# Language equality in the digital age
## Towards a Human Language Project

The EU is a unique endeavour involving more than 500 million citizens sharing about 80 different languages. While multilingualism is one of the biggest assets of Europe, it is also one of the most substantial challenges for the creation of a truly integrated EU. In the digital era, the invisible separating power of language barriers is a major challenge hindering European citizens and businesses from drawing upon the possibilities of a truly integrated Europe. Language barriers particularly affect the less educated and older population, as well as speakers of smaller languages, thus creating a noticeable language divide. It has a profound effect on cross-border public services, fostering a common European identity, workers mobility and cross-border e-commerce and trade.

The emergence of new technological approaches, based on increase computational power and access to sizeable amounts of data, such as deep-learning neural networks, are making human language technologies (HLT) a real solution to overcoming language barriers. However, several challenges, such as market fragmentation and unsubstantial and uncoordinated funding strategies, are hindering the European HLT community, including research and industry. The consequence is that more than 20 European languages are in danger of digital extinction and that the enormous potential that HLT can have for Europe remains largely untapped.

Moreover, language technologies are not properly represented in the agenda of European policy-makers, although they are likely to be crucial for the construction of a fair and truly integrated European Union.

## The Human Language Project

Technology is the only possible way of addressing Europe and its multilingual setup. However, multilingualism in Europe is a complex topic involving many stakeholders in different countries with intertwined interests. Therefore, no single policy will tackle the aforementioned problems. On the contrary, to truly seize the opportunities of a multilingual Europe, the authors of this paper believe that a joint and coordinated action is required at the European, national and regional levels, involving a variety of stakeholders. Therefore, we propose launching a multidisciplinary European Human Language Project (HLP) including a set of policies involving stakeholders from the public sector, civil society, research institutions, and the industry, at the European, national and regional levels, in a coordinated and joint effort to move Europe into the pole position in this field. The ambitious goal of the HLP is that Europe can develop and fully benefit from the next generation of language technologies.

Research in Europe should focus on creating the new paradigm of HLT, combining a fresh look at linguistics with the power of current artificial intelligence (AI) methods that are based on vast knowledge bases, themselves created with the help of sophisticated language technologies. Not only European institutions but also national and regional governments should be involved in creating resources for their own languages. Talent scarcity and drain brain should be transformed into talent creation and brain gain. Resources for the industry should be provided in a seamless, open and effective way through existing European platforms. The public sector should provide their contents and services for all European languages while promoting the growth of the HLT market through public procurement of innovative technology. Mechanisms to facilitate the scaling-up of European innovative HLP companies should be promoted. Eventually, policies supporting firms across Europe to sell cross-border by providing their contents, products and services in the different European languages, should be enacted to create a fully integrated digital single market (DSM).

## Policies within the HLP

To achieve these goals, the authors have proposed and assessed different policy options that are structured into five groups: institutional policies, research policies, industry policies, market policies and public services

policies. Institutional policies involve initiatives aimed at developing the current institutional framework by drawing upon emerging technology trends in order to better fit the challenges of a multilingual Europe, while properly assessing the results. Research policies focus on encouraging Europe to develop the next generation of language technologies. Research policies also aim at integrating research and industry, providing Europe with the tools to share resources so as to effectively compete with other markets, while contributing to the equality of all European citizens in their everyday digital experience, regardless of their language. Industry policies foster the creation and growth of competitive European firms while increasing the availability of highly qualified workers. Market policies seek to improve the HLT sector in Europe by raising awareness among European stakeholders of the relevance of these technologies to further increase the demand of services. There is a specific policy targeting small web merchants so they can benefit from accessing a much bigger market by translating their web shops through HLT. Public service policies intend to promote multilingual public services and contents in European, national, regional and local institutions, while contributing to increasing the innovative HLT sector by using public procurement tools.

