D2.4
Report from LT-Innovate

Version: 1.0 (FINAL)
Date: 16-FEB-2022

Table of Content

Table of Figures:

About this document

Project	European Language Equality (ELE)
Grant Agreement no.	LC-01641480 – 101018166 ELE
Coordinator	Prof. Andy Way (DCU)
Coordinator	Dr. Georg Rehm (DFKI)
Start Date:	01-01-2021, 18 months
Deliverable number	D2.4
Deliverable title	Report from LT-innovate
Type	Report
Number of pages	26
Status and version	Final
Dissemination level	Public
Date of delivery	Contractual: 28-02-2022 – Actual: XX-XX-2022
Authors	Andrew Rufener (Crosslang) with contribution from Philippe Wacker (LT-Innovate)

1 Introduction

1.1 Background & Context

This deliverable is one of five (D2.2 to D2.6) within the first task (Task 2.1) of WP2 of the European Language Equality (ELE) project. WP2 is aimed at collecting input for the strategic agenda and producing several reports by a broad spectrum of stakeholders – from research through industry to users – about their views, needs and perspectives related to language technologies and digital language equality, while having in mind expected developments over the next 10 years that aim at the identification of the needs and visions from research and industrial networks and associations for digital language equality in 2030.

More specifically, Task 2.1 "The perspective of European LT developers (industry and research)" aims at investigating the ideas, demands, future visions and predictions of this diverse group of stakeholders. To this end, it explored the factors that drive their development plans and investments and the perceived obstacles to overcome to achieve digital language equality.

In this particular document D2.4, the needs and visions of the LT industry were collected and analysed, using the membership of LT-Innovate, the Language Technology Industry Association (www.lt-innovate.org) and network of CrossLang (https://www.crosslang.com) respectively. The fact finding conducted by CrossLang, with support from LT-Innovate, was aimed at coalescing the views, suggestions and needs of the European Language Industry through surveys and focused interviews with senior industry representatives to help establish a coherent and consolidated view. As such, the D2.4 findings contribute the industry's views to T2.1 and the overall ELE project.

1.2 About LT Innovate and CrossLang

LT-Innovate,[1] founded in 2012, is the Language Technology Industry Association. Its founders and members are listed on its website (http://www.lt-innovate.org/about/founders & http://www.lt-innovate.org/directory).

LT-Innovate supports the LT industry and its members by:

- Promoting the industry as a whole in the most promising target markets;
- Promoting Language Intelligence as driver of economic success, societal well-being and cultural integrity;
- Supporting capacity development, portfolio pooling, productisation and platformisation to allow LT companies to stand a better chance in global markets;
- Sharing knowledge, leveraging the usage of key resources / infrastructures / tools, and pooling complementary skills / capacities for market development;
- Supporting the public policy-making process and influencing all relevant high-level policy documents, regulations and support programmes that affect the LT sector and its markets;
- Building the case for a language-neutral digital content and communication market in Europe and globally.

Membership in LT-Innovate is open to any organisation or individual with an interest in language technology. LT Innovate currently has around 300 members from around the globe.

[1] www.lt-innovate.org

CrossLang

CrossLang[2] is a European leader in translation automation and language technology since 2002. CrossLang provides consulting, systems integration, operation, and research to help organizations modernize not just their translation infrastructure, but their translation approaches as well. It’s a holistic approach to “translation modernization” encompassing translation automation, infrastructure, people, and processes, enabling our clients to be faster, more cost effective and more capable in the process.

Operating both in research and public sector as well as European and global corporations and their technology suppliers, CrossLang has a unique network within the commercial language technology sector.

[2] **www.crosslang.com**

2 Methodology & Instruments

Elicitation of the LT industry representative views was carried out utilising a two-pronged approach as follows:

1. **ELE Online Survey**
   The online survey developed under T2.1 was distributed via email, newsletters and social media requesting online feedback. The survey was distributed using the ELE's own mailing list as well as the CrossLang's and LT-INNOVATE's mailing lists and social media. Unfortunately, the response rate from the CrossLang and LT-INNOVATE contacts was low, with only 18 respondents of which 7 where from an academic institution. With a response rate of around 0.3% the response rate was disappointing, yet not totally unexpected given the fact that the target audience for CrossLang and LT-INNOVATE were industry executives and senior managers mainly.

2. **Industry-specific Questionnaire and Interviews (Annex 1)**
   Provided the limited industry response rate to the ELE Survey, the project team proceeded to developing and distributing an industry-specific questionnaire to a target list of 48 companies (see list in Annex 3). This questionnaire was sent through individual and personalised e-mails. As the responses to this questionnaire were laborious to obtain, the questionnaire was complemented by interviews with senior industry executives, in particular where no or not enough information was elicited from the written feed-back. These interviews lasted on average one hour. Altogether, CrossLang and LTI carried out 29 on-hour interviews.

2.1 ELE Online Survey

The ELE Online Survey was distributed to approx. 6000 recipients (total number of people reached via the combined CrossLang and LT-Innovate contacts databases and social networks).

Herewith an example of the announcement included in the LT-Innovate NewsFlash in June, August and September 2021:

OTHER NEWS

Invitation to participate in the European Language Equality consultation process through an online questionnaire.

The European Language Equality (ELE) project (supported by LT-Innovate) comes in response to the European Parliament's Resolution on Language Equality in the Digital Age. Its primary goal is the preparation of a European Strategic Research and Innovation Agenda and Roadmap, in order to tackle the striking imbalance between European languages in terms of the support they receive through language technologies.

To this end, ELE is reaching out to all European stakeholders involved in Digital Language Equality through a series of consultation rounds. This questionnaire is specifically addressed to **practitioners in LT including NLP, Speech Technologies and Language-centric AI**.

The questionnaire takes approx. 20 minutes to fill in. You are asked to evaluate the current situation with respect to the level of LT support for European languages, to indicate challenges and to share your needs and expectations for the future. Your contributions will be carefully taken into account when preparing the ELE strategic agenda and roadmap.

This is a joint pan-European effort that will impact the field of LT in Europe for the next 10-15 years, including the funding situation.

Take part in the survey

3. Unfortunately, it turned out that obtaining responses to this questionnaire was an uphill battle. All in all, 18 respondents declaring they were members of LT-Innovate filled in the survey. 7 questionnaires were filed by research organisations. Since the purpose of this deliverable was to supply an industry viewpoint, the research organizations responses were not considered. The remaining 9 responses were processed alongside the feedback obtained through the industry-specific questionnaire/interviews and have been integrated in the findings and recommendations outlined in the present report.

2.2 Industry-specific Questionnaire/Interviews

To supplement the ELE survey, it was decided to attempt an approach specifically designed for industry, targeting a shortlist of 51 industry representatives (see list in Annex 3), contacted via personalised e-mails. This restricted e-mail campaign produced another 12 questionnaires. To obtain more detailed feedback and insight, these questionnaires were then followed-up conduct interviews. All in all, it was possible to conduct 29 interviews and the most senior industry representatives only made themselves available to interviews but did not respond to any questionnaire.

