

Author	Marlies Thönnissen
Dissemination level	Public
Date	28-02-2022

About this document

Project Grant agreement no. Coordinator Co-coordinator Start date, duration	European Language Equality (ELE) LC-01641480 – 101018166 ELE Prof. Dr. Andy Way (DCU) Prof. Dr. Georg Rehm (DFKI) 01-01-2021, 18 months
Deliverable number Deliverable title	D2.2 Report from CLAIRE
Type Number of pages Status and version Dissemination level Date of delivery Work package Task Author Reviewers	Report 38 Final Public Contractual: 28-02-2022 – Actual: 28-02-2022 WP2: European Language Equality – The Future Situation in 2030 Task 2.1 The perspective of European LT developers (industry and research) Marlies Thönnissen Maria Giagkou, Jan Hajič
EC project officers	Susan Fraser, Miklos Druskoczi
Contact	European Language Equality (ELE) ADAPT Centre, Dublin City University Glasnevin, Dublin 9, Ireland
	Prof. Dr. Andy Way – andy.way@adaptcentre.ie
	European Language Equality (ELE) DFKI GmbH Alt-Moabit 91c, 10559 Berlin, Germany
	Prof. Dr. Georg Rehm – georg.rehm@dfki.de
	http://www.european-language-equality.eu
	© 2022 ELE Consortium

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	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 Arna Magnússonar í íslenskum fræðum SAM (Arni 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
30	Lieuwių Kalbos Institutas (institute of the Linuanian Language)	LKI	
3/	Luxembourg institute of Science and Technology	LISI	LU
38	Universita ta Maita (University of Maita)	UM	MI
39	Suchting Instituti voor de Nederlandse Taal (Duich Language Institute)	INI	NL
40	Sprakradet (Language Council of Norway)	LUNOK	NU
41	Instytut Polstaw Informatyki Polskiej Akademii Natk (Polsta Academy of Sciences)	IPIPAN FOULishan	PL
42	Universidade de Lisboa, Faculdade de Ciencias (University of Lisbon, Faculty of Science)	FULISDOIL	PI
43	Institutu de Cercetari Pentru Intelgenja Artificial (containan Academy)	ICIA	CV
44	University of Cyprus, French and European Studies	UCY	CI
45	Jazykovedny ustav Ludovita Stura Slovenskej akademie vied (Slovak Academy of Sciences)	JULS	SK
46	Institut Jozef Stefan (Jozef Stefan Institute)	JSI	SI
4/	Centro Nacional de Supercomputación (Barcelona Supercomputing Center)	B2C	ES
48	Kungliga Tekniska nogskolan (Royal Institute of Technology)	KIH	SE
49	Universitat Zurich (University of Zurich)	UZH	CH
50	University of Snemela	U2ED	UK
21	Universidad de vigo (University of vigo)	UVIGU	£5
э∠	bangor university	BNGK	UK

Contents

1.	Introduction	1
	1.1. About CLAIRE	2
2.	Methodology and instruments 2.1. Online survey. 2.2. Interviews	2 2 4
3.	Analysis of responses to survey questions3.1. Respondents profile3.2. Language coverage3.3. Evaluation of current situation3.4. Predictions and visions for the future	5 6 7 10
4.	Analysis of interviews	13
5.	Conclusions	17
Α.	The LT researchers and developers full survey	20
В.	Additional tables and graphs	28

List of Figures

1. 2. 3. 4.	Type of organisation LT areas in which the respondents conduct research or develop tools and services Languages that organisations plan to support in the short- and medium-term . Level of agreement that basic research is still needed as one of the main challenges and obstacles the European LT community currently faces	5 6 7 8
5.	Results for measure "Initiate large-scale, long-term funding programme for Eu-	U
	ropean LT development"	11
6.	Results for measure "Continuous investment in the research infrastructures	
	that support LT"	12
7.	Full survey as published (page 1/9)	20
8.	Full survey as published (page 2/9)	21
9.	Full survey as published (page 3/9)	22
10.	Full survey as published (page 4/9)	23
11.	Full survey as published (page 5/9)	24
12.	Full survey as published (page 6/9)	25
13.	Full survey as published (page 7/9)	26
14.	Full survey as published (page 8/9)	26
15.	Full survey as published (page 9/9)	27

List of Tables

1.	Types of survey questions	3
2.	Breakdown of answers to "Which of the following best describes the type of	
	organisation you work for?" (mandatory closed question)	28
3.	Breakdown of answers to "Where is your organisation's headquarter based?"	
	(mandatory closed question, plus "if other" as optional open-ended question) .	28
4.	Full list of answers to "Which LT areas do you mainly work in? If "Other",	
	please specify." (optional open-ended question)	29
5.	Breakdown of answers to "What languages does your organisation conduct re-	
	search in and/ or for what languages do you offer services, software, resources,	
	models etc.?" (mandatory multiple choice question, plus "if other" as optional	• •
0	open-ended question)	29
6.	Breakdown of answers to "Are there any languages that your organisation does	
	not yet support, but you plan to support in the next three years?" (mandatory	20
7	Montions of the top drivers for the desision to support additional languages	20
/. 0	Answers to the question: "Dease indicate if you agree with the following state	30
0.	ments: "One of the main challenges and obstacles the European LT community	
	currently faces is " (mandatory closed question answers provided on a four-	
	noint scale nlus "I don't know/No answer")	31
9	Answers to the question: "In your opinion, how effective can the following poli-	51
	cies/instruments be in speeding up the development and deployment of LT in	
	Europe equally for all languages?" (mandatory closed question, answers pro-	
	vided on a five-point scale, plus "I don't know/No answer")	32



List of Acronyms

AAAI	Association for the Advancement of Artificial Intelligence
Adra	AI, Data and Robotics Association
AI	Artificial Intelligence
AI4EU	AI4EU (EU project, 2019-2021)
CEF AT	Connecting Europe Facility, Automated Translation
CH	Cultural Heritage
CLAIRE	Confederation of Laboratories for AI Research in Europe
CLARIN	Common Language Resources and Technology Infrastructure
CRACKER	Cracking the Language Barrier (EU project, 2015–2017)
DH	Digital Humanities
EC	European Commission
ECSPM	European Civil Society Platform for Multilingualism
EFNIL	European Federation of National Institutes for Language
ELE	European Language Equality (this project)
ELE Programme	European Language Equality Programme (the long-term, large-scale fund-
	ing programme specified by the ELE project)
ELG	European Language Grid (EU project, 2019-2022)
ELRC	European Language Resource Coordination
ELLIS	European Laboratory for Learning and Intelligent Systems
EP	European Parliament
ESA	European Space Agency
EurAI	The European Association for Artificial Intelligence
euRobotics	euRobotics AISBL
GIZ	German Association for Internal Collaboration
HPC	High-Performance Computing
LR	Language Resources/Resources
LT	Language Technology/Technologies
META-NET	EU Network of Excellence to foster META
ML	Machine Learning
MT	Machine Translation
NLP	Natural Language Processing
NLU	Natural Language Understanding
R&D	Research and Development
SME	Small- and medium-sized enterprises
SRA	Strategic Research Agenda
SRIA	Strategic Research and Innovation Agenda
SSH	Social Science Humanities
STOA	Science and Technology Options Assessment

Abstract

In 2018, the European Parliament published its resolution "Language equality in the digital age", which was passed in a landslide vote with 592 votes in favour and 45 against it. Its more than 40 recommendations include the establishment of a large-scale, long-term coordinated funding programme for research, development and innovation in the field of Language Technology, at European, national and regional levels, tailored specifically to Europe's needs and demands as well as securing Europe's leadership in language-centric AI.¹

According to the results of this survey, there is a strong agreement for the need to support European basic research in language-centric AI in a variety of research areas, where research and development fall behind USA and Asia – a concern that CLAIRE also raises for all areas of AI. Closely linked to this, a major challenge that the European Language Technology community currently faces is the strong competition with key players outside of Europe and the market disruption through them. Furthermore, the results of the survey show that the distinctly unique feature of multilinguality is not adequately recognised in Europe. CLAIRE has consulted its (LT) community on the topic of European Language Equality in order to support the European Parliament's ambition for digital language equality by 2030 and to join forces as part of the European Language Equality project to make a meaningful contribution to this goal.

