and others
EC, DG Translation	DG Translation is the EC's in-house translation service, they also provide the eTranslation Machine Translation platform used by the EU, member states public organisations and European SMEs.	European LT community-EU body	–
EC JRC	The Joint Research Centre is the EC's science and knowledge service which employs scientists to carry out research to provide independent scientific advice and support to EU policy. The two relevant groups are the Competence Centre on Text Mining and Analysis and the HUMAINT project, an interdisciplinary project aiming to understand the impact of AI on human behaviour.	European LT community (EU body)	–

The stakeholders described in this section, either participating in the ELE consortium or not, cover to a large extent the European LT community as partially visualised in Figure 3.

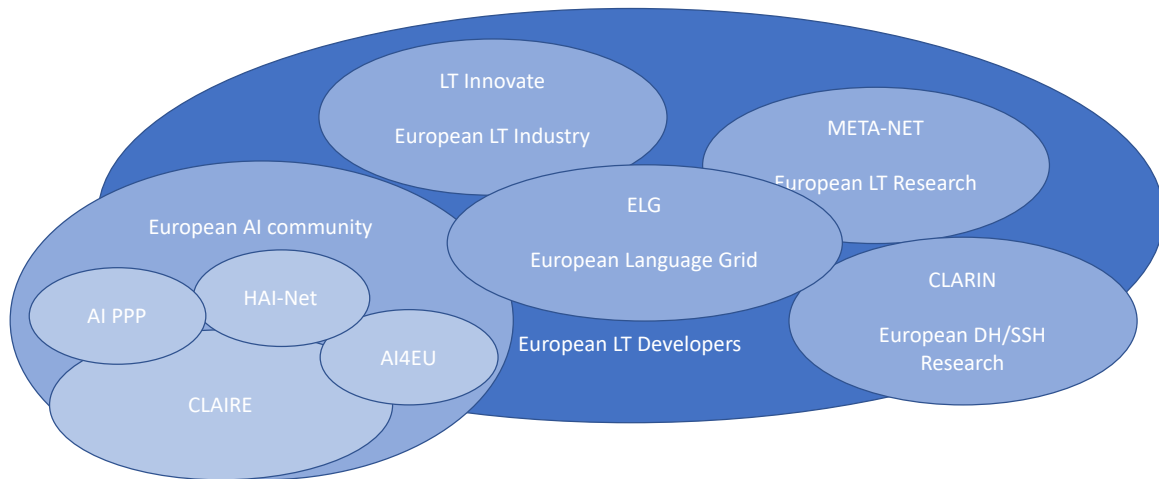


Figure 3: European LT developers communities (excerpt)

3.2. Instruments

To collect and analyse the LT developers' views, demands, visions and predictions, we adopt an inclusive and participatory approach, through which every voice will find its way into the strategic agenda. We reach out to as many representatives of the LT community as possible and elicit their educated views in a structured, yet unconstrained, way. Two main instruments will be used to collect the views of the European LT developers: Surveys and questionnaires (Section 3.2.1) as well as interviews and focused consultation meetings (Section 3.2.2).

3.2.1. Survey

The LT developer survey seeks to elicit views in a structured way that lends itself to the efficient analysis, consolidation and integration of the feedback in the respective WP2 deliverables (D2.2-D2.6, D2.17) which, in turn, will feed into the strategic agenda and roadmap. Driven by the envisaged topics that the final strategic agenda will cover, the survey encompasses closed and open-ended questions to inquire about the LT developers' views of the current situation and their future predictions and visions. The survey has four parts.

- **Part A. Respondents' profiling:** The first part includes questions regarding the demographic profile¹ of respondents with an emphasis on characteristics relevant to the task at hand, i. e.,
 - country
 - professional experience and role
 - affiliation
 - type of organisation
 - LT areas that the respondent is mainly active in
 - participation or membership in networks or associations
 - sectors or domains that the respondent is active in (if relevant)
- **Part B. Language coverage:** This part investigates the degree of coverage of the European languages by the respondents' current research and development activities, i. e.,
 - languages currently supported in research, products or services
 - languages planned to be supported in the short- or medium-term (3-10 years)
 - factors that influence the respondents' decision with regard to language coverage or support
- **Part C. Evaluation of current situation:** This part includes questions regarding the respondents' evaluation of the current situation of LT research and development, the strengths, gaps and challenges that the European LT community is facing, i. e.,
 - gaps in terms of: a) technologies, b) tools or applications, and c) resources, especially with regard to specific languages
 - main perceived challenges and obstacles that should be overcome
- **Part D. Predictions and visions for the future:** The last part is forward-looking and investigates ideas, predictions and wishes of the LT community about how the LT field as a whole will be able to equally support all European languages by 2030, i. e.,

¹ The ELE project is fully GDPR-compliant. Demographic data will be used for research purposes only and destroyed as soon as the project is over.

- prediction of future opportunities for LT in basic and applied research (scientific vision) and in innovation and industry
- policies or instruments that could contribute to speed up the effective deployment of LT in Europe equally for all languages
- expectations with regard to the challenges the ELE programme can address by 2030

The survey will be circulated through the members of the networks and associations identified and described in Section 3.1 and through the additional channels described in Section 6. It will be set up as an online form for easy distribution as well as collection and analysis of responses. Appendix A (p. 21 ff.) contains the complete pre-final survey.

3.2.2. Interviews and focused consultation meetings

To supplement the survey responses and to collect more detailed feedback, where appropriate, we will conduct consultation meetings with targeted informants who will be selected based on either the quality of their input to the survey or their prominence in and impact on the European LT landscape. Operationally, the selection of stakeholders to be interviewed will be based on the following criteria.

1. The respondent has partially filled in the survey and some essential input is missing in order to have a more complete understanding of his/her views; or
2. No member of a network/association identified in Section 3.1 has filled in the survey.