The industry-specific questionnaire was structured around the following key themes:

Q1: Company Details: It is obvious that company type, and size plays a role in its activities and therefore, in determining needs and outlook towards the market. At the same time, the current outreach of a company allows for a more or less accurate and objective prognostics of where we would stand in 10 years' time, in particular in its area of activities, both sectoral and geographical. Therefore, the question was subdivided into turnover (optional), staff (FTE), and keywords to describe the activities. Furthermore, where possible, the web address of audio-visual and social channels was requested, since experience shows that these channels are more effective in determining the activities of a company than a mere website.

Q2: Markets: This question is split into four sub-questions: Firstly, regarding the markets in which the respective company is active (sectoral and geographical), secondly, which sector will be most in need for multilingual solutions and, thirdly, how they, from their position, expect that the relation between AI and LT/LI evolve. The latter question is of strategic importance for the future of LT: should it be/remain a technology domain on its own (with all the subsequent consequences such as a specific funding portfolio, specific statistics regarding market penetration and market opportunities) or will it be subsumed under the umbrella of AI?

Q3: Challenges: By the nature of the survey, this was the most important yet complex question, divided into 3 sub-questions whereby answers could be ticked but subsequently explained (open ended questions). It was important to distil meaningful indicators from the answers for future developments and specifically for the roadmap for the next 10 years. The sub-questions/answer options were the following and provided guidance also for the interviews conducted:

Q3/1: What are the biggest challenges you encounter in developing and growing your business?
Answers could be selected among the following thematic groups:

- Legal/regulatory
- Bureaucratic hurdles/delays
- Costs of doing business
- Lack of financing
- Potential clients who are badly informed about available solutions
- Potential clients who are risk averse and not prepared to invest
- Access to potential clients blocked off by intermediaries
- (Unfair) competition from abroad
- Technical challenges, e.g., scarcity of data, language resources

Q3/2: Which languages remain challenging for you to integrate in your offerings? Why?

Q3/3: Which research breakthroughs would you like to see happening in the next 5-10 years?

And at the end an option for “Other” was provided in case a situation did not fit any of the above challenges.

Q4: Collaborative Initiatives
Are you involved in any collaborative initiatives around NLP/LI, AI, Cloud?

This question was added to allow for the assessment of how collaborative the sector acts, be it on a commercial basis, be it in EC or other government funded projects or other collaborative initiatives. It also provided an indicator of the industry’s appetite for dedicating funding and resources to joint development programs as well as the rationale for doing so.

Q5: European Union Policies & Programmes

This question was again sub-divided into topics relevant for the outlook for the next 10 years at EU level regarding actions to be taken.

Q5/1: To what extent are you aware of EU language related projects and resources?

Q5/2: To what extent do you think EU initiatives are sustainable, i.e. able to deliver structural benefits to businesses over the coming years?

Q5/3: What do you expect from the EU in the next 5 years?
These questions, again, allowed for multiple responses plus open-ended explanations:

- More funding programmes for companies?
  - If yes, what should be improved in the existing funding programmes?
  - What types of projects should the EU support?
- Focus on 'finishing' the Internal Market - what is missing from your point of view?
- Financial guarantees for small companies that want to levy investment
- A baseline language infrastructure in all EU languages
- Do you know the ELG project [https://www.european-language-grid.eu/]? Are you participating in it? Do you consider it useful? Why? How could it become (more) useful to you?
  - Make available multilingual data assets held by European institutions (Euramis, EuroVoc, IATE, etc.), if necessary, by anonymising them
  - Require that all public services become available in all EU languages
  - Require that public institutions buy mainly/only from European suppliers
  - Other – explanation

These questions were included to gauge the level of awareness and understanding of EU initiatives, the perception thereof and the willingness to engage and/or apply results of such projects in a commercial setting.

Please note that a confidentiality clause was included in the questionnaire, to warrant that responses could not to be allocated to a specific company or individual. Participants were reassured that their answers would be anonymised and released only as an amalgamated industry view. Neither would any answer be published with name/company name in any of the project's documents. This clause was essential to ensure that the senior industry representatives would be able to respond freely and openly.

The interviews conducted with senior representatives, took on average 60 minutes (sometimes longer). This allowed for additional feedback on the questions allowing the interviewer to gain a more in-depth understanding of the interviewee's positions, opinions and needs.

2.4 Consolidation of all input

The input received via the ELE Online Survey, the Industry-specific Questionnaire and the Interviews produced feedback from 46 organisations (after deduction of the research organisations). It was consolidated in the **findings & analysis** below (section 4). Furthermore, LT-Innovate contributed a set of **recommendations** incorporating the current industry consensus on 'Languages, Digital Sovereignty and Artificial Intelligence' (section 5). This document was compiled over the last year. It was also released publicly on 25 January 2022 in the form of an Open Letter to European Decision Makers.

3 Respondents' Profiles

An overview of the profiles of the respondents to the two questionnaires and participants in the interviews is provided in the figures below.