1. Introduction

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

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

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

Due to the **multi- and interdisciplinary** nature of Language Technology, which stands at the intersection of Linguistics and Computational Linguistics, Computer Science and Artificial Intelligence, while at the same time it encompasses methods and findings from Cognitive Science and Psychology, Mathematics, Statistics, Philosophy and more, the ELE stakeholders group of LT developers also includes neighbouring disciplines, especially AI and Digital Humanities/Social Science and Humanities (DH/SSH). To reach out to this diverse and extensive group of stakeholders, the partners of the ELE consortium mobilised various European networks, associations, initiatives and projects, covering both research and industry.

See recommendations 25 and 27 of the EP's resolution "Language equality in the digital age" (https://eur-lex. europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018IP0332&from=EN)

1.1. About CLAIRE

The Confederation of Laboratories for Artificial Intelligence Research in Europe (CLAIRE)³ is an organisation created by the European AI community that seeks to strengthen European excellence in AI research and innovation, with a strong focus on human-centred AI. CLAIRE aims to ensure that societies and citizens across all of Europe, and beyond, benefit from AI as a major driver of innovation, future growth and competitiveness, and to achieve world-wide brand recognition for "AI made in Europe".

Founded in 2018, CLAIRE has garnered the support of more than 3,700 AI experts and stakeholders, who jointly represent the vast majority of Europe's AI community, spanning academia and industry, research and innovation. Among them are more than 140 fellows from various key scientific associations.

CLAIRE's membership network consists of over 430 research groups and institutions, covering jointly more than 24,000 employees in 37 countries. Furthermore, CLAIRE has recently set up an Innovation Network that, together with the established Research Network, will foster a strong link between research and industry.

The CLAIRE vision⁴ is officially supported by the governments of nine European countries,⁵ 28 scientific associations across all of Europe, the European Association for Artificial Intelligence (EurAI), the Association for the Advancement of Artificial Intelligence (AAAI), and the European Space Agency (ESA).

CLAIRE is also actively liaising with other important AI-organisations, including the AI, Data and Robotics Association (Adra), AI4EU, the Big Data Value Association (BDVA), the European Laboratory for Learning and Intelligent Systems (ELLIS) and euRobotics.

CLAIRE strongly believes that Language Technology and NLP play a key role not only in Europe but also around the world. CLAIRE supports the mission to leverage the capabilities and potential in that area for the benefit of everyone. According to a survey recently conducted among the CLAIRE member groups, 44.9% indicated to have (some) expertise in NLP. CLAIRE has an Advisory Group on NLP⁶ consisting of high-caliber European NLP and LT scientists for advice on the needs of the community and on how to connect to it.

2. Methodology and instruments

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

2.1. Online survey

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

ELE

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

³ https://claire-ai.org

⁴ The CLAIRE vision, https://claire-ai.org/wp-content/uploads/2019/10/CLAIRE-vision.pdf (October 2019).

⁵ The CLAIRE vision is officially supported by Belgium, Czech Republic, Finland, Greece, Italy, Luxembourg, Netherlands, Slovak Republic and Spain

⁶ https://claire-ai.org/iags

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

Table 1: Types of	survey questions
-------------------	------------------

The survey was structured in four main parts:

- **Part A. Respondents' profiling**: The first part of the survey included 13 questions for the demographic profiling of respondents with emphasis on characteristics relevant to the task at hand, i. e.
 - Country
 - Affiliation
 - Type of organisation
 - LT areas that the respondent is mainly active in
 - Participation/membership in networks/associations
 - Sectors/domains that the respondent is active in (if relevant)
- **Part B. Language coverage**: The second part investigated the degree of coverage of the European languages by the respondents' current research and development activities, i. e.
 - languages currently supported in research/products/services
 - languages planned to be supported in the short-/middle-term
 - factors that influence the respondents' decision with regard to language coverage/support
- **Part C. Evaluation of current situation**: This part included questions that sought to elicit the respondents' evaluation of the current situation of the LT research and development, the strengths, gaps and challenges that the European LT community is facing, i.e.
 - gaps in terms of: a) technologies, b) tools/applications, and c) resources, especially with regard to specific languages
 - LT areas where the European LT community excels
 - main perceived challenges and obstacles that should be overcome
- **Part D. Predictions and visions for the future**: The fourth part of the survey is the forward-looking section that investigated ideas, predictions and wishes of the LT community about how the LT field as a whole will achieve to equally support all European languages by 2030, i. e.
 - policies/instruments that could contribute to speed up the effective deployment of LT in Europe equally for all languages

- prediction of future opportunities for LT in basic and applied research (scientific vision) and in innovation and the industry
- expectations of the community with regard to the challenges an ELE programme can address by 2030
- **Follow-up**: The last three questions asked the respondent's permission to be contacted for an interview and, given an affirmative answer, his/her contact details.

The survey was designed, set up and published on the EU Survey platform.⁷ The full survey, as published online, is presented in Appendix A (p. 20 ff.).

The LT developers survey was distributed via email to 83 CLAIRE members who had expertise in Natural Language Processing (NLP), ranging from some expertise through significant expertise to world-class expertise, according to the results of a previously conducted internal CLAIRE questionnaire. The survey was also sent to 177 additional members who indicated that they work in the field of NLP at the time they signed up to CLAIRE. In addition, it was advertised on the ELE, ELG and ELT websites.⁸

The survey was opened on 17 June 2021 and closed on 18 October 2021. In total, 320 responses have been collected, out of which 38 from respondents who indicated that they are members of CLAIRE. This subset of responses, representing the views of CLAIRE, is analysed in this report.

2.2. Interviews

The interviews were conducted with the aim to further engage with the LT community as well as to spread the message about European Language Equality's mission and, most importantly, to elicit some views and opinions about how the LT community should pursue the goal of Digital Language Equality. The selection process for the interview candidates was based on a geographical balance, including different affiliations. In total, eight survey participants were contacted in early November 2021 via e-mail, with the option to conduct the interview in oral or written form. The selected organisations are located in 8 different countries, namely France, Germany, Greece, the Netherlands, Portugal, Slovak Republic, Spain and Switzerland. Furthermore, two members of CLAIRE's Advisory Group on NLP were contacted. The selected candidates represent 3 different organisations, out of which 50% are universities, 37.5% are research centers (independent from any other academic organisation), one represents a large company.

- **Part 1. Introduction to the European Language Equality project:** The first part of the interview consisted of a short introduction of the ELE project, including
 - Overview
 - Background
 - Objectives
- **Part 2. Introduction of the interviewee:** The second part of the interview consisted of a short introduction of the selected interviewee, i. e.
 - Name
 - Affiliation
 - Expertise in LT, NLP and/or language-centric AI

⁷ https://ec.europa.eu/eusurvey/runner/ELE-LTdevs

⁸ https://european-language-equality.eu, https://www.european-language-grid.eu, https://www.european-language-technology.eu as well as through the ELT social media accounts on Twitter and LinkedIn.

- **Part 3. Evaluation of current and future situation:** The third part of the interview consisted of selected questions of the LT developers survey with the request to further elicit the interviewee's answer, i. e.
 - Main challenges and obstacles the European LT community currently faces 10 questions
 - Policies and instruments to speed up the development and deployment of LT in Europe equally for all languages – 10 questions
 - Key challenges a future ELE funding programme should address 3 questions

One oral interview has been conducted in the beginning of December 2021, via an online meeting with one of our Advisory Group NLP members. Two interviews have been conducted in written form with the coordinator of the European Language Resource Coordination (ELRC) and the Head of the Data Intelligence research group.

3. Analysis of responses to survey questions

3.1. Respondents profile

One major goal of the survey was to bring the European LT community together and hence reach a broad and demographically distributed audience. The respondents represent 27 different organisations, out of which 95% are research or academic institutions, while the rest are practitioners from industry (Figure 1). The headquarters of these organisations are located in 15 different countries, with most responses from 1) Spain, 2) Greece, 3) Switzerland, 4) Portugal and 5) Germany. Detailed statistics of the breakdown of organisation types and countries are provided in Appendix B (Tables 2 and 3).