In the first case, we will ask for a short and focused meeting with the respondent to elicit the missing information, based on the answers the respondent has already provided in the survey. In the second case, when a network or association that is considered a stakeholder for ELE is not represented, we will identify key persons and conduct an interview, the script will follow the structure of the survey and utilise the questions from parts B, C and D.

4. The Perspective of European LT Users and Consumers (Demand Side)

Complementing Section 3, in the following we describe our approach at gathering the voices of the highly heterogeneous and diverse group of European LT users and consumers as the final beneficiaries of LT with regard to the necessary and desired developments supporting digital equality for all European languages over the next decade. This activity will require engagement with individuals, representative public bodies and government units, organisations and businesses, including SMEs as well as larger companies, that use LT. We will investigate the expectations, anticipated developments, needs and wishes of this diverse group of end-user stakeholders, to make sure that their voices play a decisive role in the pursuit of digital language equality supported by LT. We will also explore the factors that can promote language equality in the users' and consumers' view, especially with regard to encouraging the uptake of missing or poor LT that can solve real communication problems for the members of all European language communities. Special attention will be paid to the speakers of lesser-served languages, particularly those that face digital extinction or neglect, eliciting from the LT users of such language communities indications of necessary or desirable developments that are expected to put their own languages on equal footing with the dominant ones by 2030; a complementary focus will consider the perceived obstacles that hinder full digital language equality, so that effective remedial action can be promptly taken. We

will follow the same approach as for the supply side (Section 3), i.e., based on surveys and structured templates several reports will be produced by the relevant ELE consortium members who represent relevant stakeholder groups. Additional feedback will be collected from external stakeholders through surveys.

4.1. Stakeholders

LT users and consumers comprise an expansive group of stakeholders from a wide variety of domains and sectors. We will reach out to individuals, representatives from public administration (public bodies and government units), organisations and businesses, including SMEs as well as larger companies, that currently use and benefit from LT. Six relevant stakeholders were already identified in the ELE proposal writing stage with a special focus upon speakers of lesser-served languages, particularly those that face digital extinction or neglect. Table 4 shows a summary of these six ELE consortium members.

Table 4: LT users and consumers (core ELE stakeholders)

Initiative	Short description	Stakeholder group	Represented by ELE consortium partner(s)
ECSPM	The European Civil Society Platform for Multilingualism is an alliance for the languages spoken in Europe (national/official, minority, regional and autochthonous, as well as the languages of immigrant communities). It includes networks of more than 200 European associations, societies and organisations that view multilingualism as an asset for European economic, social and cultural development, as well as a facilitator for intellectual and personal growth. It is a fervent voice of Europe's civil society promoting languages, language policies and research on multilingualism, by way of focusing on people, and on their ability to use a variety of semiotic resources to access education, social affairs and culture, to participate as active citizens in Europe, shaping its making, to benefit from better communication, wider employment and study opportunities.	European Platform for Multilingualism	ECSPM
EFNIL	The European Federation of National Institutions of Language is a pan-European organization that was founded in 2003. EFNIL has 41 members from 27 countries and provides a forum for these institutions to exchange information about their work and to gather and publish information about language use and language policy within the EU. In addition, EFNIL encourages the study of the official European languages and a coordinated approach towards mother-tongue and foreign-language learning, as a means of promoting linguistic and cultural diversity within the EU.	European National Languages	EFNIL

Continued on next page

Table 4 – Continued from previous page

Initiative	Short description	Stakeholder group	Represented by ELE consortium partner(s)
ELEN	The European Language Equality Network is the international NGO for the protection and promotion of European lesser-used languages gathering 166 member organisations representing 46 languages in 23 European states. Founded in 2012, it represents the voice of grassroots European RML civil society.	European Regional, Minority and Endangered Languages	ELEN
LIBER	The Association of European Research Libraries is Europe's principle association of research libraries, consisting of nearly 450 national, university and other libraries from more than 40 countries. It was established in 1971 and became a Foundation under Dutch law in 2009. LIBER helps European research libraries to support a functional network across national boundaries to ensure the preservation of European cultural heritage, to improve access to collections in European research libraries, and to provide more efficient information services. Enabling Open Science is a major priority, as is promoting innovative scholarly communication, fostering digital skills and services, and engaging with world-class e-infrastructures.	European Research Libraries	LIBER
NEM	New European Media is the leading European Network for Media and Creative Industries with the mission to foster the impact of interactive technologies on the future of new media through interaction between media, content, creative industries, social media, broadcasting and telecom sectors as well as consumer electronics, represented by more than 1,000 members. NEM wants to develop a common innovation environment for the new European media landscape. The application of the newest technologies in respect to equal access to media for all is one of its higher priorities.	European New Media Community	ERSCM
Wikipedia	Wikimedia Deutschland is an independent, charitable membership-based non-profit organization that serves as the German chapter of the global Wikimedia movement. With more than 140 employees it is the oldest and largest of approx. 40 independent chapters.	European Free Knowledge Community	WMD

In addition to the deliverables to be produced by these six core representative bodies and ELE partners, other relevant external stakeholders will be consulted as well. Including additional groups will ensure the widest possible coverage and promote our inclusive approach to build a comprehensive, accurate and all-encompassing picture. While these external stakeholders will (obviously) not provide reports, their feedback will be regarded as equally important. It will be collected through surveys and focused meetings. A current snapshot of our list of potential additional LT user stakeholders can be found in Appendix C. This list is not meant to be definitive, as more stakeholders can be added or removed. Modifications of this list will be discussed by the ELE core group.

4.2. Instruments

In a similar way as already described in Section 3.2 for the stakeholder class of LT developers, surveys and focused consultation meetings will be used to collect and analyse the perspective of European LT users and consumers, i. e., the views, ideas, demands, future visions and predictions of LT users and consumers with regard to digital language equality. Our goal is to consult with as many representatives of this stakeholder class as possible to collect their opinions in a structured, yet unconstrained, way.