3.1 Respondents – Organisation Types

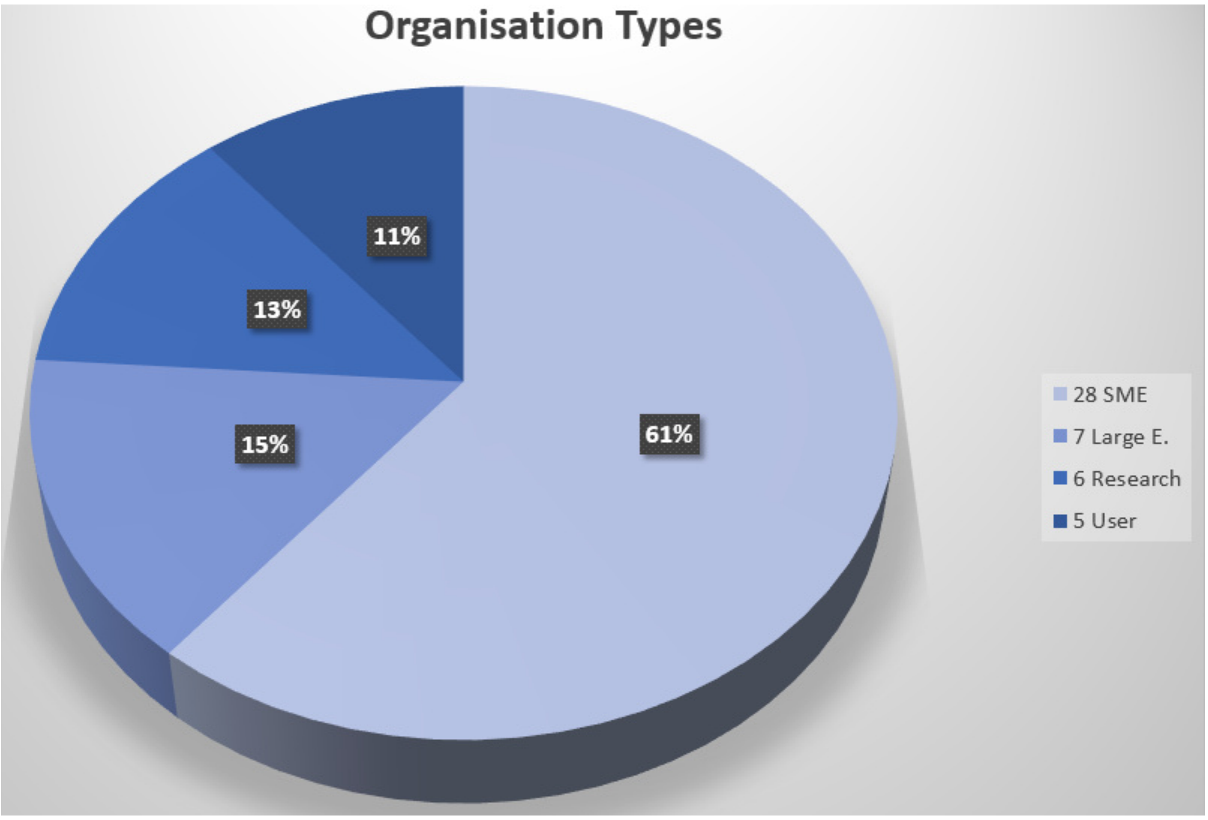

Figure 1: Organisation Types

N.B. The ELE online survey also included 6 responses from universities / research institutes. While their responses are interesting and highly relevant for the ELE project, their feedback was, however, not taken into consideration for D2.4, as this deliverable intends to elicit the views of the LT Industry.

3.2 Sectoral representations & LT areas

The overall total of 46 respondents to the online questionnaire, the industry-specific questionnaire and the interviews declared to cover within their market segments nearly all vertical sectors, from media, finance, insurance to food, agriculture, humanities & health, culture & art, green economy to logistics, construction, legal etc. However, it was not possible to obtain specific information about the size of business conducted within these market segments.

The respondents cover the three domains of LT/LI that LT-Innovate uses as top-level classification, i.e., Translation Technologies, Interaction Technologies, and Analytics Technologies. A majority declared themselves as providers of Translation Technologies (19), while Interaction Technologies and Analytics Technologies were represented respectively by 7 and 9 respondents. In addition, there were 5 large corporate users of language technology that also develop their own internal systems and applications. Figure 2 provides a graphical representation of the type of respondents.

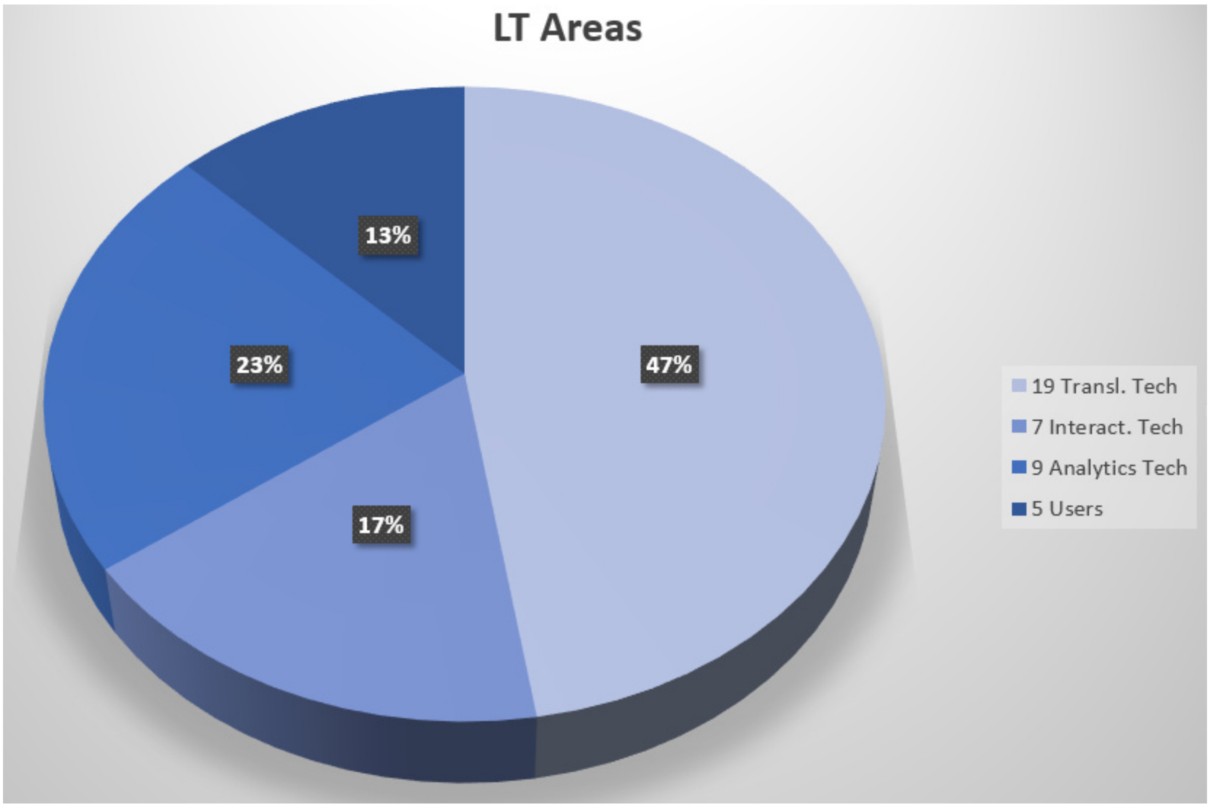

Figure 2: LT Areas

4. Findings & Analysis

4.1 The concept of "language equality" from a business perspective

Before delving into the findings of the survey conducted, it is worth noting that the term 'language equality' was questioned by respondents, particularly during interviews. It appears that 'language equality' as a concept is, a priori, not understood in a business context. Languages just like geographical markets are obviously not equal; at the very least, some are more equal than others. Language markets are of various sizes and therefore of various interest/value from a commercial point of view. It is therefore important to note that, while "language equality" may be a stated goal by the European Parliament, it is a concept that is less relevant to the industry as such. At the same time the industry does consider languages as well as country specific regulation as factors that complicate cross-border business and in some cases also is a consideration when considering what new markets to enter.