Figure 1: Type of organisation

The respondents are mainly active in the following LT areas, by order of frequency: 1) Text analytics and mining, information extraction, text classification, 2) Basic natural language processing services (PoS tagging, parsing, named entity recognition etc.) and 3) Conversational systems (Figure 2; see also Appendix B, Table 4).

The technologies, products or services offered by the respondents' organisations are used in a diverse set of domains, a finding that demonstrates the applicability of LT in practically

ELE



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

all economic sectors. The top 3 domains indicated by the respondents were Information and Communication Technologies, Education and Media.

3.2. Language coverage

A total of 18 languages are currently supported by survey participants in research, products or services. The strongest support is first and foremost represented by English, followed by French, German, Spanish and Portuguese. (For a more detailed overview, see also Appendix B, Table 5). Languages participants indicated as "Other" include Arabic, Chinese, Korean, Russian and Swiss German.

With 36 out of the 38 respondents claiming to conduct research or develop LT services for English (Figure 3) and considering that none of the participating organisations are located in a country with English as an official language (cf. Table 3), a possible confirmation can be seen in the statement that the technological support for other European languages are characterised by a strong imbalance and many resources and technologies exist for English while other languages experience a lack of technological support. The results also reflect the key results from the META-NET White Paper series Europe's Languages in the Digital Age (Rehm and Uszkoreit, 2012) that languages with more speakers had better support through Language Technology.

Fourteen organisations indicated 33 languages that they plan to support in the short and medium term. The most chosen languages are French and Polish, followed by Italian and Spanish (Figure 3). Languages listed as "Other" include Afrikaans, Chinese, Japanese, Russian and Swiss German. For a more detailed overview, see also Appendix B, Table 6.

The participants were asked to select the main factors that influence their decision to support additional languages. The respondents' top three drivers were research/scientific interest, available funding/investment and market interest/demand by users or customers. Further selected factors were the availability of language resources, the availability of tech-



Figure 3: Languages that organisations plan to support in the short- and medium-term

nologies/tools and the availability of human experts for other languages. In addition, the provision of services for minority languages and the assurance of language rights in the digital economy, digital services and applications were mentioned as further drivers to support additional languages (see also Appendix B, Table 7).

3.3. Evaluation of current situation

In order to further grasp the main challenges and obstacles the European LT community faces, the survey participants were asked to indicate their level of agreement with a set of statements. The respondents were also asked to further elicit their evaluation for LT areas where the European LT community excels as well as the main perceived challenges and obstacles that should be overcome. Generally, over half of the participants agreed with each of the given statements, differing in the proportion of strong (dis-)agreement and the wish to not answer (I don't know/no answer). The following section is aiming to give a more detailed evaluation of the current situation by analysing the responses individually. A detailed list and more exhaustive summary of all answers can be found in Appendix B, Table 8.

The strongest support for one of the challenges the respondents currently see for the European LT community is that basic research is still needed. With approximately 90% of agreement, the statement represents the highest level of support for all questions asked in this regard (Figure 4). Further information on the results can be found in Appendix B, Table 8). In an open question where the participants had the option to list further obstacles or challenges, it was stressed that basic research is not necessarily seen as an obstacle, since basic research is always needed, but that Europe needs to invest in basic research in order to achieve digital self-determination and sovereignty.

The second highest level of agreement (81.6%) was obtained on the statement that one of the main challenges is an inadequate recognition of the importance of multilinguality. 82% of the respondents agreed or strongly agreed to this statement, making up more than



Figure 4: Level of agreement that basic research is still needed as one of the main challenges and obstacles the European LT community currently faces

three quarters of all responses (cf. Appendix B, Table 8). In the open question of this section, participants strongly ask the European Union to provide a high level of digital support for minority, regional or co-official languages through language technologies and resources in order to achieve the goal of digital language equality by 2030 in Europe. Furthermore, it is suggested to join forces for language families to overcome the challenge for developing smaller languages and to focus more on minority languages.

The third highest level of agreement is represented by the competition with big non-European companies and market disruption by global players. 79% of the respondents indicated that they agree or strongly agree, whereas only about 16% disagreed or strongly disagreed with this statement (cf. Appendix B, Table 8). It was also expressed by the participants that, in order to overcome Europe's fragmented AI LT landscape, conditions need to be created that allow for an ecosystem of growth based on Europe's current situation. This is also a mission that is strongly supported by CLAIRE and has been advocated in CLAIRE's vision for several years, asking not only for a European strategy, but also bringing together the still fragmented AI research activities and expertise in Europe, while at the same time creating centres of excellence and a structure that can efficiently focus on research and distribute results (Hoos et al., 2019, p. 4).

Europe is a continent that is represented by different talents, languages and cultures that have created and continue to create great achievements through interaction, joint forces and uniting its diversity. **CLAIRE strongly supports the EU's goal of standing "United in Diversity" and considers it crucial that Europe builds on its existing strengths not only in NLP but across the full spectrum of AI to foster AI excellence across Europe, also by ensuring diversity and inclusion across languages, cultures and gender.**

A lack of coordination and missing links between research, LT vendors, integrators and customers was also considered a main challenge for the European LT community with almost 66% agreement. It is worth mentioning that about 16% of the participants responded with "I don't know / No answer". This almost equals the proportion of disagreement where 18%

strongly disagreed or disagreed with the statement. Comparable results were obtained in the same category of questions with a focus on the fragmentation of the European LT industry, where almost 76% (strongly) agreed and 13% (strongly) disagreed (11% no answer).

A phenomenon that has been viewed very critically for several years by CLAIRE for all areas of AI also applies in part to LT. A total of 50% of the participants are convinced that lack of talent or brain drain across European borders is one of the main problems that the LT community in Europe currently faces. Apart from 8% of participants who decided not to answer the question, 42% disagree or strongly disagree that the lack of talent or brain drain represent one of the key challenges that the European LT community currently faces.

As also mentioned in the recently published CLAIRE response (CLAIRE, 2021) to the European Commission's proposed regulation on AI,⁹ Europe has experienced an outflow of great AI talents from all kinds of AI research areas to countries outside of Europe. After graduating, top young European AI researchers are often appointed elsewhere, attracted by highly paid internships and other more appealing offers. **CLAIRE, among others, has been warning against this dangerous trend since 2018.** According to **CLAIRE's assessment, Europe critically needs to stop the AI brain drain at all levels by rapidly introducing competitive conditions for attracting and retaining AI talent.**

One reason for the disagreement to the statement might be that participants do not see the problem in missing talent in Europe. CLAIRE, for example, believes that Europe is filled with great universities, great companies, great promise and potential. Furthermore, CLAIRE is convinced that Europe has one of the best AI labs, some of the best AI scientists – actually most of the best scientists in the world. But Europe trails other regions in turning new enabling technologies like AI into global products, services and platforms. **Therefore, CLAIRE recommends creating and rapidly deploying effective, light-weight, EU-wide mechanisms for attracting and retaining talent, such as an ERC programme in AI.**

Insufficient public procurement is seen as another challenge by over half (58%) of the participants but only 13% disagree with the statement. An important fact to note in this context is that one third of the participants decided not to answer the question. In the open question, one organisation mentioned that it agrees with the statement mainly in the context of minority languages. For more details on the results, see Appendix B, Table 8.

There is also a common agreement of just over half (53%) of the participants that insufficient markets to justify investments in LTs for smaller languages are one of the main challenges for the European LT community. On the other hand, 37% of the respondents do not agree. Looking at the answer categories, more people generally disagree than agree – a quarter of the respondents strongly agree. Four participants decided not to answer. Moreover, the cost of access to compute infrastructure is generally seen as an obstacle for the community, with 50% of the participants agreeing on the statement. The proportion of participants who disagree is almost as large as the proportion who agree.

Further challenges and obstacles that were mentioned by the participants include the lack of publicly available high-quality resources like datasets, corpora or lexica as well as the wish for more shared open resources.