4.2.1. Survey

Similarly to the survey for the LT developers, feedback from the LT users and consumers will be collected in a structured way that lends itself to the efficient analysis, consolidation and integration of the feedback in the respective WP2 deliverables (D2.7-D2.12, D2.17) which, in turn, will feed into the strategic agenda and roadmap. Driven by the envisaged topics that the final strategic agenda will cover, the survey encompasses closed and open-ended questions to understand the LT users and consumers' views of the current situation and their future predictions and visions with regard to digital language equality. The survey has four parts.

- **Part A. Respondents' profiling:** The first part includes questions regarding the demographic profile¹ of respondents with an emphasis on characteristics relevant to the task at hand, i. e.,
 - country
 - professional experience and role
 - affiliation
 - type of organisation they work for
 - sector, segment or domain they represent
- **Part B. Language coverage:** The second part investigates the level of coverage LT users of the different European languages experience, i. e.,
 - languages they currently work with
 - languages they plan to work with
 - whether they work with minority, regional or lesser-used languages
- **Part C. Evaluation of current situation:** This part includes questions regarding the respondents' evaluation of the current situation of the LT tools they use, their gaps (if any), technological support for the languages they work with and whether the gaps they perceive are related to specific languages they work with.
- **Part D. Predictions and visions for the future:** The last part is forward-looking and investigates ideas, predictions and wishes of the LT users and consumers about how LT providers could achieve language equality by 2030.
 - policies or instruments that could contribute to speed up the effective deployment of LT in Europe equally for all languages
 - expectations with regard to the challenges an ELE programme can address by 2030

The survey will be circulated through the members of the networks and associations identified and described in Section 3.1 and through the additional channels described in Section 6. It will be set up as an online form for easy distribution as well as collection and analysis of responses. Appendix B (p. 24 ff.) contains the complete pre-final survey.

4.2.2. Interviews and focused consultation meetings

To complement the survey responses of the six LT user and consumer stakeholder groups represented in the ELE consortium, we will conduct a substantial number of consultation meetings with targeted informants who will be selected from the list of relevant additional stakeholder groups (see Appendix C), as described in Section 4.1.

Furthermore, to collect more detailed feedback we will conduct additional interviews and consultation meetings with informants based on either the quality of their input to the original survey or their prominence in and impact on the European LT landscape. Operationally, the selection of stakeholders to be interviewed will be based on the following criteria.

1. The respondent has partially filled in the survey and some essential input is missing in order to have a more complete understanding of his/her views; or
2. No member of a network/association identified in Section 3.1 has filled in the survey.

In the first case, we will ask for a short and focused meeting with the respondent to elicit the missing information, based on the answers the respondent has already provided in the survey. In the second case, when a network or association that is considered a stakeholder for ELE is not represented, we will identify key persons and conduct an interview, the script will follow the structure of the survey and utilise the questions from parts B, C and D.

5. Predicting LT in 2030: Technology Deep Dives

The targeted, forward-looking Task 2.3 (Science – Technology – Society: Language Technology in 2030) aims at collecting, analysing and consolidating the views of European LT industrial and academic stakeholders on future technological progress, innovations and impact on society ten years from now, with a special emphasis on technologies, resources, approaches, coverage and performance needed to achieve DLE by 2030.

The task is set up to seek an agreement among these stakeholders to pinpoint novel or significantly extended or adapted technologies that will ultimately enable or contribute to DLE, and consequently enable true equality in European society. To achieve these goals, such new technologies will have to take into account the state-of-the-art in various LT and AI areas, including the reasons why current technologies do not perform equally or near equally well for all languages (e. g., due to lack of data, poor-quality data, language properties, knowledge collectively and indirectly acquired for only some languages in the past etc.) as well as the reasons for biased results in some areas. Focusing on possible methods, technologies and processes for bringing all European languages on par both technologically and in consumer usage, there will be a unifying theme, namely, to discover and explore ways to convert the unique challenges of a diverse European multilingual society into opportunities and technologies, processes and services superior to those developed in the context of largely homogeneous linguistic societies.

We will also take a fresh look at deployment, i. e., how LT will be made available to the different stakeholders and end-users, from machines to household appliances to mobile devices and maybe even “invisible” devices. To achieve these goals, structured document templates and also surveys oriented towards technological development and technology forecasting along the aforementioned lines (complementary to Task 2.1 and Task 2.2, see Sections 3 and 4, respectively) will be worked on by both industrial and research stakeholders, and then assembled into deliverables D2.13-D2.16, reflecting the major technology areas (Machine Translation; Speech-related technologies; Text Analytics and Text and Data Mining; Data creation, annotation, curation, and preservation; Table 5). The final report of WP2 (D2.18) will summarise the landscape and will also attempt a forecast of the situation in 2030

with regard to the question how citizens and the European society make use of ICT and LT in 2030 and how a genuinely “European approach” can be developed until then.

Four ELE partners that are among the European leaders in their respective areas were selected to lead the development of the four technology deep dives. They will collaborate closely with those ELE partners that also work in the respective field (see Table 5). All deep dives will be developed using the same efficient yet flexible methodology. First, a structured document template is provided to the lead deliverable partners who can then extend the structuring and provide additional levels of detail. Informed by the intermediate results from Tasks 2.1 and 2.2 (as well as other tasks in WP1 and WP3) and guided by the ELE core partners, the lead deep dive partners will then develop the content, together with the collaborating partners involved in the respective deep dives. They can also (possibly) make use of additional instruments and materials, such as surveys and also structured and targeted interviews to make sure that the documents include a broad representation of voices and stakeholders but avoiding reduplication of work already carried out in other ELE tasks, especially Tasks 2.1 and 3.1. Furthermore, the involved partners will make use of existing studies, reports and foresight studies as well as science and technology predictions (provided by D3.1). The idea is that the respective groups of experts develop a consolidated opinion as regards the direction in which the relevant field is moving, what the current gaps and roadblocks as well as industry’s needs from research are, and what they can contribute to DLE.²

The deep dives will comprise the following sections.³ By adhering to this high-level structure and to the indicative number of pages indicated for each section, we can make sure that the four deliverables will fit the eventual structure of the strategic agenda and roadmap.