4.2 Findings from the questionnaires & interviews

In summary, the findings of our consultation processes were interesting and, while in most cases not surprising, some noticeably strong opinions were voiced. With regard to the EU programmes and initiatives, the responses were particularly sobering. Responses show that that a majority of executives are largely unaware of EU initiatives (including outputs they are sometimes benefiting from without understanding that these have originated from EU initiatives). While the more detailed

analysis of feedback reveals a more balanced picture, it is clear that industry often perceives that the benefit of EU initiatives is largely limited and has to a large extent developed a “blind spot” in that EU initiatives are not tracked or considered when considering new product or service developments. The perception that many executives seem to have is that these programs are largely academic in nature and have limited practical application. This finding is particularly concerning given that this feedback has been provided by language technology executives, not simply industry in general.
To that extent, it is apparent that a move towards creating more awareness of the initiatives and their practical benefits and applications as well as a more active involvement and inclusion of a broader group of industry stakeholders in future EU initiatives would appear to be essential to stimulate the use by commercial users of the assets developed under various EU programmes as well as empowering the industry as a whole.

Some notable points raised by participants related to the following:

4.2.1 The EU as catalyst for collaboration

Notwithstanding the above critics and/or disappointment with EU initiatives, companies see some clear roles that the EU can and even should play. Respondents generally agree that the EU should support mechanisms to coordinate collaboration between vendors, integrators, customers etc. to strengthen the (European) industry and its competitiveness. In their view, the most important benefits of EU support would be the fostering of cross-border and cross-organisations collaboration:

- “Support international (Europe-wide) business networks (not only academic ones)”
- “Funding initiatives to benefit a wider commercial ecosystem across Europe”
- “Funding should cover more cross-sector collaboration and global market development”
- “The EU should mediate discussions and set up facilitating frameworks”

4.2.2 Help better define market needs to raise competitiveness

Respondents clearly see and are exposed to the competition of the “GAFAM”[3] and other big tech companies from outside Europe homing in on them. In addition, companies frequently have only very limited information on markets outside of their (national or local) home market. It has often been stated that the EU should increasingly support market research, particularly for the benefit of SMEs to enable cross border growth.

- One respondent brings it to the point: “We are operating in a ‘Red Ocean’: too many players to stand out from. Focus on specific market segments may be an answer but access is not easily available for all”.
- “Better knowledge of markets / verticals – why not fund market research?”
- “Solutions need to be customised for specific markets and specific business processes within these markets (generic pilots are of little value)”
- “Insufficient public procurement including pre-commercial public Procurement

4.2.3 Support skills development in the LT/LI domain

[3] Google, Apple, Facebook, Amazon, Microsoft

The online questionnaire asked whether the lack of talent or "brain drain" represents a challenge for the LT/LI sector. Interestingly, half of the respondents answered in the affirmative, i.e., that it is a challenge, while the other half did not see it as a challenge at all. The answers from the industry-specific questionnaire/interviews were more specific and indicated the lack of skilled personnel as a challenge, in particular:

- "lack of technically qualified staff"
- "lack of linguists for some languages, e.g. agglutinative languages like Turkish, or ambiguous ones like Arabic"
- "not able to employ in EU w/o local entity (no Internal Market!)"
- "publicly funded projects attract staff away from industry"

Interestingly respondents who stated they had limited skill issues benefitted either from their geographic location in that there were few other employment opportunities for staff unless they were willing to relocate, or they actively developed skill in-house.

4.2.4 Make available more language resources

Respondents agreed that developing language technology solutions for some markets may not be commercially viable. Only three respondents disagreed, of which two declared to have no problem with language data. The majority however agrees that more open language resources are needed and that that public sector support is required to guarantee a level playing field for all languages, including those deemed to be of lesser commercial viability. Some specific requirements stated were:

- "More language resources for scarce(r) EU languages (e.g. HU, FI, HR)"
- "More language resources for Asian languages"
- "EU should make available high level language resources & 'WordNets' (like Princeton for EN)"
- "Availability of language data sets (via ELG) positive, but should be increased (e.g. Euramis)"
- "Data availability is only useful if its use for commercial projects is possible; if useful, companies are willing to pay for such services"
- "Solution to making more data available exists: anonymization"

4.2.5 Support research on additional LT/LI breakthroughs

It should not come as a surprise that respondents with a research/academia background stress that more basic research is needed. However, companies see also the need for further – applied – research to better serve the markets in the following areas. The following 'breakthroughs' appear on the wish list of respondents:

- "LT with better language understanding and reasoning"
- "Advances in text summarisation & classification"
- "Language understanding: put together neural, statistical, rule-based approaches"
- "LOD, unsupervised automatic knowledge graphs"
- "NLP self-training from continuous user feed-back"
- "Overcoming CAT deficiencies"

- “Better human-machine interaction”

4.2.6 Address regulatory challenges

When asked what the biggest challenges in growing their business are, almost all respondents mentioned legal or regulatory challenges.

Highlighted **regulatory issue**s that should be addressed urgently are:

- GDPR: while some businesses have built strategies on compliance with GDPR, others see it as seriously “hampering data analytics and encouraging its offshoring”.
- GDPR is negatively affecting the ability machine learning applications due to restrictions. Defined and acceptable anonymization standards and supporting technology would be an example of a domain where EC initiatives could help businesses.
- Copyright: “there is (still!) no equivalent to US fair use”

4.3 Analysis

Overall, four main themes can be extracted from all input received:

4.3.1 Theme 1: The need for EC initiatives as perceived by the LT/LI industry

Even though the LT/LI industry at large has a commercial and, in some respect, short(er)-term focus, all respondents have clearly indicated the need for language technology and services in a multilingual continent such as Europe. Without exception, respondents have indicated that they believe the EU can and should play a pivotal role in providing a base infrastructure and should continue to invest in initiatives that foster the delivery of basic language technologies, services, data assets, human resources, as well as innovation pipelines, i.e. innovation and pilot projects aimed at supporting the validation of technologies in a commercial context. Ideally these initiatives would be either generic in nature or specific to industry challenges as well as focused around solving specific business rather than technology challenges by possibly combining a range of innovative offerings. However, respondents were equally aligned in their view that predictable and reliable government (particularly at EU level) requires a clearly formulated and communicated vision and strategy to foster a successful private-public partnership as well as supporting business strategies and associated investment decisions. The vison and strategy should clearly indicate the stakeholder communities that have been identified as well as how, when and in what manner they should be involved and/or supported by EU initiatives.

Importantly, a key requirement identified by LT/LI industry representatives was that any deliverables, be they technology assets, products or services - such as e.g. the European Language Grid (ELG), need to meet business level requirements specifically for services in order for them to be adopted by businesses. This means that they need to provide clear Service Levels as part of a Service Level Agreement (SLA), guarantee data privacy requirements under the EU GDPR (which is clearly preferred over US cloud service providers services) and provide open integration options. For products, clear license structures, standards compliance, maintenance of the systems over time are just a few examples of prerequisites to envisage building offerings on top and around deliverables from EU initiatives.