In summary, these results do not only call for more coordination, but also for a more sophisticated exchange among research and industry. The respondents further express their wish to strengthen initiatives that focus on an exchange between these groups. According to the open question answers, initiatives in NLP, such as European Language Grid (ELG) and European Language Equality (ELE), need to be strengthened. A similar initiative to promote the exchange between industry and research has been launched by CLAIRE, namely the Innovation Network.¹⁰ Also, CLAIRE is strongly fostering the collaboration between industry and

ELE

⁹ https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0001.02/DOC_1& format=PDF

¹⁰ https://claire-ai.org/innovation-network/

research through VISION.¹¹ with a dedicated work package on joint forces between academia and industry and other activities like the recently held Joint Series of Theme Development Workshops.¹²

3.4. Predictions and visions for the future

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

- 1. initiate a large-scale and long-term funding programme for European LT development,
- 2. continuous investment in the research infrastructures that support LT,
- 3. investment in the development of new methodologies for transfer/adaptation of resources/technologies to other domains and languages and
- 4. increase of the availability of qualified personnel on LT and incentives for talent retention.

The following section will provide a more detailed analysis of the individual questions. A more exhaustive summary of the results can be found in the Appendix B, Table 9.

Results from the survey have shown that the proposal to initiate a large-scale and longterm funding programme for European LT development was very well received. The results show that close to 90% of the AI community CLAIRE surveyed believed that a large-scale longterm funding program in European Language technology would be effective (ranging from very to moderately effective) to speed up the development and deployment of LT in Europe equally for all languages (Figure 5). Only 2.6% each found the initiation of a large-scale longterm funding programme moderately or slightly effective. Detailed statistics of the respondents' assessments on a large-scale long-term funding programme for European LT development are provided in Appendix B, Table 9. Furthermore, the results are also in line with CLAIRE's recommendation 9 in its response to the EC's proposed AI regulation, namely to implement key parts of the coordinated plan through longer-term mission-oriented investments, including a 7- and 10-year funding for research networks (CLAIRE, 2021).

Equally, the members of CLAIRE who responded to the survey support the proposal for continuous investment in the research infrastructures for LT. 92% considered this measure very effective or effective. None of the respondents found that this measure is not effective, only 3% answered with "moderately effective" and 5% with "I don't know / No answer" (Figure 6).

A very similar support and call for continuous investment in research infrastructures can also be found in the CLAIRE vision that was released in 2018, where it is noted that discussions within the European AI community have led to a very clear message that the research activity level, the coordination and collaboration in Europe needs to increase (Hoos et al., 2019, p. 3). **To achieve this goal, investment in AI research and in structures that allow effective collaboration and transfer of results is required.** Participants furthermore noted that the continuous investment does not only apply to research infrastructure but also

¹¹ https://vision4ai.eu

¹² https://www.vision4ai.eu/tdw/



Figure 5: Results for measure "Initiate large-scale, long-term funding programme for European LT development"

on basic research. It was also suggested to stimulate the development of large shared LT resources by making the releasing of data part of the investment in startups and research.

The CLAIRE LT community also supports the idea for investment in the development of new methodologies for transfer and adaptation of resources and technologies to other domains and languages. Almost 80% of the respondents considered this measure effective, and almost 40% even very effective. Only one organisation found that this measure might not be effective at all.

Also, the increase of availability of qualified personnel on LT and incentives for talent retention were indicated as an (very) effective tool to speed up the development and deployment of LT in Europe equally for all languages by almost three quarters of the participants. In fact, all participants considered this instrument as somehow effective (including slightly and moderately effective).

Further measures that the LT community found effective were the reinforcement of training and education initiatives, including undergraduate and master programs and vocational training in LT with almost 65% as well as initiating investment instruments and accelerator programs targeting LT start-ups with almost 62% votes on effective and very effective. With regard to education initiatives, the wish to include topics related to man-machine interfaces in school programs for data science or to inspire young generations at an early age by demonstrating talking robots was expressed in the open question at the end of the respective survey's section.

A total of approximately 58% of the participants thought that raising awareness of the benefits for companies, public bodies and citizens of the availability of online services, contents and products in multiple languages is an effective or very effective instrument to speed up the development and deployment of LT in Europe for all languages. Almost one third of the respondents found this measure to be slightly or moderately effective and 5% voted for "not effective at all". Further recommendations from the community were to extend the public appearances of LT to all kinds of media and to promote public releases of language resources,



Figure 6: Results for measure "Continuous investment in the research infrastructures that support LT"

like translation memories, for the public sector and others.

Two of the suggested instruments or policies in this section of the survey were overall considered to be only moderately or slightly effective. Half of the participants think that imposing content accessibility regulations¹³ are a (very) effective measure. More than one third of the answers indicated that the implementation would only be moderately or slightly effective. Equally, almost 8% believe that this might not be effective at all or withdraw their answer.

The second question that received only half of the votes for (high) effectiveness is the public procurement of innovative technology and pre-commercial public procurement. An observation similar to the question related to public procurement in Section 3.3 can also be made for the predictions and visions for the future situation, namely that a relatively high number of participants (26%) decided not to answer or do not know what to answer. The rest of the participants considered this measure to be slightly or moderately effective. None of the answers included "not effective at all". These results are also consistent with a widely held view in the research community that regulation can and should not be the primary means towards ensuring the success of "AI made in Europe".

Additionally, the participants of the survey were asked in two open questions where they see the key challenges that Europe needs to concentrate on with regard to 1) basic applied research and 2) innovation and the LT industry for a large-scale long-term funding programme dedicated to European Language Technology research.

Both questions were answered by only approximately 15% (i. e. 6 out of 38) of the survey participants. The key challenges that were listed cover a broad variety of areas, including transfer learning, i. e. cross-lingual learning, multimodal approaches to document understanding, i. e. video, image, speech and text, as well as language modeling. Also, multilingual large-scale language resources and infrastructures for large-scale model development and training were considered challenges that need to be overcome. Further-

¹³ Multimedia subtitling, readability, dubbing, availability of content in multiple languages etc.

more, participants suggest a promotion or adoption of standards, the harmonisation of resources and the implementation of interoperability. One of the suggestions was to address the fragmentation in basic research by offering long-term programmes to consortia of high-level teams with the possibility of new successor programmes that build on the outcomes of previous ones. As for the LT industry, it was suggested to start building on common data and product spaces and LT business network spaces. Small and medium LT businesses should be fostered through public procurement and the forces of academia and industry need to be strengthened. Conditions for growth and scaling of LT SMEs should be created beyond their national or language foci to transnational players. Another perception of the challenges that Europe needs to concentrate on in case of a large-scale long-term funding programme dedicated to European Language Technology research is that all participating countries in Europe need to be brought together under the same umbrella programme and in reaching an agreement upon a number of strategic objectives along a small number of axes and dimensions. Further, R&D activities need to be synchronised, sufficient computing resources need to be provided as well as a suitable digital infrastructure to support the above-mentioned umbrella programme.

4. Analysis of interviews

The interviews allowed for a more detailed view and explanation for the chosen answers. As in the LT developers survey, the interviewees were asked to respond to the current main challenges for the European LT community, the policies and instruments to speed up the development and deployment of LT in Europe equally for all languages as well as the key challenges a future ELE funding programme should address. Due to the small number of interviews conducted, conclusions drawn from this small number of in-depth interviews should be treated with caution.

Three interviews have been conducted from November to December 2021. One interview was conducted via a video meeting. The remaining two interviews were conducted by answering the questions in writing through a tailor-made survey questionnaire with the same questions as in the oral interview. The selected interviewees represent high-level experts in the field of Language Technology, including a professor for Language Technologies at University Pompeu Fabre and Senior Researcher at the Institute of Applied Linguistics at the same university, who has been involved in key EU projects for many years. Furthermore, the coordinator and project manager of the European Language Resource Coordination was interviewed, who is also involved in initiatives under CEF focusing on the provision of LT tools and services for the European public sector, activities within H2020/Horizon Europe focusing on MT/NLP development for low-resourced languages, as well as consultation activities with regard to NLP evaluation and development in Africa and India for the German Association for International Collaboration (GIZ). The third interview was conducted with the head of the Data Intelligence Group in France.

There has been a strong agreement from all interviewees on the statement that the competition with big non-European companies and market disruption by global players is one of the main challenges that the European LT community currently faces. It was mentioned that it is difficult for European LT providers, often SMEs, to compete with big global players in terms of resources and manpower. It is also stated that, in comparison to the US, there is no credible European leader that is competitive in terms of data availability, which in turn prevents Europe from being a world leader in this field. Furthermore, key players on the market possess the agility and manpower to address given problems and provide effective solutions in a short period of time, whereas European companies currently cannot compete.

