



THE FORUM FOR EUROPE'S LANGUAGE TECHNOLOGY INDUSTRY

March 2012

POSITION PAPER: THE STATE AND EVOLUTION OF THE LANGUAGE INDUSTRIES TO 2020 AND THEIR ROLE IN "HORIZON 2020"

Language Technologies for Innovation and Inclusion – a Window of Opportunity

Europe has a unique opportunity to build upon its expertise and strength in language technologies.

Europe is characterised by 62 languages, the EU has 23 official working languages. The EU is host to a unique mix of world and less-favoured languages. Many "small" European languages are under-represented online.

A wide variety of languages from other parts of the world are spoken by immigrant communities in EU countries – among the most frequent are Arabic, Berber, Chinese, Russian, Turkish, Balkan languages and languages from the Indian sub-continent. The role of English as lingua franca in the digital world is constantly decreasing – it now represents less than 30% of the total web content. English accounts on average for no more than 25% of the communications flowing through global channels like Facebook or Twitter. New regions and economies are progressing fast (the fastest growing languages are Chinese, Brazilian Portuguese and Arabic). Eventually, Internet penetration and online languages will just follow GDP evolution.

60% of the 30 top translation companies are based in Europe, but as in other market segments consolidation will favour large international groups, with headquarters across the Atlantic. The worldwide market for translation products and services accounts for 15 billion €, Europe having around 50% share in this market. The worldwide market for speech products and services is also estimated at around 15 billion €. The analytics and business intelligence software market is estimated at around 8 billion € worldwide. Worldwide demand for language services continues to grow much faster than the general economy, estimated at 10-13% per year in recent years.

Language Technologies enable people to access and make sense of online content and services in Europe's many languages. The unprecedented demand for information - for businesses, consumers and citizens - is opening and establishing new market opportunities for European language technology developers and suppliers. Managing, searching for and accessing information on the Web is only possible through exploiting the language-encoded information embedded in digital objects, using language technology.

The language industry is at the leading edge of technological development of on-line and communications services. The European language industry has a solid scientific and engineering base and the opportunity at this moment to harness its technological leadership, to challenge for global market leadership in internet-related application services. Most currently evolving features of this market, social networks, big data, need language technologies, which can bring considerable added value, both to consumers and to professional services on the internet.

This fast moving market will not wait. The window of opportunity is likely to be short. Key language technology players will need to move quickly in this market in the short-term to exploit their leading positions. To aid industry, European public policies need to provide a clear and stable strategy supporting openness and diversity. The EU's Multilingualism policy is an important foundation for such an approach. European programmes can usefully nurture new innovative ideas in the domain, to maximize the strength of the industry in the mid-term.

Language Technologies as Key Enabler for Industrial Technologies

The proposed Council Decision establishing the Specific Programme Implementing Horizon 2020 (Section II - Industrial Technologies of Annex 1 recognizes this challenge in Section 1.1.4, "Content Technologies and Information Management", which identifies language technologies as key enablers for innovation I¹. ***The language technology companies represented in LT-Innovate strongly support this position***, seeing it firmly anchored, as described in this position paper, in a true analysis of the current market situation.

While new challenges, such as big data and social networks, need new hardware architectures and the control software for these, they equally need enhanced tools for access, analysis and interpretation of the content of big data. Indeed, there is great scope for innovation into many new applications that will use content management support, such as language technologies. Language technologies emerge in this scenario as a key strategic enabler for innovation in new Internet applications.

Language and Related Technological Tools can help addressing Societal Challenges by removing obstacles to inclusion

Language technologies are also key to social aspects of our Internet-based society. They are an essential component of the practical toolset needed to ensure that internet society is truly inclusive. Many characteristics of the internet already allow this. Universal open standards, ever-increasing interconnection, powerful user-friendly search tools, development programmes to make computing and communications devices available in emerging and developing countries, all take us closer to a world in which every individual on our Earth has physical access and the real ability to use the internet.

Globally, many of the obstacles to universal access are gradually being removed. Nevertheless, in all our societies, even within Europe itself, language is still a barrier. Basic IT support for various script forms (e.g. Cyrillic, Chinese, Arabic) is today available. But making all tools available for use in all languages is progressing gradually. On-line translation is seldom available and not sufficiently sophisticated and reliable to offer useful machine-aided support to interaction between individuals across languages.