Finally, a key statement supported by a vast majority of the respondents was a clear willingness to pay for value-generating services, meaning businesses are not expecting these services to be free. One specific example provided by one of the respondents was the European Patent Office (EPO) which provides a range of services to commercial players in the Intellectual Property (IP) ecosystem, allowing them to innovate on top of its services. These kinds of models are favoured by the industry since they provide predictability and added value services that can be embedded into commercial parties' value chains. In the context of language technology and related services such as the ELG as an example, one could envision that access to the translation infrastructure at volume would be chargeable yet would come with clear Service Levels and Service Guarantees.

4.3.2 Theme 2: Dissemination & Perception

Very few of the respondents stated they were ‘fully’ aware of EU initiatives, meaning that part of the target audience (businesses in language technology) of EC initiatives are not aware of the opportunities presented to them. A small minority (from start-up to established corporate) that has been involved in EU projects, claimed awareness. However, when presented with examples of EU initiatives, during interviews, respondents at large indicated clear interest. To what extent that interest will result in activity remains to be seen, but as indicated in the previous section, the industry at large is seeking readily available products and services that can be easily and effectively embedded into their respective business processes. In the broader context it is also important to remind ourselves that for the vast majority of users of language technology and localization (i.e., the customers of language technology providers) is not a core business activity but a requirement to allow access to a European and/or global marketplace. Consequently, most organizations also have limited LT/LI knowledge or expertise in-house. They are user’s dependent on the capability and services of their suppliers, but not experts. At the same time, they are potential consumers and beneficiaries of technology and services developed under the EC programs by proxy.

Furthermore, all respondents indicated that they felt that active stakeholder involvement (active outreach) and broader dissemination was key to spreading the message, raising awareness, and broadening engagement and that clarification of strategy and vision was essential to building understanding and confidence.

Those respondents who had some exposure to EU initiatives in the past felt that the strong “academic” focus and spirit and the fact that (by and large) the “same” players seem to emerge in projects regularly, led to the impression and reputation of (rightly or wrongly) subsidised projects with no commercial bearing rather than initiatives that serve real-world purposes. This negative perception as well as the difficulty of finding and applying deliverables effectively was frequently indicated as a reason not to engage. Most commercial organizations view EU initiatives as a parallel universe that provides no benefits to them - which is clearly a missed opportunity.

The usability of results was underlined in one answer: “EU and national organisations should promote LT by showing that they are really using it themselves”.

4.3.3 Theme 3: Vision, Strategy, and Direction

Most respondents indicated that they did not understand the vision and strategy of the EC with regards to language related initiatives. Industry representatives underline that the mere "administration" of projects without a clear vision and direction affected the understanding and usefulness of project deliverables and, ultimately, discouraged the industry's willingness to engage.

More specifically, the key concerns raised by respondents in this regard, were the following:

1. Clarity on the vision, strategy, and direction / roadmap. Considering the clear interest of the industry in applying innovation and services delivered under EC initiatives, business leaders lack understanding and confidence to engage and invest. The perceived lack of a well communicated and understood strategy and vision, outreach and engagement and business-level available services and products affects the willingness to engage and the confidence substantially. In an increasingly global and fast-moving ecosystem, business leaders select their partners carefully and based on a match in terms of both vision as well as service level, but also longer-term confidence in the relationship. To engage businesses more effectively and achieve the goal of European Language Equality, respondents felt that there was still a way to go to provide them with the confidence to engage.

   Where services, tooling, products, and more are provided, increased clarity on scope is required as well as clarity on how support for these deliverables is designed will impact businesses willingness to engage. For example, key software components delivered under EC initiatives will only be applied by organizations outside of the language technology space if there is a "community" that maintains and supports these components. Most businesses do not have the ability to maintain such tools by themselves. Most businesses (within and beyond the language technology industry) want to be able to quickly and effectively apply products and services as part of their value chain, and do not possess the skills or willingness to perform complex integration and operation themselves. The increasing levels of outsourcing and Cloud usage by European businesses are a key indication thereof. Where services are provided, they need not be free, but should be priced reasonably, should come with clear SLA's and should effectively deal with issues such as data privacy so that they can be applied by businesses quickly and effectively. Tooling and products are useless if they are not maintained and supported. In the context of EC initiatives targeted at the language technology industry and its customers who are largely operating on a European or even global scale, this means that productization of core services in support of EU businesses is essential to ensuring adoption and broader application.

2. Openness and integration are essential. Integration with existing key platforms and solutions used within the industry as well as Cloud integration are key to increased technology take-up. In the case of Machine Translation Services for example integrations with commercial TMS or "gateway" systems would assist in driving substantial adoption. Overall, industry leaders clearly stated the need for an ecosystem and increased productization and platformization, stating also that an alternative to the dominant US providers would be more than welcome. A truly European yet competitive alternative.

3. A number of respondents proposed that the EU clarify how its approach compares to US and Chinese initiatives and ensures a business-driven collaborative approach. The general

perception is that the US and even the Chinese focus drives more (business) opportunity and that these markets are generally more uniform and attractive than the fragmented EU market.

4. A key theme is the request to simplify and ease access to deliverables from EC initiatives; ideally a single portal should provide access to all relevant information, tools, and services. Business grade SLA's, legal terms and an ecosystem approach should complement the initiatives and help ensure easier and broader application.

5. With the industry perception being that EC initiatives are largely benefitting academia and "insiders", the suggestion is to ensure clearer delineation between the necessary academic programs and more business focused programs, yet ensure that, where possible, they are extensions of each other. Suggestions include set-up outreach initiatives at a national and an EU level to foster adoption and collaboration as well as general understanding within the business community.

4.3.4 Theme 4: Standards

Standards were a key theme emerging from the interviews, ranging from the need to clear data formats, to open API standards, standards for data anonymization and a range of other guidelines and standards, generic and vertical market specific (e.g. medical), that support an ecosystem-based approach as well as accelerated time to market on offerings, a key requirement for the industry. The respondents indicated a clear need for and support for standardization as well as a commitment to developing open standards.

5. Recommendations

In terms of what the European Institutions should do concretely to reach the goal of 'language equality' by 2030, LT-Innovate has elaborated an LT/LI industry consensus around the following recommendations:

1. Complete the Internal Market

The European Internal Market is far from being completed. The resulting market 'fragmentation' is what bears most heavily on the competitiveness of European companies compared to their rivals based in large homogeneous markets. Language barriers are one of the main factors still to be overcome to reduce market fragmentation - without penalising individual languages. As a matter of fact, if 'language equality' is properly implemented, it should lead to language 'neutrality', i.e. a state of affairs that will allow communication independently of the languages used by the various parties involved. The technologies to achieve this 'neutrality' are now largely available. Their deployment appears to be solely a matter of objective-driven and consequent investment, backed-up by political will (particularly in Europe).