## Institutional policies

1. **Reinforce the role of HLT within the institutional framework of multilingualism-related bodies**

   To successfully achieve the goals of the HLP a joint effort is required between multiple stakeholders (public, private and civil society) across Europe at the European, national and regional levels. Therefore, a strong coordination is required. We suggest two actions: (i) reorganise HLT activities within the EU, by having a unit at the highest level possible to take care of multilingualism and language technologies and (ii) attach the 'multilingualism and language technology' topic to a Commissioner.

2. **Create tools to properly evaluate HLT policies**

   In defining new policies it is very important to learn by experience. To assess the policies, it is necessary to create open and homogeneous European longitudinal datasets including extensive information about language barriers, their economic and social impact, and the effect of HLT on overcoming these barriers. This policy is particularly relevant in Europe, which can act as a natural policy lab, where different policies and approaches can be tested at the national and regional levels, providing a valuable experience to help other countries to define their policies.

## Research policies

3. **Refocus and strengthen research in the Human Language Project**

   The first and probably most crucial research policy is to set up, under the umbrella of the HLP, a large-scale, long-term research and development and innovation (R&D&I) funding programme, in which basic research, applied research and development, as well as innovation and commercialisation, work closely together in order to develop technologies for 'deep natural language understanding' by the year 2030. Fields needed to contribute to this goal are computational linguistics, linguistics, artificial intelligence, language technology, computer science and cognitive science.

   The key goal of the research activities of the HLP is deep natural language understanding (including language generation). Current LT approaches are still rather shallow; many research and technology breakthroughs are needed in order to move any closer to the goal of full natural language understanding. Nevertheless, recent breakthroughs in artificial intelligence, and a fresh look at recent results in linguistics, can bring about the needed technology shift for the next generation of LT. With regard to research, key ingredients of the HLP must be artificial intelligence, language technology, machine learning and knowledge technology. Technological goals should be broad coverage, high quality and high precision. It is worth mentioning that the 'Multilingual Value Programme', as suggested by the Strategic Research and Innovation Agenda 2016 (SRIA), can be conceptualised as part of the Human Language Project. In terms of funding sources, a mix can be envisaged between Horizon 2020 (Work Programme 2018-2020) and the next Framework Programme (2021+), as well as national and regional funding sources.

4. **Promote the European LT Platform of data and services**

   It is necessary to draw upon existing infrastructures and platforms to promote an open, cloud-based architecture, enabling the sharing and further development of LT-related resources developed at the

different European levels. Because of the diversity of the LT applications and services, it may be more flexible to actually talk of connected existing platforms instead of creating a single platform, at least in a first phase, but it is important that they are conceptualised as a single European-based infrastructure for data and services.

Existing resource-exchange infrastructures, such as META-SHARE and CLARIN, have shown the necessity of storing and sharing of linguistic data and resources in a single place. These existing infrastructures, along with the platforms developed in the Human Brain Project (High Performance Analytics and Computing Platform and Neuromorphic Computing Platform), can play an important role in the construction of a European HLP platform.

Also, regulation of the use of such data should be made much more open, and core language resources (annotated corpora, lexicons and ontologies) should be standardised and shared in this open environment. The regulation about crawling data should be carefully updated. Currently, this area is a legal minefield, while making use of data sets collected from the web for the purpose of building language technologies is crucial for the development of these technologies.

5. **Bridge the technology gap between European languages**

In order to bridge the technology gap, policies should focus on fostering basic research, applied research and development, and innovation for language technologies for all official European and many additional languages. Research and technology transfer among languages, along with increased collaboration across linguistic communities, must receive more attention. Specific programmes within funding schemes, such as Horizon 2020 and the successor funding programme, should boost the knowledge and technology transfer between countries that already have excellent research and innovation hubs in language technology and those that do not; the goal would be to enable the less resourced countries to develop technologies for their respective languages.

## Industry policies

6. **Foster and support the development of investment instruments and accelerator programmes targeting HLT start-ups**

One of the biggest issues that companies must address when bringing their innovations to the market is the lack of private financing. While the research process is usually supported by public funding, both at national and European level, the following steps in the business lifecycle (development of the product, commercial activities, etc.) lack coordinated public-private support in terms of funding.