1. Introduction (2 pages)
2. Scope of this Deep Dive and Relationship to the others (2 pages)⁴
3. *Technology Field*: Main Components (6-8 pages)
4. *Technology Field*: Current State of Play (6-8 pages)⁵
5. *Technology Field*: Main Gaps (4-6 pages)
6. *Technology Field*: Contribution to DLE and Impact on Society (6-8 pages)
7. *Technology Field*: Main Breakthroughs Needed (4-6 pages)
8. *Technology Field* in 2030: Main Technology Visions and Development Goals (6-8 pages)⁶
9. *Technology Field* in 2030: Towards Deep Natural Language Understanding (6-8 pages)
10. Conclusions (2-3 pages)

² This overall approach has been inspired by the methodology followed in META-NET, in which a small number of vision groups has worked on similar documents (“vision papers”, see Rehm and Uszkoreit, 2013). The ELE approach is shorter and significantly more targeted due to the umbrella goal of achieving DLE as well as due to the clear distribution of responsibilities.

³ In the outline, *Technology Field* is a placeholder for the four different technology fields (see Table 5).

⁴ The ELE core team will provide guidance with regard to this distinction.

⁵ This part will be informed by Task 1.2 and Task 3.1.

⁶ These main technology visions and development goals need to be powerful, convincing, visionary, long-term and also realistic as well as achievable. The visions need to serve the goal of achieving DLE and also address the various societal and economical challenges Europe and the EU are facing. In that regard, we will attempt to develop a “European way” of doing LT/NLP/Speech research and development, for example, “Human-centric LT”, “Inclusive LT”, “Deep LT”. This conceptualisation will be aligned with AI and it will also be grounded in areas, in which Europe is considered a clear leader, for example, Knowledge Graphs and Linked Data as one technical foundation (combination of symbolic and sub-symbolic methods, combination of Knowledge Graphs and Deep Learning approaches etc.).

	Technology Field	Responsible	Contributing Partners	Instruments
D2.13	Machine Translation	TILDE	CRSLNG, KNTN, PAN, SAP, DCU, CUNI, UPV/EHU	Template, Surveys
D2.14	Speech technologies (incl. conversational systems etc.)	HENS	CRSLNG, UPV/EHU	TILDE, Template, Surveys
D2.15	Text Analytics, Text and Data Mining, Natural Language Understanding	EXPSYS	TILDE, SAP, ILSP, UPV/EHU	Template, Surveys
D2.16	Data, Language Resources, Knowledge Graphs (data creation, annotation, curation, preservation, representation)	SWC	DFKI, ELDA, ONTO, ILSP, WMD, UPV/EHU	Template, Surveys

Table 5: Technological advances necessary to achieve DLE in 2030 documented in four technology deep dives (all deliverables due in M14)

Finally, we are not ruling out the possibility that a small number of additional technology deep dives will be developed under the umbrella of ELE. Likely candidates are deep dives on AI and/or machine learning, multimodal processing, robotics, deep dives on the use of LT in media or (say) journalism or perhaps also a deep dive on the deployment of LT (AI DevOps, MLOps, dynamic machine learning etc.). In that regard, we will explore the possibility of collaborating with the relevant networks and initiatives, such as CLAIRE (for LT and AI) or AI4MEDIA (for LT and media). Additional deep dives will only be developed if deemed necessary and also feasible by all partners involved.

6. Collecting Additional Input and Feedback

Complementing the instruments described in the previous sections, we will set up additional ways to collect input for and feedback to the emerging strategic agenda and roadmap and to the intermediate results of the project. In terms of the rationale for implementing these additional instruments, we want to enable all interested stakeholders to get in touch with the project easily and efficiently so that their voices can be heard and their opinions and ideas integrated into our final strategic agenda and recommendations. The emerging results of the project will be disseminated through various channels (e.g., project website, publications, talks and presentations, social media etc.) so that we are eager to keep a number of incoming communication channels open constantly. This relates to two main ways of soliciting external feedback, either by actively asking involved stakeholders for feedback (e.g., feedback to certain draft documents), or by actively listening, especially on social media, to identify additional opinions regarding our topic of DLE including LT and language-centric AI. The three main instruments described below are the ELE website (Section 6.1), the ELE social media channels (Section 6.2), and email surveys (Section 6.3).

6.1. Instrument: Website

An interactive contact form is already available on the ELE website through which interested stakeholders can get in touch with the ELE consortium.⁷ A similar interactive form is

⁷ <https://european-language-equality.eu/contact/>

currently being implemented for deployment through the upcoming online list of languages under investigation in the ELE project. Through this form, external stakeholders are able to select a language for which they want to report additional information. In addition, as soon as the first intermediate results of the project are available online (in Q3/Q4 2021), we will set up an online form with an open invitation for others to provide any kind of input or feedback on our first results. Later on, this page may be extended to include a small number of targeted questions for which we are seeking responses (Q4/2021, Q1/2022).

6.2. Instrument: Social Media

Social media activities in ELE concentrate on two different networks: LinkedIn (Section 6.2.1) and Twitter (Section 6.2.2). We use LinkedIn primarily to address professional stakeholders of all kinds including LT developers and users. In contrast, while Twitter is primarily used to reach out to European citizens, it is also used by many stakeholders for professional communication purposes.