- **The European Institutions should relaunch the completion of the Internal Market and encourage the deployment of technologies to remove all linguistic barriers to the latter.**

2. Mandate that all cross-border services are made 'native' for everyone

An obvious step that would greatly contribute to the Internal Market and language ‘neutrality’ becoming a reality, would be to require that all cross-border services (i.e. services that are not strictly of relevance to one particular language community) should, by definition, be multilingual – in particular, public services. Language neutrality should thus become a corollary of the rights of free movement across the EU. In other words, wherever there is a right to free movement, the latter should not be restricted because of language barriers. In concrete terms, the European Commission should draw up an exhaustive catalogue of cross-border services that (by a certain date e.g. 2030) will be mandatorily required to be made ‘native’ for everyone because they are corollaries of the right to free movement and/or required for the proper functioning of the Internal Market.

- **The European Institutions should mandate that cross-border services become fully multilingual across the EU.**

3. Remain in control of the digital life of our languages

Pervasive digitisation has so far played into the hands of the English language and has established the latter not only as main programming language, but as source language for a disproportionate amount of the services and content available via the internet and, overall, has established it as *lingua franca* for international communication. It has been recently recognised that this poses a challenge for all non-native English-speakers and, ultimately, for the digital sovereignty of all non-English-speaking countries. While investment in semiconductors, cloud computing, cryptography and ‘data’ are all very important, digital sovereignty requires, first and foremost, that Europe remains in control of the ‘digital life’ of its languages, as 50% of all data is textual (i.e. expressed in human languages).

- **The European Institutions (in collaboration with Member States governments and Regional authorities) should invest in the digital life of all European languages, making sure that they have equal access to digitisation and making the internet and/or the ‘cloud’ a language-neutral level playing field. More broadly, it should make the case that languages are not only cultural assets, but also economic assets.**

4. Understand that human-centred artificial intelligence (AI) must be multilingual, of necessity

The core of artificial intelligence is related to the processing of human languages as illustrated in the graphic below:

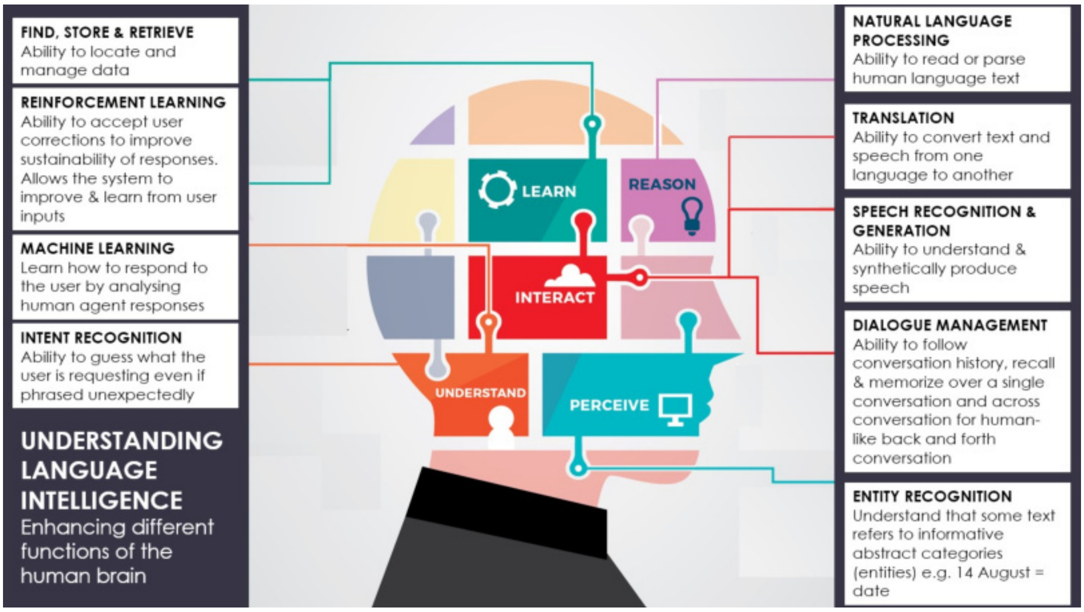

Figure 3: Understanding Language Intelligence

In this context, language technologies and language intelligence (the technologies listed in the boxes above) are key components for a human-centred AI strategy. However, this is hardly reflected in current EU strategy documents.

- **The European Institutions should position language technology and language intelligence at the very heart of their 'data strategy'[4] (from which it is currently absent).**
- **The European Institutions should position multilingual AI as the way forward to (re)position Europe on the global AI map (in particular, by training big Natural Language Processing AI models, similar to GPT-3, on all EU languages).**

5. Strengthen the LT/LI corporate ecosystem across Europe

European companies have been pioneers of language technologies and language intelligence for many decades. This was recognised by the European Commission in the 1990s when programmes geared at reinforcing the SME ecosystem in this domain were launched. However, since 2014, these programmes have been quasi-interrupted, resulting in a weakening of the corporate ecosystem and the slowing down of its internationalisation. In the meantime, large US companies (Google, Amazon, Microsoft, Apple, IBM, etc.) have invested heavily in this field, often acquiring European companies or commercializing open technology developed in the EU under EU programmes. In recent years, Asian companies have also recognised the importance of language technologies and language intelligence and are now seeking to acquire assets.

- **The European Commission should (hopefully via the Digital Europe programme) aim at remobilising and galvanising the European LT/LI corporate ecosystem around an industrial strategy taking into account the manifold statements, studies and roadmaps[5] elaborated by LT-Innovate over the last decade. Said strategy should be elaborated in close collaboration with the core group of European scale-up companies (i.e. that have already expanded beyond their**

[4] https://digital-strategy.ec.europa.eu/en/policies/strategy-data
[5] http://www.lt-innovate.org/public-positions

national borders) and should aim at supporting innovative companies who empower Europeans to generate more multilingual content.

- **The European Commission should provide support for the application of EXISTING TECHNOLOGIES to use cases of European and global dimension, and speed up their deployment, in particular through 'productisation' and 'vertical platformisation' (see below). In doing so, the European Commission should fund (via its Digital Europe programme) only projects with clear paths towards industrialisation and European / global market penetration. Such projects should be evaluated – early in their life – as to the viability and sustainability of their business plans (which may stay confidential but MUST be put to the test). Projects that do not give satisfaction with regard to the latter should see their funding interrupted.**

- **The European Institutions should ensure that European companies acquired by entities of non-European origin refund all EU financial support received by the acquired company within the 5 years prior to its acquisition.**

6. Build, maintain and operate a plug-and-play baseline language infrastructure

Since 2014, key industry players and senior researchers have called for a plug-and-play Language Cloud[6] and/or baseline language infrastructure[7] to be implemented at European level. The European Language Grid (ELG)[8] project (building on previous EU-funded initiatives such as META-NET, ELRC, etc.) goes some way into the direction of addressing this. However, ELG is both too ambitious and not ambitious enough. Too ambitious in substituting itself to (and/or undermining) existing commercial initiatives. Not ambitious enough in supplying an operational baseline infrastructure that would, instead, support commercial players as well as underpinning the goal of creating the fundament for a language-neutral Internal Market.