Currently, there are no specific accelerator programmes targeting HLT start-ups. The strategic relevance of HLT for reinforcing the European common identity and citizenship engagement, as well as the preservation of cultural heritage of less-resourced languages and the effective development of the digital single market, requires a specific approach of investments instruments for all stages (seed capital, early stage, expansion stage, etc.).

7. **Increase the availability of qualified personnel on HLT**

The digital economy in Europe is facing a big challenge: the shortage of technical professionals, which may endanger the economic growth expected through the intensive use of digital technologies in all productive sectors.

This fact is more evident in the HLT European sector, as it is at the forefront of research and innovation and must face the fierce competition of US players, not only in the market but also in the recruitment of qualified professionals. European policies on this topic should be oriented towards: (i) incentivising models to retain talent in Europe; (ii) analysing, together with research and industry, the current needs of LT-related education; and (iii) raising awareness among students about the career opportunities in the HLT industry.

## Market policies

8. **Raise awareness of the benefits for companies, public bodies, and citizens of the availability of on-line services, contents and products in multiple languages**

European society as a whole must be aware of the challenges of a multilingual Europe and how the fast improvement of HLT can effectively help to overcome language barriers. The main targets of this policy

are: (i) the public bodies, namely policy-makers and public officials; (ii) the firms, particularly SMEs that could benefit from access to a pan-European market thanks to an integrated DSM without language borders; and (iii) the citizens, who can benefit from multilingual public services and private offerings in a Europe without language barriers, regardless of the languages they speak. It is crucial to show European stakeholders the benefits that these technologies bring to their businesses and daily lives, to further increase the demand for HLT.

9. **Promote the automated translation of e-commerce websites of European SMEs**

Within European SMEs, retail trade is one of the most crucial sectors that could particularly benefit from increasing their market size by providing their web shops in multiple European languages. However, language barriers and technical issues are hindering European SME web merchants in selling to other countries. Therefore, the authors propose providing economic incentives to: (i) SMEs, to translate their web shops using HLT, and (ii) ICT integrators, to develop specific cloud-based services allowing a smooth integration of HLT in the e-commerce applications used by SMEs. European structural and investment funds for the period 2014-2020 for the theme 'Competitiveness of SMEs' could be used in this regard.

## Public service policies

10. **Public procurement of innovative technology and pre-commercial public procurement**

Public procurement of innovation (PPI) has been recognised by the European Union as an important tool to foster innovation, reducing time to market new solutions, improving the quality of public services and increasing the competitiveness of European enterprises, particularly SMEs. Pre-commercial procurement (PCP) is an R&D tool that applies when innovative goods and services are not yet available in the market. This procurement process implies that the buyer (the public administration) and the supplier share risks and benefits under market conditions.

The application of PPI and PCP when implementing multilingualism across public services, contents and products in the EU in ways such as intelligent assistants for migrants, translation services in justice and health, multilingual e-government services, and legal documents available in all European languages, can become a strategic step in fostering the development and growth of the HLT industry in Europe.

11. **Foster the translation of national and regional public websites and documents into other EU languages by using HLT**

To create a truly integrated Europe, it is important that EU citizens and businesses can access relevant public information and use public online services of any of the Member States in their own language. Public digital services throughout Europe will never be truly effective, efficient and inclusive in the current context of language fragmentation. Public administrations at all levels – local, regional, national and European – should be encouraged to use HLT and provide information and services in all European languages. Additionally, the availability of public services around Europe in all languages would contribute to increasing citizens' engagement and to further improving trust in public bodies.

Having legal support to promote the translation of public websites into all of the European languages could speed up the deployment of this policy. It could be ensured by establishing a legal mandate or recommendations to national and regional governments to provide their web pages in several languages when receiving structural funds. As a first goal, the authors recommend that certain relevant public services in all countries be available in, at the very least, all of the official EU languages. European structural and investment funds for the period 2014-2020 for the theme 'Efficient Public Administration' could be used in this regard.