Also, there is consent that insufficient markets to justify investments in LTs for smaller

languages represents an obstacle for the LT community in Europe. It was again stressed that fewer resources exist for smaller languages, but, on the other hand, that there are also several initiatives, e.g. by the European Commission and other stakeholders, to provide support for these languages.

For LT, the interviewees further agree that the European industry is very fragmented, especially in comparison to other countries outside of Europe, like the US, where the market is dominated by few global players. In their opinion, it is a problem of the market. The investments of European industries are smaller and have to invest in a market with relatively limited benefits. It is stated that larger companies tend to behave conservatively in the market and take few risks. Smaller companies often do not have the needed capacities to invest in European languages, due to the limited expected market gain. While the statement that basic research is still needed received the strongest support in the LT survey (Section 3.3), the interviewees differ in their opinions. On the one hand, there is support for the statement, especially with regard to low-resource languages, that there is the need to find suitable ways for development and to understand how to do more with less. On the other hand, there is a disagreement with the explanation that the history demonstrates that the problem in Europe is not basic research but the application and the use of the researchers' resources by industry. Furthermore, the transfer and uptake of research results by the industry was criticised as being too slow.

Moreover, different opinions on the challenge "cost of access to compute infrastructure" have been expressed. One part agreed with current insufficiency in this regard, but stated that the situation might be improved through initiatives, i. e. the European Commission's HPC cluster under DIGITAL. The other interviewee disagreed saying that the European Union is already funding large infrastructures, i. e. by buying large computing machines.

In terms of inadequate recognition of the importance of multilingualism, it was only partly agreed that it is a main challenge for the European LT community, but all interviewees agreed that there is a market-related problem. On the one hand, it was seen from an economical and market perspective: It was perceived that in Europe, there is more emphasis on multilingual identity and cultural assets than on market needs. On the other hand, the awareness and recognition is perceived as existent, whereas the means to bring the knowledge to the market is seen as missing.

Furthermore, lack of talent is generally not seen as a key challenge for Europe in Language Technology. As already assumed in the analysis of the results from the survey (Section 3.3), there is no agreement that there is a lack of talent; on the contrary, there is a belief that talent is existing, but there is a lack of innovative ideas and missing investments for more offers to attract and keep talents in Europe.

There is also common disagreement that a lack of coordination and missing links between research, LT vendors, integrators and customers represent an obstacle for the European LT community. It is argued that it is not a lack of coordination as links exist, i. e. through various initiatives like ELG, but that these links are not being exploited and used. Further, it was criticised that any official effort to coordinate is extremely slow and that there is a big discrepancy between the coordination efforts and the real market. Many coordinating projects are trying to link research with industry developers, but repeatedly face the problem that official coordination efforts are too slow in creating change. This puts smaller innovative companies and start-ups into a disadvantaged position in comparison to larger companies.

Further criticism and disagreement is expressed with regard to insufficient procurement. Several schemes and examples have been named by the interviewees, both at a European and national level. The problem is rather seen in the fact that solutions are often not provided by the European LT community but by big international players. It is also attributed to the long bureaucratic processes, for example in terms of contract and implementation. One of the interviewees suggested addressing the market by conducting a study on the balance of public support for small languages through the development of resources and data-sets as a source of revenue for the market in order to support these languages. An additional challenge was identified in the fact that board members, top managers in large companies and officials in key governmental organisations do not have the knowledge and background required to assess technological developments. It was therefore suggested to re-evaluate the requirements that the above-mentioned positions have to meet in order to be able to make the right decisions for future investments.

One of the proposals to speed up the development and deployment of Language Technology in Europe equally for all languages that was considered effective by all interviewees was the reinforcement of training and education initiatives. It was noted that this would be an effective approach, but that it does not prevent the brain drain that is happening in Europe. Education is considered an effective way to spread LT around the world, especially for rare languages and to make young people aware of the notion of ethics in the field of AI and NLP, specifically with regard to bias and fairness.

Another approach that was considered slightly or moderately effective is initiating investment instruments and accelerator programs targeting LT start-ups to speed up the development and deployment of LT in Europe equally for all languages. On the one hand, it is believed that this approach could make a change, on the other hand it is stated that such schemes have been existing for a long time, but that they did not make a significant difference in Europe compared to other big markets outside Europe. It was also seen as moderately effective to invest in the development of new methodologies for transfer and adaptation of resources and technologies to other domains and languages. The respondents criticised that even though it is considered effective, it is often not translated into products by European SMEs. It was suggested to invest in methods and technologies that require less data.

The proposal to raise awareness of the benefits for companies, public bodies, and citizens of the availability of online services, contents and products in multiple languages was considered only slightly effective in speeding up the development and deployment of LT in Europe. In one of the interviewee's opinions, raising awareness needs to be targeted, and technical knowledge needs to be requested with an option to retrain bodies in key positions for a more profound understanding of existing technologies and dissemination campaigns need to simplify certain aspects. One of the problems that are seen in this regard is that expectations are created that are far from reality causing harm for the market. It is again stressed that there is no problem of general dissemination but a question of having the right people in the right positions.

Furthermore, one of the interviewees supported the approach to continuously invest in the research infrastructures that support LT. Again, with regard to the development and the deployment of LT in Europe it was commented that results are often not translated to actual market products as it is the case in other markets outside of Europe. The other interviewee preferred not to answer due to a missing concrete definition of the term "research infrastructures" which is often used in different contexts.

A similar outcome can be observed with regard to how effective imposing content accessibility regulations can be in speeding up the development and deployment of LT in Europe. One interviewee considered the proposal as slightly effective, although such a measure may positively impact language equality, there's a risk of overburdening and obstructing European content providers. The other interviewee is convinced that Europe is too small to harm big markets by imposing content regulations. Furthermore, it was stated that the question is not so much related to language technology but rather to cultural aspects.

Public procurement of innovative technology was considered slightly effective by one of the interviewees, who argued that several relevant schemes under DIGITAL Europe (formerly CEF programme) do exist, but still solutions are often provided not by the European LT community, but rather by big international players. Another interviewee considered such a measure not effective at all, because of the slow administrative processes that both the national and European programmes impose. Additionally, the problem is seen in a lack of

means for writing technical specifications for LT calls by public administration, causing a great challenge for speeding up the development and deployment for LT in Europe. The initiation of a large-scale, long-term funding programme for European LT development was strongly supported by one of the interviewees, as such an umbrella programme could reserve and ensure the means and resources required for the LT and language communities in Europe, in order to equally support all languages. The second respondent considered an LT funding programme not effective anymore, arguing that such a discussion and consideration for an LT-dedicated programme has been considerably delayed as it has been more than three years since the European Parliament's resolution "Language equality in the digital age".¹⁴ While the discussion in Europe is only starting, in the meantime, other key players outside Europe have already started investing. An approach that was suggested during the interviews was to move away from cultural and identity aspects to concentrate on the industry and the market.

To speed up the development and deployment of LT in Europe equally for all languages it was also mentioned that, with data as a core element of LT, research in LT should also be assisted by a massive digitalisation and storage of specific resources for all languages. These resources would then need to be localised nationally instead of centralised. This way, Europe could help countries worldwide to develop digitalisation and design standards that could lead to open-source resources. Europe should also play an important role in enforcing that resources created by collecting human interactions on the web (e.g. search engines, chatbots, etc.) are open data, freely available to any user who uses them, according to an internationally recognised contract.

One of the interviewees identified sustainable LT, the development of hybrid techniques combining symbolic and numerical knowledge representation, and the quality of data as the three main challenges for basic and applied research on which a large-scale, long-term funding programme for research, development and innovation in European language technology should focus.

For two interviewees, the main challenges that a future ELE funding programme should address in terms of basic and applied research were that methods need to be developed that learn with less data. Furthermore, the challenge to create and collect relevant language resources needs to be overcome. Moreover, a suggestion made by one of the respondents was to establish an evaluation framework at a European level to trace the development in LT industry and infrastructure, referring to national programmes like Technolangue¹⁵ (France) or evalita¹⁶ (Italy), setting specific objectives for the evaluation that fit the needs of the European market.