- **The European Commission should re-orient the European Language Grid to providing resources for one-stop-shop language interoperability (at the exclusion of any other goals that compete with existing commercial offers) and become truly operational and dependable in a commercial context (i.e become a plug-and-play resource supporting commercial solutions). Pervasive Named Entity Recognition and data Anonymization are examples of base line services that would allow the ELG to address one of the biggest AI concerns of citizens, data privacy, and turning the GDPR constraints into a competitive advantage. Moreover, the ELG should enable crucial cross-border projects (such as the 'Data Spaces' foreseen in the Digital Europe programme) by developing an underlying information infrastructure of multilingual knowledge graphs for applications requiring legally valid terminology and supporting the semantic layer of the European Interoperability Framework, becoming thus a key resource for Europe's software industry. Finally, the ELG could play a crucial role in driving standardization efforts (particularly of interfaces).**

- **The European Institutions (in collaboration with Member States and Regions) should intensify the language resources collection effort initiated in the European projects such as ELRC[9] and**

[6] http://www.lt-innovate.org/lt-observe/document/call-action-industry-driven-european-language-cloud-more-necessary-ever
[7] http://www.lt-innovate.org/lt-observe/document/citia-baselayer-multilingual-speech-technology
[8] https://www.european-language-grid.eu/
[9] European Language Resource Coordination: https://www.lr-coordination.eu

MLIA[10] providing actionable (i.e. commercially usable) multilingual language resources for all EU languages (including lesser-used and underserved languages, not forgetting languages of indigenous people) as well as the languages of the EU's main trading partners.

- **The European Commission should ensure that its linguistic assets (e.g. Euramis) and other multilingual knowledge, dormant in databases such as EuroVoc and IATE, are formalised and made API-accessible in multilingual data spaces; it should furthermore encourage (and where possible mandate) other public institutions to do likewise with their linguistic assets. All of these assets should be made accessible for all players (including commercial ones).**

- **The European Institutions should encourage their own services as well as Member States to buy/use European products and services in the language processing domain, this would include replacing non-European products and services where possible and adding (largely) automated multi-lingual services to all digital aspects of the EU's and Member States' communication (e.g. web-pages).**

7. Encourage the emergence of multilingual vertical AI 'platforms' at European level

LT/LI technologies are true enabling technologies. They are ubiquitous but need to be customised for a myriad of application domains and use cases.

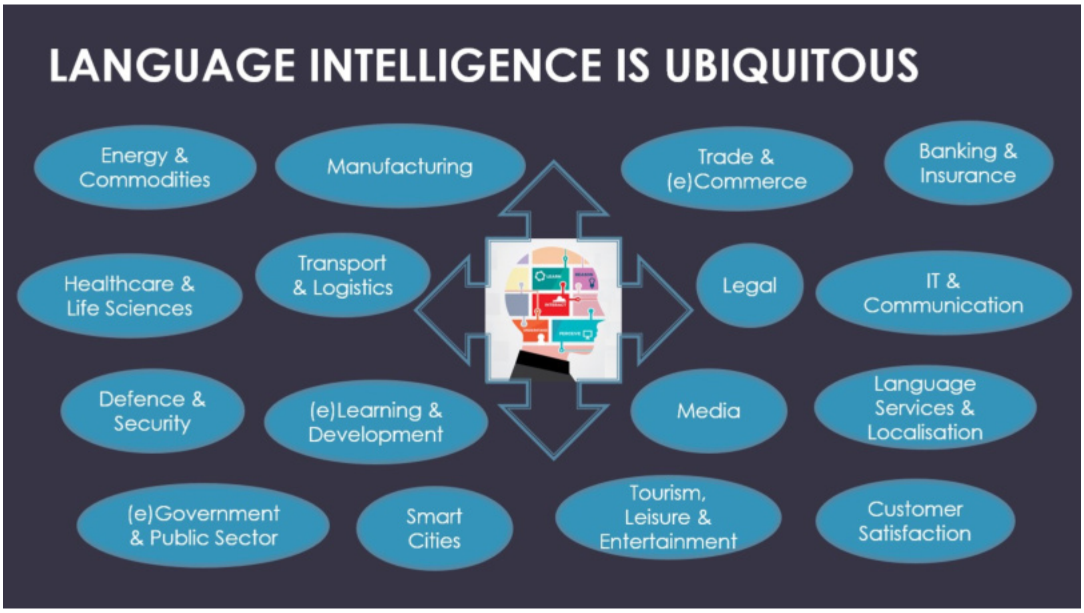

Figure 4: Language Intelligence is ubiquitous

LT/LI companies have started homing in on the needs of various domains (such as those listed in the graph above) in the last 3-4 years (witness the domain-specific pages on their websites[11]). This trend is a sign of maturity, but it also pitches the European LT/LI industry against large international competitors (global IT companies, system integrators, IT project management consultancies, etc.) who traditionally support large accounts involved in process re-engineering and change management. In parallel, many language service providers (LSPs) have embarked on reinventing themselves from translation bureaux to specialised value-added content management service providers. The latter trend represents an important window of opportunity for smart, technology-savvy players. The

[10] COVID-19 Multilingual Information Access: http://eval.covid19-mlia.eu/
[11] See: http://www.lt-innovate.org/page/domain-specific-solutions

availability of multilingual AI (the components listed in fig. 1) will be crucial, although not sufficient, to allow the European LT/LI/LSP ecosystem to avail itself of this window of opportunity in order to leverage their technical expertise into a new generation of content management products and services of global reach.

- **The European Commission should facilitate the consolidation of the LT/LI/LSP corporate ecosystem by encouraging 'productisation' and 'vertical platformisation'. By 'productisation', we mean the transformation of technologies and components into solutions clearly aimed at specific use-cases in specific markets. By 'vertical platformisation', we mean the assembly of such products and their combination with value-added services addressing domain-specific processes. This could be achieved via the AI Data Spaces envisaged in the Digital Europe programme, provided that these make the link with an explicit industrial LT/LI strategy (as mentioned above).**

- **The European Commission should fund a number of pilot projects (in various application domains) to pioneer this approach and deliver the proof of concept, particularly in terms of helping commercial productisation and platformisation projects to 'europeanise' and/or 'globalise'. The budgets specifically earmarked for this purpose in the Digital Europe programme appear as dramatically insufficient in view of meeting the challenges and opportunities outlined above.**