For this reason, LT-Innovate believes that Section III (Societal Challenges) of Annex I of the proposed Council Decision establishing the Specific Programme Implementing Horizon 2020 would be an appropriate additional context for the role of language, based upon the EU's multilingualism policy. ***We recommend that the following texts in paragraphs 6.1 (new text in bold italics) are amended accordingly²***;

"It will develop tools for a better assessment of the contextual and mutual impacts of such evolutions and policy options in areas such as employment, taxation, inequalities, poverty, social inclusion, education and skills, ***multilingualism***, community development, competitiveness and the Internal Market." (para 6.1.1).

"Research will analyse how societies and politics become more European in a broad sense through evolutions of identities, cultures and values, ***the emergence and acceptance of multilingualism***, the circulation of ideas and beliefs and combinations of principles and practices of reciprocity, commonality and equality." (para 6.1.2).

"This requires a greater understanding of the history, cultures, ***languages*** and political-economic systems of other world regions, as well as of the role and influence of transnational actors." (para 6.1.3).

¹ See COM (2011)811final, page 39, section 1.1.4. Content technologies and information management: ICT for digital content and creativity:

"The objective is to provide professionals and citizens with new tools to create, exploit and preserve all forms of digital content in any language and to model, analyse, and visualise vast amounts of data, including linked data. This includes new technologies for language, learning, interaction, digital preservation, content access and analytics; intelligent information management systems based on advanced data mining, machine learning, statistical analysis and visual computing technologies".

² See COM (2011)811final, pages 76-78.

ANNEX – Key Facts on the State and Potential of the Language Industries

The cost of language services (translation and interpretation) to the European Institutions is nearly 1.1B€ per year. The European Commission has the largest translation service in the world, yet does not fully exploit the potential of (multi-)language technologies. DG Translation decided only recently to launch a state-of-the-art machine translation programme (MT@EC) that will progressively build machine translation engines for all EU languages to serve internal and external clients. MT@EC is based on a statistical machine translation toolkit codenamed MOSES, and developed within one of the projects supported by DG INFSO (Euromatrix project).

Over half of all European staffing costs are now allocated to “information work” with employees suffering from both information overload (too much information to be digested productively) and “underload” (not able to find the right information); 80% of time spent using information involves locating it, with only 20% working on it.

If you offer services in English alone, you stand to lose up to two thirds of potential online sales. Only a third of the European buyers of online goods or services are prepared to do so in a language other than their own, according to a Eurobarometer study. Research suggests that nearly one million European SMEs may be losing trade every year as a result of lack of language competence and resources.

Due to the cost and complexity of the process, very few European online shopping sites are localised. According to a study 82% of European online shops are published in a single language; 11% offer two languages, and only 2% publish in five or more languages. None could be considered fully multilingual.

Successful business globalisation/internationalisation requires translation to be embedded in a “global content management” process that extends from authoring through to publishing. Language technology plays a particularly significant role in this market because automation benefits buyers; yet these savings are often available only to large companies with appropriate resources at their disposal.

The high costs of multilingualism can be cut by more autonomous and high-quality machine translation systems for all languages. Language technologies provide solutions that help to win new markets by being multilingual e.g. opinion mining allows companies to find out rapidly what people think of their products, and take quick corrective action. News mining and business intelligence help companies to find new market opportunities, learn about their competitors. All the information is already there – the winner is the one who finds it and adapts to it before all the others. Multilingual content management helps to keep the entire lifecycle of information under control even if you have to replicate the information in 50 languages.

Localisation is a driving factor for cross-border e-commerce and the integrated digital market. And it applies to websites as well as to media (e.g. games) and audio-visual products.

Language Technologies enable people to access and make sense of online content and services in Europe's many languages. The unprecedented demand for information - for businesses, consumers and citizens - is opening new market opportunities for European LT developers and suppliers. Managing, searching for and accessing information on the Web is only possible through exploiting the language-encoded information embedded in digital objects, using language technology.

As smartphones increasingly dominate telephony markets (expected 76% penetration by 2015), there are significant opportunities for service providers who evolve their value-added services in these markets to include features such as advanced speech recognition, natural language understanding and voice-to-text technologies.

Businesses leverage social media to create brand communities and crowdsourcing models, gain consumer insights, enhance product/brand awareness, improve search engine optimisation, reduce the cost of customer acquisition/ service/maintenance, and to optimise overall marketing communication. Managing – or even tracking – relationships with customers across language barriers is a significant marketing challenge, again one uniquely addressable by language technologies.

New modes of interactivity demand new and better interfaces, and both speech and dialogue modelling technologies will be scaled up for application in emerging markets for interactive systems; addressing interactivity in all EU languages is a particularly acute need.