A similar approach was suggested with regard to innovation and LT industry with the additional idea to develop interoperability standards for small players, enabling them to plug into different Europe-wide applications. To achieve this, the European Committee for Standardisation plays a modest role and by establishing these interoperability standards, it would be possible for smaller providers to create business alliances for different languages to come up with bigger multilingual technologies for a variety of different applications. This support for entrepreneurship was also mentioned by the other interviewee together with the challenge to ensure that research results are exploited and brought to market. Finally, it was stated that a long-term ELE funding programme, endorsed at the national and European levels, should not be a general programme but one that is targeted at the development of Language Resources and Language Technology for public services to enable inclusion.

¹⁴ Language equality in the digital age, European Parliament resolution of 11 September 2018 (https://eur-lex. europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018IP0332&from=EN)

¹⁵ http://www.technolangue.net

¹⁶ https://www.evalita.it



5. Conclusions

The results from the LT developers survey and the conducted interviews have shown clear results and statements not only for the main challenges in the current situation, but have also provided a clear outlook for the future of LT in Europe. Furthermore, they also reflect what CLAIRE believes is necessary to ensure the success of "AI made in Europe". The challenges and obstacles that the LT community currently sees can be summarised as follows:

1. Research and Technology:

- a) There is very strong support for the statement that **basic research is still needed**, also with regard to low-resource languages. To achieve digital self-determination and sovereignty, Europe needs to invest in basic research.
- b) There is a **lack of coordination and missing links between research and industry**. To solve this, (existing) resources from research need to be used and applied by industry, and the transfer and uptake of these resources need to be coordinated more quickly, which would also support smaller innovative companies or startups.

2. Data:

- a) The **lack of publicly available high-quality resources**, such as data sets, corpora or lexica, poses an obstacle for the LT community. The survey participants ask for more shared open resources, and the competitiveness in Europe in terms of data availability needs to be strengthened to maintain global competitiveness and leadership in this field.
- b) Furthermore, to overcome this obstacle, there is a need for research on how more results can be obtained with less data resources.

3. Market:

- a) The **competition with big non-European companies and market disruption by global players** is seen as one of the main challenges that need to be overcome, as it is difficult for European LT providers, often SMEs, to compete with big global players in terms of resources and manpower. Key players on the market address given problems and provide effective solutions quickly, whereas in Europe, companies currently cannot compete.
- b) **Insufficient markets to justify investments in LTs for smaller languages** are a major obstacle for the European LT community. There are several initiatives by the EC and other stakeholders to provide support for smaller languages, still, fewer resources exist for these languages.

4. Innovation:

- a) One of the major current challenges for the European LT community is the **strong fragmentation of the European LT industry**, especially with regard to other countries outside of Europe, with few global players on the market with higher investments. In Europe, smaller companies often do not have the needed capacities to invest in European languages. To overcome the fragmented AI LT landscape, conditions need to be created that allow for an ecosystem of growth based on Europe's current situation to foster innovation.
- b) The ongoing **brain drain** represents a challenge to retain AI talent in Europe. More offers, innovative ideas and investments are needed to attract these talents in Europe.

5. Legal and organisational matters:

- a) The **inadequate recognition of the importance of multilinguality** represents a major challenge for the LT community, which is also considered to be related to the market. The European Union is asked to give a high level of digital support for minority, regional or co-official languages by providing language technologies and resources to achieve the goal of digital language equality in Europe by 2030.
- b) **Insufficient public procurement** is seen as another challenge. Solutions are often not provided by the European LT community but by big international players, which is also attributed to the long bureaucratic processes, for example in terms of contract and implementation. Europe needs to act faster to make a change.

For a more detailed analysis of the survey responses, please see Section 3 and the analysis of the interviews in Section 4. The five most effective instruments to solve the abovementioned challenges and to speed up the development and deployment of Language Technology in Europe, as seen by the LT community, are listed in the following according to the level of support:

- 1. Initiation of a large-scale and long-term funding programme
- 2. Continuous investment in the research infrastructures for LT
- 3. Investment in the development of new methodologies for transfer and adaptation of resources and technologies to other domains and languages
- 4. Increased availability of qualified personnel on LT and incentives for talent retention
- 5. Reinforcement of training and education initiatives, including undergraduate and master programs and vocational training in LT

Many of the concerns, challenges and recommendations expressed by the various opinions of the LT community in this survey are also reflected in the many statements made by CLAIRE, not only related to NLP, but to the whole AI landscape in Europe.

CLAIRE sees the main obstacles that need to be overcome in the complex and highly fragmented nature of the AI ecosystem and the measures put into place to strengthen it but also in a lack of effective funding instruments for stimulating research and innovation, a lack of strategic EU-level activities that attract global attention or generate global impact, and a lack of coordination between different activities in order to reach larger goals. To preserve Europe's unique selling point, namely its manifold diversity, more language resources, data-sets and methods need to be collected, shared and developed to maintain this diversity not only on a cultural level but also on an economic level. In order to prevent European languages from becoming digitally extinct, to achieve European language equality by 2030 and not to lag behind other markets, not only must attractive offers be made for top researchers in Europe, but forces must be united, existing resources must be optimally used, future resources must be sustainably promoted and financed, but action must also be taken immediately.

CLAIRE strongly believes that Language Technology and NLP play a key role not only in Europe but also around the world. CLAIRE highly supports the mission to leverage the capabilities and potential in LT and NLP for the benefit of everyone. These results are therefore calling for raising awareness for the LT potential in Europe on a political level as well as for a strong commitment from the European Union as well as from the European Commission.



References

- CLAIRE. CLAIRE Response to the European Commission Proposal for AI Regulation and 2021 Coordination Plan for AI, August 2021. https://claire-ai.org/wp-content/uploads/2021/08/CLAIRE-EC-AI-Regulation-Feedback.pdf.
- Holger Hoos, Morten Irgens, and Philipp Slusallek. A European Vision for AI Call for the Establishment of a Confederation of Laboratories for Artificial Intelligence Research in Europe (CLAIRE), October 2019. https://claire-ai.org/wp-content/uploads/2019/10/CLAIRE-vision.pdf.
- Georg Rehm and Hans Uszkoreit, editors. *META-NET White Paper Series: Europe's Languages in the Digital Age*, 32 volumes on 31 European languages, Heidelberg etc., 2012. Springer.



A. The LT researchers and developers full survey

Figures 7 to 15 show the complete LT research and developers survey.



Figure 7: Full survey as published (page 1/9)