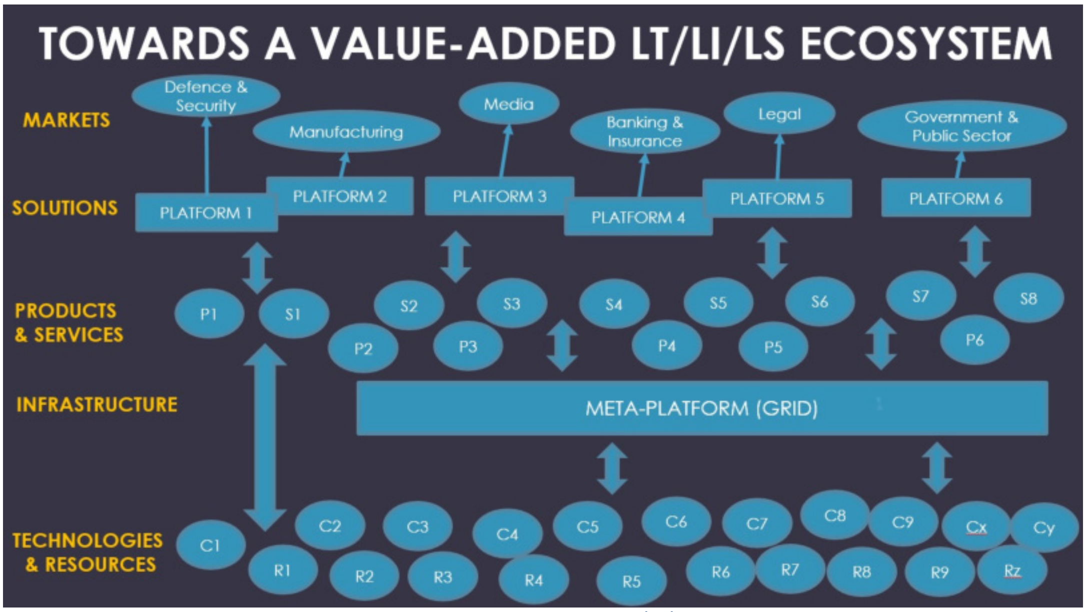

Figure 5: Towards a Value-added LT/LI/LS ecosystem

A strategy for the LT/LI/LS ecosystem should be based on an in depth analysis of the state of the art, capabilities and needs of all stakeholders in the ecosystem (technical components and resources providers, meta-platform operators, products and services providers, commercial platform operators, domain-specific application markets/users) and their interaction (as shown in Fig. 6). Furthermore, it should define clear paths to accelerate this interaction in view of delivering innovative solutions into domain-specific markets. The latter could be subject to trials via a number of pilot projects.

ANNEXES

Annex 1: Industry-specific Questionnaire

LANGUAGE INTELLIGENCE MARKET SURVEY QUESTIONNAIRE

Q1: COMPANY DETAILS
Please review your corporate profile on the LT-Innovate website [http://www.lt-innovate.org/directory/organisation/readspeaker] and let us know if any changes / updates need to be made.
In addition, please provide the following (to the extent you feel comfortable disclosing the information - see confidentiality clause *in fine*):
- Size of your company:
- Turnover (NLP/LI technologies and related services): Click here to enter text. Year: Click here to enter text.
- Corporate headcount (in FTE): Click here to enter text.
- 5-10 Keywords which best characterise your business: Click here to enter text.
- Web address of your corporate YouTube channel if available: Click here to enter text.

Q2: MARKETS
In which specific domains (markets) do you see the greatest need for multilingual solutions from your point of view? Click here to enter text.
In which markets / industries / domains / verticals are you mostly active today? Click here to enter text.
Into which countries did you sell in the last 3 years? Click here to enter text.
Will NLP/LI be subsumed in AI (or something else) or will it continue to have its own life? Why?

Q3: CHALLENGES
What are the biggest challenges you encounter in your growing your business?

☐ Laws/regulations – explanation: Click here to enter text.

☐ Bureaucratic hurdles/delays – explanation: Click here to enter text.

☐ Costs of doing business – explanation: Click here to enter text.

☐ Lack of financing – explanation: Click here to enter text.

☐ Potential clients who are badly informed about available solutions – explanation: Click here to enter text.

☐ Potential clients who are risk averse and not prepared to invest – explanation: Click here to enter text.

☐ Access to potential clients blocked off (e.g. by LSPs, integrators, large accounts...) – explanation: Click here to enter text.

☐ (Unfair) competition from abroad – explanation: Click here to enter text.

☐ Technical challenges, e.g. scarcity of data, language resources – explanation: Click here to enter text.

- Which languages remain challenging for you to integrate in your offerings? Click here to enter text. Why? Click here to enter text.
- Which research breakthroughs would you like to see happen in the next 5-10 years? Click here to enter text.

☐ Other – explanation: Click here to enter text.

Q5: COLLABORATIVE INITIATIVES
Are you involved in any collaborative initiatives around NLP/LI, AI, Cloud (such as Gaia-X)? Click here to enter text.

Q6: EUROPEAN UNION POLICIES & PROGRAMMES
To what extent are you aware of EU language related projects and resources? Click here to enter text.
To what extent do you think EU initiatives are sustainable, i.e able to deliver structural benefits to businesses over the coming years? Click here to enter text.
What do you expect from the EU in the next 5 years?

☐ More funding programmes for companies? – explanation: Click here to enter text.

- If yes, what should be improved in the existing funding programmes? Click here to enter text.
- What types of projects should the EU support? Click here to enter text.

☐ Focus on 'finishing' the Internal Market - what is missing from your point of view? – explanation: Click here to enter text.

☐ Financial guarantees for small companies that want to levy investment – explanation: Click here to enter text.

☐ A baseline language infrastructure in all EU languages – explanation: Click here to enter text.

- Do you know the ELG project [https://www.european-language-grid.eu/]? Are you participating in it? Do you consider it useful? Why? How could it become (more) useful to you? Click here to enter text.

☐ Make available multilingual data assets held by European institutions (Euramis, EuroVoc, IATE, etc.), if necessary by anonymising them – explanation: Click here to enter text.

☐ Require that all public services become available in all EU languages – explanation: Click here to enter text.

☐ Require that public institutions buy mainly/only from European suppliers – explanation: Click here to enter text.

☐ Other – explanation: Click here to enter text.

Annex 3: Target list for industry-specific questionnaire

24Translate
Acapela
ACT Translations
Almawave
Artificial Solutions
AT Language Solutions
Basis Technology
Cedat85
ChapsVision/BertinIT
Coreon
Cortical.io
e2f
ESTeam
Exfluency

Expert.ai
EyeOnText
Hensoldt Analytics
Interverbum
IQVIA
KDictionaries / Lexicala
LexisNexis
Lingsoft
MeaningCloud
Medtronic
MemoQ
Memsource
Mondragon Lingua
NEN
Ontotext
Pangeanic
Phonexia
Priberam
Proxem
Replaywell
Reverso
RWS
Semantic Web Company
Semantix
Summa Linguae Technologies
Syllabs
SYSTRAN
Taiger
Textgain
Toppan Digital Language
Unbabel
ViaDialog
Vocavio
VoiceInteraction
WCS Group
XTM International
XTRF