6.2.1. LinkedIn

Under the umbrella of WP4, we will set up a LinkedIn page for the ELE project through which we will be able to reach out to a large variety of stakeholders relevant for ELE (as further described in the upcoming Deliverable D4.2). Through this LinkedIn page and corresponding LinkedIn posts, which will also be shared and boosted by all consortium partners through their own personal and organisational LinkedIn accounts to extend their reach into the LinkedIn community, we will be able to address a large, five to six figure number of relevant stakeholders, asking them for their opinions, ideas, needs and wishes with respect to DLE, and thereby encouraging them to contribute to the project. The key approach will be to inform users on LinkedIn through LinkedIn posts in order to drive them (eventually) to the ELE website where additional information about the topic of the respective post is provided including ways of getting in touch with the ELE consortium (see Section 6.1). In addition to the goal of driving users to the ELE website, we will also make use of dedicated features of the LinkedIn platform that are of relevance to our project. One challenge will be the identification of the right stakeholders. For this purpose, we will make use of LinkedIn sub-communities and groups as well as hashtags.

6.2.2. Twitter

Just like for LinkedIn, WP4 will set up a Twitter page for the ELE/ELG projects through which we will be able to reach out to a large variety of stakeholders relevant for ELE and ELG (as further described in the upcoming Deliverable D4.2). Through this Twitter page and corresponding Twitter posts, which will also be shared and boosted by all consortium partners through their own personal and organisational Twitter accounts to extend their reach into the Twitter community, we will be able to address a large, five to six figure number of relevant stakeholders, asking them for their opinions, ideas, needs and wishes with respect to DLE, encouraging them to contribute to the project. The key approach will be to inform users on Twitter through Twitter posts in order to drive them, eventually, to the ELE website where additional information about the topic of the respective post is provided including ways of getting in touch with the ELE consortium (see Section 6.1). In addition to the goal of driving users to the ELE website, we will also make use of dedicated features of the Twitter platform that are of relevance to our project, for example, online polls. One challenge will be the identification of the right stakeholders. For this purpose, we will make use of Twitter hashtags and finding the right multipliers to benefit from their own reach.

6.3. Instrument: Email Surveys

The last additional instrument to be used for soliciting feedback are online surveys to which we will invite selected stakeholders. These final surveys will be short, and prepared once we have assembled the main pillars of the strategic agenda and roadmap including our overall strategic recommendations. Especially if far-reaching decisions have to be made when selecting between competing ideas or visions, we will consider organising additional online surveys that may perhaps function as tie-breakers (in Q1/2022). Technically, these surveys will be realised in the same way as the online surveys organised under the umbrella of Task 2.1 and Task 2.2.

7. Summary and Next Steps

This report specifies the consultation process to be carried out under the umbrella of WP2 (European Language Equality – The Future Situation in 2030). The deliverable is meant to be used as an agreed upon set of guidelines and instructions with regard to the implementation of our internal processes and instruments to be applied by all actively involved partners, especially with regard to Task 2.1 (The perspective of European LT developers – industry and research) and Task 2.2 (The perspective of European LT users and consumers), but also with regard to Task 2.3 (Science – Technology – Society: Language Technology in 2030).

WP2, with its three different tasks, has an important role within the ELE project: it defines all aspects of the future situation that we want to have reached with regard to digital language equality by 2030. Due to this important, mission-critical role in the grander scheme of the project, all involved stakeholders have to be aware of the various different aspects and dimensions that WP2 needs to provide input for when it comes to assembling the strategic agenda and roadmap. Eventually, the Strategic Research, Innovation and Deployment Agenda and Roadmap must be able to demonstrate:

- *Interdisciplinarity*: Language Technology is not “just Artificial Intelligence”, neither is it “just Computational Linguistics” or “just Linguistics plus Computer Science”. Instead, it is an interdisciplinary field that includes and benefits from various areas and disciplines that are all vital and essential for the overall success of the strategic agenda. If there is no dedicated support for all different fields involved, the digital language equality topic will simply fall by the wayside.
- *Breadth and completeness*: Due to the complexity and potential size and extension of the involved (eco)system of stakeholders, the project must make sure that all relevant and important communities – especially the different LT user communities – are included in the preparation and discussion of the strategic agenda and also in its promotion.
- *Societal and economic relevance*: The strategic plan to be produced by the project needs to position itself with regard to society and economy, especially concerning the potential and need of creating new markets, new economic hubs and new jobs, which will eventually all contribute to the emergence and sustainability of digital language equality in Europe by 2030.
- *Potential*: Overall, the strategic agenda and roadmap need to be able to demonstrate enormous potential for the development of future and emerging technologies.⁸

⁸ Just as a comparison, the potential needs to be emphasised and demonstrated in the order of an EU FET Flagship project (the FET Flagship Programme has been discontinued in 2019).

- *Competence of the community:* The European Language Technology community, comprised of hundreds of research organisations and hundreds of companies, is not only world-class but also standing by to tackle the issue of creating digital language equality in Europe by 2030. In addition, the community is also competitive, especially when it comes to the offerings of the big international technology companies.
- *Urgency:* Attempting to achieve Digital Language Equality in Europe is not only extremely important but also very urgent. If Europe is unable to make a coordinated push into this direction in a substantial and timely manner, there is a danger that the fabric of the European society will be further eroding, maybe even beyond repair.
- *Adequacy:* Our eventual plans for achieving Digital Language Equality need to emphasise the potential of not only partially addressing but actually solving a larger number of crucial European technical, social, economic and scientific challenges.
- *Fit for Europe:* In other words, the strategic agenda and roadmap need to explain why Digital Language Equality is a *European* topic.