No personal da The names and respondents' vi in the project's etc., without an	nvite responde ta of the respo d emails of the iews and opinio deliverables or y reference to	nts to follow-up interviews or to the ELE conference or other project events. andents will be made available to any third-party, beyond the ELE consortiur respondents will not be reported in any project public document. The bons, as expressed through this questionnaire, may be reported anonymous r in other public documents, e.g. scientific publications, dissemination mater the individual's personally identifiable information.
Please read the	e ELE Privacy	policy to get informed about the processing of your personal data when
ning in this qu	estionnaire.	
1 Introduce	e yourself a	and your organisation
Which of the foll	owing best desc	cribes the type of organisation you work for?
O Universit	y or other acade	emic research organisation
Researcl	h center (indepe	endent from any other academic organisation)
SME		
Large en	terprise	
Other		
lf "Other" place	a specify	
	o opeony.	
What is the nam	e of the organis	ation you work for?
• What is the nam If applicable, pleas	e of the organis	ation you work for? ee of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics
• What is the nam If applicable, pleas /School of Philolog	e of the organis se provide the nam gy/University of Ath	tation you work for? le of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics nens.
What is the nam If applicable, pleas /School of Philolog	e of the organis se provide the nam gy/University of Ath	vation you work for? ne of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics nens.
What is the nam If applicable, pleas //School of Philolog	e of the organis se provide the nam gy/University of Ath	ation you work for? e of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics eens.
What is the nam If applicable, pleas /School of Philolog . Where is your of	e of the organis se provide the nam gy/University of Ath rganisation's he	ation you work for? te of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based?
• What is the nam If applicable, pleas /School of Philolog	e of the organis se provide the nam gy/University of Ath rganisation's he Germany	ation you work for? le of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based? Netherlands
• What is the nam If applicable, pleas /School of Philolog • Where is your or © Austria © Belgium	e of the organis se provide the nam gy/University of Ath rganisation's he Germany Greece	ation you work for? le of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based? Netherlands Norway Protocol
• What is the nam If applicable, pleas /School of Philolog • Where is your or	e of the organis se provide the nam gy/University of Ath rganisation's he Germany Greece Hungary	ation you work for? te of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based? Netherlands Norway Poland Poland
• What is the nam If applicable, pleas /School of Philolog • Where is your or	e of the organis se provide the nam gy/University of Ath rganisation's he Germany Greece Hungary Iceland	ation you work for? te of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based? Netherlands Norway Poland Portugal Romania
• What is the nam If applicable, pleas /School of Philolog • Where is your or	e of the organis se provide the nam gy/University of Ath rganisation's he Germany Greece Hungary Iceland Ireland	ation you work for? te of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based? Netherlands Norway Poland Portugal Romania Slovek Popublic
• What is the nam If applicable, pleat /School of Philolog • Where is your or	e of the organis se provide the nam gy/University of Ath rganisation's he Germany Greece Hungary Iceland Ireland Ireland	ation you work for? te of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based? Netherlands Norway Poland Portugal Romania Slovak Republic Slovaenia
What is the nam If applicable, pleat /School of Philolog /////////////////////////////////	e of the organis se provide the nam gy/University of Ath rganisation's he Germany Greece Hungary Iceland Ireland Ireland Lativa Lativa	ation you work for? te of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based? Netherlands Norway Poland Portugal Romania Slovak Republic Slovenia Soain
• What is the nam If applicable, pleat /School of Philolog • Where is your or	e of the organis se provide the nam gy/University of Ath rganisation's he Germany Greece Hungary Iceland Ireland Iteland Italy Latvia Lithuania	ation you work for? te of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based? Netherlands Norway Poland Portugal Romania Slovak Republic Slovenia Spain rg Swerden
• What is the nam If applicable, pleat /School of Philolog • Where is your or	e of the organis se provide the nam py/University of Ath rganisation's he @ Germany @ Greece @ Hungary @ Iceland @ Ireland @ Italy & Latvia @ Luxembour @ Luxembour @ Luxembour	ation you work for? te of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based? Netherlands Norway Poland Portugal Romania Slovak Republic Slovenia Spain g Sweden Other
What is the nam If applicable, pleat /School of Philolog Where is your or Austria Belgium Bulgaria Croatia Cyprus Czechia Denmark Estonia Finland France	e of the organis se provide the nam gy/University of Ath rganisation's he Germany Greece Hungary Iceland Ireland Italy Latvia Latvia Luxembour Malta	ation you work for? te of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based? Netherlands Norway Poland Portugal Romania Slovak Republic Slovenia Spain rg Sweden Other
What is the nam If applicable, pleat /School of Philolog Where is your or Austria Belgium Bulgaria Croatia Cyprus Czechia Denmark Estonia Finland France	e of the organis se provide the nam yy/University of Ath Greece Germany Iceland Ireland Italy C Latvia Lithuania Luxembour Malta	ation you work for? te of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based? Netherlands Norway Poland Portugal Romania Silovak Republic Silovenia Spain rg Sweden Other
• What is the nam If applicable, pleat /School of Philolog • Where is your or	e of the organis se provide the nam y/University of Ath Garence Hungary Iceland Ireland Lithuania Luxembour Malta e specify.	ation you work for? te of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based? Netherlands Norway Poland Portugal Romania Slovak Republic Slovenia Spain rg Sweden Other
What is the nam If applicable, pleat /School of Philolog Where is your or Austria Belgium Bulgaria Croatia Cyprus Czechia Denmark Estonia Finland France	e of the organis se provide the nam yy/University of Ath " Gareace Hungary Iceland Ireland Lithuania Luxembour Malta e specify.	ation you work for? te of the LT-specific group within the organisation first, e.g. NLP group/Department of Linguistics tens. adquarters based? Netherlands Norway Poland Portugal Romania Slovak Republic Slovenia Spain rg Sweden Other

Figure 8: Full survey as published (page 2/9)



Which LT areas do you r	nainly work in?
Basic natural lang	juage processing services (PoS tagging, parsing, named entity recognition etc.)
Search and information	nation retrieval technologies
Text analytics an	d mining, information extraction, text classification
Translation techn	ologies (Machine Translation, translation memories management, CAT tools)
Speech technolo	jies
Conversational s	ystems
Language resour	ces: data production, data aggregation
Language resour	ces: data distribution, data marketplace
Research infrastr	uctures (e.g. catalogue, repository)
Other	
If "Other", please specify	·
Are you/your organisatio	TA member of one or more of the following associations/networks/projects?
META-NET	Al4Media
ELG	
	Al4Copernicus
LT-Innovate	AlPlan4EU
AI4EU	BonsAPPs
	DIH4AI
BDVA	I-NERGY
AI PPP	StairwAl
🔲 HumanE AI Netw	ork 🔲 Other
Nexus Linguarun	None of the above
ELISE	
If "Other", please specify	
How many organisations	participate in your national CLARIN consortium?
How many LT researche	rs/experts/students are employed and/or actively contribute to the national CLARIN
Please do not report the pur	nher of students using the resources in education only. Only the number of active contributors is releva
here.	nor or stations using instreadures in education only. Only the number of active contributions is releva
	tooppologios, products or sorvices used?
In which sectors are you	

Figure 9: Full survey as published (page 3/9)



Digital Humar	ities, arts, culture and	other services 🔲 Justice and legal
Broadcasting		Media
Business server	vices	Public administration
Construction		Publishing
eCommerce		Security (threat detection in general)
Education		Social Sciences
Energy/green	economy/environment	t 🔲 Tourism, accommodation and food services
Finance/bank	ing	Trade and repair
Health		Transportation, logistics and storage
Industry and i	nanufacturing	Other
Information a	nd Communication Tec	chnologies
If "Other", please spe	cify.	
2 Language c	overage	
L Languago o	oronago	
What languages doe	s your organisation con	nduct research in and/ or for what languages do you offer service
software, resources,	models etc.?	
Basque	Galician	Norwegian
Bulgarian	🔲 German	Delish
Catalan; Vale	ncian 🔲 Greek	Portuguese
Croatian	🔲 Hungarian	Romanian
Czech	lcelandic	Serbian
Danish	Irish	Slovak
Dutch	ltalian	Slovenian
English	Latvian	Spanish Spanish
Estonian	Lithuanian	Swedish
Finnish	Luxembourgi	ish 🔲 Welsh
French	Maltese	Other
If "Other", please spe	cify.	
Please separate multiple	e languages with a comma	l (,).
	ges that your organisat	tion does not yet support, but you plan to support in the next three
Are there any langua		
Are there any langua years?	Galician	Norwegian
Are there any langua years? Basque		Polish
Are there any langua years? Basque Bulgarian	German	
Are there any langua years? Basque Bulgarian Catalan; Vale	German ncian Greek	Portuguese
Are there any langua years? Basque Bulgarian Catalan; Vale Croatian	German ncian Greek Hungarian	Portuguese
Are there any langua years? Basque Bulgarian Catalan; Vale Croatian Czech	German ncian Greek Hungarian Icelandic	PortugueseRomanianSerbian
Are there any langua years? Basque Bulgarian Catalan; Vale Croatian Czech Danish	German ncian Greek Hungarian Icelandic Irish	 Portuguese Romanian Serbian Slovak
Are there any langua years? Basque Bulgarian Catalan; Vale Croatian Czech Danish Dutch	German ncian Greek Hungarian Icelandic Irish Italian	 Portuguese Romanian Serbian Slovak Slovenian

Figure 10: Full survey as published (page 4/9)