Month	WP/Task		Description
May 2021	M5	WP2	Submission of this deliverable (actual date)
June 2021	M6	T2.1	Surveys sent to leads of Deliverables D2.2-D2.6
		T2.2	Surveys sent to leads of Deliverables D2.7-D2.12
July 2021	M7	T2.1	Structure of D2.2-D2.6 sent to Deliverable leads
		T2.2	Structure of D2.7-D2.12 sent to Deliverable leads
		T2.3	Structure of D2.13-D2.16 sent to Deliverable leads
Aug. 2021	M8	T2.1, T2.2	Completed surveys collected from Deliverable leads
Sept. 2021	M9	T2.1, T2.2	Overall survey results compiled
		T2.1, T2.2	Interviews to be performed determined, started
Oct. 2021	M10	T2.1, T2.2	Interviews completed
Nov. 2021	M11	T2.1	D2.2-2.6 drafts ready and communicated to WP3
		T2.2	D2.7-2.12 drafts ready and communicated to WP3
Dec. 2021	M12	T2.3	D2.13-D2.16 drafts ready and communicated to WP3
Jan. 2022	M13	WP2	Recommendations extracted from D2.2-D2.16 drafts
		WP2	Analysis of strategic reports (WP3) for further activities
Feb. 2022	M14	WP2	D2.2-D2.16 submitted
		WP2	Surveys prepared in M13 sent to selected stakeholders
Mar. 2022	M15	WP2	D2.17 submitted; all results to be summarised in D2.18
Apr. 2022	M16	WP2	D2.18 submitted

Table 6: Consolidated WP2 timeline covering Task 2.1, Task 2.2 and Task 2.3

As a summary and overview of the ELE consultation process, Table 6 provides a consolidated timeline for the main WP2 activities, which functions as the WP-internal roadmap, especially for the ELE core team and for all deliverable leads.

References

- LT-Innovate. The lt-innovate innovation agenda, 2016. URL http://www.lt-innovate.org/sites/default/files/2904-LTi_Innovation_Agenda.pdf.
- 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. *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, 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 Vasiljevs, Orians Anvari, Andis Lagzdīņš, Jūlija Melņika, Gerhard Backfried, Erinc Dikici, Miroslav Janosik, Katja Prinz, Christoph Prinz, Severin Stampler, Dorothea Thomas-Aniola, José Manuel Gómez Pérez, Andres Garcia Silva, Christian Berrio, Ulrich Germann, Steve Renals, and Ondrej Klejch. European Language Grid: An Overview. In Nicoletta Calzolari, Frédéric Béchet, Philippe Blache, Christopher Cieri, Khalid Choukri, Thierry Declerck, Hitoshi Isahara, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, pages 3359–3373, Marseille, France, 5 2020. European Language Resources Association (ELRA).
- Georg Rehm, Stelios Piperidis, Kalina Bontcheva, Jan Hajic, Victoria Arranz, Andrejs Vasiljevs, Gerhard Backfried, José Manuel Gómez Pérez, Ulrich Germann, Rémi Calizzano, Nils Feldhus, Stefanie Hegele, Florian Kintzel, Katrin Marheinecke, Julian Moreno-Schneider, Dimitris Galanis, Penny Labropoulou, Miltos Deligiannis, Katerina Gkirtzou, Athanasia Kolovou, Dimitris Gkoumas, Leon Voukoutis, Ian Roberts, Jana Hamrlová, Dusan Varis, Lukáš Kačena, Khalid Choukri, Valérie Mapelli, Mickaël Rigault, Jūlija Melņika, Miro Janosik, Katja Prinz, Andres Garcia-Silva, Cristian Berrio, Ondrej Klejch, and Steve Renals. European Language Grid: A Joint Platform for the European Language Technology Community. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations (EACL 2021)*, pages 221–230, Kyiv, Ukraine, 4 2021. Association for Computational Linguistics (ACL).
- Andrejs Vasiljevs, Khalid Choukri, Luc Meertens, and Stefania Aguzzi. Final study report on CEF Automated Translation value proposition in the context of the European LT market/ecosystem, 2019. DOI 10.2759/142151. A study prepared for the European Commission, DG Communications Networks, Content & Technology by Crosslang, Tilde, ELDA, IDC.

Appendix

A. LT Developers: Survey Questions (Draft)

Table 7: LT developers survey (draft)

Section	Question	Question type	Answers
Introduction	What is your name?	Open-ended	–
	What is your email address?	Open-ended	–
	In which country is your organisation headquartered?	Single choice	list of European countries
	What is the name of the group, department, organisation you work for?	Open-ended	–
	Which of the following best describes the type of organisation you work for?	Single choice	a) University or academic research organization b) Research center c) SME d) Large enterprise e) Other
	Which LT areas do you mainly work in?	Multiple choice	a) Basic natural language processing services b) Text analytics and mining, Information extraction, Text classification c) Machine Translation d) Speech technologies e) Conversational systems f) Language resources, data production, aggregation g) Language resources, data distribution, data marketplace h) Research infrastructures (e. g., repository) i) Other
	In which sectors are your technologies, products, services used?	Multiple choice	list of domains, sectors
Language coverage	Which languages does your organisation conduct research in or for which languages do you offer services, software, resources, models etc.?	Multiple choice	list of languages
Continued on next page			

Table 7 – Continued from previous page

Section	Question	Question type	Answers
	Are there any languages that you do not yet support, but you plan to support in the next three years?	Multiple choice	list of languages
	Considering your development plans with respect to language coverage, what are the main drivers for your decision to support additional languages?	Multiple choice	a) No market interest b) Research/scientific interest c) Not available funding/investment d) No expertise, human experts for other languages e) Lack of language resources f) Lack of technologies or tools g) Other
	Where do you see crucial gaps in terms of tools or applications, especially with regard to specific languages? For instance, do you see crucial gaps in speech technologies, basic NLP components, conversational systems, machine translation etc. for specific languages?	Open-ended	–
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 ...”	Four-point scale (strongly agree, agree, disagree, strongly disagree)	a) basic research is still needed b) inadequate recognition of the importance of multilinguality c) lack of talent, brain drain d) fragmentation of the European LT industry e) lack of coordination and missing links between research, LT vendors, integrators and customers f) too little public procurement (in contrast to other continents, where government sectors are the main drivers of LT industry) g) insufficient market in smaller countries to justify investments in LTs for smaller languages h) low understanding of customer needs while, at the same time, integrators do not know the value of LT and cannot educate the customer i) cost for access to compute infrastructure j) competition with non-European big companies and market disruption by global players

Continued on next page

Table 7 – Continued from previous page

Section	Question	Question type	Answers
	Please provide more details on the obstacles and challenges indicated in the previous question and/or add any other obstacle or challenge that was not previously listed.	Open-ended	–
Predictions and visions for the future	In your opinion, how effectively can the following policies or instruments contribute to speeding up the development and deployment of LT in Europe equally for all languages?	Likert-scale (very effective – not at all)	<ul style="list-style-type: none"> a) Initiate large-scale, long-term funding programme for European LT development b) Initiate investment instruments and accelerator programs targeting LT start-ups c) Increase the availability of qualified personnel on LT and incentives for talent retention d) Public procurement of innovative technology and pre-commercial public procurement e) Raise awareness of the benefits for companies, public bodies, and citizens of the availability of online services, contents and products in multiple languages f) Impose content accessibility regulations, e.g., multimedia subtitling, readability, dubbing, availability of content in multiple languages etc. g) Invest in the development of new (scientific, technological) methodologies for transfer or adaptation of resources or technologies to other domains and languages h) Reinforce training and education initiatives, including undergraduate and masters programs and vocational training in LT i) Continuous investment to Resources Infrastructures that support LT j) Other
Continued on next page			

Table 7 – Continued from previous page

Section	Question	Question type	Answers
	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?	Open-ended	–
	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 work in with regard to innovation and the LT industry?	Open-ended	–
	Do you have any other additional suggestions or recommendations with regard to ELE?	Open-ended	–
	Can we contact you to arrange a possible follow-up discussion?	Single choice	Yes/No

B. LT Users and Consumers: Survey Questions (Draft)

Table 8: LT users and consumers survey (draft)

Section	Question	Question type	Answers
Introduction	What is your name?	Open-ended	–
	What is your email address?	Open-ended	–
	In which country are you based?	Single choice	list of European countries
<i>Continued on next page</i>			

Table 8 – Continued from previous page

Section	Question	Question type	Answers
	Which association/community/organisation of users and consumers do you represent?		
	How many members are there in the association/community/organisation of users and consumers you represent in this survey? (total number of full-time employees)	Single choice	a) 1-10 b) 11-100 c) 100-500 d) More than 500 e) n. a.
	What is your main role in the organisation you work for?	Open-ended	–
	Which of the following best describes the type of organisation you work for?	Single choice	a) Professional association b) Government department, unit c) SME d) Large enterprise e) Independent contractor or consultant e) Education, research f) Other
	If “other”, please specify	Open-ended	–
Language coverage	Which language(s) listed below do you or your organization work with?	Multiple choice	List of European Languages
	If “other”, please specify	Open-ended	–
	Do you or your organisation plan to include additional languages in your workflow in the next 3 years?	Yes/No	a) Yes b) No
	Which language(s)?	Multiple choice	List of European countries
	If “other”, please specify	Open-ended	–
	Do you or your organisation work with any minority/regional/lesser-used language(s) not included in the list of EU languages provided above?	Yes/No	a) Yes b) No
Evaluation of the current situation	Which minority/regional language(s)?	Open-ended	–
	Which language technology tools or applications listed below are available for the European language(s) you or your organisation work with?	Multiple choice	List of NLP tools and applications
Continued on next page			

Table 8 – Continued from previous page

Section	Question	Question type	Answers
	If “other”, please specify	Open-ended	–
	Do you perceive gaps in technological support for the EU language(s) you work with?	Yes/No	a) Yes b) No
	Please, indicate the language(s) you perceive the gaps listed below.	Matrix	List of European languages a) Gaps in the amount and variety of available applications b) Gaps in the quality of the tool or application (delays in responding, difficulties with special characters, language-related errors in the output etc.) c) Gaps in the variety of linguistic phenomena or text types covered d) Gaps in adaptability to different systems (e.g., adaptability to iOS system) e) Other
	Please choose the option that best describes the level of language technology support for the language(s) you or your organisation work with.	Matrix	List of European languages a) Excellent support b) Good support c) Moderate support d) Poor support e) No support
	In general terms, how do you evaluate the performance of the tools you use with the EU language(s) you work with? Please evaluate based on a four-point scale.	Four-point likert scale	List of NLP tools and applications 1) Very poor 2) Poor 3) Good 4) Excellent 5) n. a.
	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.	Five-point scale	List of NLP tools and applications 1) never 2) rarely 3) sometimes 4) frequently 5) every day
Continued on next page			

Table 8 – Continued from previous page

Section	Question	Question type	Answers
	Please indicate for which language(s) you or your organisation use the Language Technology tools/applications listed below.	Matrix	List of NLP tools and applications – list of European languages
	If “other”, please specify	Open-ended	–
	Are there language technology tools or applications available for any minority/regional/lesser-used language(s) you or your organisation work with?	Yes/No	a) Yes b) No
	Which tools or applications do you use with these minority/regional/lesser-used languages?	Multiple-choice	List of NLP tools and applications
	Do you perceive gaps in technological support for the minority/regional/lesser-used language (s) you work with?	multiple choice	a) Amount and variety of available applications b) Quality of the tool/application (delays in responding, difficulties with special characters, language-related errors in the output etc.) c) Variety of linguistic phenomena/text types covered d) Adaptability to systems (e. g., adaptability to iOS system) e) Other
	If “other”, please specify	Open-ended	–
	In general terms, how do you evaluate the performance of the tools for the minority/regional/lesser-used language(s) you work with? Please evaluate based on a four-point scale.	Four-point likert scale	List of NLP tools and applications 1) Very poor 2) Poor 3) Good 4) Excellent
	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.	number-slider	1(Very poor) - 5 (Excellent)
	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.	Five-point likert scale	1) never 2) rarely 3) sometimes 4) frequently 5) every day