English Latvian S Estonian Lithuanian S Finnish Luxembourgish W French Maltese O	oanish wedish 'elsh ther				
f "Other", please specify. Please separate multiple language with a comma (,).					
Considering your development plans with respect to decision to support additional languages? <i>at most 3 choice(s)</i> Please choose a maximum of 3. Market interest/demand by users or custome Research/scientific interest Available funding/investment Availability of human experts for other langua Availability of language resources Availability of technologies/tools	language cove rs ages	erage, what	are the top t	hree drivers f	ior your
Conter Other Tother", please specify. Content Sector Se	nents: "One of	the main o	challenges a	nd obstacles	; the
Conter Conte	nents: " One of Strongly agree	the main of Agree	challenges at Disagree	nd obstacles Strongly disagree	the
Conter Conte	nents: "One of Strongly agree	the main of Agree	Disagree	nd obstacles Strongly disagree	the
Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter	nents: "One of Strongly agree	the main of Agree	Disagree	nd obstacles Strongly disagree	s the
Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter	nents: "One of Strongly agree	Agree	Disagree	nd obstacles Strongly disagree	the
Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter	nents: "One of Strongly agree © 0 0	the main of Agree	Disagree	nd obstacles Strongly disagree © ©	the
Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter Conter	nents: "One of Strongly agree © © © ©	the main of Agree	Challenges and Disagree	nd obstacles	the

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



 insufficient markets to justify investments in LTs for smaller languages." 	0	۲	\odot	O	C
*cost of access to compute infrastructure."	0	0	\odot	0	C
 competition with non-European big companies and market disruption by global players." 	O	۲	۲	O	C
If you wish, please elaborate on the obstacles and cha other obstacle/challenge that was not previously listed	allenges indic	ated in the	previous que	stion and/or a	add any
4 Predictions and visions for the fut	ure				
	ure				

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

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

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



Figure 14: Full survey as published (page 8/9)



hein	ere any other policies/instruments not listed in the previous question, which in your opinion can be effect peeding up the development and deployment of LT in Europe equally for all languages?
	peeding up the development and deployment of LT in Europe equally for an languages :
If ther	e is a large-scale, long-term funding programme dedicated to European Language Technology research
develo challe	pment and innovation running for approx. ten years, what are, in your opinion, the (up to) five key nges Europe needs to concentrate on with regard to basic and applied research?
If there develo	e is a large-scale, long-term funding programme dedicated to European Language Technology research, pment and innovation running for approx. ten years, what are, in your opinion, the (up to) five key nges Europe needs to concentrate on with regard to innovation and the LT industry?
Do yo	a have any other additional suggestions or recommendations with regard to European Language Equalit
Do yo	u have any other additional suggestions or recommendations with regard to European Language Equalit e contact you to arrange a possible follow-up discussion?
Do yo	a have any other additional suggestions or recommendations with regard to European Language Equalit e contact you to arrange a possible follow-up discussion? Yes No
Do yo Can w O What	u have any other additional suggestions or recommendations with regard to European Language Equalit e contact you to arrange a possible follow-up discussion? Yes No s your email address?
Do yo Can w What	a have any other additional suggestions or recommendations with regard to European Language Equalit e contact you to arrange a possible follow-up discussion? Yes No s your email address?
Do yo Can w What What	u have any other additional suggestions or recommendations with regard to European Language Equalit e contact you to arrange a possible follow-up discussion? Yes No s your email address?
Do yo	a have any other additional suggestions or recommendations with regard to European Language Equalit e contact you to arrange a possible follow-up discussion? Yes No s your email address? s your email address? clicking on 'Submit', I agree that my personal data (email address and/or name) can be used according vacy Policy of the European Language Equality (ELE) project.
Do yo	a have any other additional suggestions or recommendations with regard to European Language Equalit e contact you to arrange a possible follow-up discussion? Yes No s your email address? s your name? clicking on 'Submit', I agree that my personal data (email address and/or name) can be used according vacy Policy of the European Language Equality (ELE) project. ELE Privacy Policy.pdf

Figure 15: Full survey as published (page 9/9)

B. Additional tables and graphs

Type of organisation	Answe	ers (Perc.)
Research center (independent)	8	21%
University or academic research	28	74%
Large enterprise	1	3%
SME	1	3%
Other	0	0%
Total	38	

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

Country	Respondents (Perc.)			
Austria	1	2.6%		
Belgium	1	2.6%		
Bulgaria	1	2.6%		
Estonia	1	2.6%		
France	2	5.3%		
Germany	3	7.9%		
Greece	5	13.2%		
Italy	2	5.3%		
Luxembourg	1	2.6%		
Netherlands	3	7.9%		
Poland	1	2.6%		
Portugal	4	10.5%		
Slovak Republic	2	5.3%		
Spain	6	15.8%		
Switzerland	5	13.2%		
Total	38			

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

Basic natural language processing services (PoS tagging, parsing, named entity recognition etc.)	26
Search and information retrieval technologies	19
Text analytics and mining, information extraction, text classification	32
Translation technologies (Machine Translation, translation memories management, CAT tools)	11
Speech technologies	14
Conversational systems	19
Language resources: data production, data aggregation	18
Language resources: data distribution, data marketplace	8
Research infrastructures (e.g. catalogue, repository)	11
Other	2

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

Language	Mentions (count)
Basque	3
Bulgarian	1
Catalan	4
Dutch	4
English	36
Estonian	1
French	12
Galician	2
German	12
Greek	5
Italian	5
Luxembourgish	1
Other	7
Polish	1
Portuguese	6
Slovak	2
Spanish	8

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



Maltese

Other

Polish

Norwegian

Portuguese

Romanian

Slovenian

Spanish

Swedish

Welsh

Serbian

Slovak

Language	Mentions (count)
Basque	1
Bulgarian	1
Catalan	1
Croatian	1
Czech	3
Danish	1
Dutch	2
English	1
Estonian	1
Finnish	1
French	5
Galician	2
German	1
Greek	1
Hungarian	1
Icelandic	1
Irish	1
Italian	4
Latvian	1
Lithuanian	1
Luxembourgish	1
Maltese	1

1

4

5

3

1

1

1

1

4

1

1

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

Drivers	Mentions (count)
Market interest/demand by users or customers	19
Research/scientific interest	30
Available funding/investment	18
Availability of human experts for other languages	5
Availability of language resources	12
Availability of technologies/tools	5
Other	6

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

	Strongly agree	Agree	Disagree	Strongly dis- agree	I don't know / No an- swer
basic research is still needed	18	16	3	0	1
inadequate recognition of the importance of multilinguality	13	18	4	1	2
lack of talent/brain drain	8	11	9	7	3
fragmentation of the Euro- pean LT industry	9	20	3	2	4
lack of coordination and miss- ing links between research, LT vendors, integrators and cus- tomers	8	17	6	1	6
insufficient public procure- ment	10	12	4	1	11
insufficient markets to justify investments in LTs for smaller languages	10	10	11	3	4
cost of access to compute in- frastructure	8	11	13	3	3
competition with non- European big companies and market disruption by global players	15	15	6	0	2

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

	Very effec- tive	Effective	Moderately effective	Slightly effec- tive	Not effec- tive at all	I don't know / No answer
Initiate long-scale, long-term funding programme for Euro- pean LT development	21	13	1	1	0	3
Initiate investment instru- ments and accelerator pro- grams targeting LT start-ups	6	18	10	1	1	3
Continuous investment in the Research Infrastructures that support LT	18	18	1	0	0	2
Increase availability of quali- fied personnel on LT and in- centives for talent retention	13	16	6	0	2	2
Public procurement of inno- vative technology and pre- commercial public procure- ment	6	13	8	2	0	10
Raise awareness of the bene- fits for companies, public bod- ies, and citizens of the avail- ability of on-line services, con- tents and products in multiple languages	5	17	9	2	2	2
Impose content accessibility regulations, e.g., multimedia subtitling, readability, dub- bing, availability of content in multiple languages etc.	5	14	8	6	3	3
Invest in the development of new (scientific/technological) methodologies for trans- fer/adaptation of re- sources/technologies to other domains and languages	15	16	4	2	1	1
Reinforce training and educa- tion initiatives, including un- dergraduate and masters pro- grams and vocational training in LT	11	14	8	4	1	1

Table 9: Answers to the question: "In your opinion, how effective can the following policies/instruments be in speeding up the development and deployment of LT in Europe equally for all languages?" (mandatory closed question, answers provided on a five-point scale, plus "I don't know/No answer")