Continued on next page

Table 8 – Continued from previous page

Section	Question	Question type	Answers
Predictions and visions for future	What could be done to increase the use of language tools for the specific languages you or the community you represent use?	Multiple-choice	a) wider range of language tools for the languages I work with b) Higher-quality tools for the languages I work with c) More training of personnel working with such tools d) Other
	If “other”, please specify	Open-ended	–
	Which tools or applications that substantially use language technology do you want to see in the constituency you represent that are not available today (we welcome any suggestion, even ideas that are not possible with current technology)?	Open-ended	–
	Please, indicate the best option that describes your vision for the future of languages technology	Four-point likert scale	1) Strongly disagree 2) Disagree 3) Agree 4) Strongly Agree a) In the next 10 years, there will be higher-quality language tools that deal with all the languages that concern me, including minority languages b) In the next 10 years, there will be a wider range of language tools for European Languages c) Language technology tools will help prevent the loss of linguistic diversity
<i>Continued on next page</i>			

Table 8 – Continued from previous page

Section	Question	Question type	Answers
	In your opinion, what would be the most relevant benefits of improving technologies for the languages you or your organisation work with (including minority or regional languages)?	Multiple choice	a) Increase individuals' exposure to these languages b) Prevent minority or regional languages from disappearing c) Increase the number of speakers of those languages, including minority or regional languages d) Improve communication between native speakers e) Improve literacy for minority or regional languages f) Enhance the communication capabilities of people with disabilities g) Increase engagement with social, leisure and work activities in their own languages h) Improve online trade in countries where those languages are spoken i) Improve offline trade (i.e., not e-commerce) in countries where those languages are spoken j) Other
	If "other", please specify	Open-ended	–
	If you have any comments or suggestions let us know.	Open-ended	–

C. LT Users and Consumers: Additional Stakeholders

Table 9: LT users and consumers survey – additional stakeholders

Acronym	Full name	Stakeholder group
–	Europeana	European Digital Heritage and Library Community
Ecommerce	Ecommerce Europe	European Ecommerce Community
EBU	European Public Broadcasters	European Public Broadcasters
–	European Telecom Operators	
–	European (Digital/Technology) Health Sector	
–	European Educational (Technology) Sector	
–	European Legal IT Sector	
–	European Publishers	
–	European Banks	
–	European Mobility Association	
–	European Car Manufacturer Association	
–	European Green Deal Association	
–	European Smart City Association	
Tekom	European Technical Documentation Community	
EUPAN-Network	European Public Administration Network	
DISPA	Network of Directors of Institutes and Schools of Public Administration	
EASA	European Union Aviation Safety Agency	
ACI Europe	Airports Council International Europe	
EUROCHAMBERS	Association of European Chambers of Commerce and Industry	
EBF	European Banking Federation	
EPC	European Publishers Council	
ENPA	European Newspaper Publishers Association	
EPSU	European Federation of Public Service Unions	
–	European DIGITAL SME Alliance	
–	Enterprise Europe Network	
–	European Digital Marketing Association	Digital Marketing Association
–	IAB Europe	Digital Marketing Association
–	Translators without Borders	NGO
EDMO	European Digital Media Observatory	
EPRS	European Parliamentary Research Service	European Research
EPD	European Partnership for Democracy	
ERC	European Research Council	
ESF	European Services Forum	European Service Sector
EUA	European University Association	European Education Community
Continued on next page		

Table 9 – Continued from previous page

Acronym	Full name	Stakeholder group
EACEA	Education, Audiovisual and Culture Executive Agency	European Education and cultural exchange
Cefic	European Chemical Industry Council	
FAEP	European Federation of Magazine Publishers	European Publishers
NATO	North Atlantic Treaty Organization	International Military Alliance
PostEurop	Association of European Public Postal Operators	European Postal Operators
AFME	Association for Financial Markets in Europe	European Banks
–	Insurance Europe	European Insurance
ERTICO	European Road Transport Telematics Implementation Coordination Organisation	European Transport and Mobility
EnR	European Energy Network	European Energy Association
EFPIA	European Federation of Pharmaceutical Industries and Assoc.	European Pharmacy
–	The Open Group	Global open standardization
DIGITAL SME	European DIGITAL SME Alliance	European economy
EMMA	European Magazine Media Association	European Publishers
EAA	European Acoustics Association	European Research
ENPA	European Newspaper Publishers' Association	European Publishers
UITP	The International Association of Public Transport	European Mobility Association
ACT	Association of Commercial Television in Europe	European Public Broadcasters
AER	Association of European Radios	European Radios
EACA	European Association of Communication Agencies	European Communication Sector
FEDMA	Federation of European Direct and Interactive Marketing	European Marketing
ACEA	European Automobile Manufacturers' Association	European Car Manufacturer Association
DTN	Digital Tourism Network	European Digital Tourism Community
GDC	Green Digital Charter	European Smart City and Green Deal Association
WGIN	World Green Infrastructure Network	Global Green Deal Association
STC	Society for Technical Communication	European Technical Documentation Community
TAFTIE	Applied Research Funders Association	Government Funding Agencies
DARIAH	Research Infrastructure in Digital Humanities and Arts	Research Infrastructure
EHRI	Research Infrastructure for Holocaust Research	Research Infrastructure
OECD	Organisation for Economic Co-operation and Development	Intergovernmental economic organisation
UNESCO	UN Educational, Scientific and Cultural Organization	Intergovernmental cultural organisation
WTO	World Tourism Organization	Intergovernmental touristic organisation
WHO	World Health Organization	Intergovernmental health organisation
FAO	Food and Agriculture Organization of the UN	Intergovernmental food organisation